CONTRIBUTIONS
TO
ECONOMIC ANALYSIS

110

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

PRODUCTION ECONOMICS: A DUAL APPROACH TO THEORY AND APPLICATIONS

Volume 1
The Theory of Production

Editors:
MELVYN FUSS and DANIEL McFADDEN

1978

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

ISBN North-Holland for this series: 0 7204 3100 X
ISBN North-Holland for this volume: 0 444 85012 0

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

Sole distributors for the U.S.A. and Canada:

ELSEVIER NORTH-HOLLAND INC.
52 VANDERBILT AVENUE
NEW YORK, N.Y. 10017

Library of Congress Cataloging in Publication Data
Main entry under title:

Production economics.

(Contributions to economic analysis ; 110-111)
Includes bibliographies and indexes.
CONTENTS: v. 1. The theory of production.--v. 2. Applications of the theory of production.
1. Production (Economic theory) I. Fuss, Melvyn A. II. McFadden, Daniel. III. Series.
HB241.P74 338'.001 78-19128
ISBN 0-444-85014-7 (set)
Vol 1; 0 444 85012 0
Vol 2; 0 444 85013 9

PRINTED IN THE NETHERLANDS

INTRODUCTION TO THE SERIES

This series consists of a number of hitherto unpublished studies, which are introduced by the editors in the belief that they represent fresh contributions to economic science.

The term 'economic analysis' as used in the title of the series has been adopted because it covers both the activities of the theoretical economist and the research worker.

Although the analytical methods used by the various contributors are not the same, they are nevertheless conditioned by the common origin of their studies, namely theoretical problems encountered in practical research. Since for this reason, business cycle research and national accounting, research work on behalf of economic policy, and problems of planning are the main source of the subjects dealt with, they necessarily determine the manner of approach adopted by the authors. Their methods tend to be 'practical' in the sense of not being too far remote from application to actual economic conditions. In addition they are quantitative rather than qualitative.

It is the hope of the editors that the publication of these studies will help to stimulate the exchange of scientific information and to reinforce international cooperation in the field of economics.

The Editors

PREFACE

The traditional starting point of production theory is a set of physical technological possibilities, often described by a production or transformation function. The development of the theory then parallels the process of firm operation, with the firm seeking to achieve its goals subject to the limitation of its technology and of the economic environment. The results are *constructed* input demands and output supplies, expressed as functions of the technology and the economic environment.

An alternative approach to production theory is to start directly from observed economic data – supplies, demands, prices, costs, and profits. The advantage of such an attack is that the theory can be formulated directly in terms of the causal *economic* relationships that are presumed to hold, without the intervening constructive steps required in the traditional theory. Because this approach is not bound by computational tractability in the step from production technology to economic observations, the prospect is opened for more satisfactory models of complex production problems.

It would at first appear that a theory of production couched in terms of economic observables would be less fundamental than one based on the physical technology, and that one could never be sure in an economic theory of consistency with a physical model. However, the theory of *production duality* establishes that the two approaches are equivalent and equally fundamental. Using duality, the technology underlying an economic model can be reconstructed and tested for compatibility with physical laws, as necessary. Then, the main thrust of analysis can be devoted to developing the structure and relationships of observed economic variables.

The purpose of these volumes is to develop the theory of production from the standpoint of the "dual" – the relationships between economic observables which are dual to the physical technology. The spirit of our treatment is the view that the end purpose of production theory is econometric study of economic problems involving technological limitations. The volumes emphasized the empirical implications of the theory, and therefore the development of the theoretical concepts proceeds with an eye towards the econometric framework inherent in empirical applications. We hold the view that there is an intimate, symbiotic relationship between theory and econometrics, and that development of a fully successful economic analysis of production requires an integration of theoretical and econometric ideas in a unified approach. The papers in the two volumes of *Production Economics* represent an attempt to achieve this ideal.

The theory of production duality had its beginnings in the work of Hotelling (1932), Hicks (1946), Roy (1942), and Samuelson (1947). A pioneering book by Shephard (1953) provided the first comprehensive treatment of the subject and proof of the basic duality of cost and production. Extensions of the formal theory of duality were later made by McFadden (1962), Uzawa (1964), Shephard (1970), and Diewert (1971). Many of the basic duality results were also obtained by Gorman (1970), working independently. In a paper on the estimation of returns to scale, Nerlove (1963) utilized a cost function to derive econometric estimating equations. Subsequent work by McFadden (1964), Diewert (1969a,b), Christensen, Jorgenson and Lau (1971), and others have established the use of dual cost and profit functions as a basic tool in econometric production analysis.

It is possible to trace the origins of the present volumes back to 1961 when D. McFadden worked as a research assistant to M. Nerlove and H. Uzawa at Stanford University. The contributions of Uzawa (1962, 1964), McFadden (1962, 1963), and Nerlove (1963) date from that period. The empirical implications of duality theory were developed in McFadden (1964 and 1966). The first explicit empirical application of dual flexible functional forms appeared in Diewert's (1969a) study of labor demand functions for the Canadian Department of Manpower and Immigration. The generalized Leontief function [Diewert (1971)] was introduced in that study. The subsequent generation and empirical application of flexible functional forms received their major impetus from McFadden (1966) and Diewert (1969a,b).

Applications of the basic duality concepts continued to evolve at the University of California, Berkeley, during the years 1968–1970 under the auspices of the Project for the Optimization and Evaluation of Economic Growth. The introduction of the translog function by Christensen, Jorgenson, and Lau (1971, 1973), the nested generalized Leontief form by Fuss (1970, 1977b), the hybrid generalized Leontief form by Hall (1973), and the generalized CES form by Denny (1974a) all result from research begun at that time. A. Belinfante, T. Cowing, and P. Frenger also were associated with the Economic Growth Project at various times. M. Bruno was a visiting scholar at M.I.T., together with D. McFadden, in 1971 when his chapter was written.

The idea of collecting a group of studies in duality under a common cover grew out of a seminar series held at the Economic Growth Project during the summer of 1969. A tentative title, *An Econometric Approach to Production Theory*, was chosen at that time. A number of the papers which appear in this volume have been referenced under that title. Since that time, the contents of the volumes evolved through several additions and deletions and M. Fuss joined D. McFadden as a co-editor. We feel that the current title more accurately reflects the spirit and content of the books.

Production Economics is divided into two main parts. Volume 1 contains the basic theoretical analysis of the duality of cost, profit, and production and a number of investigations of specific functional forms. Volume 2 contains the empirical applications. In keeping with the spirit of this work, these applications draw heavily on the analysis of Volume 1. Details of the contents of both volumes can be found in the two introductions.

The editors have been unable to standardize notation throughout the volumes; however, the notation in each chapter is self-contained. In almost all cases, upper case boldface letters denote sets, lower case boldface letters denote vectors. Upper and lower case Roman and Greek letters are used variously to denote scalars and functions. Derivatives are denoted variously by subscripts (the symbol for the variable with respect to which derivatives are being taken, or the ordinal position of this variable among the arguments), primes, the ∇ operator, or the usual notation $\partial f/\partial x$.

The editors wish to acknowledge the contributions that many individuals have made to the preparation of *Production Economics*. Dale Jorgenson and Zvi Griliches have provided encouragement and ideas. A large intellectual debt is owed to K.J. Arrow, W.M. Gorman, L. Hurwicz, M. Nerlove, and H. Uzawa, whose work provided the background for most of the developments in these books. We thank the contributors, who have displayed stoic patience and goodwill in the lengthy process of refereeing and publication. We also wish to acknowledge the help of several scholars who participated in the early planning, and who have published related work elsewhere: T. Cowing (1974), W. E. Diewert (1971, 1974a), R.E. Hall (1973), C.K. Liew (1976), and M. Ohta (1975).

To G. Katagiri and N. Katagiri goes the credit for careful typing and editing of the manuscript.

The editors accept responsibility for all errors not allocatable to individual contributors. Finally, we thank our wives, Beverlee and Susan, for tolerance and encouragement through the lengthy process of bringing these volumes to completion.

CONTENTS OF VOLUME 1

Introduction to the Series v
Preface vii
Introduction xix

Part I. Duality of Production, Cost, and Profit Functions 1

I.1 Daniel McFadden: Cost, Revenue, and Profit Functions 3

1. Introduction 3

Part I. Cost Functions 5

2. History 5
3. Production technologies 5
4. The cost function 10
5. The derivative property 14
6. Duality 19
7. Distance functions and economic transformation functions 24
8. Extensions of duality 29
9. Cobb–Douglas and C.E.S. cost functions 38
10. The geometry of two-input cost functions 39
11. Comparative statics for the cost minimizing firm 46
12. Composition of distance and cost functions 49

Part II. Restricted Profit Functions 60

13. The general representation of production possibilities 60
14. The general restricted profit function 66
15. The derivative property of the restricted profit function 74
16. The gauge function for production possibilities 76
17. Duality for the restricted profit function 81
18. Examples of restricted profit functions 93
19. Composition rules for profit functions 95
20. Profit saddle-functions 101

I.2 Giora Hanoch: Symmetric Duality and Polar Production Functions 111

1. Introduction 111
2. A symmetric formulation of cost and production functions 112
3. Polar production and cost functions 118
4. A symmetric formulation of profit functions and production frontiers 121
5. The polar profit and production functions 126
6. Some extensions and an application 129

I.3 Lawrence J. Lau: Applications of Profit Functions 133

1. The profit function – An alternative derivation 133
2. The structure of normalized profit functions 151
3. Extensions to multiple outputs 169
4. Examples of normalized profit functions 190
5. Applications of the normalized profit function 197
6. Summary and conclusions 215

Part II. Functional Forms in Production Theory 217

II.1 Melvyn Fuss, Daniel McFadden, and Yair Mundlak: A Survey of Functional Forms in the Economic Analysis of Production 219

1. The context and objectives of production analysis 219
2. Criteria for the design of functional forms 222
3. Dual transformation, cost and profit functions – Maintained hypotheses on the technology and its representations 225
4. A general approach – Forms linear-in-parameters 230
5. Special non-linear forms 240
6. Separability: Functional implications and tests 244
7. Econometric estimation of production parameters 249
8. Overview of empirical analysis 267
9. Conclusion 268

II.2 Daniel McFadden: The General Linear Profit Function 269

1. Introduction 269
2. The basic model 270
3. General linear profit functions 272
4. The dual technology of the general linear profit function 281
5. Applications of the linear profit function 283

II.3 Giora Hanoch: Polar Functions with Constant Two Factors – One Price Elasticities 287

1. Introduction 287
2. Elasticities of substitution and transformation 289
3. A summary of DRES and CDE (implicitly additive) models 294
4. Multiproduct production frontiers with constant TOES or TOET 301

II.4 Melvyn Fuss and Daniel McFadden: Flexibility versus Efficiency in *Ex Ante* Plant Design 311

1. Introduction 311
2. Historical background 313
3. A model of the firm with an *ex ante–ex posi* technology 320
4. Functional forms for the *ex ante–ex post* production structures 330
5. Derivation of the two-level structure of technology 345
6. Separable technology across states 356
7. A general linear-in-parameters *ex ante–ex post* technology 359
8. Concluding remarks 363

Appendices 365

A.1 Daniel McFadden: Definite Quadratic Forms Subject to Constraints 365

1. Conditions for a matrix to be positive definite 365
2. Conditions for a matrix to be positive definite subject to constraint 368

A.2 Daniel McFadden: Necessary and Sufficient Conditions for the Classical Programming Problem 375

A.3 Daniel McFadden: Convex Analysis 383

1. Introduction 383
2. Notation 383
3. Hyperplanes 384
4. Convex sets 384
5. Affine subspaces 384
6. Separation of sets 384
7. Cones 384
8. Polar and normal cones 385
9. Convex functions 385
10. Properties of convex sets 385
11. Semi-bounded sets 387
12. Properties of convex functions 390
13. Properties of maximand correspondences 392
14. Exposed sets 396
15. Conjugate correspondences 397
16. Differential properties 401

A.4 Lawrence J. Lau: Testing and Imposing Monotonicity, Convexity and Quasi-Convexity Constraints 409

1. Introduction 409
2. Hessian matrices of convex and quasi-convex functions 414
3. The Cholesky factorizability of semidefinite and indefinite matrices 421
4. Estimation 437
5. Testing of hypotheses 447
6. Conclusion 452

Bibliography 455

Author Index 471

Subject Index 475

CONTENTS OF VOLUME 2

Introduction to the Series v
Preface vii
Introduction xix

Part III. An Analysis of the Concept of Real Value-Added 1

III.1 Michael Bruno: Duality, Intermediate Inputs and Value-Added 3

1. Introduction 3
2. Duality and value-added 4
3. Value-added in constant prices 7
4. The measurement of total productivity 11
5. Introducing imperfect competition 15
6. Conclusions 16

III.2 W. Erwin Diewert: Hicks' Aggregation Theorem and the Existence of a Real Value-Added Function 17

1. Introduction 17
2. Hicks' aggregation theorem in the consumer context 22
3. Duality between direct and conditional indirect utility functions 24
4. Hicks' aggregation theorem in the producer context 34
5. The duality between value-added "production" and cost functions 37
6. Concluding remarks 40
7. Proofs of theorems 43

III.3 Michael Denny and J. Douglas May: Homotheticity and Real Value-Added in Canadian Manufacturing 53

1. Introduction 53
2. Real value-added and the production structure 54

3. The approximate translog cost function 56
4. Homotheticity and homogeneity 58
5. The Canadian manufacturing data 59
6. The estimation of the production structure 61
7. An alternative real value-added formulation 63
8. Weak homotheticity 65
9. Double deflation and real value-added 68
10. Conclusion 70

Part IV. Empirical Applications of Production Theory: Micro-Economic Data 71

IV.1 Daniel McFadden: Estimation Techniques for the Elasticity of Substitution and Other Production Parameters 73

1. Introduction 73
2. Cost functions 75
3. The elasticity of substitution 78
4. The specification of the ACMS model 80
5. Properties of the elasticity of substitution 82
6. An econometric model of the electricity generation industry 95
7. Appendix 118

IV.2 Peter Diamond, Daniel McFadden, and Miguel Rodriguez: Measurement of the Elasticity of Factor Substitution and Bias of Technical Change 125

1. Introduction 125
2. Notation 127
3. Growth equations 129
4. Production relations 131
5. A non-identification theorem 132
6. Necessary conditions for factor augmenting technical change 135
7. Sufficient conditions for factor augmenting technical change 137
8. Identification for capital augmentation 141

9. Identification for finite parameter families of augmentation functions 143
10. Extensions and conclusions 145
11. Appendix: Generalization to non-constant returns 145

IV.3 Alexander Belinfante: The Identification of Technical Change in the Electricity Generating Industry 149

1. Introduction 149
2. Notation 151
3. Identification problems in the measurement of technical change 152
4. The industry and the data 157
5. The measurement of disembodied technical change 162
6. The measurement of embodied technical change 172
7. The measurement of total technical change 177
8. The bias of technical change 181
9. Concluding remarks 185

IV.4 Melvyn Fuss: Factor Substitution in Electricity Generation: A Test of the Putty-Clay Hypothesis 187

1. Introduction 187
2. The basic model 189
3. The model applied to electricity generation 190
4. Tests of the structure of technology 200
5. Estimation of the putty-clay model 208
6. Conclusion 210
7. Appendices 210

IV.5 Thomas G. Cowing: The Effectiveness of Rate-of-Return Regulation: An Empirical Test Using Profit Functions 215

1. Introduction 215
2. The Averch–Johnson model of a regulated firm 217
3. The profit function for a regulated firm 222
4. Statistical tests of regulatory effectiveness 225
5. Comparison with previous studies 237
6. Conclusions 243
7. Appendix 245

Part V. Empirical Applications of Production Theory: Macro-Economic Data 247

V.1 Michael Denny and Cheryl Pinto: An Aggregate Model with Multi-Product Technologies 249

1. Introduction 249
2. Multi-product technologies and final demand 250
3. Jointness and separability of the cost function 253
4. The approximate translog cost function 253
5. Approximate separability and non-jointness of the translog cost function 255
6. An aggregate Canadian production model 258
7. Estimation and hypothesis testing 260
8. Appendix: Data 266

V.2 Petter Frenger: Factor Substitution in the Interindustry Model and the Use of Inconsistent Aggregation 269

1. Introduction 269
2. The model 271
3. The data 273
4. Estimation 275
5. Empirical results 277
6. 1961 predictions 283
7. Inconsistent shadow elasticities of substitution 287
8. A directional shadow elasticity of substitution 291
9. Inconsistent directional shadow elasticity of substitution 296
10. Appendix 300

Bibliography 311

Author Index 327

Subject Index 331

INTRODUCTION

Volume 1 develops the theory of production from the standpoint of economic observables – prices, demands, supplies, cost, and profit – utilizing duality to relate this approach to the underlying production technology. The papers of Part I set out models of production and basic duality theorems, and discuss theoretical applications of these models.

In Chapter I.1, McFadden provides an introduction to cost, revenue, and profit functions. The first twelve sections of his chapter provide a detailed description of properties of production and cost functions, duality, the geometry of cost functions, and the comparative statics of the firm using cost functions. The remainder of the chapter introduces the concept of the restricted profit function – of which cost, revenue, and total profit functions are special cases – and utilizes the mathematical theory of convex conjugate and polar reciprocal forms to deduce the properties implied for the restricted profit function by various properties on the technology, and vice versa. Of particular interest are Tables 1, 3, and 4 listing dual properties; Tables 2 and 5 listing composition rules for concave functions which can be used to construct functional forms or deduce theorems on production structure; and Tables 6 and 7 summarizing the duality mappings holding for restricted profit functions.

In Chapter I.2, Hanoch shows how formal duality theory can be used to generate new functional forms for cost and production functions. This chapter explores the use and implications of structural assumptions on technology, cost, and profit in the specification of functional forms.

In Chapter I.3, Lau applies the restricted profit function to a variety of theoretical production problems. Using the classical theory of Legendre transformations, he develops a convenient formal calculus for working with derivatives of dual production and profit functions. Lau establishes the implications for the profit function of various homotheticity and separability properties, and develops a number of specific functional forms. He considers the formulation in terms of the profit function of measures of the elasticity of technical substitution and rates of technical change. Finally, he explores the structure of production in multiple-output firms and its implications.

Part II concentrates on the development of functional forms for econometric analysis, and the interaction of functional and stochastic specification. In Chapter II.1, Fuss, McFadden, and Mundlak set out the criteria that might be used to choose among functional forms, and use these criteria to compare many of the econometric forms appearing in the literature. The issue of stochastic specification is surveyed in the context of an extended example.

In Chapter II.2, McFadden outlines a general procedure for generating linear-in-parameters functional forms, and establishes conditions under which an arbitrary restricted profit function can be approximated to the second order by a specified approximating form.

In Chapter II.3, Hanoch applies the concepts of symmetric duality and polar production functions to develop specific functional forms for the study of substitutability in multiple-factor production functions.

In Chapter II.4, Fuss and McFadden develop a nested generalized Leontief functional form for the econometric representation of an *ex ante–ex post* production structure, and suggest methods for the analysis of technological flexibility within this structure.

This volume has a series of mathematical appendices which develop some of the concepts and tools used. Appendix A.1 gives a self-contained treatment of the theory of definite quadratic forms subject to constraint. Appendix A.2 surveys necessary and sufficient conditions for the use of classical Lagrangian methods for constrained optimization. The third appendix is a survey of convex analysis – the mathematical theory of convex sets and functions. In addition to outlining the standard theory, this appendix develops new results on the behavior of polar reciprocal convex correspondences. Appendix A.4 develops methods for imposing or testing concavity on a fitted production or cost function.

Part I

Duality of Production, Cost, and Profit Functions

Chapter I.1

COST, REVENUE, AND PROFIT FUNCTIONS

DANIEL McFADDEN*

University of California, Berkeley

1. Introduction

In the classical theory of cost and production, the firm is assumed to face fixed technological possibilities and competitive input markets, and to choose an input bundle to minimize the cost of producing each possible output. For fixed input prices, this behavior determines minimum cost as a function of output, yielding the standard cost curves of elementary textbooks. An immediate generalization is to allow input prices to vary and consider minimum cost as a function of both input prices and output. With this minor modification, the cost function becomes a powerful analytic tool in the theory of production, particularly in econometric applications.

The principal practical advantage of the cost function lies in its computationally simple relation to the cost minimizing input demand functions: the partial derivatives of the cost function with respect to input prices yield the corresponding input demand functions, and the sum of the values of the input demands equals cost. The useful analytic properties of the cost function derive from a fundamental duality

*This research was supported in part by the National Science Foundation Grants GS-2345, GS-35890X, SOC-05551A02, and SOC75-22657. I am indebted to M. Nerlove, H. Uzawa, and S. Winter, who stimulated this work; to S. Cosslett, G. Debreu, E. Diewert, D. Gale, and D. Jorgenson for many useful discussions; and to G. Katagiri, N. Katagiri, L. Chichilnisky, and C. Gilchrist for editorial assistance. I retain sole responsibility for errors.

between this function and the underlying production possibilities. The definition of the cost function as the result of an optimization yields strong mathematical properties, and establishes the cost function as a "sufficient statistic" for all the economically relevant characteristics of the underlying technology.

In econometric applications, use of the cost function as the starting point for developing models avoids the difficulty of deriving demand systems constructively from production possibilities, while at the same time insuring consistency with the hypothesis of competitive cost minimization. Further, under a number of econometric specifications of firm behavior, the cost function and its derivatives define the reduced form of the model.

Properties of the cost function can also be used to generalize and simplify the qualitative implications of cost minimization. In particular, a number of comparative statics results can be derived without assuming divisibility of commodities, or convexity and smoothness of production possibilities.

Two concepts are closely related to the cost function and also are useful in theory and applications. One is the revenue function of a multiple-product firm facing competitive markets, defining maximum revenue as a function of output prices and inputs. The second is the profit function of a firm facing competitive markets for inputs and outputs, defining maximum profits as a function of input and output prices. Cost, revenue, and profit functions can all be considered as special cases of a restricted profit function, defining maximum profits over a subset of inputs and outputs with competitive prices when quantities of the remaining inputs and outputs are fixed.

This chapter can be divided into two parts. The first part, consisting of Sections 2 to 12, is a self-contained treatment of the theory of cost functions and its applications. Mathematical rigor and generality are deemphasized for pedagogic simplicity, and economic interpretations are stressed. These sections will be accessible to readers with modest technical backgrounds. Proofs of more difficult results are postponed. The second part, consisting of Sections 13 to 20, gives a formal analysis of the properties of restricted profit functions for the more technical reader. Examples of restricted profit functions are discussed in Section 18. Appendix A.3 gives a self-contained survey of properties of convex sets and functions used in this chapter.

PART I. COST FUNCTIONS

2. History

The cost curve is a classical concept in economics, antedating even the concept of a production function. However, the systematic analysis of the properties of price derivatives of the cost function seems to have originated in a paper of Hotelling (1932) on the mathematically equivalent problem of minimizing consumer expenditure subject to a utility level constraint. The cost function and its properties were discussed in Samuelson (1947), and later led Samuelson to develop the concept of a factor–price frontier (which is a level curve of a cost function).

The properties of consumer expenditure functions were developed further by Roy (1942) and McKenzie (1957). McKenzie seems to have first noted that the properties of expenditure functions can be obtained as a consequence of optimization using the mathematical theory of convex functions with much weaker assumptions than were employed by the earlier authors.

The theory establishing the dual relation between cost functions and production functions was introduced into economics by Shephard (1953), who drew heavily on properties of convex sets discovered by Fenchel (1953). Additional contributions to economic applications of duality theory have been made by Uzawa (1964), McFadden (1962), Diewert (1974a), Hanoch (1975a), and Lau (1976a).

Perhaps because the theoretical results on cost functions were scattered and relatively inaccessible, their potential worth in econometric analysis was not recognized until Nerlove (1963) employed the Cobb–Douglas case in a study of returns to scale in electric utilities. Since the mid-1960s, a series of empirical studies, including papers by Diewert (1969a), and Jorgenson and Lau (1974a), have made systematic use of duality concepts.

3. Production Technologies

Basic to a model of the firm are descriptions of the commodities with which it deals and the technological limits on its actions. Following Debreu (1959), the concept of a commodity is taken generally to include both physical goods, such as wheat and fuel, and services, such as transportation and labor. Further, commodities are distinguished by location and date; e.g., trucks delivered at different locations and/or in different months will be considered distinct commodities. In particular, dated commodities extend over the planning horizon of the firm, and static and intertemporal theories of the producer are formally equivalent.

Occasionally, the same good will appear on both the input and output ledgers of a firm. If inputs are delivered temporally prior to outputs, these quantities are properly recorded as distinct commodities. However, if the ledgers of the firm also record intermediate goods in the production process (and this is particularly likely to be true if "intertemporal decentralization of accounts" is imposed on a firm having a lengthy production process), the same good may appear simultaneously as an input and an output. In this case, it is sometimes adequate for an economic problem to record net output. For other problems, it is convenient to treat the input and output as separate commodities in the firm's accounts. In the following analysis, we shall treat inputs and outputs as distinct commodities, making the artificial accounting distinction above necessary.

We consider a firm which uses N inputs indexed $n = 1,2,...,N$, to produce M outputs, indexed $m = 1,2,...,M$. An input bundle is an N-tuple of non-negative real numbers, $\mathbf{v} = (v_1,...,v_N)$, as is an input price vector $\mathbf{r} = (r_1,...,r_N)$. An output bundle is an M-tuple of non-negative real numbers, $\mathbf{y} = (y_1,...,y_M)$. The cost of an input bundle $\mathbf{v}$ at an input price vector $\mathbf{r}$ is given by the inner product of $\mathbf{v}$ and $\mathbf{r}$, $c = \mathbf{r}\cdot\mathbf{v} = r_1v_1 + r_2v_2 + \cdots + r_Nv_N$.

The technological limits on the actions of the firm can be described by the set $\mathbf{Y}$ of pairs of input and output bundles $(\mathbf{v},\mathbf{y})$ which are possible, in the sense that the firm can deliver the prescribed output bundle $\mathbf{y}$ by using the input bundle $\mathbf{v}$; $\mathbf{Y}$ is termed the *production possibility set* of the firm. For example, a Cobb–Douglas production function $y_1 = v_1^{1/2}v_2^{1/2}$ corresponds to a production possibility set with one output and two inputs, $\mathbf{Y} = \{(v_1,v_2,y_1) | v_1,v_2 \geqq 0 \& v_1^{1/2}v_2^{1/2} = y_1\}$.

The production possibility set of a firm is determined first by the state of technological knowledge and physical laws. For example, the outputs of chemical refining processes are limited by chemical laws and the current knowledge of chemical engineers. There may be further limitations on the availability of techniques due to imperfect information and legal restrictions (e.g., patent agreements, pollution control regulations, safety standards). Non-transferable commodities, such as "managerial capacity", climate, and environmental factors, may also enter the determination of production possibilities. Finally, in most economic problems, the firm will be required to meet restrictions on some input and output quantities due to prior contracts, quotas, rationing, or "hardening" of commodities following ex ante design decisions. Common examples are commitments to fixed plant and equipment inputs, and contracts to purchase inputs (e.g., labor services) or supply outputs. It should be noted that "fixed" inputs or outputs can be either included or excluded from the commodity list facing the firm, depending on the economic problem. The sources of restrictions on the firm's production possibilities will be important in determining the economic interpretation of the cost function and its generalizations, but can be left undefined in the derivation of the formal properties of these functions.

With virtually no loss of economic generality, we usually assume that the production possibility set of a firm is non-empty and closed, and that a non-zero output bundle requires a non-zero input bundle. The condition that the production possibility set be closed requires that there be no "thresholds" at which discontinuities in required inputs or attainable outputs occur.[1] A production possibility set with these properties will be called *regular*.

In examining the cost function, it is convenient to work with "isoquants" rather than the production possibility set itself. First define the *producible output set* $\mathbf{Y}^*$ containing all the output bundles $\mathbf{y}$ which appear in some pair of input and output bundles in the production possibility set; i.e., $\mathbf{Y}^* = \{\mathbf{y} | (\mathbf{v},\mathbf{y}) \in \mathbf{Y} \text{ for some } \mathbf{v}\}$. Next, for each $\mathbf{y}$ in $\mathbf{Y}^*$, define the *input requirement* set $\mathbf{V}(\mathbf{y})$ containing all the input bundles $\mathbf{v}$ which can produce $\mathbf{y}$; i.e., $\mathbf{V}(\mathbf{y}) = \{\mathbf{v} | (\mathbf{v},\mathbf{y}) \in \mathbf{Y}\}$. The input requirement set corresponds to the conventional notion of an isoquant, except that it may include "inefficient" input bundles. Note that the input requirement set is well-defined in both the single-output and multiple-output cases. For the earlier example of the Cobb–Douglas production possibility set $\mathbf{Y} = \{(v_1,v_2,y_1) | v_1,v_2 \geqq 0 \& v_1^{1/2} v_2^{1/2} = y_1\}$, the producible output set is the non-negative real line and the input requirement sets are the isoquants $\mathbf{V}(y_1) = \{(v_1,v_2) | v_1,v_2 \geqq 0,\ v_1^{1/2} v_2^{1/2} = y_1\}$.

A production possibility set $\mathbf{Y}$ will be termed *input-regular* if (1) the

[1] A set is closed if it contains its boundaries; i.e., if the limit of each convergent sequence of points from the set is also contained in the set. Closedness does not rule out the possibility of lumpy (integer-valued) commodities. For example, the set $\mathbf{Y} = \{(v_1,y_1) | v_1 = 0,1,\dots \ \& \ v_1 \geqq y_1 \geqq 0\}$ is closed, as is the set $\mathbf{Y} = \{(v_1,y_1) | v_1 \geqq 0 \ \& \ [v_1] \geqq y_1 \geqq 0\}$, where $[v]$ denotes the largest integer less than or equal to v.

set of producible outputs $\mathbf{Y}^*$ is non-empty, and (2) for each $\mathbf{y}$ in the set of producible outputs, the input requirement set $\mathbf{V}(\mathbf{y})$ is closed, and for a non-zero output bundle does not contain the zero input bundle. Clearly, if a production possibility set is regular, then it is also input-regular.

In the conventional theory of the firm, marginal products of inputs are assumed to be non-negative, and marginal rates of substitution between inputs are assumed to be non-increasing. Stated in terms of the input requirement sets, these conditions become:

> *Assumption A.* There is free disposal of inputs; i.e., if an input bundle $\mathbf{v}$ can produce an output bundle $\mathbf{y}$, and a second input bundle $\mathbf{v}'$ is at least as large as $\mathbf{v}$ in every component, then $\mathbf{v}'$ can also produce $\mathbf{y}$.

> *Assumption B.* The input requirement sets are convex from below; i.e., if two input bundles $\mathbf{v}$ and $\mathbf{v}'$ are in an input requirement set $\mathbf{V}(\mathbf{y})$, then for any weighted combination of $\mathbf{v}$ and $\mathbf{v}'$, say $\mathbf{v}'' = \theta\mathbf{v} + (1-\theta)\mathbf{v}'$ with θ a scalar, $0 < \theta < 1$, there exists an input bundle $\mathbf{v}^*$ in the input requirement set such that $\mathbf{v}''$ is at least as large as $\mathbf{v}^*$ in every component.

In set notation, Assumption A is sometimes written $\mathbf{V}(\mathbf{y}) + \mathbf{E}_+^N \subseteq \mathbf{V}(\mathbf{y})$, where $\mathbf{E}_+^N$ is the non-negative orthant of the N-dimensional input commodity space, and the algebraic sum of sets is defined by $\mathbf{V}(\mathbf{y}) + \mathbf{E}_+^N = \{\mathbf{v} + \mathbf{v}' | \mathbf{v} \in \mathbf{V}(\mathbf{y})\ \&\ \mathbf{v}' \in \mathbf{E}_+^N\}$. Geometrically, $\mathbf{V}(\mathbf{y}) + \mathbf{E}_+^N$ is the set formed from $\mathbf{V}(\mathbf{y})$ by adding all points northeast of each point in $\mathbf{V}(\mathbf{y})$; $\mathbf{V}(\mathbf{y}) + \mathbf{E}_+^N$ is called the *free disposal hull* of $\mathbf{V}(\mathbf{y})$. The assumption is then that $\mathbf{V}(\mathbf{y})$ contains its free disposal hull. A set is said to be *convex* if it contains the line segment connecting any two of its elements. Assumption B can be restated as requiring that the free disposal hull $\mathbf{V}(\mathbf{y}) + \mathbf{E}_+^N$ be convex.

Justifications for these assumptions appear in most textbooks. Free disposal holds if firms can stockpile or refuse delivery of inputs, or if the technology is such that application of an additional unit of input always yields some non-negative amount of additional output and outputs can be disposed freely if necessary. Convexity from below holds if the technology is such that substitution of one input combination for a second, keeping output constant, results in a diminishing marginal reduction in the second input combination, or if production activities can be operated side by side (or sequentially) without interfering with each other. However, the importance of Assumptions A and B in traditional production analysis lies in their analytic convenience rather than in their economic realism; they provide the groundwork for application of calculus tools to the firm's cost minimization problem. One of the useful observations resulting from the analysis of cost functions is that the standard qualitative implications for supply and demand by the competitive firm can be obtained *without* imposing these conditions. Observed input demand functions for a cost minimizing firm facing positive input prices can be treated *as if* they come from input requirement sets satisfying Assumptions A and B even if these conditions fail to hold for the true technology.

Figure 1 illustrates Assumptions A and B. In (a), the input requirement set contains all the points northeast of any point in the set, thus satisfying Assumption A. In (b), the bundle **v** is in the set while the larger bundle **v**′

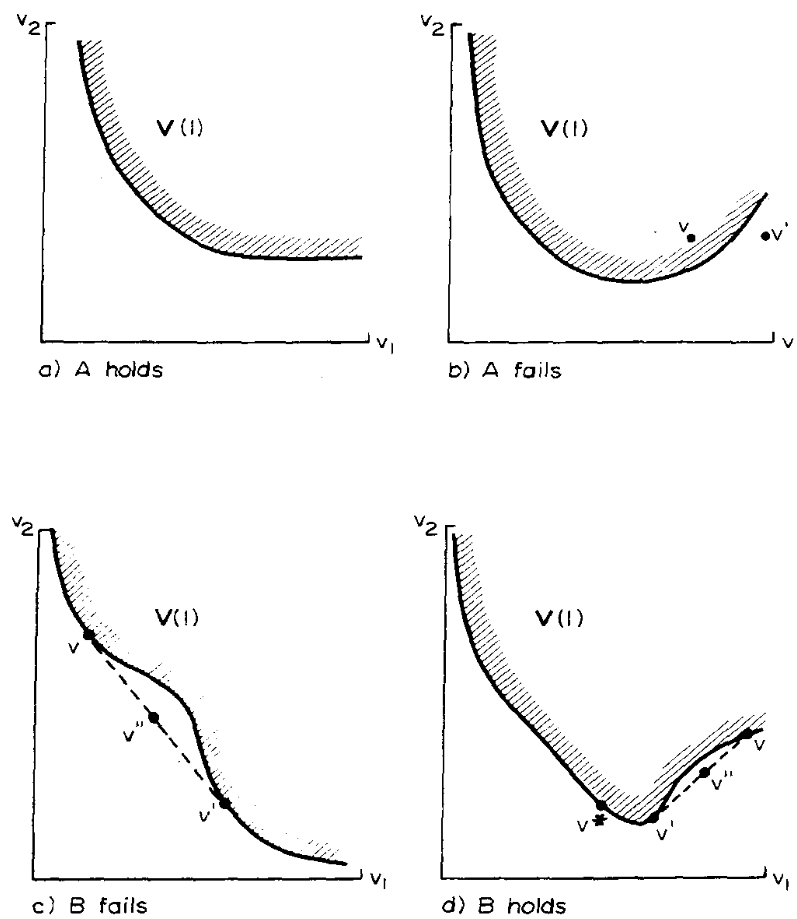


FIGURE 1

is not, and Assumption A fails. Assumption B fails in (c), where $\mathbf{v}''$ is an average of two points $\mathbf{v}$ and $\mathbf{v}'$ in the set, but is not itself northeast of any point in the set. In (d), on the other hand, Assumption B holds. Even though the weighted average $\mathbf{v}''$ of $\mathbf{v}$ and $\mathbf{v}'$ is not in the set, it lies northeast of $\mathbf{v}^*$ and the definition of convexity from below is satisfied.

A regular production possibility set satisfying Assumptions A and B will be termed *conventional.* Thus, in summary, a conventional production possibility set is non-empty and closed, with non-zero outputs requiring non-zero inputs, and has input requirement sets satisfying free disposal and convexity from below. An input-regular production possibility set satisfying Assumptions A and B will be termed *input-conventional.*

4. The Cost Function

Suppose that a firm has an input-regular production possibility set with a producible output set $\mathbf{Y}^*$ and input requirement sets $\mathbf{V}(\mathbf{y})$ for $\mathbf{y}$ in $\mathbf{Y}^*$. Suppose that the firm faces competitive input markets with strictly positive prices $\mathbf{r} = (r_1,\ldots,r_N)$, and chooses an input bundle $\mathbf{v}$ to minimize the cost $c = \mathbf{r}\cdot\mathbf{v} = r_1v_1 + \cdots + r_Nv_N$ of producing a given producible output bundle $\mathbf{y} = (y_1,\ldots,y_M)$. The *cost function* is then defined by

$$c = C(\mathbf{y},\mathbf{r}) = \mathrm{Min}\{\mathbf{r}\cdot\mathbf{v}|\mathbf{v} \in \mathbf{V}(\mathbf{y})\}, \tag{1}$$

and specifies the least cost of producing $\mathbf{y}$ with input prices $\mathbf{r}$.

We first verify that the cost function exists for all $\mathbf{y}$ in the producible output set and all strictly positive $\mathbf{r}$, using a mathematical theorem that a continuous function on a non-empty, closed, bounded set achieves a minimum in the set. The linear function $\mathbf{r}\cdot\mathbf{v}$ is continuous in $\mathbf{v}$. Since $\mathbf{V}(\mathbf{y})$ is non-empty, it contains at least one input bundle $\mathbf{v}'$, and the search for a minimizing bundle can be confined to the points in $\mathbf{V}(\mathbf{y})$ satisfying $\mathbf{r}\cdot\mathbf{v} \leqq \mathbf{r}\cdot\mathbf{v}'$. But this set is closed and bounded since $\mathbf{r}$ is strictly positive (see Figure 2), and the mathematical theorem above implies that $\mathbf{r}\cdot\mathbf{v}$ achieves a minimum on this set (at $\mathbf{v}''$ in the figure).

Since $\mathbf{v}$ and $\mathbf{r}$ are non-negative, the cost function is clearly non-negative. Further, if the output bundle $\mathbf{y}$ is non-zero, then every input bundle $\mathbf{v}$ which can produce $\mathbf{y}$ is non-zero. Since $\mathbf{r}$ is strictly positive, this implies that the cost function is strictly positive for non-zero output bundles.

We next show that for a fixed producible output bundle $\mathbf{y}$, the cost function is non-decreasing in input prices. Consider any strictly positive

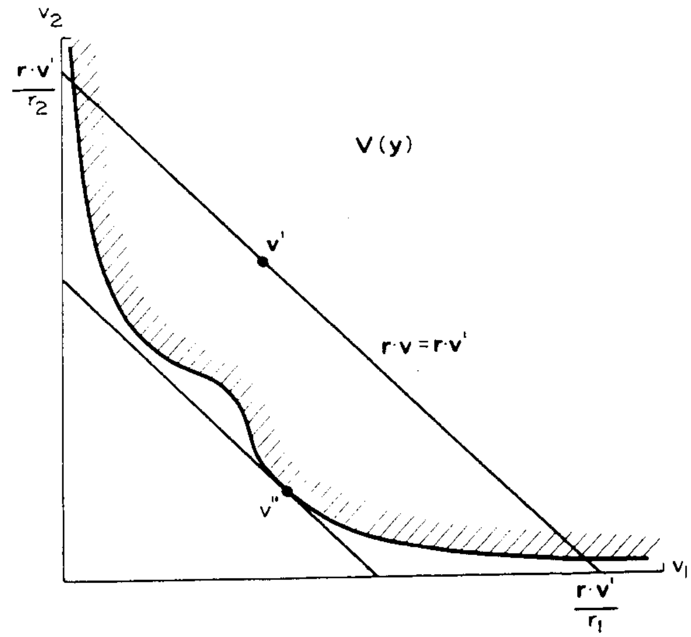


FIGURE 2

input price vector **r** and a second price vector **r**′ which is at least as large in every component. Suppose for the price vector **r**′, cost is minimized at some input bundle **v**′. Then, minimum cost at the input price vector **r** can be no higher than **r**·**v**′, which in turn can be no higher than **r**′·**v**′, which is the minimum cost at the input price vector **r**′.

Note that if an input bundle **v** is cost minimizing at a strictly positive price vector **r**, and if all prices are multiplied by a positive scalar θ, then **v** remains a cost minimizing bundle and the level of minimum cost is multiplied by θ. A function with this property is termed *positively linear homogeneous.*[2]

[2]A function $C(\mathbf{r})$ is said to be homogeneous of degree k in **r** if $C(\lambda \mathbf{r}) = \lambda^k C(\mathbf{r})$ for all $\lambda > 0$, and linear homogeneous if $k = 1$. If C is differentiable in r, then C is homogeneous of degree k if and only if $r_1(\partial C/\partial r_1) + \cdots + r_n(\partial C/\partial r_n) \equiv kC$ for all **r**. This is *Euler's law*. To demonstrate its validity, first differentiate the identity $C(\lambda \mathbf{r}) = \lambda^k C(\mathbf{r})$ with respect to λ, obtaining $r_1 C_1(\lambda \mathbf{r}) + \cdots + r_n C_n(\lambda \mathbf{r}) = k\lambda^{k-1}C(\mathbf{r})$, and set $\gamma = 1$. [We let $C_i(\mathbf{r}) = \partial C/\partial r_i$.] Second, evaluate the formula $r_1 C_1(\mathbf{r}) + \cdots + r_n C_n(\mathbf{r}) = kC(\mathbf{r})$ at $\lambda \mathbf{r}$ for a fixed vector **r**, obtaining $\lambda[r_1 C_1(\lambda r) + \cdots + r_n C_n(\lambda r)] = kC(\lambda r)$. Treating C as a function of λ, the term in brackets is just $dC/d\lambda$, and we have $(1/C)(dC/d\lambda) = (k/\lambda)$. This differential equation has the solution $C(\lambda \mathbf{r}) = \lambda^k A$, where A is a term independent of λ but depending in general on **r**. Setting $\lambda = 1$ implies $A = C(\mathbf{r})$, and hence $C(\lambda \mathbf{r}) = \lambda^k C(\mathbf{r})$.

An implication of homogeneity is that if $C(\mathbf{r})$ is homogeneous of degree k, then its derivatives $C_i(\mathbf{r})$ are homogeneous of degree $k-1$, and second derivatives $C_{ij}(\mathbf{r}) = \partial^2 C/\partial r_i \partial r_j$ are homogeneous of degree $k-2$.

A function is concave if it has the curvature of an overturned bowl.[3] We next show the cost function to be concave in input prices for each fixed output level. Consider any pair of strictly positive input price vectors $\mathbf{r}^0$ and $\mathbf{r}'$, and a weighted average of these vectors, $\mathbf{r}^* = \theta\mathbf{r}^0 + (1-\theta)\mathbf{r}'$, with $0 < \theta < 1$. Let $\mathbf{v}^0$, $\mathbf{v}'$, and $\mathbf{v}^*$ be cost minimizing input bundles corresponding to $\mathbf{r}^0$, $\mathbf{r}'$, and $\mathbf{r}^*$, respectively. Then, $\mathbf{r}^0 \cdot \mathbf{v}^* \geqq C(\mathbf{y},\mathbf{r}^0)$ and $\mathbf{r}' \cdot \mathbf{v}^* \geqq C(\mathbf{y},\mathbf{r}')$, implying $C(\mathbf{y},\mathbf{r}^*) = \mathbf{r}^* \cdot \mathbf{v}^* = \theta(\mathbf{r}^0 \cdot \mathbf{v}^*) + (1-\theta)(\mathbf{r}' \cdot \mathbf{v}^*) \geqq \theta C(\mathbf{y},\mathbf{r}^0) + (1-\theta)C(\mathbf{y},\mathbf{r}')$. This inequality is just the algebraic definition of a concave function, requiring that the chord between any two points in the graph of the function is no higher than the graph itself. Hence, the cost function is concave in input prices.

It is possible to obtain a further result that the cost function is continuous in input prices for fixed output, as a mathematical consequence of the concavity of the function.[4]

The property that the cost function is positively linear homogeneous in prices is one form of the old adage that only relative prices enter the economic calculus. The concavity of the cost function in prices is less intuitive economically, despite the almost trivial argument by which it was demonstrated. The reader's intuition may be helped by the following example: if the price of an input, say input 1, is raised by one infinitesimal unit, the cost of production is raised by v_1 units, where v_1 is the quantity of this input used. (One might expect an offsetting effect due to compensating adjustments in the input mix. However, this effect turns out to be a higher order infinitesimal which can be neglected.) At a

[3]A real-valued function f on $\mathbf{E}^n$ is *concave* if for every pair of points $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbf{E}^n$ and every scalar θ satisfying $0 < \theta < 1$, $f(\theta\mathbf{x} + (1-\theta)\mathbf{x}') \geqq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{x}')$. Geometrically, this requires that the chord between any two points in the graph of the function be no higher than the graph itself. f is *quasi-concave* if $f(\theta\mathbf{x} + (1-\theta)\mathbf{x}') \geqq \min\{f(\mathbf{x}), f(\mathbf{x}')\}$ for $0 < \theta < 1$. Geometrically, this requires that upper contour sets, $\{\mathbf{x} \in \mathbf{E}^n | f(\mathbf{x}) \geqq \alpha\}$, be convex for all real α. A function f is (quasi-) convex if $-f$ is (quasi-) concave.

[4]See Fenchel (1949, p. 75) or Rockafellar (1970, p. 82). We can also give a direct argument for this result. Suppose a sequence of strictly positive prices $\mathbf{r}^i$ converges to a strictly positive price $\mathbf{r}^0$. Then, there exist strictly positive price vectors $\mathbf{r}'$ and $\mathbf{r}''$ bounding the $\mathbf{r}^i$, i.e., $\mathbf{r}' \geqq \mathbf{r}^i \geqq \mathbf{r}''$ for each i. Let $\mathbf{v}^i$ and $\mathbf{v}^0$ be cost minimizing bundles for $\mathbf{r}^i$ and $\mathbf{r}^0$, respectively. Since $\mathbf{r}'' \cdot \mathbf{v}^i \leqq \mathbf{r}^i \cdot \mathbf{v}^i \leq \mathbf{r}^i \cdot \mathbf{v}^0$, the set of minimizing bundles $\mathbf{v}^i$ lie in the closed and bounded set of non-negative $\mathbf{v}$ satisfying $\mathbf{r}'' \cdot \mathbf{v} \leqq \mathbf{r}' \cdot \mathbf{v}^0$. Hence, the sequence of $\mathbf{v}^i$ will have a subsequence converging to $\mathbf{v}^*$ in the input requirement set. Retaining the notation $\mathbf{v}^i$ for *any* convergent subsequence, we then have the inequalities $C(\mathbf{y},\mathbf{r}^i) \leqq \mathbf{r}^i \cdot \mathbf{v}^0$ and $C(\mathbf{y},\mathbf{r}^0) \leqq \mathbf{r}^0 \cdot \mathbf{v}^*$ and the limits $\mathbf{r}^i \cdot \mathbf{v}^0 \to \mathbf{r}^0 \cdot \mathbf{v}^0 = C(\mathbf{y},\mathbf{r}^0)$ and $\mathbf{r}^i \cdot \mathbf{v}^i \to \mathbf{r}^0 \cdot \mathbf{v}^*$. The first inequality and limit imply $\lim C(\mathbf{y},\mathbf{r}^i) \leqq C(\mathbf{y},\mathbf{r}^0)$, while the second inequality and limit imply $\lim C(\mathbf{y},\mathbf{r}^i) = \mathbf{r}^0 \cdot \mathbf{v}^* \geqq C(\mathbf{y},\mathbf{r}^0)$. Since these inequalities hold for *every* limit point $\mathbf{v}^*$ of the original sequence, the result $\lim_{\mathbf{r} \to \mathbf{r}^0} C(\mathbf{y},\mathbf{r}^i) = C(\mathbf{y},\mathbf{r}^0)$ is established.

higher price of input 1, a lower quantity of the input will be used at the cost minimum, and the effect on cost of an infinitesimal unit increase in the price will be less than previously. This declining marginal effect is a classical characterization of the concavity property.

Thus far, the cost function has been defined only for strictly positive input prices. We can extend the definition (1) to the case in which some prices are zero, provided we relax the requirement that a minimum cost input bundle actually be achievable. This is done for a non-negative price vector **r** by defining

$$C(\mathbf{y},\mathbf{r}) = \operatorname{Inf}\{\mathbf{r}\cdot\mathbf{v} \mid \mathbf{v} \in \mathbf{V}(\mathbf{y})\}, \tag{1a}$$

where "Inf" denotes the infinum, or greatest lower bound, of the numbers in the set. For positive **r**, this definition coincides with (1). For non-negative **r** with some zero components, if a cost minimizing input bundle exists, the definition (1a) will yield a cost equal to the value of this input bundle. Alternately, no cost minimizing input bundle may exist (this is the case, for example, in the Cobb–Douglas input requirement sets illustrated in Section 3), and the cost $C(\mathbf{y},\mathbf{r})$ in (1a) is approached by the values of an unbounded sequence of input bundles. With minor variations, the arguments we gave earlier that the cost function is positively linear homogeneous and concave in positive input prices for a fixed output bundle can be applied to the extended definition (1a) to establish these properties for all non-negative prices. A more difficult argument [see Rockafellar (1970, p. 85) or Appendix A.3, Section 12.7] establishes that the extended cost function is continuous in all non-negative input prices for a fixed output bundle.

The basic properties of the cost function demonstrated in this section are summarized in the following result.

Lemma 1. Suppose that a firm has an input-regular production possibility set with a producible output set $\mathbf{Y}^*$ and input requirement sets $\mathbf{V}(\mathbf{y})$ for $\mathbf{y} \in \mathbf{Y}^*$. Suppose that the firm faces competitive input markets with a non-negative input price vector **r**. Then, the cost function defined by (1) exists for all $\mathbf{y} \in \mathbf{Y}^*$ and all strictly positive **r**, and coincides with the extended cost function defined by (1a), which exists for all $\mathbf{y} \in \mathbf{Y}^*$ and all non-negative **r**. Further, for each $\mathbf{y} \in \mathbf{Y}^*$, the (extended) cost function as a function of **r** is non-negative, positive when **r** is strictly positive and **y** is non-zero, non-decreasing, positively linear homogeneous, concave, and continuous.

5. The Derivative Property

The cost function is related to the cost minimizing input demand functions through its partial derivatives with respect to input prices. Again consider a firm with an input-regular production possibility set, and let $c = C(\mathbf{y},\mathbf{r})$ denote its cost function, with $\mathbf{r} = (r_1,\dots,r_N)$ a vector of positive input prices. When the partial derivative of the cost function with respect to an input price r_n exists at an argument $(\mathbf{y},\mathbf{r})$, it will be denoted by $C_n(\mathbf{y},\mathbf{r}) = \partial C/\partial r_n$. We now establish the following result: If $C_n(\mathbf{y},\mathbf{r})$ exists, then it equals the unique cost minimizing input of good n at the argument $(\mathbf{y},\mathbf{r})$; and if there is a unique cost minimizing input of good n at the argument $(\mathbf{y},\mathbf{r})$, then $C_n(\mathbf{y},\mathbf{r})$ exists. This property, known as *Shephard's lemma*, was first noted by Hotelling (1932) and established formally by Shephard (1953). The demonstration given below was first used by McKenzie (1957).

Suppose $\mathbf{y}$ is a producible output bundle and $\mathbf{r}^0$ is a strictly positive input price vector, and suppose $\mathbf{v}^0$ is a corresponding cost minimizing input bundle. Consider any vector of input price increments $\Delta\mathbf{r} = (\Delta r_1,\dots,\Delta r_N)$. For any scalar θ which is sufficiently small to make $\mathbf{r}^0 + \theta\Delta\mathbf{r}$ strictly positive, the definition of the cost function implies the inequality $C(\mathbf{y},\mathbf{r}^0 + \theta\Delta\mathbf{r}) \leqq (\mathbf{r}^0 + \theta\Delta\mathbf{r})\cdot\mathbf{v}^0$. Since $\mathbf{r}^0\cdot\mathbf{v}^0 = C(\mathbf{y},\mathbf{r}^0)$, this inequality can be rewritten as

$$C(\mathbf{y},\mathbf{r}^0 + \theta\Delta\mathbf{r}) - C(\mathbf{y},\mathbf{r}^0) \leqq \theta(\Delta\mathbf{r})\cdot\mathbf{v}^0. \tag{2}$$

Single out one commodity, say the first, and define $\Delta r_1 = 1$ and $\Delta r_2 = \cdots = \Delta r_N = 0$. Define the ratio

$$g(\theta) = [C(\mathbf{y},r_1^0 + \theta,r_2^0,\dots,r_N^0) - C(\mathbf{y},r_1^0,r_2^0,\dots,r_N^0)]/\theta,$$

for $\theta \neq 0$. If θ is positive, (2) can be written

$$g(\theta) \leqq v_1^0. \tag{3a}$$

If θ is negative, the inequality reverses to give

$$g(\theta) \geqq v_1^0. \tag{3b}$$

If the partial derivative $C_1(\mathbf{y},\mathbf{r}^0)$ exists, then by its definition $g(\theta)$ has a limiting value, as θ approaches zero from above or below, equal to $C_1(\mathbf{y},\mathbf{r}^0)$. The inequalities then imply $C_1(\mathbf{y},\mathbf{r}^0) = v_1^0$. Since this equality must hold for any cost minimizing input vector, the cost minimizing input of good 1 is unique. This proves the first half of the lemma, and shows that differentiability of the cost function in input prices rules out the existence of flat segments in isoquants where multiple minima can occur.

The second half of Shephard's lemma requires a more advanced mathematical argument; see Appendix A.3, Lemma 13.8, or Rockafellar (1970, p. 265(e)).

A second justification of the derivative property of cost functions can be given using classical calculus arguments provided we add some facilitating assumptions on the technology. The following argument is due to Samuelson (1938). Suppose for a given producible output bundle $\mathbf{y}$, the input requirement set is defined by the input bundles $\mathbf{v}$ satisfying $F(\mathbf{y},\mathbf{v}) \geqq 1$, where F is a transformation function which is twice continuously differentiable in $\mathbf{v}$. The problem of cost minimization can then be restated as a classical constrained minimization problem: Minimize $\mathbf{r}\cdot\mathbf{v}$ subject to $F(\mathbf{y},\mathbf{v}) \geqq 1$. Form the Lagrangian $L = \mathbf{r}\cdot\mathbf{v} - \lambda(F(\mathbf{y},\mathbf{v}) - 1)$. Ignoring for simplicity the possibility of a corner solution or non-binding constraint, the first-order conditions for a mimimum are given by equating to zero the partial derivatives of the Lagrangian with respect to $\mathbf{v}$ and λ. (See Appendix A.2.) This procedure yields $N+1$ equations, the constraint $F(\mathbf{y},\mathbf{v}) = 1$ plus the marginal conditions $r_n = \lambda \partial F/\partial v_n$, $n = 1,\ldots,N$. Suppose this system has a unique solution for $\mathbf{v}$ and λ as a function of $(\mathbf{y},\mathbf{r})$, and let $v_n = h^n(\mathbf{y},\mathbf{r})$ denote the solution for v_n. Assume the h^n are continuously differentiable in $\mathbf{r}$. From the definition $C(\mathbf{y},\mathbf{r}) = \sum_{n=1}^N r_n h^n(\mathbf{y},\mathbf{r})$, we obtain the condition $C_1(\mathbf{y},\mathbf{r}) = h^1(\mathbf{y},\mathbf{r}) + \sum_{n=1}^N r_n \partial h^n/\partial r_1$. But $r_n = \lambda\partial F/\partial v_n$ and $F(\mathbf{y},h^1(\mathbf{y},\mathbf{r}),\ldots,h^N(\mathbf{y},\mathbf{r})) = 1$ imply, by differentiation,

$$\frac{1}{\lambda}\sum_{n=1}^{N} r_n(\partial h^n/\partial r_1) = \sum_{n=1}^{N}(\partial F/\partial v_n)(\partial h^n/\partial r_1) = 0, \tag{4}$$

and hence $C_1(\mathbf{y},\mathbf{r}) = h^1(\mathbf{y},\mathbf{r})$.

Several stronger derivative properties of the cost function can be obtained as corollaries of the mathematical theory of convex functions. For each producible output bundle, the cost function can be shown to possess first and second differentials for almost all strictly positive input price vectors (i.e., for all positive input price vectors except those in a set of Lebesgue measure zero). This implies that for almost all input price vectors there is a unique input bundle demanded under cost minimization. Further, the second partial derivatives of the cost function with respect to input prices are found to be independent of the order of differentiation whenever the second differential exists. Since these second differentials are the first partial derivatives of the cost minimizing input demands, this result implies a production analogue of the symmetry of the Slutsky substitution effects in consumer theory. It should be noted that these properties hold without any assumptions on the structure of the technology beyond the condition that it be regular. In particular, they hold even if the underlying technology exhibits non-convexities, indivisible inputs or outputs, or failures of free disposal. Lemma 12.1 in Appendix A.3 states these results formally.

In many economic applications, particularly comparative statics, it is convenient to know that the cost minimizing input demands are unique for all positive input prices (Shephard's lemma then implies that the cost function possesses a first differential in input prices for all positive values of these prices). A stronger version of Assumption B on the convexity of the input requirement sets from below is necessary and sufficient to give this property. Define a *plane* (or *hyperplane*) in input space to be a set of "isocost" points; i.e., a set **H** of points **v** satisfying $\mathbf{r}\cdot\mathbf{v} = r_1v_1 + \cdots + r_Nv_N = r_0$ for some fixed non-zero vector **r** and some scalar r^0. The vector **r** gives the direction numbers of the plane, and is termed a *normal* to the plane. A plane **H** *bounds* a set **V** if the set is contained in one of the closed half-spaces defined by the plane; i.e., if $\mathbf{r}\cdot\mathbf{v} = r_0$ for **v** in **H**, and $\mathbf{r}\cdot\mathbf{v} \geqq r_0$ for $\mathbf{v} \in \mathbf{V}$, then **H** bounds **V**. A plane **H** *supports* a set **V** if it bounds **V**, and **H** and **V** meet.

> *Assumption B-2.* The input requirement sets are strictly convex from below; i.e., if **H** is any plane with a strictly positive normal which supports **V(y)** from below,[5] then **H** meets **V(y)** at exactly one point.

This assumption states that if **v** and **v**′ are in **V(y)**, with $\mathbf{v} \neq \mathbf{v}'$, and $\mathbf{v}'' = \theta\mathbf{v} + (1-\theta)\mathbf{v}'$, $0 < \theta < 1$, then there exists $\mathbf{v}^*$ in **V(y)** such that $\mathbf{v}^* \leqq \mathbf{v}''$ and either (i) $\mathbf{v}^* \neq \mathbf{v}''$ or (ii) there exists no plane **H** with strictly positive normal which contains **v** and **v**′ and which supports **V(y)** from below. If **V(y)** or its free disposal hull is a strictly convex set, then Assumption B-2 holds, and condition (i) above is always satisfied.

Figure 3 illustrates this assumption. In (a), the weighted average $\mathbf{v}''$ of two points **v** and **v**′ lies northeast of $\mathbf{v}^*$ in the set. The points $\mathbf{v}^3$ and $\mathbf{v}^4$ satisfy condition (ii) above since the only plane through them is parallel to the $\mathbf{v}_2$ axis, and hence has a zero direction number. In (b), the assumption fails because the isoquant contains a flat segment. A three-

[5]The plane **H** bounds (or supports) **V(y)** *from below* if $\mathbf{r}\cdot\mathbf{v} = r_0$ for $\mathbf{v} \in \mathbf{H}$, and $\mathbf{r}\cdot\mathbf{v} \geqq r_0$ for $\mathbf{v} \in \mathbf{V(y)}$.

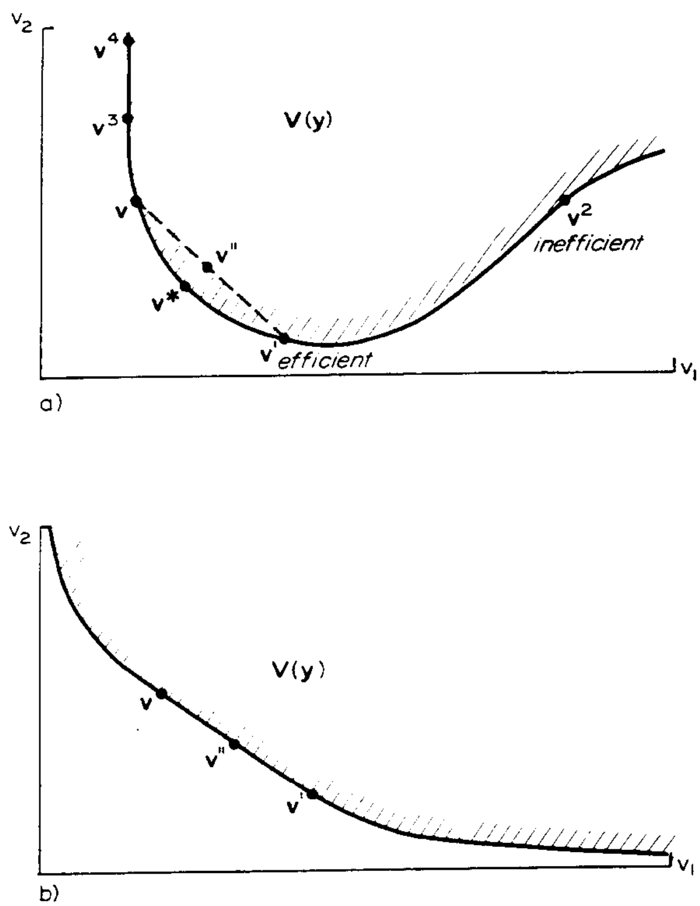


FIGURE 3. (a) Assumption B-2 holds. (b) Assumption B-2 fails.

input requirement set is illustrated in (c). The points **v** and **v**′ both lie in a plane, identified by the rectangle ABCD, parallel to the v_3 coordinate axis and bounding **V**(**y**) from below. This plane has a zero direction number in the direction v_3. Every other plane containing **v** and **v**′ cuts through the input requirement set rather than bounding it. Hence, condition (ii) above holds for **v** and **v**′. For distinct pairs of points such as $\mathbf{v}^3$ and $\mathbf{v}^4$, the input requirement set contains a point $\mathbf{v}^6$ no greater than and unequal to a linear combination $\mathbf{v}^5 = \theta\mathbf{v}^3 + (1-\theta)\mathbf{v}^4$, $0 < \theta < 1$. Hence, $\mathbf{v}^3$ and $\mathbf{v}^4$ satisfy condition (i) above.

It is clear that this condition implies that minimum cost is achieved by a unique input bundle for any strictly positive input price vector: if two distinct input bundles simultaneously minimized cost, then a weighted average of them would also have this minimum cost, and Assumption B-2 would imply the existence of another bundle in the input require-

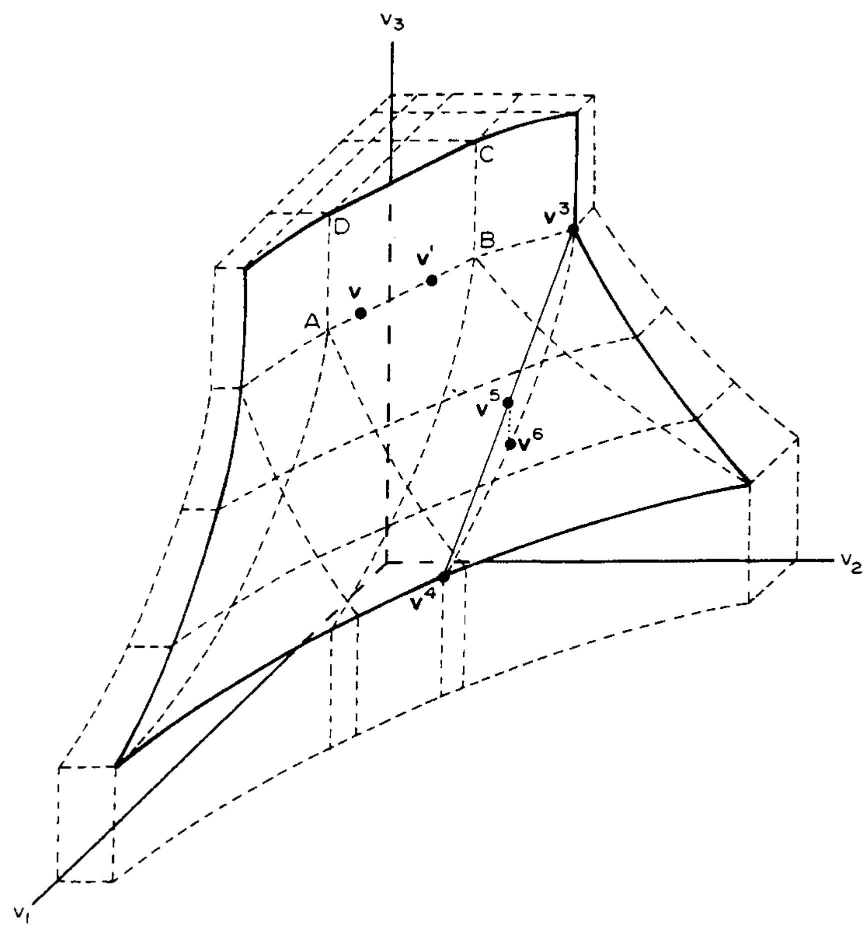


FIGURE 3(c)

ment set costing less, and thus contradict the initial supposition. As noted earlier, this uniqueness of the cost minimizing input demand bundle guarantees that the cost function has a differential in input prices for *all* positive input price vectors. The mathematical properties of convex functions then imply that the cost function is *continuously* differentiable in input prices.

Then, Assumption B-2 and Shephard's lemma imply that (1) unique cost minimizing input demands exist for all positive input prices and are given by the price derivatives of the cost function, and (2) the input demands vary continuously with input prices.

Returning to the case of input-regular production possibility sets without added assumptions on structure, it is possible to generalize the concept of a vector of partial derivatives of the cost function in a mathematically meaningful way so that (1) this generalized derivative, called the *sub-differential*, always exists and is a *set* of N-dimensional vectors, (2) the vectors in the sub-differential correspond to the cost minimizing input bundles in a sense elaborated below, and (3) in the case where there is a unique cost minimizing input bundle, the sub-differential contains exactly the ordinary vector of partial derivatives. This concept is developed formally in Appendix A.3, Sections 13.7–13.9. Expanding informally on the conclusions of this construction, the sub-differential will contain a single vector if and only if there is a unique cost minimizing input bundle, in which case the definition of the sub-differential reduces to the ordinary definition of a vector of partial derivatives and the vector in the sub-differential coincides with the cost minimizing input bundle. More generally, the sub-differential will contain all the cost minimizing input bundles. If the input requirement sets satisfy Assumption B of convexity from below, then the sub-differential equals the set of cost minimizing bundles at any $(\mathbf{y},\mathbf{r})$ argument. When no convexity assumptions are imposed on the input requirement sets, then the sub-differential may contain, in addition to the true cost minimizing input bundles, some input bundles which lie outside the input requirement set. However, all these latter bundles can be written as weighted averages of a finite number of true cost minimizing input vectors.

6. Duality

We have established that corresponding to every input-regular production possibility set is a cost function with the properties summarized in Lemma 1. We now pose the converse question: given a function with the properties specified in Lemma 1, does there exist an input-regular production possibility set such that this function is its minimum cost function? A duality between input-conventional production possibility sets and cost functions first proved by Shephard (1953) and Uzawa (1962) provides an affirmative answer. This theoretical result is of considerable practical importance. It allows the economist to write down cost functions and their input demand systems and verify their consistency with the cost minimization hypothesis without difficult constructive arguments. Further, it establishes that the cost function contains all the information necessary to reconstruct the structure of production possibilities. It is in a sense a "sufficient statistic" for the technology. Thus, corresponding to every hypothesis the economist might impose on the structure of a conventional production possibility set is a hypothesis on the form of the cost function.

We begin the discussion of duality with several definitions. An *input-conventional cost structure* is defined by (1) a non-empty set of non-negative M-dimensional vectors, denoted by $\mathbf{Y}^*$ and interpreted as a producible output set, and (2) a real-valued function $c = C(\mathbf{y},\mathbf{r})$, defined on the domain consisting of $\mathbf{y} \in \mathbf{Y}^*$ and strictly positive N-dimensional price vectors $\mathbf{r}$, this function being non-negative, non-decreasing, positively linear homogeneous, and concave in $\mathbf{r}$ for each fixed $\mathbf{y} \in \mathbf{Y}^*$, and positive for non-zero $\mathbf{y}$.

Consider an input-conventional cost structure $C(\mathbf{y},\mathbf{r})$ defined for $\mathbf{y}$ in a set $\mathbf{Y}^*$. For each $\mathbf{y} \in \mathbf{Y}^*$, define an *implicit input requirement set*

$$\mathbf{V}^*(\mathbf{y}) = \{\mathbf{v} \in \mathbf{E}^N \,|\, \mathbf{v} \geqq 0,\ \mathbf{r}\cdot\mathbf{v} \geqq C(\mathbf{y},\mathbf{r}) \quad \text{for all strictly positive } \mathbf{r}\}. \quad (5)$$

The implicit input requirement sets will be shown to be non-empty, allowing the definition of an *implicit production possibility set*

$$\mathbf{Y} = \{(\mathbf{y},\mathbf{v}) \in \mathbf{E}^{M+N} | \mathbf{y} \in \mathbf{Y}^*,\ \mathbf{v} \in \mathbf{V}^*(\mathbf{y})\}. \quad (6)$$

The first duality result establishes that each input-conventional cost function determines an implicit production possibility set which is input-conventional (i.e., is input-regular and satisfies Assumptions A and B).

> *Lemma* 2. If $C(\mathbf{y},\mathbf{r})$ is an input-conventional cost function defined for $\mathbf{y}$ in a set $\mathbf{Y}^*$, then the implicit input requirement sets $\mathbf{V}^*(\mathbf{y})$ are non-empty for each $\mathbf{y} \in \mathbf{Y}^*$, and the implicit production possibility set $\mathbf{Y}$ is input-conventional.

Proof: The lemma will be proved in three steps. First, the implicit input requirement sets are shown to be non-empty for each $\mathbf{y} \in \mathbf{Y}^*$. This allows the implicit production possibility set (6) to be defined. Second, this production possibility set is shown to be input-regular. Third, Assumptions A and B are shown to hold.

Step 1. By hypothesis, $\mathbf{Y}^*$ is non-empty. Consider any $\mathbf{y} \in \mathbf{Y}^*$. Let $\mathbf{r}^0 = (1,1,\dots,1)$ be an N-vector of ones, and define an input bundle $\mathbf{v}^0 = c\mathbf{r}^0$ with $c = C(\mathbf{y},\mathbf{r}^0)$. Let $|\mathbf{r}| = \sum_{n=1}^{N} |r_n|$ denote the norm of an N-vector. Since the function $C(\mathbf{y},\mathbf{r})$ is non-decreasing and positively linear homogeneous in $\mathbf{r}$, we have for any strictly positive $\mathbf{r}$ the inequality

$$C(\mathbf{y},\mathbf{r}) = C(\mathbf{y},\mathbf{r}/|\mathbf{r}|)\cdot|\mathbf{r}| \leqq C(\mathbf{y},\mathbf{r}^0)\cdot|\mathbf{r}| = C(\mathbf{y},\mathbf{r}^0)(\mathbf{r}\cdot\mathbf{r}^0) = \mathbf{r}\cdot\mathbf{v}^0.$$

Then by (5), $\mathbf{v}^0$ is contained in the implicit input requirement set, which is thus non-empty.

Step 2. To show that the implicit production possibility set is input-regular, we must show that each implicit input requirement set is closed and does not contain the zero input bundle when the output bundle is non-zero. Consider any $\mathbf{y} \in \mathbf{Y}^*$. To show that $\mathbf{V}^*(\mathbf{y})$ is closed, consider any sequence $\mathbf{v}^k \in \mathbf{V}^*(\mathbf{y})$ converging to a bundle $\mathbf{v}^0$. For any positive $\mathbf{r}$, the $\mathbf{v}^k$ satisfy $\mathbf{r}\cdot\mathbf{v}^k \geqq C(\mathbf{y},\mathbf{r})$ by (5). Then this inequality must hold also in the limit, $\mathbf{r}\cdot\mathbf{v}^0 \geqq C(\mathbf{y},\mathbf{r})$. But (5) then implies $\mathbf{v}^0 \in \mathbf{V}^*(\mathbf{y})$. Hence, $\mathbf{V}^*(\mathbf{y})$ is closed. If the zero input bundle is in $\mathbf{V}^*(\mathbf{y})$, then by (5), $0 = C(\mathbf{y},\mathbf{r})$, implying $\mathbf{y} = \mathbf{0}$ by hypothesis.

Step 3. We first establish that $\mathbf{Y}$ satisfies Assumption A, free disposal of inputs. If a bundle $\mathbf{v}$ is in $\mathbf{V}^*(\mathbf{y})$, and a second bundle $\mathbf{v}'$ is at least as large in every component, then for any positive $\mathbf{r}$, $\mathbf{r}\cdot\mathbf{v}' \geqq \mathbf{r}\cdot\mathbf{v} \geqq C(\mathbf{y},\mathbf{r})$, implying $\mathbf{v}' \in \mathbf{V}^*(\mathbf{y})$. Hence, Assumption A holds. We next establish Assumption B, convexity from below of $\mathbf{V}^*(\mathbf{y})$. If $\mathbf{v}$, $\mathbf{v}'$ are input bundles in $\mathbf{V}^*(\mathbf{y})$ and for a scalar θ, $0 < \theta < 1$, $\mathbf{v}'' = \theta\mathbf{v} + (1-\theta)\mathbf{v}'$ is a weighted combination of these bundles, then for any positive $\mathbf{r}$ the inequalities $\mathbf{r}\cdot\mathbf{v} \geqq C(\mathbf{y},\mathbf{r})$ and $\mathbf{r}\cdot\mathbf{v}' \geqq C(\mathbf{y},\mathbf{r})$ imply $\mathbf{r}\cdot\mathbf{v}'' \geqq C(\mathbf{y},\mathbf{r})$. Hence, $\mathbf{v}'' \in \mathbf{V}^*(\mathbf{y})$, and $\mathbf{V}^*(\mathbf{y})$ is convex. Q.E.D.

The next result, called the Shephard–Uzawa duality theorem [Shephard (1970), Uzawa (1962)], establishes a one-to-one relationship between input-conventional production possibility sets and input-conventional cost structures. Let us call the procedure (1) which obtains a minimum cost function from a production possibility set the *cost mapping*, and the procedure (5) which obtains an implicit production possibility set from a cost function the *technology mapping*. Lemma 1 establishes that the cost mapping is a function from the class of input-conventional (actually, more generally, input-regular) production possibility sets into the class of input-conventional cost structures. Lemma 2 establishes that the technology mapping is a function from the class of input-conventional cost structures into the class of input-conventional production possibility sets. The duality theorem establishes that on the two input-conventional classes above, the cost mapping and technology mapping are mutual inverses; i.e., applying the cost mapping to an input-conventional production possibility set yields a cost function, and applying the technology mapping to this cost function yields the initial production possibility set; and similarly, applying the technology mapping to an input-conventional cost structure yields a production possibility set, and applying the cost mapping to this production possibility set yields the initial cost function. Consequently, all structural features of the production possibilities are embodied in the functional specification of the cost function and are recovered by the technology mapping. As a corollary, distinct input-conventional technologies yield distinct input-conventional cost functions, and vice versa.

It should be noted that the one-to-one link between the input-conventional classes described above does not hold between input-conventional cost structures and input-regular production possibility sets. Distinct input-regular production possibility sets may yield the same input-conventional cost function. However, while going from the production possibility set to the cost function can entail a real loss of technological information in this case, the information lost is precisely that which is superfluous to the determination of observed competitive cost minimizing behavior. Figure 4 illustrates input-regular technologies which yield the same cost structure. In this example, under cost minimization the portions of the isoquant labeled "alternative 1" and "alternative 2" are never utilized, and hence cannot be distinguished on the basis of the behavior of the firm.

Lemma 3. Application of the cost mapping (1) to an input-conventional production possibility set yields an input-conventional cost structure. Application of the technology mapping (5) to this cost structure yields the initial production possibility set. Conversely, application of the technology mapping (5) to an input-

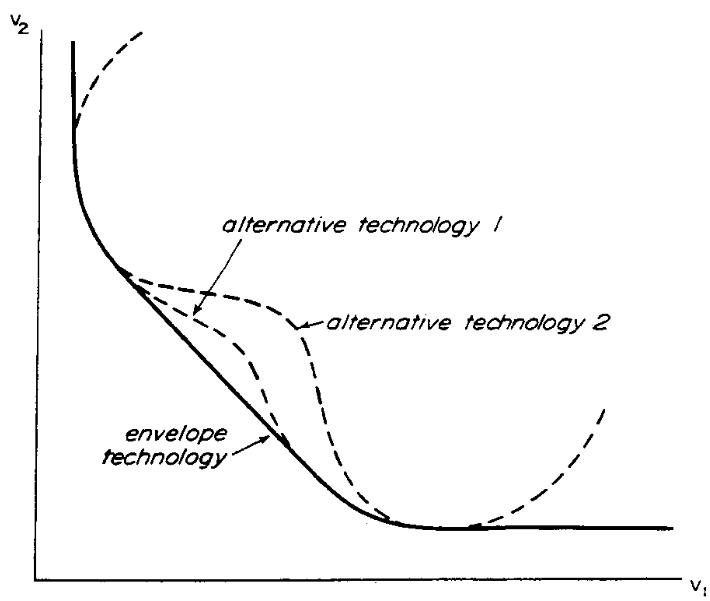


FIGURE 4

conventional cost structure yields an input-conventional production possibility set. Application of the cost mapping (1) to this production possibility set yields the initial cost structure.

Proof: Consider an input-conventional production possibility set defined by a producible output set $\mathbf{Y}^*$ and input requirement sets $\mathbf{V}(\mathbf{y})$ for $\mathbf{y} \in \mathbf{Y}^*$. The cost mapping yields a cost function $C(\mathbf{y},\mathbf{r})$, and the technology mapping applied to this cost function yields implicit input requirement sets $\mathbf{V}^*(\mathbf{y})$ for $\mathbf{y} \in \mathbf{Y}^*$. By Lemmas 1 and 2, the $\mathbf{V}^*(\mathbf{y})$ are input-conventional. We now show that $\mathbf{V}(\mathbf{y}) = \mathbf{V}^*(\mathbf{y})$.

If $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{v}^0 \in \mathbf{V}(\mathbf{y})$, then $\mathbf{r}\cdot\mathbf{v}^0 \geqq C(\mathbf{y},\mathbf{r})$ for all positive $\mathbf{r}$ by (1), and (5) then implies $\mathbf{v}^0 \in \mathbf{V}^*(\mathbf{y})$. Alternately, suppose $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{v}^0 \notin \mathbf{V}(\mathbf{y})$. We can apply a strict separating hyperplane theorem (Appendix A.3, 10.13) to establish the existence of a non-zero N-vector $\mathbf{r}$ and a positive scalar θ such that $\mathbf{r}\cdot\mathbf{v}^0 + \theta \leqq \mathbf{r}\cdot\mathbf{v}$ for all $\mathbf{v} \in \mathbf{V}(\mathbf{y})$. Since $\mathbf{V}(\mathbf{y})$ satisfies the free disposal Assumption A, this inequality implies that $\mathbf{r}$ is non-negative. Choose $\mathbf{r}^0$ larger than $\mathbf{r}$ in every component and sufficiently close to $\mathbf{r}$ to satisfy $|\mathbf{r}\cdot\mathbf{v}^0 - \mathbf{r}^0\cdot\mathbf{v}^0| < \theta/2$. Then, $\mathbf{r}^0\cdot\mathbf{v}^0 + \theta/2 \leqq \mathbf{r}\cdot\mathbf{v} \leqq \mathbf{r}^0\cdot\mathbf{v}$ for all $\mathbf{v} \in \mathbf{V}(\mathbf{y})$, implying $\mathbf{r}^0\cdot\mathbf{v}^0 < C(\mathbf{y},\mathbf{r}^0)$. By (5), $\mathbf{v}^0 \notin \mathbf{V}^*(\mathbf{y})$. This establishes $\mathbf{V}(\mathbf{y}) = \mathbf{V}^*(\mathbf{y})$.

To prove the second half of the lemma, consider an input-conventional cost structure given by a function $C(\mathbf{y},\mathbf{r})$ defined on a set $\mathbf{y} \in \mathbf{Y}^*$. The technology mapping yields implicit input requirement sets $\mathbf{V}^*(\mathbf{y})$, and the cost mapping applied to these input requirement sets yields a cost function $C^*(\mathbf{y},\mathbf{r})$ for $\mathbf{y} \in \mathbf{Y}^*$. By Lemmas 1 and 2, $C^*(\mathbf{y},\mathbf{r})$ is input-conventional. We now show that $C(\mathbf{y},\mathbf{r}) = C^*(\mathbf{y},\mathbf{r})$ for $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{r}$ positive.

Since $\mathbf{v} \in \mathbf{V}^*(\mathbf{y})$ implies $\mathbf{r}\cdot\mathbf{v} \geqq C(\mathbf{y},\mathbf{r})$, we have immediately the inequality $C^*(\mathbf{y},\mathbf{r}) \geqq C(\mathbf{y},\mathbf{r})$. The proof is completed by supposing that $C^*(\mathbf{y},\mathbf{r}^0) > C(\mathbf{y},\mathbf{r}^0)$ for some $\mathbf{y} \in \mathbf{Y}^*$ and positive $\mathbf{r}^0$, and showing a contradiction results. Define the set $\mathbf{B} = \{(\mathbf{r},\xi) \in \mathbf{E}^{N+1} | \mathbf{r} \text{ positive}, \ \xi \geqq -C(\mathbf{y},\mathbf{r})\}$. Since C is concave and positively linear homogeneous in $\mathbf{r}$, the set $\mathbf{B}$ is a non-empty, convex cone. The point $(\mathbf{r}^0,\xi^0)$ with $\xi^0 = -C^*(\mathbf{y},\mathbf{r}^0)$ is by supposition not contained in $\mathbf{B}$. Further, by the continuity of C established in Lemma 1, $(\mathbf{r}^0,\xi^0)$ is not contained in the closure of $\mathbf{B}$. Then, the strict separating hyperplane theorem (Appendix A.3, 10.13) establishes the existence of a non-zero vector $(\mathbf{v}^0,\lambda) \in \mathbf{E}^{N+1}$ and a positive scalar θ such that $(\mathbf{v}^0,\lambda)\cdot(\mathbf{r}^0,\xi^0) + \theta \leqq (\mathbf{v}^0,\lambda)\cdot(\mathbf{r},\xi)$ for all $(\mathbf{r},\xi) \in \mathbf{B}$. Since $\mathbf{B}$ satisfies "free disposal", this inequality implies $\mathbf{v}^0$ and λ non-negative. If λ were zero, then the inequality would be violated by a point

$(\mathbf{r}^0,\xi) \in B$. Hence, we can assume without loss that $\lambda = 1$. Since **B** is a cone, the inequality can be written

$$\mathbf{r}^0\cdot\mathbf{v}^0 - C^*(\mathbf{y},\mathbf{r}^0) + \theta \leqq 0 \leqq \mathbf{r}\cdot\mathbf{v}^0 - C(\mathbf{y},\mathbf{r}), \tag{7}$$

for all positive $\mathbf{r}$. By (5), $\mathbf{r}\cdot\mathbf{v}^0 \geqq C(\mathbf{y},\mathbf{r})$ for all positive $\mathbf{r}$ implies $\mathbf{v}^0 \in \mathbf{V}^*(\mathbf{y})$, and hence $\mathbf{r}^0\cdot\mathbf{v}^0 \geqq C^*(\mathbf{y},\mathbf{r}^0)$. But this contradicts (7). Hence, $C(\mathbf{y},\mathbf{r}) = C^*(\mathbf{y},\mathbf{r})$. Q.E.D.

7. Distance Functions and Economic Transformation Functions

Frequently economists characterize production possibility sets implicitly using transformation functions or, in the one-output case, production functions. We will now give a straightforward restatement of the basic duality theorem of Section 6 in terms of the cost function and a form of a transformation function known as the distance function. The concept of a distance function comes from the mathematical theory of convex sets, and was introduced into economics by Shephard (1970). While the reformulation of duality in terms of distance functions is potentially useful in applications, its primary appeal comes from the fact that it allows us to establish a full, formal mathematical duality between transformation and cost functions, in the sense that both can be thought of as drawn from the same class of functions and having the same properties. We can exploit this formal duality to get "double our money" in further investigations of production and cost structures: if we can prove that a property "P" on a transformation function implies a property "Q" on a cost function, we can conclude by duality that property "P" on a cost function implies property "Q" on a transformation function. Hanoch's Chapter I.2 in this volume develops and applies this formal duality to functional forms in production theory.

Consider an input-conventional production possibility set characterized by a producible output set $\mathbf{Y}^*$ and input requirement sets $\mathbf{V}(\mathbf{y})$ for $\mathbf{y} \in \mathbf{Y}^*$. For this technology, define the *distance function*

$$F(\mathbf{y},\mathbf{v}) = \operatorname{Max}\left\{\lambda > 0 \,\middle|\, \frac{1}{\lambda}\mathbf{v} \in \mathbf{V}(\mathbf{y})\right\}, \tag{8}$$

for $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{v}$ strictly positive. In Lemma 4 below, we show that this formula defines a unique function which is finite valued for non-zero

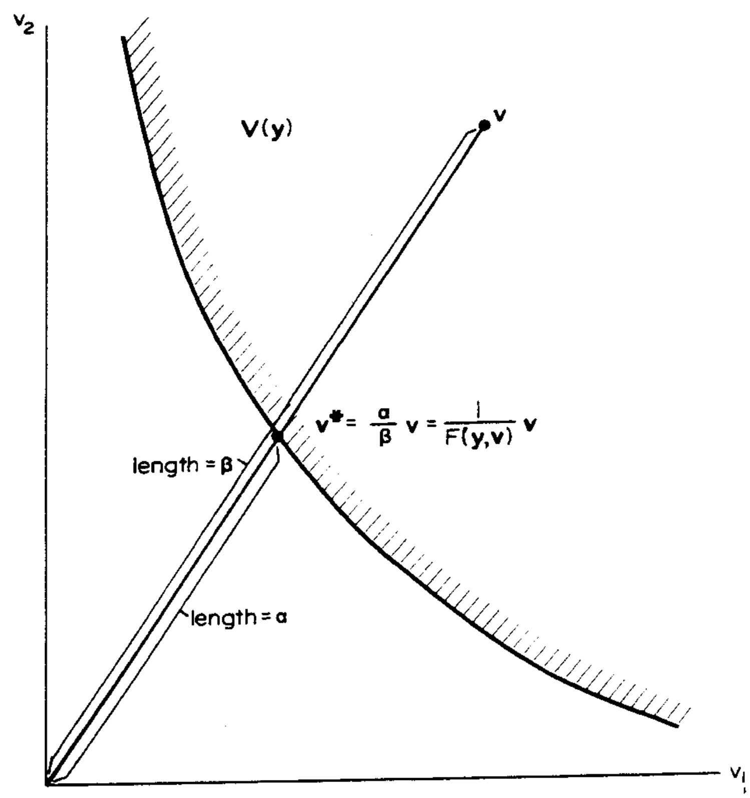


FIGURE 5

$\mathbf{y} \in \mathbf{Y}^*$. For $\mathbf{y} = \mathbf{0}$, the vector $\mathbf{0}$ may be in $\mathbf{V}(\mathbf{0})$, in which case $F(\mathbf{y},\mathbf{v})$ is defined to take the extended value $+\infty$. As illustrated in Figure 5, the value of $F(\mathbf{y},\mathbf{v})$ is given by the ratio of the length of the vector $\mathbf{v}$ to the length of a vector $\mathbf{v}^*$ defined by the intersection of the "y-isoquant" and the ray through $\mathbf{v}$.

For $\mathbf{y} \in \mathbf{Y}^*$, the strictly positive vectors $\mathbf{v}$ in the input requirement set $\mathbf{V}(\mathbf{y})$ are exactly those satisfying $F(\mathbf{y},\mathbf{v}) \geqq 1$. From the definition of the distance function, $\mathbf{v}/F(\mathbf{y},\mathbf{v})$ is contained in $\mathbf{V}(\mathbf{y})$, but no point southwest of it is in $\mathbf{V}(\mathbf{y})$. If $F(\mathbf{y},\mathbf{v}) \geqq 1$, then $\mathbf{v} \geqq \mathbf{v}/F(\mathbf{y},\mathbf{v})$, and $\mathbf{v} \in \mathbf{V}(\mathbf{y})$ by free disposal. If $F(\mathbf{y},\mathbf{v}) < 1$, then $\mathbf{v} < \mathbf{v}/F(\mathbf{y},\mathbf{v})$ is not in $\mathbf{V}(\mathbf{y})$.

Suppose the technology has a single output, and is defined by a production function $y = f(\mathbf{v})$. Then, at any point $(y,\mathbf{v})$, the distance function $F(y,\mathbf{v})$ takes on the value necessary to satisfy $y = f(\mathbf{v}/F(y,\mathbf{v}))$. This formula has a particularly simple form when the production function f is homothetic; i.e., $f(\mathbf{v}) = \phi(h(\mathbf{v}))$, where $h(\mathbf{v})$ is a linear homogeneous function and ϕ is a strictly monotone increasing function with

$\phi(0) = 0$. Then,

$$y = \phi\left[h\left(\frac{\mathbf{v}}{F(y,\mathbf{v})}\right)\right] = \phi\left[\frac{1}{F(y,\mathbf{v})}h(\mathbf{v})\right] \quad \text{or} \quad F(y,\mathbf{v}) = \frac{h(\mathbf{v})}{\phi^{-1}(y)},$$

where ϕ^{-1} is the inverse function of ϕ.

In the case of multiple outputs and a technology described by a tranformation function $G(\mathbf{y},\mathbf{v}) = 0$, the distance function is defined for $(\mathbf{y},\mathbf{v})$ by the value necessary to make $G(\mathbf{y}, \mathbf{v}/F(\mathbf{y},\mathbf{v})) = 0$. The distance function is then itself one representation of the transformation function for the technology, $F(\mathbf{y},\mathbf{v}) = 1$.

A distance function $F(\mathbf{y},\mathbf{v})$, defined for $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{v}$ positive, will be termed *input-conventional* if for each $\mathbf{y} \in \mathbf{Y}^*$, F as a function of $\mathbf{v}$ is positive, non-decreasing, positively linear homogeneous, concave, and continuous and if $F(\mathbf{y},\mathbf{v}) = +\infty$ implies $\mathbf{y} = 0$. Generally, we expect a cost function $C(\mathbf{y},\mathbf{r})$ to be increasing in the output bundle $\mathbf{y}$ and a distance function $F(\mathbf{y},\mathbf{v})$ to be decreasing in the output bundle $\mathbf{y}$. However, input-conventional cost structures and distance functions are defined to have identical mathematical properties with respect to their second arguments, input prices or inputs respectively. It is this formal duality that proves useful in obtaining further results. We first establish the relation between input-conventional production possibilities and input-conventional distance functions.

Lemma 4. Suppose a producible output set $\mathbf{Y}^*$ and input requirement sets $\mathbf{V}(\mathbf{y})$ for $\mathbf{y} \in \mathbf{Y}^*$ define an input-conventional technology. Then, the distance function $F(\mathbf{y},\mathbf{v})$ defined by (8) exists and is input-conventional. Conversely, given a non-empty set $\mathbf{Y}^*$ and an input-conventional distance function $F(\mathbf{y},\mathbf{v})$ defined for $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{v}$ positive, the relation

$$\mathbf{V}^*(\mathbf{y}) = \text{Closure}\ \{\mathbf{v} | \mathbf{v} \text{ positive}, F(\mathbf{y},\mathbf{v}) \geqq 1\} \tag{9}$$

defines indirect input requirement sets for $\mathbf{y} \in \mathbf{Y}^*$ which are input-conventional. If F is the distance function of an input-conventional technology with input requirement sets $\mathbf{V}(\mathbf{y})$, then $\mathbf{V}^*(\mathbf{y}) = \mathbf{V}(\mathbf{y})$ for $\mathbf{y} \in \mathbf{Y}^*$.

Proof: The first two steps of the proof verify that F exists and is input-conventional. Step 3 verifies that $\mathbf{V}^*(\mathbf{y})$ defined by (9) is input-conventional. Step 4 verifies the last result of the lemma, $\mathbf{V}^*(\mathbf{y}) = \mathbf{V}(\mathbf{y})$.

Step 1. We first show that $F(\mathbf{y},\mathbf{v})$ exists. Since $\mathbf{V}(\mathbf{y})$ is non-empty for

$\mathbf{y} \in \mathbf{Y}^*$ and free disposal holds, there is for each positive $\mathbf{v}$ some positive scalar λ' such that $(1/\lambda')\mathbf{v}$ is at least as large in every component as some fixed vector in $\mathbf{V}(\mathbf{y})$. Then, $(1/\lambda')\mathbf{v} \in \mathbf{V}(\mathbf{y})$. If $\mathbf{0} \in \mathbf{V}(\mathbf{y})$ in the case $\mathbf{y} = \mathbf{0}$, then $F(\mathbf{y},\mathbf{v}) = +\infty$ by definition. Suppose $\mathbf{0} \notin \mathbf{V}(\mathbf{y})$. Since $\mathbf{V}(\mathbf{y})$ is closed by hypothesis, there is an upper bound on the set of λ satisfying $(1/\lambda)\mathbf{v} \in \mathbf{V}(\mathbf{y})$, and the maximum in (8) is attained.

Step 2. That F is positive and positively linear homogeneous in positive $\mathbf{v}$ for each $\mathbf{y} \in \mathbf{Y}^*$ follows directly from (8). To show that F is non-decreasing in $\mathbf{v}$, note that if $\mathbf{v}^0,\mathbf{v}^1$ are positive input bundles with $\mathbf{v}^1 \geqq \mathbf{v}^0$, then $\mathbf{v}^1/F(\mathbf{y},\mathbf{v}^0) \geqq \mathbf{v}^0/F(\mathbf{y},\mathbf{v}^0) \in \mathbf{V}(\mathbf{y})$, implying $\mathbf{v}^1/F(\mathbf{y},\mathbf{v}^0) \in \mathbf{V}(\mathbf{y})$ by free disposal. Hence, $F(\mathbf{y},\mathbf{v}^1)/F(\mathbf{y},\mathbf{v}^0)) \geqq 1$, and $F(\mathbf{y},\mathbf{v}^1) \geqq F(\mathbf{y},\mathbf{v}^0)$ by the positive linear homogeneity of F. To show F concave in $\mathbf{v}$, it is sufficient (because of linear homogeneity) to show for any positive $\mathbf{v}^0$ and $\mathbf{v}^1$ that $F(\mathbf{y},\mathbf{v}^0 + \mathbf{v}^1) \geqq F(\mathbf{y},\mathbf{v}^0) + F(\mathbf{y},\mathbf{v}^1)$. Since $\mathbf{v}^i/F(\mathbf{y},\mathbf{v}^i) \in \mathbf{V}(\mathbf{y})$ for $i = 0,1$, the convexity of $\mathbf{V}(\mathbf{y})$ implies

$$\alpha \frac{\mathbf{v}^0}{F(\mathbf{y},\mathbf{v}^0)} + (1-\alpha)\frac{\mathbf{v}^1}{F(\mathbf{y},\mathbf{v}^1)} \in \mathbf{V}(\mathbf{y}),$$

for any α satisfying $0 \leqq \alpha \leqq 1$. In particular, for $\alpha = F(\mathbf{y},\mathbf{v}^0)/[F(\mathbf{y},\mathbf{v}^0) + F(\mathbf{y},\mathbf{v}^1)]$, one obtains

$$\frac{\mathbf{v}^0 + \mathbf{v}^1}{F(\mathbf{y},\mathbf{v}^0) + F(\mathbf{y},\mathbf{v}^1)} \in \mathbf{V}(\mathbf{y}),$$

implying

$$F\left[\mathbf{y}, \frac{\mathbf{v}^0 + \mathbf{v}^1}{F(\mathbf{y},\mathbf{v}^0) + F(\mathbf{y},\mathbf{v}^1)}\right] \geqq 1.$$

By linear homogeneity, $F(\mathbf{y},\mathbf{v}^0 + \mathbf{v}^1) \geqq F(\mathbf{y},\mathbf{v}^0) + F(\mathbf{y},\mathbf{v}^1)$. The continuity of F in positive $\mathbf{v}$ is an implication of concavity. This verifies that F is input-conventional.

Step 3. Suppose $F(\mathbf{y},\mathbf{v})$ defined for $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{v}$ positive is input-conventional. Consider the indirect input requirement sets $\mathbf{V}^*(\mathbf{y})$ defined by (9). If $F(\mathbf{y},\mathbf{v}) = +\infty$, then $\mathbf{y}=\mathbf{0}$ and $\mathbf{V}^*(\mathbf{0})$ is the non-negative orthant. Consider $F(\mathbf{y},\mathbf{v}) < +\infty$. From (9), the $\mathbf{V}^*(\mathbf{y})$ are closed. By the positive linear homogeneity of F, $\mathbf{0} \notin \mathbf{V}^*(\mathbf{y})$. Since F is concave, the contour set $\mathbf{V}^*(\mathbf{y})$ is convex. Since F is non-decreasing in $\mathbf{v}$, the free disposal condition is satisfied by $\mathbf{V}^*(\mathbf{y})$. Hence, the indirectly defined technology is input-conventional.

Step 4. By (9), if $\mathbf{v}$ is positive, then $\mathbf{v} \in \mathbf{V}(\mathbf{y})$ if and only if $F(\mathbf{y},\mathbf{v}) \geqq 1$, and hence if and only if $\mathbf{v} \in \mathbf{V}^*(\mathbf{y})$. Since $\mathbf{V}(\mathbf{y})$ and $\mathbf{V}^*(\mathbf{y})$ are closed and the convexity of $\mathbf{V}(\mathbf{y})$ implies that it equals the closure of its interior, the equality $\mathbf{V}^*(\mathbf{y}) = \mathbf{V}(\mathbf{y})$ follows. Q.E.D.

It is sometimes useful to extend the definition of the distance function to all non-negative input bundles $\mathbf{v}$ by applying the formula (8) provided $\mathbf{v}/\lambda$ is in $\mathbf{V}(\mathbf{y})$ for some positive scalar λ, and setting $F(\mathbf{y},\mathbf{v}) = 0$ otherwise. Appealing to the arguments used to establish Lemma 1, one can show that this extended distance function is a positively linear homogeneous, non-decreasing, concave, continuous function of non-negative $\mathbf{v}$ for each $\mathbf{y} \in \mathbf{Y}^*$ when the hypotheses of Lemma 4 hold. In applications, it is sometimes useful to employ this extended definition of the distance function.

We can now restate the duality conditions of Lemmas 2 and 3 in terms of the distance function. This form of the duality theorem is due to Shephard (1970), who has made an exhaustive examination of the implications of the resulting formal mathematical duality.

Lemma 5. Consider (a) the family of input-conventional cost structures and (b) the family of input-conventional distance functions. For a cost structure $C(\mathbf{y},\mathbf{r})$, $\mathbf{y} \in \mathbf{Y}^*$, in family (a), define a technology mapping

$$F(\mathbf{y},\mathbf{v}) = \operatorname{Max}\{\lambda > 0 | \mathbf{r}\cdot\mathbf{v} \geqq \lambda C(\mathbf{y},\mathbf{r}) \text{ for all } \mathbf{r} \text{ positive}\}. \tag{10}$$

For a distance function $F(\mathbf{y},\mathbf{v})$, $\mathbf{y} \in \mathbf{Y}^*$, in the family (b), define a cost mapping $C(\mathbf{y},\mathbf{r}) = 0$ if $\mathbf{y} = 0$, and for $\mathbf{y} \neq 0$,

$$C(\mathbf{y},\mathbf{r}) = \operatorname{Max}\{\lambda > 0 | \mathbf{r}\cdot\mathbf{v} \geqq \lambda F(\mathbf{y},\mathbf{v}) \text{ for all } \mathbf{v} \text{ positive}\}. \tag{11}$$

Then, the function $F(\mathbf{y},\mathbf{v})$ defined by (10) is in family (b), and the function $C(\mathbf{y},\mathbf{r})$ defined by (11) is in family (a). The technology mapping (10) is equivalent to application of the mapping (5) to obtain implicit input requirement sets, and application of the mapping (8) to these sets to obtain a distance function. The cost mapping (11) is equivalent to application of the mapping (9) to obtain indirect input requirement sets, and application of the mapping (1) to these sets to obtain a cost function. Hence, the technology and cost mappings (10) and (11) are mutual inverses on the families (a) and (b).

Corollary. For all positive $\mathbf{r}$ and $\mathbf{v}$,

$$F(\mathbf{y},\mathbf{v})C(\mathbf{y},\mathbf{r}) \leqq \mathbf{r}\cdot\mathbf{v},$$

with equality if and only if $\mathbf{v}$ is a cost minimizing input vector for the argument $(\mathbf{y},\mathbf{r})$.

Proof: The first step of the proof shows that the mapping (10) is the composition of the mappings (5) and (8). The second step shows that mapping (11) is the composition of the mappings (9) and (1). Then, Lemmas 2–4 will establish the implications of this lemma.

Step 1. Suppose an input-conventional cost structure $C(\mathbf{y},\mathbf{r})$, $\mathbf{y} \in \mathbf{Y}^*$, is given. The mapping (5) defines implicit input requirement sets $\mathbf{V}^*(\mathbf{y})$ with $\mathbf{v} \in \mathbf{V}^*(\mathbf{y})$ if and only if $\mathbf{r}\cdot\mathbf{v} \geqq C(\mathbf{y},\mathbf{r})$ for all positive $\mathbf{r}$. The mapping (8) defines an implicit distance function

$$F^*(\mathbf{y},\mathbf{v}) = \operatorname{Max}\left\{\lambda > 0 \,\middle|\, \frac{1}{\lambda}\mathbf{v} \in \mathbf{V}^*(\mathbf{y})\right\}$$
$$= \operatorname{Max}\left\{\lambda > 0 \,\middle|\, \mathbf{r}\cdot\left(\frac{1}{\lambda}\mathbf{v}\right) \geqq C(\mathbf{y},\mathbf{r}) \text{ for all } \mathbf{r} \text{ positive}\right\}.$$

But this is the technology mapping (10), and $F^*(\mathbf{y},\mathbf{v}) = F(\mathbf{y},\mathbf{v})$.

Step 2. Given an input-conventional distance function $F(\mathbf{y},\mathbf{v})$, $\mathbf{y} \in \mathbf{Y}^*$, the mapping (9) defines indirect input requirement sets $\mathbf{V}^*(\mathbf{y})$, and the mapping (1) defines a minimum cost function $C^*(\mathbf{y},\mathbf{r})$ for these indirect input requirement sets. We need consider only $\mathbf{y} \neq \mathbf{0}$.

$$C^*(\mathbf{y},\mathbf{r}) = \operatorname{Min}\{\mathbf{r}\cdot\mathbf{v} \mid \mathbf{v} \in \mathbf{V}^*(\mathbf{y})\}$$
$$= \operatorname{Max}\{\lambda > 0 \mid \mathbf{r}\cdot\mathbf{v} \geqq \lambda \text{ for all } \mathbf{v} \in \mathbf{V}^*(\mathbf{y})\}$$
$$= \operatorname{Max}\{\lambda > 0 \mid \mathbf{r}\cdot\mathbf{v} \geqq \lambda \text{ for all } \mathbf{v} \in \mathbf{V}^*(\mathbf{y}),\ \mathbf{v} \text{ positive}\}.$$

Now, for all positive $\mathbf{v}$, $\mathbf{v}/F(\mathbf{y},\mathbf{v}) \in \mathbf{V}^*(\mathbf{y})$. Further, $\mathbf{v} \in \mathbf{V}^*(\mathbf{y})$ implies $F(\mathbf{y},\mathbf{v}) \geqq 1$, and hence $\mathbf{r}\cdot\mathbf{v} \geqq \mathbf{r}\cdot\mathbf{v}/F(\mathbf{y},\mathbf{v})$. Therefore,

$$C^*(\mathbf{y},\mathbf{r}) = \operatorname{Max}\{\lambda > 0 \mid \mathbf{r}\cdot\mathbf{v}/F(\mathbf{y},\mathbf{v}) \geqq \lambda \text{ for all } \mathbf{v} \text{ positive}\}.$$

But this is the cost mapping (11), and $C^*(\mathbf{y},\mathbf{r}) = C(\mathbf{y},\mathbf{r})$. Q.E.D.

8. Extensions of Duality

The duality theorem established in Section 6 provides a basis for relating structural properties of production possibilities to structural properties of the cost function. In applications, it is useful to have a large family of duality relationships of the form: "the production possibility set has property 'P' if and only if the cost function has property 'Q'." Using the formal duality of cost and distance functions derived in the preceding section, we will be able to establish also the validity of such propositions with the properties "P" and "Q" interchanged. Through the remainder of this section, we shall assume that production possibility sets are described by distance functions, and that all cost structures and distance functions are input-conventional. We begin with a series of definitions.

A positive input bundle $\mathbf{v}$ is *efficient* for an output bundle $\mathbf{y}$ and distance function F if $F(\mathbf{y},\mathbf{v}) = 1$ and any distinct positive input bundle $\mathbf{v}'$ with $\mathbf{v}' \leqq \mathbf{v}$ has $F(\mathbf{y},\mathbf{v}') < 1$. Alternately, define an input bundle $\mathbf{v}$ to be *efficient* for an input requirement set $\mathbf{V}(\mathbf{y})$ if any distinct input bundle $\mathbf{v}'$ with $\mathbf{v}' \leqq \mathbf{v}$ has $\mathbf{v}' \notin \mathbf{V}(\mathbf{y})$. The reader can verify that for positive input bundles, these definitions of efficient input bundles are equivalent. In (a) of Figure 3, the points $\mathbf{v}^*$ and $\mathbf{v}^3$ are efficient, while $\mathbf{v}^2$ and $\mathbf{v}^4$ are not.

Recall that the distance function F is concave in $\mathbf{v}$, by (12) and linear homogeneity. Define F to be *strictly quasi-concave from below* if its upper contour sets $\{\mathbf{v} \in \mathbf{E}_+^N | F(\mathbf{y},\mathbf{v}) \geqq 1\}$ are strictly convex from below (see Assumption B-2) for all $\mathbf{y} \in \mathbf{Y}^*$. This property can be restated as requiring, for any positive, distinct points $\mathbf{v}^0$ and $\mathbf{v}^1$ and output $\mathbf{y} \in \mathbf{Y}^*$, that either (i) every plane which contains $\mathbf{v}^0/F(\mathbf{y},\mathbf{v}^0)$ and $\mathbf{v}^1/F(\mathbf{y},\mathbf{v}^1)$ and bounds $\{\mathbf{v} \in \mathbf{E}_+^N | F(\mathbf{y},\mathbf{v}) \geqq 1\}$ from below is parallel to a coordinate axis, or else (ii) for every weighted average $\mathbf{v}'' = \theta\mathbf{v}^0 + (1-\theta)\mathbf{v}^1$, with $0 < \theta < 1$.

$$F(\mathbf{y},\mathbf{v}'') > \text{Min}\{F(\mathbf{y},\mathbf{v}^0),\ F(\mathbf{y},\mathbf{v}^1)\}. \tag{12}$$

Figure 3 illustrates the geometry of this condition, which guarantees that the "efficient" boundary of each input requirement set is rotund, containing no "flat segments".

A stronger version of strict quasi-concavity from below will also be used. When the transformation function $F(\mathbf{y},\mathbf{v})$ is differentiable in the inputs, let $F_v(\mathbf{y},\mathbf{v})$ denote the vector of partial derivatives $F_n(\mathbf{y},\mathbf{v}) \equiv \partial F/\partial v_n$, $n = 1,\ldots,N$, evaluated at $(\mathbf{y},\mathbf{v})$. This vector is termed the *gradient* of F. Let $F_{vv}(\mathbf{y},\mathbf{v})$ denote the N-dimensional matrix of second partial derivatives $\partial^2 F/\partial v_n \partial v_m$, $n,m = 1,\ldots,N$, evaluated at $(\mathbf{y},\mathbf{v})$. This array is termed the *Hessian matrix* of F. The transformation function F is *strictly differentiably quasi-concave from below* in positive $\mathbf{v}$ if for any positive efficient $\mathbf{v}^0,\mathbf{v}^1$ and weighted average $\mathbf{v}'' = \theta\mathbf{v}^0 + (1-\theta)\mathbf{v}^1$, $0 < \theta < 1$, it follows that the Hessian matrix $F_{vv}(\mathbf{v}'')$ is negative semi-definite of rank $N-1$.

A remark on the relation of these definitions is in order. The conditions that F is concave and positively linear homogeneous in $\mathbf{v}$ imply that when the Hessian of F exists, it is symmetric, negative semi-definite, and singular, with a zero characteristic root corresponding to the characteristic vector $\mathbf{v}''$. Hence, strict differential quasi-concavity from below requires that the quadratic form,

$$Q(\mathbf{v},F_{vv}(\mathbf{v}'')) = \sum_{n=1}^{N}\sum_{m=1}^{N} v_n v_m F_{v_n v_m}(\mathbf{y},\mathbf{v}''), \tag{13}$$

be negative for any non-zero vector $\mathbf{v}$ not proportional to $\mathbf{v}''$. It is shown in Appendix A.3 that strict differential quasi-concavity from below implies strict quasi-concavity from below. As a partial converse it is shown that continuous second-order differentiability plus strict quasi-concavity from below implies that the condition of strict differential quasi-concavity from below holds on a subset of $\mathbf{v}$ which is open and dense[6] relative to the set of efficient $\mathbf{v}$.

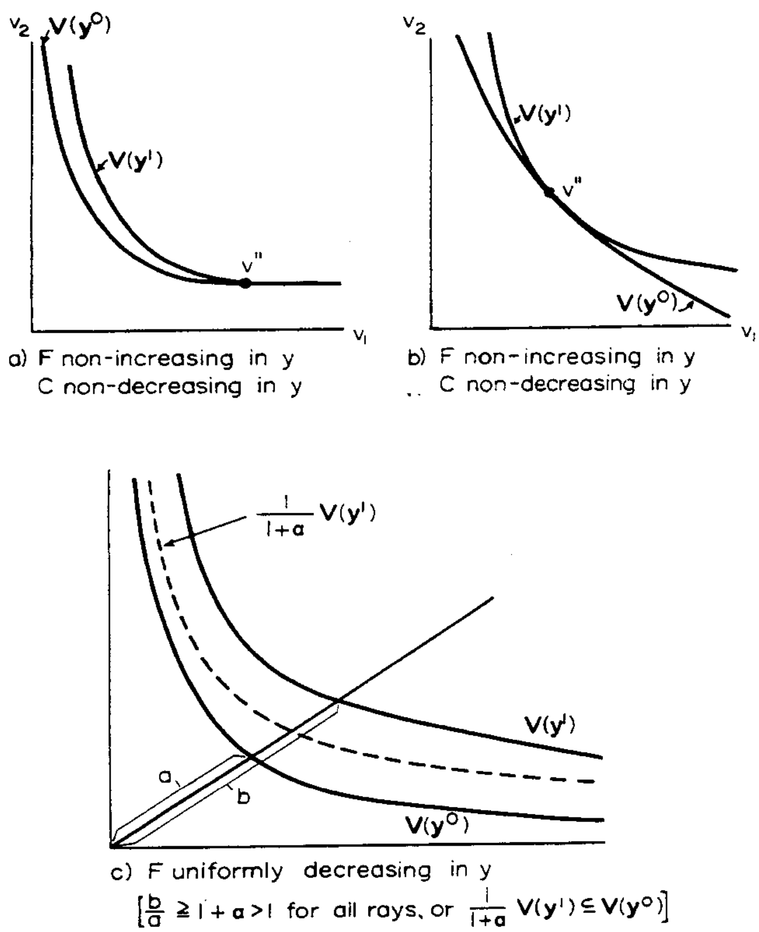
The distance function $F(\mathbf{y},\mathbf{v})$ is *non-increasing* in the output bundle $\mathbf{y}$ if for any $\mathbf{y}^0$, $\mathbf{y}^1 \in \mathbf{Y}^*$ with $\mathbf{y}^0 \leqq \mathbf{y}^1$, it follows that $F(\mathbf{y}^0,\mathbf{v}) \geqq F(\mathbf{y}^1,\mathbf{v})$. This property is equivalent to the condition on input requirement sets that $\mathbf{y}^0$, $\mathbf{y}^1 \in \mathbf{Y}^*$, $\mathbf{y}^0 \leqq \mathbf{y}^1$ implies that $\mathbf{V}(\mathbf{y}^1)$ is contained in $\mathbf{V}(\mathbf{y}^0)$. Similarly, the cost function $C(\mathbf{y},\mathbf{r})$ is non-decreasing in $\mathbf{y}$ if for any $\mathbf{y}^0$, $\mathbf{y}^1 \in \mathbf{Y}^*$ with $\mathbf{y}^0 \leqq \mathbf{y}^1$, it follows that $C(\mathbf{y}^0,\mathbf{r}) \leqq C(\mathbf{y}^1,\mathbf{r})$.

The distance function $F(\mathbf{y},\mathbf{v})$ is *uniformly decreasing* in the output bundle $\mathbf{y}$ if for any distinct $\mathbf{y}^0$, $\mathbf{y}^1 \in \mathbf{Y}^*$ with $\mathbf{y}^0 \leqq \mathbf{y}^1$, there exists a small positive scalar α such that $F(\mathbf{y}^0,\mathbf{v})/F(\mathbf{y}^1,\mathbf{v}) \geqq 1+\alpha$ for all positive $\mathbf{v}$. In terms of the input requirement sets, this condition is equivalent to the property that distinct $\mathbf{y}^0$, $\mathbf{y}^1 \in \mathbf{Y}^*$ with $\mathbf{y}^0 \leqq \mathbf{y}^1$ implies $\mathbf{V}(\mathbf{y}^1)$ a proper subset of $\mathbf{V}(\mathbf{y}^0)$, with each input bundle in $\mathbf{V}(\mathbf{y}^1)$ at least as large as a $(1+\alpha)$-multiple of an input bundle in $\mathbf{V}(\mathbf{y}^0)$. When the set of efficient input bundles in $\mathbf{V}(\mathbf{y}^0)$ is bounded, this condition reduces to the requirement that $\mathbf{V}(\mathbf{y}^1)$ not contain the efficient bundles in $\mathbf{V}(\mathbf{y}^0)$.

The cost function $C(\mathbf{y},\mathbf{r})$ is *uniformly increasing* in the output bundle $\mathbf{y}$ if for any distinct $\mathbf{y}^0$, $\mathbf{y}^1 \in \mathbf{Y}^*$ with $\mathbf{y}^0 \leqq \mathbf{y}^1$, there exists a small positive scalar α such that $C(\mathbf{y}^1,\mathbf{r})/C(\mathbf{y}^0,\mathbf{r}) > 1+\alpha$ for all positive $\mathbf{r}$.

The distance function $F(\mathbf{y},\mathbf{v})$ is *strongly upper semicontinuous* in $(\mathbf{y},\mathbf{v})$ if for any sequence $(\mathbf{y}^i,\mathbf{v}^i)$ with $\mathbf{y}^i \in \mathbf{Y}^*$ and $\mathbf{v}^i$ positive which converges to a point $(\mathbf{y}^0,\mathbf{v}^0)$, two properties hold: (a) If $F(\mathbf{y}^i,\mathbf{v}^*)$ is bounded away from zero for some positive $\mathbf{v}^*$, then $\mathbf{y}^0 \in \mathbf{Y}^*$. (b) If $\mathbf{y}^0 \in \mathbf{Y}^*$ and $\mathbf{v}^0$ is positive, then $F(\mathbf{y}^0,\mathbf{v}^0) \geqq \limsup_i F(\mathbf{y}^i,\mathbf{v}^i)$. The cost function $C(\mathbf{y},\mathbf{r})$ is *strongly lower semicontinuous* in $(\mathbf{y},\mathbf{r})$ if for any sequence $(\mathbf{y}^i,\mathbf{r}^i)$ with $\mathbf{y}^i \in \mathbf{Y}^*$ and $\mathbf{r}^i$ positive which converges to a point $(\mathbf{y}^0,\mathbf{r}^0)$, two properties

[6]A set is *open* if it contains a neighborhood of each point in the set, and is *dense* if every neighborhood contains some point of the set.



a) F non-increasing in y
C non-decreasing in y

b) F non-increasing in y
C non-decreasing in y

c) F uniformly decreasing in y
$\left[\frac{b}{a} \geq 1 + a > 1 \text{ for all rays, or } \frac{1}{1+a} V(y^1) \subseteq V(y^0)\right]$

FIGURE 6

hold: (a) If $C(\mathbf{y}^i,\mathbf{r}^*)$ is bounded for some positive $\mathbf{r}^*$, then $\mathbf{y}^0 \in \mathbf{Y}^*$. (b) If $\mathbf{y}^0 \in \mathbf{Y}^*$ and $\mathbf{r}^0$ is positive, then $C(\mathbf{y}^0,\mathbf{r}^0) \leqq \liminf_i C(\mathbf{y}^i,\mathbf{r}^i)$.

Figures 6 and 7 illustrate these concepts. In (a) of Figure 6 the cost of producing $\mathbf{y}^1$ exceeds the cost of producing $\mathbf{y}^0$ at any strictly positive prices. However, at $\mathbf{v}''$ one has $F(\mathbf{y}^0,\mathbf{v}'') = F(\mathbf{y}^1,\mathbf{v}'')$ and F is not strictly decreasing in $\mathbf{y}$. In (b), F is again not strictly decreasing in $\mathbf{y}$ at $(\mathbf{y}^0,\mathbf{v}'')$. At the price vector $\mathbf{r}''$ at which $\mathbf{v}''$ is optimal, C is not strictly increasing in $\mathbf{y}$. Both (a) and (b) of Figure 6 correspond to pathological technologies which are unlikely to arise in practice. (c) illustrates the assumption of uniform monotonicity. This condition requires that isoquants not converge (when the distance between them is measured along rays). In Figure 7, (a) illustrates upper semicontinuity of a function F. At the argument $\mathbf{y}^0$, the function takes the largest of the limiting values. In this

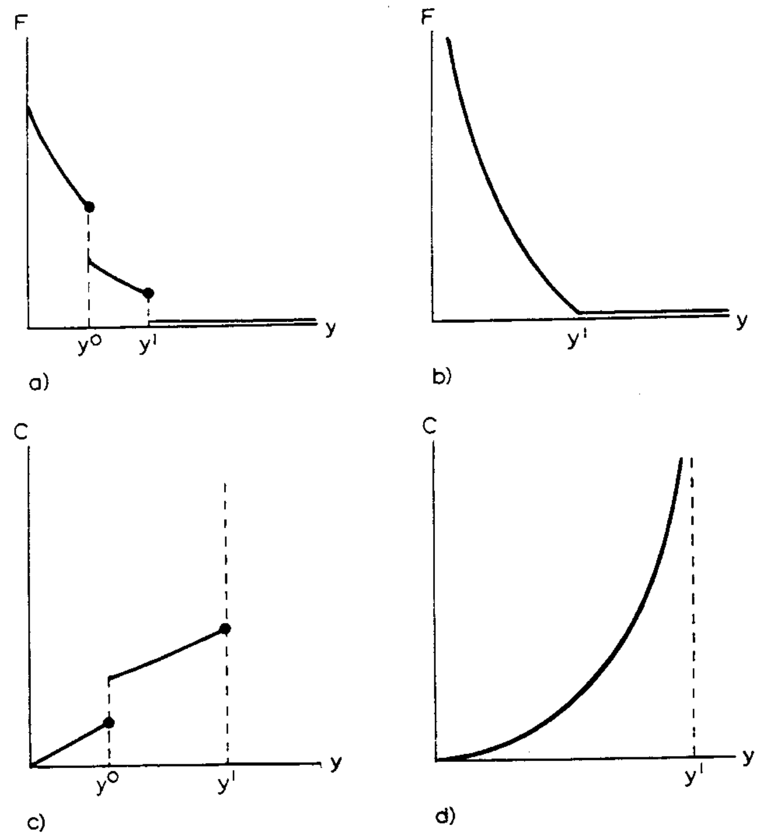


FIGURE 7

graph, F is bounded away from zero for y in the closed interval $[0,y^1]$; hence strong upper semicontinuity implies $\mathbf{Y}^* = [0,y^1]$. In (b), F approaches zero as y approaches y^1, implying $\mathbf{Y}^* = [0,y^1]$. [In general, $\mathbf{Y}^* = \{\mathbf{y}|F(\mathbf{y},\mathbf{v}) > 0$ for some $\mathbf{v} \gg 0\}$.] (c) of Figure 7 illustrates lower semicontinuity of a function C. At the argument y^0, C takes the smallest of the limiting values. At y^1, C is bounded, implying by strong lower semicontinuity that $\mathbf{Y}^* = [0,y^1]$. In (d), C is unbounded as y approaches y^1, implying $\mathbf{Y}^* = [0,y^1)$. The next result relates the strong upper semicontinuity of the distance function to a property of the production possibility set.

Lemma 6. Consider an input-conventional production possibility set $\mathbf{Y}$, and let $F(\mathbf{y},\mathbf{v})$ be its distance function, so that (8) and (9) hold. Then, the set $\mathbf{Y}$ is closed if and only if F is strongly upper semicontinuous in $(\mathbf{y},\mathbf{v})$.

Proof: First, suppose F is strongly upper semicontinuous in (y,v). Consider a sequence $(\mathbf{y}^i,\mathbf{v}^i)\in\mathbf{Y}$ with $(\mathbf{y}^i,\mathbf{v}^i)\rightarrow(\mathbf{y}^0,\mathbf{v}^0)$. Choose $\mathbf{v}^*$ strictly larger than $\mathbf{v}^0$. Then, for i large, $\mathbf{v}^i\leqq\mathbf{v}^*$, implying $F(\mathbf{y}^i,\mathbf{v}^*)\geqq 1$. This implies $\mathbf{y}^0\in\mathbf{Y}^*$. Let $\mathbf{w}$ be an arbitrarily small positive vector. Then, $(\mathbf{y}^i,\mathbf{v}^i+\mathbf{w})\in\mathbf{Y}$ and $(\mathbf{y}^i,\mathbf{v}^i+\mathbf{w})\rightarrow(\mathbf{y}^0,\mathbf{v}^0+\mathbf{w})$, implying, since $F(\mathbf{y}^i,\mathbf{v}^i+\mathbf{w})\geqq 1$, that $F(\mathbf{y}^0,\mathbf{v}^0+\mathbf{w})\geqq 1$. Letting $\mathbf{w}\rightarrow 0$, (9) implies $\mathbf{v}^0\in\mathbf{V}(\mathbf{y}^0)$, and hence $(\mathbf{y}^0,\mathbf{v}^0)\in\mathbf{Y}$.

Next, suppose $\mathbf{Y}$ is closed. Consider a sequence $(\mathbf{y}^i,\mathbf{v}^i)$ with $\mathbf{y}^i\in\mathbf{Y}^*$ and $\mathbf{v}^i$ positive which converges to a point $(\mathbf{y}^0,\mathbf{v}^0)$. Then, $(\mathbf{y}^i,\mathbf{v}^i/F(\mathbf{y}^i,\mathbf{v}^i))\in\mathbf{Y}$. If $F(\mathbf{y}^i,\mathbf{v}^i)$ is unbounded, then the closedness of $\mathbf{Y}$ implies $(\mathbf{y}^0,\mathbf{0})\in\mathbf{Y}$, implying $\mathbf{y}^0=\mathbf{0}$ and $F(\mathbf{y}^0,\mathbf{v}^0)=+\infty\geqq\lim_i F(\mathbf{y}^i,\mathbf{v}^i)$. Alternately, assume $F(\mathbf{y}^i,\mathbf{v}^i)$ bounded. Then $F(\mathbf{y}^i,\mathbf{v}^i)$ has a subsequence (retain notation) converging to a scalar α. If α is positive, it follows that $(\mathbf{y}^0,\mathbf{v}^0/\alpha)\in\mathbf{Y}$, implying $\mathbf{y}^0\in\mathbf{Y}^*$ and, if $\mathbf{v}^0$ is positive, $F(\mathbf{y}^0,\mathbf{v}^0)\geqq\alpha$. If α is zero, but $\mathbf{y}^0\in\mathbf{Y}^*$ and $\mathbf{v}^0$ positive, then $F(\mathbf{y}^0,\mathbf{v}^0)>0=\lim_i F(\mathbf{y}^i,\mathbf{v}^i)$ for the subsequence. In either case, the condition for strong upper semicontinuity of F is met. Q.E.D.

The following result relates properties of the distance function and the cost function.

> *Lemma* 7. Consider (a) the family of input-conventional cost structures $C(\mathbf{y},\mathbf{r})$, $\mathbf{y}\in\mathbf{Y}^*$, and (b) the family of input-conventional distance functions $F(\mathbf{y},\mathbf{v})$, $\mathbf{y}\in\mathbf{Y}^*$. Suppose these families are related by the mutually inverse technology and cost mappings (10) and (11). Then, in Table 1, the distance function has property "P" if and only if the cost structure has the corresponding property "Q".

Proof: A detailed proof of this lemma is tedious and of minimal inherent interest. Hence, only outlines of proofs will be given, and mathematically difficult points will be deferred to Appendix A.3. The steps of this proof correspond to the eight results in Table 1. In each step, we first show that "P" implies "Q", and then show that "Q" implies "P".

Step 1. Suppose F is non-increasing in y, so that $\mathbf{y}^0,\mathbf{y}^1\in\mathbf{Y}^*$ and $\mathbf{y}^0\leqq\mathbf{y}^1$ imply $F(\mathbf{y}^1,\mathbf{v})\leqq F(\mathbf{y}^0,\mathbf{v})$. By (11), for any positive price vector $\mathbf{r}$ and any $\epsilon>0$, there exists a positive vector $\mathbf{v}$ such that $C(\mathbf{y}^1,\mathbf{r})F(\mathbf{y}^1,\mathbf{v})\geqq(\mathbf{r}\cdot\mathbf{v})/(1+\epsilon)$. Further, $C(\mathbf{y}^0,\mathbf{r})F(\mathbf{y}^0,\mathbf{v})\leqq\mathbf{r}\cdot\mathbf{v}$. Hence, $C(\mathbf{y}^0,\mathbf{r})F(\mathbf{y}^0,\mathbf{v})\leqq$

TABLE 1

Property "P" holds for an input-conventional transformation function, $F(\mathbf{y},\mathbf{v})$, if and only if property "Q" holds for its input-conventional cost function, $C(\mathbf{y},\mathbf{r})$.[a]

	"P" on $F(\mathbf{y},\mathbf{v})$	"Q" on $C(\mathbf{y},\mathbf{r})$
1.	Non-increasing in **y**	Non-decreasing in **y**
2.	Uniformly decreasing in **y**	Uniformly increasing in **y**
3.[b]	Strongly upper semicontinuous in (**y**,**v**)	Strongly lower semicontinuous in (**y**,**r**)
4.[c]	Strongly lower semicontinuous in (**y**,**v**)	Strongly upper semicontinuous in (**y**,**r**)
5.[d]	Strongly continuous in (**y**,**v**)	Strongly continuous in (**y**,**r**)
6.[e]	Strictly quasi-concave from below in **v**	Continuously differentiable in positive **r**
7.[f]	Continuously differentiable in positive **v**	Strictly quasi-concave from below in **r**
8.[g]	Twice continuously differentiable and strictly differentiably quasi-concave from below in **v**	Twice continuously differentiable and strictly differentiably quasi-concave from below in **r**

[a]By the formal duality of cost and transformation functions, the implications of this table continue to hold when properties "P" and "Q" are reversed; i.e., "P" holds for the cost function and "Q" holds for the transformation function.

[b]Recall that this property is equivalent to the condition that the production possibility set be a closed set.

[c]Input requirement sets $\mathbf{V}(\mathbf{y})$ form a ***strongly lower hemicontinuous correspondence*** if two properties hold: (a) If $\mathbf{y}^i \in \mathbf{Y}^*$, $\mathbf{y}^i \to \mathbf{y}^0 \notin \mathbf{Y}^*$ and **A** is any bounded set in $\mathbf{E}^N$, then for sufficiently large *i*, $\mathbf{V}(\mathbf{y}^i)$ does not meet **A**. (b) If $\mathbf{y}^0 \in \mathbf{Y}^*$, $\mathbf{v}^0 \in \mathbf{V}(\mathbf{y}^0)$, and $\mathbf{y}^i \in \mathbf{Y}^*$ $\mathbf{y}^i \to \mathbf{y}^0$, then there exist $\mathbf{v}^i \in \mathbf{V}(\mathbf{y}^i)$ such that $\mathbf{v}^i \to \mathbf{v}^0$. This condition implies that the cost function is strongly upper semicontinuous in (**y**,**r**). To show this, note first that $\mathbf{y}^i \in \mathbf{Y}^*$, $\mathbf{y}^i \to \mathbf{y}^0 \notin \mathbf{Y}^*$ implies $C(\mathbf{y}^i,\mathbf{r}^*) \to +\infty$ for $\mathbf{r}^*$ positive. Hence, $C(\mathbf{y}^i,\mathbf{r}^*)$ bounded implies $\mathbf{y}^0 \in \mathbf{Y}^*$. Next, note that if $(\mathbf{y}^i,\mathbf{r}^i) \to (\mathbf{y}^0,\mathbf{r}^0)$ with $\mathbf{y}^i,\mathbf{y}^0 \in \mathbf{Y}^*$, and $\mathbf{r}^i,\mathbf{r}^0$ positive, there exists $\mathbf{v}^0 \in \mathbf{V}(\mathbf{y}^0)$ such that $C(\mathbf{y}^0,\mathbf{r}^0) = \mathbf{r}^0 \cdot \mathbf{v}^0$ and there exist $\mathbf{v}^i \in \mathbf{V}(\mathbf{y}^i)$ such that $\mathbf{v}^i \to \mathbf{v}^0$. Then $C(\mathbf{y}^i,\mathbf{r}^i) \leqq \mathbf{r}^i \cdot \mathbf{v}^i \to \mathbf{r}^0 \cdot \mathbf{v}^0$ implies $\limsup_i C(\mathbf{y}^i,\mathbf{r}^i) \leqq C(\mathbf{y}^0,\mathbf{r}^0)$. A more difficult argument, given in Appendix A.3. 15.5, establishes the converse implication from C to **V**, and consequently the equivalence of the condition that the distance function be strongly lower semicontinuous and the condition that the input requirement sets define a strongly lower hemicontinuous correspondence.

[d]A function is strongly continuous if it is strongly upper and strongly lower semicontinuous. This property is equivalent to a requirement that the input requirement sets $\mathbf{V}(\mathbf{y})$ define a strongly continuous correspondence (Appendix A.3, 13.2).

[e]This property guarantees that isoquants are rotund, with no flat segments.

[f]This property guarantees that isoquants have no "kinks".

[g]An input-conventional transformation function with these properties is termed *neoclassical*. This result then provides a formal duality theorem for neoclassical distance functions and neoclassical cost functions.

$(1+\epsilon)C(\mathbf{y}^1,\mathbf{r})F(\mathbf{y}^1,\mathbf{v})$, or

$$\frac{C(\mathbf{y}^0,\mathbf{r})}{C(\mathbf{y}^1,\mathbf{r})} \leqq (1+\epsilon)\frac{F(\mathbf{y}^1,\mathbf{v})}{F(\mathbf{y}^0,\mathbf{v})} \leqq 1+\epsilon, \tag{14}$$

implying $C(\mathbf{y}^0,\mathbf{r}) \leqq C(\mathbf{y}^1,\mathbf{r})$.

Next suppose C is non-decreasing in $\mathbf{y}$, so that $\mathbf{y}^0,\mathbf{y}^1 \in \mathbf{Y}^*$ and $\mathbf{y}^0 \leqq \mathbf{y}^1$ imply $C(\mathbf{y}^0,\mathbf{r}) \leqq C(\mathbf{y}^1,\mathbf{r})$. Analogously to the preceding argument, (10) implies, for any positive $\mathbf{v}$ and any $\epsilon > 0$, the existence of $\mathbf{r}$ such that $C(\mathbf{y}^0,\mathbf{r})F(\mathbf{y}^0,\mathbf{v}) \geqq (\mathbf{r}\cdot\mathbf{v})/(1+\epsilon)$. Since $C(\mathbf{y}^1,\mathbf{r})F(\mathbf{y}^1,\mathbf{v}) \leqq \mathbf{r}\cdot\mathbf{v}$,

$$\frac{F(\mathbf{y}^1,\mathbf{v})}{F(\mathbf{y}^0,\mathbf{v})} \leqq (1+\epsilon)\frac{C(\mathbf{y}^0,\mathbf{r})}{C(\mathbf{y}^1,\mathbf{r})} \leqq 1+\epsilon, \tag{15}$$

implying $F(\mathbf{y}^0,\mathbf{v}) \geqq F(\mathbf{y}^1,\mathbf{v})$.

Step 2. Suppose F is uniformly decreasing in $\mathbf{y}$, so that distinct $\mathbf{y}^0,\mathbf{y}^1 \in \mathbf{Y}^*$ with $\mathbf{y}^0 \leqq \mathbf{y}^1$ imply $F(\mathbf{y}^0,\mathbf{v})/F(\mathbf{y}^1,\mathbf{v}) \geqq 1+\alpha$ for some positive scalar α, uniformly in $\mathbf{v}$. In (14), this implies $C(\mathbf{y}^1,\mathbf{r})/C(\mathbf{y}^0,\mathbf{r}) \geqq 1+\alpha$ uniformly in $\mathbf{r}$. Conversely, suppose C is uniformly increasing in $\mathbf{y}$. Then, a similar argument applied to (15) yields the result that F is uniformly decreasing in $\mathbf{y}$.

Step 3. Suppose F is strongly upper semicontinuous. By Lemmas 4 and 6, the indirectly defined production possibility set $\mathbf{Y}$ is closed. Consider a sequence $(\mathbf{y}^i,\mathbf{r}^i)$ with $\mathbf{y}^i \in \mathbf{Y}^*$, $\mathbf{r}^i$ positive which converges to a point $(\mathbf{y}^0,\mathbf{r}^0)$. Then there exist $\mathbf{v}^i \in \mathbf{V}(\mathbf{y}^i)$ such that $C(\mathbf{y}^i,\mathbf{r}^i) = \mathbf{r}^i\cdot\mathbf{v}^i$. If $C(\mathbf{y}^i,\mathbf{r}^i)$ is bounded and $\mathbf{r}^0$ is positive, this equality implies that $\mathbf{v}^i$ is bounded and has at least one limit point $\mathbf{v}^0$. The closedness of $\mathbf{Y}$ implies $(\mathbf{y}^0,\mathbf{v}^0) \in \mathbf{Y}$. Hence, $\mathbf{y}^0 \in \mathbf{Y}^*$ and $C(\mathbf{y}^0,\mathbf{r}^0) \leqq \mathbf{r}^0\cdot\mathbf{v}^0$. Since this inequality holds for each limit point, $C(\mathbf{y}^0,\mathbf{r}^0) \leqq \liminf_i C(\mathbf{y}^i,\mathbf{r}^i)$. This establishes condition (a) for C to be lower semicontinuous, and condition (b) in the case that $C(\mathbf{y}^i,\mathbf{r}^i)$ has a bounded subsequence. Finally, if $C(\mathbf{y}^i,\mathbf{r}^i)$ has no bounded subsequence, but $\mathbf{y}^0 \in \mathbf{Y}^*$ and $\mathbf{r}^0$ is positive, then obviously $C(\mathbf{y}^0,\mathbf{r}^0) \leqq \liminf_i C(\mathbf{y}^i,\mathbf{r}^i)$. Hence, C is strongly lower semicontinuous.

Next, suppose C is strongly lower semicontinuous. By (5), we have $\mathbf{V}(\mathbf{y}) = \{\mathbf{v} \geqq 0 \mid \mathbf{r}\cdot\mathbf{v} \geqq C(\mathbf{y},\mathbf{r})$ for all positive $\mathbf{r}\}$ for $\mathbf{y} \in \mathbf{Y}^*$ and by (6) a production possibility set $\mathbf{Y}$. Consider a sequence $(\mathbf{y}^i,\mathbf{v}^i) \in \mathbf{Y}$ converging to a point $(\mathbf{y}^0,\mathbf{v}^0)$. For each positive $\mathbf{r}$, $\mathbf{r}\cdot\mathbf{v}^0 = \lim_i \mathbf{r}\cdot\mathbf{v}^i \geqq \lim_i C(\mathbf{y}^i,\mathbf{r})$, implying $\mathbf{y}^0 \in \mathbf{Y}^*$ and $\lim_i C(\mathbf{y}^i,\mathbf{r}) \geqq C(\mathbf{y}^0,\mathbf{r})$ by strong lower semicontinuity. This implies $\mathbf{v}^0 \in \mathbf{V}(\mathbf{y}^0)$, and hence $(\mathbf{y}^0,\mathbf{v}^0) \in \mathbf{Y}$. Therefore, $\mathbf{Y}$ is closed, and Lemma 6 implies that F is strongly upper semicontinuous.

Step 4. Utilizing the formal duality of F and C, properties "P" and "Q" in Step 3 can be reversed to yield result 4.

Step 5. This result is implied by the results 3 and 4.

Step 6. Note that a concave function which is differentiable on an open set is continuously differentiable on that set, and that the negative of a concave function is a convex function. Then, a lengthy argument given in the Appendix A.3, 16.7(7) and 16.7(10), yields this result.

Step 7. This result is implied by result 6 using the formal duality of C and F.

Step 8. This result is established in the Appendix A.3, 16.7(11). Q.E.D.

One implication of the duality theory developed above is that the input requirement sets have image sets in the space of input prices, defined for $\mathbf{y} \in \mathbf{Y}^*$ by

$$\mathbf{R}(\mathbf{y}) = \{\mathbf{r} \geqq \mathbf{0} | \mathbf{r} \cdot \mathbf{v} \geqq F(\mathbf{y},\mathbf{v}) \text{ for all positive } \mathbf{v}\} \\ = \text{Closure}\,\{\mathbf{r} | \mathbf{r} \text{ positive}, C(\mathbf{y},\mathbf{r}) \geqq 1\}. \tag{16}$$

This set is termed the *factor price requirement set*, and its boundary is termed the *factor price frontier*, for the output bundle $\mathbf{y}$. This concept has been employed in applications by Samuelson (1953–54), Bruno (1968), and others.

The properties of the cost function – concavity, monotonicity, linear homogeneity, and continuity – imply that the factor price requirement set $\mathbf{R}(\mathbf{y})$ is closed, is non-empty for $\mathbf{y} \neq \mathbf{0}$, and satisfies the free disposal and convexity assumptions A and B. Therefore, there is a formal mathematical duality between input requirement sets $\mathbf{V}(\mathbf{y})$ and factor price requirement sets $\mathbf{R}(\mathbf{y})$; they are termed *polar reciprocal sets*, and can be characterized directly by the relationship $\mathbf{r} \cdot \mathbf{v} \geqq 1$ for all $\mathbf{r} \in \mathbf{R}(\mathbf{y})$ and $\mathbf{v} \in \mathbf{V}(\mathbf{y})$.

The factor price frontier is a solution $r_1 = c(r_2,\ldots,r_n,\mathbf{y})$, of the equation $C(\mathbf{y},\mathbf{r}) = 1$. The frontier c is a convex, non-increasing function of $(r_2,\ldots,r_n)$, and a non-increasing function of $\mathbf{y}$. In the case of a single output, the factor price frontier is usually defined for unit output, $r_1 = c(r_2,\ldots,r_n,1)$. When the technology exhibits constant returns to scale, it is completely determined once the input requirement set for unit output is specified. Then duality implies that an input-conventional constant returns technology is completely characterized by the factor price frontier $r_1 = c(r_2,\ldots,r_n,1)$.

9. Cobb–Douglas and C.E.S. Cost Functions

In econometric applications of production theory, one normally works with parametric families of transformation or distance functions. Cobb–Douglas and C.E.S. (or, Arrow–Chenery–Minhas–Solow) production functions are widely used cases. Cost functions are derived in this section for these two families. Dual functions for other parametric families are derived elsewhere in this volume (Diewert, Chapter III.2; Hanoch, Chapter II.3; Lau, Chapter I.3).

Consider a technology with N inputs, $\mathbf{v} = (v_1,...,v_N)$, producing a single output y. The technology is of the Cobb–Douglas form if it has the distance function

$$F(y,\mathbf{v}) = Dv_1^{\theta_1}v_2^{\theta_2}\cdots v_N^{\theta_N}/\gamma(y), \tag{17}$$

where D is a positive efficiency parameter, the θ_i are positive distribution parameters satisfying $\theta_1 + \theta_2 + \cdots + \theta_N = 1$, and γ is a function from a subset $\mathbf{Y}^*$ of the non-negative real line onto the non-negative real line. In case $\gamma(y)$ has the special form $\gamma(y) = y^{1/\mu}$, production possibilities exhibit returns to scale of degree μ. The cost function obtained by applying (1) to the technology defined by (17) has the functional form

$$C(y,\mathbf{r}) = D^*\gamma(y)r_1^{\theta_1}r_2^{\theta_2}\cdots r_N^{\theta_N}, \tag{18}$$

where $D^* = D^{-1}\theta_1^{-\theta_1}\theta_2^{-\theta_2}\cdots\theta_N^{-\theta_N}$, and is called the *Cobb–Douglas cost function.*

The technology is of the C.E.S. form if it has the distance function

$$F(y,\mathbf{v}) = [(v_1/D_1(y))^{1-1/\sigma} + (v_2/D_2(y))^{1-1/\sigma} + \cdots + (v_N/D_N(y))^{1-1/\sigma}]^{1/(1-1/\sigma)}, \tag{19}$$

where σ is a positive elasticity of substitution parameter, $\sigma \neq 1$, and the $D_i(y)$ are positive (non-decreasing) functions of positive y. The cost function obtained by applying (1) to the technology defined by (19) has the functional form

$$C(y,\mathbf{r}) = [(r_1D_1(y))^{1-\sigma} + (r_2D_2(y))^{1-\sigma} + \cdots + (r_ND_N(y))^{1-\sigma}]^{1/(1-\sigma)}, \tag{20}$$

and is called the C.E.S. cost function.

Two limiting cases of the C.E.S. transformation function are most easily treated separately. In the limit $\sigma \rightarrow 0$, one obtains the Leontief transformation function

$$F(y,\mathbf{v}) = \text{Min}\{(v_1/D_1(y)), (v_2/D_2(y)),...,(v_N/D_N(y))\}, \tag{21}$$

which has the corresponding cost function

$$C(y,\mathbf{r}) = r_1D_1(y) + r_2D_2(y) + \cdots + r_ND_N(y). \tag{22}$$

Alternately, in the limit $\sigma \to +\infty$, one obtains the perfect substitute transformation function

$$F(y,\mathbf{v}) = (v_1/D_1(y)) + (v_2/D_2(y)) + \cdots + (v_N/D_N(y)), \tag{23}$$

which has the corresponding cost function

$$C(y,\mathbf{r}) = \text{Min}\{(r_1D_1(y)), (r_2D_2(y)),\ldots,(r_ND_N(y))\}. \tag{24}$$

These formulae can be verified by indirect methods (Lau, Chapter I.3), or by direct computation of the minimizing input bundle. For the C.E.S. case, the steps in the direct computation are the following: (1) obtain as a first-order condition for cost minimization the expression $r_i/r_j = (v_i/v_j)^{-1/\sigma}(D_i(y)/D_j(y))^{1/\sigma-1}$; (2) reverse this expression to obtain the expression $r_iv_i/r_jv_j = (r_iD_i(y)/r_jD_j(y))^{1-\sigma}$; (3) sum this expression over i to obtain $r_jv_j/C(y,\mathbf{r}) = (r_jD_j(y))^{1-\sigma}/[(r_1D_1(y))^{1-\sigma} + \cdots + (r_ND_N(y))^{1-\sigma}]$; (4) solve this expression for v_j, substitute the result into (19) with $F(y,\mathbf{v}) = 1$, and simplify to obtain (20).

10. The Geometry of Two-Input Cost Functions

Dual distance and cost functions have a geometric structure which can be used to establish qualitative relationships between these functions. Consider the case of two inputs $\mathbf{v} = (v_1,v_2)$, and suppose production possibilities are defined by input-conventional input requirement sets $\mathbf{V(y)}$, $\mathbf{y} \in \mathbf{Y}^*$. Let $F(\mathbf{y},\mathbf{v})$ and $C(\mathbf{y},\mathbf{r})$ denote the transformation and cost functions, respectively, for this technology, and let $\mathbf{R(y)} = \{\mathbf{r} \geqq \mathbf{0} | C(\mathbf{y},\mathbf{r}) \geqq 1\}$ denote the factor price requirement set.

Figure 8 illustrates a typical input requirement set $\mathbf{V(y)}$ and corresponding factor price requirement set $\mathbf{R(y)}$. Hereafter, we shall refer to the boundaries of these sets as the isoquant and the factor price frontier, respectively. Let $\mathbf{v}^0$ denote an input bundle in the isoquant, and let $\mathbf{r}^0$ be a normal to a plane tangent to $\mathbf{V(y)}$ at $\mathbf{v}^0$. Choose the magnitude of $\mathbf{r}^0$ to make $\mathbf{r}^0 \cdot \mathbf{v}^0 = 1$. Using Lemma 5 and the derivative property of the cost function, one can conclude that $\mathbf{r}^0$ is in the factor price frontier, and that $\mathbf{v}^0$ is a normal to a plane tangent to $\mathbf{R(y)}$ at $\mathbf{r}^0$. Furthermore, this geometric relationship is completely dual: starting from $\mathbf{r}^0$ in the factor price frontier, one can proceed in the opposite direction to locate $\mathbf{v}^0$ in

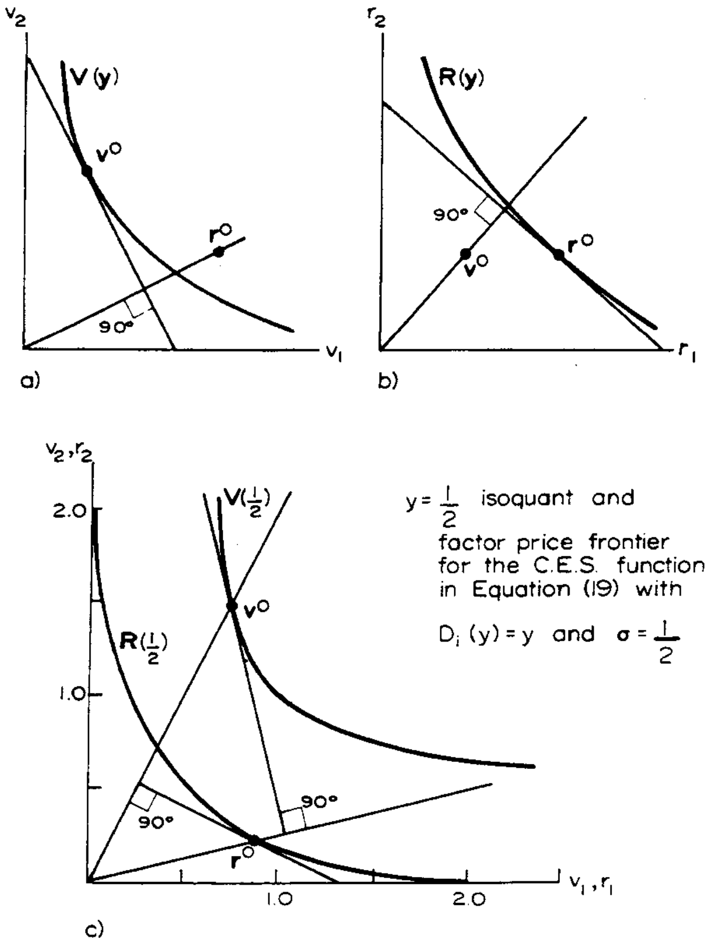


FIGURE 8

the isoquant. In Figure 8, the mapping between points in the isoquant and factor price frontier is one to one, and as $\mathbf{v}^0$ moves from northwest to southeast along the isoquant, its image $\mathbf{r}^0$ moves from southeast to northwest along the factor price frontier.

These movements correspond to a rise in the relative intensity of use of factor 1 and a rise in the marginal rate of substitution of factor 1 per unit of factor 2 (i.e., a rise in the relative price of factor 2). Thus, the value of a factor rises as its relative scarcity rises.

Employing this geometric mapping rule, we can establish a simple inverse relationship between the degree of curvature of the isoquant and the degree of curvature of the factor price frontier, as illustrated in Figure 9. Curves A, B, C, D denote dual isoquants and factor price frontiers. A straight line isoquant (A) maps into a rectangular factor

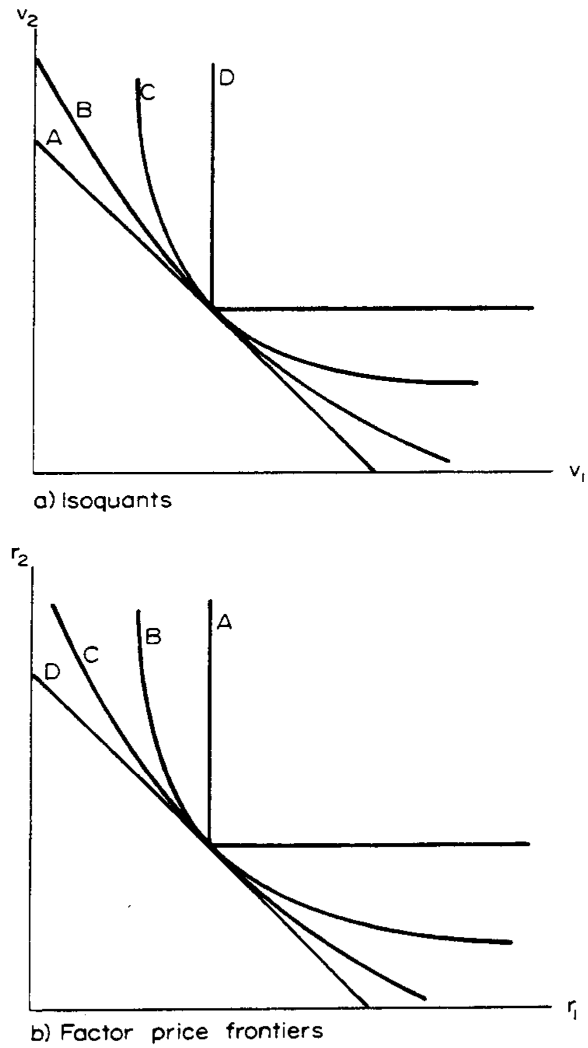


FIGURE 9

price frontier, and a rectangular isoquant (D) maps into a linear factor price frontier. Isoquant (C), with a sharper curvature than isoquant (B), maps into the factor price frontier with less sharp curvature. Using the elasticity of factor substitution as an index of curvature, this inverse curvature relationship can be made quantitative. Assume the distance function to be twice continuously differentiable and strictly differentiably quasi-concave from below. Then, the cost function also has these properties, by Lemma 7, and the dual points $\mathbf{v}^0$, $\mathbf{r}^0$ in Figure 8 satisfy

$$r_1^0/r_2^0 = F_1(\mathbf{y},\mathbf{v}^0)/F_2(\mathbf{y},\mathbf{v}^0), \tag{25}$$

and

$$v_1^0/v_2^0 = C_1(\mathbf{y},\mathbf{r}^0)/C_2(\mathbf{y},\mathbf{r}^0), \tag{26}$$

where $F_i(\mathbf{y},\mathbf{r}^0)$ denotes the partial derivative $\partial F(\mathbf{y},\mathbf{v}^0)/\partial v_i$, and $C_i(\mathbf{y},\mathbf{v}^0)$ denotes $\partial C(\mathbf{y},\mathbf{r}^0)/\partial r_i$. Define the elasticity of input substitution at $(\mathbf{y},\mathbf{v}^0)$,

$$\sigma(\mathbf{y},\mathbf{v}^0) = -\left.\frac{\mathrm{d}\log(v_1^0/v_2^0)}{\mathrm{d}\log(r_1^0/r_2^0)}\right|_{\mathbf{y}\text{ fixed and }F(\mathbf{y},\mathbf{v}^0)=1} \tag{27}$$

From (a) in Figure 8, (r_1^0/r_2^0) falls as (v_1^0/v_2^0) rises, at a rate which increases in magnitude as the curvature of the isoquant rises. Then, $\sigma(\mathbf{y},\mathbf{v}^0)$ is positive, is near zero if the isoquant has high curvature and is nearly rectangular, and is near infinity if the isoquant has low curvature and is nearly linear. A formula for the elasticity can be obtained by logarithmic differentiation of (25):

$$\begin{aligned}
\mathrm{d}\ln\frac{r_1}{r_2} &= \left[\frac{F_{11}}{F_1} - \frac{F_{21}}{F_2}\right]\mathrm{d}v_1 + \left[\frac{F_{12}}{F_1} - \frac{F_{22}}{F_2}\right]\mathrm{d}v_2 \\
&= \left[-\frac{v_2F_{12}}{v_1F_1} - \frac{F_{21}}{F_2}\right]\mathrm{d}v_1 + \left[\frac{F_{12}}{F_1} + \frac{v_1F_{12}}{v_2F_2}\right]\mathrm{d}v_2 \\
&= -\frac{F_{12}}{F_1F_2}\left[\frac{F}{v_1}\mathrm{d}v_1 - \frac{F}{v_2}\mathrm{d}v_2\right] = -\frac{FF_{12}}{F_1F_2}\mathrm{d}\ln\frac{v_1}{v_2}.
\end{aligned}$$

The second equation uses the homogeneity conditions $v_1F_{11} + v_2F_{12} = 0$ and $v_1F_{12} + v_2F_{22} = 0$, while the third uses the condition $F = v_1F_1 + v_2F_2$. Substituting this formula in (27) yields

$$\sigma(\mathbf{y},\mathbf{v}^0) = \frac{F_1(\mathbf{y},\mathbf{v}^0)F_2(\mathbf{y},\mathbf{v}^0)}{F(\mathbf{y},\mathbf{v}^0)F_{12}(\mathbf{y},\mathbf{v}^0)}.$$

Alternately, logarithmic differentiation of (26) yields

$$\begin{aligned}
\mathrm{d}\ln\frac{v_1}{v_2} &= \left[\frac{C_{11}}{C_1} - \frac{C_{21}}{C_2}\right]\mathrm{d}r_1 + \left[\frac{C_{12}}{C_1} - \frac{C_{22}}{C_2}\right]\mathrm{d}r_2 \\
&= \left[-\frac{r_2C_{12}}{r_1C_1} - \frac{C_{21}}{C_2}\right]\mathrm{d}r_1 + \left[\frac{C_{12}}{C_1} + \frac{r_1C_{21}}{r_2C_2}\right]\mathrm{d}r_2 \\
&= -\frac{CC_{12}}{C_1C_2}\mathrm{d}\ln\frac{r_1}{r_2},
\end{aligned}$$

where the same homogeneity arguments are used as in the preceding derivation. Then,

$$\sigma(\mathbf{y},\mathbf{v}^0) = \frac{C(\mathbf{y},\mathbf{r}^0)C_{12}(\mathbf{y},\mathbf{r}^0)}{C_1(\mathbf{y},\mathbf{r}^0)C_2(\mathbf{y},\mathbf{r}^0)},$$

where $\mathbf{r}^0$ is the vector dual to $\mathbf{v}^0$ [i.e., $\mathbf{r}^0 = F_v(\mathbf{y},\mathbf{v}^0)/F(\mathbf{y},\mathbf{v}^0)$].

Define a similar curvature index for the factor price frontier at $(\mathbf{y},\mathbf{r}^0)$,

$$\rho(\mathbf{y},\mathbf{r}^0) = -\left.\frac{\mathrm{d}\log(r_1^0/r_2^0)}{\mathrm{d}\log(v_1^0/v_2^0)}\right|_{\mathbf{y}\text{ fixed and }C(\mathbf{y},\mathbf{r}^0)=1} \tag{28}$$

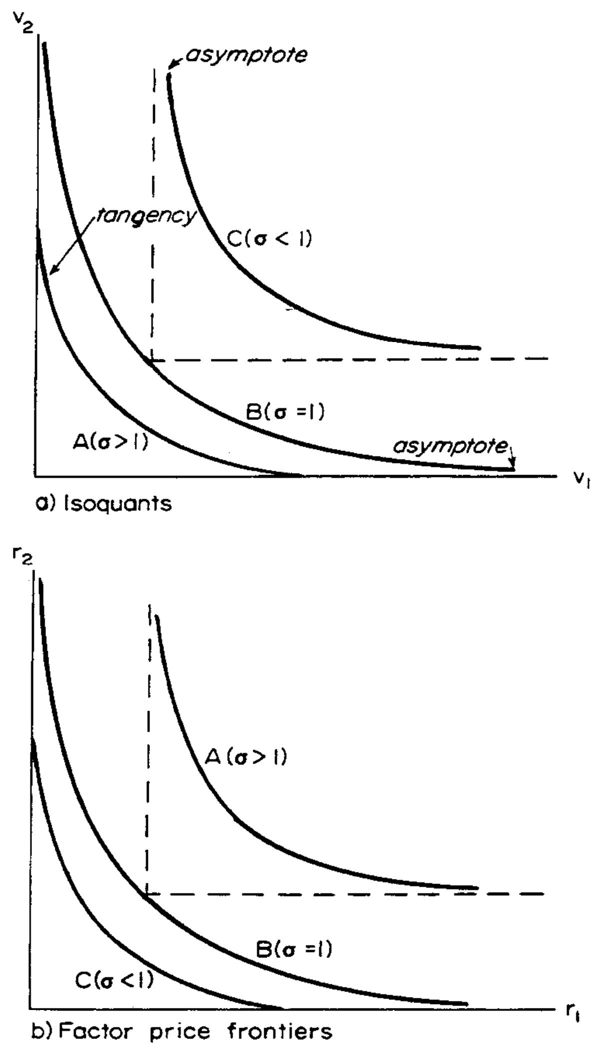


FIGURE 10

Then we obtain in the same manner as above the formula

$$\rho(\mathbf{y},\mathbf{r}^0) = C_1(\mathbf{y},\mathbf{r}^0)C_2(\mathbf{y},\mathbf{r}^0)/C(\mathbf{y},\mathbf{r}^0)C_{12}(\mathbf{y},\mathbf{r}^0).$$

Comparing the formulae for $\sigma(\mathbf{y},\mathbf{v}^0)$ and $\rho(\mathbf{y},\mathbf{r}^0)$, we obtain the condition $\rho(\mathbf{y},\mathbf{r}^0) = 1/\sigma(\mathbf{y},\mathbf{v}^0)$. Thus, an isoquant with an elasticity of substitution equal to one is dual to a factor price frontier with a curvature index $\rho(\mathbf{y},\mathbf{r}^0)$ equal to one, and an isoquant with an elasticity of substitution less (greater) than one has a factor price frontier with a curvature index greater (less) than one. Figure 10 illustrates this relationship for C.E.S. isoquants in (a) with an elasticity greater than one (A) and an elasticity less than one (C), and a Cobb–Douglas isoquant with an elasticity equal to one (B). The corresponding factor price frontiers are given in (b) of Figure 10.

Figure 11 illustrates the mapping of Figure 8 when there is a "kink" in the isoquant at $\mathbf{v}^0$. The image of this point is a line segment in the factor price frontier from $\mathbf{r}^0$ to $\mathbf{r}^1$. Any vector $\mathbf{r}$ in this line segment is a normal to a plane "supporting" the input requirement set at $\mathbf{v}^0$. Then, $\mathbf{r}^0$ and $\mathbf{r}^1$ are normals to the extreme supporting planes, as illustrated. Proceeding in the opposite direction, we note that each $\mathbf{r}$ in the line segment $\mathbf{r}^0$ to $\mathbf{r}^1$ in the factor price frontier has the same normal vector $\mathbf{v}^0$, and hence maps into the "kink" $\mathbf{v}^0$. Since, by duality, we can interchange $\mathbf{r}$ and $\mathbf{v}$ in this figure, we can show that flat segments in the isoquant map into "kinks" in the factor price frontier. Thus, we can conclude generally that "kinks" (or, lack of differentiability) in one function map into "flats" (or, lack of strict quasi-concavity) in the dual function, and vice versa. In the special case of an activity analysis model[7] of the technology, this duality is complete, with each "kink" ("flat") in an isoquant mapping into a "flat" ("kink") in the factor price frontier.

Our discussion of the geometry of two-factor cost functions will be concluded with an examination of the behavior of isoquants and factor price frontiers near the boundaries of the non-negative orthant. Five classes of boundary behavior can be distinguished:

A. The curve is asymptotic to an axis.
B. The curve is asymptotic to a line parallel to an axis.
C. The curve is tangent to an axis.
D. The curve meets an axis, but is not tangent to the axis.

[7]An input requirement set $\mathbf{V}(\mathbf{y})$ comes from an activity analysis model if it can be obtained from a *finite* set of input vectors by forming convex combinations and/or using free disposal of inputs.

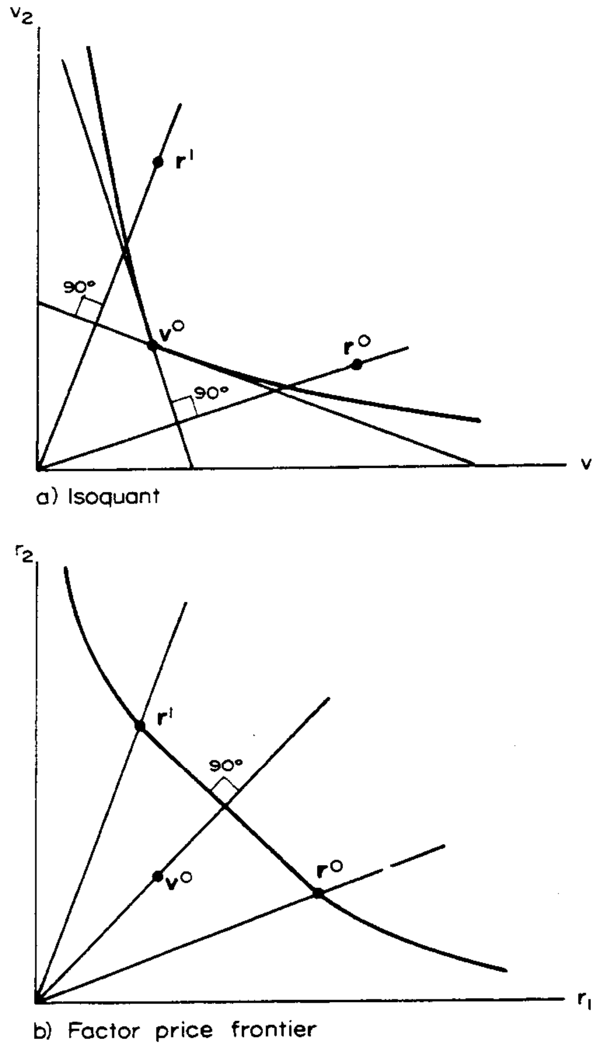


FIGURE 11

E. The curve meets and does not extend beyond a line parallel to an axis.

Figures 12 and 13 illustrate these classes of behavior, and the following geometric duality relationships between then:

1. A curve satisfies A on one axis if and only if the dual curve satisfies A on the other axis.
2. A curve satisfies B on one axis if and only if the dual curve satisfies C on the other axis.
3. A curve satisfies D on one axis if and only if the dual curve satisfies E on the other axis.

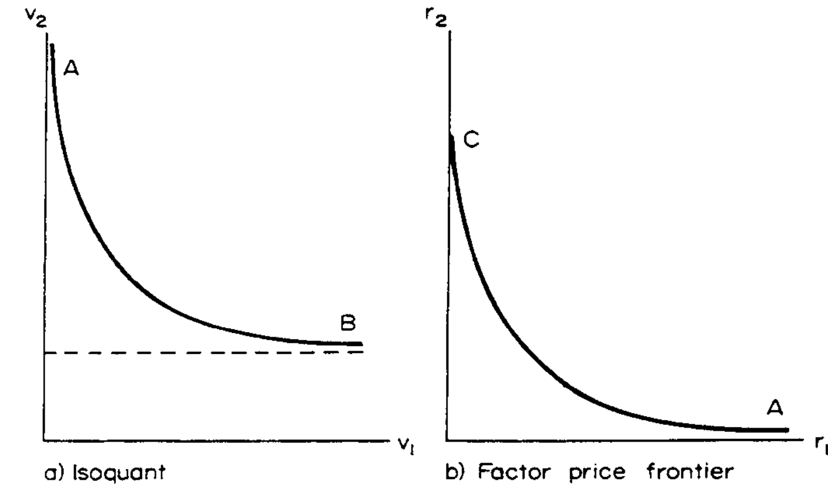


FIGURE 12

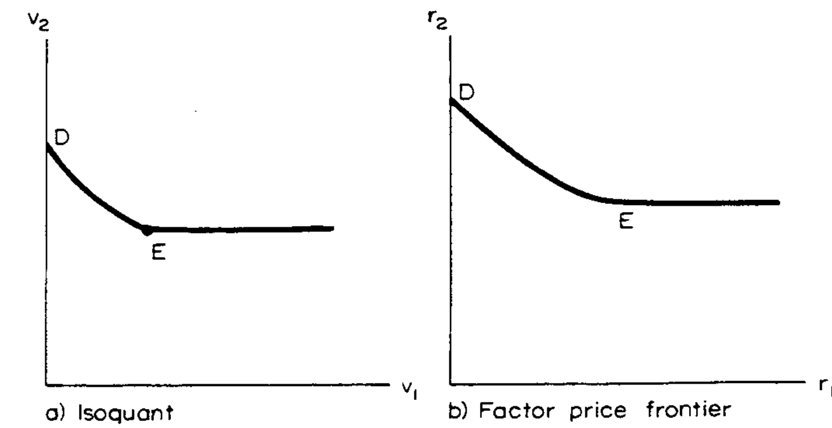


FIGURE 13

11. Comparative Statics for the Cost Minimizing Firm

The basic qualitative questions in the theory of the cost minimizing firm, as formulated by Samuelson (1947, p. 59) are the effects on an input demand of a change in its own price, in the price of another input, or in the output bundle, and the effects on total cost and marginal costs of changes in input prices or the output bundle.

We have noted in Section 5 on the derivative property of the cost function that for an input-regular production possibility set, the cost function has first and second derivatives with respect to input prices for almost all positive input prices. Since these first derivatives equal the cost minimizing demands when they exist, concavity of the cost function implies that an input demand function is non-increasing in its own price, and that the matrix of partial derivatives of inputs with respect to input prices is negative semi-definite and symmetric. It should be emphasized that these results hold with only the weak input-regular conditions imposed on production possibilities. In particular, some inputs may be non-divisible, or "isoquants" may fail to be convex, without altering this conclusion. This observation was first noticed by Samuelson (1953, p. 359), and first deduced formally in an economic application by McKenzie (1957).

For further comparative statics results, we shall for the remainder of this section impose classical assumptions on production possibilities: the technology is input-conventional and can be represented by a distance function $F(\mathbf{y},\mathbf{v})$, which is strongly continuous in $(\mathbf{y},\mathbf{v})$, twice continuously differentiable in $(\mathbf{y},\mathbf{v})$, uniformly decreasing in $\mathbf{y}$, and strictly differentiably quasi-concave from below in $\mathbf{v}$. (We term a technology satisfying these conditions *input-classical.*) Lemma 7 then implies that the cost function $C(\mathbf{y},\mathbf{r})$ is strongly continuous in $(\mathbf{y},\mathbf{r})$, twice continuously differentiable in $\mathbf{r}$, uniformly increasing in $\mathbf{y}$, and strictly differentiably quasi-concave from below in $\mathbf{r}$. A classical calculus argument using the implicit function theorem establishes that $C(\mathbf{y},\mathbf{r})$ is continuously differentiable in $\mathbf{y}$.[8] Under these conditions the input demands $v_i = D^i(\mathbf{y},\mathbf{r}) = C_i(\mathbf{y},\mathbf{r})$ are continuously differentiable in $(\mathbf{y},\mathbf{r})$, with a negative own price effect

$$\partial v_i/\partial r_i = D^i_i(\mathbf{y},\mathbf{r}) = C_{ii}(\mathbf{y},\mathbf{r}) < 0, \tag{29}$$

and symmetric cross-price effects

$$\partial v_i/\partial r_j = C_{ij}(\mathbf{y},\mathbf{r}) = C_{ji}(\mathbf{y},\mathbf{r}) = \partial v_j/\partial r_i. \tag{30}$$

[8]The cost function satisfies $C(\mathbf{y},\mathbf{r}) = \min_{\mathbf{v}} \mathbf{r}\cdot\mathbf{v}$ subject to $F(\mathbf{y},\mathbf{v}) = 1$. For $\mathbf{r}$ such that the minimum is achieved at strictly positive $\mathbf{v}$, the first-order conditions for minimization are $\lambda F_v(\mathbf{y},\mathbf{v}) = \mathbf{r}$ and $F(\mathbf{y},\mathbf{v}) = 1$, where λ is a Lagrangian multiplier. From the assumptions on F, these equations have a total differential which is continuous in $\mathbf{y}$.

$$\begin{bmatrix} \lambda F_{vv}(\mathbf{y},\mathbf{v}) & F_v(\mathbf{y},\mathbf{v}) \\ F_v(\mathbf{y},\mathbf{v}) & 0 \end{bmatrix} \begin{bmatrix} d\mathbf{v} \\ d\lambda \end{bmatrix} = \begin{bmatrix} -\lambda F_{vy} \\ -F_y \end{bmatrix} d\mathbf{y}.$$

The left-hand-side matrix is non-singular by the assumption of strict differential quasi-concavity of F in $\mathbf{v}$. Therefore, dv/dy exists and is continuous in $(\mathbf{y},\mathbf{r})$, implying $C(\mathbf{y},\mathbf{r}) = \mathbf{r}\cdot\mathbf{v}$ continuously differentiable in $\mathbf{y}$.

The matrix of price effects $[\partial v_i/\partial r_j] = [C_{ij}(\mathbf{y},\mathbf{r})]$ is symmetric, negative semi-definite, and of rank $N-1$, with

$$r_1\partial v_i/\partial r_1 + r_2\partial v_i/\partial r_2 + \cdots + r_N\partial v_i/\partial r_N = 0. \tag{31}$$

Inputs i and j are termed *substitutes* if $\partial v_i/\partial r_j > 0$, and *complements* if $\partial v_i/\partial r_j < 0$.

The effect on input i of an increase in output k is given by

$$\partial v_i/\partial y_k = C_{iy_k}(\mathbf{y},\mathbf{r}) = C_{y_ki}(\mathbf{y},\mathbf{r}) = \partial m_k/\partial r_i, \tag{32}$$

where $m_k = M^k(\mathbf{y},\mathbf{r}) = C_{y_k}(\mathbf{y},\mathbf{r})$ is the marginal cost of producing output k. Input i is termed *normal* for output k at $(\mathbf{y},\mathbf{r})$ if $\partial v_i/\partial y_k$ is positive, and is termed *regressive* for output k otherwise. Equation (32) shows that the marginal cost of output k rises when the price of a normal input rises, but falls when the price of a regressive input rises.

Since the cost function is uniformly increasing in $\mathbf{y}$, the marginal cost of output k, $M^k(\mathbf{y},\mathbf{r})$, is non-negative, and is positive for almost all y_k, given any values for the remaining arguments. The effect on total cost of an increase in input price i is non-negative, and is positive when the demand for input i is positive, since $C_i(\mathbf{y},\mathbf{r}) = v_i$.

Next, we examine the effects of output changes on marginal costs, $\partial m_k/\partial y_l = C_{y_ky_l}(\mathbf{y},\mathbf{r})$. Outputs k and l are termed *substitutes* if $\partial m_k/\partial y_l > 0$ and *complements* if $\partial m_k/\partial y_l < 0$. A production possibility set $\mathbf{Y}$ is said to exhibit *generally non-increasing returns* if $\mathbf{Y}$ is a convex set. We say that $\mathbf{Y}$ exhibits *eventually diminishing returns to scale* if $(\lambda\mathbf{y},\lambda\mathbf{v}) \in \mathbf{Y}$ for all $\lambda > 0$ implies $\mathbf{y} = 0$.

A cost function $C(\mathbf{y},\mathbf{r})$ is said to exhibit *generally non-decreasing costs* if C is a convex function of $\mathbf{y}$ for each positive $\mathbf{r}$. We say $C(\mathbf{y},\mathbf{r})$ exhibits *eventually increasing costs* if $\lim_{\lambda\to\infty} C(\lambda\mathbf{y},\mathbf{r})/\lambda = +\infty$ for all positive $\mathbf{r}$ and all $\mathbf{y} \neq \mathbf{0}$. [A competitive profit maximum exists for all strictly positive output prices if and only if $C(\mathbf{y},\mathbf{r})$ exhibits eventually increasing costs.]

Lemma 8. Assume the production possibility set $\mathbf{Y}$ to be input-conventional. Then the following implications hold:

(i) $\mathbf{Y}$ exhibits generally non-increasing returns if and only if $C(\mathbf{y},\mathbf{r})$ exhibits generally non-decreasing costs.

(ii) $\mathbf{Y}$ exhibits eventually diminishing returns to scale if and only if $C(\mathbf{y},\mathbf{r})$ exhibits eventually increasing costs.

Proof: (i) If $\mathbf{Y}$ is convex, and costs are minimized for $(\mathbf{y}^i,\mathbf{r}^0)$ at a bundle $\mathbf{v}^i$ with $(\mathbf{y}^i,\mathbf{v}^i)\in\mathbf{Y}$, then for $0<\theta<1$ and $(\mathbf{y}^0,\mathbf{v}^0)=\theta(\mathbf{y}^1,\mathbf{v}^1)+(1-\theta)(\mathbf{y}^2,\mathbf{v}^2)\in\mathbf{Y}$, we have

$$C(\mathbf{y}^0,\mathbf{r}^0)\leqq \mathbf{r}^0\cdot\mathbf{v}^0=\theta\mathbf{r}^0\cdot\mathbf{v}^1+(1-\theta)\mathbf{r}^0\cdot\mathbf{v}^2=\theta C(\mathbf{y}^1,\mathbf{r}^0)+(1-\theta)C(\mathbf{y}^2,\mathbf{r}^0).$$

Hence C is a convex function of $\mathbf{y}$.

Alternately, suppose $C(\mathbf{y},\mathbf{r})$ convex in $\mathbf{y}$ for fixed $\mathbf{r}$. Given $(\mathbf{y}^i,\mathbf{v}^i)\in\mathbf{Y}$ and $(\mathbf{y}^0,\mathbf{v}^0)=\theta(\mathbf{y}^1,\mathbf{v}^1)+(1-\theta)(\mathbf{y}^2,\mathbf{v}^2)$ for $0<\theta<1$, we have for any positive $\mathbf{r}$,

$$C(\mathbf{y}^0,\mathbf{r})\leqq\theta C(\mathbf{y}^1,\mathbf{r})+(1-\theta)C(\mathbf{y}^2,\mathbf{r})\leqq\theta\,\mathbf{r}\cdot\mathbf{v}^1+(1-\theta)\mathbf{r}\cdot\mathbf{v}^2=\mathbf{r}\cdot\mathbf{v}^0,$$

implying by Lemma 3 that $(\mathbf{y}^0,\mathbf{v}^0)\in\mathbf{Y}$. Hence $\mathbf{Y}$ is convex.

(ii) Suppose that for some $\mathbf{y}\neq\mathbf{0}$ and positive $\mathbf{r}$, $C(\lambda\mathbf{y},\mathbf{r})/\lambda$ fails to converge to $+\infty$ as $\lambda\to+\infty$. Then there exists a sequence $\lambda_i\to\infty$ such that $\{C(\lambda_i\mathbf{y},\mathbf{r})/\lambda_i\}$ is bounded. Let $\mathbf{v}^i$ be such that $C(\lambda_i\mathbf{y},\mathbf{r})=\mathbf{r}\cdot\mathbf{v}^i$. Then $\{\mathbf{v}^i/\lambda_i\}$ is bounded, and we can choose $\mathbf{v}'$ such that $(\mathbf{v}_i/\lambda_i)\leqq\mathbf{v}'$ for all i. Then, $(\lambda_i\mathbf{y},\mathbf{v}^i)\in\mathbf{Y}$ implies $(\lambda_i\mathbf{y},\lambda_i\mathbf{v}')\in\mathbf{Y}$, and the production possibility set fails to exhibit eventually diminishing returns to scale.

Alternately, suppose there exists $(\mathbf{y},\mathbf{v})\in\mathbf{Y}$, $\mathbf{y}\neq\mathbf{0}$, and $\lambda_i\to+\infty$ such that $(\lambda_i\mathbf{y},\lambda_i\mathbf{v})\in\mathbf{Y}$. Then $C(\lambda_i\mathbf{y},\mathbf{r})/\lambda_i\leqq\mathbf{r}\cdot(\lambda_i\mathbf{v})/\lambda_i=\mathbf{r}\cdot\mathbf{v}$, and C fails to exhibit eventually increasing cost. Q.E.D.

12. Composition of Distance and Cost Functions

For some simple parametric families of distance and cost functions, such as the Cobb–Douglas and C.E.S. cases analyzed in Section 9, it is possible to perform the cost and technology mappings constructively. However, many applications require more complex parametric specifications. One method of forming such functions is to build them up from simple functions for which the duality mappings are known. The primary result of this section gives a series of rules for the composition of these functions and the implications for their duals.

Theorem 9. Consider a producible output set $\mathbf{Y}^*$, and input-conventional input requirement sets $\mathbf{V}^j(\mathbf{y})\subseteq\mathbf{E}^N_+$, defined for $\mathbf{y}\in\mathbf{Y}^*$ and $j=1,\ldots,J$. Also, let $\mathbf{V}^*(\mathbf{y})\subseteq\mathbf{E}^J_+$ be an input-conventional input requirement set for $\mathbf{y}\in\mathbf{Y}^*$. Let $F^j(\mathbf{y},\mathbf{v})$ and $F^*(\mathbf{y},\mathbf{z})$ be the distance functions, and $C^j(\mathbf{y},\mathbf{r})$ and $C^*(\mathbf{y},\mathbf{q})$ be the cost functions, for $\mathbf{V}^j(\mathbf{y})$

TABLE 2

Composition rules for distance functions (property P), cost functions (property Q), input requirement sets (property S), and factor price requirement sets (property T).

---

1. *Neutral Scaling*[a]

For an arbitrary positive real-valued function $\alpha(\mathbf{y})$ defined in $\mathbf{Y}^*$,

P: $F^0(\mathbf{y},\mathbf{v}) = F^1(\mathbf{y},\mathbf{v})/\alpha(\mathbf{y}) = F^1(\mathbf{y},\mathbf{v}/\alpha(\mathbf{y}))$

Q: $C^0(\mathbf{y},\mathbf{r}) = \alpha(\mathbf{y})C^1(\mathbf{y},\mathbf{r}) = C^1(\mathbf{y},\alpha(\mathbf{y})\mathbf{r})$

S: $\mathbf{V}^0(\mathbf{y}) = \alpha(\mathbf{y})\mathbf{V}^1(\mathbf{y}) = \{\alpha(\mathbf{y})\mathbf{v}|\mathbf{v} \in \mathbf{V}^1(\mathbf{y})\}$

T: $\mathbf{R}^0(\mathbf{y}) = \mathbf{R}^1(\mathbf{y})/\alpha(\mathbf{y}) = \{\mathbf{r}/\alpha(\mathbf{y})|\mathbf{r} \in \mathbf{R}^1(\mathbf{y})\}$

---

2. *Non-neutral Scaling*[b]

For an arbitrary diagonal N-dimensional matrix $\mathbf{A}(\mathbf{y})$, where the diagonal elements of $\mathbf{A}(\mathbf{y})$ are positive real-valued functions defined in $\mathbf{Y}^*$,

P: $F^0(\mathbf{y},\mathbf{v}) = F^1(\mathbf{y},\mathbf{A}(\mathbf{y})^{-1}\mathbf{v})$

Q: $C^0(\mathbf{y},\mathbf{r}) = C^1(\mathbf{y},\mathbf{A}(\mathbf{y})\mathbf{r})$

S: $\mathbf{V}^0(\mathbf{y}) = \{\mathbf{A}(\mathbf{y})\mathbf{v}|\mathbf{v} \in \mathbf{V}^1(\mathbf{y})\}$

T: $\mathbf{R}^0(\mathbf{y}) = \{\mathbf{A}(\mathbf{y})^{-1}\mathbf{r}|\mathbf{r} \in \mathbf{R}^1(\mathbf{y})\}$

---

3. *Union of Input Requirement Sets*

P: $F^0(\mathbf{y},\mathbf{v}) = \mathrm{Sup}\left\{\sum_{j=1}^{J} F^j(\mathbf{y},\mathbf{v}^j)|\mathbf{v}^j \text{ positive}, \sum_{j=1}^{J} \mathbf{v}^j = \mathbf{v}\right\}$

Q: $C^0(\mathbf{y},\mathbf{r}) = \mathrm{Min}_{j=1,\dots,J} C^j(\mathbf{y},\mathbf{r})$

S: $\mathbf{V}^0(\mathbf{y}) = \text{Convex hull of } \bigcup_{j=1}^{J} \mathbf{V}^j(\mathbf{y})$

T: $\mathbf{R}^0(\mathbf{y}) = \bigcap_{j=1}^{J} \mathbf{R}^j(\mathbf{y})$

---

4. *Intersection of Input Requirement Sets*

P: $F^0(\mathbf{y},\mathbf{v}) = \mathrm{Min}_{j=1,\dots,J} F^j(\mathbf{y},\mathbf{v})$

Q: $C^0(\mathbf{y},\mathbf{r}) = \mathrm{Sup}\left\{\sum_{j=1}^{J} C^j(\mathbf{y},\mathbf{r}^j)|\mathbf{r}^j \text{ positive}, \sum_{j=1}^{J} \mathbf{r}^j = \mathbf{r}\right\}$

S: $\mathbf{V}^0(\mathbf{y}) = \bigcap_{j=1}^{J} \mathbf{V}^j(\mathbf{y})$

T: $\mathbf{R}^0(\mathbf{y}) = \text{Convex hull of } \bigcup_{j=1}^{J} \mathbf{R}^j(\mathbf{y})$

---

[a]The function $\alpha(\mathbf{y})$ may depend upon exogenous factors such as technical change, and may be independent of $\mathbf{y}$. It is convenient to include the value $\alpha(\mathbf{y}) = 0$ in this rule by defining $\mathbf{V}^0(\mathbf{y}) = \mathbf{E}_+^N$, $F^0(\mathbf{y},\mathbf{v}) = +\infty$, $C^0(\mathbf{y},\mathbf{r}) = 0$, and $\mathbf{R}^0(\mathbf{y}) = \emptyset$. Then Rule 1 holds in the limit as $\alpha(\mathbf{y}) \to 0$. Note that for $\mathbf{V}^0(\mathbf{y})$ to be input-conventional in this case, one must have $\mathbf{y} = \mathbf{0}$.

[b]The matrix $\mathbf{A}(\mathbf{y})$ may depend upon exogenous variables such as technical change.

TABLE 2 (*continued*)

5. *Summation of Input Requirement Sets*[c]

P: $F^0(\mathbf{y},\mathbf{v}) = \text{Sup}\left\{\text{Min}_{j=1,\ldots,J} F^j(\mathbf{y},\mathbf{v}^j) | \mathbf{v}^j \text{ positive}, \sum_{j=1}^{J} \mathbf{v}^j = \mathbf{v}\right\}$

Q: $C^0(\mathbf{y},\mathbf{r}) = \sum_{j=1}^{J} C^j(\mathbf{y},\mathbf{r})$

S: $\mathbf{V}^0(\mathbf{y}) = \sum_{j=1}^{J} \mathbf{V}^j(\mathbf{y})$

T: $\mathbf{R}^0(\mathbf{y}) = \bigcup_{\substack{z_j \geq 0 \\ \sum_{j=1}^{J} z_j = 1}} \bigcap_{j=1}^{J} z_j \mathbf{R}^j(\mathbf{y})$

6. *Convolution of Input Requirement Sets*[d]

P: $F^0(\mathbf{y},\mathbf{v}) = \sum_{j=1}^{J} F^j(\mathbf{y},\mathbf{v})$

Q: $C^0(\mathbf{y},\mathbf{r}) = \text{Sup}\left\{\text{Min}_{j=1,\ldots,J} C^j(\mathbf{y},\mathbf{r}^j) | \mathbf{r}^j \text{ positive}, \sum_{j=1}^{J} \mathbf{r}^j = \mathbf{r}\right\}$

S: $\mathbf{V}^0(\mathbf{y}) = \bigcup_{\substack{z_j \geq 0 \\ \sum_{j=1}^{J} z_j = 1}} \bigcap_{j=1}^{J} z_j \mathbf{V}^j(\mathbf{y})$

T: $\mathbf{R}^0(\mathbf{y}) = \sum_{j=1}^{J} \mathbf{R}^j(\mathbf{y})$

7. *General Concave Composition of Distance Functions*[e]

P: $F^0(\mathbf{y},\mathbf{v}) = F^*(\mathbf{y},F^1(\mathbf{y},\mathbf{v}),\ldots,F^J(\mathbf{y},\mathbf{v}))$

Q: $C^0(\mathbf{y},\mathbf{r}) = \text{Sup}\left\{C^*(\mathbf{y},C^1(\mathbf{y},\mathbf{r}^1),\ldots,C^J(\mathbf{y},\mathbf{r}^J)) | \mathbf{r}^j \text{ positive}, \sum_{j=1}^{J} \mathbf{r}^j = \mathbf{r}\right\}$

S: $\mathbf{V}^0(\mathbf{y}) = \text{Closure} \bigcup_{\mathbf{z} \in \mathbf{V}^*(\mathbf{y})} \bigcap_{j=1}^{J} z_j \mathbf{V}^j(\mathbf{y})$

T: $\mathbf{R}^0(\mathbf{y}) = \bigcup_{\mathbf{q} \in \mathbf{R}^*(\mathbf{y})} \sum_{j=1}^{J} q_j \mathbf{R}^j(\mathbf{y})$

8. *General Concave Composition of Cost Functions*[e]

P: $F^0(\mathbf{y},\mathbf{v}) = \text{Sup}\left\{F^*(\mathbf{y},F^1(\mathbf{y},\mathbf{v}^1),\ldots,F^J(\mathbf{y},\mathbf{v}^J)) | \mathbf{v}^j \text{ positive}, \sum_{j=1}^{J} \mathbf{v}^j = \mathbf{v}\right\}$

Q: $C^0(\mathbf{y},\mathbf{r}) = C^*(\mathbf{y},C^1(\mathbf{y},\mathbf{r}),\ldots,C^J(\mathbf{y},\mathbf{r}))$

S: $\mathbf{V}^0(\mathbf{y}) = \bigcup_{\mathbf{z} \in \mathbf{V}^*(\mathbf{y})} \sum_{j=1}^{J} z_j \mathbf{V}^j(\mathbf{y})$

T: $\mathbf{R}^0(\mathbf{y}) = \text{Closure} \bigcup_{\mathbf{q} \in \mathbf{R}^*(\mathbf{y})} \bigcap_{j=1}^{J} q_j \mathbf{R}^j(\mathbf{y})$

[c] By convention, for $z_j = 0$ we define $z_j \mathbf{R}^j(\mathbf{y}) = \mathbf{E}_+^N$, even if $\mathbf{R}^j(\mathbf{y})$ is empty.
[d] By convention, for $z_j = 0$ we define $z_j \mathbf{V}^j(\mathbf{y}) = \mathbf{E}_+^N$.
[e] Any of the functions F^* or F^j may, as a special case, be independent of $\mathbf{y}$.

and $\mathbf{V}^*(\mathbf{y})$, respectively. Let $\mathbf{R}^j(\mathbf{y})$ and $\mathbf{R}^*(\mathbf{y})$ be the factor price requirement sets for $\mathbf{V}^j(\mathbf{y})$ and $\mathbf{V}^*(\mathbf{y})$, respectively. Then, the composition rules in Table 2 hold, defining dual input-conventional input requirement sets $\mathbf{V}^0(\mathbf{y})$, factor price requirement sets $\mathbf{R}^0(\mathbf{y})$, distance functions $F^0(\mathbf{y},\mathbf{v})$, and cost functions $C^0(\mathbf{y},\mathbf{r})$.

Proof: *Rules* 1 *and* 2 – Given the positive diagonal matrix $\mathbf{A}(\mathbf{y})$, the set $\mathbf{V}^0(\mathbf{y}) = \{\mathbf{A}(\mathbf{y})\mathbf{v}|\mathbf{v} \in \mathbf{V}^1(\mathbf{y})\}$ is obviously input-conventional. From equation (8),

$$F^0(\mathbf{y},\mathbf{v}) = \text{Max}\left\{\lambda > 0 \,\middle|\, \frac{1}{\lambda}\mathbf{v} \in \mathbf{V}^0(\mathbf{y})\right\} = \text{Max}\left\{\lambda > 0 \,\middle|\, \frac{1}{\lambda}\mathbf{A}(\mathbf{y})^{-1}\mathbf{v} \in \mathbf{V}^1(\mathbf{y})\right\}$$
$$= F^1(\mathbf{y},\mathbf{A}(\mathbf{y})^{-1}\mathbf{v}).$$

From equation (1),

$$C^0(\mathbf{y},\mathbf{r}) = \text{Min}\{\mathbf{r}\cdot\mathbf{v}|\mathbf{v} \in \mathbf{V}^0(\mathbf{y})\} = \text{Min}\{\mathbf{r}\mathbf{A}(\mathbf{y})\mathbf{v}|\mathbf{v} \in \mathbf{V}^1(\mathbf{y})\} = C^1(\mathbf{y},\mathbf{r}\mathbf{A}(\mathbf{y})).$$

Finally,

$$\mathbf{R}^0(\mathbf{y}) = \{\mathbf{r}|C^0(\mathbf{y},\mathbf{r}) \geqq 1\} = \{\mathbf{r}|C^1(\mathbf{y},\mathbf{r}\mathbf{A}(\mathbf{y})) \geqq 1\}$$
$$= \{\mathbf{r}\mathbf{A}(\mathbf{y})^{-1}|C^1(\mathbf{y},\mathbf{r}) \geqq 1\} = \{\mathbf{r}\mathbf{A}(\mathbf{y})^{-1}|\mathbf{r} \in \mathbf{R}^1(\mathbf{y})\}.$$

Duality then implies that each of the composition rules P, Q, S, and T holds for Rule 2. Taking all the diagonal elements of $A(\mathbf{y})$ to be the scalar function $\alpha(\mathbf{y})$ implies Rule 1.

Rules 3 *and* 4 – Consider Rule 3, and suppose that S holds, defining $\mathbf{V}^0(\mathbf{y})$ as the convex hull of the union of the $\mathbf{V}^j(\mathbf{y})$. The minimum of a linear function on a convex hull of a closed set can always be attained at some point in the original set. Hence,

$$C^0(\mathbf{y},\mathbf{r}) = \text{Min}\{\mathbf{r}\cdot\mathbf{v}|\mathbf{v} \in \mathbf{V}^0(\mathbf{y})\} = \text{Min}\{\mathbf{r}\cdot\mathbf{v}|\mathbf{v} \in \mathbf{V}^j(\mathbf{y}), \text{ some } j\}$$
$$= \text{Min}_{j=1,\dots,J} C^j(\mathbf{y},\mathbf{r}).$$

Using duality, this establishes the equivalence of Q and S.

For a positive $\mathbf{v}$, one has $\mathbf{v}/F^0(\mathbf{y},\mathbf{v}) \in \mathbf{V}^0(\mathbf{y})$, implying the existence of scalars $z_j \geqq 0$, $\sum_{j=1}^J z_j = 1$ and points $\mathbf{v}^j/F^0(\mathbf{y},\mathbf{v}) \in \mathbf{V}^j(\mathbf{y})$ such that $\sum_{j=1}^J z_j\mathbf{v}^j = \mathbf{v}$. But this implies $F^j(\mathbf{y},\mathbf{v}^j) \geqq F^0(\mathbf{y},\mathbf{v})$, and hence, using the linear homogeneity of F^j in $\mathbf{v}$, $\sum_{j=1}^J F^j(\mathbf{y},z_j\mathbf{v}^j) \geqq F^0(\mathbf{y},\mathbf{v})$.

Alternately, consider the relation $F^0(\mathbf{y},\mathbf{v}) = \text{Max}\{\lambda|\mathbf{r}\cdot\mathbf{v} \geqq \lambda C^0(\mathbf{y},\mathbf{r})$ for all positive $\mathbf{r}\}$. Take any positive $\mathbf{w}^j$ with $\sum_{j=1}^J \mathbf{w}^j = \mathbf{v}$. By Lemma 5, $C^j(\mathbf{y},\mathbf{r})F^j(\mathbf{y},\mathbf{w}^j) \leqq \mathbf{r}\cdot\mathbf{w}^j$. For $\lambda = \sum_{j=1}^J F^j(\mathbf{y},\mathbf{w}^j)$, one has

$$\lambda C^0(\mathbf{y},\mathbf{r}) = \lambda\ \text{Min}_{j=1,\dots,J} C^j(\mathbf{y},\mathbf{r}) = \sum_{j=1}^{J} F^j(\mathbf{y},\mathbf{w}^j)\ \text{Min}_{i=1,\dots,J} C^i(\mathbf{y},\mathbf{r})$$
$$\leqq \sum_{j=1}^{J} F^j(\mathbf{y},\mathbf{w}^j) C^j(\mathbf{y},\mathbf{r}) \leqq \sum_{j=1}^{J} \mathbf{r}\cdot\mathbf{w}^j = \mathbf{r}\cdot\mathbf{v}.$$

Hence,

$$F^0(\mathbf{y},\mathbf{v}) \geqq \lambda = \sum_{j=1}^{J} F^j(\mathbf{y},\mathbf{w}^j),$$

for all $\mathbf{w}^j$ with $\sum_{j=1}^{J} \mathbf{w}^j = \mathbf{v}$. With the inequality in the preceding paragraph, this establishes P. Then P and S are equivalent by duality.

Given $F^0(\mathbf{y},\mathbf{v})$ from P, note that

$$\mathbf{R}^0(\mathbf{y}) = \{\mathbf{r} | \mathbf{r}\cdot\mathbf{v} \geqq F^0(\mathbf{y},\mathbf{v}) \text{ for all positive } \mathbf{v}\}$$
$$= \left\{\mathbf{r} | \mathbf{r}\cdot\mathbf{v} \geqq \sum_{j=1}^{J} F^j(\mathbf{y},\mathbf{v}^j) \text{ for all positive } \mathbf{v}^j, \sum_{j=1}^{J} \mathbf{v}^j = \mathbf{v}\right\}$$
$$= \left\{\mathbf{r} \;\middle|\; \sum_{j=1}^{J} \mathbf{r}\cdot\mathbf{v}^j \geqq \sum_{j=1}^{J} F^j(\mathbf{y},\mathbf{v}^j) \text{ for all positive } \mathbf{v}^j\right\}$$
$$= \{\mathbf{r} | \mathbf{r}\cdot\mathbf{v}^j \geqq F^j(\mathbf{y},\mathbf{v}^j), \text{ all } j \text{ and all positive } \mathbf{v}^j\}$$
$$= \bigcap_{j=1}^{J} \mathbf{R}^j(\mathbf{y}).$$

Hence, P and T are equivalent. This establishes Rule 3.

Rule 4 can be deduced from Rule 3 using the formal duality of C and F.

Rules 5 and 6 – Consider Rule 5: Given $\mathbf{V}^0(\mathbf{y}) = \sum_{j=1}^{J} \mathbf{V}^j(\mathbf{y})$, we see that $\mathbf{V}^0(\mathbf{y})$ is input-conventional, and that equation (1) implies $C^0(\mathbf{y},\mathbf{r}) = \sum_{j=1}^{J} C^j(\mathbf{y},\mathbf{r})$. Then Q and S are equivalent by duality. Next consider

$$F^0(\mathbf{y},\mathbf{v}) = \text{Max}\left\{\lambda \;\middle|\; \frac{1}{\lambda}\mathbf{v} \in \mathbf{V}^0(\mathbf{y})\right\} = \text{Max}\left\{\lambda \;\middle|\; \frac{1}{\lambda}\mathbf{v} \in \sum_{j=1}^{J} \mathbf{V}^j(\mathbf{y})\right\}$$
$$= \text{Max}\left\{\lambda \;\middle|\; \frac{1}{\lambda}\mathbf{v}^j \in \mathbf{V}^j(\mathbf{y}) \text{ for some } \mathbf{v}^j \text{ with } \sum_{j=1}^{J} \mathbf{v}^j = \mathbf{v}\right\}.$$

Given a small positive scalar α, there exist positive $\mathbf{v}^j$ with $\sum_{j=1}^{J} \mathbf{v}^j = \mathbf{v}$ such that $(\mathbf{v}^j/(F^0(\mathbf{y},\mathbf{v}) - \alpha)) \in \mathbf{V}^j(\mathbf{y})$, and hence $F^j(\mathbf{y},\mathbf{v}^j) \geqq F^0(\mathbf{y},\mathbf{v}) - \alpha$. Conversely, for any positive $\mathbf{w}^j$ with $\sum_{j=1}^{J} \mathbf{w}^j = \mathbf{v}$ and $\lambda = \text{Min}_{j=1,\dots,J} F^j(\mathbf{y},\mathbf{w}^j)$ one has $\mathbf{w}^j/\lambda \in \mathbf{V}^j(\mathbf{y})$, implying $F^0(\mathbf{y},\mathbf{v}) \geqq \lambda =$

$\text{Min}_{j=1,\dots,J}F^j(\mathbf{y},\mathbf{w}^j)$. With the previously established inequality, this implies $F^0(\mathbf{y},\mathbf{v}) = \text{Sup}\{\text{Min}_{j=1,\dots,J}F^j(\mathbf{y},\mathbf{v}^j)|\mathbf{v}^j \text{ positive}, \sum_{j=1}^J \mathbf{v}^j = \mathbf{v}\}$. By duality, P and S are then equivalent.

The factor price requirement set satisfies $\mathbf{R}^0(\mathbf{y}) = \text{Closure }\{\mathbf{r}|\mathbf{r} \text{ positive}, \sum_{j=1}^J C^j(\mathbf{y},\mathbf{r}) \geqq 1\}$. Then positive $\mathbf{r}$ is contained in $\mathbf{R}^0(\mathbf{y})$ if and only if there exist non-negative scalars z_j such that $\sum_{j=1}^J z_j = 1$ and $C^j(\mathbf{y},\mathbf{r}) \geqq z_j$, or $\mathbf{r} \in \bigcap_{j=1}^J(z_j\mathbf{R}^j(\mathbf{y}))$. Hence $\mathbf{R}^0(\mathbf{y}) = \text{Closure } \mathbf{A}$, where

$$\mathbf{A} = \bigcup_{\substack{z_j \geqq 0 \\ \sum_{j=1}^J z_j = 1}} \bigcap_{j=1}^{J} (z_j\mathbf{R}^j(\mathbf{y})).$$

Now suppose $\mathbf{r}^i \in \mathbf{A}$ and $\mathbf{r}^i \to \mathbf{r}^0$. Then there exist $z_{ji} \geqq 0$ such that $\sum_{j=1}^J z_{ji} = 1$ and $\mathbf{r}^i \in z_{ji}\mathbf{R}^j(\mathbf{y})$ for each j. Choose a subsequence of $(z_{1i},\dots,z_{Ji})$ converging to $(z_{10},\dots,z_{J0})$. Retain the index notation i for the subsequence. If $z_{j0} > 0$, then $\mathbf{r}^i/z_{ji} \to \mathbf{r}^0/z_{j0} \in \mathbf{R}^j(\mathbf{y})$, since $\mathbf{R}^j(\mathbf{y})$ is closed. If $z_{j0} = 0$, then $\mathbf{r}^0 \in z_{j0}\mathbf{R}^j(\mathbf{y}) = \mathbf{R}_+^N$. Hence, $\mathbf{r}^0 \in \bigcap_{j=1}^J (z_{j0}\mathbf{R}^j(\mathbf{y})) \subseteq \mathbf{A}$. Therefore, $\mathbf{A}$ is a closed set, and $\mathbf{R}^0(\mathbf{y}) = \mathbf{A}$. Duality then implies the equivalence of Q and T. This establishes Rule 5.

Rule 6 follows from Rule 5 by the formal duality of the distance and cost functions and of the input and factor price requirement sets.

Rules 7 *and* 8 – Consider Rule 7, and $F^0(\mathbf{y},\mathbf{v}) = F^*(\mathbf{y},\mathbf{v}),\dots, F^J(\mathbf{y},\mathbf{v}))$. Since $F^*(\mathbf{y},\mathbf{z})$ is non-decreasing, linear homogeneous, and concave in positive $\mathbf{z}$, and the $F^j(\mathbf{y},\mathbf{v})$ have the same properties in positive $\mathbf{v}$, it is immediate that $F^0(\mathbf{y},\mathbf{v})$ is non-decreasing and linear homogeneous in positive $\mathbf{v}$. Consider positive $\mathbf{v},\mathbf{v}'$ and $0 < \theta < 1$. Then $F^j(\mathbf{y},\theta\mathbf{v} + (1-\theta)\mathbf{v}') \geqq \theta F^j(\mathbf{y},\mathbf{v}) + (1-\theta)F^j(\mathbf{y},\mathbf{v}')$, implying

$$\begin{aligned} &F^0(\mathbf{y},\theta\mathbf{v} + (1-\theta)\mathbf{v}') \\ &= F^*(\mathbf{y},F^1(\mathbf{y},\theta\mathbf{v} + (1-\theta)\mathbf{v}'),\dots,F^J(\mathbf{y},\theta\mathbf{v} + (1-\theta)\mathbf{v}')) \\ &\geqq F^*(\mathbf{y},\theta F^1(\mathbf{y},\mathbf{v}) + (1-\theta)F^1(\mathbf{y},\mathbf{v}'),\dots,\theta F^J(\mathbf{y},\mathbf{v}) + (1-\theta)F^J(\mathbf{y},\mathbf{v}')) \\ &\geqq \theta F^*(\mathbf{y},F^1(\mathbf{y},\mathbf{v}),\dots,F^J(\mathbf{y},\mathbf{v})) + (1-\theta)F^*(\mathbf{y},F^1(\mathbf{y},\mathbf{v}'),\dots,F^J(\mathbf{y},\mathbf{v}')), \end{aligned}$$

with the second inequality following from the concavity property of F^*. The value $F^0(\mathbf{y},\mathbf{v}) = +\infty$ can occur only if $F^*(\mathbf{y},\mathbf{z}) = +\infty$ for some positive finite $\mathbf{z}$, or $F^j(\mathbf{y},\mathbf{v}) = +\infty$ for some j. Since F^j and F^* are input-conventional, either case implies $\mathbf{y} = \mathbf{0}$. This establishes that $F^0(\mathbf{y},\mathbf{v})$ is input-conventional.

From equation (9),

$$\begin{aligned}\mathbf{V}^0(\mathbf{y}) &= \text{Closure } \{\mathbf{v} \gg \mathbf{0} | F^*(\mathbf{y}, F^1(\mathbf{y},\mathbf{v}),\ldots,F^J(\mathbf{y},\mathbf{v})) \geqq 1\} \\ &= \text{Closure } \{\mathbf{v} \gg \mathbf{0} | \text{ there exists } \mathbf{z} \gg \mathbf{0} \text{ such that } F^j(\mathbf{y},\mathbf{v}) \\ &\quad \geqq z_j \text{ and } F^*(\mathbf{y},\mathbf{z}) \geqq 1\} \\ &= \text{Closure } \bigcup_{\substack{\mathbf{z}\in \mathbf{V}^*(\mathbf{y}) \\ \mathbf{z}\gg\mathbf{0}}} \bigcap_{j=1}^{J} \{\mathbf{v} \gg \mathbf{0} | F^j(\mathbf{y},\mathbf{v}) \geqq z_j\}.\end{aligned}$$

Define

$$\tilde{\mathbf{V}}^0(\mathbf{y}) = \bigcup_{\mathbf{z}\in \mathbf{V}^*(\mathbf{y})} \bigcap_{j=1}^{J} (z_j \mathbf{V}^j(\mathbf{y})).$$

Clearly $\mathbf{V}^0(\mathbf{y}) \subseteq$ Closure $\tilde{\mathbf{V}}^0(\mathbf{y})$. If $\mathbf{v} \in \tilde{\mathbf{V}}^0(\mathbf{y})$, then $\mathbf{v} \in z_j \mathbf{V}^j(\mathbf{y})$ for some $\mathbf{z} \in \mathbf{V}^*(\mathbf{y})$. For a small positive scalar α, the vector $\mathbf{v} + \alpha \mathbf{e}_N$, where $\mathbf{e}_N$ is an N-vector of ones, is in the interior of $z_j \mathbf{V}^j(\mathbf{y})$. Hence, there exists a small positive scalar β such that $F^j(\mathbf{y},\mathbf{v} + \alpha \mathbf{e}_N) \geqq z_j + \beta$. Since $\mathbf{z} + \beta \mathbf{e}_J \in \mathbf{V}^*(\mathbf{y})$, this implies $\mathbf{v} + \alpha \mathbf{e}_N \in \mathbf{V}^0(\mathbf{y})$. Hence, closure $\tilde{\mathbf{V}}^0(\mathbf{y}) \subseteq \mathbf{V}^0(\mathbf{y})$. This establishes $\mathbf{V}^0(\mathbf{y}) =$ Closure $\tilde{\mathbf{V}}^0(\mathbf{y})$. Duality implies the equivalence of P and S.

Next consider the cost function defined by

$$\begin{aligned}C^0(\mathbf{y},\mathbf{r}) &= \text{Min}\{\mathbf{r}\cdot\mathbf{v} | \mathbf{v} \in \mathbf{V}^0(\mathbf{y})\} = \inf\{\mathbf{r}\cdot\mathbf{v} | \mathbf{v} \in \tilde{\mathbf{V}}^0(\mathbf{y})\} \\ &= \inf_{\mathbf{z}\in \mathbf{V}^*(\mathbf{y})} \inf\{\mathbf{r}\cdot\mathbf{v} | \mathbf{v} \in z_j \mathbf{V}^j(\mathbf{y}) \text{ for all } j\}.\end{aligned}$$

For fixed $\mathbf{z} \in \mathbf{V}^*(\mathbf{y})$,

$$\inf\left\{\mathbf{r}\cdot\mathbf{v} | \mathbf{v} \in \bigcap_{j=1}^{J} (z_j \mathbf{V}^j(\mathbf{y}))\right\} = \sup\left\{\sum_{j=1}^{J} z_j C^j(\mathbf{y},\mathbf{r}^j) | \mathbf{r}^j \geqq \mathbf{0}, \sum_{j=1}^{J} \mathbf{r}^j = \mathbf{r}\right\},$$

by Rules 1 and 4. The function $f(\mathbf{z},(\mathbf{r}^j);\mathbf{y}) = \sum_{j=1}^{J} z_j C^j(\mathbf{y},\mathbf{r}^j)$, defined for $\mathbf{z} \in \mathbf{V}^*(\mathbf{y})$ and $(\mathbf{r}^j)$ in the set $\mathbf{A} = \{(\mathbf{r}^j) | \mathbf{r}^j \geqq \mathbf{0}, \sum_{j=1}^{J} \mathbf{r}^j = \mathbf{r}\}$, is continuous on $\mathbf{V}^*(\mathbf{y}) \times \mathbf{A}$, concave in $(\mathbf{r}^j)$ for each $\mathbf{z}$, and linear (and thus convex) in $\mathbf{z}$ for each $(\mathbf{r}^j)$. Since $\mathbf{A}$ is bounded, the general minimax theorem [Rockafellar (1970, Corollary 37.3.1)] implies

$$\inf_{\mathbf{z}\in \mathbf{V}^*(\mathbf{y})} \sup_{(\mathbf{r}^j)\in \mathbf{A}} f(\mathbf{z},(\mathbf{r}^j);\mathbf{y}) = \sup_{(\mathbf{r}^j)\in \mathbf{A}} \inf_{\mathbf{z}\in \mathbf{V}^*(\mathbf{y})} f(\mathbf{z},(\mathbf{r}^j);\mathbf{y}).$$

But

$$\inf_{\mathbf{z}\in \mathbf{V}^*(\mathbf{y})} f(\mathbf{z},(\mathbf{r}^j);\mathbf{y}) = \inf_{\mathbf{z}\in \mathbf{V}^*(\mathbf{y})} \sum_{j=1}^{J} z_j C^j(\mathbf{y},\mathbf{r}^j) = C^*(\mathbf{y}, C^1(\mathbf{y},\mathbf{r}^1),\ldots,C^J(\mathbf{y},\mathbf{r}^J)),$$

by the definition of C^*, implying

$$C^0(\mathbf{y},\mathbf{r}) = \inf_{\mathbf{z}\in\mathbf{V}^*(\mathbf{y})} \sup_{(\mathbf{r}^j)\in\mathbf{A}} f(\mathbf{z},(\mathbf{r}^j);\mathbf{y}) = \sup_{(\mathbf{r}^j)\in\mathbf{A}} C^*(\mathbf{y},C^1(\mathbf{y},\mathbf{r}^1),\ldots,C^J(\mathbf{y},\mathbf{r}^J)).$$

Using duality, this establishes the equivalence of Q and S.

The factor price requirement set satisfies

$$\begin{aligned}\mathbf{R}^0(\mathbf{y}) &= \{\mathbf{r}|C^0(\mathbf{y},\mathbf{r}) \geqq 1\} = \left\{\sum_{j=1}^{J} \mathbf{r}^j | C^*(\mathbf{y},C^1(\mathbf{y},\mathbf{r}^1),\ldots,C^J(\mathbf{y},\mathbf{r}^J)) \geqq 1\right\} \\ &= \left\{\sum_{j=1}^{J}\mathbf{r}^j | C^*(\mathbf{y},\mathbf{q}) \geqq 1 \text{ and } C^j(\mathbf{y},\mathbf{r}^j) \geqq q_j; \text{ for some } q_j \geqq 0\right\} \\ &= \bigcup_{\mathbf{q}\in\mathbf{R}^*(\mathbf{y})} \sum_{j=1}^{J} \left\{\mathbf{r}^j | C^j(\mathbf{y},\mathbf{r}^j) \geqq q_j\right\} = \bigcup_{\mathbf{q}\in\mathbf{R}^*(\mathbf{y})} \sum_{j=1}^{J} q_j \mathbf{R}^j(\mathbf{y}).\end{aligned}$$

With duality, this establishes the equivalence of Q and T. Hence, Rule 7 is established.

The formal duality of the distance and cost functions yields Rule 8 from Rule 7. Q.E.D.

A variety of implications for technological structure can be drawn from these composition rules. First, using Rule 7 and the Cobb–Douglas distance and cost functions given in equations (17) and (18), we obtain a *Cobb–Douglas composition of distance functions*: For $\alpha_1,\ldots,\alpha_J > 0$, $\sum_{j=1}^{J} \alpha_j = 1$,

$$\text{P:} \quad F^0(\mathbf{y},\mathbf{v}) = F^1(\mathbf{y},\mathbf{v})^{\alpha_1}\cdot\cdots\cdot F^J(\mathbf{y},\mathbf{v})^{\alpha_J},$$

$$\text{Q:} \quad C^0(\mathbf{y},\mathbf{r}) = \alpha_1^{-\alpha_1}\cdot\cdots\cdot\alpha_J^{-\alpha_J} \cdot\sup\left\{C^1(\mathbf{y},\mathbf{r}^1)^{\alpha_1}\cdot\cdots\cdot C^J(\mathbf{y},\mathbf{r}^J)^{\alpha_J} | \mathbf{r}^j \text{ positive}, \sum_{j=1}^{J} \mathbf{r}^j = \mathbf{r}\right\},$$

$$\text{S:} \quad \mathbf{V}^0(\mathbf{y}) = \bigcup_{\sum_{j=1}^{J} z_j \geqq 0} \bigcap_{j=1}^{J} e^{z_j/\alpha_j}\mathbf{V}_j(\mathbf{y}),$$

$$\text{T:} \quad \mathbf{R}^0(\mathbf{y}) = \bigcup_{\sum_{j=1}^{J} z_j \geqq 0} \sum_{j=1}^{J} \alpha_j e^{z_j/\alpha_j}\mathbf{R}^j(\mathbf{y}).$$

Formal duality gives *Cobb–Douglas composition of cost functions*: For

$\alpha_1,\ldots,\alpha_J > 0$, $\sum_{j=1}^{J} \alpha_j = 1$,

$$\text{P:}\quad F^0(\mathbf{y},\mathbf{v}) = \alpha_1^{-\alpha_1}\cdot\cdots\cdot\alpha_J^{-\alpha_J}\sup\left\{F^1(\mathbf{y},\mathbf{v}^1)^{\alpha_1}\cdot\cdots\cdot F^J(\mathbf{y},\mathbf{v}^J)^{\alpha_J} \cdot \middle| \mathbf{v}^j \text{ positive}, \sum_{j=1}^{J}\mathbf{v}^j = \mathbf{v}\right\},$$

$$\text{Q:}\quad C^0(\mathbf{y},\mathbf{r}) = C^1(\mathbf{y},\mathbf{r})^{\alpha_1}.\ldots.C^J(\mathbf{y},\mathbf{r})^{\alpha_J},$$

$$\text{S:}\quad \mathbf{V}^0(\mathbf{y}) = \bigcup_{\sum_{j=1}^{J} z_j \geqq 0}\ \bigcap_{j=1}^{J} \alpha_j e^{z_j/\alpha_j}\mathbf{V}^j(\mathbf{y}),$$

$$\text{T:}\quad \mathbf{R}^0(\mathbf{y}) = \bigcup_{\sum_{j=1}^{J} z_j \geqq 0}\ \sum_{j=1}^{J} e^{z_j/\alpha_j}\mathbf{R}^j(\mathbf{y}).$$

Using Rule 7 and the C.E.S. distance and cost functions given in equations (19) and (20), we obtain a *C.E.S. composition of distance functions*:

$$\text{P:}\quad F^0(\mathbf{y},\mathbf{v}) = \left(\sum_{j=1}^{J}(F^j(\mathbf{y},\mathbf{v})/D_j(\mathbf{y}))^{1-1/\sigma}\right)^{1/(1-1/\sigma)},$$

$$\text{Q:}\quad C^0(\mathbf{y},\mathbf{r}) = \sup\left\{\left(\sum_{j=1}^{J}(C^j(\mathbf{y},\mathbf{r}^j)D_j(\mathbf{y}))^{1-\sigma}\right)^{1/(1-\sigma)} \middle| \mathbf{r}^j \geqq 0, \sum_{j=1}^{J}\mathbf{r}^j = \mathbf{r}\right\},$$

$$\text{S:}\quad \mathbf{V}^0(\mathbf{y}) = \bigcup_{\substack{z_j \geqq 0 \\ \sum_{j=1}^{J} z_j = 1}}\ \bigcap_{j=1}^{J}(z_j^{\sigma/(\sigma-1)}D_j(\mathbf{y})\mathbf{V}^j(\mathbf{y})),$$

$$\text{T:}\quad \mathbf{R}^0(\mathbf{y}) = \bigcup_{\substack{q_j \geqq 0 \\ \sum_{j=1}^{J} q_j = 1}}\ \sum_{j=1}^{J}(\mathbf{R}^j(\mathbf{y})q_j^{1/(1-\sigma)}/D_j(\mathbf{y})).$$

Again, application of formal duality gives a *C.E.S. composition of cost functions*:

$$\text{P:}\quad F^0(\mathbf{y},\mathbf{v}) = \sup\left\{\left(\sum_{j=1}^{J}(F^j(\mathbf{y},\mathbf{v}^j)/D_j(\mathbf{y}))^{1-1/\sigma}\right)^{1/(1-1/\sigma)} \middle| \mathbf{v}^j \geqq 0, \sum_{j=1}^{J}\mathbf{v}^j = \mathbf{v}\right\},$$

$$\text{Q:}\quad C^0(\mathbf{y},\mathbf{r}) = \left(\sum_{j=1}^{J}(C^j(\mathbf{y},\mathbf{r})D_j(\mathbf{y}))^{1-\sigma}\right)^{1/(1-\sigma)},$$

$$\text{S:}\quad \mathbf{V}^0(\mathbf{y}) = \bigcup_{\substack{z_j \geqq 0 \\ \sum_{j=1}^{J} z_j = 1}}\ \sum_{j=1}^{J}(\mathbf{V}^j(\mathbf{y})z_j^{1/(1-1/\sigma)}D_j(\mathbf{y})),$$

$$\text{T:}\quad \mathbf{R}^0(\mathbf{y}) = \bigcup_{\substack{q_j \geqq 0 \\ \sum_{j=1}^{J} q_j = 1}}\ \bigcap_{j=1}^{J}(q_j^{1/(1-\sigma)}\mathbf{R}^j(\mathbf{y})/D_j(\mathbf{y})).$$

A production possibility set is said to be input-homothetic[9] if there exists a positive function $\alpha(\lambda,\mathbf{y})$ of $\lambda \in \mathbf{E}_+$ and $\mathbf{y} \in \mathbf{E}_+^M$, increasing in λ, with $\alpha(0,\mathbf{y}) = 0$, such that for $\mathbf{y} \neq \mathbf{0}$, $\mathbf{V}(\mathbf{y}) = \alpha(|\mathbf{y}|, \mathbf{y}/|\mathbf{y}|)\ \mathbf{V}(\mathbf{y}/|\mathbf{y}|)$, where $|\mathbf{y}|$ is the norm of $\mathbf{y}$ and we assume $\mathbf{y}/|\mathbf{y}| \in \mathbf{Y}^*$. In the case of a single output, this reduces to the textbook definition $\mathbf{V}(\mathbf{y}) = \alpha(\mathbf{y},1)\mathbf{V}(1)$ of homotheticity. More generally, it satisfies the textbook definition for any fixed output proportions, and allows the shape of the scaling of inputs versus output to vary with the output proportions. A property of an input-homothetic technology is that for fixed output proportions, the cost minimizing input mix is determined solely by input prices, independent of the scale of output. Rule 1 in Lemma 9 yields the following conclusion, where $|\mathbf{y}|$ is the norm of $\mathbf{y}$.

For an input-conventional production possibility set, the following conditions are equivalent:

(a) *The production possibility set is input-homothetic.*

(b) *The distance function has the form*

$$F(\mathbf{y},\mathbf{v}) = F(\mathbf{y}/|\mathbf{y}|,\mathbf{v})/\alpha(|\mathbf{y}|,\mathbf{y}/|\mathbf{y}|) \quad \text{for} \quad \mathbf{y} \neq \mathbf{0}. \tag{33}$$

(c) *The cost function has the form*

$$C(\mathbf{y},\mathbf{r}) = \alpha(|\mathbf{y}|,\mathbf{y}/|\mathbf{y}|)C(\mathbf{y}/|\mathbf{y}|,\mathbf{r}) \quad \text{for} \quad \mathbf{y} \neq \mathbf{0}. \tag{34}$$

A technology is *input-output separable* if it can be defined by a condition of the form $\beta(\mathbf{v})\gamma(\mathbf{y}) \geqq 1$. The distance function for this technology satisfies $\beta(\mathbf{v}/F(\mathbf{y},\mathbf{v}))\gamma(\mathbf{y}) = 1$, and hence can be written in the form $F(\mathbf{y},\mathbf{v}) = f(\gamma(\mathbf{y}),\mathbf{v})$, with f linear homogeneous in $\mathbf{v}$. Then, the cost function can be written $C(\gamma(\mathbf{y}),\mathbf{r})$, and the input requirement set $\mathbf{V}(\gamma(\mathbf{y}))$, with $\gamma(\mathbf{y})$ interpretable as the level of a single intermediate output. From the preceding result, a technology is both input-homothetic and input–output-separable if and only if the distance and cost functions can be written in the separable forms $F(\mathbf{y},\mathbf{v}) = F^1(\mathbf{v})F^2(\mathbf{y})$ and $C(\mathbf{y},\mathbf{r}) = C^1(\mathbf{r})C^2(\mathbf{y})$. Note also that these forms are related directly by composition Rule 1 (with F^1 and C^1 independent of $\mathbf{y}$).

Composition Rule 2 can be used to deduce the implications of factor augmenting technical change, or output change, on the distance and cost functions. Composition Rules 3–6 allow the geometric or algebraic construction of cost functions and factor price requirement sets. For example, in (a) of Figure 14, suppose given a Cobb–Douglas input

[9]This definition is due to G. Hanoch.

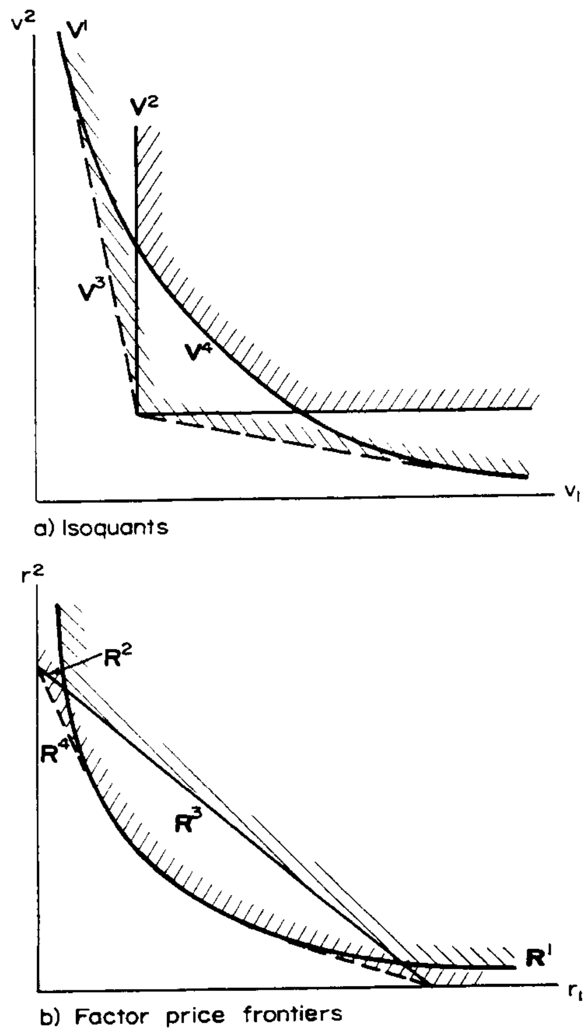


FIGURE 14

requirement set $\mathbf{V}^1$ and a Leontief input requirement set $\mathbf{V}^2$. The duals of these sets are the Cobb–Douglas factor price requirement set $\mathbf{R}^1$ and the linear factor price requirement set $\mathbf{R}^2$, respectively, illustrated in (b). The (convex hull of the) union of $\mathbf{V}^1$ and $\mathbf{V}^2$ is the set $\mathbf{V}^3$ in (a) of Figure 14. By Rule 3, the dual of $\mathbf{V}^3$ is the intersection $\mathbf{R}^3$ of $\mathbf{R}^1$ and $\mathbf{R}^2$. The intersection $\mathbf{V}^4$ of $\mathbf{V}^1$ and $\mathbf{V}^2$ has by Rule 4 the dual $\mathbf{R}^4$, given by the convex hull of the union of $\mathbf{R}^1$ and $\mathbf{R}^2$.

Composition Rules 7 and 8 yield a general result on separable distance and cost functions. Suppose the input vector $\mathbf{v}$ can be partitioned into sub-vectors, $\mathbf{v} = (\mathbf{v}_{(1)},\ldots,\mathbf{v}_{(J)})$, with a commensurate partition $\mathbf{r} = (\mathbf{r}_{(1)},\ldots,\mathbf{r}_{(J)})$ of the input price vector. Suppose a distance function F^j depends only on the sub-vector of inputs $\mathbf{v}_{(j)}$. Then, we can with a slight change of notation write the distance function $F^j(\mathbf{y},\mathbf{v}_{(j)})$. The dual cost function C^j then depends only on the sub-vector of prices $\mathbf{r}_{(j)}$, and can be written $C^j(\mathbf{y},\mathbf{r}_{(j)})$.

Lemma 10. Let $F^j(\mathbf{y},\mathbf{v}_{(j)})$ and $F^*(\mathbf{y},(z_1,\ldots,z_J))$ be input-conventional distance functions, and $C^j(\mathbf{y},\mathbf{r}_{(j)})$ and $C^*(\mathbf{y},(q_1,\ldots,q_J))$ their respective cost functions. Then $F^0(\mathbf{y},\mathbf{v}) = F^*(\mathbf{y},F^1(\mathbf{y},\mathbf{v}_{(1)}),\ldots,F^J(\mathbf{y},\mathbf{v}_{(J)}))$ is an input-conventional distance function with the dual cost function $C^0(\mathbf{y},\mathbf{r}) = C^*(\mathbf{y},C^1(\mathbf{y},\mathbf{r}_{(1)}),\ldots,C^J(\mathbf{y},\mathbf{r}_{(J)}))$; i.e., the distance function is separable if and only if the cost function is separable.

Proof: The general concave composition Rule 7 implies that $F^0(\mathbf{y},\mathbf{v})$ is input-conventional, and that its cost function is

$$C^0(\mathbf{y},\mathbf{r}) = \sup\left\{C^*(\mathbf{y},C^1(\mathbf{y},\mathbf{r}^1_{(1)}),\ldots,C^J(\mathbf{y},\mathbf{r}^J_{(J)}))\middle|\sum_{j=1}^{J}\mathbf{r}^j = \mathbf{r}\right\},$$

where $\mathbf{r}^j = (\mathbf{r}^j_{(1)},\ldots,\mathbf{r}^j_{(j)},\ldots,\mathbf{r}^j_{(J)})$. Since the cost functions are non-decreasing in prices, the supremum is achieved by $\mathbf{r}^j = (\mathbf{0},\ldots,\mathbf{0},\mathbf{r}_{(j)},\mathbf{0},\ldots,\mathbf{0})$ for $j = 1,\ldots,J$, or $C^0(\mathbf{y},\mathbf{r}) = C^*(\mathbf{y},C^1(\mathbf{y},\mathbf{r}_{(1)}),\ldots,C^J(\mathbf{y},\mathbf{r}_{(J)}))$. Q.E.D.

In the case of a single output and input-homotheticity, the separability property of the distance function implies a corresponding separability of the production function. However, in the absence of input-homotheticity, there is no simple relation between separability properties of the production and distance functions.

Composition Rules 7 and 8 can provide rules for computing distance or cost functions in cases of incomplete separability. For example, if the distance function is separable except for one input common to each F^j, then the cost function dual to the composite distance function will be given by a supremum involving the price of the one common input.

PART II. RESTRICTED PROFIT FUNCTIONS

13. The General Representation of Production Possibilities

In the previous sections of this chapter, the implications of input cost minimization with fixed outputs have been explored. More generally, optimization by the firm over any set of variable inputs and outputs can be analyzed. This approach leads to a general concept, the restricted profit function, which in special cases reduces to the cost function, a maximum revenue function, or an unrestricted maximum profit function.

Consider an environment for a firm in which N commodities, indexed $n = 1,2,...,N$, can be traded in competitive markets at a price vector $\mathbf{p} = (p_1,p_2,...,p_N)$. The firm treats these commodities as possible variable inputs or outputs. A *production plan* for the firm is an N-tuple of real numbers $\mathbf{x} = (x_1,x_2,...,x_N)$, with x_n interpreted as the quantity of net output (or, for compactness, *netput*) of commodity n, negative if the commodity is an input and positive if it is an output. The profit associated with a production plan $\mathbf{x}$ is given by $\pi = \mathbf{p}\cdot\mathbf{x}$, the inner product of $\mathbf{p}$ and $\mathbf{x}$.[10]

The technological limits on the actions of the firm can be described by a set $\mathbf{T}$ of possible production plans. Generally, the firm's possibilities are influenced by prior contracts to hire inputs or deliver outputs, and by the physical and economic environment. It is convenient to suppose that these effects can be summarized in an M-dimensional real vector $\mathbf{z} = (z_1,z_2,...,z_M)$ which can vary within some allowable set $\mathbf{Z}$. The production possibility set $\mathbf{T} = \mathbf{T}(\mathbf{z})$ then depends on the value of the vector $\mathbf{z}$.

Several examples will illustrate the generality of this formulation. If $\mathbf{z}$ is an output bundle, and all the commodities in the netput bundle are inputs, then $\mathbf{T}(\mathbf{z})$ is an input requirement set (with a negative sign) and $\pi = \mathbf{p}\cdot\mathbf{x}$ is the negative of cost. Under the appropriate interpretation of $\mathbf{T}(\mathbf{z})$, the problems of ex ante or ex post, or long or short run, cost minimization can be treated in this model.

If a firm is maximizing profit with a fixed input, then this input can be included in the parameter vector $\mathbf{z}$, and maximization can be carried out in terms of the variable commodities, yielding a maximum variable profit, net of the cost of the fixed input. Alternately, the fixed input can be included in the netput vector, with the production possibility set specifying its level. Maximization in this case yields a maximum total profit.

[10]The commodity price vector $\mathbf{p}$ is defined so that the prices of most commodities are non-negative. Then, output of a positively priced commodity contributes to revenue, and input of such a commodity contributes to cost. However, we do not rule out the possibility of negatively priced commodities. While this generalization is largely definitional, it proves useful in dealing with commodities for which there is no free disposal and for which net supply in an economy at zero price may be positive or negative. (Sawdust is an example.)

If the parameter vector $\mathbf{z}$ contains all inputs to the firm and all commodities in the netput bundle $\mathbf{x}$ are outputs, maximization leads to a maximum revenue for fixed inputs. If all inputs and outputs of the firm are in the netput bundle, then maximization leads to maximum unrestricted profits. Components of the vector $\mathbf{z}$ may be environmental or behavioral parameters other than commodity levels. The state of technical progress or the degree of learning by the firm may be included in $\mathbf{z}$. If the possibilities of the firm are influenced by ex ante decisions, then design parameters or anticipated prices and quantities can be included in $\mathbf{z}$. The parameter vector may include, in the case that the firm is subject to externalities, the production plans of other firms. Finally, $\mathbf{z}$ may include parameters introduced by the economic analyst to characterize the technology.

Three basic axioms on production possibilities, which involve little loss of economic generality, will be imposed in further analysis.

> *Axiom* 1. The set $\mathbf{Z}$ of possible production parameters is a non-empty subset of an M-dimensional Euclidean space $\mathbf{E}^M$. For each $\mathbf{z} \in \mathbf{Z}$, the set $\mathbf{T} = \mathbf{T}(\mathbf{z})$ of possible production plans is a closed non-empty subset of an N-dimensional Euclidean space $\mathbf{E}^N$.

The next axiom requires several definitions. The *normal cone* (barrier cone) of $\mathbf{T}(\mathbf{z})$, denoted by $\mathbf{P}(\mathbf{z})$, is the set of all price vectors $\mathbf{p} \in \mathbf{E}^N$ such that $\mathbf{p}\cdot\mathbf{x}$ is bounded above for $\mathbf{x} \in \mathbf{T}(\mathbf{z})$. Clearly, the normal cone will be the largest set of prices on which we can hope to define a maximum profit function. We denote the interior of the normal cone by $\mathbf{P}^0(\mathbf{z})$, and its closure by $\bar{\mathbf{P}}(\mathbf{z})$. The set $\mathbf{T}(\mathbf{z})$ is said to be *semi-bounded* if $\mathbf{P}^0(\mathbf{z})$ is non-empty. An example will illustrate the restriction placed on the structure of $\mathbf{T}(\mathbf{z})$ by a condition that it be semi-bounded: If $\mathbf{T}(\mathbf{z})$ contains both $\theta\mathbf{x}$ and $-\theta\mathbf{x}$ for some $\mathbf{x}$ and all large positive scalars θ, then the requirement that $\mathbf{p}\cdot(\theta\mathbf{x})$ and $\mathbf{p}\cdot(-\theta\mathbf{x})$ be bounded above for $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ implies that $\mathbf{p}$ must satisfy $\mathbf{p}\cdot\mathbf{x} = 0$. But this implies that $\mathbf{P}(\mathbf{z})$ is contained in a hyperplane, so that $\mathbf{P}^0(\mathbf{z})$ is empty and $\mathbf{T}(\mathbf{z})$ fails to be semi-bounded. Thus, the condition that $\mathbf{T}(\mathbf{z})$ be semi-bounded requires that at sufficiently large scale levels, production plans be irreversible in the sense that starting from a possible production plan $\mathbf{x}$, it is not feasible to reverse the role of inputs and outputs and produce the plan $-\mathbf{x}$. Most technologies can be expected to satisfy irreversibility, and hence be semi-bounded. This will be the case if labor cannot be produced, and all non-zero production plans require some labor input. Alternately, this will be the case if non-zero outputs in a production plan always require chronologically prior inputs.

An alternative definition of the semi-boundedness property of a set $\mathbf{T}(\mathbf{z})$, a condition that the asymptotic cone (recession cone) of $\mathbf{T}(\mathbf{z})$ be pointed,[11] is discussed in Appendix A.3. Result 11.3 in this appendix shows this definition to be equivalent to the requirement that $\mathbf{P}^0(\mathbf{z})$ be non-empty.

> *Axiom* 2. For each $\mathbf{z} \in \mathbf{Z}$, the production possibility set $\mathbf{T}(\mathbf{z})$ is semi-bounded.

In investigating the effects on profit levels of shifts in the parameter vector $\mathbf{z}$, it is useful to require that $\mathbf{T}(\mathbf{z})$ vary "regularly" with $\mathbf{z}$. We define $\mathbf{T}(\mathbf{z})$ to be *strongly continuous* on $\mathbf{Z}$ if for each $\mathbf{z}^0 \in \mathbf{Z}$ and sequence $\mathbf{z}^k \in \mathbf{Z}$ converging to $\mathbf{z}^0$, the following three conditions hold:

(i) If a sequence $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$ converges to $\mathbf{x}^0$, then $\mathbf{x}^0 \in \mathbf{T}(\mathbf{z}^0)$.
(ii) If $\mathbf{x}^0 \in \mathbf{T}(\mathbf{z}^0)$, then there exists a sequence $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$ which converges to $\mathbf{x}^0$.
(iii) If a sequence $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$ and a sequence of positive scalars θ_k have θ_k converging to zero and $\theta_k \mathbf{x}^k$ converging to $\mathbf{x}^0$, then there exists a sequence $\hat{\mathbf{x}}^k \in \mathbf{T}(\mathbf{z}^0)$ and a sequence of positive scalars $\hat{\theta}_k$ with $\hat{\theta}_k$ converging to zero and $\hat{\theta}_k \hat{\mathbf{x}}^k$ converging to $\mathbf{x}^0$.

In mathematical terminology, $\mathbf{T}(\mathbf{z})$ is a *correspondence*, conditions (i) and (ii) define *upper* and *lower hemicontinuous* correspondences, respectively, and (iii) states that the asymptotic cone of $\mathbf{T}(\mathbf{z})$ is an upper hemicontinuous correspondence.

Condition (i) requires that $\mathbf{T}(\mathbf{z})$ not "shrink" discontinuously as $\mathbf{z}$ varies, while (ii) requires that it not "expand" discontinuously. Condition (iii) requires that the set of directions in which $\mathbf{T}(\mathbf{z})$ is unbounded not "shrink" discontinuously as $\mathbf{z}$ varies. When $\mathbf{T}(\mathbf{z})$ satisfies (i) and (ii) alone, it is termed *continuous*. We shall show later that when the production possibility set is convex, the upper and lower hemicontinuity conditions (i) and (ii) imply condition (iii). Hence, a continuous convex production possibility correspondence is strongly continuous.

Define the set $\mathbf{Y} = \{(\mathbf{z},\mathbf{x}) \in \mathbf{E}^M \times \mathbf{E}^N | \mathbf{z} \in \mathbf{Z},\ \mathbf{x} \in \mathbf{T}(\mathbf{z})\}$. Note that $\mathbf{Y}$ is a

[11]The asymptotic cone (recession cone) of a set can be defined informally as the set of directions in which the set is unbounded. A cone is pointed if it contains no lines.

closed set if and only if the following two conditions hold: (i) $\mathbf{T(z)}$ is an upper hemicontinuous correspondence, and (iv) $(\mathbf{z}^k,\mathbf{x}^k) \in \mathbf{Y}$ and $(\mathbf{z}^k,\mathbf{x}^k)$ converging to $(\mathbf{z}^0,\mathbf{x}^0)$ with $\mathbf{x}^0$ finite implies $\mathbf{z}^0 \in \mathbf{Z}$. Then, strong continuity of $\mathbf{T(z)}$ is neither necessary nor sufficient for the set $\mathbf{Y}$ to be closed. However, (i) is common to both conditions.

Figure 15 illustrates the geometry of these continuity conditions. In each case except (c), $\mathbf{T(z)}$ is an upper hemicontinuous correspondence. In case (c), the point $\mathbf{x}^0$, obtained as a limit of points in $\mathbf{T(z)}$ for $\mathbf{z}$ approaching $\mathbf{z}^0$ from below, is not contained in $\mathbf{T}(\mathbf{z}^0) = \{\mathbf{x}|\mathbf{x} \leqq \mathbf{x}^1\}$. In cases (a), (b), (d), (e) the set $\mathbf{Y}$ is closed. Note that the set $\mathbf{Z}$ may or may not be closed [cases (a), (b), respectively] even though $\mathbf{Y}$ is closed. In case (c), $\mathbf{Y}$ fails to be closed because upper hemicontinuity of $\mathbf{T(z)}$ is absent. In case (f), $\mathbf{Y}$ fails to be closed because property (iv) fails to hold.

Lower hemicontinuity holds in Figure 15 in every case except (d), where it fails at the point $(\mathbf{z}^0,\mathbf{x}^0)$ for $\mathbf{z}$ approaching $\mathbf{z}^0$ from above. Finally, case (e) gives an example in which condition (iii) on the upper hemicontinuity of the asymptotic cone fails, since a sequence $\mathbf{z}^k$ converging to $\mathbf{z}^0$ from above and $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$ with $\mathbf{x}^k \to -\infty$ has $\theta_k \mathbf{x}^k = -1$ when $\theta_k = 1/|\mathbf{x}^k| \to 0$, whereas $\mathbf{T}(\mathbf{z}^0)$ is bounded, and any sequence $\hat{\theta}_k \hat{\mathbf{x}}^k$ with $\hat{\theta}_k$ converging to zero and $\hat{\mathbf{x}}^k \in \mathbf{T}(\mathbf{z}^0)$ must also converge to zero.

Insight into the economic restrictions imposed by strong continuity can be gained by two interpretations of the examples in Figure 15. First, suppose that $\mathbf{T(z)}$ determines the level of a single input (x) required to produce a specified output level (z). The normally imposed condition that the overall production possibility set $\mathbf{Y}$ be closed will imply upper hemicontinuity. Strong upper hemicontinuity seems to rule out only pathological cases such as (e). Lower hemicontinuity rules out cases such as (d) where there is a "plateau" at which additional input fails to yield more output, and in a multiple-input case rules out "thick" isoquants. Intuitively then, lower hemicontinuity implies that some small change in the input bundle must be productive.

For the second interpretation, suppose that $\mathbf{T(z)}$ specifies the level of a single input (x) required to produce a unit of output at different levels of technological knowledge (z) (with low z corresponding to advanced technology). Strong continuity then requires a steady progression of the state of the arts, without "breakthroughs" such as at z^0 in case (d).

Axiom 3. The production possibility set $\mathbf{T(z)}$ is strongly continuous on $\mathbf{Z}$.

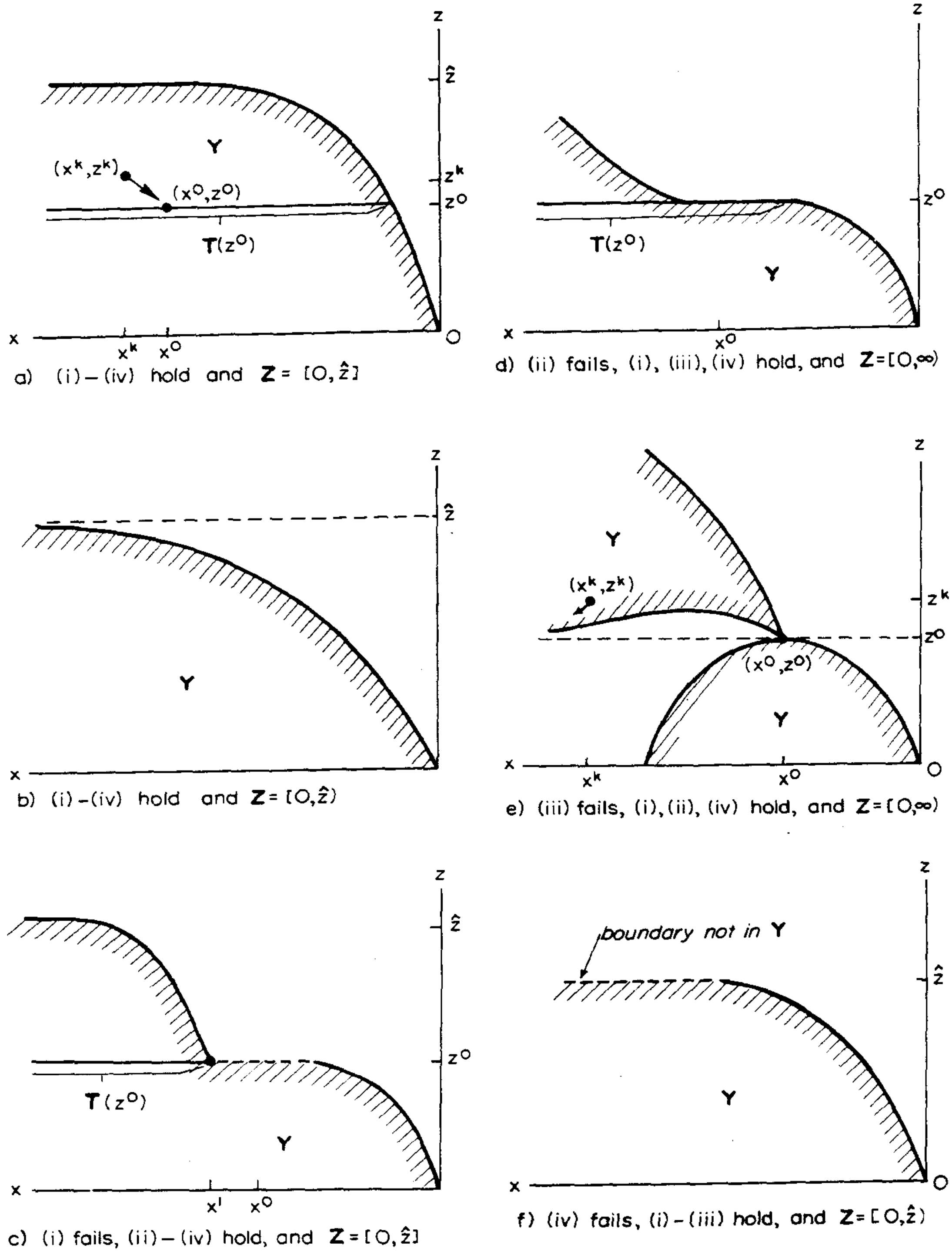

z
ẑ
Y
(xᵏ,zᵏ)
(x⁰,z⁰)
zᵏ
z⁰
T(z⁰)
x
xᵏ x⁰
O
a) (i)–(iv) hold and Z = [O, ẑ]
z
z⁰
T(z⁰)
Y
x
x⁰
d) (ii) fails, (i), (iii), (iv) hold, and Z=[O,∞)
z
ẑ
Y
x
b) (i)–(iv) hold and Z = [O,ẑ)
z
Y
(xᵏ,zᵏ)
zᵏ
z⁰
(x⁰,z⁰)
Y
x
xᵏ
x⁰
O
e) (iii) fails, (i), (ii), (iv) hold, and Z=[O,∞)
z
ẑ
z⁰
T(z⁰)
Y
x
x¹ x⁰
c) (i) fails, (ii)–(iv) hold, and Z = [O, ẑ]
z
boundary not in Y
ẑ
Y
x
O
f) (iv) fails, (i)–(iii) hold, and Z= [O,ẑ)

FIGURE 15

The basic Axioms 1–3 play the same role in the following analysis as did the assumption of input-regularity in the development of the cost function. The reader will recall that in the previous treatment, further assumptions of free disposal and convexity were often used with the argument that the economic behavior implied by cost minimization would always be consistent with these conditions. The same argument will be used to justify the following axiom.

Axiom 4. For each $\mathbf{z} \in \mathbf{Z}$, the technology $\mathbf{T}(\mathbf{z})$ is convex; i.e., for any netput bundles $\mathbf{x}^0$, $\mathbf{x}^1 \in \mathbf{T}(\mathbf{z})$ and weighted average $\mathbf{x}^* = \theta\mathbf{x}^0 + (1-\theta)\mathbf{x}^1$, $0 < \theta < 1$, it follows that $\mathbf{x}^* \in \mathbf{T}(\mathbf{z})$.

Lemma 13.3(2) in Appendix A.3 establishes that a technology satisfying Axioms 1, 2, and 4, plus the upper and lower hemicontinuity conditions (i) and (ii) in the definition of strong continuity, must also satisfy condition (iii), and hence Axiom 3.

The technology $\mathbf{T}(\mathbf{z})$ is said to exhibit *free disposal of inputs and outputs* if $\mathbf{x} \in \mathbf{T}(\mathbf{z})$ and $\mathbf{x}^1 \leqq \mathbf{x}$ imply $\mathbf{x}^1 \in \mathbf{T}(\mathbf{z})$. When $\mathbf{T}(\mathbf{z})$ satisfies this condition, all price vectors $\mathbf{p}$ in the normal cone $\mathbf{P}(\mathbf{z})$ of $\mathbf{T}(\mathbf{z})$ must be non-negative. Conversely, the existence of price vectors in $\mathbf{P}(\mathbf{z})$ with negative components indicates a lack of free disposability of the corresponding commodities. Free disposal and related assumptions will be discussed in the next section.

14. The General Restricted Profit Function

Consider a firm with a technology $\mathbf{T}(\mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, satisfying Axioms 1–3. Suppose that the firm faces a competitive price vector $\mathbf{p} \in \mathbf{E}^N$ for the commodities in its production plan, and desires to maximize its profit $\pi = \mathbf{p}\cdot\mathbf{x}$ over $\mathbf{x} \in \mathbf{T}(\mathbf{z})$. Recalling that $\mathbf{P}(\mathbf{z})$ is the set of price vectors $\mathbf{p}$ for which $\pi = \mathbf{p}\cdot\mathbf{x}$ is bounded above over $\mathbf{x} \in \mathbf{T}(\mathbf{z})$, define the *restricted profit function* of the firm by

$$\pi = \Pi(\mathbf{z},\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{x} \mid \mathbf{x} \in \mathbf{T}(\mathbf{z})\} \quad \text{for} \quad \mathbf{p} \in \mathbf{P}(\mathbf{z}). \tag{35}$$

The restricted profit function gives the least upper bound on the level of profits that can be attained with a parameter vector $\mathbf{z}$ and a price vector $\mathbf{p}$.

The restricted profit function Π is *convex* in $\mathbf{p}$ for fixed $\mathbf{z}$ if for any $\mathbf{p}$, $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z})$ and weighted average $\mathbf{p}^1 = \theta\mathbf{p} + (1-\theta)\mathbf{p}^0$, with $0 < \theta < 1$, it follows that $\mathbf{p}^1 \in \mathbf{P}(\mathbf{z})$ and $\Pi(\mathbf{z},\mathbf{p}^1) \leqq \theta\Pi(\mathbf{z},\mathbf{p}) + (1-\theta)\Pi(\mathbf{z},\mathbf{p}^0)$. This function is *positively linear homogeneous* in $\mathbf{p}$ for fixed $\mathbf{z}$ if for any $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ and $\lambda > 0$, it follows that $\lambda\mathbf{p} \in \mathbf{P}(\mathbf{z})$ and $\Pi(\mathbf{z},\lambda\mathbf{p}) = \lambda\Pi(\mathbf{z},\mathbf{p})$. This function is *closed* in $\mathbf{p}$ for fixed $\mathbf{z}$ if for any sequence $\mathbf{p}^k \in \mathbf{P}(\mathbf{z})$ converging to $\mathbf{p}^0$, either (a) $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z})$ and $\Pi(\mathbf{z},\mathbf{p}^0) = \lim_{\mathbf{p}\in\mathbf{P}(\mathbf{z}),\mathbf{p}\to\mathbf{p}^0} \inf \Pi(\mathbf{z},\mathbf{p})$, or (b) $\mathbf{p}^0 \notin \mathbf{P}(\mathbf{z})$ and $\lim_{\mathbf{p}\in\mathbf{P}(\mathbf{z}),\mathbf{p}\to\mathbf{p}^0} \inf \Pi(\mathbf{z},\mathbf{p}) = +\infty$.[12] Restating the last condition less formally, the set of prices for which profit is bounded above contains a boundary point $\mathbf{p}^0$ if and only if profits are uniformly bounded above for some sequence of prices in $\mathbf{P}(\mathbf{z})$ approaching $\mathbf{p}^0$.

The next result establishes the basic properties of the restricted profit function.

Lemma 11. Suppose a technology $\mathbf{T}(\mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, satisfies Axioms 1 and 2. Then, the following conclusions hold:

(1) The set $\mathbf{P}(\mathbf{z})$ is a convex cone, and its interior $\mathbf{P}^0(\mathbf{z})$ is non-empty.

(2) For each $\mathbf{z} \in \mathbf{Z}$, $\Pi(\mathbf{z},\mathbf{p})$ is a convex, positively linear homogeneous, closed function of $\mathbf{p} \in \mathbf{P}(\mathbf{z})$.

(3) For each $\mathbf{z} \in \mathbf{Z}$, $\Pi(\mathbf{z},\mathbf{p})$ is a continuous function of $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$, and satisfies

$$\Pi(\mathbf{z},\mathbf{p}) = \mathrm{Max}\{\mathbf{p}\cdot\mathbf{x} \mid \mathbf{x} \in \mathbf{T}(\mathbf{z})\}, \tag{36}$$

i.e., a profit maximizing netput bundle can be attained for each $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$.

(4) The closed convex hull of $\mathbf{T}(\mathbf{z})$ is equal to the set

$$\tilde{\mathbf{T}}(\mathbf{z}) = \{\mathbf{x} \in \mathbf{E}^N \mid \mathbf{p}\cdot\mathbf{x} \leqq \Pi(\mathbf{z},\mathbf{p}) \quad \text{for all} \quad \mathbf{p} \in \mathbf{P}(\mathbf{z})\}. \tag{37}$$

If Axiom 4 holds, then $\mathbf{T}(\mathbf{z}) = \tilde{\mathbf{T}}(\mathbf{z})$.

Proof: In the terminology of Appendix A.3, Sections 8 and 9, $\mathbf{P}(\mathbf{z})$ is the normal cone of $\mathbf{T}(\mathbf{z})$ and $\Pi(\mathbf{z},\mathbf{p})$ is the support function of $\mathbf{T}(\mathbf{z})$. Axiom 2 implies $\mathbf{P}^0(\mathbf{z})$ non-empty, and the convexity of the normal cone is a standard result (Appendix A.3, Section 10.18). Lemma 12.4 in the

[12]The notation $\lim_{\mathbf{p}\in\mathbf{P}(\mathbf{z}),\mathbf{p}\to\mathbf{p}^0} \inf \Pi(\mathbf{z},\mathbf{p})$ means the greatest lower bound of the set of all limit points of $\Pi(\mathbf{z},\mathbf{p})$ for all sequences in $\mathbf{P}(\mathbf{z})$ converging to $\mathbf{p}^0$.

appendix establishes results (2) and (4).[13] Appendix A.3, Lemma 13.5(1) implies that a profit maximizing netput bundle can be attained for each $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$, and Appendix A.3, Lemma 12.1(1) implies that Π is continuous in $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$. Q.E.D.

Figure 16 illustrates several simple technologies and the corresponding profit functions. (a) gives a case in which the normal cone of the technology fails to be closed. Note that the profit function for this case is not continuous at $\mathbf{p} = \mathbf{0}$. [E.g., $p_1 = p_2 + p_2^2$ and $p_2 \to 0$ implies $\Pi \to 1/4 \neq \Pi(0)$]. Hence, the conclusion of Lemma 11 that the profit function is continuous for prices in the interior of the normal cone and lower semicontinuous on all prices in the normal cone cannot be strengthened without further hypotheses.

A netput bundle $\mathbf{x} \in \mathbf{T}(\mathbf{z})$ is *exposed* if there exists a price vector $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ such that $\mathbf{p}\cdot\mathbf{x} > \mathbf{p}\cdot\mathbf{x}^1$ for all distinct $\mathbf{x}^1 \in \mathbf{T}(\mathbf{z})$. The next result gives conditions under which the profit function is continuous on $\mathbf{P}(\mathbf{z})$.

Lemma 12. If Axioms 1 and 2 hold, then any one of the following conditions is sufficient to imply that for each $\mathbf{z} \in \mathbf{Z}$, the restricted profit function is continuous in $\mathbf{p}$ on $\mathbf{P}(\mathbf{z})$:

(1) $\mathbf{P}(\mathbf{z})$ is closed and can be represented as the convex cone spanning a finite number of points.
(2) $\mathbf{T}(\mathbf{z})$ is bounded.
(3) The set of exposed points in $\mathbf{T}(\mathbf{z})$ is bounded.

Proof: Appendix A.3, Lemma 12.7 implies (1). If $\mathbf{T}(\mathbf{z})$ is bounded, then its asymptotic cone is empty and thus $\mathbf{P}(\mathbf{z}) = \mathbf{E}^N$ by result 10.16 in this appendix, and (2) is implied by (1). Finally, we prove (3).

[13]These proofs employ the fundamental mathematical theory of convex conjugate functions, from which many other implications can be easily derived. Alternately, Result (2) can be proved directly using the simple, pedagogically appealing arguments employed in deriving the properties of cost functions: If $\mathbf{p}, \mathbf{p}^0 \in \mathbf{P}(\mathbf{z})$ and $\mathbf{p}^1 = \theta\mathbf{p} + (1-\theta)\mathbf{p}^0$, $0 < \theta < 1$, and if $\mathbf{x}^k \in \mathbf{T}(\mathbf{z})$ is a sequence with $\mathbf{p}^1\cdot\mathbf{x}^k \to \Pi(\mathbf{z},\mathbf{p}^1)$, then $\mathbf{p}^1\cdot\mathbf{x}^k = \theta\mathbf{p}\cdot\mathbf{x}^k + (1-\theta)\mathbf{p}^0\cdot\mathbf{x}^k \leqq \theta\Pi(\mathbf{z},\mathbf{p}) + (1-\theta)\Pi(\mathbf{z},\mathbf{p}^0)$, implying in the limit $\mathbf{p}^1 \in \mathbf{P}(\mathbf{z})$ and $\Pi(\mathbf{z},\mathbf{p}^1) \leqq \theta\Pi(\mathbf{z},\mathbf{p}) + (1-\theta)\Pi(\mathbf{z},\mathbf{p}^0)$. Positive linear homogeneity is immediate from the definition of the profit function. Finally, a simple argument shows that Π is closed. Suppose $\mathbf{p}^0$ is in the boundary of $\mathbf{P}(\mathbf{x})$, and let $\mathbf{x}^j \in \mathbf{T}(\mathbf{z})$ be a sequence such that $\mathbf{p}^0\cdot\mathbf{x}^j \to \sup\{\mathbf{p}^0\cdot\mathbf{x} | \mathbf{x} \in \mathbf{T}(\mathbf{z})\}$. Then for any sequence $\mathbf{p}^k \in \mathbf{P}(\mathbf{z})$, $\mathbf{p}^k \to \mathbf{p}^0$, the inequality $\Pi(\mathbf{z},\mathbf{p}^k) \geqq \mathbf{p}^k\cdot\mathbf{x}^j$ implies in the limit $\lim_{k\to\infty} \inf \Pi(\mathbf{z},\mathbf{p}^k) \geqq \mathbf{p}^0\cdot\mathbf{x}^j$, and hence letting $j \to \infty$, $\lim_{k\to\infty} \inf \Pi(\mathbf{z},\mathbf{p}^k) \geqq \Pi(\mathbf{z},\mathbf{p}^0)$. Choosing the sequence $\mathbf{p}^k = k^{-1}\mathbf{p}^1 + (1-k^{-1})\mathbf{p}^0$ for some $\mathbf{p}^1 \in \mathbf{P}^0(\mathbf{z})$, one obtains from the convexity condition the opposite inequality $\lim_{k\to\infty} \Pi(\mathbf{z},\mathbf{p}^k) \leqq \Pi(\mathbf{z},\mathbf{p}^0)$. Hence, we conclude that either $\mathbf{p}^0 \notin \mathbf{P}(\mathbf{z})$ and $\lim_{\mathbf{p}\in\mathbf{P}(\mathbf{z}),\mathbf{p}\to\mathbf{p}^0} \inf \Pi(\mathbf{z},\mathbf{p}) = +\infty$, or $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z})$ and $\lim_{\mathbf{p}\in\mathbf{P}(\mathbf{z}),\mathbf{p}\to\mathbf{p}^0} \inf \Pi(\mathbf{z},\mathbf{p}) = \Pi(\mathbf{z},\mathbf{p}^0)$.

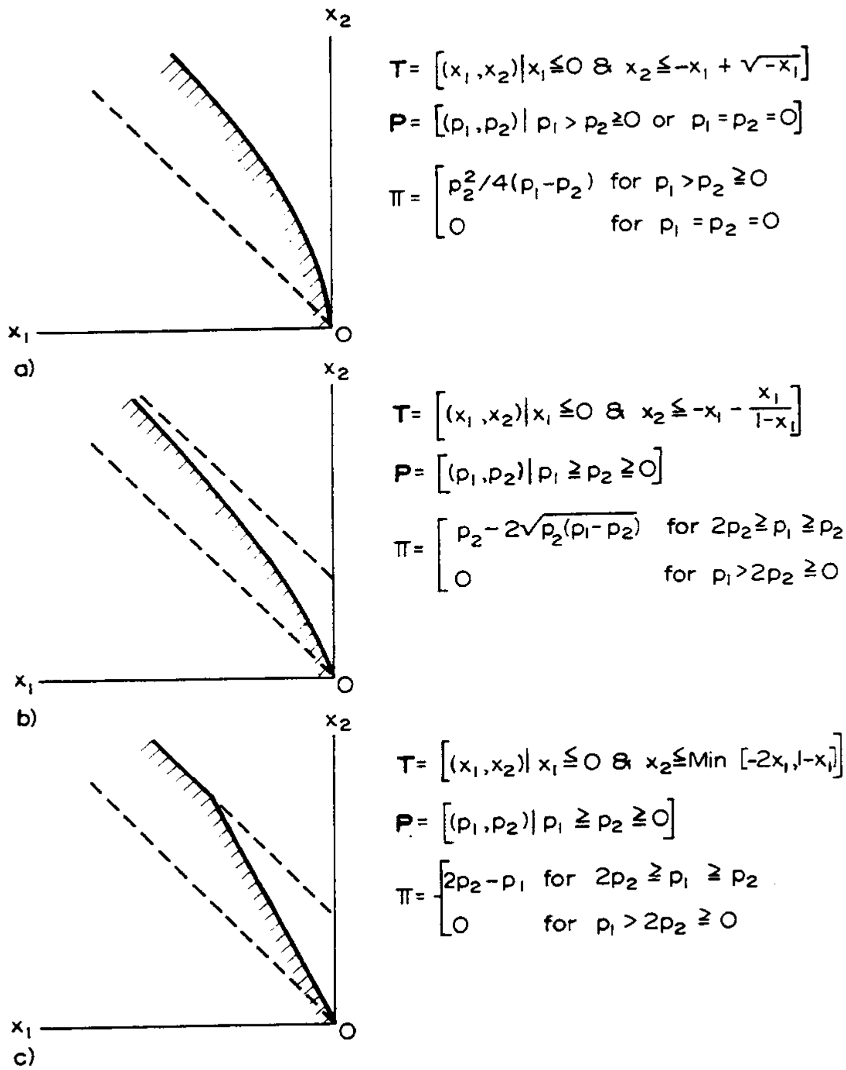


FIGURE 16

Consider a sequence $\hat{\mathbf{p}}^k \in \mathbf{P}^0(\mathbf{z})$ converging to $\mathbf{p}^0$. By results 12.1(2), 13.8(3), and 13.9 in Appendix A.3, there exists a vector $\mathbf{p}^k \in \mathbf{P}^0(\mathbf{z})$ arbitrarily close to $\hat{\mathbf{p}}^k$ and an exposed bundle $\mathbf{x}^k \in \mathbf{T}(\mathbf{z})$ such that $\Pi(\mathbf{z},\mathbf{p}^k) = \mathbf{p}^k \cdot \mathbf{x}^k$. By continuity of Π on $\mathbf{P}^0(\mathbf{z})$, $\mathbf{p}^k$ can be chosen so that $|\mathbf{p}^k - \hat{\mathbf{p}}^k| < k^{-1}$ and $|\Pi(\mathbf{z},\mathbf{p}^k) - \Pi(\mathbf{z},\hat{\mathbf{p}}^k)| < k^{-1}$. Since the exposed points $\mathbf{x}^k$ are bounded, we can extract a subsequence converging to a point $\mathbf{x}^0 \in \mathbf{T}(\mathbf{z})$ such that $\Pi(\mathbf{z},\mathbf{p}^0) \geqq \mathbf{p}^0 \cdot \mathbf{x}^0 = \lim_{k\to\infty} \sup \Pi(\mathbf{z},\hat{\mathbf{p}}^k)$. Since Π is lower semicontinuous in $\mathbf{p}$ on $\mathbf{P}(\mathbf{z})$, we have established $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z})$ and $\Pi(\mathbf{z},\mathbf{p}^0) = \lim_{\mathbf{p}\in\mathbf{P}^0(\mathbf{z})\&\mathbf{p}\to\mathbf{p}^0} \Pi(\mathbf{z},\mathbf{p})$. Finally, note that for any sequence $\hat{\mathbf{p}}^k \in \mathbf{P}(\mathbf{z})$ converging to $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z})$, we have, by the result just proved, a sequence

$\mathbf{p}^k \in \mathbf{P}^0(\mathbf{z})$ with $|\mathbf{p}^k - \hat{\mathbf{p}}^k| < k^{-1}$ and $|\Pi(\mathbf{z},\mathbf{p}^k) - \Pi(\mathbf{z},\hat{\mathbf{p}}^k)| < k^{-1}$. Then, $\Pi(\mathbf{z},\mathbf{p}^0) = \lim_{\mathbf{p}\in\mathbf{P}(\mathbf{z})\ \&\ \mathbf{p}\to\mathbf{p}^0} \Pi(\mathbf{z},\mathbf{p})$. Q.E.D.

We note from the proof of Lemma 12 that condition (3) implies $\mathbf{P}(\mathbf{z})$ closed. This lemma has one immediate corollary: *If* $N = 2$ *and* $\mathbf{P}(\mathbf{z})$ *is closed, then* (1) *is satisfied and* Π *is continuous in* $\mathbf{p}$ *on* $\mathbf{P}(\mathbf{z})$. Figure 16, (b) and (c), illustrates cases in which $\mathbf{P}(\mathbf{z})$ is closed, and the resulting profit functions are continuous in $\mathbf{p}$ on $\mathbf{P}(\mathbf{z})$ in accord with this corollary.

When the technology exhibits free disposal of inputs and outputs, it is convenient to distinguish bundles $\mathbf{x} \in \mathbf{T}(\mathbf{z})$ which are *efficient* in that no distinct bundle $\mathbf{x}^1 \in \mathbf{T}(\mathbf{z})$ has $\mathbf{x}^1 \geqq \mathbf{x}$. Since under this assumption any exposed $\mathbf{x}$ must be efficient, (3) in Lemma 12 implies the following corollary: *If Axioms* 1 *and* 2 *hold, the technology exhibits free disposal of inputs and outputs, and the set of efficient points is bounded, then the profit function is continuous in* $\mathbf{p}$ *on* $\mathbf{P}(\mathbf{z})$.

A technology $\mathbf{T}(\mathbf{z})$ is *bounded above* if there exists $\mathbf{x}^0$ such that $\mathbf{x} \leqq \mathbf{x}^0$ for all $\mathbf{x} \in \mathbf{T}(\mathbf{z})$. In the case of cost minimization, in which the netput bundles in $\mathbf{T}(\mathbf{z})$ are the negative of input bundles, $\mathbf{T}(\mathbf{z})$ is bounded above by the origin. In the case of profit maximization with all inputs fixed, or with some input fixed which is essential to production, the technology is generally bounded above by some positive $\mathbf{x}^0$. For these cases, the following corollary is useful: *If Axioms* 1 *and* 2 *hold, and the technology is bounded above and exhibits free disposal of inputs and outputs, then the profit function is continuous in* $\mathbf{p}$ *on* $\mathbf{P}(\mathbf{z})$. Under the hypotheses of this corollary, the normal cone of the technology is the non-negative orthant of $\mathbf{E}^N$, and (1) in Lemma 12 implies the conclusion.

For $N > 2$, $\mathbf{P}(\mathbf{z})$ closed is not in general sufficient to imply continuity of the profit function over $\mathbf{P}(\mathbf{z})$. A counter-example of Gale, Klee, and Rockafellar (1968, Lemma and proof of Theorem 2) gives a convex function which is lower semicontinuous (closed), but not upper semicontinuous. The domain of this function can be taken to be the set of $\mathbf{p} \in \mathbf{E}^N$ satisfying $\mathbf{p}\cdot\mathbf{p} \leqq 1$ and $\sum_{n=1}^{N} p_n = 1/2$. Form the closed convex semi-bounded cone spanned by this domain, and define a linear homogeneous extension of this function on the cone. Then, Lemma 12.5 in Appendix A.3 implies that the resulting function is the profit function for a technology satisfying Axioms 1 and 2.

Lemma 11(3) establishes that a profit maximizing netput bundle can be attained for $\mathbf{p}$ in the interior of the normal cone of the technology. (b) in Figure 16 illustrates a case in which a maximum cannot be achieved at the boundary price vector $\mathbf{p} = (1,1)$, and (c) illustrates a case in which a maximum can be achieved at this boundary price vector. Let $\mathbf{P}^1(\mathbf{z})$ denote the set of $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ for which a profit maximizing netput bundle can be attained. These examples show that $\mathbf{P}^1(\mathbf{z})$ may be neither open nor closed in general. A further example given in Appendix A.3, 13.6, shows that $\mathbf{P}^1(\mathbf{z})$ need not be convex, although its interior, equal to $\mathbf{P}^0(\mathbf{z})$, is convex. The following result gives one condition under which the normal cone of the technology is closed and a profit maximizing bundle can be achieved for each price vector in this cone.

Lemma 13. If Axioms 1 and 2 hold, and the set of exposed points in $\mathbf{T}(\mathbf{z})$ is bounded, then $\mathbf{P}(\mathbf{z})$ is closed and $\mathbf{P}^1(\mathbf{z}) = \mathbf{P}(\mathbf{z})$.

Proof: Let $\mathbf{T}^1$ denote the closed convex hull of $\mathbf{T}(\mathbf{z})$, $\mathbf{T}^2$ denote the asymptotic cone of $\mathbf{T}^1$, and $\mathbf{T}^3$ denote the closed convex hull of the set of exposed points in $\mathbf{T}(\mathbf{z})$. By hypothesis, $\mathbf{T}^3$ is closed and bounded. By Appendix A.3, Lemma 14.3, $\mathbf{T}^1 = \mathbf{T}^2 + \mathbf{T}^3$. For any $\mathbf{p}^0$ in the closure of $\mathbf{P}(\mathbf{z})$, $\mathbf{p}^0 \cdot \mathbf{x} \leqq 0$ for any $\mathbf{x} \in \mathbf{T}^2$. Hence, the supremum of $\mathbf{p}^0 \cdot \mathbf{x}$ for $\mathbf{x} \in \mathbf{T}^1$ is approached by $\mathbf{x} \in \mathbf{T}^3$, and is therefore achieved at some $\mathbf{x}^0 \in \mathbf{T}^3 \subseteq \mathbf{T}^1$. But if a linear function achieves a maximum on the convex hull of a closed set, then it achieves a maximum on the set. Hence, a profit maximizing bundle for $\mathbf{p}^0$ can be found in $\mathbf{T}(\mathbf{z})$, implying $\mathbf{p}^0 \in \mathbf{P}^1(\mathbf{z})$ and $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z})$. Q.E.D.

Further conditions for the convexity and closedness of $\mathbf{P}^1(\mathbf{z})$ have been given by Winter (forthcoming). We next establish several additional properties of profit maximizing netput bundles. For $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$, let $\mathbf{\Phi}(\mathbf{z},\mathbf{p})$ denote the set of netput bundles in $\mathbf{T}(\mathbf{z})$ which maximize profit.

Lemma 14. Suppose Axioms 1 and 2 hold. Then, $\mathbf{\Phi}(\mathbf{z},\mathbf{p})$ has the following properties:

(1) For $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$, $\mathbf{\Phi}(\mathbf{z},\mathbf{p})$ is closed, and bounded.

(2) For any closed, bounded subset $\mathbf{R}$ of $\mathbf{P}^0(\mathbf{z})$, the set $\bigcup_{\mathbf{p}\in\mathbf{R}} \mathbf{\Phi}(\mathbf{z},\mathbf{p})$ is bounded.

(3) $\mathbf{\Phi}(\mathbf{z},\mathbf{p})$ is an upper hemicontinuous correspondence in $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$ for each $\mathbf{z} \in \mathbf{Z}$; i.e., if $\mathbf{p}^k \in \mathbf{P}^0(\mathbf{z})$, $\mathbf{p}^k \to \mathbf{p}^0 \in \mathbf{P}^0(\mathbf{z})$, $\mathbf{x}^k \in \mathbf{\Phi}(\mathbf{z},\mathbf{p}^k)$, $\mathbf{x}^k \to \mathbf{x}^0$, then $\mathbf{x}^0 \in \mathbf{\Phi}(\mathbf{z},\mathbf{p}^0)$.

(4) If Axiom 4 holds, then $\mathbf{\Phi}(\mathbf{z},\mathbf{p})$ is convex set.

(5) $\mathbf{\Phi}(\mathbf{z},\mathbf{p})$ is positively homogeneous of degree zero in $\mathbf{p}$; i.e., for $\lambda > 0$, $\mathbf{\Phi}(\mathbf{z},\lambda\mathbf{p}) = \mathbf{\Phi}(\mathbf{z},\mathbf{p})$.

Proof: Appendix A.3, Lemma 13.5 establishes results (1)–(4); the proof of result (5) is trivial.

Now consider the behavior of the restricted profit function under joint variation of the parameter vector $\mathbf{z}$ and the price vector $\mathbf{p}$. The first result establishes the behavior of the normal cone of the technology under variations in $\mathbf{z}$.

Lemma 15. Suppose Axioms 1–3 hold. Then, $\mathbf{P}(\mathbf{z})$ has the following properties:

(1) $\mathbf{P}(\mathbf{z})$ is a lower hemicontinuous correspondence on $\mathbf{Z}$; i.e., $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z}^0)$ implies the existence of $\mathbf{p}^k \in \mathbf{P}(\mathbf{z}^k)$ such that $\mathbf{p}^k \to \mathbf{p}^0$.

(2) If $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, and $\mathbf{R}$ is a non-empty, closed, bounded subset of $\mathbf{P}^0(\mathbf{z}^0)$, then there exists a k_0 such that $\mathbf{R} \subseteq \mathbf{P}^0(\mathbf{z}^k)$ for $k \geqq k_0$ and the set $\cup_{k \geqq k_0} \cup_{\mathbf{p} \in \mathbf{R}} \mathbf{\Phi}(\mathbf{z}^k, \mathbf{p})$ is bounded.

Proof: Appendix A.3, Lemma 15.2.

Figure 17 illustrates the results of Lemmas 14 and 15 for a simple case in which $\mathbf{Z}$ is the non-negative real line and

$$\mathbf{T}(z) = \{(x_1, x_2) \in \mathbf{E}^2 | x_1 \leqq -\theta, x_2 \leqq \theta - z\theta^2 \text{ for some } \theta \geqq 0\}. \tag{38}$$

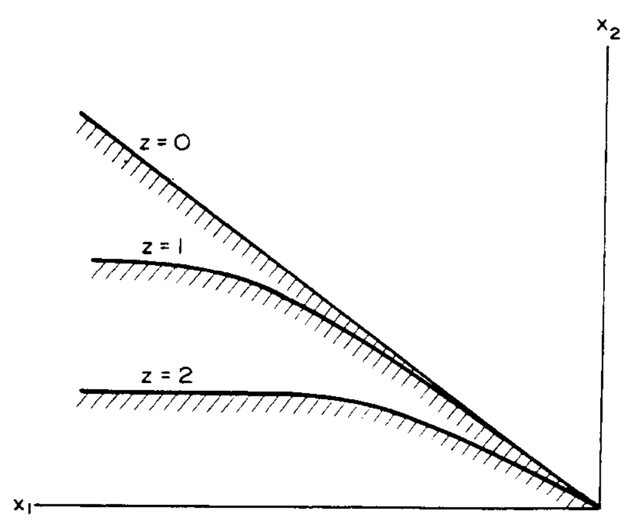


FIGURE 17

The normal cone of this technology is the non-negative orthant of $\mathbf{E}^2$ when z is positive, and is the set of (p_1,p_2) satisfying $0 \leqq p_2 \leqq p_1$ when $z = 0$. Note that this cone is a continuous correspondence at any positive z, but is only a lower hemicontinuous correspondence at $z = 0$. The restricted profit function for this example is

$$\begin{aligned} \Pi(z,\mathbf{p}) &= (p_2 - p_1)^2/4zp_2 \quad \text{if} \quad z > 0 \quad \text{and} \quad p_2 > p_1 \geqq 0, \\ &= 0 \quad \text{if} \quad 0 \leqq p_2 \leqq p_1 \quad \text{and} \quad z \geqq 0. \end{aligned} \tag{39}$$

The set of profit maximizing netput bundles for $z > 0$ is

$$\begin{aligned} \mathbf{\Phi}(z,\mathbf{p}) &= \{((p_1 - p_2)/2zp_2, (p_2^2 - p_1^2)/4zp_2^2)\} \quad \text{for} \quad p_2 > p_1 \geqq 0, \\ &= \{(0,0)\} \quad \text{for} \quad 0 < p_2 \leqq p_1 \\ &\qquad \text{or} \quad 0 = p_2 < p_1, \\ &= \mathbf{T}(\mathbf{z}) \quad \text{for} \quad p_1 = p_2 = 0. \end{aligned} \tag{40a}$$

The set of profit maximizing netput bundles for $z = 0$ is

$$\begin{aligned} \mathbf{\Phi}(0,\mathbf{p}) &= \{(0,0)\} \quad \text{for} \quad 0 \leqq p_2 < p_1, \\ &= \{(-\theta,\theta) \mid 0 \leqq \theta < +\infty\} \quad \text{for} \quad 0 < p_1 = p_2, \\ &= \mathbf{T}(0) \quad \text{for} \quad 0 = p_1 = p_2. \end{aligned} \tag{40b}$$

For each $z \in \mathbf{Z}$, the set of maximands is seen to be bounded for $\mathbf{p} \in \mathbf{P}^0(z)$, and to be upper semicontinuous on $\mathbf{P}(z)$. Note in this example that for a point such as $\mathbf{p} = (1,2) \in \mathbf{P}(z)$ for $z > 0$ with $\mathbf{p} \notin \mathbf{P}(0)$, one has $\lim_{z \to 0} \Pi(z,\mathbf{p}) = +\infty$, while for a point such as $\mathbf{p} = (2,1) \in \mathbf{P}^0(z)$ for $z \geqq 0$, one has $\lim_{z \to 0} \Pi(z,\mathbf{p}) = \Pi(0,\mathbf{p})$. This property that the profit function is continuous in z and $\mathbf{p}$ jointly at a point with $\mathbf{p} \in \mathbf{P}^0(z)$, and that $\Pi(z,\mathbf{p})$ approaches infinity as $(z,\mathbf{p})$ approach a point at which the price vector is not contained in the corresponding normal cone, is a general one, as the following result shows.

Lemma 16. Suppose Axioms 1–3 hold. Then, the restricted profit function $\Pi(\mathbf{z},\mathbf{p})$ is continuous jointly in $\mathbf{z}$ and $\mathbf{p}$ at each $\mathbf{z}^0 \in \mathbf{Z}$ and $\mathbf{p}^0 \in \mathbf{P}^0(\mathbf{z}^0)$. Further, at any $\mathbf{z}^0 \in \mathbf{Z}$ and $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z}^0)$, $\Pi(\mathbf{z},\mathbf{p})$ is lower semicontinuous in $(\mathbf{z},\mathbf{p})$; i.e., if $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z}^0)$, then $\Pi(\mathbf{z}^0,\mathbf{p}^0) = \lim_{\mathbf{p}^k \in \mathbf{P}(\mathbf{z}^k),(\mathbf{z}^k,\mathbf{p}^k) \to (\mathbf{z}^0,\mathbf{p}^0)} \inf \Pi(\mathbf{z}^k,\mathbf{p}^k)$. Finally, if a sequence $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{p}^k \in \mathbf{P}(\mathbf{z}^k)$ converges to $\mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{p}^0 \notin \mathbf{P}(\mathbf{z}^0)$, then $\lim_k \Pi(\mathbf{z}^k,\mathbf{p}^k) = +\infty$.

Proof: Appendix A.3, Lemma 15.3.

15. The Derivative Property of the Restricted Profit Function

In Section 5, the cost function was shown to have the useful property that its partial derivatives with respect to input prices were equal, when they existed, to the corresponding cost minimizing input demands. We now establish a similar property for the restricted profit function: the vector of partial derivatives of this function with respect to commodity prices, when it exists, equals a unique profit maximizing netput bundle. Further, the vector of partial derivatives is found to exist for almost all commodity vectors. Finally, employing a generalization of the ordinary concept of a derivative, the identification of the "derivative" with the set of profit maximizing netput bundles can be shown to hold for all commodity prices. The first result concerns the differentiability of the restricted profit function.

> *Lemma* 17. Suppose Axioms 1 and 2 hold. For fixed $\mathbf{z} \in \mathbf{Z}$, the profit function $\Pi(\mathbf{z},\mathbf{p})$, considered as a function of $\mathbf{p}$, possesses a first and second differential on a set $\mathbf{P}^2(\mathbf{z}) \subseteq \mathbf{P}^0(\mathbf{z})$, where the set of points in $\mathbf{P}^0(\mathbf{z})$, but not in $\mathbf{P}^2(\mathbf{z})$, has Lebesgue measure zero. The vector of first order partial derivatives of Π with respect to $\mathbf{p}$, denoted by $\Pi_p(\mathbf{z},\mathbf{p})$ and termed the *gradient*, is continuous in $\mathbf{P}^2(\mathbf{z})$. At each $\mathbf{p} \in \mathbf{P}^2(\mathbf{z})$, the matrix of second-order partial derivatives of Π with respect to $\mathbf{p}$, denoted by $\Pi_{pp}(\mathbf{z},\mathbf{p})$ and termed the *Hessian*, is symmetric and non-negative definite.

Proof: Appendix A.3, Lemma 12.1.

From the derivative property of the cost function and its relation to the curvature of the boundary of an input requirement set, as illustrated in Figures 3 and 11; it is clear that the *set* of minimizing bundles coincides with the *set* of normals to "tangent planes", or supporting planes, to the cost function, appropriately scaled. To generalize this concept, we define the *sub-differential* of Π with respect to $\mathbf{p}$ at a point $\mathbf{z} \in \mathbf{Z}$ and $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$ as the set of points $\mathbf{x} \in \mathbf{E}^N$ such that for all $\mathbf{q} \in \mathbf{E}^N$, it follows that

$$\mathbf{q}\cdot\mathbf{x} \leqq \lim_{\theta \to 0^+} \inf(\Pi(\mathbf{z},\mathbf{p}+\theta\mathbf{q}) - \Pi(\mathbf{z},\mathbf{p}))/\theta. \tag{41}$$

When Π is differentiable in $\mathbf{p}$ at $(\mathbf{z},\mathbf{p})$, then the limit of the right-hand side of (41) exists and equals $\mathbf{q}\cdot\Pi_p(\mathbf{z},\mathbf{p})$. Hence, the sub-differential equals $\{\Pi_p(\mathbf{z},\mathbf{p})\}$ when the gradient $\Pi_p(\mathbf{z},\mathbf{p})$ exists. In Figure 11, the sub-differential of a function with the illustrated contour at the "kink" $\mathbf{v}^0$ is the closed line segment joining $\mathbf{r}^0$ and $\mathbf{r}^1$. The next result establishes the properties of the sub-differential of Π with respect to $\mathbf{p}$, which will be denoted hereafter by $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$.

Lemma 18. Suppose Axioms 1 and 2 hold. Then, the sub-differential $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$ exists for all $\mathbf{z} \in \mathbf{Z}$, $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$, with the following properties:

(1) $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$ is a non-empty, convex, closed, and bounded set, with $\mathbf{x} \in \boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$ if and only if $\boldsymbol{\Pi}(\mathbf{z},\mathbf{p}) = \mathbf{p}\cdot\mathbf{x}$ and $\boldsymbol{\Pi}(\mathbf{z},\mathbf{q}) \geqq \mathbf{q}\cdot\mathbf{x}$ for all $\mathbf{q} \in \mathbf{P}(\mathbf{z})$.

(2) $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$ is an upper hemicontinuous correspondence in $\mathbf{p}$; i.e., if $\mathbf{p}^k \in \mathbf{P}^0(\mathbf{z})$, $\mathbf{x}^k \in \boldsymbol{\Gamma}(\mathbf{z},\mathbf{p}^k)$, $(\mathbf{p}^k,\mathbf{x}^k) \to (\mathbf{p}^0,\mathbf{x}^0)$ with $\mathbf{p}^0 \in \mathbf{P}^0(\mathbf{z})$, then $\mathbf{x}^0 \in \boldsymbol{\Gamma}(\mathbf{z},\mathbf{p}^0)$.

Proof: Appendix A.3, Lemma 13.8.

We can now state the basic derivative property of the restricted profit function. Recall that $\boldsymbol{\Phi}(\mathbf{z},\mathbf{p})$ denotes the set of profit maximizing netput bundles for $\mathbf{z} \in \mathbf{Z}$ and $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$.

Lemma 19. Suppose Axioms 1 and 2 hold. Then, for $\mathbf{z} \in \mathbf{Z}$ and $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$, the sub-differential $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$ equals the convex hull of the set of profit maximizing netput bundles $\boldsymbol{\Phi}(\mathbf{z},\mathbf{p})$. If, for any given $(\mathbf{z},\mathbf{p})$, the sub-gradient contains the unique vector $\Pi_p(\mathbf{z},\mathbf{p})$ [i.e., Π is differentiable at $(\mathbf{z},\mathbf{p})$], then there is a unique profit maximizing netput vector equal to $\Pi_p(\mathbf{z},\mathbf{p})$. If Axiom 4 holds, then $\boldsymbol{\Phi}(\mathbf{z},\mathbf{p}) = \boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$.

Proof: Appendix A.3, Corollary 13.9, except the last statement, which follows from Lemma 14(4) above.

Figure 18 illustrates the relation established in this result. For the price vector $\mathbf{p}$, the set of profit maximizing netput bundles for this technology is $\boldsymbol{\Phi}(\mathbf{z},\mathbf{p}) = \{\mathbf{x}^1,\mathbf{x}^2\}$, whereas the sub-differential is the set $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p}) = \{\mathbf{x}|\mathbf{x} = \theta\mathbf{x}^1 + (1-\theta)\mathbf{x}^2,\ 0 \leqq \theta \leqq 1\}$. Hence, all extreme points in $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$ are also in $\boldsymbol{\Phi}(\mathbf{z},\mathbf{p})$, but $\boldsymbol{\Gamma}(\mathbf{z},\mathbf{p})$ may contain additional non-extreme points which are not possible to produce. However, if $\mathbf{T}(\mathbf{z})$ is convex by Axiom 4, these sets coincide exactly.

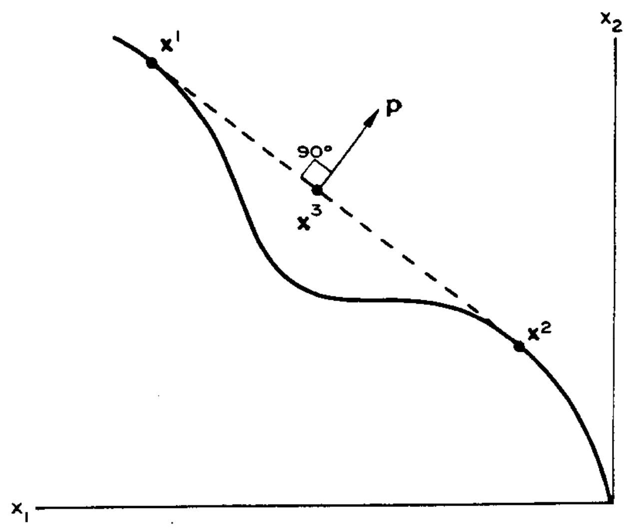


FIGURE 18

16. The Gauge Function for Production Possibilities

The introduction of a distance function to define input requirement sets yielded a convenient symmetry in the treatment of cost functions. A similar concept, the *gauge function* of the production possibility set, plays the same role in analysis of the restricted profit function.

Consider a family of production possibility sets $\mathbf{T}'(\mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, satisfying Axioms 1–4. By Result 15.6 in Appendix A.3, there exists a continuous function $\mathbf{x}^*(\mathbf{z})$ from $\mathbf{Z}$ into $\mathbf{E}^N$ with $\mathbf{x}^*(\mathbf{z}) \in \mathbf{T}'(\mathbf{z})$. [Further, $\mathbf{x}^*(\mathbf{z})$ can be chosen so that for each $\mathbf{x}' \in \mathbf{T}'(\mathbf{z})$, it follows that $\theta\mathbf{x}' + (1-\theta)\mathbf{x}^*(\mathbf{z}) \in \mathbf{T}'(\mathbf{z})$ for any sufficiently small positive or negative scalar θ. Then $\mathbf{x}^*(\mathbf{z})$ is said to be in the *relative interior* of $\mathbf{T}'(\mathbf{z})$.] Suppose one now redefines quantities of commodities by measuring them from $\mathbf{x}^*(\mathbf{z})$, so that $\mathbf{x} = \mathbf{x}' - \mathbf{x}^*(\mathbf{z})$ becomes a "translated" commodity bundle and $\mathbf{T}(\mathbf{z}) = \{\mathbf{x}' - \mathbf{x}^*(\mathbf{z}) | \mathbf{x}' \in \mathbf{T}'(\mathbf{z})\}$ becomes the "translated" technology. The translated technology continues to satisfy Axioms 1–4, and has the property that $\mathbf{0} \in \mathbf{T}(\mathbf{z})$ for all $\mathbf{z} \in \mathbf{Z}$. If, further, $\mathbf{x}^*(\mathbf{z})$ is in the relative interior of $\mathbf{T}'(\mathbf{z})$, then for any $\mathbf{x} \in \mathbf{T}(\mathbf{z})$, one has $\theta\mathbf{x} \in \mathbf{T}(\mathbf{z})$ for any sufficiently small positive or negative scalar θ, and $\mathbf{0}$ is in the *relative interior* of $\mathbf{T}(\mathbf{z})$. If $\Pi'(\mathbf{z},\mathbf{p})$, $\mathbf{p} \in \mathbf{P}'(\mathbf{z})$, is the restricted profit function of the original technology and $\Pi(\mathbf{z},\mathbf{p})$, $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, is the restricted profit function of the translated technology, for *any* translation $\mathbf{x}^*(\mathbf{z})$, then these functions are related by $\mathbf{P}'(\mathbf{z}) = \mathbf{P}(\mathbf{z})$ and $\Pi'(\mathbf{z},\mathbf{p}) = \Pi(\mathbf{z},\mathbf{p}) + \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})$.

Consider a translated technology $\mathbf{T}(\mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, satisfying Axioms 1–4, and containing the origin. Define the set

$$\mathbf{W}(\mathbf{z}) = \left\{ \mathbf{x} \in \mathbf{E}^N \,\middle|\, \frac{1}{\lambda}\mathbf{x} \in \mathbf{T}(\mathbf{z}) \text{ for some } \lambda > 0 \right\}. \tag{42}$$

Define the *gauge function* of the translated technology by

$$\begin{aligned} H(\mathbf{z},\mathbf{x}) &= +\infty && \text{if} \quad \mathbf{x} \notin \mathbf{W}(\mathbf{z}), \\ &= \inf\left\{ \lambda > 0 \,\middle|\, \frac{1}{\lambda}\mathbf{x} \in \mathbf{T}(\mathbf{z}) \right\} && \text{if} \quad \mathbf{x} \in \mathbf{W}(\mathbf{z}). \end{aligned} \tag{43}$$

Figure 19 illustrates the definition of this function. The following lemma gives its basic properties.

Lemma 20. Suppose a translated technology $\mathbf{T}(\mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, satisfies Axioms 1, 2, and 4 and contains the origin. Then, the following properties hold:

(1) $\mathbf{W}(\mathbf{z})$ is a convex cone in $\mathbf{E}^N$ for each $\mathbf{z} \in \mathbf{Z}$, and if Axiom 3 holds is a lower hemicontinuous correspondence [i.e., $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{x}^0 \in \mathbf{W}(\mathbf{z}^0)$ implies the existence of $\mathbf{x}^k \in \mathbf{W}(\mathbf{z}^k)$, $\mathbf{x}^k \to \mathbf{x}^0$]. If the origin is in the relative interior of $\mathbf{T}(\mathbf{z})$, then $\mathbf{W}(\mathbf{z})$ is a linear subspace.

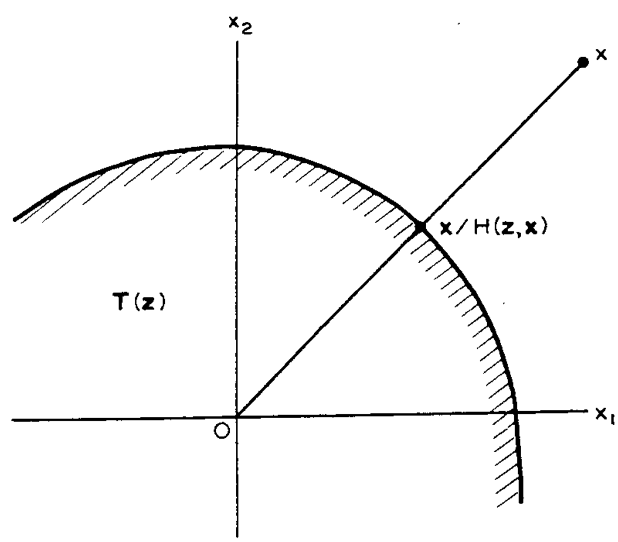


FIGURE 19

(2) For each $\mathbf{z} \in \mathbf{Z}$, $H(\mathbf{z},\mathbf{x})$ is a non-negative convex, closed,[14] positively linear homogeneous function of $\mathbf{x} \in \mathbf{W}(\mathbf{z})$.
(3) For each $\mathbf{z} \in \mathbf{Z}$, $\mathbf{T}(\mathbf{z}) = \mathbf{T}'(\mathbf{z})$, where

$$\mathbf{T}'(\mathbf{z}) = \{\mathbf{x} \in \mathbf{W}(\mathbf{z}) | H(\mathbf{z},\mathbf{x}) \leqq 1\}, \tag{44}$$

and $\mathbf{T}'(\mathbf{z})$ is semi-bounded; i.e., its asymptotic cone, given by the set of $\mathbf{x}$ with $H(\mathbf{z},\mathbf{x}) = 0$, is semi-bounded.
(4) For each $\mathbf{z} \in \mathbf{Z}$, $H(\mathbf{z},\mathbf{x})$ is continuous in $\mathbf{x}$ in the relative interior of $\mathbf{W}(\mathbf{z})$.
(5) If Axiom 3 holds, then $H(\mathbf{z},\mathbf{x})$ is lower semicontinuous jointly in $\mathbf{z}$ and $\mathbf{x}$ at each $\mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{x}^0 \in \mathbf{W}(\mathbf{z}^0)$; i.e., for each sequence $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0$, one has $H(\mathbf{z}^0,\mathbf{x}^0) = \lim_{\mathbf{x}^k \in \mathbf{W}(\mathbf{z}^k), \mathbf{x}^k \to \mathbf{x}^0} \inf H(\mathbf{z}^k,\mathbf{x}^k)$.
(6) If Axiom 3 holds, and if $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{x}^k \in \mathbf{W}(\mathbf{z}^k)$, $\mathbf{x}^k \to \mathbf{x}^0 \notin \mathbf{W}(\mathbf{z}^0)$, then $\lim_k H(\mathbf{z}^k,\mathbf{x}^k) = +\infty$.

Proof: By Lemma 16.2 in Appendix A.3, $\mathbf{W}(\mathbf{z})$ is a cone, and if $\mathbf{0}$ is in the relative interior of $\mathbf{T}(\mathbf{z})$, $\mathbf{W}(\mathbf{z})$ is a linear subspace. The convexity of $\mathbf{W}(\mathbf{z})$ is trivial. Hence (2), (3) hold. To show that $\mathbf{W}(\mathbf{z})$ is lower hemicontinuous, note that $\mathbf{x}^0 \in \mathbf{W}(\mathbf{z}^0)$ implies $\mathbf{x}^0/\lambda \in \mathbf{T}(\mathbf{z}^0)$ for some $\lambda > 0$. By Axiom 3, there exist $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$ for $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0$ such that $\mathbf{x}^k \to \mathbf{x}^0/\lambda$. Since $\mathbf{T}(\mathbf{z}^k) \subseteq \mathbf{W}(\mathbf{z}^k)$, this establishes lower hemicontinuity. Result (4) is an implication of Appendix A.3, Lemma 12.1. To show (5), consider a sequence of positive scalars λ_k satisfying $\lambda_k > H(\mathbf{z}^k,\mathbf{x}^k) > \lambda_k - k^{-1}$, where $(\mathbf{z}^k,\mathbf{x}^k)$ is a sequence with $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{x}^k \in \mathbf{W}(\mathbf{z}^k)$ converging to $\mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{x}^0 \in \mathbf{W}(\mathbf{z}^0)$. Then, $\mathbf{x}^k/\lambda_k \in \mathbf{T}(\mathbf{z}^k)$. By strong continuity, if $\lim_k \lambda_k = 0$, then $\mathbf{x}^0$ is in the asymptotic cone of $\mathbf{T}(\mathbf{z}^0)$, implying $H(\mathbf{z}^0,\mathbf{x}^0) = 0$ by the definition of H. If $\lim_k \lambda_k = +\infty$, then $H(\mathbf{z}^0,\mathbf{x}^0) \leqq \lim_k H(\mathbf{z}^k,\mathbf{x}^k)$ trivially. Finally, if $\lim_k \lambda_k = \lambda_0 > 0$, then $\mathbf{x}^0/\lambda_0 \in \mathbf{T}(\mathbf{z}^0)$ by strong continuity, implying $H(\mathbf{z}^0,\mathbf{x}^0) \leqq \lim_k H(\mathbf{z}^k,\mathbf{x}^k)$. We next show that there exist $\{\mathbf{x}^k\}$ for which equality is achieved in this limit; this will prove lower hemicontinuity. If $H(\mathbf{z}^0,\mathbf{x}^0) = \lambda > 0$, then $\mathbf{x}^0/\lambda \in \mathbf{T}(\mathbf{z}^0)$, and Axiom 3 implies the existence of $\mathbf{x}^k/\lambda \in \mathbf{T}(\mathbf{z}^k)$, $\mathbf{x}^k \to \mathbf{x}^0$. Then, $\lim_k H(\mathbf{z}^k,\mathbf{x}^k) \leqq H(\mathbf{z}^0,\mathbf{x}^0)$, implying with the previously established inequality that equality is achieved. If $H(\mathbf{z}^0,\mathbf{x}^0) = 0$, then $j\mathbf{x}^0 \in \mathbf{T}(\mathbf{z}^0)$ for each integer j. By Axiom 3, there exists $\mathbf{x}^{jk} \in \mathbf{T}(\mathbf{z}^k)$ with $\mathbf{x}^{jk} \to j\mathbf{x}^0$. Then, $\mathbf{x}^k = \mathbf{x}^{kk}/k$ has $\mathbf{x}^k \to \mathbf{x}^0$ and $H(\mathbf{z}^k,\mathbf{x}^k) \leqq k^{-1}$, and $H(\mathbf{z}^0,\mathbf{x}^0) = \lim_k H(\mathbf{z}^k,\mathbf{x}^k)$.

To show (6), note that $\mathbf{x}^k \in \mathbf{W}(\mathbf{z}^k)$ implies the existence of $\lambda_k > 0$ such

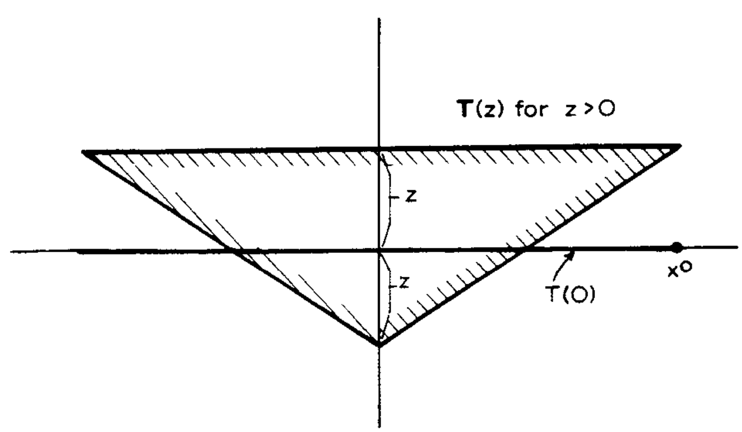
[14]The function $H(\mathbf{z},\mathbf{x})$ is *closed* in $\mathbf{x}$ (for fixed $\mathbf{z}$) if the set $\{(\mathbf{x},h) \in \mathbf{E}^N \times \mathbf{E} | h \geqq H(\mathbf{z},\mathbf{x})\}$ is closed.

that $\mathbf{x}^k/\lambda_k \in \mathbf{T}(\mathbf{z}^k)$ and $\lambda_k < H(\mathbf{z}^k,\mathbf{x}^k) + k^{-1}$. If λ_k is bounded, then there exists a subsequence (retain notation) converging to a scalar λ. If $\lambda > 0$, then $\mathbf{x}^0/\lambda \in \mathbf{T}(\mathbf{z}^0)$ by continuity of $\mathbf{T}$, implying $\mathbf{x}^0 \in \mathbf{W}(\mathbf{z}^0)$. If $\lambda = 0$, then $\mathbf{x}^0$ is contained in the asymptotic cone of $\mathbf{T}(\mathbf{z}^0)$ by the strong continuity of $\mathbf{T}$, implying $\mathbf{x}^0 \in \mathbf{W}(\mathbf{z}^0)$. Hence, $\mathbf{x}^0 \notin \mathbf{W}(\mathbf{z}^0)$ implies λ_k unbounded, and $\lim_k H(\mathbf{z}^k,\mathbf{x}^k) = +\infty$. Q.E.D.

Figure 20 shows that conclusion (5) cannot in general be strengthened to imply continuity. However, the following corollary can be established.

Corollary 21. Suppose a translated technology $\mathbf{T}(\mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, satisfies Axioms 1–4 and contains the origin in its interior. Then $\mathbf{W}(\mathbf{z}) = \mathbf{E}^N$ for all $\mathbf{z} \in \mathbf{Z}$, and $H(\mathbf{z},\mathbf{x})$ is continuous jointly in $\mathbf{z}$ and $\mathbf{x}$ at each $\mathbf{z} \in \mathbf{Z}$, $\mathbf{x} \in \mathbf{E}^N$.

Proof: Suppose $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, and $\mathbf{x}^k \to \mathbf{x}^0$. For any $\lambda > H(\mathbf{z}^0,\mathbf{x}^0)$, it follows from the continuity of H in $\mathbf{x}$ that $\lambda > H(\mathbf{z}^0,\mathbf{x}^k)$ for $k \geqq k_1$, with k_1 some large integer. The set $\mathbf{R} = \{\mathbf{x}^0/\lambda, \mathbf{x}^{k_1}/\lambda, \mathbf{x}^{k_1+1}/\lambda, \ldots\}$ is then closed and bounded, and is contained in the interior of $\mathbf{T}(\mathbf{z}^0)$. By Lemma 13.3(3) in Appendix A.3, there then exists $k_0 \geqq k_1$ such that $\mathbf{R}$ is contained in the interior of $\mathbf{T}(\mathbf{z}^k)$ for $k \geqq k_0$. But this implies $H(\mathbf{z}^k,\mathbf{x}^k) \leqq \lambda$, and hence, taking $\lambda \to H(\mathbf{z}^0,\mathbf{x}^0)$, $\lim_k H(\mathbf{z}^k,\mathbf{x}^k) \leqq H(\mathbf{z}^0,\mathbf{x}^0)$. Combined with lower semi-



$\mathbf{Z} = [0,1]$. Define $\mathbf{x}^1(z) = (1,z)$, $\mathbf{x}^2(z) = (-1,z)$, $\mathbf{x}^3(z) = (0,-z)$, and define $\mathbf{T}(z)$ to be the convex hull of these three points. Then, for $\mathbf{x}^0 = (1,0)$, one has $H(z,\mathbf{x}^0) = 2$ for $z > 0$ and $H(0,\mathbf{x}^0) = 1$.

FIGURE 20

continuity established in Lemma 20(5), this establishes the result. Q.E.D.

It should be noted that the gauge function has many of the same mathematical properties as the restricted profit function (compare Lemmas 11 and 15 versus 20). In particular, under the hypotheses of Corollary 21, the gauge function has all the properties that were demonstrated for the restricted profit function. The next result establishes that all the implications that can be drawn for the gauge function from Axioms 1–4 can also be drawn from the properties given by Lemma 20.

Lemma 22. Let $\mathbf{Z}$ be a non-empty subset of $\mathbf{E}^M$, $\mathbf{W}(\mathbf{z})$ be a linear subspace in $\mathbf{E}^N$ which is a lower hemicontinuous correspondence, and $H(\mathbf{z},\mathbf{x})$ be a non-negative convex, closed, positively linear homogeneous function of $\mathbf{x} \in \mathbf{W}(\mathbf{z})$ for each $\mathbf{z} \in \mathbf{Z}$. Suppose that $H(\mathbf{z},\mathbf{x})$ satisfies properties (5) and (6) of Lemma 20, and that the set of $\mathbf{x}$ satisfying $H(\mathbf{z},\mathbf{x}) = 0$ is semi-bounded for each $\mathbf{z} \in \mathbf{Z}$. Then, the correspondence $\mathbf{T}(\mathbf{z}) = \{\mathbf{x} \in \mathbf{W}(\mathbf{z}) | H(\mathbf{z},\mathbf{x}) \leqq 1\}$ satisfies Axioms 1–4.

Proof: It is immediate that $\mathbf{T}(\mathbf{z})$ satisfies Axioms 1, 2, and 4. To verify that Axiom 3 holds, consider $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$. Suppose $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$, $\mathbf{x}^k \to \mathbf{x}^0$. Then, $H(\mathbf{z}^k,\mathbf{x}^k) \leqq 1$ implies by property (6) in Lemma 20 that $\mathbf{x}^0 \in \mathbf{W}(\mathbf{z}^0)$, and implies by property (5) in Lemma 20 that $\mathbf{x}^0 \in \mathbf{T}(\mathbf{z}^0)$. Alternately, suppose $\mathbf{x}^0 \in \mathbf{T}(\mathbf{z}^0)$. By property (5) of Lemma 20, there exist $\mathbf{y}^k \in \mathbf{W}(\mathbf{z}^k)$ such that $\mathbf{y}^k \to \mathbf{x}^0$ and $\lim_k H(\mathbf{z}^k,\mathbf{y}^k) = H(\mathbf{z}^0,\mathbf{x}^0)$. Choose $\lambda_k > 0$ such that $\lambda_k - k^{-1} < H(\mathbf{z}^k,\mathbf{x}^k) < \lambda_k$. Then, $\lim \lambda_k \leqq 1$. Define $\mathbf{x}^k = \mathbf{y}^k/\text{Max}\,(\lambda_k,1)$. Then, $\lim_k \mathbf{y}^k = \lim_k \mathbf{x}^k = \mathbf{x}^0$ and $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$. Thus, $\mathbf{T}(\mathbf{z})$ is continuous. Lemma 13.3(2) in Appendix A.3 then implies that $\mathbf{T}(\mathbf{z})$ is strongly continuous. Q.E.D.

Using Lemmas 20 and 22, we can take the production possibility set and the gauge function as interchangeable descriptions of the technology when Axioms 1–4 hold. We conclude our analysis of the gauge function by noting its relation to the "gauge function" of the original technology $\mathbf{T}'(\mathbf{z})$. Recalling that $\mathbf{T}(\mathbf{z}) = \{\mathbf{x}' - \mathbf{x}^*(\mathbf{z}) \in \mathbf{E}^N | \mathbf{x}' \in \mathbf{T}'(\mathbf{z})\}$, define the *gauge function* relative to $\mathbf{x}^*(\mathbf{z})$ of the original technology $\mathbf{T}'(\mathbf{z})$ by

$$H'(\mathbf{z},\mathbf{x}') = H(\mathbf{z},\mathbf{x}' - \mathbf{x}^*(\mathbf{z})), \tag{45}$$

where $H(\mathbf{z},\mathbf{x})$ is the gauge function of the translated technology $\mathbf{T}(\mathbf{z})$. The reader can verify that H' is finite on an affine subspace of $\mathbf{x} \in \mathbf{E}^N$

for each $\mathbf{z} \in \mathbf{Z}$, that $\mathbf{T}'(\mathbf{z}) = \{\mathbf{x}' \in \mathbf{E}^N | H'(\mathbf{z},\mathbf{x}') \leqq 1\}$, and that H' has the same mathematical properties as H, with the exception of linear homogeneity.

17. Duality for the Restricted Profit Function

It has been established that a technology satisfying Axioms 1 and 2 yields a restricted profit function with the properties stated in Lemma 11, and that the technology can be completely recovered from a knowledge of the restricted profit function, provided in addition Axiom 4 holds, by use of the mapping (37). We will now show, conversely, that a function with the properties of a restricted profit function will yield, via the mapping (37), a technology satisfying Axioms 1, 2, and 4, and that this technology returns the original function via the mapping (35). From these results, we will have obtained a generalization of the Shephard–Uzawa duality theorem to the case of the restricted profit function. This generalization will allow the use of the duality principle in applications for a broad range of environments of the competitive firm.

Lemma 23. Suppose $\mathbf{Z}$ is a non-empty subset of $\mathbf{E}^M$, and that for each $\mathbf{z} \in \mathbf{Z}$, $\mathbf{P}(\mathbf{z})$ is a convex cone with a non-empty interior, and $\Pi(\mathbf{z},\mathbf{p})$ is a convex, positively linear homogeneous, closed function of $\mathbf{p} \in \mathbf{P}(\mathbf{z})$. Then, $\mathbf{T}(\mathbf{z}) = \{\mathbf{x} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{x} \leqq \Pi(\mathbf{z},\mathbf{p}) \text{ for all } \mathbf{p} \in \mathbf{P}(\mathbf{z})\}$ satisfies Axioms 1, 2, and 4, and $\Pi(\mathbf{z},\mathbf{p})$ is the restricted profit function of $\mathbf{T}(\mathbf{z})$ as defined by equation (35).

Proof: Appendix A.3, Lemma 12.5.

Note that the basic duality theorem expressed by Lemmas 11 and 23 is an immediate consequence of the basic mathematical theory of convex conjugate functions. As in the case of cost functions, the restricted profit function will fail to distinguish between distinct technologies which have the same convex hull, reflecting the fact that two competitive firms with these respective technologies may exhibit identical behavior.

Utilizing the concept of a gauge function introduced in the previous section, we can summarize these duality relations in the following formal duality theorem.

Theorem 24. Suppose $\mathbf{Z}$ is a non-empty subset of $\mathbf{E}^M$ and $\mathbf{x}^*(\mathbf{z})$ is a function from $\mathbf{Z}$ into $\mathbf{E}^N$.

(a) Let $\mathscr{T}$ denote the class of sets $\mathbf{T}(\mathbf{z})$, $\mathbf{z} \in \mathbf{Z}$, which satisfy Axioms 1, 2, and 4 and have $\mathbf{x}^*(\mathbf{z})$ in the relative interior of $\mathbf{T}(\mathbf{z})$; i.e., for each $\mathbf{x} \in \mathbf{T}(\mathbf{z})$, it follows that $\theta\mathbf{x} + (1-\theta)\mathbf{x}^*(\mathbf{z}) \in \mathbf{T}(\mathbf{z})$ for small positive or negative scalars θ.

(b) Let $\mathscr{H}$ denote the class of pairs $\langle H,\mathbf{X}\rangle$, where for each $\mathbf{z} \in \mathbf{Z}$, the set $\{\mathbf{x} - \mathbf{x}^*(\mathbf{z}) | \mathbf{x} \in \mathbf{X}(\mathbf{z})\}$ is a linear subspace of $\mathbf{E}^N$ and $H(\mathbf{z},\mathbf{x})$ is a non-negative convex, closed function of $\mathbf{x} \in \mathbf{X}(\mathbf{z})$ with $H(\mathbf{z},\mathbf{x}^*(\mathbf{z}) + \theta(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))) = \theta H(\mathbf{z},\mathbf{x})$ for $\mathbf{x} \in \mathbf{X}(\mathbf{z})$, $\theta > 0$, and with the set of $\mathbf{x} \in \mathbf{X}(\mathbf{z})$ for which $H(\mathbf{z},\mathbf{x}) = 0$ a semi-bounded set.

(c) Let $\mathscr{P}$ denote the class of pairs $\langle \Pi,\mathbf{P}\rangle$, where for each $\mathbf{z} \in \mathbf{Z}$, the set $\mathbf{P}(\mathbf{z})$ is a convex cone in $\mathbf{E}^N$ with a non-empty interior and $\Pi(\mathbf{z},\mathbf{p})$ is a convex, closed, positively linear homogeneous function of $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ such that if $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ and either $-\mathbf{p} \notin \mathbf{P}(\mathbf{z})$, or $-\mathbf{p} \in \mathbf{P}(\mathbf{z})$ and $\Pi(\mathbf{z},\mathbf{p}) \neq -\Pi(\mathbf{z},-\mathbf{p})$, then $\Pi(\mathbf{z},\mathbf{p}) > \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})$.

(Note that $\mathscr{T}$ is a class of production possibility sets, $\mathscr{H}$ is a class of non-translated gauge functions, and $\mathscr{P}$ is a class of restricted profit functions.)

(d) Define a *profit mapping* from $\mathbf{T} \in \mathscr{T}$ to pairs $\langle \Pi,\mathbf{P}\rangle$ satisfying $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ if and only if $\sup\{\mathbf{p}\cdot\mathbf{x} | \mathbf{x} \in \mathbf{T}(\mathbf{z})\} < +\infty$, and $\Pi(\mathbf{z},\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{x} | \mathbf{x} \in \mathbf{T}(\mathbf{z})\}$ for $\mathbf{p} \in \mathbf{P}(\mathbf{z})$. Then, the image of the profit mapping is a unique element in $\mathscr{P}$.

(e) Define an *implicit technology mapping* from $\langle \Pi,\mathbf{P}\rangle \in \mathscr{P}$ to sets $\mathbf{T}(\mathbf{z}) \subseteq \mathbf{E}^N$ satisfying $\mathbf{T}(\mathbf{z}) = \{\mathbf{x} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{x} \leqq \Pi(\mathbf{z},\mathbf{p}) \text{ for all } \mathbf{p} \in \mathbf{P}(\mathbf{z})\}$. Then, the image of the implicit technology mapping is a unique element in $\mathscr{T}$. Further, the profit and implicit technology mappings are mutual inverses.

(f) Define a *gauge mapping* from $\mathbf{T} \in \mathscr{T}$ to pairs $\langle H,\mathbf{X}\rangle$ satisfying $\mathbf{x} \in \mathbf{X}(\mathbf{z})$ if and only if $\theta\mathbf{x} + (1-\theta)\mathbf{x}^*(\mathbf{z}) \in \mathbf{T}(\mathbf{z})$ for small positive θ, and $H(\mathbf{z},\mathbf{x}) = \inf\{\lambda > 0 | \mathbf{x}^*(\mathbf{z}) + (\mathbf{x} - \mathbf{x}^*(\mathbf{z}))/\lambda \in \mathbf{T}(\mathbf{z})\}$ for $\mathbf{x} \in \mathbf{X}(\mathbf{z})$. Then, the image of the gauge mapping is a unique element in $\mathscr{H}$.

(g) Define an *inverse gauge mapping* from $\langle H,\mathbf{X}\rangle \in \mathscr{H}$ to sets $\mathbf{T}(\mathbf{z}) \subseteq \mathbf{E}^N$ satisfying $\mathbf{T}(\mathbf{z}) = \{\mathbf{x} \in \mathbf{X}(\mathbf{z}) | H(\mathbf{x},\mathbf{z}) \leqq 1\}$. Then, the image of the inverse gauge mapping is a unique element in $\mathscr{T}$. Further, the gauge and inverse gauge mappings are mutual inverses.

(h) Define an *implicit profit mapping* from $\langle H,\mathbf{X}\rangle \in \mathscr{H}$ to pairs $\langle \Pi,\mathbf{P}\rangle$ satisfying $\Pi(\mathbf{z},\mathbf{p}) = \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \inf\{\lambda > 0 | \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq \lambda H(\mathbf{z},\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbf{X}(\mathbf{z})\}$ for $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, with $\mathbf{P}(\mathbf{z})$ defined as the set of $\mathbf{p} \in \mathbf{E}^N$

for which the set of λ in the right-hand side of this expression is non-empty. Then, the implicit profit mapping is the composition of the inverse gauge mapping and the profit mapping, and its image is a unique element in $\mathcal{P}$.

(i) Define an *implicit gauge mapping* from $\langle \Pi,\mathbf{P}\rangle \in \mathcal{P}$ to pairs $\langle H,\mathbf{X}\rangle$ satisfying $H(\mathbf{z},\mathbf{x}) = \inf\{\lambda > 0 | \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq \lambda(\Pi(\mathbf{z},\mathbf{p}) - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}))$ for all $\mathbf{p} \in \mathbf{P}(\mathbf{z})\}$ for $\mathbf{x} \in \mathbf{X}(\mathbf{z})$, with $\mathbf{X}(\mathbf{z})$ defined as the set of $\mathbf{x} \in \mathbf{R}^N$ for which the set of λ in the right-hand side of this expression is non-empty. Then, the implicit gauge mapping is the composition of the implicit technology mapping and the gauge mapping, and its image is a unique element in $\mathcal{H}$. Further, the implicit profit and implicit gauge mappings are mutual inverses.

Corollary. Suppose in Theorem 24, condition (a) is modified to require only that $\mathbf{x}^*(\mathbf{z})$ be contained in $\mathbf{T}(\mathbf{z})$ rather than in its relative interior, condition (b) is modified to require that the set $\{\mathbf{x} - \mathbf{x}^*(\mathbf{z}) | \mathbf{x} \in \mathbf{X}(\mathbf{z})\}$ be a convex cone rather than a linear subspace, and condition (c) is modified to weaken the implication $\Pi(\mathbf{z},\mathbf{p}) > \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})$ to $\Pi(\mathbf{z},\mathbf{p}) \geqq \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})$. Then the conclusions of the theorem continue to hold.

Proof: Consider statement (d). Lemma 11 establishes $\langle \Pi,\mathbf{P}\rangle$ to be a unique element in $\mathcal{P}$, provided that Π satisfies the last property in (c). But (a) implies that if the conclusion of this property is false, so that $\Pi(\mathbf{z},\mathbf{p}) = \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})$ for some $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, then $\mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) = \mathbf{p}\cdot\mathbf{x}$ for all $\mathbf{x} \in \mathbf{T}(\mathbf{z})$. Then, $\Pi(\mathbf{z},\mathbf{p}) = \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) = -\Pi(\mathbf{z},-\mathbf{p})$, and the hypothesis for the last property in (c) is also false.

Consider statement (e). Lemma 23 and Lemma 11(4) imply all the results, provided we show that $\mathbf{x}^*(\mathbf{z})$ is in the relative interior of $\mathbf{T}(\mathbf{z})$. If this conclusion failed to hold, then by Result 10.12 in Appendix A.3, there would exist $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ with $\mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) \geqq \Pi(\mathbf{z},\mathbf{p}) > \mathbf{p}\cdot\mathbf{x}$ for some $\mathbf{x} \in \mathbf{T}(\mathbf{z})$, and the last condition defining the class $\mathcal{P}$ would be contradicted. Hence, $\mathbf{x}^*(\mathbf{z})$ is in the relative interior of $\mathbf{T}(\mathbf{z})$.

Consider statements (f) and (g). Except for two propositions, Lemmas 20 and 22 imply the results directly. First, we need to show that the composition of the mappings from $\mathcal{H}$ to $\mathcal{T}$ and from $\mathcal{T}$ to $\mathcal{H}$ is the identity mapping. If not, there exist distinct gauge functions H^1 and H^2 such that H^1 maps into T^1 in $\mathcal{T}$ and T^1 maps into H^2. But Lemma 20(3) implies that H^2 maps into T^1. Hence $H^1(\mathbf{z},\mathbf{x}) = 1$ implies $H^2(\mathbf{z},\mathbf{x}) = 1$, and positive linear homogeneity in $\mathbf{x}$ implies $H^1 = H^2$. Second, we need to show that $\mathbf{x}^*(\mathbf{z})$ is in the image of the inverse gauge mapping, and in its relative interior when $\mathbf{X}(\mathbf{z})$ is an affine linear subspace. The definition of H implies $H(\mathbf{z},\mathbf{x}^*(\mathbf{z})) = 0$, implying $\mathbf{x}^*(\mathbf{z})$ is in $\mathbf{T}(\mathbf{z}) = \{\mathbf{x} \in \mathbf{X}(\mathbf{z}) | H(\mathbf{z},\mathbf{x}) \leqq 1\}$. Since H is convex and closed, it is continuous on $\mathbf{X}(\mathbf{z})$ when $\mathbf{X}(\mathbf{z})$ is an affine linear subspace, implying $\mathbf{x}^*(\mathbf{z})$ is in the relative interior of $\mathbf{T}(\mathbf{z})$.

Finally, consider statements (h) and (i). In (h), $\mathbf{x} \in \mathbf{X}(\mathbf{z})$ implies either $H(\mathbf{z},\mathbf{x}) = 0$ and $\mathbf{p}\cdot\mathbf{x} \leqq \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})$ for all $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, or $H(\mathbf{z},\mathbf{x}) > 0$ and $\mathbf{x}^*(\mathbf{z}) + (\mathbf{x} - \mathbf{x}^*(\mathbf{z}))/H(\mathbf{z},\mathbf{x}) \in \mathbf{T}(\mathbf{z})$, with $\mathbf{T}$ defined by the inverse gauge mapping. Then,

$$\begin{aligned}\Pi(\mathbf{z},\mathbf{p}) &= \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \sup\{\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))/H(\mathbf{z},\mathbf{x}) | \mathbf{x} \in \mathbf{X}(\mathbf{z}),\ H(\mathbf{z},\mathbf{x}) > 0\} \\ &= \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \inf\{\lambda > 0 | \lambda > \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))/H(\mathbf{z},\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbf{X}(\mathbf{z}) \\ &\qquad \text{such that } H(\mathbf{z},\mathbf{x}) > 0\} \\ &= \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \inf\{\lambda > 0 | \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq \lambda H(\mathbf{z},\mathbf{x}) \text{for all } \mathbf{x} \in \mathbf{X}(\mathbf{z})\}.\end{aligned}$$

Hence, the implicit profit mapping is a composition of the inverse gauge and profit mappings. In (i),

$$\begin{aligned}H(\mathbf{z},\mathbf{x}) &= \inf\{\lambda > 0 | \mathbf{x}^*(\mathbf{z}) + (\mathbf{x} - \mathbf{x}^*(\mathbf{z}))/\lambda \in \mathbf{T}(\mathbf{z})\} \\ &= \inf\{\lambda > 0 | \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))/\lambda \leqq \Pi(\mathbf{z},\mathbf{p}) \text{for all } \mathbf{p} \in \mathbf{P}(\mathbf{z})\} \\ &= \inf\{\lambda > 0 | \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq \lambda(\Pi(\mathbf{z},\mathbf{p}) - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})) \text{for all } \mathbf{p} \in \mathbf{P}(\mathbf{z})\}.\end{aligned}$$

Hence, the implicit gauge mapping is a composition of the implicit technology mapping and the gauge mapping. Q.E.D.

Note that the implicit profit mapping and the implicit gauge mapping have similar formal structures, making the "duality" essentially complete. If one takes $\mathbf{x}^*(\mathbf{z}) = \mathbf{0}$ and considers a subclass of $\mathscr{H}$ for which $\mathbf{X}(\mathbf{z}) = \mathbf{E}^N$ and a subclass of $\mathscr{P}$ for which $\mathbf{P}(\mathbf{z})$ is semi-bounded, then the duality is complete in the sense that the members of H and Π have symmetric properties, and the implicit profit and gauge mappings are identical, except for the change of variables. For dual profit and gauge functions $\Pi(\mathbf{z},\mathbf{p})$ and $H(\mathbf{z},\mathbf{x})$, the inequality

$$\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq H(\mathbf{z},\mathbf{x})(\Pi(\mathbf{z},\mathbf{p}) - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})) \tag{46}$$

holds for all $\mathbf{p}$ and $\mathbf{x}$ in $\mathbf{E}^N$, with equality when $\mathbf{x} \in \mathbf{\Phi}(\mathbf{z},\mathbf{p})$.

For the classes of gauge and profit functions defined in Theorem 24, we can establish further structural relations of the form "the gauge function has property 'P' if and only if the restricted profit function has property 'Q'," in the way that Lemma 7 established structural relationships for cost and distance functions. Hereafter, we shall assume the translation of production possibility sets described in Section 16 has been carried out, so that $\mathbf{x}^*(\mathbf{z}) \equiv \mathbf{0}$ in the hypothesis of Theorem 24, the origin is contained in the relative interior of sets $\mathbf{T} \in \mathscr{T}$, $\mathbf{X}(\mathbf{z})$ is a linear subspace of $\mathbf{E}^N$ for each $\langle H,\mathbf{X}\rangle \in \mathscr{H}$, and each $\langle \Pi,\mathbf{P}\rangle \in \mathscr{P}$ satisfies $\Pi(\mathbf{z},\mathbf{p}) > 0$ for all $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ such that either $\Pi(\mathbf{z},\mathbf{p}) \neq -\Pi(\mathbf{z},-\mathbf{p})$ or $-\mathbf{p} \notin \mathbf{P}(\mathbf{z})$. Let $\mathscr{T}^0$, $\mathscr{H}^0$, and $\mathscr{P}^0$ denote the classes $\mathscr{T}$, $\mathscr{H}$, $\mathscr{P}$, respectively, for this case of $\mathbf{x}^*(\mathbf{z}) \equiv \mathbf{0}$. The qualitative structural relationships derived under this assumption continue to hold in the case of a general non-zero $\mathbf{x}^*(\mathbf{z})$, with an appropriate modification of the definition of the sub-differential for the gauge function and of an exposed value of $\mathbf{x}$ for this function (defined below). This last stage of generalization, which considerably complicates notation without adding new results, will be left to the reader.

Recalling that $\mathbf{z} \in \mathbf{Z}$ is a vector in $\mathbf{E}^M$, let $\mathbf{z} = (\mathbf{z}_{(1)},\mathbf{z}_{(2)})$ denote a partition of $\mathbf{z}$ into sub-vectors. The gauge function $H(\mathbf{z},\mathbf{x})$ is defined to be *non-increasing* in $\mathbf{z}_{(1)}$ if for any $\mathbf{z}',\mathbf{z}'' \in \mathbf{Z}$ with $\mathbf{z}'_{(1)} \geqq \mathbf{z}''_{(1)}$ and $\mathbf{z}'_{(2)} = \mathbf{z}''_{(2)}$, it follows that $H(\mathbf{z}',\mathbf{x}) \leqq H(\mathbf{z}'',\mathbf{x})$, where $H(\mathbf{z},\mathbf{x})$ is allowed to assume the value $+\infty$ if $\mathbf{x} \notin \mathbf{X}(\mathbf{z})$. The gauge function is defined to be *uniformly decreasing* in $\mathbf{z}_{(1)}$ if for any distinct $\mathbf{z}'$, $\mathbf{z}'' \in \mathbf{Z}$ with $\mathbf{z}'_{(1)} \geqq \mathbf{z}''_{(1)}$ and $\mathbf{z}'_{(2)} = \mathbf{z}''_{(2)}$, it follows that there exists a positive scalar λ such that $(1+\lambda)H(\mathbf{z}',\mathbf{x}) \leqq H(\mathbf{z}'',\mathbf{x})$, where again the value $+\infty$ is allowed for $H(\mathbf{z},\mathbf{x})$ when $\mathbf{x}$ fails to lie in $\mathbf{X}(\mathbf{z})$. Clearly, H is non-increasing in $\mathbf{z}_{(1)}$ if and only if the production possibility set $\mathbf{T}(\mathbf{z})$ satisfies $\mathbf{T}(\mathbf{z}'') \subseteq \mathbf{T}(\mathbf{z}')$. This could be expected to be the case, for example, if the components of $\mathbf{z}_{(1)}$ are fixed inputs to the production process (measured with a positive sign) or are indices of the level of technical progress. Analogous definitions can be made for the gauge function non-decreasing or uniformly increasing in a sub-vector $\mathbf{z}_{(1)}$, or for the restricted profit function weakly or uniformly monotone in $\mathbf{z}_{(1)}$.

Theorem 25. Suppose $\mathbf{Z}$ is a non-empty subset of $\mathbf{E}^M$, and consider the class of gauge functions $\mathscr{H}^0$ and the class of restricted profit functions $\mathscr{P}^0$. Then, for gauge and restricted profit functions in these classes which are dual under the implicit profit and implicit gauge mappings of Theorem 24, the structural relationships given in Table 3 hold; i.e., the gauge function has the property "P" if and only if the restricted profit function has the property "Q".

Proof: *Result* 1 – If $\mathbf{X}(\mathbf{z}) = \mathbf{E}^N$, then for any $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, $\mathbf{p} \neq \mathbf{0}$, one has $H(\mathbf{z},\mathbf{p}) < +\infty$, implying $H(\mathbf{z},\lambda\mathbf{p}) \leqq 1$ for some $\lambda > 0$, and hence $\Pi(\mathbf{z},\mathbf{p}) \geqq$

TABLE 3
Property "P" holds for a gauge function $\langle H,\mathbf{X}\rangle \in \mathscr{H}^0$ if and only if property "Q" holds for its restricted profit function $\langle \Pi,\mathbf{P}\rangle \in \mathscr{P}^0$.[a]

	"P" on the gauge function $H(\mathbf{z},\mathbf{x})$, $\mathbf{x} \in \mathbf{X}(\mathbf{z})$	"Q" on the restricted profit function $\Pi(\mathbf{z},\mathbf{p})$, $\mathbf{p} \in \mathbf{P}(\mathbf{z})$
1.[b]	$\mathbf{X}(\mathbf{z}) = \mathbf{E}^N$	$\Pi(\mathbf{z},\mathbf{p}) > 0$ for all $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, $\mathbf{p} \neq 0$
2.[c]	$H(\mathbf{z},\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbf{X}(\mathbf{z}), \mathbf{x} \neq \mathbf{0}$	$\mathbf{P}(\mathbf{z}) = \mathbf{E}^N$
3.	Non-increasing (non-decreasing) in a subset of variables $\mathbf{z}_{(1)}$ of $\mathbf{z}$	Non-decreasing (non-increasing) in a subset of variables $\mathbf{z}_{(1)}$ of $\mathbf{z}$
4.	Uniformly decreasing (uniformly increasing) in a subset of variables $\mathbf{z}_{(1)}$ of $\mathbf{z}$	Uniformly increasing (uniformly decreasing) in a subset of variables $\mathbf{z}_{(1)}$ of $\mathbf{z}$
5.	$\langle H,\mathbf{X}\rangle$ has the joint continuity property in $(\mathbf{z},\mathbf{x})$ defined by statements (1), (4), (5), and (6) in Lemma 20	$\langle \Pi,\mathbf{P}\rangle$ has the joint continuity property in $(\mathbf{z},\mathbf{p})$ defined by statement (1) in Lemma 15 and the conclusions of Lemma 16

[a]By formal duality, the implications of this table continue to hold when properties "P" and "Q" are reversed.

[b]This condition is equivalent to requiring that the origin be an interior point of the production possibility set $\mathbf{T} \in \mathscr{T}^0$.

[c]This condition is equivalent to requiring that the production possibility set $\mathbf{T} \in \mathscr{T}^0$ be bounded.

$\lambda\mathbf{p}\cdot\mathbf{p} > 0$. Alternately, if $\mathbf{X}(\mathbf{z}) \neq \mathbf{E}^N$, then taking $\mathbf{p} \neq \mathbf{0}$ in the subspace orthogonal to $\mathbf{X}(\mathbf{z})$ implies $\mathbf{p} \in \mathbf{P}(\mathbf{z})$ and $\Pi(\mathbf{z},\mathbf{p}) = 0$.

Result 2 – This result follows from Result 1 by formal duality.

Result 3 – Consider $\mathbf{z}'$, $\mathbf{z}'' \in \mathbf{Z}$ with $\mathbf{z}'_{(1)} \geqq \mathbf{z}''_{(1)}$ and $\mathbf{z}'_{(2)} = \mathbf{z}''_{(2)}$. Since either H non-increasing or Π non-decreasing are equivalent to $\mathbf{T}(\mathbf{z}'') \subseteq \mathbf{T}(\mathbf{z}')$ under the inverse gauge or implicit technology mappings, respectively, the result follows.

Result 4 – Consider distinct $\mathbf{z}'$, $\mathbf{z}'' \in \mathbf{Z}$ with $\mathbf{z}'_{(1)} \geqq \mathbf{z}''_{(1)}$ and $\mathbf{z}'_{(2)} = \mathbf{z}''_{(2)}$. As in the previous result, each condition is equivalent to $(1+\lambda)\mathbf{T}(\mathbf{z}'') \subseteq \mathbf{T}(\mathbf{z}')$ for some $\lambda > 0$, and the result follows.

Result 5 – We first establish that property "Q" on restricted profit functions in $\mathscr{P}^0$ implies that dual $\mathbf{T} \in \mathscr{T}^0$ satisfy Axiom 3; the remaining properties of $\mathbf{T}$, Axioms 1, 2, and 4, are immediate from Lemmas 15 and 16 and Theorem 24. Suppose $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$, $\mathbf{x}^k \to \mathbf{x}^0$. Consider any $\mathbf{p}^0 \in \mathbf{P}^0(\mathbf{z}^0)$. The lower hemicontinuity of $\mathbf{P}(\mathbf{z})$ implies there exist $\mathbf{p}^k \in \mathbf{P}(\mathbf{z}^k)$ such that $\mathbf{p}^k \to \mathbf{p}^0$. By definition of $\mathbf{T}(\mathbf{z})$, $\mathbf{p}^k\cdot\mathbf{x}^k \leqq \Pi(\mathbf{z}^k,\mathbf{p}^k)$. Then, $\lim_k \mathbf{p}^k\cdot\mathbf{x}^k = \mathbf{p}^0\cdot\mathbf{x}^0 \leqq \lim_k \Pi(\mathbf{z}^k,\mathbf{p}^k)$ by the continuity of Π. The condition $\mathbf{p}\cdot\mathbf{x}^0 \leqq \Pi(\mathbf{z}^0,\mathbf{p})$ for $\mathbf{p} \in \mathbf{P}^0(\mathbf{z}^0)$ and the lower semicontinuity of Π on $\mathbf{P}(\mathbf{z})$ implies $\mathbf{p}\cdot\mathbf{x}^0 \leqq \Pi(\mathbf{z}^0,\mathbf{p})$ for all $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, and hence $\mathbf{x}^0 \in \mathbf{T}(\mathbf{z}^0)$.

Next suppose $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, $\mathbf{x}^0 \in \mathbf{T}(\mathbf{z}^0)$. Suppose there exists $\epsilon > 0$ such that $(1-\epsilon)\mathbf{x}^0 \not\in \mathbf{T}(\mathbf{z}^k)$ for k large. Then there exists $\mathbf{p}^k \in \mathbf{P}(\mathbf{z}^k)$, $|\mathbf{p}^k| = 1$ such that $(1-\epsilon)\mathbf{p}^k \cdot \mathbf{x}^0 > \Pi(\mathbf{z}^k,\mathbf{p}^k)$. Then $\mathbf{p}^k$ has a subsequence (retain notation) converging to $\mathbf{p}^0$. Then $\Pi(\mathbf{z}^k,\mathbf{p}^k) < (1-\epsilon)\mathbf{p}^k \cdot \mathbf{x}^0 \to (1-\epsilon)\mathbf{p}^0 \cdot \mathbf{x}^0$ implies $\mathbf{p}^0 \in \mathbf{P}(\mathbf{z}^0)$ by the last continuity property of $\Pi(\mathbf{z},\mathbf{p})$ given in the statement of Lemma 16. Hence, $\mathbf{p}^0 \cdot \mathbf{x}^0 \leqq \Pi(\mathbf{z}^0,\mathbf{p}^0)$, implying $\Pi(\mathbf{z}^k,\mathbf{p}^k) < (1-\epsilon)\Pi(\mathbf{z}^0,\mathbf{p}^0)$. But this contradicts the lower semicontinuity property of Π. Hence there exist $\mathbf{x}^k \in \mathbf{T}(\mathbf{z}^k)$, $\mathbf{x}^k \to \mathbf{x}^0$. Therefore, $\mathbf{T}$ is continuous and convex-valued, and hence (Appendix A.3, Lemma 13.3) strongly continuous. Hence, $\mathbf{T}$ satisfies Axiom 3. Similarly, Lemmas 20 and 22 along with Theorem 24 imply that property "P" on gauge functions in $\mathscr{H}^0$ is equivalent to $\mathbf{T} \in \mathscr{T}^0$ satisfying Axioms 1–4. Hence, the result follows. Q.E.D.

The next result relates differentiability properties of the gauge function to curvature properties of the restricted profit function, and vice versa. We shall now add to the conditions determining the class of gauge functions $\mathscr{H}^0$ the assumption that $\mathbf{X}(\mathbf{z}) = \mathbf{E}^N$ for each $\mathbf{z} \in \mathbf{Z}$. From Result 1 in Table 3, this is equivalent to assuming that the origin of corresponding production possibility sets is an interior point of these sets, or that the restricted profit function is positive for all non-zero $\mathbf{p} \in \mathbf{P}(\mathbf{z})$. Let $\mathscr{T}^1$, $\mathscr{H}^1$, $\mathscr{P}^1$ denote the subclasses of $\mathscr{T}^0$, $\mathscr{H}^0$, $\mathscr{P}^0$, respectively, on which this added restriction is satisfied.

By Lemma 19, the sub-differential $\mathbf{\Gamma}(\mathbf{z},\mathbf{p})$ of $\Pi(\mathbf{z},\mathbf{p})$ with $\langle \Pi,\mathbf{P}\rangle \in \mathscr{P}^1$ exists for $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$, and satisfies $\mathbf{\Gamma}(\mathbf{z},\mathbf{p}) = \mathbf{\Phi}(\mathbf{z},\mathbf{p})$, where $\mathbf{\Phi}(\mathbf{z},\mathbf{p})$ is the set of maximands of $\mathbf{p}\cdot\mathbf{x}$ for $\mathbf{x} \in \mathbf{T}(\mathbf{z})$. Further, $\mathbf{x} \in \mathbf{\Phi}(\mathbf{z},\mathbf{p})$ implies $H(\mathbf{z},\mathbf{x}) = 1$. Define $\mathbf{X}^*(\mathbf{z}) = \{\mathbf{x} \in \mathbf{E}^N | \mathbf{x} \in \mathbf{\Phi}(\mathbf{z},\mathbf{p}) \text{ for some } \mathbf{p} \in \mathbf{P}^0(\mathbf{z})\}$.

The gauge function $H(\mathbf{z},\mathbf{x})$, $\langle H,\mathbf{E}^N\rangle \in \mathscr{H}^1$, is *exposed* at $\mathbf{x} \in \mathbf{E}^N$ if $H(\mathbf{z},\mathbf{x}') - H(\mathbf{z},\mathbf{x}) > \mathbf{p}\cdot(\mathbf{x}' - \mathbf{x})$ for all $\mathbf{x}' \in \mathbf{E}^N$, $\mathbf{x}'$ not proportional to $\mathbf{x}$, and all $\mathbf{p}$ in the relative interior of $\mathbf{\Lambda}(\mathbf{z},\mathbf{x})$, where $\mathbf{\Lambda}(\mathbf{z},\mathbf{x})$ is the sub-differential of $H(\mathbf{z},\mathbf{x})$. (a) in Figure 21 illustrates a case in which H is not exposed at $\mathbf{x}^0$, since the point $\mathbf{x}^1$ has $H(\mathbf{z},\mathbf{x}^1) = H(\mathbf{z},\mathbf{x}^0)$, $\mathbf{p}^0\cdot\mathbf{x}^1 = \mathbf{p}^0\cdot\mathbf{x}^0$, where $\mathbf{p}^0$ is the price vector satisfying $\{\mathbf{p}^0\} = \mathbf{\Lambda}(\mathbf{z},\mathbf{x}^0)$. In (b), (c), (d), the point $\mathbf{x}^0$ is exposed. Note that in case (c), the set $\mathbf{\Lambda}(\mathbf{z},\mathbf{x}^0)$ is the closed line segment connecting $\mathbf{p}^1$ and $\mathbf{p}^2$, and the relative interior of $\mathbf{\Lambda}(\mathbf{z},\mathbf{x}^0)$ is the open line segment connecting, but excluding, $\mathbf{p}^1$ and $\mathbf{p}^2$. For any $\mathbf{p}^0$ in this open line segment, note that the strict inequality $H(\mathbf{z},\mathbf{x}') - H(\mathbf{z},\mathbf{x}^0) > \mathbf{p}^0\cdot(\mathbf{x}' - \mathbf{x}^0)$ holds for $\mathbf{x}'$ not proportional to $\mathbf{x}^0$, but that for the end points $\mathbf{p}^1$ and $\mathbf{p}^2$, points such as $\mathbf{x}^1$ and $\mathbf{x}^2$ will result in equality holding. In case (d), the strict inequality $H(\mathbf{z},\mathbf{x}') - H(\mathbf{z},\mathbf{x}^0) > \mathbf{p}^0\cdot(\mathbf{x}' - \mathbf{x}^0)$ holds for $\mathbf{x}'$ not propor-

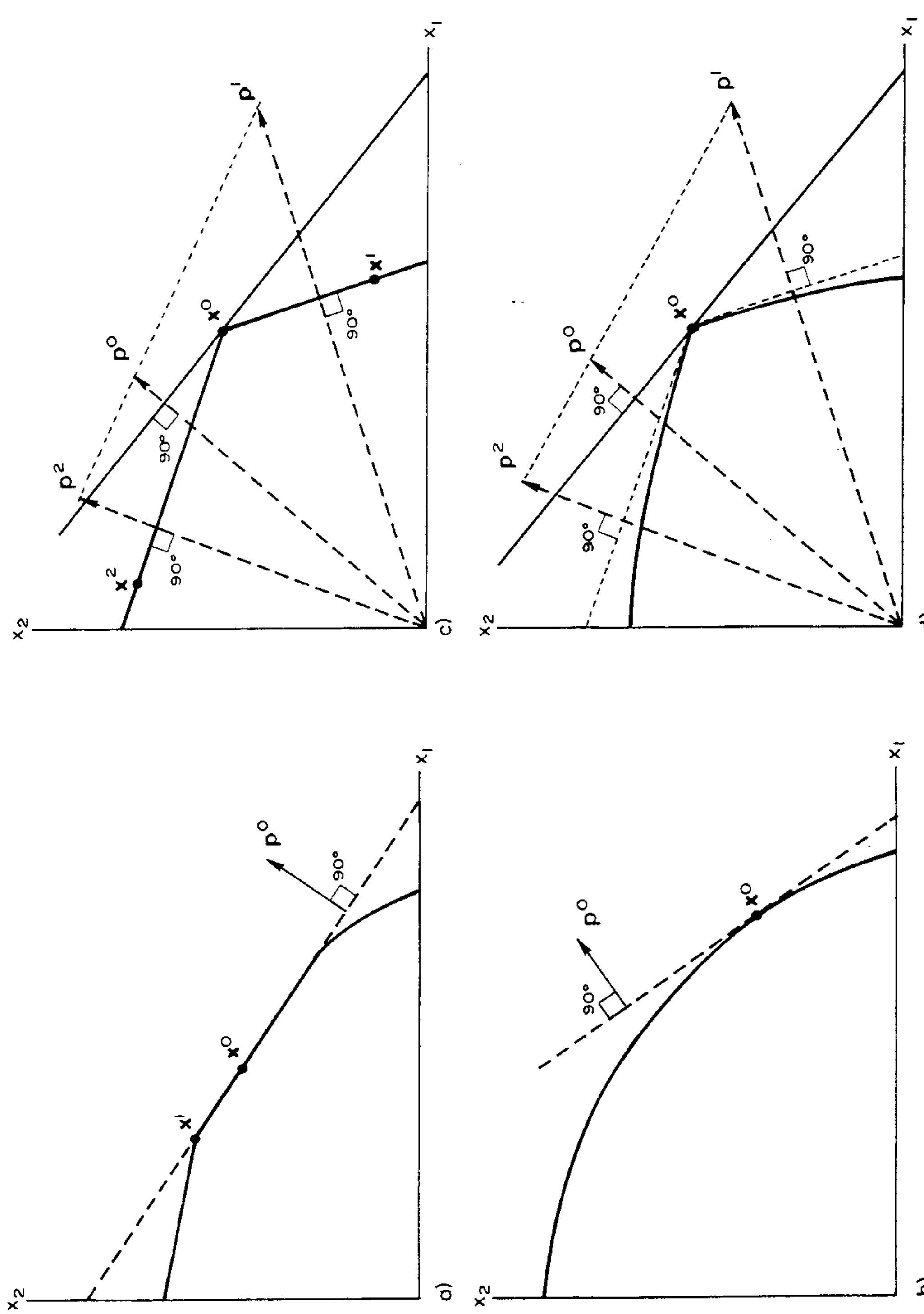


FIGURE 21

tional to $\mathbf{x}$ and all points $\mathbf{p}^0$ in $\Lambda(\mathbf{z},\mathbf{x}^0)$, including the end points $\mathbf{p}^1$ and $\mathbf{p}^2$ which are not in the relative interior of $\Lambda(\mathbf{z},\mathbf{x}^0)$.

The gauge function $H(\mathbf{z},\mathbf{x})$, $\langle H,\mathbf{E}^N\rangle \in \mathscr{H}^1$, is *strictly quasiconvex at* $\mathbf{x}\in\mathbf{E}^N$ if $H(\mathbf{z},\theta\mathbf{x}+(1-\theta)\mathbf{x}') < \theta H(\mathbf{z},\mathbf{x})+(1-\theta)H(\mathbf{z},\mathbf{x}')$ for all $\mathbf{x}'$ not proportional to $\mathbf{x}$, $0<\theta<1$.[15] (a) and (c) of Figure 21 illustrate cases in which the gauge function fails to be strictly quasi-convex at $\mathbf{x}^0$, since in each case an average of $\mathbf{x}^0$ and $\mathbf{x}^1$ has $H(\mathbf{z},\theta\mathbf{x}^0+(1-\theta)\mathbf{x}^1)=\theta H(\mathbf{z},\mathbf{x}^0)+(1-\theta)H(\mathbf{z},\mathbf{x}^1)$. Cases (b) and (d) in this figure have H strictly quasi-convex at $\mathbf{x}^0$. Lemma 16.5 in Appendix A.3 establishes that H strictly quasi-convex at $\mathbf{x}$ implies H exposed at $\mathbf{x}$, and that the converse implication holds provided the strict inequality required for H to be exposed at $\mathbf{x}$ holds for all points in the sub-differential of H, and not just those points in its relative interior.

The gauge function $H(\mathbf{z},\mathbf{x})$, $\langle H,\mathbf{E}^N\rangle\in\mathscr{H}^1$, is *strictly differentially quasi-convex* at $\mathbf{x}^0\in\mathbf{E}^N$ if it has a first and second differential in $\mathbf{x}$ in a neighborhood of $\mathbf{x}^0$, and its Hessian matrix of second partial derivatives in $\mathbf{x}$ is non-negative definite and is of rank $N-1$. Lemma 16.5 in Appendix A.3 establishes that H strictly differentiably quasi-convex implies H strictly quasi-convex, and also gives a partial converse. The following result relates the structural properties of curvature and differentiability.

> *Theorem* 26. Suppose $\mathbf{Z}$ is a non-empty subset of $\mathbf{E}^M$, and consider the class of gauge functions $\mathscr{H}^1$ and the class of restricted profit functions $\mathscr{P}^1$. Then, for the gauge and restricted profit functions in these classes which are dual under the implicit profit and implicit gauge mappings of Theorem 24, the structural relationships given in Table 4 hold; i.e., the gauge function has the property "P" if and only if the restricted profit function has the property "Q".

Proof: Most of the results are established in Lemma 16.7 in Appendix A.3, as follows: Result 1, "P" implies "Q", is established by 16.7(3), and the converse is trivial. Result 2, "Q" implies "P", is established by 16.7(4), and the converse is trivial. Result 3, "Q" implies "P", is established by 16.7(6). The converse is established by the proof of 16.7(9), using Appendix A.3, Lemma 16.5(1) rather than 16.5(2). Result 4 follows from an argument dual to that for Result 3. Result 5, "P" implies "Q", is

[15]Since H is linear homogeneous in $\mathbf{x}$, this definition is equivalent to the requirement that an open line segment between $\mathbf{x}$ and any point $\mathbf{x}'$ in the lower contour set $\{\mathbf{x}'|H(\mathbf{z},\mathbf{x}')\leqq H(\mathbf{z},\mathbf{x})\}$ be contained in the interior of this set. The condition that H be strictly quasi-convex for all $\mathbf{z}\in\mathbf{E}^N$ is then equivalent to the usual definition of strict quasi-convexity.

TABLE 4
Property "P" holds for the gauge function $\langle H,\mathbf{X}\rangle \in \mathscr{H}^1$ if and only if property "Q" holds for its restricted profit function $\langle \Pi,\mathbf{P}\rangle \in \mathscr{P}^1$.

	"P" on the gauge function $H(\mathbf{z},\mathbf{x})$, $\mathbf{x} \in \mathbf{R}^N$	"Q" on the restricted profit function $\Pi(\mathbf{z},\mathbf{p})$, $\mathbf{p} \in \mathbf{P}(\mathbf{z})$
1.	$H(\mathbf{z},\mathbf{x})$ differentiable at $\mathbf{x} \in \mathbf{X}^*(\mathbf{z})$	$\Pi(\mathbf{z},\mathbf{p})$ exposed at $\mathbf{p}$, where $\{\mathbf{p}\} = \Lambda(\mathbf{z},\mathbf{x})$, $\mathbf{x} \in \mathbf{X}^*(\mathbf{z})$
2.	$H(\mathbf{z},\mathbf{x})$ exposed at $\mathbf{x}$, where $\{\mathbf{x}\} = \Gamma(\mathbf{z},\mathbf{p})$, $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$	$\Pi(\mathbf{z},\mathbf{p})$ differentiable at $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$
3.	$H(\mathbf{z},\mathbf{x})$ differentiable at all $\mathbf{x}$ in the relative interior of $\Gamma(\mathbf{z},\mathbf{p})$ where $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$	$\Pi(\mathbf{z},\mathbf{p})$ exposed at $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$
4.	$H(\mathbf{z},\mathbf{x})$ exposed at $\mathbf{x} \in \mathbf{X}^*(\mathbf{z})$	$\Pi(\mathbf{z},\mathbf{p})$ differentiable at all $\mathbf{p}$ in the relative interior of $\Lambda(\mathbf{z},\mathbf{x})$, where $\mathbf{x} \in \mathbf{X}^*(\mathbf{z})$
5.	$H(\mathbf{z},\mathbf{x})$ strictly quasi-convex at $\mathbf{x} \in \mathbf{X}^*(\mathbf{z})$	$\Pi(\mathbf{z},\mathbf{p})$ differentiable at all $\mathbf{p} \in \Lambda(\mathbf{z},\mathbf{x})$, where $\mathbf{x} \in \mathbf{X}^*(\mathbf{z})$
6.	$H(\mathbf{z},\mathbf{x})$ differentiable at all $\mathbf{x} \in \Gamma(\mathbf{z},\mathbf{p})$, where $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$	$\Pi(\mathbf{z},\mathbf{p})$ strictly quasi-convex at $\mathbf{p} \in \mathbf{P}^0(\mathbf{z})$
7.	$H(\mathbf{z},\mathbf{x})$ possesses a continuous first and second differential in $\mathbf{x}$ in a neighborhood of $\mathbf{x}^0 \in \mathbf{X}^*(\mathbf{z})$, and is strictly differentially quasi-convex on this neighborhood	$\Pi(\mathbf{z},\mathbf{p})$ possesses a continuous first and second differential in $\mathbf{p}$ in a neighborhood of $\mathbf{p}^0$, where $\{\mathbf{p}^0\} = \Lambda(\mathbf{z},\mathbf{x}^0)$, and is strictly differentially quasi-convex on this neighborhood
8.	$H(\mathbf{z},\mathbf{x})$ possesses a continuous first and second differential in $\mathbf{x}$ in a neighborhood of $\mathbf{x}^0$, where $\{\mathbf{x}^0\} = \Gamma(\mathbf{z},\mathbf{p}^0)$, and is strictly differentially quasi-convex on this neighborhood	$\Pi(\mathbf{z},\mathbf{p})$ possesses a continuous first and second differential in $\mathbf{p}$ in a neighborhood of $\mathbf{p}^0 \in \mathbf{P}^0(\mathbf{z})$, and is strictly differentially quasi-convex on this neighborhood

established by 16.7(7), and the converse is established by 16.7(10). Result 6 follows from an argument dual to that for Result 5. Results 7 and 8 follow from 16.7(11) and 16.7(12), plus the observation that $\{\mathbf{p}^0\} = \Lambda(\mathbf{z},\mathbf{x}^0)$ and $\{\mathbf{x}^1\} = \Gamma(\mathbf{z},\mathbf{p}^0)$ imply $\mathbf{x}^1 = \mathbf{x}^0$. Q.E.D.

The geometry of the structural relationships in Theorem 26 is summarized in Figure 22. As in the geometry of the two-input cost function, "kinks" are mapped into "flats", and vice versa. The linear segment $\mathbf{x}^1\mathbf{x}^2$ maps into the point $\mathbf{p}^1$. Note that the "price frontier" is not horizontal at $\mathbf{p}^1$, allowing this price frontier to be supported at $\mathbf{p}^1$ by planes with normals in the line segment through $\mathbf{x}^1\mathbf{x}^2$ (and extending indefinitely to

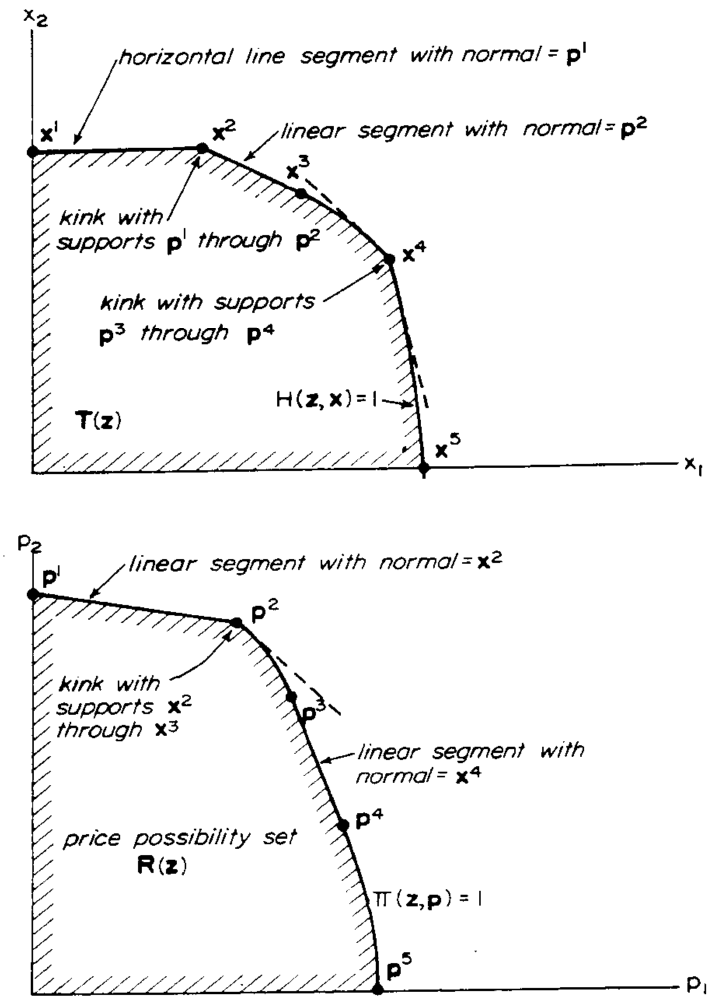


FIGURE 22

the left). The kink $\mathbf{x}^2$ maps into the linear segment $\mathbf{p}^1\mathbf{p}^2$, and the normal to this linear segment is $\mathbf{x}^2$. The linear segment $\mathbf{x}^2\mathbf{x}^3$ maps into the point $\mathbf{p}^2$. Note that the price frontier is linear to the left of $\mathbf{p}^2$ and curved to the right of $\mathbf{p}^2$, reflecting the fact that a kink occurred at the $\mathbf{x}^2$ end of the line segment $\mathbf{x}^2\mathbf{x}^3$, but the boundary of $\mathbf{T}$ is smooth at the $\mathbf{x}^3$ end of the line segment. The curve $\mathbf{x}^3\mathbf{x}^4$ maps into the curve $\mathbf{p}^2\mathbf{p}^3$, and the kink $\mathbf{x}^4$ maps into the line segment $\mathbf{p}^3\mathbf{p}^4$. Finally, the curve $\mathbf{x}^4\mathbf{x}^5$ maps into the curve $\mathbf{p}^4\mathbf{p}^5$, with the vertical tangent at $\mathbf{p}^5$ corresponding to the absence of a "flat" above $\mathbf{x}^5$. Note that the boundary of $\mathbf{T}$ is differentiable at $\mathbf{x}^1$, $\mathbf{x}^3$, and $\mathbf{x}^5$, exposed at $\mathbf{x}^2$, and strictly quasi-convex at $\mathbf{x}^4$. Correspondingly, the price frontier is exposed at $\mathbf{p}^1$ and $\mathbf{p}^2$, and differentiable at $\mathbf{p}^3$, $\mathbf{p}^4$, and $\mathbf{p}^5$.

The *price possibility set* used in Figure 22 can be introduced formally as

$$\begin{aligned}\mathbf{R}(\mathbf{z}) &= \{\mathbf{p} \in \mathbf{P}(\mathbf{z}) | \Pi(\mathbf{z},\mathbf{p}) \leqq 1\} \\ &= \{\mathbf{p} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{x} \leqq 1 \text{ for all } \mathbf{x} \in \mathbf{T}(\mathbf{z})\}. \end{aligned} \tag{47}$$

This set plays the same role as the factor price requirement set in the analysis of cost functions, and can be viewed as the formal dual of the set $\mathbf{T}(\mathbf{z})$. Figure 23 gives two further illustrations of this relationship. In case (a), with commodity 1 an input to production and commodity 2 an output, a translation of the production frontier has been made so that the

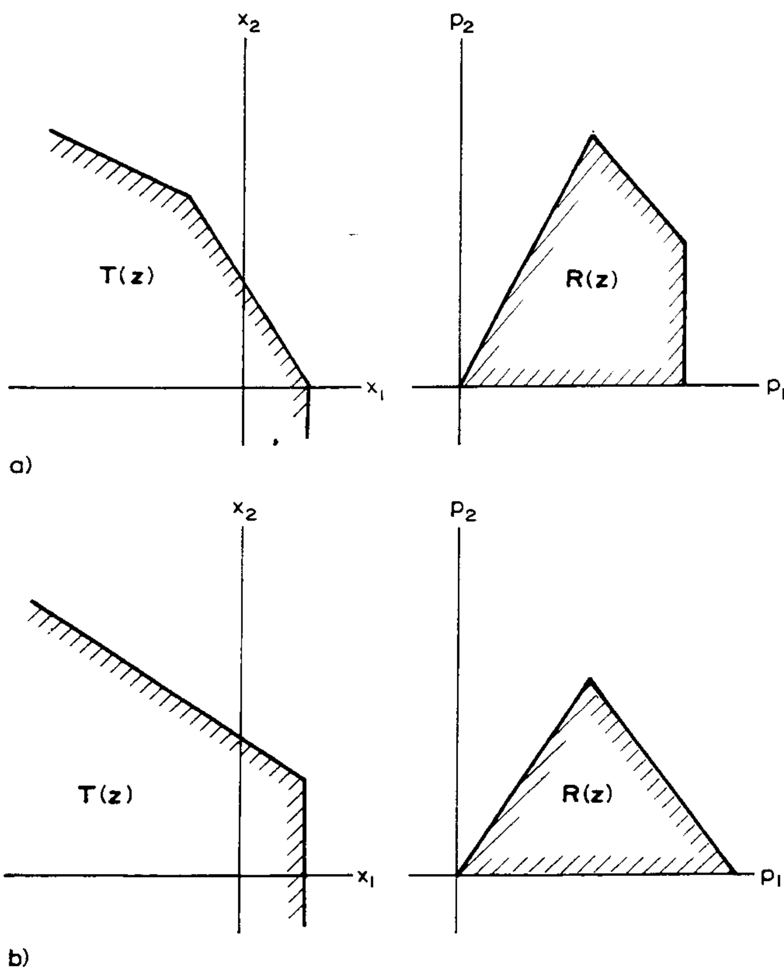


FIGURE 23

origin is interior to the possibility set. Correspondingly, the restricted profit function is positive for all non-zero $\mathbf{p} \in \mathbf{P}(\mathbf{z})$, and the price frontier is bounded. Case (b) illustrates a constant returns technology [i.e., the set $\mathbf{T}(\mathbf{z})$ is a convex cone] translated again so that the origin is an interior point.

18. Examples of Restricted Profit Functions

Examples of restricted profit functions can be given for the Cobb–Douglas and C.E.S. production functions introduced in Section 9, and for a C.E.S. production frontier. Consider a homogeneous version of the Cobb–Douglas transformation function (17), defined for some $\mathbf{z}^0 \in \mathbf{Z}$,

$$x_{N+1}^{1/\mu} \leqq D(-x_1)^{\theta_1}(-x_2)^{\theta_2} \cdots (-x_N)^{\theta_N}, \tag{48}$$

where $x_1,\ldots,x_N \leqq 0$, $x_{N+1} \geqq 0$, $\theta_i > 0$, and $\sum_{i=1}^N \theta_i = 1$, and the scale parameter μ is less than one. The profit function when all inputs and output are variable is

$$\Pi(\mathbf{z}^0, p_1,\ldots,p_{N+1}) = D'' p_1^{-\theta_1\eta} p_2^{-\theta_2\eta} \cdots p_N^{-\theta_N\eta} p_{N+1}^{1+\eta}, \tag{49}$$

where $\eta = \mu/(1-\mu)$ and $D'' = (1-\mu)D^\eta \mu^\eta \theta_1^{\eta\theta_1} \cdots \theta_N^{\eta\theta_N}$.

When inputs $S+1,\ldots,N$ are held fixed, and the remaining inputs and output are variable, the restricted profit function is

$$\Pi(\mathbf{z}^0, x_{S+1},\ldots,x_N, p_1,\ldots,p_S, p_{N+1}) = D^* p_1^{-\theta_1\nu} \cdots p_S^{-\theta_S\nu} p_{N+1}^{1+\nu\nu'}, \tag{50}$$

where $\nu = \mu/(1-\mu\nu')$, $\nu' = \sum_{i=1}^S \theta_i$, and $D^* = (1-\mu\nu')D^\nu \mu^{\nu\nu'} \theta_1^{\nu\theta_1} \cdots \theta_S^{\nu\theta_S}(-x_{S+1})^{\nu\theta_{S+1}} \cdots (-x_N)^{\nu\theta_N}$. Note that this function is defined for all positive p_{N+1} provided $\mu\nu' < 1$ ($\mu < 1$ is not required).

Consider a homogeneous version of the C.E.S. transformation function (19), defined for some $\mathbf{z}^0 \in \mathbf{Z}$,

$$x_{N+1}^{1/\mu} \leqq [(-x_1/D_1)^{1-1/\sigma} + \cdots + (-x_N/D_N)^{1-1/\sigma}]^{1/(1-1/\sigma)}, \tag{51}$$

where $x_1,\ldots,x_N \leqq 0$, $x_{N+1} \geqq 0$, $D_1,\ldots,D_N > 0$, $\sigma > 0$, $\sigma \neq 1$, and $\mu < 1$. The profit function when all inputs and output are variable is

$$\Pi(\mathbf{z}^0, p_1,\ldots,p_{N+1}) = (1-\mu)\mu^\eta p_{N+1}^{1+\eta}[(p_1 D_1)^{1-\sigma} + \cdots + (p_N D_N)^{1-\sigma}]^{-\eta/(1-\sigma)}. \tag{52}$$

When inputs $S+1,\ldots,N$ and output $N+1$ are held fixed and the remaining inputs are variable, then the restricted profit function, equal to the negative of the cost function for the variable inputs, has been shown by Knut Mork (1976) to have the form

$$\pi^* = \Pi^*(\mathbf{z}^0, x_{S+1},\dots,x_{N+1},p_1,\dots,p_S) \quad \text{if} \quad x_{N+1}^{1/\mu(1-1/\sigma)} > \sum_{i=S+1}^{N} (-x_i/D_i)^{1-1/\sigma},$$

$$= 0 \quad \text{if} \quad x_{N+1}^{(1-1/\sigma)/\mu} \leqq \sum_{i=S+1}^{N} (-x_i/D_i)^{1-1/\sigma} \quad \text{and} \quad \sigma > 1, \tag{52a}$$

$$= -\infty \quad \text{if} \quad x_{N+1}^{(1-1/\sigma)/\mu} \leqq \sum_{i=S+1}^{N} (-x_i/D_i)^{1-1/\sigma} \quad \text{and} \quad 0 < \sigma < 1,$$

where the last alternative corresponds to a non-producible value of x_{N+1}, and where

$$\Pi^* = -\left[\sum_{i=1}^{S} (p_i D_i)^{1-\sigma}\right]^{1/(1-\sigma)} \left\{ x_{N+1}^{(1-1/\sigma)/\mu} - \sum_{i=S+1}^{N} (-x_i/D_i)^{1-1/\sigma} \right\}^{-\sigma/(1-\sigma)}. \tag{52b}$$

Suppose now that x_{N+1} is made variable, so that the restricted profit function becomes

$$\pi' = \Pi'(z^0, x_{S+1},\dots,x_N,p_1,\dots,p_S,p_{N+1}) = \operatorname*{Max}_{x_{N+1}} [p_{N+1}x_{N+1} + \pi^*], \tag{52c}$$

with π^* given in (52a). For $\mu = 1$, Mork shows this restricted profit function to have the form

$$\pi' = \Pi'(\mathbf{z}^0, x_{S+1},\dots,x_N,p_1,\dots,p,p_{N+1}) \quad \text{if} \quad p_{N+1}^{1-\sigma} > \sum_{i=1}^{S} (D_i p_i)^{1-\sigma},$$

$$= 0 \quad \text{if} \quad p_{N+1}^{1-\sigma} \leqq \sum_{i=1}^{S} (D_i p_i)^{1-\sigma} \quad \text{and} \quad 0 < \sigma < 1, \tag{52d}$$

$$= +\infty \quad \text{if} \quad p_{N+1}^{1-\sigma} \leqq \sum_{i=1}^{S} (D_i p_i)^{1-\sigma} \quad \text{and} \quad \sigma > 1,$$

where

$$\Pi' = \left[p_{N+1}^{1-\sigma} - \sum_{i=1}^{S} (D_i p_i)^{1-\sigma}\right]^{1/(1-\sigma)} \left[\sum_{i=S+1}^{N} (-x_i/D_i)^{1-1/\sigma}\right]^{1/(1-1/\sigma)}. \tag{52e}$$

For $\mu \neq 1$, the maximand of (52c) satisfies

$$\mu p_{N+1} \left\{ x_{N+1}^{(1-1/\sigma)/\mu} - \sum_{i=S+1}^{N} (-x_i/D_i)^{1-1/\sigma} \right\}^{1/(1-\sigma)}$$

$$= x_{N+1}^{-1+(1-1/\sigma)/\mu} \left[\sum_{i=1}^{S} (p_i D_i)^{1-\sigma}\right]^{1/(1-\sigma)}. \tag{52f}$$

In general, (52f) does not have a closed form solution for x_{N+1}, and consequently (52c) fails to have a closed form.

Consider a C.E.S. production frontier, defined for some $\mathbf{z}^0 \in \mathbf{Z}$ by

$$(-x_{N+1})^{\mu} \geqq [(x_1/D_1)^{1+1/\sigma} + \cdots + (x_N/D_N)^{1+1/\sigma}]^{1/(1+1/\sigma)}, \tag{53}$$

where $x_1,\ldots,x_N \geqq 0$, $x_{N+1} \leqq 0$, $D_1,\ldots,D_N > 0$, $\sigma > 0$, and $\mu < 1$. The revenue function when all outputs are variable and the input is fixed is

$$\Pi(\mathbf{z}^0, x_{N+1}, p_1,\ldots,p_N) = (-x_{N+1})^{\mu}[(p_1D_1)^{1+\sigma} + \cdots + (p_ND_N)^{1+\sigma}]^{1/(1+\sigma)}. \tag{54}$$

The profit function when all outputs and input are variable is

$$\Pi(\mathbf{z}^0, p_1,\ldots,p_{N+1}) = (1-\mu)\mu^{\eta}[(p_1D_1)^{1+\sigma} + \cdots + (p_ND_N)^{1+\sigma}]^{(1+\eta)/(1+\sigma)}p_{N+1}^{-\eta}, \tag{55}$$

where $\eta = \mu/(1-\mu)$.

19. Composition Rules for Profit Functions

Composition rules of the sort established for cost functions can also be derived for restricted profit functions. These rules allow the construction of complex functions from simple known forms.

Throughout the remainder of this section, we consider a set of parameters $\mathbf{Z}$ and a family of production possibility sets $\mathbf{T}^j(\mathbf{z})$, $j = 1,\ldots,J$, each satisfying Axioms 1–4, containing the origin in its interior, and having the free disposal property. Let $\mathbf{P}^j(\mathbf{z})$ denote the normal cone, $\Pi^j(\mathbf{z},\mathbf{p})$ the profit function, $H^j(\mathbf{z},\mathbf{x})$ the gauge function, and $\mathbf{R}^j(\mathbf{z})$ the price possibility set of $\mathbf{T}^j(\mathbf{z})$. The first result is due to Fenchel (1949, Result 41), and is also proved by Karlin (1959, 7.6.1).

Lemma 27. If $\mathbf{T}'(\mathbf{z}) = \bigcap_{j=1}^{J} \mathbf{T}^j(\mathbf{z})$, then $\mathbf{T}'(\mathbf{z})$ satisfies Axioms 1–4, contains the origin in its interior, has the free disposal property, and satisfies

(i) $H'(\mathbf{z},\mathbf{x}) = \mathrm{Max}_j\, H^j(\mathbf{z},\mathbf{x})$.

(ii) $\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z}) \subseteq \mathbf{P}'(\mathbf{z}) \subseteq \text{Closure } (\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z}))$.

(iii) For $p \in \sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$, $\Pi'(\mathbf{z},\mathbf{p}) = \inf\{\sum_{j=1}^{J} \Pi^j(\mathbf{z},\mathbf{p}^j) | \mathbf{p}^j \in \mathbf{P}^j(\mathbf{z}), \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\}$.

(iv) $\mathbf{R}'(\mathbf{z}) = \text{Convex Hull of } \bigcup_{j=1}^{J} \mathbf{R}^j(\mathbf{z})$.

The next result gives one general-purpose composition rule. Let $\mathbf{W}^+(\mathbf{z})$

denote a convex closed bounded subset of the non-negative orthant of $\mathbf{E}^J$ which contains a strictly positive vector, and let $\mathbf{W}(\mathbf{z})$ denote the set obtained from $\mathbf{W}^+(\mathbf{z})$ by free disposal; i.e., $\mathbf{W}(\mathbf{z}) = \{\mathbf{w} \in \mathbf{E}^J | \mathbf{w} \leqq \mathbf{w}' \in \mathbf{W}^+(\mathbf{z})\}$. Then, the origin is in the interior of $\mathbf{W}(\mathbf{z})$. For $\mathbf{q} \in \mathbf{R}^J$, $\mathbf{q}\cdot\mathbf{w}$ is bounded above on $\mathbf{W}(\mathbf{z})$ if and only if $\mathbf{q}$ is non-negative. Define $\Pi^*(\mathbf{z},\mathbf{q})$ to be the restricted profit function of $\mathbf{W}(\mathbf{z})$ for $\mathbf{z} \in \mathbf{Z}$, $\mathbf{q} \geqq \mathbf{0}$. Define $H^*(\mathbf{z},\mathbf{w})$ to be the gauge function of $\mathbf{W}(\mathbf{z})$ (centered at the origin of $\mathbf{E}^J$). Define $\mathbf{R}^*(\mathbf{z}) = \{\mathbf{q} \in \mathbf{E}^J | \Pi^*(\mathbf{z},\mathbf{q}) \leqq 1\}$ to be the price possibility set of $\mathbf{W}(\mathbf{z})$.

Theorem 28. If $H'(\mathbf{z},\mathbf{x}) = H^*(\mathbf{z},H^1(\mathbf{z},\mathbf{x}),\ldots,H^J(\mathbf{z},\mathbf{x}))$, then

(i) $H'(\mathbf{z},\mathbf{x})$ is the gauge function of a convex set in $\mathbf{E}^N$,

(ii) $\mathbf{T}'(\mathbf{z}) = \text{Closure } \bigcup_{\mathbf{w}\in\mathbf{W}^+(\mathbf{z})} \bigcap_{j=1}^J (w_j \mathbf{T}^j(\mathbf{z}))$.

(iii) $\sum_{j=1}^J \mathbf{P}^j(\mathbf{z}) \subseteq \mathbf{P}'(\mathbf{z}) \subseteq \text{Closure } (\sum_{j=1}^J \mathbf{P}^j(\mathbf{z}))$.

(iv) For $\mathbf{p} \in \sum_{j=1}^J \mathbf{P}^j(\mathbf{z})$, $\Pi'(\mathbf{z},\mathbf{p}) = \inf\{\Pi^*(\mathbf{z},\Pi^1(\mathbf{z},\mathbf{p}^1),\ldots,\Pi^J(\mathbf{z},\mathbf{p}^J)) | \mathbf{p}^j \in \mathbf{P}^j(\mathbf{z}), \sum_{j=1}^J \mathbf{p}^j = \mathbf{p}\}$.

(v) $\mathbf{R}'(\mathbf{z}) = \text{Closure } \bigcup_{\mathbf{q}\in\mathbf{R}^*(\mathbf{z})} \sum_{j=1}^J (q_j \mathbf{R}^j(\mathbf{z}))$.

Proof: It is immediate that $H'(\mathbf{z},\mathbf{x})$ is positively linear homogeneous in $\mathbf{x}$, with $H'(\mathbf{z},\mathbf{0}) = 0$. By convexity, $H^*(\mathbf{z},\mathbf{w}) \leqq H^*(\mathbf{z},\mathbf{w}-\mathbf{w}') + H^*(\mathbf{z},\mathbf{w}')$. If $\mathbf{w}' \geqq \mathbf{w}$, then $H^*(\mathbf{z},\mathbf{w}-\mathbf{w}') = 0$ by the free disposal property of $\mathbf{W}(\mathbf{z})$. Hence, $H^*(\mathbf{z},\mathbf{w})$ is non-decreasing in $\mathbf{w}$. Now consider $\mathbf{x}'$, $\mathbf{x}'' \in \mathbf{R}^N$, $\mathbf{x}^* = \theta\mathbf{x}' + (1-\theta)\mathbf{x}''$ for some $0 < \theta < 1$. Then, $H^j(\mathbf{z},\mathbf{x}^*) \leqq \theta H^j(\mathbf{z},\mathbf{x}') + (1-\theta)H^j(\mathbf{z},\mathbf{x}'')$. Hence,

$$\begin{aligned}
&H^*(\mathbf{z},H^1(\mathbf{z},\mathbf{x}^*),\ldots,H^J(\mathbf{z},\mathbf{x}^*))\\
&\leqq H^*(\mathbf{z},\theta H^1(\mathbf{z},\mathbf{x}')+(1-\theta)H^1(\mathbf{z},\mathbf{x}''),\ldots,\theta H^J(\mathbf{z},\mathbf{x}')+(1-\theta)H^J(\mathbf{z},\mathbf{x}''))\\
&\leqq \theta H^*(\mathbf{z},H^1(\mathbf{z},\mathbf{x}'),\ldots,H^J(\mathbf{z},\mathbf{x}'))+(1-\theta)H^*(\mathbf{z},H^1(\mathbf{z},\mathbf{x}''),\ldots,H^J(\mathbf{z},\mathbf{x}'')),
\end{aligned}$$

and $H'(\mathbf{z},\mathbf{x})$ is convex in $\mathbf{x}$. This establishes (i).

By definition, $\mathbf{T}'(\mathbf{z}) = \{\mathbf{x} | H'(\mathbf{z},\mathbf{x}) \leqq 1\}$. But $H'(\mathbf{z},\mathbf{x}) \leqq 1$ if and only if there exists $\mathbf{w} \in \mathbf{E}^J$, $w \geqq 0$, such that $H^*(\mathbf{z},\mathbf{w}) \leqq 1$ and $H^j(\mathbf{z},\mathbf{x}) \leqq w_j$. Then, we may without loss of generality choose $\mathbf{w} \in \mathbf{W}^+(\mathbf{z})$. If $w_j = 0$, then $\{\mathbf{x} | H^j(\mathbf{z},\mathbf{x}) = 0\} = \mathbf{AT}^j(\mathbf{z})$, the asymptotic cone of $\mathbf{T}^j(\mathbf{z})$. If $w_j > 0$, then $\{\mathbf{x} | H^j(\mathbf{z},\mathbf{x}) \leqq w_j\} = w_j\mathbf{T}^j(\mathbf{z})$, and $w_j\mathbf{T}^j(\mathbf{z})$ contains $\mathbf{AT}^j(\mathbf{z})$. Since $\mathbf{W}^+(\mathbf{z})$ contains a positive vector, every vector in $\mathbf{W}^+(\mathbf{z})$ is a limit of strictly positive vectors in $\mathbf{W}^+(\mathbf{z})$. Hence,

$$\begin{aligned}\mathbf{T}'(\mathbf{z}) &= \bigcup_{\mathbf{w}\in\mathbf{W}^+(\mathbf{z})} \bigcap_{j=1}^{J} \{\mathbf{x}|H^j(\mathbf{z},\mathbf{x}) \leqq w_j\} \\ &= \bigcup_{\mathbf{w}\in\mathbf{W}^+(\mathbf{z})} \bigcap_{j=1}^{J} (w_j\mathbf{T}^j(\mathbf{z}) + \mathbf{AT}^j(\mathbf{z})) \\ &= \text{Closure} \bigcup_{\mathbf{w}\in\mathbf{W}^+(\mathbf{z})} \bigcap_{j=1}^{J} (w_j\mathbf{T}^j(\mathbf{z})).\end{aligned}$$

This proves (ii).

To verify (iii), note first that the normal cone of $w_j\mathbf{T}^j(\mathbf{z}) + \mathbf{AT}^j(\mathbf{z})$ is $\mathbf{P}^j(\mathbf{z})$ for every $w_j \geqq 0$. Hence by Lemma 27, the normal cone of $\bigcap_{j=1}^{J} (w_j\mathbf{T}^j(\mathbf{z}) + \mathbf{AT}^j(\mathbf{z}))$ contains $\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$ and is contained in the closure of this set. Now, the normal cone of the union of an arbitrary collection of sets is contained in the intersection of the normal cones of its members. Hence, the normal cone of $\mathbf{T}'(\mathbf{z})$ is contained in the closure of $\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$. Finally, since $\mathbf{W}^+(\mathbf{z})$ is bounded, there exists a vector $\bar{\mathbf{w}} \geqq \mathbf{w}$ for all $\mathbf{w} \in \mathbf{W}^+(\mathbf{z})$, implying $\mathbf{T}'(\mathbf{z}) \subseteq \bigcap_{j=1}^{J} (\bar{w}_j\mathbf{T}^j(\mathbf{z}))$. The normal cone of this last set contains $\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$, and is contained in the normal cone of $\mathbf{T}'(\mathbf{z})$. Hence, (iii) holds.

Consider $\mathbf{p} \in \sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$. By Lemma 27,

$$\sup\left\{\mathbf{p}\cdot\mathbf{x}|\mathbf{x} \in \bigcap_{j=1}^{J} (w_j\mathbf{T}^j(\mathbf{z}))\right\} = \inf\left\{w_1\Pi^1(\mathbf{z},\mathbf{p}^1) + \cdots + w_J\Pi^J(\mathbf{z},\mathbf{p}^J)|\mathbf{p}^j \in \mathbf{P}^j(\mathbf{z}), \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\right\}.$$

Hence,

$$\begin{aligned}\Pi'(\mathbf{z},\mathbf{p}) &= \sup_{\mathbf{w}\in\mathbf{W}^+(\mathbf{z})} \inf\left\{w_1\Pi^1(\mathbf{z},\mathbf{p}^1) + \cdots + w_J\Pi^J(\mathbf{z},\mathbf{p}^J)|\mathbf{p}^j \in \mathbf{P}^j(\mathbf{z}), \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\right\} \\ &= \inf\left\{\sup_{\mathbf{w}\in\mathbf{W}^+(\mathbf{z})} (w_1\Pi^1(\mathbf{z},\mathbf{p}^1) + \cdots + w_J\Pi^J(\mathbf{z},\mathbf{p}^J))|\mathbf{p}^j \in \mathbf{P}^j(\mathbf{z}), \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\right\} \\ &= \inf\left\{\Pi^*(\mathbf{z},\Pi^1(\mathbf{z},\mathbf{p}^1),\ldots,\Pi^J(\mathbf{z},\mathbf{p}^J))|\mathbf{p}^j \in \mathbf{P}^j(\mathbf{z}), \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\right\},\end{aligned}$$

with the first equality holding by the minimax theorem [Rockafellar

(1970, Corollary 37.3.2)] since $\mathbf{W}^{+}(\mathbf{z})$ is bounded, and the second equality holding by the definition of Π^*. This proves (iv).

If $\mathbf{p} \in \bigcup_{\mathbf{q}\in\mathbf{R}^*(\mathbf{z})} \sum_{j=1}^{J} q_j \mathbf{R}^j(\mathbf{z})$, then there exist $\mathbf{q} \in \mathbf{R}^*(\mathbf{z})$ and $\mathbf{p}^j \in \mathbf{R}^j(\mathbf{z})$ such that $\mathbf{p} = \sum_{j=1}^{J} q_j \mathbf{p}^j$, implying $\Pi'(\mathbf{z},\mathbf{p}) \leqq \Pi^*(\mathbf{z}, q_1 \Pi^1(\mathbf{z},\mathbf{p}^1),\ldots,q_J \Pi^J(\mathbf{z},\mathbf{p}^J)) \leqq \Pi^*(\mathbf{z},\mathbf{q}) \leqq 1$, and hence $\mathbf{p} \in \mathbf{R}'(\mathbf{z}) = \{\mathbf{p} \in \mathbf{R}^N | \Pi'(\mathbf{z},\mathbf{p}) \leqq 1\}$. Alternately, suppose $\mathbf{p} \in \mathbf{R}'(\mathbf{z})$. Then for any $\epsilon > 0$, there exist $\mathbf{p}^j \in \mathbf{P}^j(\mathbf{z})$ such that $\Pi^*(\mathbf{z}, \Pi^1(\mathbf{z},\mathbf{p}^1),\ldots,\Pi^J(\mathbf{z},\mathbf{p}^J)) \leqq 1+\epsilon$ and $\mathbf{p} = \sum_{j=1}^{J} \mathbf{p}^j$. Let $q_j = \Pi^j(\mathbf{z},\mathbf{p}^j)$. Then $\Pi^*(\mathbf{z},\mathbf{q}) \leqq 1+\epsilon$ and $\Pi^j(\mathbf{z},\mathbf{p}^j) \leqq q_j$, implying $\mathbf{p}^j \in q_j \mathbf{R}^j(\mathbf{z})$. Hence, $\mathbf{p} \in \sum_{j=1}^{J} (q_j \mathbf{R}^j(\mathbf{z}))$, and therefore $\mathbf{p} \in \bigcup_{\mathbf{q}\in(1+\epsilon)\mathbf{R}^*(\mathbf{z})} \sum_{j=1}^{J} (q_j \mathbf{R}^j(\mathbf{z}))$. Letting $\epsilon \to 0$ establishes

$$\mathbf{R}'(\mathbf{z}) \subseteq \text{Closure} \bigcup_{\mathbf{q}\in\mathbf{R}^*(\mathbf{z})} \sum_{j=1}^{J} (q_j \mathbf{R}^j(\mathbf{z})).$$

With the preceding inclusion, this proves (v). Q.E.D.

Using the formal duality of the gauge and profit functions, one obtains the following corollary to Lemma 27 and Theorem 28.

Corollary 28a. Suppose the production possibility sets $\mathbf{T}^j(\mathbf{z})$, price possibility sets $\mathbf{R}^j(\mathbf{z})$, gauge functions $H^j(\mathbf{z},\mathbf{x})$, and profit functions $\Pi^j(\mathbf{z},\mathbf{p})$ satisfy the assumptions preceding Lemma 27 for $j = 1,\ldots,J$. Suppose the production possibility set $\mathbf{W}(\mathbf{z})$, price possibility set $\mathbf{R}^*(\mathbf{z})$, gauge function $H^*(\mathbf{z},\mathbf{w})$, and profit function $\Pi^*(\mathbf{z},\mathbf{q})$ satisfy the assumptions preceding Theorem 28. Then composition Rules 1–3 in Table 5 hold. Under the additional assumption that the interior of $\bigcap_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$ is non-empty, Rules 4–6 in this table hold. (In Rules 2 and 4, $\mathbf{\Delta}$ is the unit simplex.)

Proof: Rules 1 and 3 are restatements of Lemma 27 and Theorem 28, while Rule 2 is a special case of Rule 3 when $\mathbf{W}^{+}(\mathbf{z}) = \mathbf{\Delta}$. Rule 4 is a formal dual of Rule 2, Rule 5 is a formal dual of Rule 1, and Rule 6 is a formal dual of Rule 3, with the expression for $\mathbf{P}'(\mathbf{z})$ following in these rules from application of Rockafellar (1970, Corollary 8.3.3 and Corollary 16.4.2) to the dual asymptotic and normal cones of the $\mathbf{T}^j(\mathbf{z})$. The assumption that $\mathbf{P}'(\mathbf{z})$ has a non-empty interior in Rules 4, 5, 6 implies that the technologies $\mathbf{T}'(\mathbf{z})$ are semi-bounded. Hence, they satisfy Axiom 1, and the set $\mathbf{T}'(\mathbf{z})$ defined in Rules 4 and 5 is closed. Q.E.D.

An example illustrates the use of these composition rules to generate new functional forms. Let $\mathbf{A} = (a_{ij})$ denote a symmetric, positive definite,

TABLE 5
Composition rules for gauge and profit functions.

	Rule 1	Rule 2	Rule 3
$\mathbf{T}'(\mathbf{z})$	$\bigcap_{j=1}^{J} \mathbf{T}^j(\mathbf{z})$	Closure $\bigcup_{\mathbf{w}\in\Delta} \bigcap_{j=1}^{J} \mathbf{T}^j(\mathbf{z})w_j$	Closure $\bigcup_{\mathbf{w}\in\mathbf{W}^*(\mathbf{z})} \bigcap_{j=1}^{J} \mathbf{T}^j(\mathbf{z})w_j$
$H'(\mathbf{z},\mathbf{x})$	$\mathrm{Max}_j\, H^j(\mathbf{z},\mathbf{x})$	$\sum_{j=1}^{J} H^j(\mathbf{z},\mathbf{x})$	$H^*(\mathbf{z},H^1(\mathbf{z},\mathbf{x}),\ldots,H^J(\mathbf{z},\mathbf{x}))$
The interior of $\mathbf{P}'(\mathbf{z})$ equals the interior of the following set	$\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$	$\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$	$\sum_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$
$\Pi'(\mathbf{z},\mathbf{p})$	$\inf\left\{\sum_{j=1}^{J} \Pi^j(\mathbf{z},\mathbf{p}^j) \,\middle\vert\, \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\right\}$	$\inf\left\{\mathrm{Max}_j\, \Pi^j(\mathbf{z},\mathbf{p}^j) \,\middle\vert\, \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\right\}$	$\inf\left\{\Pi^*(\mathbf{z},\Pi^1(\mathbf{z},\mathbf{p}^1),\ldots,\Pi^J(\mathbf{z},\mathbf{p}^J)) \,\middle\vert\, \sum_{j=1}^{J} \mathbf{p}^j = \mathbf{p}\right\}$
$\mathbf{R}'(\mathbf{z})$	Convex Hull of $\bigcup_{j=1}^{J} \mathbf{R}^j(\mathbf{z})$	$\sum_{j=1}^{J} \mathbf{R}^j(\mathbf{z})$	Closure $\bigcup_{\mathbf{q}\in\mathbf{R}^*(\mathbf{z})} \sum_{j=1}^{J} (q_j\mathbf{R}^j(\mathbf{z}))$

	Rule 4	Rule 5	Rule 6
$\mathbf{T}'(\mathbf{z})$	$\sum_{j=1}^{J} \mathbf{T}^j(\mathbf{z})$	Convex Hull of $\bigcup_{j=1}^{J} \mathbf{T}^j(\mathbf{z})$	Closure $\bigcup_{\mathbf{w}\in\mathbf{W}^*(\mathbf{z})} \sum_{j=1}^{J} (w_j\mathbf{T}^j(\mathbf{z}))$
$H'(\mathbf{z},\mathbf{x})$	$\inf\left\{\mathrm{Max}_j\, H^j(\mathbf{z},\mathbf{x}^j) \,\middle\vert\, \sum_{j=1}^{J} \mathbf{x}^j = \mathbf{x}\right\}$	$\inf\left\{\sum_{j=1}^{J} H^j(\mathbf{z},\mathbf{x}^j) \,\middle\vert\, \sum_{j=1}^{J} \mathbf{x}^j = \mathbf{x}\right\}$	$\inf\left\{H^*(\mathbf{z},H^1(\mathbf{z},\mathbf{x}^1),\ldots,H^J(\mathbf{z},\mathbf{x}^J)) \,\middle\vert\, \sum_{j=1}^{J} \mathbf{x}^j = \mathbf{x}\right\}$
The interior of $\mathbf{P}'(\mathbf{z})$ equals the interior of the following set	$\bigcap_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$	$\bigcap_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$	$\bigcap_{j=1}^{J} \mathbf{P}^j(\mathbf{z})$
$\Pi'(\mathbf{z},\mathbf{p})$	$\sum_{j=1}^{J} \Pi^j(\mathbf{z},\mathbf{p})$	$\mathrm{Max}_j\, \Pi^j(\mathbf{z},\mathbf{p})$	$\Pi^*(\mathbf{z},\Pi^1(\mathbf{z},\mathbf{p}),\ldots,\Pi^J(\mathbf{z},\mathbf{p}))$
$\mathbf{R}'(\mathbf{z})$	Closure $\bigcup_{\mathbf{q}\in\Delta} \bigcap_{j=1}^{J} (q_i\mathbf{R}^i(\mathbf{z}))$	$\bigcap_{j=1}^{J} \mathbf{R}^j(\mathbf{z})$	Closure $\bigcup_{\mathbf{q}\in\mathbf{R}^*(\mathbf{z})} \bigcap_{j=1}^{J} \mathbf{R}^j(\mathbf{z})q_i$

non-negative matrix of order J. Then,

$$\begin{aligned} \Pi^*(\mathbf{q}) &= (\mathbf{q}'\mathbf{A}\mathbf{q})^{1/2} \quad \text{if} \quad \mathbf{q} \geqq \mathbf{0}, \\ &= +\infty \qquad \text{otherwise}, \end{aligned} \tag{56}$$

is a profit function which is non-decreasing in $\mathbf{q} \geqq \mathbf{0}$. The dual gauge function is

$$H^*(\mathbf{w}) = \inf\{(\hat{\mathbf{w}}'A^{-1}\hat{\mathbf{w}})^{1/2} | \hat{\mathbf{w}} \geqq \mathbf{w}\}. \tag{57}$$

[The duality of the functions $(\mathbf{q}'\mathbf{A}\mathbf{q})^{1/2}$ and $(\mathbf{w}'\mathbf{A}^{-1}\mathbf{w})^{1/2}$ is established from equation (46) using Schwartz's inequality; see Rockafellar (1970, p. 136). The modification in equations (56) and (57) can be obtained from composition Rule 5.] For a sequence of non-negative profit functions $\Pi^i(\mathbf{z},\mathbf{p})$ and dual gauge functions $H^i(\mathbf{z},\mathbf{w})$, the composites

$$\Pi^0(\mathbf{z},\mathbf{p}) = \left[\sum_{i,j=1}^{J} a_{ij}\Pi^i(\mathbf{z},\mathbf{p})\Pi^j(\mathbf{z},\mathbf{p})\right]^{1/2}, \tag{58}$$

and

$$H^0(\mathbf{z},\mathbf{x}) = \left[\sum_{i,j=1}^{J} b_{ij}H^i(\mathbf{z},\mathbf{x})H^j(\mathbf{z},\mathbf{x})\right]^{1/2}, \tag{59}$$

where the b_{ij} are elements of $\mathbf{A}^{-1}$, define new profit and gauge functions. Taking the coefficients a_{ij} to be non-negative parameters and the Π^i to be concrete functions (such as Cobb–Douglas profit functions in all prices or subsets of prices), one obtains a broad parametric class of profit functions with netput supply functions

$$\mathbf{x}_k = \frac{1}{\Pi^0} \sum_{i,j=1}^{J} a_{ij}\Pi^i \frac{\partial \Pi^j}{\partial p^k},$$

which are linear in the coefficients a_{ij}. Such linear-in-parameters forms have obvious econometric applications; this topic is discussed further in Chapter II.2.

Composition rules can also be used to deduce results on the structure of technology and separability properties of netputs. In Chapter I.3, Lau has several applications, including conditions for non-joint production (Theorem III.6) and for separability of inputs (Theorem II.6). In Chapter V.1, Denny establishes a condition for separability of inputs and outputs.

20. Profit Saddle-Functions

Suppose a production possibility set **Y** contains production plans (**z**,**x**), where $\mathbf{z} = (z_1,\ldots,z_M)$ and $\mathbf{x} = (x_1,\ldots,x_N)$ are netput vectors which are distinguished from each other in an application. For example, **z** may denote netputs which are fixed in the short run and variable in the long run, while **x** denotes short run variable netputs. Alternately, **z** may represent netputs of primary goods (capital and labor), with **x** representing all other goods; or **z** may represent outputs and **x** inputs (expressed as netputs). It is possible to define gauge and profit functions with

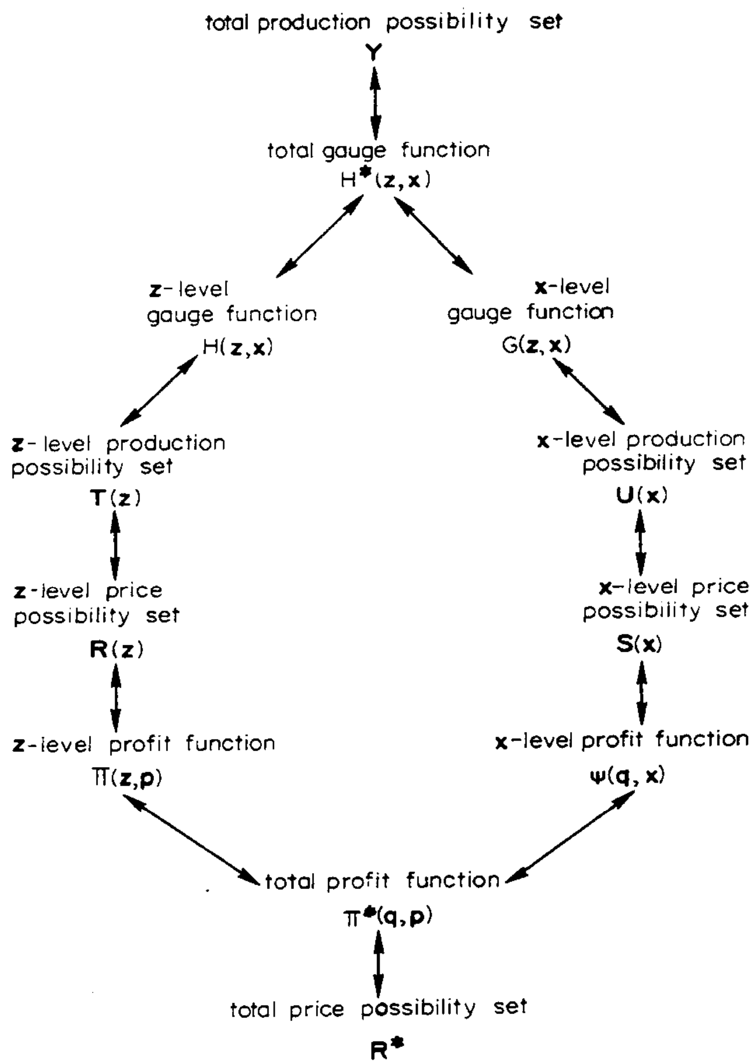


FIGURE 24

respect to either or both the netput vectors **z** and **x**. These functions and the mappings between them provide a basis for the analysis of problems such as the relation of short and long run profit maximizing behavior, or the relation of value added to revenue and profit.

Figure 24 outlines the various possibility sets and gauge and profit functions we shall consider. These sets and functions will be assumed to lie in classes defined by the properties in Table 6. We shall establish that for members of these classes, the mappings in Table 7 hold.

> *Theorem* 29. The mappings of Table 7 hold for members of the classes of functions and sets defined in Table 6, and for these classes are one to one onto. That is, for each member of a class in Table 6, the images of the mappings in Table 7 have the associated properties in Table 6; each member of a class in Table 6 is the image of a unique member of each of the remaining classes in Table 6 under the mappings of Table 7; and the mappings are invertible in the sense that any composition of Table 7 mappings which leads from a class in Table 6 back to the same class reduces to the identity mapping.

TABLE 6
Properties of possibility sets and profit functions.

1. Domains and centering functions

$\mathbf{Z}$ = a non-empty convex subset of $\mathbf{E}^M$ (60)

$\mathbf{X}$ = a non-empty convex subset of $\mathbf{E}^N$ (61)

$(\bar{\mathbf{z}},\bar{\mathbf{x}})$ = a point in $\mathbf{Z}\times\mathbf{X}$ (62)

$\mathbf{x}^*:\mathbf{Z}\to\mathbf{X}$ = a continuous function (63)

$\mathbf{z}^*:\mathbf{X}\to\mathbf{Z}$ = a continuous function (64)

2. The production possibility set $\mathbf{Y}\subseteq\mathbf{E}^M\times\mathbf{E}^N$

(i) $\mathbf{Y}$ is non-empty, convex, closed, and semi-bounded
(ii) $\mathbf{Z}=\{\mathbf{z}\in\mathbf{E}^M|(\mathbf{z},\mathbf{x})\in\mathbf{Y}$ for some $\mathbf{x}\in\mathbf{E}^N\}$
$\mathbf{X}=\{\mathbf{x}\in\mathbf{E}^N|(\mathbf{z},\mathbf{x})\in\mathbf{Y}$ for some $\mathbf{z}\in\mathbf{E}^M\}$; $(\bar{\mathbf{z}},\bar{\mathbf{x}})\in\mathbf{Y}$; $(\mathbf{z},\mathbf{x}^*(\mathbf{z}))\in\mathbf{Y}$ for $\mathbf{z}\in\mathbf{Z}$;
$(\mathbf{z}^*(\mathbf{x}),\mathbf{x})\in\mathbf{Y}$ for $\mathbf{x}\in\mathbf{X}$

3. The total gauge function $H^*:\mathbf{E}^M\times\mathbf{E}^N\to\bar{\mathbf{E}}$ [see equations (72) and (65)][a]

(i) H^* is non-negative, convex, and closed
(ii) H^* is finite on a non-empty convex cone with vertex at $(\bar{\mathbf{z}},\bar{\mathbf{x}})$
(iii) $\{(\mathbf{z},\mathbf{x})\in\mathbf{E}^M\times\mathbf{E}^N|H^*(\mathbf{z},\mathbf{x})=0\}$ is semi-bounded
(iv) For $\lambda\geqq 0$, $H^*(\bar{\mathbf{z}}+\lambda(\mathbf{z}-\bar{\mathbf{z}}),\ \bar{\mathbf{x}}+\lambda(\mathbf{x}-\bar{\mathbf{x}}))=\lambda H^*(\mathbf{z},\mathbf{x})$

TABLE 6 (*continued*)

4. The z-level gauge function $H:\mathbf{Z}\times\mathbf{E}^N\rightarrow\bar{\mathbf{E}}$ [see equations (80) and (66)][a,b]
   (i) For each $\mathbf{z}\in\mathbf{Z}$, $H(\mathbf{z},\mathbf{x})$ is non-negative, convex, and closed in $\mathbf{x}$
   (ii) For $\theta\geqq 0$, $H(\mathbf{z},\mathbf{x}^*(\mathbf{z})+\theta(\mathbf{x}-\mathbf{x}^*(\mathbf{z})))=\theta H(\mathbf{z},\mathbf{x})$
   (iii) $\{(\mathbf{z},\mathbf{x})\in\mathbf{Z}\times\mathbf{E}^N|H(\mathbf{z},\mathbf{x})\leqq 1\}$ is semi-bounded
   (iv) $\{(\mathbf{z},\mathbf{x})\in\mathbf{Z}\times\mathbf{E}^N|H(\mathbf{z},\mathbf{x})\leqq 1\}$ is convex and closed[f]
   (v) $H(\bar{\mathbf{z}},\bar{\mathbf{x}})\leqq 1$

5. The z-level production possibility set $\mathbf{T}(\mathbf{z})\subseteq\mathbf{E}^N$ [see equations (92) and (67)][c]
   (i) $\mathbf{T}(\mathbf{z})$ is non-empty, convex, closed, and semi-bounded for each $\mathbf{z}\in\mathbf{Z}$
   (ii) If $\mathbf{z}^k\rightarrow\mathbf{z}^0$, $\mathbf{x}^k\in\mathbf{T}(\mathbf{z}^k)$, $\mathbf{x}^k\rightarrow\mathbf{x}^0$, then $\mathbf{z}^0\in\mathbf{Z}$ and $\mathbf{x}^0\in\mathbf{T}(\mathbf{z}^0)$
   (iii) $\{(\mathbf{z},\mathbf{x})\in\mathbf{Z}\times\mathbf{E}^N|\mathbf{x}\in\mathbf{T}(\mathbf{z})\}$ is convex and semi-bounded
   (iv) $\bar{\mathbf{x}}\in\mathbf{T}(\bar{\mathbf{z}})$ and $\mathbf{x}^*(\mathbf{z})\in\mathbf{T}(\mathbf{z})$

6. The z-level price possibility set $\mathbf{R}(\mathbf{z})\subseteq\mathbf{E}^N$ [see equations (103) and (68)][d]
   (i) $\mathbf{R}(\mathbf{z})$ is convex and closed, with a non-empty interior, for each $\mathbf{z}\in\mathbf{Z}$
   (ii) $\{(\mathbf{z},\mathbf{x})\in\mathbf{Z}\times\mathbf{E}^N|(\forall\mathbf{p}\in\mathbf{R}(\mathbf{z}))\mathbf{p}\cdot(\mathbf{x}-\mathbf{x}^*(\mathbf{z}))\leqq 1\}$ is convex, closed, and semi-bounded
   (iii) For all $\mathbf{p}\in\mathbf{R}(\bar{\mathbf{z}})$, $\mathbf{p}\cdot(\bar{\mathbf{x}}-\mathbf{x}^*(\bar{\mathbf{z}}))\leqq 1$

7. The z-level profit function $\Pi:\mathbf{Z}\times\mathbf{E}^N\rightarrow\bar{\mathbf{E}}$ [see equations (119) and (69)][a,e]
   (i) For each $\mathbf{z}\in\mathbf{Z}$, Π is convex, closed, and positively linear homogeneous in $\mathbf{p}$
   (ii) $\mathbf{P}(\mathbf{z})=\{\mathbf{p}\in\mathbf{E}^N|\Pi(\mathbf{z},\mathbf{p})<+\infty\}$ is a convex cone with a non-empty interior for each $\mathbf{z}\in\mathbf{Z}$
   (iii) For each $\mathbf{p}\in\mathbf{E}^N$, Π is concave and closed in $\mathbf{z}$
   (iv) $\{(\mathbf{z},\mathbf{x})\in\mathbf{Z}\times\mathbf{E}^N|(\forall\mathbf{p}\in\mathbf{P}(\mathbf{z}))\mathbf{p}\cdot\mathbf{x}\leqq\Pi(\mathbf{z},\mathbf{p})\}$ is convex, closed, and semi-bounded
   (v) For all $\mathbf{p}\in\mathbf{P}(\bar{\mathbf{z}})$, $\mathbf{p}\cdot\bar{\mathbf{x}}\leqq\Pi(\bar{\mathbf{z}},\mathbf{p})$ and for $\mathbf{p}\in\mathbf{P}(\mathbf{z})$, $\mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})\leqq\Pi(\mathbf{z},\mathbf{p})$.

8. The total profit function $\Pi^*:\mathbf{E}^M\times\mathbf{E}^N\rightarrow\bar{\mathbf{E}}$ [see equations (126) and (70)][a]
   (i) Π^* is convex, closed, and positively linear homogeneous
   (ii) $\{(\mathbf{q},\mathbf{p})\in\mathbf{E}^M\times\mathbf{E}^N|\Pi^*(\mathbf{q},\mathbf{p})<+\infty\}$ is a convex cone with a non-empty interior
   (iii) $\Pi^*(\mathbf{q},\mathbf{p})\geqq\mathbf{q}\cdot\bar{\mathbf{x}}+\mathbf{p}\cdot\bar{\mathbf{x}}$ for all $(\mathbf{q},\mathbf{p})\mathbf{E}^M\times\mathbf{E}^N$

9. The total price possibility set $\mathbf{R}^*$ [see equations (134) and (71)]
   (i) $\mathbf{R}^*$ is convex and closed, with a non-empty interior
   (ii) $(\mathbf{0},\mathbf{0})\in\mathbf{R}^*$

[a] $\bar{\mathbf{E}}=[-\infty,+\infty]$.

[b] The x-level gauge function $\mathbf{G}:\mathbf{E}^M\times\mathbf{X}\rightarrow\bar{\mathbf{E}}$ also has these properties with z and x interchanged.

[c] The x-level production possibility set $\mathbf{U}(\mathbf{x})\subseteq\mathbf{E}^M$, $\mathbf{x}\in\mathbf{X}$, also has these properties with z and x interchanged.

[d] The x-level price possibility set $\mathbf{S}(\mathbf{x})\subseteq\mathbf{E}^M$, $\mathbf{x}\in\mathbf{X}$, also has these properties with z and x interchanged.

[e] The x-level profit function $\Psi:\mathbf{E}^M\times\mathbf{X}\rightarrow\bar{\mathbf{E}}$ also has these properties with z and x, p and q interchanged.

[f] When $\mathbf{x}^*(\mathbf{z})$ is constant, say $\mathbf{x}^*(\mathbf{z})=\bar{\mathbf{x}}$, condition (iv) can be replaced by the requirement that $H(\mathbf{z},\mathbf{x})$ be quasi-convex, with $\{(\mathbf{z},\mathbf{x})\in\mathbf{Z}\times\mathbf{E}^N|H(\mathbf{z},\mathbf{x})\leqq 1\}$ closed.

TABLE 7
Duality mappings.

1. Mappings yielding the total production possibility set $\mathbf{Y} \subseteq \mathbf{E}^M \times \mathbf{E}^N$

$$\mathbf{Y} = \{(\mathbf{z},\mathbf{x}) \in \mathbf{E}^M \times \mathbf{E}^N | H^*(\mathbf{z},\mathbf{x}) \leqq 1\} \quad (65)$$
$$= \{(\mathbf{z},\mathbf{x}) \in \mathbf{Z} \times \mathbf{E}^N | H(\mathbf{z},\mathbf{x}) \leqq 1\} \quad (66)^b$$
$$= \{(\mathbf{z},\mathbf{x}) \in \mathbf{Z} \times \mathbf{E}^N | \mathbf{x} \in \mathbf{T}(\mathbf{z})\} \quad (67)^b$$
$$= \{(\mathbf{z},\mathbf{x}) \in \mathbf{Z} \times \mathbf{E}^N | (\forall \mathbf{p} \in \mathbf{R}(\mathbf{z}))\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\} \quad (68)^b$$
$$= \{(\mathbf{z},\mathbf{x}) \in \mathbf{Z} \times \mathbf{E}^N | (\forall \mathbf{p} \in \mathbf{R}^N)\mathbf{p}\cdot\mathbf{x} \leqq \Pi(\mathbf{z},\mathbf{p})\} \quad (69)^b$$
$$= \{(\mathbf{z},\mathbf{x}) \in \mathbf{E}^M \times \mathbf{E}^N | (\forall \mathbf{p} \in \mathbf{E}^N)(\forall \mathbf{q} \in \mathbf{E}^M)\mathbf{q}\cdot\mathbf{z} + \mathbf{p}\cdot\mathbf{x} \leqq \Pi^*(\mathbf{q},\mathbf{p})\} \quad (70)$$
$$= \{(\mathbf{z},\mathbf{x}) \in \mathbf{E}^M \times \mathbf{E}^N | (\forall (\mathbf{q},\mathbf{p}) \in \mathbf{R}^*)\mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq 1\} \quad (71)$$

2. Mappings yielding the total gauge function $H^*{:}\mathbf{E}^M \times \mathbf{E}^N \to \bar{\mathbf{E}}$[a]

$$H^*(\mathbf{z},\mathbf{x}) = \inf\left\{\lambda > 0 | (\bar{\mathbf{z}},\bar{\mathbf{x}}) + \frac{1}{\lambda}(\mathbf{z} - \bar{\mathbf{z}},\mathbf{x} - \bar{\mathbf{x}}) \in \mathbf{Y}\right\} \quad (72)$$
$$= \inf\left\{\lambda > 0 | \bar{\mathbf{z}} + \frac{1}{\lambda}(\mathbf{z} - \bar{\mathbf{z}}) \in \mathbf{Z} \;\&\; \mathbf{H}\left(\bar{\mathbf{z}} + \frac{1}{\lambda}(\mathbf{z} - \bar{\mathbf{z}}),\bar{\mathbf{x}} + \frac{1}{\lambda}(\mathbf{x} - \bar{\mathbf{x}})\right) \leqq 1\right\} \quad (73)^b$$
$$= \inf\left\{\lambda > 0 | \bar{\mathbf{z}} + \frac{1}{\lambda}(\mathbf{z} - \bar{\mathbf{z}}) \in \mathbf{Z} \;\&\; \bar{\mathbf{x}} + \frac{1}{\lambda}(\mathbf{x} - \bar{\mathbf{x}}) \in \mathbf{T}\left(\bar{\mathbf{z}} + \frac{1}{\lambda}(\mathbf{z} - \bar{\mathbf{z}})\right)\right\} \quad (74)^b$$
$$= \inf\left\{\lambda > 0 | \mathbf{z}' \equiv \bar{\mathbf{z}} + \frac{1}{\lambda}(\mathbf{z} - \bar{\mathbf{z}}) \in \mathbf{Z} \;\&\; (\forall \mathbf{p} \in \mathbf{R}(\mathbf{z}'))\mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq \lambda[1 + \mathbf{p}\cdot(\mathbf{x}^*(\mathbf{z}') - \bar{\mathbf{x}})]\right\} \quad (75)^b$$
$$= \inf\left\{\lambda > 0 | (\forall \mathbf{p} \in \mathbf{E}^N)\mathbf{z}' \equiv \bar{\mathbf{z}} + \frac{1}{\lambda}(\mathbf{z} - \bar{\mathbf{z}}) \in \mathbf{Z} \;\&\; \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq \lambda[\Pi(\mathbf{z}',\mathbf{p}) - \mathbf{p}\cdot\bar{\mathbf{x}}]\right\} \quad (76)^b$$
$$= \inf\{\lambda > 0 | (\forall \mathbf{q} \in \mathbf{E}^M)(\forall \mathbf{p} \in \mathbf{E}^N)\mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) \leqq \lambda[\Pi^*(\mathbf{q},\mathbf{p}) - \mathbf{q}\cdot\bar{\mathbf{z}} - \mathbf{p}\cdot\bar{\mathbf{x}}]\} \quad (77)$$
$$= \sup\{\mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) | \Pi^*(\mathbf{q},\mathbf{p}) - \mathbf{q}\cdot\bar{\mathbf{z}} - \mathbf{p}\cdot\bar{\mathbf{x}} \leqq 1\} \quad (78)$$
$$= \sup\{\mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) | (\mathbf{q},\mathbf{p}) \in \mathbf{R}^*\} \quad (79)$$

3. Mappings yielding the z-level gauge function $H{:}\mathbf{Z} \times \mathbf{E}^N \to \bar{\mathbf{E}}$[a,b]

$$H(\mathbf{z},\mathbf{x}) = \inf\left\{\lambda > 0 | \left(\mathbf{z},\mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))\right) \in \mathbf{Y}\right\} \quad (80)$$

TABLE 7 (*continued*)

$$= \inf\left\{\lambda > 0 \middle| H^*\left(\mathbf{z},\mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))\right) \leqq 1\right\} \quad (81)$$

$$= \inf\left\{\lambda > 0 \middle| \mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \in \mathbf{T}(\mathbf{z})\right\} \quad (82)$$

$$= \sup\{\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) | \mathbf{p} \in \mathbf{R}(\mathbf{z})\} \quad (83)$$

$$= \inf\{\lambda > 0 | (\forall \mathbf{p} \in \mathbf{E}^N)\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq \lambda[\Pi(\mathbf{z},\mathbf{p}) - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})]\} \quad (84)$$

$$= \sup\{\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) | \Pi(\mathbf{z},\mathbf{p}) - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) \leqq 1\} \quad (85)$$

$$= \inf\{\lambda > 0 | (\forall \mathbf{p} \in \mathbf{E}^N)(\forall \mathbf{q} \in \mathbf{E}^M)\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq \lambda[\Pi^*(\mathbf{q},\mathbf{p}) - \mathbf{q}\cdot\mathbf{z} - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})]\} \quad (86)$$

$$= \inf\left\{\lambda > 0 \middle| (\forall (\mathbf{q},\mathbf{p}) \in \mathbf{R}^*)\mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x}^*(\mathbf{z}) - \bar{\mathbf{x}}) + \frac{1}{\lambda}\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\right\} \quad (87)$$

$$= \inf\left\{\lambda > 0 \middle| \mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \in \mathbf{X} \;\&\; \mathbf{G}\left(\mathbf{z},\mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))\right) \leqq 1\right\} \quad (88)$$

$$= \inf\left\{\lambda > 0 \middle| \mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \in \mathbf{X} \;\&\; \mathbf{z} \in \mathbf{U}\left(\mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))\right)\right\} \quad (89)$$

$$= \inf\left\{\lambda > 0 \middle| \mathbf{x}' \equiv \mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \in \mathbf{X} \;\&\; (\forall \mathbf{q} \in \mathbf{S}(\mathbf{x}'))\mathbf{q}\cdot(\mathbf{z} - \mathbf{z}^*(\mathbf{x}')) \leqq 1\right\} \quad (90)$$

$$= \inf\left\{\lambda > 0 \middle| \mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \in \mathbf{X} \;\&\; (\forall \mathbf{q} \in \mathbf{R}^M)\mathbf{q}\cdot\mathbf{z} \leqq \Psi\left(\mathbf{q},\mathbf{x}^*(\mathbf{z}) + \frac{1}{\lambda}(\mathbf{x} - \mathbf{x}^*(\mathbf{z}))\right)\right\} \quad (91)$$

4. Mappings yielding the z-level production possibility set $\mathbf{T}(\mathbf{z}) \in \mathbf{E}^{N}$[b]

$$\mathbf{T}(\mathbf{z}) = \{\mathbf{x} \in \mathbf{E}^N | (\mathbf{z},\mathbf{x}) \in \mathbf{Y}\} \quad (92)$$

$$= \{\mathbf{x} \in \mathbf{E}^N | H^*(\mathbf{z},\mathbf{x}) \leqq 1\} \quad (93)$$

$$= \{\mathbf{x} \in \mathbf{E}^N | H(\mathbf{z},\mathbf{x}) \leqq 1\} \quad (94)$$

$$= \{\mathbf{x} \in \mathbf{E}^N | (\forall \mathbf{p} \in \mathbf{R}(\mathbf{z}))\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\} \quad (95)$$

$$= \{\mathbf{x} \in \mathbf{E}^N | (\forall \mathbf{p} \in \mathbf{E}^N)\mathbf{p}\cdot\mathbf{x} \leqq \Pi(\mathbf{z},\mathbf{p})\} \quad (96)$$

$$= \{\mathbf{x} \in \mathbf{E}^N | (\forall \mathbf{p} \in \mathbf{E}^N)(\forall \mathbf{q} \in \mathbf{E}^M)\mathbf{p}\cdot\mathbf{x} + \mathbf{q}\cdot\mathbf{z} \leqq \Pi^*(\mathbf{q},\mathbf{p})\} \quad (97)$$

$$= \{\mathbf{x} \in \mathbf{E}^N | (\forall (\mathbf{q},\mathbf{p}) \in \mathbf{R}^*)\mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) \leqq 1\} \quad (98)$$

TABLE 7 (*continued*)

$= \{\mathbf{x} \in \mathbf{X} \mid G(\mathbf{z},\mathbf{x}) \leqq 1\}$	(99)
$= \{\mathbf{x} \in \mathbf{X} \mid \mathbf{z} \in \mathbf{U}(\mathbf{x})\}$	(100)
$= \{\mathbf{x} \in \mathbf{X} \mid (\forall \mathbf{q} \in \mathbf{S}(\mathbf{x}))\mathbf{q}\cdot(\mathbf{z} - \mathbf{z}^*(\mathbf{x})) \leqq 1\}$	(101)
$= \{\mathbf{x} \in \mathbf{X} \mid (\forall \mathbf{q} \in \mathbf{E}^M)\mathbf{q}\cdot\mathbf{z} \leqq \Psi(\mathbf{q},\mathbf{x})\}$	(102)

5. Mappings yielding the z-level price possibility set $\mathbf{R}(\mathbf{z}) \subseteq \mathbf{E}^{N}$ [b]

$\mathbf{R}(\mathbf{z}) = \{\mathbf{p} \in \mathbf{E}^N \mid (\forall (\mathbf{z},\mathbf{x}) \in \mathbf{Y})\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\}$	(103)
$= \{\mathbf{p} \in \mathbf{E}^N \mid (\forall \mathbf{x} \in \mathbf{E}^N)H^*(\mathbf{z},\mathbf{x}) \leqq 1 \Rightarrow \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\}$	(104)
$= \{\mathbf{p} \in \mathbf{E}^N \mid (\forall \mathbf{x} \in \mathbf{E}^N)\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq H(\mathbf{z},\mathbf{x})\}$	(105)
$= \{\mathbf{p} \in \mathbf{E}^N \mid (\forall \mathbf{x} \in \mathbf{T}(\mathbf{z}))\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\}$	(106)
$= \{\mathbf{p} \in \mathbf{E}^N \mid \Pi(\mathbf{z},\mathbf{p}) - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) \leqq 1\}$	(107)
$= \{\mathbf{p} \in \mathbf{E}^N \mid \inf_{\mathbf{q}\in\mathbf{E}^M} [\Pi^*(\mathbf{q},\mathbf{p}) - \mathbf{q}\cdot\mathbf{z} - \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z})] \leqq 1\}$	(108)
$= \{\mathbf{p} \in \mathbf{E}^N \mid \inf_{\mathbf{q}\in\mathbf{E}^M} \inf_{\lambda>0\ni(\mathbf{q},\mathbf{p})/\lambda\in\mathbf{R}^*} [\lambda - \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}})$ $- \mathbf{p}\cdot(\mathbf{x}^*(\mathbf{z}) - \bar{\mathbf{x}})] \leqq 1\}$	(109)
$= \{\mathbf{p} \in \mathbf{E}^N \mid (\forall \mathbf{x} \in \mathbf{X})G(\mathbf{z},\mathbf{x}) \leqq 1 \Rightarrow \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\}$	(110)
$= \{\mathbf{p} \in \mathbf{E}^N \mid (\forall \mathbf{x} \in \mathbf{X})\mathbf{z} \in \mathbf{U}(\mathbf{x}) \Rightarrow \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\}$	(111)
$= \{\mathbf{p} \in \mathbf{E}^N \mid (\forall \mathbf{x} \in \mathbf{X})[(\forall \mathbf{q} \in \mathbf{S}(\mathbf{x}))\mathbf{q}\cdot(\mathbf{z} - \mathbf{z}^*(\mathbf{x}))$ $\leqq 1] \Rightarrow \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq 1\}$	(112)
$= \{\mathbf{p} \in \mathbf{E}^N \mid \inf_{\mathbf{q}\in\mathbf{E}^M} \sup_{\mathbf{x}\in\mathbf{X}} [\Psi(\mathbf{q},\mathbf{x}) + \mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) - \mathbf{q}\cdot\mathbf{z}] \leqq 1\}$	(113)

6. Mappings yielding the z-level profit function $\Pi:\mathbf{Z} \times \mathbf{E}^N \to \bar{\mathbf{E}}$ [a,b]

$\Pi(\mathbf{z},\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{x} \mid (\mathbf{z},\mathbf{x}) \in \mathbf{Y}\}$	(114)
$= \sup\{\mathbf{p}\cdot\mathbf{x} \mid H^*(\mathbf{z},\mathbf{x}) \leqq 1\}$	(115)
$= \sup\{\mathbf{p}\cdot\mathbf{x}\ H(\mathbf{z},\mathbf{x}) \leqq 1\}$	(116)
$= \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \inf\{\lambda > 0 \mid (\forall \mathbf{x} \in \mathbf{E}^N)\mathbf{p}\cdot(\mathbf{x} - \mathbf{x}^*(\mathbf{z})) \leqq \lambda H(\mathbf{z},\mathbf{x})\}$	(117)
$= \sup\{\mathbf{p}\cdot\mathbf{x} \mid H(\mathbf{z},\mathbf{x}) \leqq 1\}$	(118)
$= \mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \inf\left\{\lambda > 0 \mid \frac{1}{\lambda}\mathbf{p} \in \mathbf{R}(\mathbf{z})\right\}$	(119)
$= \inf\{\Pi^*(\mathbf{q},\mathbf{p}) - \mathbf{q}\cdot\mathbf{z} \mid \mathbf{q} \in \mathbf{E}^M\}$	(120)
$= \inf_{\mathbf{q}\in\mathbf{E}^M}\left[\mathbf{p}\cdot\bar{\mathbf{x}} - \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \inf\left\{\lambda > 0 \mid \frac{1}{\lambda}(\mathbf{q},\mathbf{p}) \in \mathbf{R}^*\right\}\right]$	(121)
$= \sup\{\mathbf{p}\cdot\mathbf{x} \mid \mathbf{x} \in \mathbf{X}\ \&\ G(\mathbf{z},\mathbf{x}) \leqq 1\}$	(122)
$= \sup\{\mathbf{p}\cdot\mathbf{x} \mid \mathbf{x} \in \mathbf{X}\ \&\ \mathbf{z} \in \mathbf{U}(\mathbf{x})\}$	(123)
$= \sup\{\mathbf{p}\cdot\mathbf{x} \mid \mathbf{x} \in \mathbf{X}\ \&\ (\forall \mathbf{q} \in \mathbf{S}(\mathbf{x}))\mathbf{q}\cdot(\mathbf{z} - \mathbf{z}^*(\mathbf{x})) \leqq 1\}$	(124)
$= \inf_{\mathbf{q}\in\mathbf{E}^M} \sup_{\mathbf{x}\in\mathbf{X}} [\Psi(\mathbf{q},\mathbf{x}) + \mathbf{p}\cdot\mathbf{x} - \mathbf{q}\cdot\mathbf{z}]$	(125)

TABLE 7 (*continued*)

7. Mappings yielding the total profit function $\Pi^*:\mathbf{E}^M \times \mathbf{E}^N \to \bar{\mathbf{E}}^a$

$$\Pi^*(\mathbf{q},\mathbf{p}) = \sup\{\mathbf{q}\cdot\mathbf{z} + \mathbf{p}\cdot\mathbf{x} | (\mathbf{z},\mathbf{x}) \in \mathbf{Y}\} \quad (126)$$
$$= \sup\{\mathbf{q}\cdot\mathbf{z} + \mathbf{p}\cdot\mathbf{x} | H^*(\mathbf{z},\mathbf{x}) \leqq 1\} \quad (127)$$
$$= \mathbf{q}\cdot\bar{\mathbf{z}} + \mathbf{p}\cdot\bar{\mathbf{x}} + \inf\{\lambda > 0 | (\forall \mathbf{z} \in \mathbf{E}^M)(\forall \mathbf{x} \in \mathbf{E}^N)\mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq \lambda H^*(\mathbf{z},\mathbf{x})\} \quad (128)$$
$$= \sup\{\mathbf{q}\cdot\mathbf{z} + \mathbf{p}\cdot\mathbf{x} | \mathbf{z} \in \mathbf{Z} \ \&\ H(\mathbf{z},\mathbf{x}) \leqq 1\} \quad (129)^b$$
$$= \sup\{\mathbf{q}\cdot\mathbf{z} + \mathbf{p}\cdot\mathbf{x} | \mathbf{z} \in \mathbf{Z} \ \&\ \mathbf{x} \in \mathbf{T}(\mathbf{z})\} \quad (130)^b$$
$$= \sup_{\mathbf{z}\in\mathbf{Z}}\left[\mathbf{p}\cdot\mathbf{x}^*(\mathbf{z}) + \mathbf{q}\cdot\mathbf{z} + \inf\left\{\lambda > 0 | \frac{1}{\lambda}\mathbf{p} \in \mathbf{R}(\mathbf{z})\right\}\right] \quad (131)^b$$
$$= \sup_{\mathbf{z}\in\mathbf{Z}}[\Pi(\mathbf{z},\mathbf{p}) + \mathbf{q}\cdot\mathbf{z}] \quad (132)^b$$
$$= \mathbf{q}\cdot\bar{\mathbf{z}} + \mathbf{p}\cdot\bar{\mathbf{x}} + \inf\left\{\lambda > 0 | \frac{1}{\lambda}(\mathbf{q},\mathbf{p}) \in \mathbf{R}^*\right\} \quad (133)$$

8. Mappings yielding the total price possibility set $\mathbf{R}^* \subseteq \mathbf{E}^M \times \mathbf{E}^N$

$$\mathbf{R}^* = \{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | (\forall (\mathbf{z},\mathbf{x}) \in \mathbf{Y})\mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq 1\} \quad (134)$$
$$= \{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | (\forall \mathbf{z} \in \mathbf{R}^M)(\forall \mathbf{x} \in \mathbf{R}^N)\mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq H^*(\mathbf{z},\mathbf{x})\} \quad (135)$$
$$= \{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | \mathbf{z} \in \mathbf{Z} \ \&\ H(\mathbf{z},\mathbf{x}) \leqq 1 \Rightarrow \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq 1\} \quad (136)^b$$
$$= \{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | \mathbf{z} \in \mathbf{Z} \ \&\ \mathbf{x} \in \mathbf{T}(\mathbf{z}) \Rightarrow \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \mathbf{p}\cdot(\mathbf{x} - \bar{\mathbf{x}}) \leqq 1\} \quad (137)^b$$
$$= \left\{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | (\forall \mathbf{z} \in \mathbf{Z})\mathbf{p}\cdot(\mathbf{x}^*(\mathbf{z}) - \bar{\mathbf{x}}) + \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) + \inf\left\{\lambda > 0 | \frac{1}{\lambda}\mathbf{p} \in \mathbf{R}(\mathbf{z})\right\} \leqq 1\right\} \quad (138)^b$$
$$= \{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | (\forall \mathbf{z} \in \mathbf{Z})\Pi(\mathbf{z},\mathbf{p}) + \mathbf{q}\cdot(\mathbf{z} - \bar{\mathbf{z}}) - \mathbf{p}\cdot\bar{\mathbf{x}} \leqq 1\} \quad (139)^b$$
$$= \{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | (\Pi^*(\mathbf{q},\mathbf{p}) - \mathbf{q}\cdot\bar{\mathbf{z}} - \mathbf{p}\cdot\bar{\mathbf{x}} \leqq 1\} \quad (140)$$

[a] $\bar{\mathbf{E}} = [-\infty, +\infty]$.

[b] A second class of mappings is obtained for x-level functions by making the substitutions

$$\mathbf{z} \leftrightarrow \mathbf{x} \qquad \mathbf{p} \leftrightarrow \mathbf{q}$$
$$H \leftrightarrow G \qquad T \leftrightarrow U$$
$$\Pi \leftrightarrow \Psi \qquad R \leftrightarrow S$$

Proof: Many of the conclusions in this theorem summarize results established earlier in the chapter, or can be deduced as simple corollaries. We provide a broad outline of the argument, leaving details to the interested reader.

Consider first the class of production possibility sets Y satisfying the conditions of Table 6, and define the total gauge function H^*, the total profit function Π^*, and the total price possibility set R^* by equations (72), (126), and (134), respectively. The remaining mappings between these functions are given in equations (65), (70), (71), (77), (79), (128), (133), (135), and (140). The properties of these mappings follow from Theorem 24 and its corollary, using the following substitution of notation:

Theorem 24 ⟶	*Theorem* 29
$\mathbf{E}^N$	$\mathbf{E}^M \times \mathbf{E}^N$
$\mathbf{Z}$	none
$\mathbf{x}$	$(\mathbf{z},\mathbf{x})$
$\mathbf{T}(\mathbf{z})$	$\mathbf{Y}$
H	H^*
$\mathbf{p}$	$(\mathbf{q},\mathbf{p})$
Π	Π^*
$\mathbf{R}$	$\mathbf{R}^*$

The price possibility set is not treated explicitly in Theorem 24; however, its properties are an immediate corollary of the definition $\mathbf{R}^* = \{(\mathbf{q},\mathbf{p}) \in \mathbf{E}^M \times \mathbf{E}^N | \Pi^*(\mathbf{q},\mathbf{p}) - \mathbf{q}\cdot\bar{\mathbf{z}} - \mathbf{p}\cdot\bar{\mathbf{x}} \leqq 1\}$ and the properties of Π^*.

Consider next the z-level production possibility set $\mathbf{T}(\mathbf{z})$, and the z-level gauge function $H(\mathbf{z},\mathbf{x})$, profit function $\Pi(\mathbf{z},\mathbf{p})$, and price possibility set $\mathbf{R}'(\mathbf{z})$ defined by equations (82), (118), and (106), respectively. The remaining mappings between the functions are given in equations (83), (84), (94), (95), (96), (105), (107), (117), and (119). The properties of these mappings are a direct restatement of Theorem 24 and its corollary.

The properties and relations of the x-level gauge function $G(\mathbf{z},\mathbf{x})$, production possibility set $\mathbf{U}(\mathbf{x})$, price possibility set $\mathbf{S}(\mathbf{x})$, and profit function $\Psi(\mathbf{q},\mathbf{x})$ can be deduced from their formal duality to the functions H, $\mathbf{T}$, $\mathbf{R}$, and Π:

Primal ⟷	*Dual*
$\mathbf{x}$	$\mathbf{z}$
$\mathbf{p}$	$\mathbf{q}$
$\mathbf{T}(\mathbf{z})$	$\mathbf{U}(\mathbf{x})$
$H(\mathbf{z},\mathbf{x})$	$G(\mathbf{z},\mathbf{x})$
$\Pi(\mathbf{z},\mathbf{p})$	$\Psi(\mathbf{q},\mathbf{x})$
$\mathbf{R}(\mathbf{z})$	$\mathbf{S}(\mathbf{x})$

The mapping in equation (92) from **Y** to **T**(**z**) and conversely in (67) from **T**(**z**) to **Y** can immediately be seen to be mutual inverses, and to be one-to-one onto for the classes of **Y** and **T**(**z**) defined in Table 6. A dual relation holds between the classes of **Y** and **U**(**x**).

The conclusions above provide a chain of mappings (via **Y**) between any two classes in Table 6 which have the properties claimed in the theorem. The remaining mappings in Table 7 are compositions of these chains.

For example, property (iii) of the **z**-level profit function, concavity and closure in **z**, can be deduced from the properties of the total profit function Π^* and the concave conjugate dual mapping (120); see Appendix A.3, Theorem 12.3. Verification of the formulae for these composite mappings is tedious, but straightforward, and is left to the reader. Q.E.D.

Several of the mappings in Table 7 deserve note. Equation (125) establishes that the **z**-level and **x**-level profit functions are conjugate saddle functions [see Rockafellar (1970, section 37)]. Equations (120) and (132) give the relation between short and long run profit functions. Equations (83) and (85) give simple dual mappings between the **z**-level gauge function and the profit function and price possibility set.

This chapter has set out the basic theory of duality in production economics, and developed the mathematical properties of dual functions. The remaining chapters of this book demonstrate the use of these methods in theoretical and empirical analysis.

Chapter I.2

SYMMETRIC DUALITY AND POLAR PRODUCTION FUNCTIONS

GIORA HANOCH*

The Hebrew University of Jerusalem

1. Introduction

The Shephard–Uzawa–McFadden duality theorems[1] relating production functions to cost and profit functions, may be utilized to generate new valid functional forms for production functions and production frontiers, or equivalently, new valid cost and profit functions. To any given standard production relation, namely one which satisfies the conditions for existence and uniqueness of the dual, there corresponds at least one other standard production relation, which satisfies the same requirements, but may exhibit rather different specific patterns. This process of getting "two for the price of one" in the search for useful functional forms is made possible by reformulating the duality relations in a perfectly symmetric way. The process has been applied before to production as well as to consumer demand theory – although it seems to have never been formalized and recognized as a generally valid procedure.

*This article constitutes a revision of a part of an earlier paper, "Generation of New Production Functions Through Duality", Discussion Paper No. 118, Harvard Institute of Economic Research, April 1970. I am thankful to Zvi Griliches, who encouraged and supported this research, and I have benefited from discussions with Kenneth Arrow, Erwin Diewert, Melvyn Fuss, Dale Jorgenson, Lawrence Lau, Daniel McFadden, Michael Rothschild, and Christopher Sims. I am indebted to the National Science Foundation for financial assistance (Grant No. 2762X), and to Harvard University, where I visited in 1969–70 and in 1973–74, while on leave from the Hebrew University, Jerusalem.

[1]See Shephard (1953), Uzawa (1964), and McFadden (Chapter I.1). More detailed presentations and modified proofs are in Diewert (1971, 1973a).

As will be clarified below, the specific formulation used in the established analysis of duality, self-duality, and related topics in utility theory,[2] has unfortunately hindered the development of analogous results in production. In particular, the choice of a utility-index representation – which is arbitrary in the context of an ordinal utility function – could not be carried over as it stood to cardinal production functions, where output is measurable and non-negative.

Heuristically, given any standard production, cost, or profit function, a standard *polar* cost, profit, or production function is obtained by a transformation from the variable quantities (prices) space into the corresponding prices (quantities) space, using the functional form of the dual relation for the new, polar, primal relation. Moreover, the fixed quantity, such as output in cost minimization analysis, is transformed into its reciprocal. In order to show this more rigorously, however, it is necessary to modify the formulation of the duality relations, so as to get a perfect symmetry between the primal and the dual – with exactly the same type of restrictions on sets and on functions appearing on both sides. This is done for cost functions in Section 2, and for profit functions and joint-production frontiers in Section 4.

Section 3 establishes the existence and uniqueness of the polar production and cost functions, as well as some special modifications for homothetic production functions and for separable production frontiers. In Section 5 similar results are stated and proved with respect to profit functions and joint production frontiers. Section 6 discusses some extensions and an application to various definitions of elasticities of substitution. Examples of specific functional forms generated by this approach are given in Chapter II.3.

2. A Symmetric Formulation of Cost and Production Functions

Suppose $y = f(\mathbf{x})$ is a standard production function, satisfying the following regularity conditions[3] for existence of a unique dual cost function $C = G(y;\mathbf{p})$:

Condition I. $f(\mathbf{x})$ is defined for all $\mathbf{x} = \{x_1,...,x_n\} \geq \mathbf{0}$ $(\mathbf{x} \in \bar{\Omega}_n)$, and is real, single-valued, right-continuous, non-decreasing in $\mathbf{x}$, quasi-

[2]E.g., in Houthakker (1965), Samuelson (1947, 1965a), and Lau (1969a).
[3]For a specification of these conditions and a proof, see Diewert (1971).

concave, finite for finite $\mathbf{x}$, and unbounded for at least some unbounded sequence $\{\mathbf{x}^N\}$, with $f(\mathbf{0}) = 0$.

For any non-negative output y, the production possibilities set $\mathbf{L}(y) \subseteq \bar{\Omega}_n$ is defined as

$$\mathbf{L}(y) = \{\mathbf{x}: f(\mathbf{x}) \geq y\}, \tag{1}$$

and satisfies the following:

Condition II.[4] For $y \geq 0$, $\mathbf{L}(y)$ is a non-empty, closed, convex set, with free disposal:

$$\mathbf{x}' > \mathbf{x} \in \mathbf{L}(y) \Rightarrow \mathbf{x}' \in \mathbf{L}(y); \quad y' > y \Rightarrow \mathbf{L}(y') \subseteq \mathbf{L}(y),$$

where for all $\mathbf{x}$ there exists a $y' > 0$ such that $\mathbf{x} \notin \mathbf{L}(y')$ [$\mathbf{x}$ is not in $\mathbf{L}(y')$]; and $\mathbf{L}(0) = \bar{\mathbf{\Omega}}_n$. If $y > 0$, then $\mathbf{0} \notin \mathbf{L}(y)$. The set $\{(y,\mathbf{x}):\mathbf{x} \in \mathbf{L}(y)\}$ (the graph of $\mathbf{L}$) is closed.

Given a positive output y, it is always possible to represent uniquely (for strictly positive vectors $\mathbf{x} \gg \mathbf{0}$) the standard production set $\mathbf{L}(y)$, and the production function equation $f(\mathbf{x}) = y$, by a normalized equation of the form: $D(1/y;\mathbf{x}) = 1$, such that the "distance function" $D(1/y;\mathbf{x})$ behaves with respect to its arguments $(1/y;\mathbf{x})$ exactly in the same manner as a standard cost function with respect to $(y;\mathbf{p})$. This statement is now formalized and proved. Define the distance function[5] as follows (where $\mathbf{\Omega}_n$ is the positive orthant, $\mathbf{\Omega}_n = \{\mathbf{x}:\mathbf{x} \gg \mathbf{0}\}$):

$$\begin{aligned} D(1/y;\mathbf{x}) &= \sup\{d:(1/d)\mathbf{x} \in \mathbf{L}(y);\mathbf{x} \in \mathbf{\Omega}_n\} \\ &= \sup\{d:f((1/d)\mathbf{x}) \geq y;\mathbf{x} \in \mathbf{\Omega}_n\}, \end{aligned} \tag{2}$$

by equation (1).

Theorem 1. If $\mathbf{L}(y)$ defined in equation (1) satisfies the conditions on standard production possibilities sets (Condition II), the function $D(1/y;\mathbf{x})$ defined in equation (2) satisfies Condition III below. The set $\mathbf{L}^*(y) = \{\mathbf{x}:D(1/y;\mathbf{x}) \geq 1;\mathbf{x} \in \mathbf{\Omega}_n\}$ coincides with the set $\mathbf{L}(y)$ for $\mathbf{x} \gg \mathbf{0}: \mathbf{L}^*(y) \equiv \mathbf{L}(y) \cap \mathbf{\Omega}_n$.

[4]These are Condition II(2.7) in Diewert (1971), with slight modifications. The notation adopted here is: $\mathbf{x}' \geq \mathbf{x}$ means $x_i' \geq x_i$ (all i); $\mathbf{x}' > \mathbf{x}$ means $\mathbf{x}' \geq \mathbf{x}$ and $\mathbf{x}' \neq \mathbf{x}$; $\mathbf{x}' \gg \mathbf{x}$ means $x_i' > x_i$ (all i). $\mathbf{x}^t$ is the transpose of $\mathbf{x}$.

[5]The distance function was first used for isoquants and unit cost functions (for the differentiable case) by Shephard (1953, p. 6). However, Shephard did not show the symmetry with respect to y, $1/y$, respectively, of the distance functions.

Condition III.[6] The function $D(1/y;\mathbf{x})$ is

(a) positive, real-valued, defined, and finite for all finite $\mathbf{x} \gg \mathbf{0}$, $1/y > 0$;
(b) non-decreasing in $1/y$ and unbounded if $1/y$ is unbounded;
(c) non-decreasing in $\mathbf{x}$;
(d) positive linear homogeneous in $\mathbf{x}$, for all finite $1/y > 0$; i.e., if $\lambda > 0$, $1/y > 0$ and finite, and $\mathbf{x} \gg \mathbf{0}$, then $D(1/y;\lambda\mathbf{x}) = \lambda D(1/y;\mathbf{x})$;
(e) concave in $\mathbf{x}$, for finite $1/y > 0$.
(f) continuous from below (left-continuous) in $1/y$.

Proof: (a) If $\mathbf{x} \in \mathbf{L}(y) \cap \mathbf{\Omega}_n$, $y > 0$, then by definition [equation (2)] D exists and $D \geq 1$, since $(\mathbf{x}/1) \in \mathbf{L}(y)$. D is finite, since $\lim_{d\to\infty}(1/d)\mathbf{x} = \mathbf{0} \notin \mathbf{L}(y)$. If $\mathbf{0} \ll \mathbf{x} \notin \mathbf{L}(y)$, then since $\mathbf{L}(y)$ is not empty, there exists an $\mathbf{x}^0$, such that $\mathbf{0} \ll \mathbf{x}^0 \in \mathbf{L}(y)$. Let $d_0 = \text{Min}_i\{x_i/x_i^0\} > 0$; then $(1/d_0)\mathbf{x} \geq \mathbf{x}^0$ and thus $(1/d_0)\mathbf{x} \in \mathbf{L}(y)$, by the free disposal assumption. Thus $D(1/y;\mathbf{x})$ exists and $D \geq d_0 > 0$. Also, $D < 1$; since if $D \geq 1$, $(1/(1-\epsilon))\mathbf{x} \geq (1/(D-\epsilon))\mathbf{x} \in \mathbf{L}(y)$, and by taking limits as $\epsilon \to 0$, the closedness of $\mathbf{L}(y)$ implies that $\mathbf{x} \in \mathbf{L}(y)$, a contradiction. We have thus proved that $D \geq 1 \Leftrightarrow \mathbf{0} \ll \mathbf{x} \in \mathbf{L}(y)$, which implies the last statement of Theorem 1:

$$\mathbf{L}^*(y) = \{\mathbf{x}: D(1/y,\mathbf{x}) \geq 1; \mathbf{x} \in \mathbf{\Omega}_n\} \equiv \mathbf{L}(y) \cap \mathbf{\Omega}_n.$$

This identity, in addition to equation (2), leads almost immediately to the proof of Conditions III(b) – III(f):

(b) Since $\mathbf{L}(0) = \bar{\mathbf{\Omega}}_n$, and $y' > y \Rightarrow \mathbf{L}(y') \subseteq \mathbf{L}(y)$, Condition III(b) follows from equation (2).

(c) Follows from the free disposal assumption.

(d) $D(1/y;\mathbf{x})$ is linear-homogeneous in $\mathbf{x}$, since for $\lambda > 0$, $(1/\lambda d)(\lambda\mathbf{x}) = (1/d)\mathbf{x}$.

(e) The convexity of $\mathbf{L}(y)$ implies that $D(1/y;\mathbf{x})$ is quasi-concave in $\mathbf{x}$. Since D is linear-homogeneous in $\mathbf{x}$, it is concave in $\mathbf{x}$.

(f) The continuity from below in $1/y$ follows from the closedness of the graph of $\mathbf{L}(y)$. Q.E.D.

It should be noted, that since $\mathbf{L}(y)$ is the closure of $\mathbf{L}^*(y)$, one could extend the definition of $D(1/y;\mathbf{x})$ to all non-negative $\mathbf{x}$, by assuming that $D = 0$ if $(1/d)\mathbf{x} \notin \mathbf{L}(y)$ for all $d > 0$, $y \geq 0$ [or $\lim_{d\to 0}(1/d)\mathbf{x} \in \mathbf{L}(y)$]. In this case,

[6] See Diewert (1971, Condition III, 2.13) for specification of similar conditions with respect to the cost function $c(y;\mathbf{p})$. Cf. also Shephard (1953) and Uzawa (1964).

$$\mathbf{L}'(y) = \{\mathbf{x}:D(1/y;\mathbf{x}) \geq 1;\mathbf{x} \geq \mathbf{0}\} \equiv \mathbf{L}(y).$$

We now analyze the dual relation, i.e., the cost function. The Shephard duality theorem, as extended by Uzawa and McFadden, states that the function

$$C = G(y;\mathbf{p}) = \mathrm{Min}_{\mathbf{x}}\{\mathbf{p}'\mathbf{x}:\mathbf{x} \in \mathbf{L}(y)\} \tag{3}$$

(where $\mathbf{p}'\mathbf{x} = \sum_i p_i x_i$) is uniquely determined by $\mathbf{L}(y)$ and satisfies Condition III above, with $(y;\mathbf{p})$ substituted for $(1/y;\mathbf{x})$. The set of $\mathbf{p}$ (for given y), defined by the equation $G(y;\mathbf{p}) = 1$, is the *unit cost frontier*. Define the *unit cost set* as follows:

$$\mathbf{V}(1/y) = \{\mathbf{p}:G(y;\mathbf{p}) \geq 1\}. \tag{4}$$

Since G satisfies Condition III, the previous discussion suffices to establish that $\mathbf{V}(1/y)$ satisfies, in $\mathbf{\Omega}_n$, exactly the same condition as $\mathbf{L}(y)$ (Condition II) with $1/y$ substituted for y, and that $G(y;\mathbf{p})$ is the "distance function" corresponding to $\mathbf{V}(1/y)$; that is,

$$G(y;\mathbf{p}) = \sup\{g:(1/g)\mathbf{p} \in \mathbf{V}(1/y);\mathbf{p} \in \mathbf{\Omega}_n)\}. \tag{5}$$

Due to the perfect symmetry established between the sets $\mathbf{L}(y)$ and $\mathbf{V}(1/y)$, and the functions $D(1/y;\mathbf{x})$ and $G(y;\mathbf{p})$, it is now possible to apply the duality theorem "in reverse", without changing the proof, to obtain the following theorem:

Theorem 2. $D(1/y;\mathbf{x}) = \mathrm{Min}_{\mathbf{p}}\{\mathbf{x}'\mathbf{p}:\mathbf{p} \in \mathbf{V}(1/y)\}$, (6)

where D and $\mathbf{V}$, are defined by equations (2) and (4), respectively.

Proof: Identical with the proof of the duality theorem on costs, with the dual variables $(y;\mathbf{p})$ substituted for the primal variables $(1/y;\mathbf{x})$, and vice-versa.[7] Q.E.D.

In addition, the functions G and D satisfy *Shephard's Lemma*;[8] that is, the first partial derivatives of G or D with respect to an input price or quantity, respectively – whenever they exist – are equal to the corresponding dual variables; i.e., if the derivatives exist,

$$\partial G(y;\mathbf{p})/\partial p_i = x_i^*, \quad \partial D(1/y;\mathbf{x})/\partial x_i = p_i^*. \tag{7}$$

[7] E.g., McFadden (Chapter I.1) or Diewert (1971, Theorem 4).
[8] Shephard (1953). Cf. also Diewert (1971).

The right-hand sides of equations (7) are the input demand and inverse demand functions, respectively (if existing), where p_i^* are the normalized (to yield unit cost) shadow prices (p_i/C):[9]

$$x_i^* = x_i^*(y;\mathbf{p}), \quad p_i^* = p_i^*(1/y;\mathbf{x}).$$

If the equation $D(1/y;\mathbf{x}) = 1$ is solved explicitly for y in terms of $\mathbf{x}$, the result $y = f(\mathbf{x})$ (for $\mathbf{x} \in \mathbf{\Omega}_n$) is the production function, satisfying the required regularity condition (Condition I).

Formally,

$$\begin{aligned} y = f(\mathbf{x}) &= \sup\{\eta : D(1/\eta;\mathbf{x}) \geq 1\} \\ &= \sup\{\eta : \mathbf{x} \in \mathbf{L}(\eta)\}. \end{aligned} \tag{8}$$

Similarly, the unit cost equation $G(y;\mathbf{p}) = 1$ may be solved explicitly for $1/y$ $(0 < y < \infty)$,

$$\begin{aligned} 1/y = g(\mathbf{p}) &= \sup\{1/\gamma : G(\gamma,\mathbf{p}) \geq 1\} \\ &= \sup\{1/\gamma : \mathbf{p} \in \mathbf{V}(1/\gamma)\} = 1/\inf\{\gamma : \mathbf{p} \in \mathbf{V}(1/\gamma)\}. \end{aligned} \tag{9}$$

The function $g(\mathbf{p})$ satisfies the same conditions with respect to $\mathbf{p}$ as $f(\mathbf{x})$ with respect to $\mathbf{x}$ (Condition I). In analogy to the accepted terminology of consumer utility theory, the function $h(\mathbf{p}) = 1/g(\mathbf{p})$ may be denoted the "*indirect production function*", corresponding to the direct production function $y = f(\mathbf{x})$ and the function $g(\mathbf{p})$ may be denoted the "reciprocal indirect production function". The equation $g(\mathbf{p}) = 1/y$, for a given y, is the "factor price frontier". Any standard production relation may be uniquely characterized by each of these functions $g(\mathbf{p})$ or $h(\mathbf{p})$.

In the discussion of duality and "self-duality" in utility theory, the accepted formulation[10] for the dual indirect form corresponding to utility $U(\mathbf{x})$ is $-V(\mathbf{p})$ [rather than $1/V(\mathbf{p})$]. However, since utility is ordinal, one could equally choose $e^{U(\mathbf{x})}$ and $e^{-V(\mathbf{p})} = 1/e^{V(\mathbf{p})}$, in analogy to the present results, without affecting the corresponding direct and indirect demand functions, or any real behavior. A similar monotone transformation is not acceptable in production theory (unless output y is replaced by $\log y$), since y is a measurable, non-negative quantity. Equation (9)

[9]The notation $\mathbf{p}^*$ implies both the optimality property of $\mathbf{p}$ (i.e., $\mathbf{p}^*$ are shadow prices) and the normalization to yield unit cost $\mathbf{p}^* = (1/c)\mathbf{p}$.

[10]E.g., Houthakker (1960, 1965), Samuelson (1947, 1965a), Lau (1969a) and Pollak (1972). The indirect utility $V(\mathbf{p})$ referred to here, is in terms of the n *normalized* prices (i.e., per unit of expenditure) and not the alternative (homogeneous) indirect utility $V(E,\mathbf{p})$ which is homogeneous of degree zero in $(n+1)$ arguments – the non-normalized prices $\mathbf{p}$ and expenditure E.

TABLE 1
Symmetric dual relations for cost and production functions.

	Primal (production)	Dual (unit costs)
Variables:	Input quantity x_i Output y	Input price p_i $1/y$
Sets (satisfying Condition II):	Production possibilities $\mathbf{L}(y)$	Unit cost $\mathbf{V}(1/y)$
Functions:		
Distance (satisfying Condition III):	$D(1/y;\mathbf{x}) = \sup\{d:(1/d)\mathbf{x} \in \mathbf{L}(y)\}$	$G(\mathbf{y};\mathbf{p}) = \sup\{g:(1/g)\mathbf{p} \in \mathbf{V}(1/y)\}$
Explicit (satisfying Condition I):	$y = f(\mathbf{x}) = \sup\{\eta:\mathbf{x} \in \mathbf{L}(\eta)\}$	$1/y = g(\mathbf{p}) = \sup\{1/\gamma:\mathbf{p} \in \mathbf{V}(1/\gamma)\}$
Minimum property:	$D(1/y;\mathbf{x}) = \mathrm{Min}_{\mathbf{p}}\{\mathbf{x}'\mathbf{p}:\mathbf{p} \in \mathbf{V}(1/y)\}$	$G(y;\mathbf{p}) = \mathrm{Min}_{\mathbf{x}}\{\mathbf{p}'\mathbf{x}:\mathbf{x} \in \mathbf{L}(y)\}$
Partial derivatives (when existing):	$\partial D/\partial x_i = p_i^*$ (factor's inverse demand)	$\partial G/\partial p_i = x_i^*$ (factor's demand)

seems therefore to be a more natural definition of the indirect production function, since it preserves the perfect symmetry between the primal and the dual representations.

Finally, if $h(y)$ is any strictly monotone function, such that $h(0) = 0$ and $h(\infty) = \infty$, one may choose the pair of dual variables to be $h(y)$ and $1/h(y)$ (rather than y and $1/y$), without affecting the results, nor the perfect symmetry of the dual relations. (The significance of this remark is clarified below, in the discussions of homothetic functions and of separable production frontiers.)

Table 1 summarizes these symmetric dual relations for the case of production and cost functions with a single output.

3. Polar Production and Cost Functions

Having demonstrated in the previous section the perfect symmetry between production and cost relations, it is now evident that new valid cost and production relations (namely the *polar* relations[11]), may be obtained by exchanging the roles of the sets $\mathbf{L}(y)$ and $\mathbf{V}(1/y)$, or equivalently the functions $D(1/y;\mathbf{x})$ and $G(y;\mathbf{p})$, through substitution of the dual variables $(1/y;\mathbf{p})$ for the primal variables $(y;\mathbf{x})$, and vice-versa (within the positive orthant $\mathbf{\Omega}_{n+1}$). The new cost and production functions thus generated are necessarily standard, satisfying the required conditions.

If the original production function is represented *implicitly* by an identity $F(y;\mathbf{x}) \equiv 0$ [where F satisfies the conditions of the implicit function theorem[12] for yielding a unique standard $y = f(\mathbf{x})$], then the unit-cost frontier of the polar production function is given by $F(1/y;\mathbf{p}) \equiv 0$, and the polar total cost $D(y;\mathbf{p})$ is defined implicitly by the identity $F(1/y;(1/D)\mathbf{p}) \equiv 0$. Conversely, if the original cost function G is given implicitly by $H(y;(1/G)\mathbf{p}) \equiv 0$ (G being linear homogeneous in $\mathbf{p}$), the polar production function is given implicitly by $H(1/y;\mathbf{x}) \equiv 0$, provided H satisfies the required conditions. Similar substitutions of variables would appear if either the cost or the production function is represented by a

[11]The term *polar* was adopted in accordance with Shephard's geometric interpretation (1953), Ch.5), where the isoquant and the unit cost surfaces are shown to be polar reciprocal to each other with respect to the unit sphere $\Sigma x_i^2 = 1$.

[12]E.g., in Hadley (1964), Courant (1936). The implicit function theorem for a single function (identity) could be made somewhat stronger, to apply to non-differentiable cases (with only strict monotonicity at 0) such as the general case discussed here.

set of parametric equations. The polar relations are now stated formally:

Theorem 3. Given a standard production function $y = f(\mathbf{x})$ which may be represented uniquely by any one of the following equivalent relations:[13]

Primal	Dual	Satisfying Conditions
(1) PF: $y = f(\mathbf{x})$	(a) InPF: $1/y = g(\mathbf{p})$	I
(2) ImPF: $F(y;\mathbf{x}) \equiv 0$	(b) IIPF: $H(1/y;\mathbf{p}) \equiv 0$	
(3) UDE: $D(1/y;\mathbf{x}) = 1$	(c) UCE: $G(y;\mathbf{p}) = 1$	III
(4) FS: $\mathbf{L}(y)$	(d) UCS: $\mathbf{V}(1/y)$	II

there exists a unique Polar Production Function $y = g(\mathbf{x})$ which is standard, and may be represented uniquely by any one of the following equivalent relations:

Primal	Dual	Satisfying Conditions
(1′) PF: $y = g(\mathbf{x})$	(a′) InPF: $1/y = f(\mathbf{p})$	I
(2′) ImPF: $H(y;\mathbf{x}) \equiv 0$	(b′) IIPF: $F(1/y;\mathbf{p}) \equiv 0$	
(3′) UDE: $G(1/y;\mathbf{x}) = 1$	(c′) UCE: $D(y;\mathbf{p}) = 1$	III
(4′) FS: $\mathbf{V}(y)$	(d′) UCS: $\mathbf{L}(1/y)$	II

Proof: By construction and by the results of Section 2, the Conditions I, II, or III are satisfied with respect to $g(\mathbf{x})$, $\mathbf{V}(y)$ or $D(y;\mathbf{p})$, respectively, for finite $y > 0$, $\mathbf{x} \gg \mathbf{0}$ [that is, $(y;\mathbf{x})$ and $(1/y;\mathbf{p})$ in $\boldsymbol{\Omega}_{n+1}$]. In order to extend these relations to all $\mathbf{x} \geq \mathbf{0}$, $\mathbf{p} \geq \mathbf{0}$, $y \geq 0$, one should put $g(\mathbf{0}) = 0$, and extend the definition of $g(\mathbf{x})$ to the (right-hand) limits as $\mathbf{x}$ approaches the boundaries of $\bar{\boldsymbol{\Omega}}_n$. Equivalently, the sets $\mathbf{V}(y)$ are to be replaced by their closures, i.e., include their boundary points in $\bar{\boldsymbol{\Omega}}_n$. Similar modifications apply to the other relations in Theorem 3.

[13]PF = production function; ImPF = implicit production function; UDE = unit distance equation; FS = feasible sets; InPF = indirect production function; IIPF = indirect implicit production function; UCE = unit cost equation; UCS = unit cost sets.

The uniqueness of the polar functions follows from their construction. The minimum properties appearing in Theorem 2 and in Table 1, assure that $D(y;\mathbf{p})$ is indeed the cost function corresponding to the production distance function $G(1/y;\mathbf{x})$ by the duality theorem; that is

$$D(y;\mathbf{p}) = \operatorname*{Min}_{\mathbf{x}}\{\mathbf{p}'\mathbf{x}:G(1/y;\mathbf{x}) \geq 1\}, \tag{10}$$

and also

$$G(1/y;\mathbf{x}) = \operatorname*{Min}_{\mathbf{p}}\{\mathbf{x}'\mathbf{p}:D(y;\mathbf{p}) \geq 1\}. \tag{11}$$

This completes our proof. A few additional remarks are in order. First, differentiability of $f(\mathbf{x})$ does not necessarily imply differentiability of the polar function $g(\mathbf{x})$, or vice-versa. This is best demonstrated by the fact that non-differentiable standard production functions may have differentiable dual cost functions, and therefore differentiable polar production functions (such as the Leontief fixed-coefficient function, with a linear differentiable cost function; see Chapter II.3). A separate duality theorem applies to the restricted class of smooth neo-classical production and cost functions,[14] such that both the original and the polar functions belong to this class, and yield everywhere continuously differentiable demand functions. It may be shown, however, that the polar transformation defined here for a more general class, transforms each isoquant surface so that any smooth, strictly convex part is transformed into a smooth counterpart; every planar section into a vertex, and every vertex into a planar section [see Shephard (1953, p. 11)].

Let us now examine the special case of *homothetic functions*. If $f(\mathbf{x})$ is homothetic, it may be written in the form $h(y) = f^*(\mathbf{x})$, where $f^*(\mathbf{x})$ is linear homogenous, and $h(y)$ strictly increases with y from 0 to ∞. The dual unit cost function is separable in this case into $G(y;\mathbf{p}) = h(y)\cdot g^*(\mathbf{p}) = 1$, or $1/h(y) = g^*(\mathbf{p})$, where $g^*(\mathbf{p})$ is the cost of producing the output $y = h^{-1}(1)$, and where the elasticity of total cost with respect to output [see Hanoch (1975b)] is $\eta_{cy} = \partial \log G/\partial \log y = yh'(y)/h(y)$, and is independent of prices! Applying our polar transformation yields the new production function given by $1/h(1/y) = g^*(\mathbf{x})$, with a corresponding unit cost equation, $1/h(1/y)\cdot f^*(\mathbf{p}) = 1$. The new output elasticity of cost is $\eta^*_{cy} = (1/y)(h'(1/y)/h(1/y))$, and the polar function is also homothetic.

[14]Cf. Lau (Chapter I.3) for a presentation and a proof of this restricted duality theorem.

However, if one wishes to impose on the polar function the same behavior of cost with respect to output as in the original – transforming only the form of any given isoquant-surface, but preserving the behavior of output along any given ray $(\lambda \mathbf{x})$ – one can modify the polar transformation so as to make $1/h(y)$ the dual variable to $h(y)$ (that is, $h^*(y) = h^{-1}[1/h(y)]$ dual to y), as explained in Section 2 above. In this case, the *homothetic-polar* transformation yields the production function $h(y) = g^*(\mathbf{x})$, with the corresponding cost function $c^* = h(y) \cdot f^*(\mathbf{p})$. These two transformations are identical, if and only if $f(\mathbf{x})$ is *homogeneous* of some degree $\mu > 0$. In this case it may be shown that $h(y) = y^{1/\mu}$, hence $h(1/y) = y^{-1/\mu} = 1/h(y)$, and thus $h^*(y) = 1/y$.

A similar approach allows extension of the polar transformation through cost functions to the case of joint production with multiple outputs, *if outputs are separable from inputs*. That is, if the production frontier is of the form $F[h(\mathbf{y});\mathbf{x}] \equiv 0$, which may be solved for $h(\mathbf{y})$ ($\mathbf{y}$ a vector of order m),

$$h(\mathbf{y}) = f(\mathbf{x}), \tag{12}$$

where $f(\mathbf{x})$ is standard, and $h(\mathbf{y})$ increasing in $\mathbf{y}$, such that $h(\mathbf{0}) = 0$; $\mathbf{y}' > \mathbf{y}^0 \Rightarrow h(\mathbf{y}') > h(\mathbf{y}^0)$ [an increase in at least one output requires an increase in $f(\mathbf{x})$, and therefore of $\mathbf{x}$, since $f(\mathbf{x})$ is single-valued and non-decreasing] and $h(\mathbf{y}^N)$ is unbounded if $\mathbf{y}^N$ is unbounded. The corresponding cost function is $G[h(\mathbf{y});\mathbf{p}]$, and the unit cost frontier is separable into $1/h(\mathbf{y}) = g(\mathbf{p})$. Hence, the previous analysis is carried over entirely, with $h(\mathbf{y})$ substituted for y as the primal variable, and $1/h(\mathbf{y})$ replacing $1/y$ as the dual variable. The cost-polar transformation now yields a new separable production frontier, of the form $h(\mathbf{y}) = g(\mathbf{x})$, with $g(\mathbf{x})$ standard, satisfying Condition I, and a new separable unit cost frontier $1/h(\mathbf{y}) = f(\mathbf{p})$.

4. A Symmetric Formulation of Profit Functions and Production Frontiers

The analysis of duality relations between profit functions and production frontiers for the case of joint production with multiple outputs and inputs, may be carried out along lines similar to the cost–production analysis of Section 2, so as to yield a perfectly symmetric formulation for the primal and the dual relations.

Suppose $\mathbf{z} = (\mathbf{y};\mathbf{x})$ are non-negative[15] vectors of m outputs $\mathbf{y}$ ($m \geq 1$), and k inputs $\mathbf{x}$ ($k \geq 1; m + k = n$). The corresponding price vector is denoted by $\mathbf{q} = (\mathbf{p};\mathbf{w}) \in \bar{\mathbf{\Omega}}_n$. The set of feasible input–output combinations for a given production process is denoted by $\mathbf{T}(\subseteq \bar{\mathbf{\Omega}}_n)$. The conditions for $\mathbf{T}$ being *regular*, namely for the existence of a unique non-negative dual profit function,

$$\pi = Q(\mathbf{p};\mathbf{w}) = \sup_{(\mathbf{y};\mathbf{x})} \{\mathbf{p}'\mathbf{y} - \mathbf{w}'\mathbf{x} : (\mathbf{y};\mathbf{x}) \in \mathbf{T} \subseteq \bar{\mathbf{\Omega}}_n\}, \tag{13}$$

are as follows:[16]

Condition B. $\mathbf{T}$ is a closed, convex set in $\bar{\mathbf{\Omega}}_n$, with free disposal:

$(\mathbf{y};\mathbf{x}') \geq (\mathbf{y};\mathbf{x}) \in \mathbf{T} \Rightarrow (\mathbf{y};\mathbf{x}') \in \mathbf{T}$,
$(\mathbf{0};\mathbf{x}) \leq (\mathbf{y}';\mathbf{x}) \leq (\mathbf{y};\mathbf{x}) \in \mathbf{T} \Rightarrow (\mathbf{y}';\mathbf{x}) \in \mathbf{T}$.

$(\mathbf{0};\mathbf{0}) \in \mathbf{T}$, and $(\mathbf{y}^0;\mathbf{x}^0) \in \mathbf{T}$ for some $\mathbf{y}^0 \gg \mathbf{0}$.

Bounded inputs $\mathbf{x}$ imply bounded outputs $\mathbf{y}$ in $\mathbf{T}$.

By McFadden's duality theorem, the profit function defined in equation (13) exists uniquely and satisfies the following:

Condition A.

(1) $Q(\mathbf{p};\mathbf{w})$ is a real, non-negative function of $\mathbf{q} = (\mathbf{p};\mathbf{w}) \geq \mathbf{0}$. $Q(\mathbf{0};\mathbf{0}) = 0$ and $Q(\mathbf{q}^0) > 0$ for some $\mathbf{q}^0 \gg \mathbf{0}$. [$Q$ may be infinite for finite $\mathbf{q}$.]
(2) Q is non-increasing in $\mathbf{w}$ and non-decreasing in $\mathbf{p}$.
(3) If $\mathbf{w} \gg \mathbf{0}$, $\lim_{d\to 0} Q(\mathbf{p};(1/d)\mathbf{w}) \leq \mathbf{p}'\mathbf{a}$, where $\mathbf{a} > \mathbf{0}$ is a vector of fixed, finite values.
(4) $Q(\mathbf{q})$ is a convex, closed function for $\mathbf{q} \geq \mathbf{0}$.
(5) Q is positive linear homogeneous in $\mathbf{q}$: $\mathbf{q} > \mathbf{0}, \lambda > 0 \Rightarrow Q(\lambda\mathbf{q}) = \lambda Q(\mathbf{q})$.

Define the *unit profit set* $\mathbf{V}$ as follows:

$$\mathbf{V} = \{\mathbf{q} : Q(\mathbf{q}) \leq 1; \mathbf{q} \geq \mathbf{0}\}; \tag{14}$$

[15]It is more convenient for our purposes to define the quantity vectors $(\mathbf{y};\mathbf{x})$ with all the arguments non-negative, rather than the "net outputs" notation $(\mathbf{y};-\mathbf{x})$. In our notation, outputs are mathematically distinguished from inputs by the direction of change of the frontier functions, rather than by their sign. Cf. McFadden (Chapter I.1).

[16]These are the conditions in Diewert (1973a), modified to imply non-negative (but not identically zero) profits.

then the following theorem holds:

> *Theorem* 4. If $Q(\mathbf{p};\mathbf{w})$ satisfies Condition A, the unit profit set $\mathbf{V}$ satisfies Condition B, where input prices $\mathbf{w}$ are to be substituted for inputs $\mathbf{x}$; output prices $\mathbf{p}$ for outputs $\mathbf{y}$, and $\mathbf{V}$ for $\mathbf{T}$.

Proof: By Condition A(4), $\mathbf{V}$ is convex and closed, since $Q(\mathbf{q})$ is a convex, closed function over the domain $\{\mathbf{q}:0 \leq Q(\mathbf{q}) \leq 1\}$.

By Condition A(2), the free disposal conditions follow immediately.

To show that $(\mathbf{p}^0;\mathbf{w}^0) \in \mathbf{V}$ for some $\mathbf{p}^0 \gg \mathbf{0}$, note that by Condition A(3), there exists a $(\mathbf{p}';\mathbf{w}') \gg \mathbf{0}$ such that $Q(\mathbf{p}';\mathbf{w}') = Q_0 < \infty$. If $Q_0 \leq 1$, choose $(\mathbf{p}^0;\mathbf{w}^0) = (\mathbf{p}';\mathbf{w}') \in \mathbf{V}$; and if $Q_0 > 1$ then $\mathbf{0} \ll (\mathbf{p}^0;\mathbf{w}^0) = ((1/Q_0)\mathbf{p}';(1/Q_0)\mathbf{w}') \in \mathbf{V}$, by Condition A(5).

It remains to be shown that bounded $\mathbf{w}$ imply bounded $\mathbf{p}$ in $\mathbf{V}$. Suppose $\{\mathbf{w}^M\} \leq \mathbf{B}$ ($\mathbf{B}$ a finite vector), but $\{\mathbf{p}^M\}$ unbounded in $\mathbf{V}$. Since there exists by Condition A(1) a strictly positive vector $(\mathbf{p}^0;\mathbf{w}^0) \gg \mathbf{0}$ such that $Q(\mathbf{p}^0;\mathbf{w}^0) > 0$, we may choose a partial sequence $\{\mathbf{p}^N;\mathbf{w}^N\}$ such that $\mathbf{p}^N \geq N\mathbf{p}^0;\mathbf{w}^N \leq \mathbf{B} \leq N\mathbf{w}^0$ for all $N \geq N_0$. Hence we get by Conditions A(2) and A(5): $\lim_{N\to\infty} Q(\mathbf{p}^N;\mathbf{w}^N) \geq \lim_{N\to\infty} Q(N\mathbf{p}^0;N\mathbf{w}^0) = \infty$, and $(\mathbf{p}^N;\mathbf{w}^N)$ cannot be all in $\mathbf{V}$, a contradiction. Q.E.D.

Theorem 4 establishes a complete symmetry between the regular production set $\mathbf{T}$ and the regular corresponding unit profit set $\mathbf{V}$, which is the dual of $\mathbf{T}$. Equipped with the duality theorem and this result, we may now state without further proof all the other symmetric results which follow by the transformation from the prices space into the quantities space and conversely.

First, applying the transformation to the duality theorem we may define the "*gauge function*" $H(\mathbf{y};\mathbf{x})$, by the following maximum property:

$$H(\mathbf{y};\mathbf{x}) = \sup_{(\mathbf{p},\mathbf{w})} \{\mathbf{y}'\mathbf{p} - \mathbf{x}'\mathbf{w}:(\mathbf{p};\mathbf{w}) \in \mathbf{V} \subseteq \bar{\mathbf{\Omega}}_n\}, \tag{15}$$

which is the dual counterpart of equation (13). The set $\mathbf{T}$ is then derivable from $H(\mathbf{y};\mathbf{x})$ by

$$\mathbf{T} = \{(\mathbf{y};\mathbf{x}):H(\mathbf{y};\mathbf{x}) \leq 1;(\mathbf{y};\mathbf{x}) \geq \mathbf{0}\}, \tag{16}$$

which is equivalent to equation (14). The function H defined here satisfies Condition A, with the appropriate substitution of variables. In particular, H is linear homogeneous in $(\mathbf{y};\mathbf{x})$ for any production frontier which is regular (but not necessarily homogeneous).

Given a regular unit profit set $\mathbf{V}$ satisfying Condition B, the profit function $Q(\mathbf{p};\mathbf{w})$ is derivable from $\mathbf{V}$ by the following relation:

$$Q(\mathbf{p};\mathbf{w}) = \inf\left\{\theta : \left(\frac{1}{\theta}\mathbf{p};\frac{1}{\theta}\mathbf{w}\right) \in \mathbf{V}; 0 < \theta \leq \infty\right\}. \tag{17}$$

This is shown as follows:

$$\begin{aligned}\inf\left\{\theta : \left(\frac{1}{\theta}\mathbf{q}\right) \in \mathbf{V}; 0 < \theta \leq \infty\right\} &= \inf\left\{\theta : Q\left(\frac{1}{\theta}\mathbf{q}\right) \leq 1; 0 < \theta \leq \infty\right\} \\ &= \inf\{\theta : Q(\mathbf{q}) \leq \theta; 0 < \theta \leq \infty\} = Q(\mathbf{q}).\end{aligned}$$

Q is the "gauge function" of the set $\mathbf{V}$, representing the distance from the origin of a point $(\mathbf{p};\mathbf{w}) \geq 0$, divided by the furthest distance from $(\mathbf{0})$ of points $((1/\theta)\mathbf{p};(1/\theta)\mathbf{w})$ which lie on the efficient boundary of the set $\mathbf{V}$.[17] In an exactly analogous derivation for the dual case, we have

$$H(\mathbf{y};\mathbf{x}) = \inf\left\{h : \left(\frac{1}{h}\mathbf{y};\frac{1}{h}\mathbf{x}\right) \in \mathbf{T}; 0 < h \leq \infty\right\}, \tag{18}$$

where $H \leq 1$ in the production set $\mathbf{T}$, and $H = 1$ on the production frontier (when it exists). In general, if the frontier exists and is given by an implicit function $F(\mathbf{y};\mathbf{x}) \equiv 0$, where F satisfies the following:

> *Condition C.* The set $\mathbf{T} = \{(\mathbf{y};\mathbf{x}) : F(\mathbf{y};\mathbf{x}) \leq 0; (\mathbf{y};\mathbf{x}) \geq \mathbf{0}\}$ satisfies Condition B above;[18]

then this frontier could equally be represented by $H(\mathbf{y};\mathbf{x}) = 1$ where H is linear homogeneous and satisfies Condition A. Symmetrically, maximum profits π may be represented by an implicit equation in the variables $((1/\pi)\mathbf{p};(1/\pi)\mathbf{w})$ [since π is linear homogeneous]: $R((1/\pi)\mathbf{p};(1/\pi)\mathbf{w}) \equiv 0$, and the *unit profit frontier* by $R(\mathbf{p}^*;\mathbf{w}^*) \equiv 0$ (R is generally *not* linear homogeneous; $\mathbf{p}^*$, $\mathbf{w}^*$ are prices normalized to yield maximum profits equal to unity).

When the partial derivatives of Q and H exist, they satisfy the relations[19] $\partial Q/\partial p_i = y_i^*(\mathbf{q})$; $\partial Q/\partial w_j = -x_j^*(\mathbf{q})$, where y_i^* and x_j^* are the

[17]This includes $Q = 0$ if $(1/\theta)\mathbf{q} \in \mathbf{V}$ for all $\theta > 0$; and $Q = \infty$, if $(1/\theta)\mathbf{q} \notin \mathbf{V}$ for all $\theta > 0$.

[18]For specification of direct conditions on the function F (rather than the set $\mathbf{T}$), see Diewert (1973a). He assumes, however, that F is normalized, i.e., solved for one argument as a dependent variable. More general conditions on F could also be given, but are omitted here.

[19]E.g., in Diewert (1973a), Lau (Chapter I.3), and others.

TABLE 2
Symmetric dual relations for regular profit and production functions.

	Primal (production)	Dual (unit profits)
Variables (non-negative):	Inputs x_i Outputs y_i	Input prices w_i Output prices p_i
Sets (satisfying Condition B):	Production **T**	Unit profit **V**
Functions:		
Gauge (satisfying Condition A):	$H(\mathbf{y};\mathbf{x})\text{:inf}\left\{h:\left(\frac{1}{h}\mathbf{y};\frac{1}{h}\mathbf{x}\right)\in\mathbf{T}\right\}$	$Q(\mathbf{p};\mathbf{w}) =\text{:inf}\left\{\theta:\left(\frac{1}{\theta}\mathbf{p};\frac{1}{\theta}\mathbf{w}\right)\in\mathbf{V}\right\}$
Implicit (satisfying Condition C):	$F(\mathbf{y};\mathbf{x})\equiv 0\ (F\leq 0 \text{ in } \mathbf{T})$	$R(\mathbf{p};\mathbf{w})\equiv 0\ (R\leq 0 \text{ in } \mathbf{V})$
Maximum property:	$H=\sup\{\mathbf{p}'\mathbf{y}-\mathbf{w}'\mathbf{x}:(\mathbf{p};\mathbf{w})\in\mathbf{V}\}$	$Q=\sup\{\mathbf{y}'\mathbf{p}-\mathbf{x}'\mathbf{w}:(\mathbf{y};\mathbf{x})\in\mathbf{T}\}$
Partial derivatives (when existing):	$\partial H/\partial y_i = p_i^*;\ \partial H/\partial x_j = -w_j^*$	$\partial Q/\partial p_i = y_i^*;\ \partial Q/\partial w_j = -x_j^*$

output supply and factor demand functions, respectively. Similarly, $\partial H/\partial y_i = p_i^*(\mathbf{z})$; $\partial H/\partial x_i = -w_i^*(\mathbf{z})$ (if existing), where p_i^* and w_i^* are the optimal *inverse* supply and demand functions, determining the normalized shadow price variables $(1/\pi)\mathbf{p},(1/\pi)\mathbf{w}$.

The Table 2 summarizes these symmetric relations.

The foregoing analysis rests heavily on the assumption that zero variable outputs and inputs are feasible; i.e., $(\mathbf{0};\mathbf{0}) \in \mathbf{T}$, since in this case maximum variable profits are non-negative, and the profit function is completely determined by the unit profit set. If, however, some minimum positive inputs are always required – either due to indivisibilities, or because some outputs are fixed or are bounded below through exogeneously determined restrictions – then $(\mathbf{0},\mathbf{0})$ is not feasible, and maximum variable profits assume negative values. The symmetric duality relations in terms of the original variables break down. However, as shown by McFadden in Chapter I.1, the variables may be translated to be measured from a point $(\boldsymbol{\eta},\boldsymbol{\xi})$ in $\mathbf{T}$, and the symmetric duality applies with respect to the translated variables $(\mathbf{y}-\boldsymbol{\eta},\mathbf{x}-\boldsymbol{\xi})$ and with respect to the corresponding modified profit and production functions. (Details and proofs of these general statements are omitted here.)

5. The Polar Profit and Production Functions

In analogy to the derivation of cost polar production functions, the perfect symmetry of the dual production–profit relations exhibited above leads to the definition, for any regular production set $\mathbf{T}$, of another regular production set $\mathbf{T}^*$, which coincides with the original unit profit set $\mathbf{V}$, if prices are transformed to the respective quantities. That is,

$$\mathbf{T}^* = \{(\mathbf{y};\mathbf{x}):(\mathbf{y};\mathbf{x}) = (\mathbf{p};\mathbf{w}) \in \mathbf{V}\}. \tag{19}$$

The set $\mathbf{T}^*$ is the *profit polar production set*, determined uniquely by $\mathbf{T}$, and satisfying the same conditions specified in Condition B. Similar results apply to all equivalent representations of $\mathbf{T}^*$, as summarized by the following theorem:

Theorem 5. Given a regular production set $\mathbf{T}$, which may be represented uniquely by any one of the following equivalent relations:[20]

[20]PS = production set; PFr = production frontier; GF = gauge function; UPS = unit profit set; IPF = implicit profit function; PFn = profit function.

Primal	Dual	Satisfying Conditions
(1) PS: $\mathbf{T}$	(a) UPS: $\mathbf{V}$	B
(2) PFr: $F(\mathbf{y};\mathbf{x}) \equiv 0$	(b) IPF: $R\left(\frac{1}{\pi}\mathbf{p};\frac{1}{\pi}\mathbf{w}\right) \equiv 0$	C
(3) GF: $H(\mathbf{y};\mathbf{x})$	(c) PFn: $Q(\mathbf{p};\mathbf{w})$	A

there exists a unique polar production set $\mathbf{T}^*$, defined by equation (19), which is regular, and may be represented uniquely by any one of the following equivalent relations:

Primal	Dual	Satisfying Conditions
(1′) PS: $\mathbf{T}^* = \mathbf{V}$	(a′) UPS: $\mathbf{V}^* = \mathbf{T}$	B
(2′) PFr: $R(\mathbf{y};\mathbf{x}) \equiv 0$	(b′) IPF: $F\left(\frac{1}{\pi}\mathbf{p};\frac{1}{\pi}\mathbf{w}\right) \equiv 0$	C
(3′) GF: $Q(\mathbf{y};\mathbf{x})$	(c′) PFn: $H(\mathbf{p};\mathbf{w})$	A

Proof: The proof is immediate, using McFadden's Theorem 24 in Chapter I.1, Theorem 4 above, and the results of Section 4, summarized in Table 2 above. Q.E.D.

The results cited in Table 2 with respect to the maximum properties of the profit and gauge functions, as well as to partial derivatives of Q and H (i.e., factor demand and output supply functions), are applicable to the new polar production frontier, if proper substitutions are made throughout. However, the specific behavior of the polar production relation may be quite different from that of the original relation, as indicated by some of the examples in Chapter II.3 and in Hanoch (1975a).

Let us examine now a few special cases. For the case of a single output with a *concave* production function $y = f(\mathbf{x})$ (such that the profit function exists), the profit polar production function defined here generally yields a different production function from the cost polar function defined in Section 3. However, if $y = f(\mathbf{x})$ is *homogeneous* of degree μ $(0 < \mu < 1)$, the two functions coincide, except for a constant scale factor. The cost function dual to $f(\mathbf{x})$ is separable in this case in the form $C = y^{1/\mu}G(\mathbf{w})$, where $G(\mathbf{w})$ is linear homogeneous. Hence the cost polar production function is given by $y = g^c(\mathbf{x}) = [G(\mathbf{x})]^{\mu}$. The profit function dual to $f(\mathbf{x})$ is derived as follows (assuming $\partial C/\partial y$ exists): $\partial C/\partial y = C/\mu y = p$, equating output price to marginal cost. The profit maximizing costs $\bar{c}^*$ then satisfies

$$\bar{c}^* = (\bar{c}^*/\mu p)^{1/\mu}G(\mathbf{w}),$$

or

$$\bar{c}^* = \{\mu p[G(\mathbf{w})]^{-\mu}\}^{1/(1-\mu)};$$

and profits are given by

$$\pi = py - \bar{c}^* = ((1-\mu)/\mu)\bar{c}^* = (1-\mu)\mu^{\mu/(1-\mu)}\{p[G(\mathbf{w})]^{-\mu}\}^{1/(1-\mu)}.$$

The profit polar production frontier function $y = g^{\pi}(\mathbf{x})$ is derived by substitution of $(1, y, \mathbf{x})$ for $(\pi, p, \mathbf{w})$, respectively, in the above expression, and solving for y gives

$$g^{\pi}(\mathbf{x}) = A[G(\mathbf{x})]^{\mu} = A \cdot g^c(\mathbf{x}),$$

where

$$A = [\mu^{\mu}(1-\mu)^{1-\mu}]^{-1} > 1. \tag{20}$$

Similarly, if $y = f(\mathbf{x})$ is *homothetic*, its cost function is separable in the form $C = h(y)G(\mathbf{w})$, where $(\partial \log h(y))/(\partial \log y) > 1$. Similar manipulations give the regular cost polar production function as $y = g^c(\mathbf{x}) = H^{(c)}[G(\mathbf{x})]$; the *homothetic cost polar* function of Section 3 is of the form $y = H^{(hc)}[G(\mathbf{x})]$; and the profit polar function is $y = H^{(\pi)}[G(\mathbf{x})]$, where the functions $H^{(c)}$, $H^{(hc)}$ and $H^{(\pi)}$ are different functions of a single variable. Therefore, the family of isoquant surfaces given by each of the three polar transformations is the same, but their output-denominations are different. (However, all three polar functions are also homothetic.)

In the case of multiple outputs with *separability of outputs from inputs*, the original production frontier is given by $h(\mathbf{y}) = f(\mathbf{x})$. The unit profit frontier of the profit polar function is given by $h(\mathbf{p}) = f(\mathbf{w})$, and is also separable. (The polar production frontier is generally not separable, however.) Hence "direct separability" implies "indirect separability" of the polar function.[21] Clearly, the converse is also true, due to the uniqueness of the profit polar function. That is, (indirect) separability of

[21]On direct and indirect separability and related concepts, see Houthakker (1960, 1965), Sameulson (1965a, 1969b), Gorman (1968b), Goldman and Uzawa (1964), Pollak (1972), and Lau (1969a and Chapter I.3 in this volume).

the original unit profit frontier, implies (direct) separability of the polar production frontier.

The separable cost polar function defined in Section 3 [namely a production frontier with costs $c = G[h(\mathbf{y});\mathbf{w}]$, where c is implicitly given by $1/h(\mathbf{y}) = f((1/c)\mathbf{w})$], is generally different from the profit polar frontier in the direct separability case. An analysis similar to the foregoing shows that these two polar transformations coincide, except for a scale factor, if and only if both $h(\mathbf{y})$ and $f(\mathbf{x})$ are *homogeneous*. The analysis of additional special cases may be carried out along similar lines.[22]

6. Some Extensions and an Application

The process of polar transformation of single-output production functions through cost functions, may be generalized further to joint-production frontiers, under two cases of *short-run profit maximization*:

(i) If either a single output $z_0 = \bar{y}_0$, or a single input $z_0 = \bar{x}_0$ is fixed. The polar transformation of the *variable profits function* $Q(\mathbf{p};\mathbf{w};z_0)$ yields then a polar production frontier $Q(\mathbf{y};\mathbf{x};1/z_0) \equiv 1$, and conversely.

(ii) If the production frontier is separable as between the variable elements $(\mathbf{y};\mathbf{x})$ and the fixed elements $\mathbf{z}_0$ (either inputs or outputs or both). That is, $F(\mathbf{y};\mathbf{x}) = h(\mathbf{z}_0)$. The *polar variable profit* function $\bar{\pi}$ is then given by $F(\mathbf{p}/\bar{\pi};\mathbf{w}/\bar{\pi}) = 1/h(\mathbf{z}_0)$.

The *Factor Requirement Function*[23] defined for the case of a single input, is an obvious special case of (i), and may yield a revenue polar transformation, in complete analogy to the cost polar analysis. Proofs of the above cursory statement are analogous to those given in Sections 2–5.

As a final example of an application of the polar transformation, consider two alternative definitions of the elasticity of substitution which are different from each other, and from the widely used Allen–Uzawa elasticities of substitution,[24] if three or more variable factors are present:

(1) *The Direct Elasticity of Substitution* D_{ij}, for a (twice continuously

[22]For additional results on relations between production and profit functions, see Lau (Chapter I.3). Modifications of such results, so as to apply to polar relations, are straightforward.

[23]See McFadden (Chapter I.1) and Diewert (1974b) for duality theorems with respect to the factor requirement function.

[24]See McFadden (1963) for definitions of these concepts. The D_{ij} were defined by Hicks (1946). See Hanoch (Chapter II.3).

differentiable) production function $y = f(\mathbf{x})$, is defined by

$$D_{ij}(\mathbf{x}) = \left(\frac{1}{x_i f_i} + \frac{1}{x_j f_j}\right) \Big/ \left(-\frac{f_{ii}}{f_i^2} + \frac{2f_{ij}}{f_i f_j} - \frac{f_{jj}}{f_j^2}\right), \tag{21}$$

where D_{ij} is interpreted as $\mathrm{d}\log(x_i/x_j)/\mathrm{d}\log(p_j/p_i)$, for constant output and other input quantities.

(2) *McFadden's Shadow Elasticity of Substitution* S_{ij}, defined through the cost function $c = G(y;\mathbf{p})$. If the unit cost function is given in the "indirect reciprocal production function" form $1/y = g(\mathbf{p})$, it may be shown that $S_{ij}(\mathbf{p})$ is given by

$$S_{ij}(\mathbf{p}) = \left(-\frac{g_{ii}}{g_i^2} + \frac{2g_{ij}}{g_i g_j} - \frac{g_{jj}}{g_j^2}\right) \Big/ \left(\frac{1}{p_i g_i} + \frac{1}{p_j g_j}\right), \tag{22}$$

where S_{ij} is interpreted as $\mathrm{d}\log(x_i/x_j)/\mathrm{d}\log(p_j/p_i)$, for other prices, output and unit cost held constant. Thus,

$$S_{ji} = S_{ij} = \frac{\mathrm{d}\log(G_j/G_i)}{\mathrm{d}\log(p_i/p_j)} = \frac{\mathrm{d}\log(g_j/g_i)}{\mathrm{d}\log(p_i/p_j)},$$

which is analogous to

$$\frac{1}{D_{ij}} = \frac{\mathrm{d}\log(f_j/f_i)}{\mathrm{d}\log(x_i/x_j)}.$$

Applying the cost polar transformation, the polar direct and indirect production functions are $y = g(\mathbf{x})$ and $1/y = f(\mathbf{p})$, respectively; hence the elasticities $\bar{D}_{ij}$, $\bar{S}_{ij}$ of the polar function satisfy $\bar{D}_{ij}(\mathbf{x}) = 1/S_{ij}(\mathbf{x})$; $\bar{S}_{ij}(\mathbf{p}) = 1/D_{ij}(\mathbf{p})$, where $S_{ij}(\ \ )$ and $D_{ij}(\ \ )$ are the functions defined by (22) and (21), respectively.

For example, the D_{ij} for CRESH [Hanoch (1971)] are given by [25]

$$D_{ij} = a_i a_j/(s_i a_i + s_j a_j),$$

where $s_i = p_i x_i / \sum p_l x_l$ are the cost shares.

Thus, the $\bar{S}_{ij}$ for the CDE polar function (Chapter II.3) are given by

$$\bar{S}_{ij} = (s_i a_i + s_j a_j)/a_i a_j = s_i(1/a_j) + s_j(1/a_i)$$

(since the cost shares s_l are symmetric in $\mathbf{x}$ and $\mathbf{p}$).

Similar applications may be used for the generalized elasticities of transformation and the profit polar production frontier. Examples of a number of particular polar pairs of functional forms are presented and

[25]See Hanoch (1971, p. 12, n. 2), and Hanoch (Chapter II.3).

discussed in Chapter II.3. Other widely used polar pairs of production function are: Diewert's (1971) Generalized Linear and Generalized Leontief Production Functions; the Transcendental-logarithmic (Trans-log) models of Christensen, Jorgenson and Lau (1973, 1975), and the polar pair of Quadratic functions [e.g., Lau (1974)]. The results of the present analysis, however, imply the existence and validity of the polar function generated by *any* functional form used previously, either in the direct or in the indirect mode. Thus the available choice of functional forms in production models is considerably enriched.

Chapter I.3

APPLICATIONS OF PROFIT FUNCTIONS

LAWRENCE J. LAU*

Stanford University

1. The Profit Function – An Alternative Derivation

1.1. *Introduction*

In a pioneering attempt, McFadden (1966) extends the concept of cost functions to revenue functions and profit functions and proves for the first time the McFadden Duality Theorem – the profit function analog of the Shephard (1953) – Uzawa (1964) Duality Theorem on cost and production functions. The purpose of this chapter is to provide an alternative derivation under conditions which guarantee twice differentiability of both the production function and the corresponding dual profit function, to characterize equivalent structural properties of the production function and the profit function, and to propose a variety of econometric applications of the profit function.

*This chapter is an extensively revised version of "Some Applications of Profit Functions", Memorandum 86A and 86B, Center for Research in Economic Growth, Stanford University, November 1969. The author is indebted to Professor Daniel McFadden for his very penetrating comments and extremely helpful suggestions which improved the chapter substantially. He has drawn heavily from his collaborative work with D.W. Jorgenson and P.A. Yotopoulos. Thanks are also due to Professors L.R. Christensen, W.E. Diewert, A.K. Dixit, W.M. Gorman, R.E. Hall, Z. Griliches, G. Hanoch, M. Nerlove and R.W. Shephard for useful discussion at various stages of this work. Financial support from the Labor Allocation Project of the Institute of International Studies, University of California, Berkeley, and the National Science Foundation through research grants to the Institute for Mathematical Studies in the Social Sciences, Stanford University, is gratefully acknowledged. Responsibility for remaining errors rests solely with the author.

Let

$$Y = F(X_1,\dots,X_m;Z_1,\dots,Z_n)$$

be the production function of a firm, where the X_i's and the Z_j's are the variable and the fixed inputs respectively. Then short-run profit, defined as revenue less variable costs, is given by

$$\begin{aligned} P &= pF(\mathbf{X},\mathbf{Z}) - \sum_{i=1}^{m} q_i^* X_i \\ &= p\left[F(\mathbf{X},\mathbf{Z}) - \sum_{i=1}^{m} q_i X_i\right] \\ &= p\,[F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}] \end{aligned}$$

where

p = nominal (money) price of output,
q_i^* = nominal price of input i,
$q_i = q_i^*/p$, normalized price of input i,

and $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{q}$ are the vectors of X_i's, Z_j's and q_i's, respectively.

It is assumed that the objective of productive activity is the maximization of short-run profit and that the firm is a price-taker in the output and variable inputs markets. Thus, the firm maximizes profit with respect to $\mathbf{X}$ taking p, $\mathbf{q}^*$ and $\mathbf{Z}$ as given. The *profit function* Π is a function of p, $\mathbf{q}^*$ and $\mathbf{Z}$ which gives for each set of values p, $\mathbf{q}^*$, $\mathbf{Z}$ the *maximized* value of profit

$$\Pi(p,\mathbf{q}^*;\mathbf{Z}) = p[F(\mathbf{X}^*,\mathbf{Z}) - \mathbf{q}'\mathbf{X}^*],$$

where the X_i^*'s are the optimized quantities of the variable inputs.

Before proceeding further, one may observe that maximization of profit is equivalent to the maximization of normalized profit, P^*,[1] defined by

$$P^* = P/p = F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X},$$

so that the X_i^*'s are identical for the two problems. It is clear that the corresponding normalized profit function is given by

$$\begin{aligned} \Pi^* &= F(\mathbf{X}^*,\mathbf{Z}) - \mathbf{q}'\mathbf{X}^* \\ &= G(\mathbf{q},\mathbf{Z}). \end{aligned}$$

The normalized profit function $G(\mathbf{q},\mathbf{Z})$ is more convenient to work with

[1]This was referred to as the "Unit-Output-Price" or "UOP" profit in Lau (1969c). The terminology "normalized profit" is due to Jorgenson and Lau (1974a and 1974b).

for the purposes at hand but the one-to-one correspondence between $\Pi(p,\mathbf{q}^*,\mathbf{Z})$ and $G(\mathbf{q},\mathbf{Z})$ should be obvious.

1.2. *Properties of the Production Function*

The production function is assumed to have certain properties. Let $\bar{R}^n_+$ and $\bar{R}^m_+$ denote the closed non-negative orthants of R^n and R^m, and R^m_+ the interior of the non-negative orthant of R^m. The assumptions on the production function are as follows:

(F.1) *Domain.* F is a finite, non-negative, real-valued function defined on $\bar{R}^m_+ \times \bar{R}^n_+$. For each $\mathbf{Z} \in \bar{R}^n_+$, $F(\mathbf{0},\mathbf{Z}) = 0$.

(F.2) *Continuity.* F is continuous on $\bar{R}^m_+ \times \bar{R}^n_+$.

(F.3) *Smoothness.* For each $\mathbf{Z} \in \bar{R}^n_+$, F is continuously differentiable on R^m_+, and the Euclidean norm of the gradient of F with respect to $\mathbf{X}$ is unbounded for any sequence of $\mathbf{X}$ in R^m_+ converging to a boundary point of $\bar{R}^m_+$. For each $\mathbf{X} \in \bar{R}^m_+$, F is continuously differentiable on R^n_+.

For each $\mathbf{Z} \in \bar{R}^n_+$, the gradient of F with respect to $\mathbf{X}$ on R^m_+ will be denoted $\nabla_X F(\mathbf{X},\mathbf{Z})$.

(F.4) *Monotonicity.* F is non-decreasing in $\mathbf{X}$ and $\mathbf{Z}$ on $\bar{R}^m_+ \times \bar{R}^n_+$ and strictly increasing in $\mathbf{X}$ and $\mathbf{Z}$ on $R^m_+ \times R^n_+$.

(F.5) *Concavity.* For each $\mathbf{Z} \in \bar{R}^n_+$, F is concave on $\bar{R}^m_+$ and locally strongly concave on R^m_+.

Definition. A function is *strongly concave* on a convex set C if there exists $\delta > 0$ such that[2]

$$F((1-\lambda)\mathbf{X}_1 + \lambda \mathbf{X}_2) \geqq (1-\lambda)F(\mathbf{X}_1) + \lambda F(\mathbf{X}_2) + \lambda(1-\lambda)\delta(\mathbf{X}_1 - \mathbf{X}_2)'(\mathbf{X}_1 - \mathbf{X}_2), \quad 0 \leqq \lambda \leqq 1, \quad \forall \mathbf{X}_1, \mathbf{X}_2 \in C.$$

An example of a strongly concave function is $F(X) = -X^2$. A function is *locally strongly concave* if there exists such a δ for every proper convex subset of C. An example of a locally strongly concave (but not strongly

[2]See Roberts and Varberg (1973, p. 268).

concave) function on $\bar{R}_+$ is $F(X) = 1 - e^{-X}$.[3] Obviously local strong concavity implies strict concavity.

(F.6) *Twice Differentiability.* For each $\mathbf{Z} \in \bar{R}^n_+$, F is twice continuously differentiable on R^m_+.

The concavity and twice differentiability assumptions together imply that for each $\mathbf{Z} \in \bar{R}^n_+$ the Hessian matrix of F with respect to $\mathbf{X}$ is negative definite on R^m_+.

(F.7) *Boundedness.* For each $\mathbf{Z} \in \bar{R}^n_+$,

$$\lim_{\lambda \to \infty} \frac{F(\lambda \mathbf{X}, \mathbf{Z})}{\lambda} = 0 \qquad \forall \mathbf{X} \in \bar{R}^m_+.$$

The boundedness assumption ensures that a bounded and attainable solution exists for the normalized profit maximization problem for all $\mathbf{q} \in R^m_+$. This assumption is sufficient even if the production function is not differentiable, that is, in the absence of (F.3) and (F.6). For the purpose at hand, one may have adopted for (F.7) the alternative assumption that for each $\mathbf{Z} \in \bar{R}^n_+$, the range of $\nabla_X F(\mathbf{X}, \mathbf{Z})$ is all of R^m_+.

An example of a function for which (F.7) fails is $F(X) = X + X^{1/2}$. For this production function there does not exist a profit maximum if $q \leqq 1$. An example of a function which satisfies (F.7), but fails (F.3) is $F(X) = 1 - e^{-X}$.

Assumptions (F.1) through (F.6) are sufficient to ensure that, if a solution $\mathbf{X}^*$ to the normalized profit maximization problem exists for a given $\mathbf{q}$ and $\mathbf{Z}$, the solution will be unique and lies in R^m_+. The additional Assumption (F.7) is needed to ensure that such a solution exists for arbitrary $\mathbf{q} \in R^m_+$ and $\mathbf{Z} \in \bar{R}^n_+$. We therefore have the following two lemmas:

Lemma I-1. Under Assumptions (F.1) through (F.6), for each $\mathbf{q} \in R^m_+$, $\mathbf{Z} \in \bar{R}^n_+$, if a vector $\mathbf{X}^*$ exists such that

$$F(\mathbf{X}^*, \mathbf{Z}) - \sum_{i=1}^{m} q_i X_i^* \geqq F(\mathbf{X}, \mathbf{Z}) - \sum_{i=1}^{m} q_i X_i, \qquad \forall \mathbf{X} \in \bar{R}^m_+,$$

then $\mathbf{X}^*$ is unique and lies in R^m_+.

[3] Under the additional assumption of twice differentiability, local strong concavity implies negative definiteness of the Hessian matrix. Compare the concept of differential strict quasi-concavity which implies that the Hessian matrix is negative semi-definite with rank $(m - 1)$. See Chapter I.1.

Proof: The Kuhn–Tucker necessary condition for a maximum implies that

$$\nabla_X F(\mathbf{X}^*,\mathbf{Z}) \leqq \mathbf{q},$$

with equality in each component of $\nabla_X F$ for which the corresponding component of $\mathbf{X}^*$ is positive. However, if any component of $\mathbf{X}^*$ is zero, by (F.3) and (F.4) $\nabla_X F$ is unbounded and positive, thus violating the Kuhn–Tucker condition. Hence $\mathbf{X}^*$ must be positive and lies in R^m_+. Finally, by (F.5) $\mathbf{X}^*$ must be unique. Q.E.D.

Lemma I-2. Under Assumptions (F.1) through (F.7), for each $\mathbf{q} \in R^m_+$, $\mathbf{Z} \in \bar{R}^n_+$, there exists a unique vector

$$\mathbf{X}^* = \mathbf{X}^*(\mathbf{q},\mathbf{Z}) \in R^m_+$$

such that

$$F(\mathbf{X}^*,\mathbf{Z}) - \mathbf{q}'\mathbf{X}^* \geqq F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}, \qquad \forall \mathbf{X} \in \bar{R}^m_+.$$

Further, the function $\mathbf{X}^*(\mathbf{q},\mathbf{Z}): R^m_+ \times \bar{R}^n_+ \to R^m_+$ is continuous on R^m_+ for each $\mathbf{Z} \in \bar{R}^n_+$ and continuous on $\bar{R}^n_+$ for each $\mathbf{q} \in R^m_+$. For each $\mathbf{Z} \in \bar{R}^m_+$, $\mathbf{X}^*$ is continuously differentiable on R^m_+.[4]

Proof: Under Assumptions (F.1) through (F.6), for each $\mathbf{Z} \in \bar{R}^n_+$, normalized profit, $P^* = F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}$, is a closed, proper concave function in $\mathbf{X}$ on $\bar{R}^m_+$ for all $\mathbf{q} \in R^m_+$.[5] For a given $\mathbf{Z}$ and $\mathbf{q}$, a finite and attainable maximum exists for this closed, proper, concave function if and only if the function P^* has no directions of recession in $\mathbf{X}$. [See Rockafellar (1970, Theorem 27.1, pp. 264–265; and also Theorem 13.3 and its corollaries, pp. 116–117).] The directions of recession of P^* are the vectors $\mathbf{y} \neq 0$, $\mathbf{y} \in \text{dom}\, P^*$ (domain of P^*), such that

$$\lim_{\lambda\to\infty} P^* \frac{(\lambda \mathbf{y},\mathbf{Z})}{\lambda} \geqq 0.$$[6]

Thus, in order for P^* to have no directions of recession one must have

$$\lim_{\lambda\to\infty} \frac{F(\lambda\mathbf{X},\mathbf{Z})}{\lambda} - \mathbf{q}'\mathbf{X} < 0, \qquad \forall \mathbf{X} \in \text{dom}\, F, \quad \mathbf{X} \neq 0.$$

Since $q \in R^m_+$, $\text{dom}\, F = \bar{R}^m_+$, $q'X > 0$. Thus, one concludes that

[4]See Appendix A.3, Lemma 15.4.

[5]A concave extended-real-valued function is proper if it nowhere takes the value of ∞ and is finite for at least one value of its arguments.

[6]See Rockafellar (1970, pp. 66–67, and especially Theorem 8.5 and its corollaries).

$\lim_{\lambda\to\infty} F(\lambda\mathbf{X},\mathbf{Z})/\lambda \leqq 0$ is necessary, But $F(\lambda\mathbf{X},\mathbf{Z}) \geqq 0$, by (F.1), thus $\lim_{\lambda\to\infty} F(\lambda\mathbf{X},\mathbf{Z})/\lambda = 0$ is necessary and sufficient to ensure that no direction of recession exists. Hence, with (F.7) a finite and attainable solution exists. But this argument works for all $\mathbf{q} \in R^m_+$. Thus, for all $\mathbf{q} \in R^m_+$, a finite and attainable solution exists.[7] By Lemma I-1, the optimal solution $\mathbf{X}^*$ is positive and unique.

Continuity properties of $\mathbf{X}^*$ follow from the continuity of $\nabla_X F$ on R^m_+ for each $\mathbf{Z} \in \bar{R}^n_+$ and on $\bar{R}^n_+$ for every $\mathbf{X} \in R^m_+$.

Implicit differentiation using the implicit function theorem guarantees the differentiability property of $\mathbf{X}^*$. The assumption of non-singularity of the Jacobian matrix so crucial in the application of the implicit function theorem is implied by the negative definiteness of the Hessian matrix of F with respect to $\mathbf{X}$. Q.E.D.

Corollary 2.1. The normalized profit function $G(\mathbf{q},\mathbf{Z}) = F(\mathbf{X}^*(\mathbf{q},\mathbf{Z}),\mathbf{Z}) - \mathbf{q}'\mathbf{X}^*(\mathbf{q},\mathbf{Z})$ is continuous on $R^m_+ \times \bar{R}^n_+$, is twice continuously differentiable on R^m_+ for each $\mathbf{Z} \in \bar{R}^n_+$, and is continuously differentiable on R^n_+ for each $\mathbf{q} \in R^m_+$. $Y^*(\mathbf{q},\mathbf{Z}) = F(\mathbf{X}^*(\mathbf{q},\mathbf{Z}),\mathbf{Z})$ is continuous on $R^m_+ \times \bar{R}^n_+$ and is continuously differentiable on R^m_+ for each $\mathbf{Z} \in \bar{R}^n_+$.

Proof: The proof follows from repeated application of the chain rule for partial differentiation and the fact that

$$\frac{\partial F}{\partial \mathbf{X}}(\mathbf{X}^*,\mathbf{Z}) - \mathbf{q} = 0 \quad \text{Q.E.D.}$$

1.3. *Duality*

The duality between production functions and normalized profit functions has been established under rather general circumstances in Chapter I.1 and elsewhere.[8] Our purpose here is to establish properties of the class of normalized profit functions which correspond to the class of production functions which satisfies our Assumptions (F.1) through (F.7) and to demonstrate that there exists a one-to-one correspondence

[7]Note that this establishes the domain of G as all of R^m_+. See Chapter I.1.

[8]See also Cass (1974), Diewert (1973a and 1974a), Jorgenson and Lau (1974a and 1974b), and Lau (1976a). Jorgenson and Lau base their duality results on the conjugacy correspondence of closed, proper convex functions.

between the members of the two classes. For every production function which satisfies Assumptions (F.1) through (F.7) one can define a normalized profit function $G(\mathbf{q},\mathbf{Z})$ on $R_+^m \times \bar{R}_+^n$

$$G(\mathbf{q},\mathbf{Z}) = \sup_{\mathbf{X}}\{F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}\}.$$

For each $\mathbf{Z} \in \bar{R}_+^n$, $G(\mathbf{q},\mathbf{Z})$ is also referred to as the conjugate of $F(\mathbf{X},\mathbf{Z})$. By Lemma I-2, a finite and attainable maximum always exists for $q \in R_+^m$. Thus,

$$G(\mathbf{q},\mathbf{Z}) = \max_{\mathbf{X}}\{F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}\}.$$

It will be shown that the normalized profit function $G(\mathbf{q},\mathbf{Z})$ corresponding to a production function satisfying Assumptions (F.1) through (F.7) possesses the following properties:

(G.1) *Domain.* G is a finite, positive, real-valued function defined on $R_+^m \times \bar{R}_+^n$.

(G.2) *Continuity.* G is continuous on $R_+^m \times \bar{R}_+^n$.

(G.3) *Smoothness.* For each $\mathbf{Z} \in \bar{R}_+^n$, G is continuously differentiable on R_+^m, and the Euclidean norm of the gradient of G with respect to $\mathbf{q}$ is unbounded for any sequence of $\mathbf{q}$ in R_+^m converging to a boundary point of $\bar{R}_+^m$. For each $\mathbf{q} \in R_+^m$, F is continuously differentiable on R_+^n.

(G.4) *Monotonicity.* $G(\mathbf{q},\mathbf{Z})$ is non-increasing $\mathbf{q}$ and non-decreasing in $\mathbf{Z}$ on $R_+^m \times \bar{R}_+^n$ and strictly decreasing in $\mathbf{q}$ and strictly increasing in $\mathbf{Z}$ on $R_+^m \times R_+^n$.

(G.5) *Convexity.* For each $\mathbf{Z} \in \bar{R}_+^n$, $G(\mathbf{q},\mathbf{Z})$ is locally strongly convex on R_+^m.

Definition. A function F is locally strongly convex if $-F$ is locally strongly concave.

(G.6) *Twice Differentiability.* For each $\mathbf{Z} \in \bar{R}_+^n$, $G(\mathbf{q},\mathbf{Z})$ is twice continuously differentiable on R_+^m.

The convexity and twice differentiability assumptions together imply that for each $\mathbf{Z} \in \bar{R}_+^n$ the Hessian matrix of G with respect to $\mathbf{q}$ is positive definite on R_+^m.

(G.7) *Boundedness.* For each $\mathbf{Z} \in \bar{R}_+^n$,

$$\lim_{\lambda \to \infty} \frac{G(\lambda \mathbf{q}, \mathbf{Z})}{\lambda} = 0, \qquad \forall \mathbf{q} \in R_+^m.$$

Lemma I-3. Under Assumptions (F.1) through (F.7), the normalized profit function satisfies Assumptions (G.1) through (G.7).

Proof:

(G.1) From Lemma I-2, $G(\mathbf{q},\mathbf{Z})$ is a finite and real-valued function defined on $R_+^m \times \bar{R}_+^n$ since a finite and attainable maximum exists. And because for each $\mathbf{Z} \in \bar{R}_+^n$, $F(\mathbf{0},\mathbf{Z}) = 0$, $G(\mathbf{q},\mathbf{Z}) \geqq 0$. If $G(\mathbf{q},\mathbf{Z}) = 0$ for any $\mathbf{q} \in R_+^m$, then a profit-maximizing vector $\mathbf{X}^*$ is $\mathbf{X}^* = 0$. However, this contradicts Lemma I-2, which states that $\mathbf{X}^* \in R_+^m$. Thus $G(\mathbf{q},\mathbf{Z})$ is positive.

(G.2) Since $G(\mathbf{q},\mathbf{Z}) = \max_{\mathbf{X}}\{F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}\}$ it follows that for each $\mathbf{q} \in R_+^m$, $G(\mathbf{q},\mathbf{Z})$ is continuous in $\mathbf{Z}$ on $\bar{R}_+^n$ by (F.2). In addition, for each $\mathbf{Z} \in \bar{R}_+^n$, $G(\mathbf{q},\mathbf{Z})$ is convex on R_+^m [proved under (G.5) below]. Thus, by a theorem in Rockafellar (1970, Theorem 10.7, pp. 89–90), $G(\mathbf{q},\mathbf{Z})$ is continuous on $R_+^m \times \bar{R}_+^n$.

(G.3) Smoothness in $\mathbf{q}$ is implied by (F.5) [see Rockafellar (1970, Theorem 26.3, pp. 253–254)]. Differentiability in $\mathbf{Z}$ follows from the fact that $F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}$ is continuously differentiable in $\mathbf{Z}$ on R_+^n and that $G(\mathbf{q},\mathbf{Z}) = \max_{\mathbf{X}}\{F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}\}$.

(G.4) Let $\mathbf{X}_1^*$ and $\mathbf{X}_2^* \in R_+^m$ be the profit-maximizing inputs at $\mathbf{q}_1$ and $\mathbf{q}_2$, respectively. Then,

$$G(\mathbf{q}_1,\mathbf{Z}) > F(\mathbf{X}_2^*,\mathbf{Z}) - \mathbf{q}_1'\mathbf{X}_2^*,$$
$$G(\mathbf{q}_2,\mathbf{Z}) > F(\mathbf{X}_1^*,\mathbf{Z}) - \mathbf{q}_2'\mathbf{X}_1^*.$$

Suppose $\mathbf{q}_1$ is strictly greater than $\mathbf{q}_2$ (in at least one component) then $F(\mathbf{X}_1^*,\mathbf{Z}) - \mathbf{q}_2'\mathbf{X}_1^* > F(\mathbf{X}_1^*,\mathbf{Z}) - \mathbf{q}_1'\mathbf{X}_1^* = G(\mathbf{q}_1,\mathbf{Z})$. Hence $G(\mathbf{q}_2,\mathbf{Z}) > G(\mathbf{q}_1,\mathbf{Z})$. Monotonicity in $\mathbf{Z}$ follows from the fact that $G(\mathbf{q},\mathbf{Z}) = \max_{\mathbf{X}}\{F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}\}$.

(G.5) Local strong convexity is implied by (F.3) and (F.5).

(G.6) Twice differentiability is implied by (F.5) and (F.6) through the chain rule.

(G.7) Boundedness follows from the fact that for each $\mathbf{Z} \in \bar{R}^n_+$, the domain of $F(\mathbf{X},\mathbf{Z})$ is $\bar{R}^m_+$, the support function of which is given by

$$\begin{aligned} \delta^*(\mathbf{X}^*|\bar{R}^m_+) &= 0, \quad \mathbf{X}^* \in \bar{R}^m_-, \\ &= +\infty, \quad \mathbf{X}^* \notin \bar{R}^m_-. \end{aligned}$$

But this is also the recession function of the conjugate of $F(\mathbf{X},\mathbf{Z})$, $G(\mathbf{q},\mathbf{Z})$, with $\mathbf{q}$ identified with $-\mathbf{X}^*$ [see Rockafellar (1970, Theorem 13.3, p. 116)]. The recession function of $G(\mathbf{q},\mathbf{Z})$ is given by

$$\lim_{\lambda \to \infty} \frac{G(\lambda \mathbf{q},\mathbf{Z})}{\lambda}, \qquad \mathbf{q} \in R^m_+.$$

Thus, one has

$$\lim_{\lambda \to \infty} \frac{G(\lambda \mathbf{q},\mathbf{Z})}{\lambda} = 0, \qquad \mathbf{q} \in R^m_+. \quad \text{Q.E.D.}$$

Given a normalized profit function $G(\mathbf{q},\mathbf{Z})$ one may define its conjugate as

$$F^*(\mathbf{X},\mathbf{Z}) = \inf_{\mathbf{q}} \{G(\mathbf{q},\mathbf{Z}) + \mathbf{q}'\mathbf{X}\}.$$

Under Assumptions (G.1) through (G.7), $G(\mathbf{q},\mathbf{Z})$ is a closed proper convex function on $R^m_+ \times \bar{R}^n_+$, hence its conjugate function is unique and equal to $F(\mathbf{X},\mathbf{Z})$ itself. [For the one-to-one correspondence between closed proper convex functions and its conjugate, see Chapter I.1 and Rockafellar (1970, ch. 12).] Hence all one needs to do is to verify that $F^*(\mathbf{X},\mathbf{Z})$ in fact satisfies Assumption (F.1) through (F.7). Thus one has:

Lemma I-4. Under Assumptions (G.1) through (G.7), the production function satisfies Assumptions (F.1) through (F.7).

Proof: The proof parallels the proof of Lemma I-3. The only exception is that of continuity of $F(\mathbf{X},\mathbf{Z})$ on the boundary of R^m_+. This follows from the fact that $F(\mathbf{X},\mathbf{Z}) = \inf_{\mathbf{q}}\{G(\mathbf{q},\mathbf{Z}) + \mathbf{q}'\mathbf{X}\}$ is a closed proper concave function and bounded below on every bounded subset of R^m_+. Hence, $F(\mathbf{X},\mathbf{Z})$ may be uniquely extended to a continuous finite concave function on $\bar{R}^m_+$. [See Rockafellar (1970, pp. 84–86) and also Lemma 12.7

in Appendix A.3 of this volume.] It is then possible to set $F(\mathbf{0},\mathbf{Z}) = 0$. Q.E.D.

We conclude this section by noting that it is possible to relax the assumption that the domain of $F(\mathbf{X},\mathbf{Z})$ is all of $\bar{R}^m_+$, or that the range of $\nabla_X F$ is all of R^m_+ as is done in Chapter I.1 and Jorgenson and Lau (1974a and 1974b). It is also possible to relax the assumption that $|\nabla_X F|$ becomes unbounded as X approaches the boundary of its domain from the interior, requiring only that the range of $\nabla_X F$ on the domain of F has a non-empty interior. Under these mild modifications, the properties of continuity, differentiability, monotonicity, concavity and twice differentiability still imply corresponding properties on the dual, only that the domains of definition are now a pair of open convex sets C and C^*, such that $C \subset \text{int}(\text{dom } F)$ and $C^* \subset \text{int } D$ where D is the range of $\nabla_X F$ on C.[9]

1.4. *The Legendre Transformation*

One way of obtaining the normalized profit function is to solve the maximization problem first for the derived demand functions and then substitute these back into the formula for normalized profit given by

$$P^* = [F(\mathbf{X},\mathbf{Z}) - \mathbf{q}'\mathbf{X}].$$

The difficulty with this method is that only for relatively simple production functions can one solve the profit-maximization problem explicitly to obtain closed form solutions for the derived demand functions. An alternative method for constructing the normalized profit function and for studying the behavior of the normalized profit function (without actually constructing it), based on the classical Legendre transformation,[10] will be given below.

The Legendre transformation is a change of variables of a function from point coordinates to plane coordinates. It is based partly on the notion that a system of partial differential equations may be used to define two or more sets of functions through transformation of variables. In the present case, it can be shown that the production function

[9]For further discussion of this point, see Rockafellar (1970, pp. 251–260).

[10]A succinct exposition of the Legendre transformation may be found in Lanczos (1966, Ch. VI). See also Courant and Hilbert (1953, Vol. II, Ch. I, pp. 32–39). For a discussion from a more modern point of view, see Rockafellar (1970).

and the normalized profit function are connected by the Legendre transformation.

Consider a given function of m variables V_i's and n parameters p_i's,

$$f = (V_1,\ldots,V_m;p_1,\ldots,p_n),$$

new variable T_i's may be introduced by means of the following transformation:

$$T_i = \frac{\partial f}{\partial V_i}, \qquad i = 1,\ldots,m, \tag{I-1}$$

which is called the Legendre transformation. The variables V_i's are replaced by the variables T_i's. f is assumed to be locally strongly concave in the V_i's so that the transformation is non-singular and hence invertible. Thus equation (I-1) may be solved, expressing the V_i's in terms of the T_i's and p_i's,

$$V_i = h_i(T_1,\ldots,T_m;p_1,\ldots,p_n), \qquad i = 1,\ldots,m.$$

A new function g may be defined as follows:

$$g(T_1,\ldots,T_m;p_1,\ldots,p_n) = \sum_{i=1}^{m} h_i(\mathbf{T},\mathbf{p})T_i - f(h_1(\mathbf{T},\mathbf{p}),\ldots,h_m(\mathbf{T},\mathbf{p});\mathbf{p}).$$

The function g is known as the Legendre's dual transformation of the primal function f.

Observe that

$$\frac{\partial g}{\partial T_i} = \sum_{j=1}^{m} \frac{\partial h_j}{\partial T_i} T_j + h_i - \sum_{j=1}^{m} \frac{\partial f}{\partial V_j}\cdot\frac{\partial h_j}{\partial T_i}, \qquad i = 1,\ldots,m.$$

But by equation (I-1), $T_j = \partial f/\partial V_j$. Thus,

$$\begin{aligned}\frac{\partial g}{\partial T_i} &= h_i(\mathbf{T},\mathbf{p}) \\ &= V_i, \qquad i = 1,2\ldots,m.\end{aligned} \tag{I-2}$$

Equation (I-2) is the inverse Legendre transformation. The variables T_i are replaced by the variables V_i's. In addition, we have

$$\begin{aligned}\frac{\partial g}{\partial p_i} &= \sum_{j=1}^{m} \frac{\partial h_j}{\partial p_i}\cdot T_j - \sum_{j=1}^{m} \frac{\partial f}{\partial V_j}\cdot\frac{\partial h_j}{\partial p_i} - \frac{\partial f}{\partial p_i}, \qquad i = 1,\ldots,n, \\ &= -\frac{\partial f}{\partial p_i},\end{aligned} \tag{I-3}$$

again by equation (I-1).

If we now compute the Legendre transformation of g, we have

$$g^*(V_1,\ldots,V_m;p_1,\ldots,p_n)=\sum_{i=1}^{m} T_i(\mathbf{V},\mathbf{p})\cdot V_i - g(T_1(\mathbf{V},\mathbf{p}),\ldots,T_m(\mathbf{V},\mathbf{p});\mathbf{p})$$
$$= f.$$

The functions f and g are linked by the following set of dual relations:

$$f(V_1,V_2,\ldots,V_m;\mathbf{p})+g(T_1,T_2,\ldots,T_m;\mathbf{p})=\sum_{i=1}^{m} V_iT_i,$$

$$\frac{\partial f}{\partial \mathbf{V}}=\mathbf{T},\qquad \frac{\partial g}{\partial \mathbf{T}}=\mathbf{V},\qquad \frac{\partial f}{\partial \mathbf{p}}+\frac{\partial g}{\partial \mathbf{p}}=0.$$

There is also a set of transformations relating the second derivatives of f and the second derivatives of g. Starting from

$$\frac{\partial f}{\partial \mathbf{V}}=\mathbf{T},$$

one may differentiate this set of dual relations with respect to $\mathbf{T}$ obtaining

$$\left[\frac{\partial V}{\partial \mathbf{T}}\right]\left[\frac{\partial^2 f}{\partial \mathbf{V}\partial \mathbf{V}'}\right]=\left[\frac{\partial^2 g}{\partial \mathbf{T}\partial \mathbf{T}'}\right]\left[\frac{\partial^2 f}{\partial \mathbf{V}\partial \mathbf{V}'}\right]=I.$$

One may also differentiate the set of dual relations with respect to $\mathbf{p}$, obtaining

$$\left[\frac{\partial V}{\partial \mathbf{p}}\right]\left[\frac{\partial^2 f}{\partial \mathbf{V}\partial \mathbf{V}'}\right]+\left[\frac{\partial^2 f}{\partial \mathbf{p}\partial \mathbf{V}'}\right]=0,$$

or

$$\left[\frac{\partial^2 g}{\partial \mathbf{p}\partial \mathbf{T}'}\right]\left[\frac{\partial^2 f}{\partial \mathbf{V}\partial \mathbf{V}'}\right]+\left[\frac{\partial^2 f}{\partial \mathbf{p}\partial \mathbf{V}'}\right]=0.$$

We note that this is also a symmetric relation because

$$\left[\frac{\partial^2 g}{\partial \mathbf{T}\partial \mathbf{T}'}\right]=\left[\frac{\partial^2 f}{\partial \mathbf{V}\partial \mathbf{V}'}\right]^{-1}.$$

In terms of our problem, the production function $F(\mathbf{X},\mathbf{Z})$ may be identified as f. The normalized profit function $G(\mathbf{q},\mathbf{Z})$ may be identified as $-g$. $\mathbf{X}$ may be identified as $\mathbf{V}$. $\mathbf{Z}$ may be identified as $\mathbf{p}$. The new variables to be introduced – the plane coordinates – are set equal to

$$\mathbf{T}=\frac{\partial F}{\partial \mathbf{X}},$$

in accordance with the Legendre transformation. However, $\partial F/\partial \mathbf{X} = q$ under the assumption of profit maximization. Thus

$$\mathbf{T} = \frac{\partial F}{\partial \mathbf{X}} = \mathbf{q}.$$

and $\mathbf{q}$ may be identified as $\mathbf{T}$. The Legendre transformation may be constructed as

$$g = \sum_{i=1}^{m} T_i X_i(\mathbf{T},\mathbf{Z}) - F(X_1(\mathbf{T},\mathbf{Z}),\ldots,X_m(\mathbf{T},\mathbf{Z}),\mathbf{Z}).$$

By recognizing that $\mathbf{T} = \mathbf{q}$, we have

$$g = \sum_{i=1}^{m} q_i X_i(\mathbf{q},\mathbf{Z}) - F(X_1(\mathbf{q},\mathbf{Z}),\ldots,X_m(\mathbf{q},\mathbf{Z}),\mathbf{Z}),$$

which is precisely equal to $-G$, the negative of the normalized profit function. Moreover, from the inverse Legendre transformation

$$\frac{\partial g}{\partial \mathbf{q}} = \frac{\partial g}{\partial \mathbf{T}} = \mathbf{X},$$

$$\frac{\partial g}{\partial \mathbf{Z}} = -\frac{\partial F}{\partial \mathbf{Z}}.$$

Hence, one has

$$\frac{\partial G}{\partial \mathbf{q}} = -\mathbf{X},$$

$$\frac{\partial G}{\partial \mathbf{Z}} = \frac{\partial F}{\partial \mathbf{Z}}.$$

This set of relations is sometimes referred to as Hotelling's (1932) Lemma and is of crucial importance in applications. We may then summarize the Legendre transformation relationships between the production function and the normalized restricted profit function:

	Primal	Dual
Function:	Production function $F(\mathbf{X},\mathbf{Z})$	Normalized function $G(\mathbf{q},\mathbf{Z})$
Active variables:	$\mathbf{X}$	$\mathbf{q}$
Passive variables:	$\mathbf{Z}$	$\mathbf{Z}$

And we get the following dual transformation relations:

(1)	$F(\mathbf{X},\mathbf{Z}) - G(\mathbf{q},\mathbf{Z}) = \mathbf{q}'\mathbf{X}$	
(2)	$\partial F/\partial \mathbf{X} = \mathbf{q}$;	$\partial G/\partial \mathbf{q} = -\mathbf{X}$
(3)	$\mathbf{X} = -\partial G/\partial \mathbf{q}$;	$\mathbf{q} = \partial F/\partial \mathbf{X}$
(4)	$\partial F/\partial \mathbf{Z} = \partial G/\partial \mathbf{Z}$;	$\partial G/\partial \mathbf{Z} = \partial F/\partial \mathbf{Z}$
(5)	$F = G - \mathbf{q}'(\partial G/\partial \mathbf{q})$;	$G = F - \mathbf{X}'(\partial F/\partial \mathbf{X})$
(6)	$\mathbf{Z}$;	$\mathbf{Z}$

Under our assumptions on $F(\mathbf{X},\mathbf{Z})$, a Legendre transformation always exists. We introduce the Legendre transformation for a number of reasons. First, its use leads to a system of partial differential equations which may be used to either construct the normalized profit function explicitly or to study its behavior, given the production function and the first order necessary conditions for a maximum (and *vice versa*). Second, the Legendre transformation may be used to deduce equivalent structures of the production function and the normalized profit function. If the production function or the normalized profit function satisfies a given partial differential equation defining a certain structural property, then the same partial differential equation must also be satisfied by a Legendre transformation of variables. This is because we have shown that the production function and the normalized profit function are Legendre transformations of each other, hence a partial differential equation for $F(\mathbf{X},\mathbf{Z})$ in $\mathbf{X}$ and $\mathbf{Z}$ becomes a partial differential equation for $G(\mathbf{q},\mathbf{Z})$ in $\mathbf{q}$ and $\mathbf{Z}$. Thus, equivalent properties may be deduced immediately. This technique is used extensively in Sections 2 and 3. Third, the Legendre transformation may be useful in the solution of certain partial differential equations which may prove intractable otherwise. Suppose we wish to establish the class of production functions such that, under profit maximization, X_1/X_2 is constant. Starting from the set of functions

$$\frac{\partial F}{\partial X_1} = q_1 \quad \text{and} \quad \frac{\partial F}{\partial X_2} = q_2,$$

$$\frac{X_1}{X_2} = \frac{k_1}{k_2}.$$

This may appear to be rather intractable. However, by using the Legendre transformation, this problem becomes

$$\frac{\partial G/\partial q_1}{\partial G/\partial q_2} = \frac{k_1}{k_2},$$

with the general solution

$$G(\mathbf{q}) = g(k_1 q_1 + k_2 q_2),$$

which has a well-known dual

$$F(\mathbf{X}) = f\left(\min\left[\frac{X_1}{k_1}, \frac{X_2}{k_2}\right]\right).$$

Another example is furnished by the partial differential equation

$$X_1 = f_1\left(\frac{\partial F}{\partial X_1}, \frac{\partial F}{\partial X_2}\right).$$

By the Legendre transformation, this equation becomes

$$-\frac{\partial G}{\partial q_1} = f_1(q_1, q_2),$$

which may be integrated. This technique is used in Section 5.

We emphasize, however, that the Legendre transformations are procedures for studying expressions that are known to exist; they are not meant to be substitutes for the fundamental existence theorems for the dual functions, which are proved in Chapter I.1 for the general case, and in Sections 1.2 and 1.3 for the locally strongly concave case.

1.5. *Comparative Statics*

We present some comparative statics results that can be obtained directly by making use of the properties of normalized profit functions.

1.5.1. *Increase in nominal price of output*

(i) The optimal output is given by

$$Y^* = G(\mathbf{q},\mathbf{Z}) - \mathbf{q}'\frac{\partial G}{\partial \mathbf{q}}(\mathbf{q},\mathbf{Z}),$$

$$\frac{\partial Y^*}{\partial p} = \sum_k \frac{\partial G}{\partial q_k}\cdot\frac{\partial q_k}{\partial p} - \frac{\partial \mathbf{q}'}{\partial p}\frac{\partial G}{\partial \mathbf{q}}(\mathbf{q},\mathbf{Z}) - \mathbf{q}'\frac{\partial^2 G}{\partial \mathbf{q}\partial \mathbf{q}'}\frac{\partial \mathbf{q}}{\partial p}$$

$$= \frac{1}{p}\mathbf{q}'\frac{\partial^2 G}{\partial \mathbf{q}\partial \mathbf{q}'}\mathbf{q} > 0,$$

since

$$\left[\frac{\partial^2 G}{\partial \mathbf{q} \partial \mathbf{q}'}\right]$$

is positive definite. Thus the effect of output price on supply is positive.

(ii) $$\frac{\partial X_i^*}{\partial p} = \frac{\partial}{\partial p} - \frac{\partial G(\mathbf{q},\mathbf{Z})}{\partial q_i} = \frac{1}{p} \sum_{k=1}^{m} \frac{\partial^2 G}{\partial q_i \partial q_k} \cdot q_k,$$

which is not definite in sign.

1.5.2. *Increase in nominal price of a variable input*

(i) $$\frac{\partial Y^*}{\partial q_i^*} = \frac{1}{p}\left[\frac{\partial G}{\partial q_i}(\mathbf{q},\mathbf{Z}) - \sum_{k=1}^{m} \frac{\partial^2 G}{\partial q_i \partial q_k} q_k - \frac{\partial G}{\partial q_i}(\mathbf{q},\mathbf{Z})\right] = -\frac{1}{p}\left[\sum_{k=1}^{m} \frac{\partial^2 G}{\partial q_i \partial q_k} \cdot q_k\right],$$

which is again not definite in sign, but equal in magnitude but opposite in sign to $\partial X_i^*/\partial p$.

(ii) $$\frac{\partial X_i^*}{\partial q_i^*} = -\frac{1}{p} \frac{\partial^2 G}{\partial q_i^2} < 0,$$

since

$$\left[\frac{\partial^2 G}{\partial \mathbf{q} \partial \mathbf{q}'}\right]$$

is positive definite. Thus, the own price effect of input price on demand is negative.

(iii) $$\frac{\partial X_i^*}{\partial q_j^*} = -\frac{1}{p} \frac{\partial^2 G}{\partial q_i \partial q_j} = \frac{\partial X_j^*}{\partial q_i^*},$$

by the twice continuous differentiability of $G(\mathbf{q},\mathbf{Z})$. This is the well-known symmetry condition on cross-price effects.

(iv) By collecting these comparative statics results, we may derive, in addition, that

$$\sum_{k=1}^{m} \frac{\partial Y^*}{\partial q_i^*} \cdot q_i^* = -\frac{\partial Y^*}{\partial p} \cdot p < 0,$$

and

$$\sum_{k=1}^{m} \frac{\partial X_i^*}{\partial p} \cdot q_i^* = \frac{\partial Y^*}{\partial p} \cdot p > 0.$$

These results summarize the basic Hicksian Laws of Production.[11]

(v) It is important to note a relationship between the Hessian matrices of the production function and the normalized profit function. By differentiating

$$\frac{\partial F}{\partial \mathbf{X}} = \mathbf{q},$$

with respect to $\mathbf{q}$, treating $\mathbf{X}$ as implicit functions of $\mathbf{q}$, we have

$$\left[\frac{\partial \mathbf{X}}{\partial \mathbf{q}}\right]\left[\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{X}'}\right] = I,$$

but

$$\frac{\partial \mathbf{X}}{\partial \mathbf{q}} = -\left[\frac{\partial^2 G}{\partial \mathbf{q} \partial \mathbf{q}'}\right],$$

Thus

$$\frac{\partial^2 G}{\partial \mathbf{q} \partial \mathbf{q}'} = -\left[\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{X}'}\right]^{-1}.$$

Also, by differentiating $\partial F/\partial \mathbf{X} = \mathbf{q}$ with respect to $\mathbf{Z}$, we have

$$\left[\frac{\partial \mathbf{X}}{\partial \mathbf{Z}}\right]\left[\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{X}'}\right] + \left[\frac{\partial^2 F}{\partial \mathbf{Z} \partial \mathbf{X}'}\right] = 0,$$

or

$$\left[\frac{\partial^2 G}{\partial \mathbf{Z} \partial \mathbf{q}'}\right] = \left[\frac{\partial^2 F}{\partial \mathbf{Z} \partial \mathbf{X}'}\right]\left[\frac{\partial^2 F}{\partial \mathbf{X} \partial \mathbf{X}'}\right]^{-1}.$$

[11]See Hicks (1946, App.).

1.6. *Econometric Implementation*

Because of the derivative property of the normalized profit function, sometimes known as Hotelling's (1932) Lemma, namely,

$$\mathbf{X} = -\frac{\partial G}{\partial \mathbf{q}}, \qquad Y = G - \mathbf{q}'\frac{\partial G}{\partial \mathbf{q}},$$

the normalized profit function is especially useful for the purpose of econometric specification of supply and demand functions. With the normalized profit function, it is not necessary to actually solve a profit maximization problem. As long as one starts out with a normalized profit function which satisfies Assumptions (G.1) through (G.7), one is assured that the supply and demand functions obtained through differentiation of G are consistent with profit maximization subject to a production function and given normalized prices. In particular, since one is free to choose the functional form of $G(q,Z)$, one may choose a parametric form that is most convenient from the point of view of econometric estimation.

There are two other points worth mentioning. First, as McFadden has stressed, convexity of the profit function is a consequence of profit maximization and does not depend at all on the concavity of the production function, so long as a proper profit function exists and is attainable for at least one set of prices. Hence, if one is willing to maintain the assumption of profit maximization, it is not necessary to insist that the production function is concave. Second, for the purpose of estimating the normalized profit function parameters, one should use all of the stochastically independent supply and demand functions for maximum efficiency. This in general entails, because of symmetry of cross-price effects, restrictions across equations.

Finally, one should also add at this point that for many empirical applications in which the observed range of normalized prices is a compact and convex set, it may not be necessary to require that the normalized profit function should satisfy Assumptions (G.1) through (G.7) globally, that is, for all possible prices. It is in many instances sufficient to have the Assumptions (G.1) through (G.7) hold locally within a compact and convex set. As long as interest is focused on this convex set, a normalized profit function, although not globally valid, may nevertheless provide an adequate local approximation. In particular, one can often modify such a function so that it satisfies globally the weak regularity conditions for normalized profit functions given in Chapter I.1.

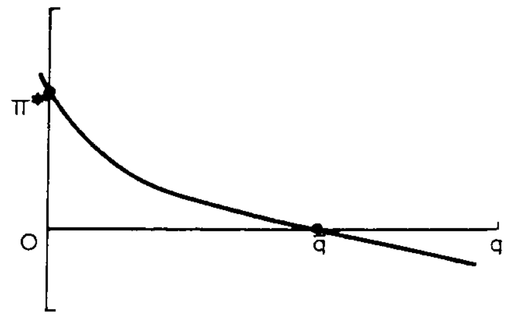


FIGURE 1

We shall illustrate the modification technique with an example. Suppose that the normalized profit function has the form shown in Figure 1 – sloping downward all the way. This function is decreasing and convex in q, but it is not non-negative as a normalized profit function should be. It is also defined for negative prices. One may modify this function so that

$$\Pi^* \text{ is not defined for } \quad q < 0,$$
$$\Pi^* = 0 \quad \text{for} \quad q \geqq \bar{q}.$$

With this modification, the normalized profit function satisfies the usual regularity conditions (such as, for instance, those given in Chapter I.1). As long as the domain of interest is contained in the open interval $(0, \bar{q})$ this normalized profit function will serve just as well as other normalized profit functions which satisfy the regularity conditions globally without modifications of the type considered here.

2. The Structure of Normalized Profit Functions

2.1. *The Case of a Single Output*

For purposes of applications, it is useful to know what are equivalent properties for the production function and the corresponding normalized profit function. To this end, we state and prove several theorems relating equivalent structures of production functions and normalized profit functions.

Theorem II-1. Under Assumptions (F.1) through (F.7), a production function is homogeneous of degree k in $\mathbf{X}$ if and only if the normalixed profit function is homogeneous of degree $-(k/(1-k))$ in $\mathbf{q}$.

Definition. A function is *homogeneous of degree k* in **X** if

$$F(\lambda \mathbf{X},\mathbf{Z}) = \lambda^k F(\mathbf{X},\mathbf{Z}), \quad \text{for} \quad \forall \lambda > 0, \quad \forall \mathbf{Z} \in \bar{R}^n_+, \quad \forall \mathbf{X} \in \bar{R}^m_+.$$

Proof: This follows directly from the dual transformation properties. By Euler's Theorem for homogeneous functions,

$$\sum_{i=1}^{m} \frac{\partial F}{\partial X_i} X_i = kF. \tag{II-1}$$

Applying the dual transformation, equation (II-1) becomes

$$-\sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i} = k\left(G - \sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i}\right).$$

Therefore,

$$\sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i} = -\frac{k}{(1-k)} G.$$

Hence, by Euler's Theorem, G is homogeneous of degree $-k/(1-k)$ in $\mathbf{q}$. The converse is proved similarly. Observe that the case $k \geqq 1$ violates the local strong concavity assumption. Q.E.D.

Corollary 1.1. Under Assumptions (F.1) through (F.7), and homogeneity of degree k of $F(\mathbf{X},\mathbf{Z})$ in $\mathbf{X}$,

$$Y^* = (1-k)^{-1} G,$$

and

$$C^* = p \frac{k}{(1-k)} G,$$

where C^* is the profit-maximizing cost of the variable inputs.

Proof: By the dual transformation

$$\begin{aligned} Y^* &= \left(G - \sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i}\right) \\ &= G + \frac{k}{(1-k)} G \\ &= (1-k)^{-1} G \end{aligned}$$

$$C^* = P \sum_{i=1}^{m} q_i \left(-\frac{\partial G}{\partial q_i}\right) = p \frac{k}{(1-k)} G. \quad \text{Q.E.D.}$$

This corollary implies that for a homogeneous production function the profit-maximizing output is proportional to normalized profit. In other words, profit-maximizing revenue is proportional to profit-maximizing profit. Likewise profit-maximizing cost is also proportional to profit-maximizing profit. These are clearly testable consequences of the homogeneity assumption.

Corollary 1.2. Under Assumptions (F.1) through (F.7), the derived demand functions are homogeneous of degree $-1/(1-k)$ in $\mathbf{q}$ if the production function is homogeneous of degree k in $\mathbf{X}$.

Proof: This follows directly from the fact that the demand functions are derivatives of the normalized profit function, which is homogeneous of degree $-(k/(1-k))$. Q.E.D.

The concept of homogeneity has been generalized by Shephard (1953 and 1970) to that of homotheticity. We give a definition that is closely related but slightly different from his.

Definition. A function is *homothetic* in $\mathbf{X}$ if it can be written in the form

$$F(H(\mathbf{X},\mathbf{Z}),\mathbf{Z}),$$

where for each $\mathbf{Z} \in \bar{R}^n_+$, F is a positive, finite, continuous and strictly monotonic function of one variable H with $F(0,\mathbf{Z}) = 0$, and H is a homogeneous function of degree one in $\mathbf{X}$.

An important property of homothetic functions is the following:

Lemma II-1. A function with strictly non-zero first partial derivatives is homothetic in $\mathbf{X}$ if and only if the ratio of each possible pair of first partial derivatives with respect to $\mathbf{X}$ is a homogeneous function of degree zero in $\mathbf{X}$.

This lemma is proved in Lau (1969a) and will not be repeated here. Based on Lemma II-1, we state and prove the following theorem:

Theorem II-2. Under Assumptions (F.1) through (F.7), a production function is homothetic in $\mathbf{X}$ if and only if the normalized profit function is homothetic in $\mathbf{q}$.

Proof: For a homothetic production function the first-order necessary conditions for a maximum imply that

$$\frac{\partial F/\partial H}{\partial F/\partial H}\frac{\partial H/\partial X_i}{\partial H/\partial X_1}=\frac{q_i}{q_1},\quad \forall i.$$

By homotheticity, the left-hand side of the equation is homogeneous of degree zero in $\mathbf{X}$. One may therefore rewrite the left-hand side as functions only of X_i/X_1. Our Assumptions (F.1) through (F.7) are sufficient to ensure that the (X_i/X_1)'s may be solved uniquely as continuously differentiable functions of (q_i/q_1)'s,

$$\frac{X_i}{X_1}=f_i^1\left(\frac{q_2}{q_1},\frac{q_3}{q_1},\ldots,\frac{q_m}{q_1},\mathbf{Z}\right),\quad \forall i,$$

which by using the dual transformation yields

$$\frac{\partial G/\partial q_i}{\partial G/\partial q_1}=f_i^1\left(\frac{q_2}{q_1},\frac{q_3}{q_1},\ldots,\frac{q_m}{q_1},\mathbf{Z}\right).$$

Since the ratios of the first partial derivatives of G with respect to $\mathbf{q}$ are homogeneous of degree zero in $\mathbf{q}$, G is homothetic in $\mathbf{q}$ by Lemma II-1. The converse is proved similarly starting from

$$\frac{\partial G/\partial H}{\partial G/\partial H}\frac{\partial H/\partial q_i}{\partial H/\partial q_1}=\frac{X_i}{X_1}.\qquad \text{Q.E.D.}$$

The next theorem shows the effect of changing the scale of measurement of output (or, as some authors prefer it, the level of technical efficiency):

> *Theorem II*-3. Let $Y=F(\mathbf{X},\mathbf{Z})$ and $\Pi^*=G(\mathbf{q},\mathbf{Z})$ be a production function satisfying Assumptions (F.1) through (F.7) and its conjugate normalized profit function, respectively. Then for any $A>0$, if the production function is given by $Y=AF(\mathbf{X},\mathbf{Z})$, the normalized profit function is given by $\Pi^*=AG(\mathbf{q}/A,\mathbf{Z})$.[12]

Proof:

$$\begin{aligned}\Pi^* &= \max\{AF(\mathbf{X},\mathbf{Z})-\mathbf{q}'\mathbf{X}\}\\ &= A\max\{F(\mathbf{X},\mathbf{Z})-\mathbf{q}'/A\mathbf{X}\}\\ &= AG(\mathbf{q}/A,\mathbf{Z}).\qquad \text{Q.E.D.}\end{aligned}$$

[12]This theorem is proved in Fenchel (1953, pp. 93–94); see also Chapter I.1, Table 2, composition rule 1. This theorem is a special case of Theorem 28 in Chapter I.1.

The next theorem shows the effect of a translation of the origin:

Theorem II-4. Let $Y = F(\mathbf{X},\mathbf{Z})$ and $\Pi^* = G(\mathbf{q},\mathbf{Z})$ be a production function satisfying Assumptions (F.1) through (F.7) and its conjugate normalized profit function, respectively. Then for any constant $\bar{Y} > 0$ and constant vector $\bar{\mathbf{X}} > 0$, if the production function is given by $Y = \bar{Y} + F(\mathbf{X} + \bar{\mathbf{X}},\mathbf{Z})$, the normalized profit function is given by $\Pi^* = \bar{Y} + G(\mathbf{q},\mathbf{Z}) + \mathbf{q}'\bar{\mathbf{X}}$.[13]

Proof:

$$\begin{aligned} \Pi^* &= \max_{\mathbf{X}}\{\bar{Y} + F(\mathbf{X} + \bar{\mathbf{X}},\mathbf{Z}) - \mathbf{q}'\mathbf{X}\} \\ &= \bar{Y} + \max_{\mathbf{X}^*}\{F(\mathbf{X}^*,\mathbf{Z}) - \mathbf{q}'(\mathbf{X}^* - \bar{\mathbf{X}})\} \\ &= \bar{Y} + \mathbf{q}'\bar{\mathbf{X}} + \max_{\mathbf{X}^*}\{F(\mathbf{X}^*,\mathbf{Z}) - \mathbf{q}'\mathbf{X}^*\} \\ &= \bar{Y} + G(\mathbf{q},\mathbf{Z}) + \mathbf{q}'\bar{\mathbf{X}}. \qquad \text{Q.E.D.} \end{aligned}$$

Theorem II-5. Under Assumptions (F.1) through (F.7), let the sum of production elasticities be given as

$$\epsilon = \sum_{i=1}^{m} \frac{\partial \ln F}{\partial \ln X_i},$$

then

$$\frac{\partial \ln G}{\partial \ln q_i} = -\frac{1}{(1-\epsilon)} \frac{\partial \ln F}{\partial \ln X_i}.$$

Proof: By the dual transformation,

$$\frac{G}{F} = (1-\epsilon).$$

Strong concavity implies that $\epsilon < 1$. Thus,

$$\begin{aligned} \frac{\partial \ln G}{\partial \ln q_i} &= \frac{q_i}{G} \frac{\partial G}{\partial q_i} \\ &= \frac{\partial F/\partial X_i}{(1-\epsilon)F}(-X_i), \end{aligned}$$

[13]This theorem is proved in Fenchel (1953, pp. 94–95).

by the dual transformation. Thus,

$$\frac{\partial \ln G}{\partial \ln q_i} = -\frac{1}{(1-\epsilon)}\frac{\partial \ln F}{\partial \ln X_i}. \qquad \text{Q.E.D.}$$

Corollary 5.1. Under Assumptions (F.1) through (F.7), let the sum of normalized profit elasticities be given as

$$\eta = \sum_{i=1}^{m} \frac{\partial \ln G}{\partial \ln q_i},$$

then

$$\frac{\partial \ln F}{\partial \ln X_i} = -\frac{1}{(1-\eta)}\frac{\partial \ln G}{\partial \ln q_i}. \qquad \text{Q.E.D.}$$

Proof: Identical to the theorem.

Corollary 5.2. Under Assumptions (F.1) through (F.7),

$$\frac{1}{\epsilon} + \frac{1}{\eta} = 1.$$

Proof: By the theorem,

$$\begin{aligned}\sum_{i=1}^{m} \frac{\partial \ln G}{\partial \ln q_i} &= -\frac{1}{(1-\epsilon)}\sum_{i=1}^{m}\frac{\partial \ln F}{\partial \ln X_i}\\ &= -\frac{1}{(1-\epsilon)}\epsilon\\ &= \eta.\end{aligned}$$

Thus,

$$-\epsilon = \eta - \eta\epsilon, \qquad \eta + \epsilon = \eta\epsilon.$$

Dividing through by $\eta\epsilon$, we obtain

$$\frac{1}{\epsilon} + \frac{1}{\eta} = 1. \qquad \text{Q.E.D.}$$

Theorem II-5 shows how estimates of production elasticities may be derived given the estimates of the normalized profit elasticities and vice versa.

Theorem II-6. Under Assumptions (F.1) through (F.7), a homogeneous production function of degree k in $\mathbf{X}$, $0<k<1$, is separable with respect to a commodity-wise partition in $\mathbf{X}$ if and only if the normalized profit function is also separable price-wise in $\mathbf{q}$.

Proof: Homogeneity implies that $F(\mathbf{X},\mathbf{Z}) = H(\mathbf{X},\mathbf{Z})$, where H is a homogeneous function of degree k in $\mathbf{X}$. Separability implies that

$$\frac{\partial}{\partial X_k}\left(\frac{(\partial H/\partial X_i)(\mathbf{X},\mathbf{Z})}{(\partial H/\partial X_j)(\mathbf{X},\mathbf{Z})}\right) = 0, \qquad \forall i,j,k, \quad i \neq j \neq k,$$

which in turn implies that $(\partial H/\partial X_i)/(\partial H/\partial X_j)$ is a function of only X_i, X_j and $\mathbf{Z}$. Homogeneity implies that this function is homogeneous of degree zero. Thus, one has

$$\begin{aligned}\frac{q_i}{q_j} = \frac{\partial H/\partial X_i}{\partial H/\partial X_j} &= h_{ij}(X_i,X_j,\mathbf{Z}) \\ &= h_{ij}\left(\frac{X_i}{X_j},1,\mathbf{Z}\right).\end{aligned}$$

If this equation can be solved for X_i/X_j as a function of q_i/q_j and $\mathbf{Z}$, then it follows immediately by a dual transformation that $(\partial G/\partial q_i)/(\partial G/\partial q_j)$ is independent of $q_k, k \neq i,j$. But Assumptions (F.1) through (F.7) are sufficient to guarantee that X_i/X_j are continuously differentiable functions of $\mathbf{q}$. Thus, the function (X_i/X_j) exists and we conclude that G is separable price-wise.

The converse of this theorem may be proved in a similar manner by observing that G is also homogeneous by Theorem II-1. This completes the proof. Q.E.D.

Corollary 6.1. Under Assumptions (F.1) through (F.7), a production function homothetic in $\mathbf{X}$ is separable commodity-wise in $\mathbf{X}$ if and only if the normalized profit function is separable price-wise in $\mathbf{q}$.

Proof: This follows directly from Theorems II-3 and II-6, and Lemma II-1. Q.E.D.

We need the following lemma to prove a generalized version of Theorem II-6 which applied to production functions in which the inputs may be grouped into several categories, such as capital and labor, each of which may consist of capital and labor of many different kinds.

> *Lemma II*-2. A strongly separable function is homothetic if and only if each category function (or quantity index) is homogeneous of the same degree, or it is a function of products of homogeneous category functions.

This lemma implies that if Y is strongly separable, that is, if

$$Y = F\left(\sum_{i=1}^{m} X^i(X_{i1},...,X_{in},\mathbf{Z})\right)$$

then Y is homothetic if and only if either each X^i is homogeneous of the same degree or $Y = F(\prod_{i=1}^{m} X^i(X_{i1},...,X_{in},\mathbf{Z}))$, where each X^i is homogeneous (not necessarily of the same degree). This is proved in Lau (1969a). We omit the proof.

> *Theorem II*-7. Under Assumptions (F.1) through (F.7), a production function is additively separable with respect to the commodity categories if and only if the normalized profit function is additively separable with respect to the corresponding price categories.

Proof: Additive separability implies

$$Y = \sum_{i=1}^{m} X^i(X_{i1},...,X_{in_i},\mathbf{Z}).$$

It is easy to see the maximization of

$$P^* = Y - \sum_{i=1}^{m} \sum_{j=1}^{n_j} q_{ij}X_{ij}$$

results in demand functions for X_{ij}'s which depend only on the normalized prices of the commodities of the ith category and $\mathbf{Z}$. Thus $G(\mathbf{q},\mathbf{Z})$ must also be additively separable in $\mathbf{q}$. The converse is proved similarly. Q.E.D.

> *Theorem II*-8. Under Assumptions (F.1) through (F.7), a production function is homogeneous and strongly separable with respect to the commodity categories if and only if the normalized profit function is homogeneous and strongly separable with respect to the corresponding price categories.

Proof: Homogeneity follows from Theorem II-1. The first-order necessary conditions for a maximum require that

$$\frac{\partial F/\partial X_{ir}}{\partial F/\partial X_{js}} = \frac{\partial X^i/\partial X_{ir}}{\partial X^j/\partial X_{js}}$$
$$= \frac{X^i_r(X_{i1},\ldots,X_{in_i},\mathbf{Z})}{X^j_s(X_{j1},\ldots,X_{jn_j},\mathbf{Z})} = \frac{q_{ir}}{q_{js}}, \qquad i \neq j, \quad i,j = 1,\ldots,m,$$
$$= \frac{X^i_r(X_{i1}/X_{j1},\ldots,X_{in_i}/X_{j1},\mathbf{Z})}{X^j_s(1,\ldots,X_{jn_j}/X_{j1},\mathbf{Z})}, \qquad r = 1,\ldots,n_i, \quad s = 1,\ldots,n_j,$$

by zero degree homogeneity. Note that there exists $n_i + n_j - 1$ independent equations for each pair (i,j) in the $n_i + n_j - 1$ unknown X_{ir}/X_{j1}'s and X_{js}/X_{j1}'s. Moreover, from Lemma I-2, the optimal factor proportions are continuously differentiable function of only q_{ir}/q_{j1}'s and q_{js}/q_{j1}'s. Hence one has

$$\frac{\partial}{\partial q_{kt}}(X_{ir}/X_{j1}) = 0, \qquad k \neq i,j,$$

which implies also that

$$\frac{\partial}{\partial q_{kt}}(X_{ir}/X_{js}) = 0, \qquad k \neq i,j. \tag{II-2}$$

On applying the dual transformation, equation (II-2) becomes

$$\frac{\partial}{\partial q_{kt}}\left(\frac{\partial G/\partial q_{ir}}{\partial G/\partial q_{js}}\right) = 0, \qquad k \neq i,j, \quad i,j,k = 1,\ldots,m, \quad r = 1,\ldots,n_i, \quad s = 1,\ldots,n_j, \quad t = 1,\ldots,n_k.$$

Hence G is strongly separable. The converse is proved similarly. Q.E.D.

Corollary 8.1. Under Assumptions (F.1) through (F.7), a production function is homothetic and strongly separable if and only if the normalized profit function is homothetic and strongly separable.

Proof: Homotheticity follows from Theorem II-2. Otherwise essentially the same proof of the theorem suffices. Q.E.D.

Note the crucial role of the homogeneity of each category function. Otherwise it will not be possible to express X_{ir}/X_{js} as a function of only $\{q_{it}, q_{jt}\}$.

Definition. A function is said to be *homothetically separable* if it is weakly separable and each category function is homothetic. (Note that the function itself need not be homothetic.)

We now introduce Lemma II-3, which is also proved in Lau (1969a).

Lemma II-3. A homothetic and weakly separable function is homothetically separable.

Theorem II-9. Under Assumptions (F.1) through (F.7), a production function is homothetically separable if and only if the normalized profit function is homothetically separable.

Proof: The first order necessary conditions are

$$\frac{\partial X^i/\partial X_{ir}}{\partial X^i/\partial X_{i1}} = \frac{X^i_r(1,\ldots,X_{in_i}/X_{i1},\mathbf{Z})}{X^i_1(1,\ldots,X_{in_i}/X_{i1},\mathbf{Z})} = \frac{q_{ir}}{q_{i1}}, \qquad r = 2,\ldots,n_i.$$

Thus by an argument similar to that in previous theorems one has

$$\frac{\partial}{\partial q_{jt}}\left(\frac{\partial G/\partial q_{ir}}{\partial G/\partial q_{is}}\right) = 0, \qquad j \neq i, \quad \forall r,s,t.$$

The converse is proved similarly. Q.E.D.

Corollary 9.1. Under Assumptions (F.1) through (F.7), a production function is homothetic and weakly separable if and only if the normalized profit function is homothetic and weakly separable.

Proof: This follows from Lemma II-3 and the Theorem II-9. Q.E.D.

Theorem II-10. Under Assumptions (F.1) through (F.7), a production function and its normalized profit function are strongly separable (but not additively separable) with respect to the commodity categories and the corresponding price categories respectively only if they are homothetic.

Proof: Strong separability of both F and G implies

$$Y = F\left(\sum_{i=1}^{m} X^i(X_{i1},\ldots,X_{in_i},\mathbf{Z})\right)$$

and

$$\Pi^* = G\left(\sum_{i=1}^{m} Q^i(q_{i1},\ldots,q_{in_i},\mathbf{Z})\right).$$

Now

$$\frac{\partial F/\partial X_{ir}}{\partial F/\partial X_{js}} - \frac{X^i_r(X_{i1},\ldots,X_{in_i},\mathbf{Z})}{X^j_s(X_{j1},\ldots,X_{jn_j},\mathbf{Z})}. \tag{II-3}$$

Applying the dual transformation to equation (II-3), we have

$$\frac{q_{ir}}{q_{js}} = \frac{X^i_r(-\partial G/\partial q_{i1},\ldots,-\partial G/\partial q_{in_i},\mathbf{Z})}{X^j_s(-\partial G/\partial q_{i1},\ldots,-\partial G/\partial q_{jn_j},\mathbf{Z})}. \tag{II-4}$$

Differentiating both sides of equation (II-4) by q_{kt}, $k \neq i,j$, and observing that

$$\frac{\partial^2 G}{\partial q_{ir}\partial q_{kt}} = G'' Q^i_r Q^k_t,$$

equation (II-4) becomes

$$X^j_s \sum_{l=1}^{n_i} X^i_{rl} \cdot G'' Q^i_l Q^k_t - X^i_r \sum_{l=1}^{n_j} X^j_{sl} \cdot G'' Q^j_l Q^k_t = 0. \tag{II-5}$$

Now $G'' \neq 0$, otherwise the production function is additive in the X^i's by Theorem II-7, which is ruled out by hypothesis. Moreover, observe that

$$G' Q^i_l = -X_{il}.$$

Hence, equation (II-5) becomes, after multiplication by $G'/G''Q^k_t$,

$$\sum_{l=1}^{n_i} \frac{\partial}{\partial X_{il}} \left(\frac{X^i_r}{X^j_s}\right) \cdot X_{il} + \sum_{l=1}^{n_j} \frac{\partial}{\partial X_{jl}} \left(\frac{X^i_r}{X^j_s}\right) \cdot X_{jl} = 0.$$

By Euler's Theorem, (X^i_r/X^j_s) is homogeneous of degree zero in $\mathbf{X}$, $\forall i,j,r,s$. By Lemma II-1, F is homothetic, and by Theorem II-1, G is also homothetic. Q.E.D.

Note that by Lemma II-2 then, the X_i's are either homogeneous of the same degree or are logarithms of homogeneous functions.

Theorem II-11. Under Assumptions (F.1) through (F.7), a production function and its corresponding normalized profit function are both weakly separable only if they are both homothetically separable.

Proof: Let the production and normalized profit functions be

$$Y = F(X^1(X_{11},\ldots,X_{1n_1},\mathbf{Z}),\ldots,X^m(X_{m1},\ldots,X_{mn_m},\mathbf{Z}),\mathbf{Z}),$$

and

$$\Pi^* = G(Q^1(q_{11},\ldots,q_{1n_1},\mathbf{Z}),\ldots,Q^m(q_{m1},\ldots,q_{mn_m},\mathbf{Z}),\mathbf{Z}).$$

It is necessary to show that each X^i and hence each G^i is homothetic. The proof is strictly analogous to that of Theorem II-10. Applying the dual transformation to the first-order necessary condition, one obtains

$$\frac{q_{ir}}{q_{is}} = \frac{X^i_r(-\partial G/\partial q_{i1},\ldots,-\partial G/\partial q_{in_i},\mathbf{Z})}{X^i_s(-\partial G/\partial q_{i1},\ldots,-\partial G/\partial q_{in_i},\mathbf{Z})}. \tag{II-6}$$

Differentiating both sides of equation (II-6) by q_{jt}, $j \neq i$, we have

$$X^i_s \sum_{l=1}^{n_i} X^i_{rl}\left(-\frac{\partial^2 G}{\partial q_{il}\partial_{jt}}\right) - X^i_r \sum_{l=1}^{n_i} X^i_{sl}\left(-\frac{\partial^2 G}{\partial q_{il}\partial q_{jt}}\right) = 0. \tag{II-7}$$

For a weakly separable normalized profit function

$$\frac{\partial^2 G}{\partial q_{il}\partial q_{jt}} = Q^i_l \frac{\partial^2 G}{\partial Q^i \partial Q^j} Q^j_t = Q^i_l \cdot G_{ij} \cdot Q^j_t.$$

Hence, equation (II-7) becomes

$$\sum_{l=1}^{n_i} (X^i_s X^i_{rl} - X^i_r X^i_{sl}) \cdot X_{il} \cdot \left(\frac{G_{ij}}{G_i}\right) Q^j_t = 0.$$

Now, $G_i \neq 0$ and $G_{ij} \neq 0$, the latter because of weak (but not strong) separability. Therefore,

$$\sum_{l=1}^{n_i} \frac{\partial}{\partial X_{il}}\left(\frac{X^i_r}{X^i_s}\right) \cdot X_{il} = 0.$$

By Euler's Theorem, (X^i_r/X^i_s) is homogeneous of degree zero, $\forall r,s$. By Lemma II-1 each X^i is homothetic. By Theorem II-9, each Q^i is homothetic. Q.E.D.

These theorems are useful in specifying technologies with multiple variable input categories. They also have application in aggregation, in the construction of quantity and price indices and in the analysis of organization and information structures.

It should be noted that homogeneity, separability and other similar properties of the normalized profit function considered here may be alternatively deduced through the cost function by utilizing the general composition rules for cost functions in Theorem 9 in Chapter I.1 along with direct arguments on maximization of $Y - C(Y,q,Z)$ where C is the cost function. Here we have relied primarily on the Legendre transformation because the proofs are more direct and immediate. Of course, the proofs only apply under conditions which allow the use of the Legendre transformation, for example, under Assumptions (F.1) through (F.7) on the production function.

2.2. *Structures Involving Fixed Inputs*

Thus far we have not examined structural properties which involve the fixed inputs **Z**. Normalized profit functions with fixed inputs are sometimes referred to as normalized restricted profit functions [see Lau (1976a)]. To analyze structures involving fixed inputs, we introduce the concept of almost homogeneity.

Definition. A function $F(X,Z)$ is *almost homogeneous of degrees* k_1 *and* k_2 in **X** and **Z**, respectively, if

$$F(\lambda \mathbf{X},\lambda^{k_2}\mathbf{Z}) = \lambda^{k_1}F(\mathbf{X},\mathbf{Z}), \qquad \forall\lambda > 0.^{14} \tag{II-8}$$

The economic interpretation of an almost homogeneous production function is the following: if a set of inputs **X** is increased by the same proportion and another set of inputs **Z** is increased by some power of that proportion, then output Y will be increased by another power of that proportion. In the special case that $k_1 = k_2 = 1$, we have constant returns to scale in all inputs.

It will be shown that an almost homogeneous function satisfies a modified Euler's Theorem.

Lemma II-4. A continuously differentiable function is almost homogeneous of degree k_1 and k_2 if and only if

$$\sum_{i=1}^{m} \frac{\partial F}{\partial X_i}\cdot X_i + k_2 \sum_{i=1}^{n} \frac{\partial F}{\partial Z_i}\cdot Z_i = k_1 F(\mathbf{X},\mathbf{Z}). \tag{II-9}$$

Proof:
Necessity. If $F(\mathbf{X},\mathbf{Z})$ is almost homogeneous it satisfies equation (II-8). Differentiation of equation (II-8) with respect to λ yields

$$\sum_{i=1}^{m} \frac{\partial F}{\partial X_i}\cdot X_i + k_2 \sum_{i=1}^{n} \frac{\partial F}{\partial Z_i}\, \lambda^{k_2-1}\cdot Z_i = k_1\lambda^{k_1-1}F(\mathbf{X},\mathbf{Z}).$$

[14] See Aczel (1966, Ch. 7) for a discussion of almost homogeneous functions. Lau (1972) defines almost homogeneity in a slightly different manner.

This must hold identically for all $\lambda > 0$ and in particular for $\lambda = 1$. Hence

$$\sum_{i=1}^{m} \frac{\partial F}{\partial X_i} \cdot X_i + k_2 \sum_{i=1}^{n} \frac{\partial F}{\partial Z_i} \cdot Z_i = k_1 F(\mathbf{X},\mathbf{Z}).$$

Sufficiency. Suppose $F(\mathbf{X},\mathbf{Z})$ satisfies equation (II-9). We note that

$$\frac{\partial F}{\partial Z_i} = \frac{\partial F}{\partial Z_i^{1/k_2}} \cdot \frac{1}{k_2} Z_i^{1/k_2 - 1},$$

one may therefore rewrite equation (II-9) in the form

$$\sum_{i=1}^{m} \frac{\partial F}{\partial X_i} X_i + \sum_{i=1}^{n} \frac{\partial F}{\partial Z_i^{1/k_2}} \cdot Z_i^{1/k_2} = k_1 F(\mathbf{X},\mathbf{Z}).$$

Thus, by Euler's Theorem, F must be homogeneous of degree k_1 in $\mathbf{X}$ and $\mathbf{Z}^{1/k_2}$. In other words,

$$F(\mathbf{X},\mathbf{Z}) = H(\mathbf{X},\mathbf{Z}^{1/k_2}),$$

where H is homogeneous of degree k_1.

$$\begin{aligned} F(\lambda \mathbf{X},\lambda^{k_2}\mathbf{Z}) &= H(\lambda \mathbf{X},(\lambda^{k_2}\mathbf{Z})^{1/k_2}) \\ &= H(\lambda \mathbf{X},\lambda \mathbf{Z}^{1/k_2}) \\ &= \lambda^{k_1} F(\mathbf{X},\mathbf{Z}). \qquad \text{Q.E.D.} \end{aligned}$$

Theorem II-12. Under Assumptions (F.1) through (F.7), a production function is homogeneous of degree k in all inputs, variable and fixed, if and only if the normalized restricted profit function is almost homogeneous of degrees $-1/(1-k)$ and $-k/(1-k)$ if $k \neq 1$, and homogeneous of degree one in $\mathbf{Z}$ if $k = 1$.

Proof: By Euler's Theorem,

$$\sum_{i=1}^{m} \frac{\partial F}{\partial X_i} \cdot X_i + \sum_{i=1}^{n} \frac{\partial F}{\partial Z_i} \cdot Z_i = kF.$$

By a dual transformation, one has

$$-\sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i} + \sum_{i=1}^{n} \frac{\partial G}{\partial Z_i} \cdot Z_i = k\left(G - \sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i}\right),$$

which simplifies to

$$\sum_{i=1}^{m} \frac{\partial G}{\partial q_i} q_i - \frac{1}{(1-k)} \sum_{i=1}^{n} \frac{\partial G}{\partial Z_i} Z_i = -\frac{k}{(1-k)} G, \quad \text{if} \quad k \neq 1,$$

or to

$$\sum_{i=1}^{n} \frac{\partial G}{\partial Z_i} \cdot Z_i = G, \qquad \text{if} \quad k = 1.$$

The converse is proved by retracing the steps. Q.E.D.

Note that $k > 1$ implies increasing returns to scale in all inputs. For the purpose of this theorem k may be either greater than or less than one.

Corollary 12.1. Under Assumptions (F.1) through (F.7), a production function is homogeneous of degree k in $\mathbf{Z}$, $k > 0$, if and only if the normalized restricted profit function is almost homogeneous of degrees 1 and $1/k$.

Proof: By Euler's Theorem,

$$\sum_{i=1}^{n} \frac{\partial F}{\partial Z_i} \cdot Z_i = kF.$$

By a dual transformation, one has

$$\sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i} + \frac{1}{k} \sum_{i=1}^{n} \frac{\partial G}{\partial Z_i} Z_i = G.$$

Thus, G is almost homogeneous of degrees 1 and $1/k$. The converse is proved similarly. Q.E.D.

Corollary 12.2. Under Assumptions (F.1) through (F.2), and homogeneity of degree k, $k \neq 1$ in all inputs, the derived demand functions are almost homogeneous of degrees $-1/(1-k)$ and $-k/(1-k)$.

Proof: This follows from differentiating the next to the last equation in the proof of the theorem. Q.E.D.

Next we wish to characterize the normalized restricted profit function corresponding to a homothetic production function. A production function is homothetic in $\mathbf{X}$ and $\mathbf{Z}$ if it can be written in the form

$$Y = F(H(\mathbf{X},\mathbf{Z})),$$

where F is a positive, finite, continuous and strictly monotonic function of one variable with $F(0) = 0$ and H is a homogeneous function of degree one in **X** and **Z**. Homogeneity of H implies that $H(\mathbf{0},\mathbf{0}) = 0$. If $F(\cdot,\cdot)$ is non-negative and strictly *monotonically* increasing on $R_+^m \times \bar{R}_+^n$, then one can always choose F and H such that $F(\cdot)$ and $H(\cdot,\cdot)$ are both non-negative and strictly increasing on the non-negative real line and $R_+^m \times \bar{R}_+^n$, respectively. Monotonicity of $F(.,.)$ implies that $F(\cdot)$ and $H(.,.)$ must be monotonic in the same direction. Subject to the convention that $H(\cdot,\cdot)$ is strictly increasing on $R_+^m \times \bar{R}_+^n$, Euler's Theorem requires that $H(\cdot,\cdot)$ be non-negative on $R_+^m \times \bar{R}_+^n$. Thus both $F(\cdot)$ and $H(\cdot,\cdot)$ can be chosen to be non-negative and strictly increasing on $\bar{R}_+$ and $R_+^m \times \bar{R}_+^n$, respectively. Given $F(0) = 0$, this implies that $F(\cdot)$ will be non-negative on $\bar{R}_+$.

We introduce Lemma II-5:

Lemma II-5. Under Assumptions (F.1) through (F.7), a production function is homothetic in **X** and **Z** if and only if

$$\sum_{i=1}^{m} \frac{\partial F}{\partial X_i} \cdot X_i + \sum_{i=1}^{n} \frac{\partial F}{\partial Z_i} \cdot Z_i = f(F(X,Z)),$$

where f is an arbitrary, finite, non-negative function of a single variable with $f(0) = 0$, continuous on $\bar{R}_+$ and continuously differentiable on R_+.

A proof of a similar result is available in Lau (1969a). We omit the proof.

Theorem II-13. Under Assumptions (F.1) through (F.7), a production function is homothetic in **X** and **Z** if and only if the normalized profit function satisfies the equation

$$-\sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i} + \sum_{i=1}^{n} \frac{\partial G}{\partial Z_i} \cdot Z_i = f\left(G - \sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i}\right), \qquad \text{(II-10)}$$

where f is an arbitrary, finite, non-negative function of a single variable with $f(0) = 0$, continuous on $\bar{R}_+$ and continuously differentiable on R_+.

Proof: The proof is immediate using Lemma II-5 and the dual transformation

$$\frac{\partial F}{\partial X_i} = q_i, \qquad \frac{\partial F}{\partial Z_i} = \frac{\partial G}{\partial Z_i},$$

$$X_i = -\frac{\partial G}{\partial q_i}, \qquad F(\mathbf{X},\mathbf{Z}) = G - \sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i}. \qquad \text{Q.E.D.}$$

There remains the question of a nomenclature for the class of functions defined by equation (II-10). We know from Theorem II-12 that homogeneous functions are dual to almost homogeneous functions. We shall refer to functions satisfying equation (II-10) as *partially homothetic* functions.

Theorem II-14. Under Assumptions (F.1) through (F.7), a production function is homothetically separable in $\mathbf{X}$, that is,

$$Y = F(H(\mathbf{X}),\mathbf{Z}),$$

where H is a homogeneous function of degree one if and only if the normalized restricted profit function is homothetically separable, that is,

$$\Pi^* = G(H^*(\mathbf{q}),\mathbf{Z}),$$

where H^* is a homogeneous function of degree one.

Proof: This follows from Theorem II-2. Q.E.D.

Corollary 14.1. Under Assumptions (F.1) through (F.7), a production function has the form

$$Y = F(H(\mathbf{X}),f(\mathbf{Z}))$$

where H is a homogeneous function of degree one if and only if the normalized restricted profit function has the form

$$\Pi^* = G(H^*(\mathbf{q}),f(\mathbf{Z})),$$

where H^* is a homogeneous function of degree one.

Proof: Obvious. Q.E.D.

Corollary 14.2 Under Assumptions (F.1) through (F.7), a production function is homothetic in $\mathbf{X}$ and $\mathbf{Z}$ and weakly separable in $\mathbf{X}$

and $\mathbf{Z}$ if and only if the normalized restricted profit function is partially homothetic in $\mathbf{q}$ and $\mathbf{Z}$ and weakly separable in $\mathbf{q}$ and $\mathbf{Z}$.

Proof: By Lemma II-3, homotheticity and weak separability of $F(\mathbf{X},\mathbf{Z})$ in $\mathbf{X}$, $\mathbf{Z}$ implies that

$$Y = F(H_1(\mathbf{X}),H_2(\mathbf{Z})),$$

where $F(H_1,H_2)$ is homothetic in H_1 and H_2 and H_1 and H_2 are homogeneous functions of degree one. By Corollary 14.1,

$$\Pi^* = G(H_1^*(\mathbf{X}),H_2(\mathbf{Z})).$$

Partial homotheticity follows from Theorem II-13. The converse is proved similarly. Q.E.D.

Theorem II-15. Under Assumptions (F.1) through (F.7), a production function and its corresponding normalized restricted profit function are both separable in $\mathbf{X}$ and $\mathbf{q}$ respectively only if either they are homothetically separable in $\mathbf{X}$ and $\mathbf{q}$ respectively or they are additive.

Proof: The first-order necessary conditions for maximization imply that

$$\frac{\partial F/\partial X_i}{\partial F/\partial X_j} = \frac{\partial f/\partial X_i(X_1,\ldots,X_m)}{\partial f/\partial X_j(X_1,\ldots,X_m)} = \frac{q_i}{q_j}.$$

By a dual transformation this becomes

$$\frac{f_i(-\partial G/\partial q_1,\ldots,-\partial G/\partial q_m)}{f_j(-\partial G/\partial q_1,\ldots,-\partial G/\partial q_m)} = \frac{q_i}{q_j}.$$

Differentiating this equation with respect to Z_k, we obtain

$$f_j \sum_l f_{il}\left(-\frac{\partial^2 G}{\partial q_l \partial Z_k}\right) - f_i \sum_l f_{jl}\left(-\frac{\partial^2 G}{\partial q_l \partial Z_k}\right) = 0. \tag{II-11}$$

Since G is separable $[\Pi^* = G(g(\mathbf{q}),\mathbf{Z})]$,

$$\frac{\partial^2 G}{\partial q_l \partial Z_k} = \frac{\partial^2 G}{\partial g_l \partial Z_k}\cdot\frac{\partial g}{\partial q_l}.$$

We know $\mathbf{X} = -\partial G/\partial g \cdot \partial g/\partial \mathbf{q}$. Thus, equation (II-11) becomes

$$f_j \sum_l f_{il} X_l - f_i \sum_l f_{jl} X_l = 0,$$

or $f(\mathbf{X})$ is homothetic.

An exceptional case arises if $G_{gk} = 0$, $\forall k$. Then

$$G(\mathbf{q},\mathbf{Z}) = g(\mathbf{q}) + h(\mathbf{Z}),$$

and then by Theorem II-4,

$$F(\mathbf{X},\mathbf{Z}) = f(\mathbf{X}) + h(\mathbf{Z}). \qquad \text{Q.E.D.}$$

Additional results on the structure of normalized restricted profit functions can be found in Lau (1976a).

Based on these theorems, one can specify $G(\mathbf{q},\mathbf{Z})$ depending on the assumptions one wishes to impose on the underlying technology. As seen in Lau (1969c), it is generally difficult to obtain closed form solutions for the normalized profit function for even simple technologies when some inputs are fixed. With the device of the normalized restricted profit function, this problem of specification is circumvented. Nonetheless, we are assured that the resulting system of conditional supply and demand functions may be derived from an underlying neoclassical technology and that all the empirically relevant assumptions have been incorporated.

3. Extensions to Multiple Outputs

3.1. *Introduction*

A natural extension of the concept of profit functions is to the case of multiple outputs. This has been accomplished by McFadden (1966). Here we shall point out certain special properties of multiple output profit functions as well as derive several theorems on the structure of such functions. Our results may be further extended to include technologies in which the same commodity may be used either as a net input or a net output, depending on the market prices. Such technologies are not infrequently found. An example is the purchase and sale of farm-produced fertilizers by agricultural households. Further examples are those of international trade, and the purchase and sale of new and used equipment. The advantage of this approach is that there need be no arbitrary partition of commodities into inputs and outputs.

The theory of the multiproduct firm has been analyzed by Mundlak (1964). The properties of profit functions have been studied by McFadden (1966), Diewert (1973a) and Jorgenson and Lau (1974a and 1974b). Christensen, Jorgenson and Lau (1971 and 1973) have also made an empirical application to the U.S. economy. In addition, Hall (1973) has approached the problem from the point of view of joint cost functions, using a generalization of the Generalized Leontief cost function due to Diewert (1971). The basic duality concepts which underly all these studies may be traced back to the pioneering work of Shephard (1953).

For a multiple-output, multiple-input firm, there is no natural numeraire commodity, such as the single output, to define the production function representation of the technology. Following Jorgenson and Lau (1974a and 1974b), we shall adopt the convention of choosing as our left-hand-side variable for the production function a variable input which is non-producible. In addition, every commodity is measured as if it were a net output. Thus, a net output is always non-negative. A net input is always non-positive. For the purposes of this paper we maintain the artificial distinction between a set of commodities which are net outputs and the set of commodities which are net inputs.[15] A more general treatment should allow a commodity to be either a net output or a net input depending on the prices and fixed factors.[16]

Let X_{m+1} be the quantity of the left-hand-side variable and non-producible net input, Y_i the quantity of the ith net output, $i = 1,\ldots,n$, and X_i the quantity of the ith net input, $i = 1,\ldots,m$. By convention then $X_{m+1} \leqq 0, X_i \leqq 0, \forall i$, and $Y_i \geqq 0, \forall i$. The production function is given by

$$L \equiv -X_{m+1} = F(\mathbf{Y},\mathbf{X}),$$

the minimum value of L for given values of $\mathbf{Y}$ and $\mathbf{X}$ such that the production plan $(\mathbf{Y},\mathbf{X},-L)$ is feasible.

It is assumed that $F(\mathbf{Y},\mathbf{X})$ possesses certain properties, which parallel similar properties of the single-output case:

(F*.1) *Domain.* F is a finite, non-negative, real-valued function defined on $\bar{R}^n_+ \times \bar{R}^m_- \cdot F(\mathbf{0},\mathbf{0}) = 0$.

(F*.2) *Continuity.* F is continuous on $\bar{R}^n_+ \times \bar{R}^m_-$.

[15]This actually involves little loss in generality since the functional consequence of an output and an input being the same "commodity" is that they are two products whose prices are in fixed proportions in the market.

[16]See Chapter I.1 and Jorgenson and Lau (1974a and 1974b).

(F*.3) *Smoothness.* F is continuously differentiable on $R_+^n \times R_-^m$, and the Euclidean norm of the gradient of F with respect to $\mathbf{Y}$ and $\mathbf{X}$ is unbounded for any sequence of $\mathbf{Y}, \mathbf{X}$ in $R_+^n \times R_-^m$ converging to a boundary point of $\bar{R}_+^n \times \bar{R}_-^m$.

(F*.4) *Monotonicity.* F is non-decreasing on $\bar{R}_+^n \times \bar{R}_-^m$ and strictly increasing on $R_+^n \times R_-^m$.

(F*.5) *Convexity.* F is convex on $\bar{R}_+^n \times \bar{R}_-^m$ and locally strongly convex on $R_+^n \times R_-^m$.

(F*.6) *Twice Differentiability.* F is twice continuously differentiable on $R_+^n \times R_-^m$.

(F*.7) *Boundedness.*

$$\lim_{\lambda \to \infty} \frac{F(\lambda \mathbf{Y}, \lambda \mathbf{X})}{\lambda} = \infty, \qquad \forall \mathbf{Y}, \mathbf{X} \in \bar{R}_+^n \times \bar{R}_-^m, \mathbf{Y}, \mathbf{X} \neq \mathbf{0}.$$

We note that one consequence of our domain assumption is that for any given vector of net inputs $\mathbf{X}$, any vector of net outputs $\mathbf{Y}$ may be produced with an appropriate choice of L. In other words, any one of the net inputs $\mathbf{X}$ may be indefinitely substituted by L. This is admittedly a restrictive assumption. For example, the technology represented by

$$L = \left[\frac{1}{Y_1^2 + Y_2^2} + \frac{1}{X}\right]^{-1}$$

violates our domain assumption, for if $Y_1^2 + Y_2^2 > -X$, L is negative.[17] However, as indicated in Section 1.3, it is a relatively straightforward matter to introduce a more restrictive domain assumption and make corresponding changes in the assumption on G. We therefore maintain our domain assumption as it stands for the sake of simplicity of exposition.

The normalized profit function is given by

$$G(\mathbf{p}, \mathbf{q}) = \sup_{\mathbf{Y}, \mathbf{X}} \{\mathbf{p}'\mathbf{Y} + \mathbf{q}'\mathbf{X} - F(\mathbf{Y}, \mathbf{X}) | \mathbf{Y}, \mathbf{X} \in \bar{R}_+^n \times \bar{R}_-^m\},$$

where $\mathbf{p}$ and $\mathbf{q}$ are respectively the normalized prices of $\mathbf{Y}$ and $\mathbf{X}$ in terms of L. The corresponding properties of the normalized profit

[17]This example is due to Daniel McFadden.

function are:

(G*.1) *Domain.* G is a finite, positive, real-valued function defined on $R^n_+ \times R^m_+$.

(G*.2) *Continuity.* G is continuous on $R^n_+ \times R^m_+$.

(G*.3) *Smoothness.* G is continuously differentiable on $R^n_+ \times R^m_+$, and the Euclidean norm of the gradient of G with respect to $\mathbf{p}$ and $\mathbf{q}$ is unbounded for any sequence of $\mathbf{p}$, $\mathbf{q}$ in $R^n_+ \times R^m_+$ converging to a boundary point of $\bar{R}^n_+ \times \bar{R}^m_+$.

(G*.4) *Monotonicity.* $G(\mathbf{p},\mathbf{q})$ is strictly increasing in $\mathbf{p}$ and strictly decreasing in $\mathbf{q}$ on $R^n_+ \times R^m_+$.

(G*.5) *Convexity.* $G(\mathbf{p},\mathbf{q})$ is locally strongly convex on $R^n_+ \times R^m_+$.

(G*.6) *Twice Differentiability.* $G(\mathbf{p},\mathbf{q})$ is twice continuously differentiable on $R^n_+ \times R^m_+$.

(G*.7) *Boundedness.*

$$\lim_{\lambda \to \infty} \frac{G(\lambda \mathbf{p}, \lambda \mathbf{q})}{\lambda} = \infty, \qquad \forall \mathbf{p},\mathbf{q} \in R^n_+ \times R^m_+.$$

It can be proved that Assumptions (F*.1) through (F*.7) imply Assumptions (G*.1) through (G*.7) and *vice versa*. The proof closely parallels the arguments used earlier in the single output case. We omit the proof. Properties of profit functions under more general conditions are derived in Chapter I.1.

As in the single output case, the Legendre transformation also holds in this case, with the following dual relationships:

$$\frac{\partial F}{\partial \mathbf{Y}} = \mathbf{p}, \qquad \frac{\partial G}{\partial \mathbf{p}} = \mathbf{Y},$$

$$\frac{\partial F}{\partial \mathbf{X}} = \mathbf{q}, \qquad \frac{\partial G}{\partial \mathbf{q}} = \mathbf{X}.$$

$$F + G = \mathbf{p}'\mathbf{Y} + \mathbf{q}'\mathbf{X}$$

$$= \frac{\partial F}{\partial \mathbf{Y}} \cdot \mathbf{Y} + \frac{\partial F}{\partial \mathbf{X}} \cdot \mathbf{X}, \qquad = \mathbf{p} \cdot \frac{\partial G}{\partial \mathbf{p}} + \mathbf{q} \cdot \frac{\partial G}{\partial \mathbf{q}}.$$

These dual relations may also be used in the study of relationships between classes of production functions and normalized profit functions.

3.2. *Homogeneity and Separability*

In the case of a multiple output and multiple input production function, the ordinary concept of homogeneity of the production function needs to be modified. Intuitively, we want to say that a production function is in some sense homogeneous of degree k if, when all net inputs are scaled by the same proportion λ, $\lambda > 0$, all net outputs are scaled by the same proportion λ^k. In other words, if

$$L = F(\mathbf{Y},\mathbf{X}),$$

then

$$\lambda L = F(\lambda^k \mathbf{Y},\lambda \mathbf{X}),$$

or

$$F(\lambda^k \mathbf{Y},\lambda \mathbf{X}) = \lambda F(\mathbf{Y},\mathbf{X}).$$

This corresponds precisely to the concept of almost homogeneity introduced in Section 2.2. The production function is almost homogeneous of degree 1 and k.[18]

Theorem III-1. Under Assumptions (F*.1) through (F*.7), a production function is almost homogeneous of degree 1 and k, $k < 1$, in outputs, if and only if the normalized profit function is homogeneous of degree $1/(1-k)$ in the normalized output prices.

Proof: By Lemma II-4 almost homogeneity implies

$$k \sum_{i=1}^{n} \frac{\partial F}{\partial Y_i} \cdot Y_i + \sum_{i=1}^{m} \frac{\partial F}{\partial X_i} \cdot X_i = F(\mathbf{Y},\mathbf{X}).$$

By a dual transformation, this equation becomes

$$k \sum_{i=1}^{n} p_i \frac{\partial G}{\partial p_i} + \sum_{i=1}^{m} q_i \frac{\partial G}{\partial q_i} = -G + \sum_{i=1}^{n} \frac{\partial G}{\partial p_i} \cdot p_i + \sum_{i=1}^{m} \frac{\partial G}{\partial q_i} \cdot q_i,$$

[18]Nothing requires that the scale effects be uniform. One may in fact have

$$F(\lambda^{k_1} Y_1, \lambda^{k_2} Y_2, \ldots, \lambda^{k_n} Y_n; \lambda \mathbf{X}) = \lambda F(\mathbf{Y},\mathbf{X}).$$

This is a straightforward generalization of almost homogeneity.

or

$$\sum_{i=1}^{n} p_i \frac{\partial G}{\partial p_i} = \frac{1}{(1-k)} G.$$

The converse may be proved by retracing the steps. Q.E.D.

Corollary 1.1. Under Assumptions (F*.1) through (F*.7), and almost homogeneity of degrees 1 and k,

$$R^* = \frac{1}{(1-k)} G,$$

and

$$C^* = \frac{k}{(1-k)} G,$$

where R^* is the profit-maximizing normalized revenue and C^* is the profit-maximizing normalized cost.

Proof: This follows from the last equation in the proof of the theorem. Q.E.D.

Corollary 1.2. Under Assumptions (F*.1) through (F*.7) and almost homogeneity of degrees 1 and k, the derived supply functions of the outputs are homogeneous of degree $1/(1-k)$ in $\mathbf{p}$.

Proof: These follow from the properties of partial derivatives of homogeneous functions. Q.E.D.

Lemma III-1. Under Assumptions (G*.1) and (G*.5), the profit function,

$$\Pi(\mathbf{p}^*,\mathbf{q}^*,w) = wG(\mathbf{p},\mathbf{q}),$$

is homogeneous of degree k in $\mathbf{p}^*$ if and only if it is homogeneous of degree $(1-k)$ in $\mathbf{q}^*$ and w.

Proof: It is well-known that $\Pi(\mathbf{p}^*,\mathbf{q}^*,w)$ is homogeneous of degree one in all prices. Hence

$$\sum_{i=1}^{n} p_i^* \frac{\partial \Pi}{\partial p_i^*} + \sum_{i=1}^{m} q_i^* \frac{\partial \Pi}{\partial q_i^*} + w \frac{\partial \Pi}{\partial w} = \Pi.$$

By hypothesis, Π is homogeneous of degree k in $\mathbf{p}^*$. Thus

$$\sum_{i=1}^{n} p_i^* \frac{\partial \Pi}{\partial p_i^*} = k\Pi = \Pi - \sum_{i=1}^{m} q_i^* \frac{\partial \Pi}{\partial q_i^*} - w \frac{\partial \Pi}{\partial w},$$

which simplifies to

$$\sum_{i=1}^{m} q_i^* \frac{\partial \Pi}{\partial q_i} + w \frac{\partial \Pi}{\partial w} = (1-k)\Pi.$$

The converse is proved similarly. Q.E.D.

Corollary 1.3. Under Assumptions (F*.1) through (F*.7), the production function is almost homogeneous of degree 1 and k if and only if the profit function is homogeneous of degree $-k/(1-k)$ in the input prices.

Proof: This follows directly from the theorem and Lemma III-1. Q.E.D.

With multiple outputs and inputs a technology is said to be separable in outputs and inputs if there exist functions $f(\cdot)$ and $g(\cdot)$ such that

$$f(\mathbf{Y}) - g(\mathbf{X},L) = 0.$$

In terms of our particular representation of the production function, it is equivalent to

$$L = F(f(\mathbf{Y}),\mathbf{X}).$$

We shall work with separability in this form.

A profit function, $\Pi(\mathbf{p}^*,\mathbf{q}^*,w)$ is said to be separable in outputs and inputs if it can be written in the form

$$\Pi(f(\mathbf{p}^*),g(\mathbf{q}^*,w)).$$

Lemma III-2. Under Assumptions (F*.1) through (F*.7), a separable production function is almost homogeneous of degree 1 and k if and only if it can be written in the form

$$L = H_1(H_2(\mathbf{Y}),\mathbf{X}),$$

where H_1 is a homogeneous function of degree one in H_2 and $\mathbf{X}$ and H_2 is a homogeneous function of degree $1/k$.

Proof: Almost homogeneity of $F(f(\mathbf{Y}),\mathbf{X})$ implies that

$$k\frac{\partial F}{\partial f}\sum_{i=1}^{n}\frac{\partial f}{\partial Y_i}\cdot Y_i+\sum_{i=1}^{m}\frac{\partial F}{\partial X_i}\cdot X_i=F.$$

This implies that

$$\sum_{i=1}^{n}\frac{\partial f}{\partial Y_i}\cdot Y_i=\frac{F(f,\mathbf{X})-\sum_{i=1}^{m}(\partial F/\partial X_i)(f,\mathbf{X})\cdot X_i}{k(\partial F/\partial f)(f,\mathbf{X})}.$$

But the left-hand side is a function of $\mathbf{Y}$ only. Thus one must have

$$\sum_{i=1}^{n}\frac{\partial f}{\partial Y_i}\cdot Y_i=g(f).$$

By Lemma II-5, f is homothetic. Without loss of generality, one may assume that f is homogeneous of degree $1/k$, making necessary accommodations in $F(f,\mathbf{X})$. Thus

$$\sum_{i=1}^{n}\frac{\partial f}{\partial Y_i}\cdot Y_i=\frac{1}{k}f.$$

Substituting this into the original differential equation we obtain

$$\frac{\partial F}{\partial f}\cdot f+\sum_{i=1}^{m}\frac{\partial F}{\partial X_i}\cdot X_i=F,$$

that is, F is homogeneous in f and $\mathbf{X}$. It may be verified immediately that

$$H_1(H_2(\lambda^k\mathbf{Y}),\lambda\mathbf{X})=\lambda H_1(H_2(\mathbf{Y}),\mathbf{X}). \qquad \text{Q.E.D.}$$

Lemma III-3. Under Assumptions (G*.1) through (G*.7), a profit function is weakly separable if and only if it can be written in the form

$$\Pi=H(H_1(\mathbf{p}^*),H_2(\mathbf{q}^*,w)),$$

where H, H_1, H_2 are all homogeneous functions of degree 1.

Proof: $\Pi(f(\mathbf{p}^*),g(\mathbf{q}^*,w))$ is homogeneous of degree one in $\mathbf{p}^*$, $\mathbf{q}^*$, w. Thus

$$\frac{\partial\Pi}{\partial f}\cdot\sum_{i=1}^{n}\frac{\partial f}{\partial p_i^*}+\frac{\partial\Pi}{\partial g}\left(\sum_{i=1}^{m}\frac{\partial g}{\partial q_i^*}q_i^*+\frac{\partial g}{\partial w}\cdot w\right)=\Pi,$$

or

$$\sum_{i=1}^{n} \frac{\partial f}{\partial p_i^*} p_i^* = \frac{\Pi(f,g) - (\partial\Pi/\partial g)\left(\sum_{i=1}^{m} (\partial g/\partial q_i^*)q_i^* + (\partial g/\partial w)\cdot w\right)}{(\partial\Pi/\partial f)}.$$

Hence

$$\sum_{i=1}^{n} \frac{\partial f}{\partial p_i^*} p_i^* = h_1(f),$$

and

$$\sum_{i=1}^{m} \frac{\partial g}{\partial q_i^*} q_i^* + \frac{\partial g}{\partial w}\cdot w = h_2(g).$$

And $\Pi(f(\mathbf{p}^*),g(\mathbf{q}^*,w))$ may be chosen so that it is a homogeneous function of degree one of two homogeneous functions of degree one. Q.E.D.

Corollary 3.1. Under Assumptions (G*.1) through (G*.7), a profit function is weakly separable if and only if the normalized profit function can be written in the form

$$G = H(H_1(\mathbf{p}),g(\mathbf{q})),$$

where H and H_1 are homogeneous functions of degree one.

Proof:

$$\begin{aligned} G &= \frac{H(H_1(\mathbf{p}^*),H_2(\mathbf{q}^*,w))}{w} \\ &= H\left(\frac{H_1(\mathbf{p}^*)}{w}, \frac{H_2(\mathbf{q}^*,w)}{w}\right) \\ &= H(H_1(\mathbf{p}),H_2(\mathbf{q},1)) \\ &= H(H_1(\mathbf{p}),g(\mathbf{q})). \end{aligned}$$

The converse is obvious. Q.E.D.

We refer to such a normalized profit function as separable.

Corollary 3.2. Under the Assumptions (G*.1) through (G*.7), a profit function is separable in the input prices if and only if the normalized profit function can be written in the form

$$G = H(\mathbf{p},f(\mathbf{q})),$$

where H is a homogeneous function of degree one in $\mathbf{p}$ and f.

Proof: We apply Lemma III-3, treating the price of each output as a separate group. Q.E.D.

Separability of the profit function implies that the optimal output proportions are independent of input prices and *vice versa*.

> *Theorem III*-2. Under Assumptions (F*.1) through (F*.7), a production function is almost homogeneous of degree 1 and k and separable if and only if the normalized profit function is homogeneous of degree $1/(1-k)$ in $\mathbf{p}$ and separable.

Proof: Almost homogeneity is equivalent to homogeneity by Theorem III-1. By Lemma III-2,

$$\frac{\partial L/\partial Y_i}{\partial L/\partial Y_j} = \frac{(\partial H_2/\partial Y_i)(\mathbf{Y})}{(\partial H_2/\partial Y_j)(\mathbf{Y})} = \frac{p_i}{p_j},$$

where H_2 is homogeneous of degree $1/k$. Then by the now familiar argument, $(\partial G/\partial p_i)/(\partial G/\partial p_j)$ is independent of q_k, $\forall k$.

Also by Lemma III-2,

$$\frac{\partial L}{\partial X_i} = \frac{\partial H_1}{\partial X_i}(H_2,\mathbf{X}) = q_i.$$

But H_1 is homogeneous of degree one which implies that $\partial H_1/\partial X_i$ is homogeneous of degree zero. Thus,

$$\frac{\partial H_1}{\partial X_i}\left(1, \frac{X_1}{H_2},\ldots,\frac{X_m}{H_2}\right) = q_i, \qquad i = 1,\ldots,m.$$

X_i/H_2 may be solved as functions of q alone. Hence X_i/X_j or $(\partial G/\partial q_i)/(\partial G/\partial q_j)$ is independent of p. Thus, we have shown that $\Pi^* = G(f(\mathbf{p}),g(\mathbf{q}))$, where G is in addition homogeneous of degree $1/(1-k)$ in $\mathbf{p}$. By Euler's Theorem,

$$\frac{\partial G}{\partial f}\cdot\sum_{i=1}^{n}\frac{\partial f}{\partial p_i}p_i = \frac{1}{(1-k)}G.$$

By the usual argument, one can choose

$$\sum_{i=1}^{n}\frac{\partial f}{\partial p_i}p_i = \frac{1}{(1-k)}G.$$

Therefore,

$$\frac{\partial \ln G}{\partial \ln f} = 1,$$

or

$$\ln G = \ln f + h(g).$$

Thus, one has

$$G = H_1(\mathbf{p})^{1/(1-k)} g^*(\mathbf{q}),$$

and hence G is separable.

To prove the converse, note that separability of the normalized profit function implies, by Lemma III-3, that it has the form

$$G = H(H_1(\mathbf{p}), g(\mathbf{q})).$$

Homogeneity of G of degree $1/(1-k)$ in $\mathbf{p}$ implies that

$$\frac{\partial G}{\partial H_1} \cdot H_1 = \frac{1}{(1-k)} G.$$

Hence

$$G(\mathbf{p},\mathbf{q}) = H_1(\mathbf{p})^{1/(1-k)} g^*(\mathbf{q}).$$

It then follows from homogeneity of $H_1(\mathbf{p})$ by the usual argument that $(\partial F/\partial Y_i)/(\partial F/\partial Y_j)$ is independent of $\mathbf{X}$. Almost homogeneity of F follows from homogeneity of $G(\mathbf{p},\mathbf{q})$ in $\mathbf{p}$. Q.E.D.

Corollary 2.1. Under Assumptions (F*.1) through (F*.7), a production function is almost homogeneous and separable if and only if the profit function can be written in the form

$$\Pi(\mathbf{p}^*,\mathbf{q}^*,w) = H_1(\mathbf{p}^*)^{1/(1-k)} H_2(\mathbf{q}^*,w)^{-k/(1-k)}.$$

Proof: This result is obtained by straightforward substitution. Q.E.D.

Theorem III-3. Under Assumptions (F*.1) through (F*.7), a production function is homothetically separable in outputs if and only if the normalized profit function is homothetically separable in output prices.

Proof: By hypothesis, $L = F(H(\mathbf{Y}),\mathbf{X})$. Thus,

$$\frac{\partial H/\partial Y_i}{\partial H/\partial Y_j} = \frac{p_i}{p_j}.$$

Hence the optimal output ratios, $(\partial G/\partial p_i)/(\partial G/\partial p_j)$ may be solved in terms of (p_i/p_j)'s alone. Moreover, they are homogeneous of degree zero in $\mathbf{p}$. The normalized profit function is therefore homothetically separable in output prices. The converse is similarly proved. Q.E.D.

Theorem III-4. Under Assumptions (F*.1) through (F*.7), a production function is homothetically separable in inputs if and only if the normalized profit function is separable in input prices, that is, has the form

$$G = g(\mathbf{q})g^*\left(\frac{\mathbf{p}}{g(\mathbf{q})}\right).$$

Proof: Homothetic separability in inputs implies that

$$L = F(\mathbf{Y},\mathbf{X}) = H(f(\mathbf{Y}),\mathbf{X}),$$

where H is homogeneous of degree one in $f(\mathbf{Y})$ and $\mathbf{X}$. This can be alternatively written as

$$L = f(\mathbf{Y})g\left(\frac{\mathbf{X}}{f(\mathbf{Y})}\right).$$

The first-order necessary conditions for a maximum are

$$\frac{\partial L}{\partial X_i} = g_i\left(\frac{X}{f(\mathbf{Y})}\right) = q_i, \qquad i = 1,\ldots,m.$$

Thus, one may solve $X_i/f(\mathbf{Y})$ as unique and continuously differentiable functions of the $\mathbf{q}$ alone. This implies

$$X_i = g_i^*(\mathbf{q})f(\mathbf{Y}).$$

Substituting this into the production function we have

$$\begin{aligned}\Pi^* &= \max_{\mathbf{Y}}\{p'\mathbf{Y} + f(\mathbf{Y})g(\mathbf{q})\}\\ &= g(\mathbf{q})g^*\left(\frac{\mathbf{p}}{g(\mathbf{q})}\right),\end{aligned}$$

By Theorem II-3. The converse is proved similarly. Q.E.D.

Theorem III-5. Under Assumptions (F*.1) through (F*.7), a production function is homothetically separable in both outputs and inputs if and only if

$$G(\mathbf{p},\mathbf{q}) = g(\mathbf{q})g^*\left(\frac{H(\mathbf{p})}{g(\mathbf{q})}\right).$$

where H is a homogeneous function of degree one.

Proof: This theorem follows directly from the two previous theorems. Q.E.D.

Theorem III-6. Under Assumption (F*.1) through (F*.7), a production function and its corresponding normalized profit function are both separable in outputs only if either they are homothetically separable in outputs or they are additive.

Proof:

$$F(\mathbf{Y},\mathbf{X}) = F(f(\mathbf{Y}),\mathbf{X}),$$
$$G(\mathbf{p},\mathbf{q}) = G(g(\mathbf{p}),\mathbf{q}).$$

The first-order necessary conditions for a maximum imply

$$\frac{f_i(Y_1,\ldots,Y_n)}{f_j(Y_1,\ldots,Y_n} = \frac{p_i}{p_j}.$$

By a dual transformation, this becomes

$$\frac{f_i(\partial G/\partial p_1,\ldots,\partial G/\partial p_n)}{f_j(\partial G/\partial p_1,\ldots,\partial G/\partial p_n)} = \frac{p_i}{p_j}.$$

Differentiating these equations with respect to q_k, we obtain

$$f_i \sum_{l=1}^{n} f_{jl}G_{lk} - f_j \sum_{l=1}^{n} f_{il}G_{lk} = 0.$$

But

$$G_{lk} = G_{gk}g_l, \qquad G_l = G_g g_l = Y_l.$$

Hence

$$f_i \sum_{l=1}^{n} f_{jl}Y_l - f_j \sum_{l=1}^{n} f_{il}Y_l = 0,$$

which means f_i/f_j is homogeneous of degree zero in $\mathbf{Y}$. Hence f is homothetic in $\mathbf{Y}$ by Lemma II-1. It follows from Theorem II-9 that $g(\mathbf{p})$ is also homothetic.

An exceptional case arises if $G_{gk} = 0$, $\forall k$. Then

$$G(\mathbf{p},\mathbf{q}) = g(\mathbf{p}) + h^*(\mathbf{q}).$$

And by Theorem II-7,

$$F(\mathbf{Y},\mathbf{X}) = f(\mathbf{Y}) + h(\mathbf{X}). \qquad \text{Q.E.D.}$$

Extension to the case with fixed inputs is straightforward and will not be repeated here.

3.3. *Non-Jointness in Production*

The problem of non-jointness has been investigated by Samuelson (1966) who derives necessary and sufficient conditions for a production function to represent a non-joint technology. Hall (1973) has approached the problem using the joint cost function. It turns out that the assumption of non-jointness of the technology implies very simple restrictions on the matrix of second partial derivatives of the normalized profit function. We shall present these results. First of all we give a definition.

Definition. A production function $L = F(\mathbf{Y},\mathbf{X})$ is said to be *non-joint in inputs* if there exist individual production functions,

$$L_i = f_i(Y_i, X_{1i}, X_{2i}, \ldots, X_{mi}), \qquad i = 1,\ldots,n.$$

such that

$$F(\mathbf{Y},\mathbf{X}) = \min\left\{ \sum_{i=1}^{n} f_i(Y_i, X_{1i}, X_{2i}, \ldots, X_{mi}) \,\middle|\, \sum_{i=1}^{n} X_{ji} \geqq X_j, \quad j = 1,\ldots,m \right\}.$$

A production function is said to be *non-joint in outputs* if there exists individual production functions,

$$L = g_0(Y_{10}, \ldots, Y_{n0}),$$
$$X_i = g_i(Y_{1i}, \ldots, Y_{ni}), \qquad i = 1,\ldots,m,$$

such that

$$F(\mathbf{Y},\mathbf{X}) = \min\left\{ g_0(Y_{10}, \ldots, Y_{n0}) | X_i \leqq g_i(Y_{1i}, \ldots, Y_{ni}), \qquad i = 1,\ldots,m, \right.$$
$$\left. \sum_{i=0}^{m} Y_{ji} \geqq Y_j, \quad j = 1,\ldots,n \right\}.$$

The minimum in these two definitions ensure that all the inputs (and outputs) are allocated amongst the individual industries so that production is efficient, that is, the output of no one industry may be increased without decreasing the output of another industry. (And no one input may be decreased without increasing another input.)

The normalized profit function of a technology characterized by non-jointness in inputs has a very simple representation: it is the sum of the individual normalized profit functions corresponding to the individual industry production functions. This is embodied in the following theorem:

Theorem III-7. Under Assumptions (F*.1) through (F*.7), a production function is non-joint in inputs if and only if its normalized profit function is additive in **p**, that is,

$$\Pi^* = \sum_{i=1}^{n} G_i(p_i, \mathbf{q}).$$

Proof:
Necessity.

$$\begin{aligned}\Pi^* &= \max_{Y_i, X_{ij}} \left\{ \sum_{i=1}^{n} p_i Y_i + \sum_{j=1}^{m} q_j \sum_{i=1}^{n} X_{ji} - \sum_{i=1}^{n} f_i(Y_i, X_{1i}, \ldots, X_{mi}) \right\} \\ &= \sum_{i=1}^{n} \max_{Y_i, X_{ij}} \left\{ p_i Y_i + \sum_{j=1}^{m} q_j X_{ji} - f_i(Y_i, X_{1i}, \ldots, X_{mi}) \right\} \\ &= \sum_{i=1}^{n} G_i(p_i, \mathbf{q}).\end{aligned}$$

Sufficiency. Given $\Pi^* = \sum_{i=1}^{n} G_i(p_i, \mathbf{q})$, one can find for each $G_i(p_i, \mathbf{q})$ a unique production function $L_i = f_i(Y_i, X_{1i}, \ldots, X_{mi})$. Thus, the technology is non-joint in inputs. Q.E.D.

Corollary 7.1. Under Assumptions (F*.1) through (F*.7), a production function is non-joint in inputs if and only if

$$\frac{\partial^2 G}{\partial p_i \partial p_j} = 0, \qquad i \neq j, \quad \forall i,j,$$

where G is the normalized profit function and **p** is the vector of normalized output prices.[19]

Proof: This follows directly from the theorem. Q.E.D.

[19]This condition is also given by Diewert (1973a).

Note that duality of $\sum_{i=1}^{n} G_i(p_i,\mathbf{q})$ to a non-joint in inputs technology is an immediate consequence of the convolution theorem for profit functions in Chapter I.1.

Corollary 7.1 provides a very useful necessary and sufficient condition for the characterization of a "non-joint in inputs" technology. In particular, it lends itself to straightforward empirical tests. In retrospect: it turns out that our conditions here are completely equivalent to the conditions stated by Samuelson (1966) on the Hessian of the production function. One needs only recall from Section 1 that the Hessian matrix of the normalized profit function $G(\mathbf{p},\mathbf{q})$ is the inverse of the Hessian matrix of $F(\mathbf{Y},\mathbf{X})$. Hence singularity conditions on the minors of the Hessian matrix of F are equivalent to zero conditions on the elements of the Hessian of $G(\mathbf{p},\mathbf{q})$.

On the other hand, in the case of non-jointness in outputs, it is easy to see that the normalized profit function is given by

$$\Pi^* = G_0(\mathbf{p}) + \sum_{i=1}^{m} q_i G_i\left(\frac{\mathbf{p}}{q_i}\right).$$

Theorem III-8. Under Assumptions (F*.1) through (F*.7), a production function is non-joint in outputs if and only if its normalized profit function can be written in the form

$$\Pi^* = G_0(\mathbf{p}) + \sum_{i=1}^{m} q_i G_i\left(\frac{\mathbf{p}}{q_i}\right).$$

Proof: Obvious. Q.E.D.

3.4. *Summary*

We may summarize the results of Sections 3.2 and 3.3 by way of a table which describes the restrictions on the normalized profit functions under alternative combinations of assumptions on the technology. The alternatives considered are as follows:[20]

(1) almost homogeneity of the production function,
(2) direct separability,

[20]This table is different from that of Lau (1972) in two respects: first, the forms are specified in terms of the normalized profit function; second, some of the errors have been corrected and "open" questions have been closed.

TABLE 1
Functional forms of normalized profit functions under alternative assumptions.

(1)	$H_p^k(\mathbf{p},\mathbf{q})$
(2)	$\sup_\lambda\{H_p(\lambda,\mathbf{p})-G(\lambda,\mathbf{q})\}$
(3)	$G(\mathbf{q})G^*\left(\frac{H(\mathbf{p})}{G(\mathbf{q})}\right)$
(4)	$\sum_{i=1}^n G_i(p_i,\mathbf{q})$
(5)	$G_0(\mathbf{p})+\sum_{i=1}^m q_iG_i\left(\frac{\mathbf{p}}{q_i}\right)$
(1)+(2)	$H^k(\mathbf{p})G(\mathbf{q})$
(1)+(3)	$H^k(\mathbf{p})G(\mathbf{q})$
(1)+(4)	$\sum_{i=1}^n p_i^kG_i(\mathbf{q})$
(1)+(5)	$H_0^k(\mathbf{p})+\sum_{i=1}^m q_iH_i^k\left(\frac{\mathbf{p}}{q_i}\right)$
(2)+(3)	Same as (3)
(2)+(4)	$F(\mathbf{q})\sum_{i=1}^n G_i\left(\frac{p_i}{F(\mathbf{q})}\right)+G(\mathbf{q})$
(2)+(5)	$G_0(H(\mathbf{p}))+\sum_{i=1}^m q_iG_i\left(\frac{H(\mathbf{p})}{q_i}\right)+H^*(\mathbf{p})$
(3)+(4)[a]	$\left[\sum_{i=1}^n \alpha_ip_i^k\right]G(\mathbf{q})^{1-k}+G(\mathbf{q})$
(3)+(5)[b]	$\left[\alpha_0+\sum_{i=1}^m \alpha_iq_i^k\right]H(\mathbf{p})^{1-k}+H(\mathbf{p})$ or $\left[\alpha_0+\sum_{i=1}^m \alpha_i\ln\left(\frac{q_i}{H(\mathbf{p})}\right)\right]H(\mathbf{p})$
(4)+(5)	$\sum_{i=1}^n\left[g_{i0}(p_i)+\sum_{j=1}^m q_jg_{ij}^*\left(\frac{p_i}{q_j}\right)\right]$
(1)+(2)+(3)	$H^k(\mathbf{p})G(\mathbf{q})$
(1)+(2)+(4)	Same as (3)+(4)
(1)+(2)+(5)	Same as (3)+(5)
(1)+(3)+(4)	Same as (3)+(4)
(1)+(3)+(5)	Same as (3)+(5)
(3)+(4)+(5)[c]	$\left[\sum_{i=1}^n \alpha_ip_i\right]+\left[\beta_0+\sum_{i=1}^m \beta_iq_i\right]$

[a]This implies that the production functions for all outputs are identical up to a multiplicative constant. Both Denny (1972) and Hall (1973) have independently discovered this result in the context of joint cost functions.

[b]Likewise, this implies that all the production functions are identical up to a multiplicative constant.

[c]This implies a fixed-coefficients type technology. Note that this violates our local strong convexity assumption. In principle,

$$\left[\alpha_0+\sum_{i=1}^m \alpha_iq_i^{1-k}\right]\left[\sum_{i=1}^n \beta_ip_i^k\right]$$

is a possible solution. However, this solution cannot satisfy the monotonicity and convexity conditions simultaneously.

(3) indirect separability,
(4) non-jointness in inputs,
(5) non-jointness in outputs.

$F(\cdot)$ and $G(\cdot)$ are used to denote arbitrary functions: $H(\cdot)$, $H^*(\cdot)$ and $H^{**}(\cdot)$ are used to denote arbitrary homogeneous functions; and $H'(\cdot)$ is used to denote arbitrary homothetic functions. A subscript denotes the set of variables in which the function is homogeneous or homothetic. A superscript denotes the degree of homogeneity when it is different from one.

Many of these combinations are obvious. We shall derive three of the relatively less obvious ones.

3.4.1. *Derivation of* (2) + (4) *and* (2) + (5)

Direct separability implies that the normalized cost function of producing $f(\mathbf{Y})$ can be written as

$$C^* = G(f(\mathbf{Y}),\mathbf{q}). \tag{III-1}$$

Non-jointness in inputs implies that the normalized cost functions can be written as

$$C^* = \sum_{i=1}^{n} f_i(Y_i,\mathbf{q}). \tag{III-2}$$

We note that C^* is characterized by $(\partial^2 C^*)/(\partial Y_i \partial Y_j) = 0$, $i \neq j$. Differentiating equation (III-1), we obtain

$$\frac{\partial^2 C^*}{\partial Y_i \partial Y_j} = \frac{\partial G}{\partial f} \cdot \frac{\partial^2 f}{\partial Y_i \partial Y_j} + \frac{\partial^2 G}{\partial f^2} \cdot \frac{\partial f}{\partial Y_i} \frac{\partial f}{\partial Y_j} = 0, \qquad i \neq j,$$

which implies

$$\frac{\partial^2 G/\partial f^2}{\partial G/\partial f} = -\frac{(\partial^2 f)/(\partial Y_i \partial Y_j)}{(\partial f/\partial Y_i)/(\partial f/\partial Y_j)}, \quad \text{if} \quad \frac{\partial^2 G}{\partial f^2} \neq 0,$$

or

$$\frac{\partial^2 f}{\partial Y_i \partial Y_j} = 0, \qquad\qquad \text{if} \quad \frac{\partial^2 G}{\partial f^2} = 0.$$

We note that in the first case the right-hand side of the equation is independent of $\mathbf{q}$. Hence the left-hand side is independent of $\mathbf{q}$. Further we observe that the left-hand side may be written as $(\partial/\partial f)\ln(\partial G/\partial f)$

which is a function of f alone. Thus, by successive integration we obtain

$$G(f,\mathbf{q}) = g(f)h_1(\mathbf{q}) + h_2(\mathbf{q}),$$

which becomes, in order to satisfy equation (III-2),

$$C^* = \sum_{i=1}^{n} g_i(Y_i)h_1(\mathbf{q}) + h_2(\mathbf{q}).$$

Note that this implies that the isoquants of each industry have the same shape although the numberings may differ.

Now the normalized profit function is given by

$$\begin{aligned}\Pi^* &= \max_{\mathbf{Y}} \left\{\sum_{i=1}^{n} p_iY_i - \sum_{i=1}^{n} g_i(Y_i)h_1(\mathbf{q}) - h_2(\mathbf{q})\right\} \\ &= \sum_{i=1}^{n} \max_{Y_i}\{p_iY_i - g_i(Y_i)h_1(\mathbf{q})\} - h_2(\mathbf{q}) \\ &= h_1(\mathbf{q}) \sum_{i=1}^{n} \max_{Y_i} \left\{\frac{p_i}{h_1(\mathbf{q})} Y_i - g_i(Y_i)\right\} - h_2(\mathbf{q}) \\ &= h_1(\mathbf{q})\sum_{i=1}^{n} g_i^*\left(\frac{p_i}{h_1(\mathbf{q})}\right) - h_2(\mathbf{q}).\end{aligned}$$ [21]

In the second case, we also have

$$C^* = \sum_{i=1}^{n} f_i(Y_i)h_1(\mathbf{q}) + h_2(\mathbf{q}).$$

The condition for (2) + (5) may be derived similarly.

3.4.2. *Derivation of* (3) + (4) *and* (3) + (5)

Indirect separability implies that the normalized profit function can be written as

$$\Pi^* = H(H^*(\mathbf{p}),f(\mathbf{q})) = f(\mathbf{q})G\left(\frac{H^*(\mathbf{p})}{f(\mathbf{q})}\right).$$

Non-jointness in inputs implies that the normalized profit function can be written as

$$\Pi^* = \sum_{i=1}^{n} F_i(p_i,\mathbf{q}).$$

[21]This is precisely the form suggested by Professor W. M. Gorman to the author in 1970. At that time the author was unable to establish the necessity of this form. See Lau (1972, p. 288, fn. 20).

We note that Π^* is characterized by $(\partial^2\Pi^*)/(\partial p_i \partial p_j) = 0$, $i \neq j$. This implies that

$$\frac{G''H_i^*H_j^*}{f} + G'H_{ij}^* = 0, \qquad G'' \neq 0,$$

or

$$\frac{G''(H^*/f)}{G'} + \frac{H_{ij}^*H^*}{H_i^*H_j^*} = 0, \qquad G'' \neq 0, \qquad \text{(III-3)}$$

and

$$H_{ij}^* = 0, \quad i \neq j, \qquad G'' = 0.$$

The second term of equation (III-3) is independent of f, which implies that the first term is independent of f. But the first term being independent of f means it is independent of H^*/f, since $G(\cdot)$ is a function of a single variable H^*/f only. Hence the first term must be constant, that is

$$\frac{G''}{G'}(H^*/f) = k,$$

k a constant. This equation may be integrated to yield,

$$\begin{aligned} G(Z) &= C_1 \frac{Z^{k+1}}{k+1} + C_2, \qquad k \neq -1, \\ &= C_1 \ln Z + C_2, \qquad k = -1. \end{aligned}$$

For $k \neq -1$, one must have

$$\frac{C_1 H^*(\mathbf{p})^{k+1}}{k+1}$$

be additive and homogeneous of degree $k+1$ in $\mathbf{p}$. This means

$$C_1 H^*(\mathbf{p})^{k+1} = \sum_{i=1}^{n} \alpha_i p_i^{k+1}.$$

For $k = -1$, one must have

$$C_1 \ln H^*(\mathbf{p}) + C_2$$

be additive and homothetic, which implies

$$C_1 \ln H^*(\mathbf{p}) = \sum_{i=1}^{n} \alpha_i \ln p_i.$$

Note, however, that this is inconsistent with the monotonicity and convexity requirements of the normalized profit function. Thus, the only possibility is that of

$$\Pi^* = f(q)\left[\alpha_0 + \sum_{i=1}^{n} \alpha_i \left(\frac{p_i}{f(\mathbf{q})}\right)^{k+1}\right].$$

If in addition we require that $\Pi^* = 0$ if $p_i = 0, \forall i$, then

$$\Pi^* = \left[\sum_{i=1}^{n} \alpha_i p_i^{k^*}\right] f(\mathbf{q})^{1-k^*}.$$

Alternatively, if $G'' = 0$, $H_{ij}^* = 0$, which implies that the normalized profit function must have the form

$$\begin{aligned}\Pi^* &= f(\mathbf{q})\left[\alpha_0 + \frac{\sum_{i=1}^{n} \alpha_i p_i}{f(\mathbf{q})}\right] \\ &= \sum_{i=1}^{n} \alpha_i p_i + f(\mathbf{q})\alpha_0.\end{aligned}$$

The condition for (3) + (5) may be derived similarly.

3.4.3. *Derivation of* (4) + (5)

$$\Pi^* = \sum_{i=1}^{n} G_i(p_i, \mathbf{q}) \tag{III-4}$$

$$= G_0(\mathbf{p}) + \sum_{i=1}^{m} q_i G_i\left(\frac{\mathbf{p}}{q_i}\right). \tag{III-5}$$

Since $p_i = 0$ implies $G_i = 0$ in equation (III-4), then $(\partial^2 \Pi^*)/(\partial q_i \partial q_j) = 0$ implies that each G_i must have the form

$$g_{i0}(p_i) + \sum_{j=1}^{m} g_{ij}(p_i, q_j).$$

Substituting this into equation (III-4) leads to

$$\Pi^* = \sum_{i=1}^{n} \left[g_{i0}(p_i) + \sum_{j=1}^{m} q_j g_{ij}^*\left(\frac{p_i}{q_j}\right)\right].$$

4. Examples of Normalized Profit Functions

4.1. *Introduction*

In this section we present a number of examples of normalized profit functions. In particular, we demonstrate how the theorems derived in Sections 2 and 3 and the Legendre transformation may be used in the construction of the normalized profit function given the production function (and *vice versa*).

4.2. *Cobb–Douglas Production Function*

Let

$$Y = \prod_{i=1}^{m} X_i^{\alpha_i}.$$

The first-order necessary conditions for a maximum are

$$\frac{\alpha_i Y}{X_i} = q_i, \qquad i = 1,\ldots,m. \tag{IV-1}$$

By Theorem (II-1), $Y = (1-\mu)^{-1}G$ where $\mu = \sum_{i=1}^{m} \alpha_i (< 1)$ because Y is homogeneous of degree μ in $\mathbf{X}$. Hence, by a dual transformation, equation (IV-1) becomes

$$\frac{\alpha_i (1-\mu)^{-1} G}{-\partial G/\partial q_i} = q_i, \qquad i = 1,\ldots,m,$$

which may be integrated as

$$G(\mathbf{q}) = A^* \prod_{i=1}^{m} q_i^{\alpha_i^*},$$

where $\alpha_i^* = -\alpha_i(1-\mu)^{-1}$, $i = 1,\ldots,m$, and A^* is a constant of integration. A^* may be determined from initial conditions. For instance, at $q_i = 1$, $i = 1,\ldots,m$, equation (IV-1) implies that $X_i = \alpha_i Y$, $i = 1,\ldots,m$. Substituting this into the production function we have

$$Y = \prod_{i=1}^{m} X_i^{\alpha_i} = \prod_{i=1}^{m} \alpha_i^{\alpha_i} Y^{\alpha_i}.$$

Therefore,

$$
\begin{aligned}
Y &= \left(\prod_{i=1}^{m} \alpha_i^{\alpha_i}\right)^{1/(1-\mu)} \\
&= (1-\mu)^{-1} G(1) \\
&= (1-\mu)^{-1} A^*.
\end{aligned}
$$

Thus,

$$
A^* = (1-\mu) \prod_{i=1}^{m} \alpha_i^{\alpha_i(1-\mu)^{-1}},
$$

and

$$
G(\mathbf{q}) = (1-\mu) \prod_{i=1}^{m} \left(\frac{q_i}{\alpha_i}\right)^{-\alpha_i(1-\mu)^{-1}}
$$

To extend this to the case with fixed inputs, we have

$$
Y = \prod_{i=1}^{m} X_i^{\alpha_i} \prod_{i=1}^{n} Z_i^{\beta_i}.
$$

Then, by applying Theorem (II-3), one has immediately,

$$
\begin{aligned}
G(\mathbf{q},\mathbf{Z}) &= \prod_{i=1}^{n} Z_i^{\beta_i} (1-\mu) \prod_{i=1}^{m} \left(\frac{q_i}{\alpha_i \prod_{i=1}^{n} Z_i^{\beta}}\right)^{-\alpha_i(1-\mu)^{-1}} \\
&= (1-\mu) \prod_{i=1}^{m} \left(\frac{q_i}{\alpha_i}\right)^{-\alpha_i(1-\mu)^{-1}} \prod_{i=1}^{n} Z_i^{\beta_i(1-\mu)^{-1}}
\end{aligned}
$$

For this latter case, the supply function is, again by Theorem (II-1),

$$
\begin{aligned}
Y(\mathbf{q},\mathbf{Z}) &= (1-\mu)^{-1} G \\
&= \prod_{i=1}^{m} \left(\frac{q_i}{\alpha_i}\right)^{-\alpha_i(1-\mu)^{-1}} \prod_{i=1}^{n} Z_i^{\beta_i(1-\mu)^{-1}},
\end{aligned}
$$

and the derived demand functions are

$$
\begin{aligned}
X_j &= -\frac{\partial G}{\partial q_j} \\
&= \frac{\alpha_j}{q_j}\left[\prod_{i=1}^{m} \left(\frac{q_i}{\alpha_i}\right)^{-\alpha_i(1-\mu)^{-1}} \prod_{i=1}^{n} Z_i^{\beta_i(1-\mu)^{-1}}\right], \qquad j = 1,\ldots,m.
\end{aligned}
$$

We note also that for the Cobb–Douglas production function the expenditure on each variable input is a constant proportion of profits. This follows from

$$
\frac{q_i X_i}{G} = -\frac{\partial \ln G}{\partial \ln q_i} = \alpha_i (1-\mu)^{-1}, \qquad i = 1,\ldots,m.
$$

4.3. *The C.E.S. Production Function*

We have

$$Y = \left[\sum_{i=1}^{m} \alpha_i X_i^{\rho}\right]^{\mu/\rho},$$

where $\mu < 1$ is a scale parameter.

The necessary conditions for a maximum are

$$\frac{\partial Y}{\partial X_j} = \mu \left[\sum_{i=1}^{m} \alpha_i X_i^{\rho}\right]^{(\mu-\rho)/\rho} \alpha_j X_j^{\rho-1}$$
$$= \mu Y^{(\mu-\rho)/\mu} \alpha_j X_j^{\rho-1} = q_j, \qquad j = 1,\ldots,m,$$

which may be rewritten as

$$-\mu^{1/(\rho-1)}[(1-\mu)^{-1}G]^{(\mu-\rho)/\mu(\rho-1)} \alpha_j^{1/(\rho-1)} \frac{\partial G}{\partial q_j} = q_j^{1/(\rho-1)}, \qquad j = 1,\ldots,m,$$

which becomes

$$-\mu^{1/(\rho-1)}(1-\mu)^{-(\mu-\rho)/\mu(\rho-1)} \cdot \frac{\mu(\rho-1)}{\rho(\mu-1)} \frac{\partial G^{\rho(\mu-1)/\mu(\rho-1)}}{\partial q_j} = \left(\frac{q_j}{\alpha_j}\right)^{1/(\rho-1)},$$

which may be integrated as[22]

$$\mu^{\rho/(\rho-1)}(1-\mu)^{\rho(1-\mu)/\mu(\rho-1)} \cdot G^{-\rho(1-\mu)/\mu(\rho-1)} = \sum_{i=1}^{m} \alpha_i \left(\frac{q_i}{\alpha_i}\right)^{\rho/(\rho-1)}.$$

Hence

$$G(\mathbf{q}) = \mu^{\mu(1-\mu)^{-1}}(1-\mu)\left[\sum_{i=1}^{m} \alpha_i \left(\frac{q_i}{\alpha_i}\right)^{\rho/(\rho-1)}\right]^{-(\mu(1-\mu)^{-1}(\rho-1))/\rho}.$$

By Theorem (II-1) the supply function is immediately given by[23]

$$Y = (1-\mu)^{-1}G(\mathbf{q})$$
$$= \mu^{\mu(1-\mu)^{-1}}\left[\sum_{i=1}^{m} \alpha_i \left(\frac{q_i}{\alpha_i}\right)^{\rho/(\rho-1)}\right]^{-(\mu(1-\mu)^{-1}(\rho-1))/\rho},$$

and the derived demand functions by[24]

$$X_i = \mu^{(1-\mu)^{-1}}\left[\sum_{i=1}^{m} \alpha_i \left(\frac{q_i}{\alpha_i}\right)^{\rho/(\rho-1)}\right]^{((1-\mu)^{-1}(\mu-\rho))/\rho} \left(\frac{q_i}{\alpha_i}\right)^{1/(\rho-1)}, \qquad i = 1,\ldots,m.$$

[22]There is no constant of integration because by Theorem (II-1), $G(\mathbf{q})$ is homogeneous.
[23]Equivalent expressions have been obtained by McFadden (1966) and Nerlove (1967).
[24]See the discussion in Lau (1969c, pp. 30–33).

However, if some inputs are fixed, then the normalized profit function corresponding to a C.E.S. production function may not have a closed form solution. It is, however, still implicitly defined.

4.4. *Combination C.E.S. – Cobb–Douglas Production Function*

In view of the analytic intractability of the C.E.S. production function with fixed input levels, functions which are hybrids of the C.E.S. and the Cobb–Douglas functions may be used. Some examples are[25]

(1) $$Y = \prod_{i=1}^{m} X_i^{\alpha_i} \left[\sum_{j=1}^{n} \beta_j Z_j \right]^{\mu_2/\rho},$$

with

$$\Pi^* = (1-\mu_1) \prod_{i=1}^{m} \left(\frac{q_i}{\alpha_i} \right)^{-\alpha_i(1-\mu_1)^{-1}} \left[\sum_{j=1}^{n} \beta_j Z_j^{\rho} \right]^{(\mu_2(1-\mu_1)^{-1})/\rho},$$

where

$$\mu_1 = \sum_{i=1}^{m} \alpha_i < 1;$$

(2) $$Y = \left[\sum_{i=1}^{m} \alpha_i X_i^{\rho} \right]^{\mu/\rho} \prod_{j=1}^{n} Z_j^{\beta_j},$$

with

$$\Pi^* = \mu^{\mu/(1-\mu)}(1-\mu) \left[\sum_{i=1}^{m} \alpha_i \left(\frac{q_i}{\alpha_i} \right)^{\rho/(\rho-1)} \right]^{-\mu/(1-\mu)\cdot(\rho-1)/\rho} \left[\prod_{j=1}^{m} Z_j^{\beta_j} \right]^{(1-\mu)^-}$$

(3) $$Y = \left[\sum_{i=1}^{m} \alpha_i X_i^{\rho_1} \right]^{\mu_1/\rho_1} \left[\sum_{j=1}^{n} \beta_j Z_j^{\rho_2} \right]^{\mu_2/\rho_2},$$

with

$$\Pi^* = \mu_1^{\mu_1(1-\mu_1)}(1-\mu_1) \left[\sum_{i=1}^{m} \alpha_i \left(\frac{q_i}{\alpha_i} \right)^{\rho_1/(\rho_1-1)} \right]^{-\mu_1(1-\mu_1)\cdot(\rho_1-1)\rho_1} \times \left[\sum_{j=1}^{n} \beta_j Z_j^{\rho_2} \right]^{(\mu_2/\rho_2)\cdot 1/(1-\mu_1)}.$$

Alternatively, one may specify the normalized restricted profit function directly.

[25]The dual functions may be derived by using either Theorem II-3 or the composition theorems given in Chapter I.1.

4.5. *Quadratic Production Function*

Thus far we have considered only those production functions which satisfy our assumptions globally. However, we shall now consider some production functions (and normalized restricted profit functions) which satisfy our assumptions only over proper convex subsets of R^m (or $R^m \times R^n$ as the case may be). First we consider the quadratic function. Adhering now to the conventions of the multiple-output, multiple-input case, we define

$$L = \alpha_0 + \sum_{i=1}^{m} \alpha_i X_i + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_{ij} X_i X_j,$$

where X_i is a net output which may be either positive or negative. If the matrix $\mathbf{B} \equiv [\beta_{ij}]$ is required to be positive definite, then L is strongly convex on R^m and in particular on any convex subset of R^m. Monotonicity requires that

$$\boldsymbol{\alpha} + \mathbf{BX} \geqq 0.$$

This system of linear inequalities defines the convex set of $\mathbf{X}$ such that L is monotonic and convex. The constant α_0 may then be adjusted so that L is non-negative on this convex set. Alternatively, one may set $\alpha_0 = 0$ and L is then non-negative, monotonic, and strongly convex on the convex set such that

$$\boldsymbol{\alpha}'\mathbf{X} + \frac{\mathbf{X}'B\mathbf{X}}{2} \geqq 0,$$

$$\boldsymbol{\alpha} + \mathbf{BX} \geqq 0.$$

The quadratic production function has the very convenient property of being *self-dual*, that is, its convex conjugate, the normalized profit function, is also a quadratic function,

$$\Pi^* = -\alpha_0 + \tfrac{1}{2}(\mathbf{p} - \boldsymbol{\alpha})'\mathbf{B}^{-1}(\mathbf{p} - \boldsymbol{\alpha}),$$

where p_i may be the price of a net output or net input. The domain of Π^* is given by the support function of the domain of L, defined above.

The derived supply and demand functions for all commodities other than L are linear functions

$$\mathbf{X} = \frac{\partial G}{\partial \mathbf{p}} = \mathbf{B}^{-1}(\mathbf{p} - \boldsymbol{\alpha}),$$

and

$$\begin{aligned} L &= \alpha_0 - \tfrac{1}{2}(\mathbf{p}-\boldsymbol{\alpha})'\mathbf{B}^{-1}(\mathbf{p}-\boldsymbol{\alpha}) + \mathbf{p}'\mathbf{B}^{-1}(\mathbf{p}-\boldsymbol{\alpha}) \\ &= \alpha_0 + \tfrac{1}{2}(\mathbf{p}-\boldsymbol{\alpha})'\mathbf{B}^{-1}(\mathbf{p}-\boldsymbol{\alpha}) + \boldsymbol{\alpha}'\mathbf{B}^{-1}(\mathbf{p}-\boldsymbol{\alpha}). \end{aligned}$$

The quadratic function may be further generalized as follows. Suppose that the production function is given by

$$L = \frac{1}{\theta}(\mathbf{X}'\mathbf{B}\mathbf{X})^{\theta/2}, \qquad 1 < \theta < +\infty.$$

Then the normalized profit function is given by

$$\Pi^* = \frac{1}{\eta}(\mathbf{p}'\mathbf{B}^{-1}\mathbf{p})^{\eta/2}, \qquad 1 < \eta < +\infty,$$

where

$$\frac{1}{\theta} + \frac{1}{\eta} = 1.$$

Again, the domain of L and its conjugate Π^* need to be appropriately restricted. If the monotonicity assumption is maintained then $\mathbf{BX} \geqq 0$ defines the domain of L.

4.6. *The Exponential Production Function*

Let

$$Y = 1 - e^{-X}, \qquad X \geqq 0.$$

Then

$$\begin{aligned} \Pi^* &= 1 - q + q \ln q, \\ X &= -\ln q. \end{aligned}$$

We note that for $q > 1$, there is no solution X such that $X \geqq 0$.

4.7. *The "Addilog" Normalized Profit Function*

We next consider normalized profit functions for which an explicit dual production function does not exist. One such example is the indirect addilog function introduced by Houthakker (1960). The normalized profit function is given by

$$\Pi^* = \sum_{i=1}^{m} \alpha_i q_i^{-\beta_i}.$$

The derived supply and demand functions are given by

$$Y = \sum_{i=1}^{m} \alpha_i q_i^{-\beta_i} + \sum_{i=1}^{m} \alpha_i \beta_i q_i^{-\beta_i}$$

$$= \sum_{i=1}^{m} \alpha_i (1+\beta_i) q_i^{-\beta_i},$$

$$X_i = \alpha_i \beta_i q_i^{-(\beta_i+1)}, \qquad i = 1,\dots,m.$$

The restrictions for monotonicity and convexity for the single-output case are

$$\alpha_i \beta_i > 0, \qquad \beta_i > -1.$$

4.8. *Reciprocal Quadratic Normalized Profit Function*

$$\Pi^* = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_{ij} q_i^{-1} q_j^{-1}.$$

The derived supply and demand functions have a remarkably simple form

$$Y = \frac{3}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_{ij} q_i^{-1} q_j^{-1},$$

$$X_i = q_i^{-2} \sum_{j=1}^{m} \beta_{ij} q_j^{-1}, \qquad i = 1,\dots,m.$$

It may be verified directly that if $\beta_{ij} \geqq 0$, $\forall i,j$, then Π^* is non-negative, non-increasing, and convex. Also, by Theorem II-1, the production function must be homogeneous of degree $\frac{2}{3}$.

A generalization of this normalized profit function exists with the exponent of q_i equal to $-\mu$, $\mu > 0$. In this latter case, the production function must be homogeneous of degree $2\mu/(1+2\mu)$.

4.9. *Transcendental Logarithmic Normalized Production Function*

For the sake of completeness, one should also mention the transcendental logarithmic normalized profit function introduced by Christensen, Jorgenson and Lau (1971 and 1973). The normalized profit function is given by

$$\ln \Pi^* = \alpha_0 + \sum_{i=1}^{m} \alpha_i \ln q_i + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_{ij} \ln q_i \ln q_j.$$

The demand functions are given by

$$\frac{q_i X_i}{\Pi^*} = -\left(\alpha_i + \sum_{j=1}^{m} \beta_{ij} \ln q_j\right), \qquad i = 1,\dots,m.$$

Note the remarkably simple estimating form. This function is not a globally valid normalized profit function as it may become non-monotonic or non-convex at some prices. However, it is possible to verify whether the function is monotonic and convex over some convex set of normalized prices. In addition, it has the advantage that it provides a second order approximation to an arbitrary normalized profit function (and hence to an arbitrary concave technology). It can also attain any value of the elasticity of substitution between any pair of inputs.

5. Applications of the Normalized Profit Function

5.1. *Elasticities of Substitution*

As is well-known, many different elasticities of substitution may be defined in the case of a technology which involves more than two inputs,[26] depending on which variables are held constant. A natural definition, however, in the spirit of the Allen–Uzawa definition of the elasticities of substitution for the case of three or more inputs, is the following:

$$\sigma_{ij} = \frac{G G_{ij}}{G_i G_j},$$

which, by the dual transformation, is equivalent to

$$\sigma_{ij} = \frac{-\left(F - \sum_{i=1}^{m} (\partial F/\partial X_i) X_i\right)}{X_i X_j} \frac{F_{ij}^{+}}{|F|},$$

where F_{ij}^{+} is the j,ith cofactor of the matrix F since $[G_{ij}] = -[F_{ij}]^{-1}$, and $|F|$ is the determinant of the Hessian matrix of F.

[26]See, for instance, McFadden (1963), Nerlove (1967) and Uzawa (1962).

In any event, when there are more than two inputs, the elasticities of substitution are not necessarily the most convenient measures of substitutibility.[27] An alternative is provided by own and cross-price elasticities of demand. In what follows, we give characterization theorems for own and cross-price elasticities of demand by solving systems of partial differential equations for the normalized profit function.

Theorem V-1. A production function is Cobb–Douglas if and only if all the own and cross-price elasticities of factor demands are constants.

Proof: Suppose

$$\frac{\partial \ln X_i}{\partial \ln q_j} = k_{ij}, \quad \text{a constant}, \quad \forall i,j.$$

Integrating this system of partial differential equations, we obtain

$$\ln X_i = \sum_l k_{il} \ln q_l + k_{i0}, \qquad i = 1,\ldots,m$$

or

$$X_i = -\frac{\partial G}{\partial q_i} = e^{k_{i0}} \prod_l q_l^{k_{il}}, \qquad i = 1,\ldots,m,$$

which upon integration yields the Cobb–Douglas normalized profit function. The k_{ij} constants of the different equations may be shown to be the same by making use of the fact that

$$\frac{\partial X_i}{\partial q_j} = -\frac{\partial^2 G}{\partial q_j \partial q_i} = \frac{\partial X_j}{\partial q_i}, \qquad i \neq j.$$

The converse is obvious. Q.E.D.

Theorem V-2. A production function is homogeneous of degree k up to an additive constant if and only if the sum of own and cross-price elasticities of demand is constant for any one commodity.

Proof: Homogeneity of the production function up to an additive constant implies homogeneity of the normalized profit function up to an

[27] Besides, they are insufficient as a description of the technology. See Lau (1976b).

additive constant by Theorems II-1 and II-4. Hence, derived demand is also homogeneous. Hence,

$$\sum_{j=1}^{m} \frac{\partial X_i}{\partial q_j} \frac{q_j}{X_i} = \sum_{j=1}^{m} G_{ij} \frac{q_j}{G_i} = k, \quad \forall i.$$

The converse is proved similarly by retracing the steps and using Euler's Theorem. If G_i is homogeneous then G must be a homogeneous function plus a constant. Q.E.D.

Theorem V-3. A production function is of the Leontief type, that is, the normalized profit function has the form

$$\Pi^* = g\left(\sum_{i=1}^{m} \alpha_i q_i\right),$$

if and only if

$$\frac{\partial \ln X_i}{\partial \ln q_k} = \phi_k(\mathbf{q}), \qquad \begin{matrix} i = 1,\ldots,m, \\ k = 1,\ldots,m. \end{matrix}$$

In other words, this means that the elasticities of each of the demand functions with respect to the kth normalized price are identical.

Proof: Necessity is obvious. To prove sufficiency, we first integrate

$$\frac{\partial \ln X_i}{\partial \ln q_k} = \phi_k(\mathbf{q}),$$

to obtain

$$\ln X_i = \int \phi_k(\mathbf{q}) d \ln q_k + \Phi_i^k(\mathbf{q}_{-k}), \qquad \forall i,$$

where $\mathbf{q}_{-k}$ is $\mathbf{q}$ reduced by q_k.

Now

$$\ln X_i - \ln X_k = \Phi_i^k(\mathbf{q}_{-k}) - \Phi_k^k(\mathbf{q}_{-k}),$$

and

$$\ln X_k - \ln X_i = \Phi_k^i(\mathbf{q}_{-i}) - \Phi_i^i(\mathbf{q}_{-i}).$$

Moreover

$$\ln X_i - \ln X_k = \Phi_i^j(\mathbf{q}_{-j}) - \Phi_k^j(\mathbf{q}_{-j}).$$

But the left-hand side is the same expression, aside from a sign change, and we have seen the right-hand side being independent of q_i, q_j and q_k. We conclude that it must be constant. Thus, $\ln (X_i/X_k) = \theta_i/\theta_k$, a constant

or

$$\frac{1}{\theta_i^*}\frac{\partial G}{\partial q_i} = \frac{1}{\theta_k^*}\frac{\partial G}{\partial q_k}, \qquad \forall i,k,$$

with the general solution

$$G(\mathbf{q}) = g\left(\sum_{i=1}^{m} \theta_i^* q_i\right). \qquad \text{Q.E.D.}$$

Corollary 3.1. If in addition $\sum_{k=1}^{m} \phi_k(q)$ is constant, then Π^* is homogeneous up to an additive constant and G has the form $G = [\sum_{i=1}^{m} \alpha_i q_i]^{\mu} + \alpha_0$.

Proof: This follows from this theorem and Theorem V-2. Q.E.D.

Theorem V-4. A normalized profit function has the form

$$\Pi^* = g\left(\sum_{i=1}^{m} g_i(q_i)\right),$$

if and only if

$$\frac{\partial \ln X_i}{\partial \ln q_k} = \phi_k(\mathbf{q}), \qquad i \neq k; \quad i,k = 1,\ldots,m.$$

Proof: Repeating the argument used in the previous theorem, one has

$$\ln X_i - \ln X_j = \Phi_i^k(\mathbf{q}_{-k}) - \Phi_j^k(\mathbf{q}_{-k}), \qquad \forall i,j,k, \quad i,j \neq k.$$

Thus, for fixed i,j, one concludes that

$$\ln X_i - \ln X_j = \psi_{ij}(q_i,q_j),$$

and for fixed l

$$\ln X_i - \ln X_l = \psi_{il}(q_i,q_l),$$
$$\ln X_j - \ln X_l = \psi_{jl}(q_j,q_l).$$

Combining the last two equations, one has

$$\ln X_i - \ln X_j = \psi_{il}(q_i,q_l) - \psi_{jl}(q_j,q_l),$$

but the right-hand side must be independent of q_l for all values of q_i and q_j.

Hence

$$\phi_{il} = \phi_i(q_i) - \phi_l(q_l), \qquad \forall i,l.$$

Thus from

$$\ln X_i/X_j = \phi_i(q_i) - \phi_j(q_j),$$

we obtain

$$\frac{\partial G/\partial q_i}{\partial G/\partial q_j} = \frac{\phi_i(q_i)}{\phi_j(q_j)},$$

which may be integrated as

$$G(\mathbf{q}) = g\left(\sum_{i=1}^{m} g_i(q_i)\right). \qquad \text{Q.E.D.}$$

Corollary 4.1. If in addition, $\sum_{k=1}^{m} (\partial \ln X_i)/(\partial \ln q_k)$ is a constant, then Π^* is homogeneous up to an additive constant, and G has the C.E.S. form

$$G = \left[\sum_{i=1}^{m} \alpha_i q_i^{\rho}\right]^{\mu/\rho} + \alpha_0.$$

Proof: This follows from Theorems V-2 and V-4, and the fact that an additive function is homogeneous if and only if it has the C.E.S. form. Q.E.D.

With these results then, one can examine directly the own and cross-price elasticities of demand, that is, the comparative statics, and obtain an idea of the degree of substitution. The aforementioned results also apply, with appropriate modification, to subsets of the inputs.

5.2. *Technical Change*

Technical change may be represented by a production function

$$Y = F(\mathbf{X},t), \qquad \frac{\partial F}{\partial t} \geqq 0.$$

This gives rise to a normalized profit function

$$\Pi^* = G(\mathbf{q},t), \qquad \frac{\partial G}{\partial t} \geqq 0.$$

By duality, $\partial F/\partial t = \partial G/\partial t$ at the profit maximum. Hence for given $\mathbf{q}$ normalized profit increases with time.

Definition. A production function is *Hicks neutral* if it can be written in the form

$$Y = F(f(\mathbf{X}),t).$$

Definition. A production function is *Harrod neutral* if it can be written in the form

$$Y = F(f(L,t),\mathbf{X}),$$

where L is labor, the primary factor of production.

Definition. A normalized profit function is *indirectly Hicks neutral* if it can be written in the form

$$\Pi^* = G(f(\mathbf{q}),t).$$

Definition. A normalized profit function is *indirectly Harrod neutral* if it can be written in the form

$$\Pi^* = G(f(w,t),\mathbf{q}).$$

The practical implication of Hicksian neutrality is that the ratio of the marginal products of any two inputs is independent of time. The practical implication of indirect Hicksian neutrality is that the ratios of the derived demands of any two inputs is independent of time.

It should be noted that in general direct Hicksian neutrality does not imply indirect Hicksian neutrality or *vice versa*. A technology is both directly and indirectly Hicksian neutral only if either it is homothetic or it is additive in t. This follows immediately from Theorem II-15. Also, under homotheticity, direct Hicksian neutrality implies and is implied by indirect Hicksian neutrality.

The practical implication of Harrod neutrality is that the ratio of the marginal productivity of labor to the rate of technical change measured in terms of output is independent of $\mathbf{X}$. The practical implication of indirect Harrod neutrality is that the ratio of the demand for labor to the rate of technical change measured in terms of normalized profit is independent of **q**. In general, direct Harrod neutrality does not imply indirect Harrod neutrality or *vice versa*. A production function is both directly and indirectly Harrod neutral only under one of the two following conditions:

(1) $f(L,t) = f(A(t)L)$,

or

(2) $Y = F(\mathbf{X}) + f(L,t)$.

That these conditions are sufficient is obvious. That they are necessary may be shown as follows:

Let $Y = F(f(L,t),\mathbf{X})$. Let $\Pi^* = G(\bar{f},\mathbf{q})$ be the normalized restricted profit function corresponding to F with $F(L,t) = \bar{f}$. The normalized profit function with f unrestricted is then given by

$$\Pi^* = \sup_{\bar{f}} \{G(\bar{f},\mathbf{q}) - wh(\bar{f},t)\},$$

where $h(\bar{f},t)$ is the inverse function of $f(L,t)$ for each given t. The necessary condition for a maximum is

$$\frac{\partial G}{\partial \bar{f}}(\bar{f},\mathbf{q}) - w\frac{\partial h}{\partial \bar{f}}(\bar{f},t) = 0.$$

Differentiating Π^* with respect to t and w we obtain

$$\frac{\partial \Pi^*}{\partial w} = -w\frac{\partial h}{\partial t}, \quad \frac{\partial \Pi^*}{\partial w} = -h(\bar{f},t).$$

Hence

$$\frac{\partial \Pi^*/\partial t}{\partial \Pi^*/\partial w} = w\frac{\partial \ln h}{\partial t},$$

$$\frac{\partial}{\partial q}\left(\frac{\partial \Pi^*/\partial t}{\partial \Pi^*/\partial w}\right) = w \cdot \frac{\partial^2 \ln h}{\partial t\, \partial \bar{f}} \cdot \frac{\partial \bar{f}}{\partial \mathbf{q}} = 0.$$

Thus, either $(\partial^2 \ln h)/(\partial t\, \partial \bar{f}) = 0$ which implies that

$$h = h^*(\bar{f})A^*(t) = L,$$

and hence

$$f = h^{*-1}(A(t)L).$$

Or $\partial \bar{f}/\partial \mathbf{q} = 0$, which implies, by differentiating the first-order condition implicitly, that

$$G(\bar{f},\mathbf{q}) = g(\bar{f}) + G(\mathbf{q}),$$

and hence

$$F(f(L,t),\mathbf{X}) = F(\mathbf{X}) + f(L,t).$$

We note that the first condition (1) corresponds precisely to that of labor-augmenting technical change. One possible specialization of the form of technical change is factor- or output-augmentation. Under factor and output augmenting technical change the production function may be written as

$$Y = A(t)F(A_1(t)X_1,\dots,A_m(t)X_m).$$

Thus $A(t)$ represents "output-augmenting" technical change and $A_i(t)$'s represent "factor-augmenting" technical change. If $A_i(t) = A^*(t), \forall i$, and F is a homothetic function, one can write

$$Y = A(t)F(A^*(t)^{-1}H(X_1,\dots,X_m)),$$

which is clearly Hicksian neutral. It reduces to Harrod neutral technical change if and only if $A(t)$ and $A_i(t)$'s are all constants except for the $A_i(t)$ associated with labor, the primary factor.

With "commodity-augmenting" technical change, the normalized profit function is given by Theorem II-3 as

$$\Pi^* = A(t)G\left(\frac{q_1}{A_1(t)}A(t),\dots,\frac{q_m}{A_m(t)}A(t)\right),$$

where $G(\mathbf{q})$ is the normalized profit function corresponding to $F(X)$. Thus, the production function is "commodity-augmenting" if and only if the normalized profit function is "commodity-augmenting". It is also clear that factor-augmentation has the same effect as price-diminution.

A technical change process is "factor-1 augmenting" if the production function may be written in the form

$$Y = F(A_1(t)X_1,X_2,\dots,X_m).$$

A given technical change process is "factor-1 saving" if

$$\begin{aligned}\frac{\partial X_1}{\partial t} &= \frac{\partial}{\partial t} - \frac{\partial G}{\partial q_1}\\ &= -G_{1t} < 0.\end{aligned}$$

Under what conditions is a technical change process simultaneously "factor-1 augmenting" and "factor-1 saving"? We note that under factor-1 augmentation the normalized profit function can be written as

$$\Pi^* = G(q_1/A_1(t), q_2, \ldots, q_m).$$

Thus

$$\begin{aligned}\frac{\partial X_1}{\partial t} &= \frac{\partial}{\partial t} - \frac{\partial G}{\partial q_1} \\ &= G_{11}\frac{q_1}{A_1^2}\frac{\dot{A}_1}{A_1} + G_1\frac{1}{A_1}\frac{\dot{A}_1}{A_1}.\end{aligned}$$

In order for this to be less than zero, we need

$$G_{11}\frac{q_1}{A_1} + G_1 < 0,$$

which implies by a dual transformation that

$$\frac{\partial X_1}{\partial q_1}\cdot q_1 + X_1 < 0,$$

or a derived demand elasticity of X_1 with respect to own price of greater than -1. Thus in general one cannot identify "factor-augmenting" technical change with "factor-saving" technical change. We note that, even with factor augmenting technical change occurring in only one factor, the derived demands of the other inputs may also change over time as

$$\begin{aligned}\frac{\partial}{\partial t}X_j &= \frac{\partial}{\partial t} - \frac{\partial G}{\partial q_j} \qquad j \neq 1, \\ &= G_{1j}q_1\frac{\dot{A}}{A^2}, \\ \frac{\partial Y}{\partial t} &= \frac{\partial}{\partial t}\left(G - \sum_{j=1}^{n} q_j \frac{\partial G}{\partial q_j}\right) \\ &= \left(\sum_{j=1}^{n} G_{1j}q_j \frac{\dot{A}}{A^2}\right).\end{aligned}$$

We note further that if G_1 were homogeneous of degree $-k$ (necessarily so because G_1 must be negative) then $\sum_{j=1}^{n} G_{1j}q_j = -kG_1$, or if G were additive, supply must be increasing. In general $\partial Y/\partial t$ is indeterminate in sign.

5.3. *Relative Efficiency*[28]

There are two dimensions to the problem of efficiency: technical efficiency and price efficiency. A firm is technically more efficient than another firm, if, and only if, it consistently produces a higher output given identical inputs for both firms. A firm is price-efficient, if, and only if, the value of the marginal product of each input is equated to its price. Any departure from this equality implies price inefficiency. It is sometimes desirable to compare the relative degree of technical efficiency and also the relative degree of price-inefficiency across two firms. If a firm is price-efficient, its profit is at a maximum for a given level of technical efficiency. Thus, a natural measure of relative price efficiency is the relative level of actual profits. A firm is considered to be more price-efficient, if, given the same prices of inputs and outputs and the same degree of technical efficiency, it is more profitable than another firm. Based on this definition, the technically more efficient firm which is also price-efficient will always be more profitable than another firm which is only price-efficient. It is important to note that relative technical efficiency need not imply relative price efficiency and *vice versa*.

Straightforward tests of relative technical and price efficiency between two firms (or groups of firms) may be devised on the basis of the normalized profit function. It is clear that given comparable endowments, identical technology, and normalized input prices, the actual normalized profits of the two firms should be identical if they both have maximized profits. To the extent that one is more price efficient, or technically more efficient, than the other, the normalized profits will differ even for the same normalized input prices and endowments of fixed inputs. The actual normalized profit functions will hence be different for the two firms.

Let us represent the situation as follows: For each firm, the marginal conditions are given by

$$\frac{\partial A_1 F(\mathbf{X}_1,\mathbf{Z}_1)}{\partial \mathbf{X}_1} = \mathbf{K}_1\mathbf{q}_1, \qquad \frac{\partial A_2 F(\mathbf{X}_2,\mathbf{Z}_2)}{\partial \mathbf{X}_2} = \mathbf{K}_2\mathbf{q}_2, \tag{V-1}$$
$$\mathbf{k}_1 = \mathrm{diag}[\mathbf{K}_1] \geqq 0, \qquad \mathbf{k}_2 = \mathrm{diag}[\mathbf{K}_2] \geqq 0,$$

where $\mathbf{K}_i$ is a diagonal matrix, $\mathbf{X}_i$, $\mathbf{Z}_i$, $\mathbf{q}_i$ and $\mathbf{k}_i$ are vectors, A_i is a scalar, and the subscript refers to the firm. If both firms are equally efficient in

[28]This section draws heavily on my joint work with P.A. Yotopoulos. See Lau and Yotopoulos (1971), Yotopoulos and Lau (1973), and Yotopoulos, Lau and Lin (1976).

optimizing with respect to all variable inputs, then $\mathbf{k}_1 = \mathbf{k}_2$. If both firms are equally efficient technically, then $A_1 = A_2$. Equation (V-1) reduced to the usual first-order conditions for profit maximization if and only if $\mathbf{k}_1 = \mathbf{k}_2 = [\mathbf{1}]$, a unit vector. Otherwise, they must be interpreted as decision rules for the individual firms. $\mathbf{k}_1$ and $\mathbf{k}_2$ may assume any non-negative values, and in particular, the special values of $[\mathbf{0}]$ and $[\mathbf{1}]$.

That the decision rules for the firm consist of equating the marginal product to a constant times the normalized price of each input may be rationalized as follows: (1) consistent over and under-valuation of the opportunity costs of the resources by the firms; (2) satisficing behavior; (3) divergence of expected and actual normalized prices; (4) divergence of the subjective probability distribution of the normalized prices from the objective distribution of normalized prices; (5) the elements of $\mathbf{k}_i$ may be interpreted as the first-order coefficients of a Taylor's series expansion of arbitrary decision rules of the type

$$\frac{\partial F_i}{\partial X_{ij}} = f_{ij}(q_{ij}),$$

where $f_{ij}(0) = 0$. A wide class of decision rules may be encompassed under (5).

Let $G(\mathbf{q},\mathbf{Z})$ be the normalized profit function corresponding to $F(\mathbf{X},\mathbf{Z})$. The firms then may be regarded to behave as if they maximize normalized profit subject to price vectors $\mathbf{K}_1\mathbf{q}_1/A_1$ and $\mathbf{K}_2\mathbf{q}_2/A_2$, respectively. Their behavior thus may be represented by the "behavioral" normalized profit functions

$$\Pi_1^b = A_1 G(k_{11}q_1/A_1,\ldots,k_{1m}q_m/A_1;Z_{11},\ldots,Z_{1n}),$$

and

$$\Pi_2^b = A_2 G(k_{21}q_1/A_2,\ldots,k_{2m}q_m/A_2;Z_{21},\ldots,Z_{2n}).$$

A test of equal relative efficiency implies a test of the hypothesis $A_1 = A_2$ *and* $\mathbf{k}_1 = \mathbf{k}_2$. The derived demand functions are given by

$$X_{ij} = -A_i \frac{\partial G(\mathbf{K}_i\mathbf{q}/A_i;\mathbf{Z}_i)}{\partial k_{ij}q_j}, \qquad i = 1,2, \quad j = 1,\ldots,m,$$

and the supply functions by

$$Y_i = A_i\left\{G(\mathbf{K}_i\mathbf{q}/A_i;\mathbf{Z}_i) - \sum_{j=1}^{m} k_{ij}q_j\left[\frac{\partial G(\mathbf{K}_i\mathbf{q}/A_i;\mathbf{Z}_i)}{\partial k_{ij}q_j}\right]\right\}.$$

The actual normalized profit functions are given by

$$\Pi_i^a = Y_i - \sum_{j=1}^{m} q_j X_{ij}$$
$$= A_i \left\{ G(\mathbf{K}_i \mathbf{q}/A_i;\mathbf{Z}_i) + \sum_{j=1}^{m} (1-k_{ij})q_j \left[\frac{\partial G(\mathbf{K}_i \mathbf{q}/A_i;\mathbf{Z}_i)}{\partial k_{ij} q_j} \right] \right\}, \quad i = 1,2.$$

Observe (1) $\partial \Pi^a / \partial A \gtreqless 0$; (2) when $\mathbf{k}_i = [\mathbf{1}]$, the actual and "behavioral" normalized profit functions coincide; and (3) if and only if $A_1 = A_2$ and $k_1 = k_2$, the actual as well as the behavioral normalized profit functions and supply and demand functions of the two firms coincide with each other. This last result is the basis of the null hypothesis for no difference in relative efficiency. When appropriate functional forms are specified for G, the joint hypothesis that $A_1 = A_2$ *and* $\mathbf{k}_1 = \mathbf{k}_2$ may be tested by comparing the coefficient estimates from either the actual profit function or the supply and demand functions, or both.

An additional test becomes relevant if we reject the joint hypothesis that $(A_1,\mathbf{k}_1) = (A_2,\mathbf{k}_2)$. In this case an overall indication of the relative efficiency between the two firms within a specified range of normalized prices for variable inputs may be obtained by comparing the actual values of the normalized profit functions within this range. If

$$\Pi_1^a \geqq \Pi_2^a,$$

for all normalized prices within a specified range, then clearly, the first firm is relatively more efficient within the price range. If some knowledge on the probability distribution of the future prices is available, a choice may be made as to the relative efficiency of the two firms.

One can also test the hypothesis that the fixed inputs command equal rent on the two firms by computing the first derivatives of the actual normalized profit functions with respect to the fixed inputs and testing for their equality. This may have important implications on the optimal form of organization.

Finally the above analysis can be easily extended to three or more firms (or groups of firms). We conclude this subsection with an example.

Example

The normalized restricted profit function corresponding to a Cobb–Douglas production function with m variable inputs and n fixed inputs is, from Section 4.2,

$$\Pi^* = (1-\mu)\left[\prod_{i=1}^{m}\left(\frac{q_i}{\alpha_i}\right)^{-\alpha_i(1-\mu)^{-1}}\right]\left[\prod_{j=1}^{n} Z_j^{\beta_j(1-\mu)^{-1}}\right],$$

$$\mu = \sum_{i=1}^{m} \alpha_i < 1.$$

By direct computation, the actual normalized profit functions and the demand functions are

$$\Pi_i^a = A_i^{(1-\mu)^{-1}}\left(1-\sum_{j=1}^{m}\frac{\alpha_j}{k_{ij}}\right)\left[\prod_{j=1}^{m} k_{ij}^{-\alpha_j(1-\mu)^{-1}}\right]\left[\prod_{j=1}^{m} \alpha_j^{\alpha_j(1-\mu)^{-1}}\right]$$
$$\times\left[\prod_{j=1}^{m} q_j^{-\alpha_j(1-\mu)^{-1}}\right]\left[\prod_{j=1}^{n} Z_j^{\beta_j(1-\mu)^{-1}}\right], \qquad i = 1,2,$$

$$X_{ij} = A_i^{(1-\mu)^{-1}}\left(\frac{\alpha_j}{k_{ij}q_j}\right)\left[\prod_{j=1}^{m} k_{ij}^{-\alpha_j(1-\mu)^{-1}}\right]\left[\prod_{j=1}^{m} \alpha_j^{\alpha_j(1-\mu)^{-1}}\right]$$
$$\times\left[\prod_{j=1}^{m} q_j^{-\alpha_j(1-\mu)^{-1}}\right]\left[\prod_{j=1}^{n} Z_j^{\beta_j(1-\mu)^{-1}}\right], \qquad i = 1,2, \quad j = 1,\ldots,m.$$

From these two equations, one may derive

$$\frac{q_j X_{ij}}{\Pi_i^a} = \frac{\alpha_j/k_{ij}}{\left(1-\sum_{j=1}^{m}\frac{\alpha_j}{k_{ij}}\right)}, \qquad i = 1,2, \quad j = 1,\ldots,m.$$

These actual profit share equations may be combined with the natural logarithm of the actual normalized profit functions to obtain estimates of A_i, $\mathbf{k}_i$ and the technological parameters.

5.4. *Monopolistic Profit Functions*

A monopolist faces a downward sloping demand curve. Let the inverse demand function be given as

$$p = D(Y),$$

where

$$D'(Y) < 0.$$

Then the profit maximization problem becomes

$$\max P = D(F(\mathbf{X},\mathbf{Z}))F(\mathbf{X},\mathbf{Z}) - \mathbf{q}^{*\prime}\mathbf{X},$$

where q^* is the vector of nominal prices of the variable inputs.

Let $S(\mathbf{X},\mathbf{Z}) \equiv D(F(\mathbf{X},\mathbf{Z}))F(\mathbf{X},\mathbf{Z})$ be the revenue function.[29] If it is assumed that $S(\mathbf{X},\mathbf{Z})$ satisfies Assumptions (F.1) through (F.7), then the profit maximization problem is isomorphic to the normalized profit maximization problem in Section 1. Thus, the profit function $G(\mathbf{q}^*,\mathbf{Z})$, is given by

$$G(\mathbf{q}^*,\mathbf{Z}) = \max_{\mathbf{X}}\{S(\mathbf{X},\mathbf{Z}) - (\mathbf{q}^{*\prime}\mathbf{X})\},$$

which satisfies Assumptions (G.1) through (G.7). Moreover, there is a one-to-one correspondence between $S(\mathbf{X},\mathbf{Z})$ and $G(\mathbf{q}^*,\mathbf{Z})$.

All the dual relationships which hold between F and G hold between S and G. As before, the demand functions for the variable inputs are given by

$$\mathbf{X}^* = -\frac{\partial G}{\partial \mathbf{q}^*}(\mathbf{q}^*,\mathbf{Z}),$$

and the optimal revenue function is given by

$$S^* = G(\mathbf{q}^*,\mathbf{Z}) - \mathbf{q}^{*\prime}\frac{\partial G}{\partial \mathbf{q}^*}(\mathbf{q}^*,\mathbf{Z}).$$

We note that G depends on nominal prices of the inputs only and hence X^* and S^* depend on only the nominal prices of the inputs.

For the purpose of econometric applications, one may just as well start with a function $G(\mathbf{q}^*,\mathbf{Z})$ which satisfies Assumptions (G.1) through (G.7) without worrying about the properties of $S(\mathbf{X},\mathbf{Z})$ since as McFadden has emphasized in Chapter I.1, one cannot in fact observe those input vectors for which $S(\mathbf{X},\mathbf{Z})$ fails to satisfy the Assumptions (F.1) through (F.7).

Example

Let

$$p = Y^{-\epsilon}, \qquad 1 > \epsilon > 0,$$

$$Y = \prod_{i=1}^{m} X_i^{\alpha_i}, \qquad \sum_{i=1}^{m} \alpha_i = \mu < 1.$$

Then

[29]Note that the revenue function as used here is different from the revenue function $R(p,\mathbf{Z})$ which gives the *maximized* value of revenue for given p and $\mathbf{Z}$.

$$S = \prod_{i=1}^{m} X_i^{\alpha_i(1-\epsilon)}.$$

The monopolistic profit function is hence

$$\Pi = (1-\mu^*) \prod_{i=1}^{m} \left(\frac{q_i^*}{\alpha_i^*}\right)^{-\alpha_i^*(1-\mu^*)^{-1}},$$

where

$$\alpha_i^* \equiv \alpha_i(1-\epsilon), \qquad i = 1,\ldots,m,$$

$$\mu^* \equiv \sum_{i=1}^{m} \alpha_i^* \equiv (1-\epsilon)\mu.$$

The derived demand functions are given by

$$X_i = \alpha_i^* q_i^{*-1} \prod_{j=1}^{m} \left(\frac{q_j^*}{\alpha_j^*}\right)^{-\alpha_j^*(1-\mu^*)^{-1}}, \qquad i = 1,\ldots,m.$$

Revenue is given by

$$R = S = \prod_{i=1}^{m} \left(\frac{q_i^*}{\alpha_i^*}\right)^{-\alpha_i^*(1-\mu^*)^{-1}}.$$

Finally, we note that while given the profit function alone one can construct an $S(\mathbf{X},\mathbf{Z})$ through the conjugacy operation, one cannot identify $F(\mathbf{X},\mathbf{Z})$ without additional information. We should also emphasize that the assumption of concavity of $S(\mathbf{X},\mathbf{Z})$ neither implies nor is implied by the concavity of $F(\mathbf{X},\mathbf{Z})$. In fact, if there is indeed a monopoly, it is likely that $F(\mathbf{X},\mathbf{Z})$ is non-concave in $\mathbf{X}$.

5.5. *Dynamic Behavior*

Dynamic models have been introduced into econometric research via two principal hypotheses – the "adaptive expectations" hypothesis and the "lagged adjustment" hypothesis. These hypotheses can be readily incorporated into the normalized profit function approach.

5.5.1. *Adaptive Expectations Hypothesis*

The firm is assumed to maximize profit for given expected normalized prices. Then, for a given technology, there is a normalized profit function of expected normalized prices which are in turn functions of current and past normalized prices. Let the price expectation formation process be

$$q_i^0(t) = \omega_i(L)q_i(t), \qquad i = 1,\dots,m,$$

where $q_i^0(t)$ is the expected normalized price of the ith input, and $\omega_i(L)$ the rational distributed lag operator for the ith price.[30]

The expected normalized profit function is then given by

$$\Pi^{*0} = G(q_1^0,\dots,q_m^0).$$

Supply and demand as functions of expected normalized prices are given by

$$Y^0 = G(q_1^0,\dots,q_m^0) - \sum_{i=1}^{m} \frac{\partial G}{\partial q_i^0} q_i^0,$$

$$X_i^0 = -\frac{\partial G}{\partial q_i^0}, \qquad i = 1,\dots,m.$$

In general, both Y^0 and X_i^0's are functions of both current and past prices, with the time structure of the effect of different input prices given by the coefficients of $\omega_i(L)$. Actual normalized profit, on the other hand, is given by

$$P^* = Y^0 - \sum_{i=1}^{m} \frac{\partial G}{\partial q_i^0} q_i.$$

5.5.2. *Lagged Adjustment Hypothesis*

Lagged adjustment models are in general based on an adjustment equation,

$$X_t - X_{t-1} = \omega(L)[X_t^* - X_{t-1}], \tag{V-2}$$

where X_t^* is the desired quantity in period t and the subscripts i are suppressed. Equation (V-2) may be rewritten as

$$(1-L)X_t = \omega(L)X_t^* - \omega(L)LX_t,$$

$$(1 - L + \omega(L)L)X_t = \omega(L)X_t^*,$$

$$X_t = \frac{\omega(L)}{(1 - L + \omega(L)L)} X_t^*$$

$$= \mu(L)X_t^*,$$

[30]For an exposition of rational distributed lag functions and rational distributed lag operators, see Jorgenson (1966a).

where

$$\mu(L) \equiv \frac{\omega(L)}{(1 - L + \omega(L)L)}.$$

Let G be the normalized profit function for the technology in period t; then

$$Y^* = G - \sum_{i=1}^{m} \frac{\partial G}{\partial q_i} q_i,$$

$$X_i^* = -\frac{\partial G}{\partial q_i}, \qquad i = 1,\ldots,m.$$

However, the actual supply and demand equations are given by

$$X_i = -\mu_i(L)\frac{\partial G}{\partial q_i}, \qquad i = 1,\ldots,m,$$

and

$$Y = F\left(\mu_1(L)\frac{\partial G}{\partial q_1},\ldots,\mu_m(L)\frac{\partial G}{\partial q_m}\right).$$

Both the "adaptive expectations" and the "lagged adjustment" models represent attempts to introduce dynamic elements into a basically static concept and are not completely satisfactory. There is, in principle, no reason why truly dynamic "profit" functions cannot be constructed. These will be functions which give the maximized value of the net worth (or equivalently the present value) of the firm for specified current and future expected prices and initial endowments.

The net worth function, or functional, may be written as

$$NW = G(\mathbf{p},\mathbf{q},t),$$

where $\mathbf{p}$ and $\mathbf{q}$ are possibly infinite dimensional vectors. The profit in period t_i is given by

$$\Pi^* = G(\mathbf{p},\mathbf{q},t_i) - G(\mathbf{p},\mathbf{q},t_i - 1).$$

The supply and demand functions in period t_i may be obtained by the usual duality relationships.

Before such a dynamic "profit" function can be constructed, however, one must have a well-developed theory of intertemporal production. The dual to the dynamic profit function is the production function that links output and input possibilities of all periods, with due recognition given to the fact that future inputs cannot contribute to present output.

Given a dynamic "profit" function, the complete optimal production and investment plan for the future may be calculated based on the expectations of future price movements. These dynamic profit functions must satisfy certain structural characteristics, e.g., at each point t in time different from the point of planning, the profit function must be expressible as a function of endowments at time t and the price at time t and in the future. The supply and demand functions at time t will be expressible as functions independent of the past prices. One may also want to impose the requirement of stationarity, a concept introduced by Koopmans et al. (1964), with regard to dynamic profit functions.

In Chapter II.4 of this volume Fuss and McFadden also analyze the problem of intertemporal production using duality concepts.

5.6. *Profit Functions and Uncertainty*

Using the normalized profit function, one can obtain an immediate proof of a well-known result that randomness in prices results in higher expected profits if the firm is able to adjust instantaneously than if the prices are constant and equal to their expected values.[31] Expected normalized profits are given by

$$E[G(\mathbf{q})].$$

Normalized profits at expected normalized prices are given by

$$G(E[\mathbf{q}]).$$

By Jensen's (1906) inequality on convex functions, one obtains immediately that

$$E[G(\mathbf{q})] \geqq G(E[\mathbf{q}]).$$

Note that this result holds true for fluctuations in all prices and not only in the output prices as the problem is customarily posed.

The effect of randomness on expected output, on the other hand, is not clear cut,

$$\begin{aligned} Y &= G - \sum_{i=1}^{m} \frac{\partial G}{\partial q_i} q_i, \\ E[Y] &= E[G(\mathbf{q})] - \sum_{i=1}^{m} E\left[\frac{\partial G}{\partial q_i} q_i\right] \\ &\gtrless G(E[\mathbf{q}]) - \sum_{i=1}^{m} E[q_i] \frac{\partial G}{\partial q_i}(E[\mathbf{q}]) \\ &= Y(E[\mathbf{q}]). \end{aligned}$$

[31] See, for example, Oi (1961).

However, if it is assumed that the production function is homogeneous of degree k, then by Theorem II-1,

$$Y = (1-k)^{-1}G.$$

Hence

$$\begin{aligned} E[Y] &= (1-k)^{-1}E[G(\mathbf{q})] \\ &\geqq (1-k)^{-1}G(E[\mathbf{q}]) \\ &\geqq Y(E[\mathbf{q}]). \end{aligned}$$

Expected output is also increased by randomness in both output and input prices.

6. Summary and Conclusions

In the preceding sections, the potential usefulness of the concept of the normalized profit function in both theoretical and empirical applications has been demonstrated. In particular, the normalized profit function provides a convenient and logical link, by virtue of its duality properties, between theoretical specification of a model and empirical implementation. By deriving a system of supply and demand functions from a normalized profit function, rather than attempting to solve the profit maximization problem itself, one avoids the potential difficulties (sometimes impossibility) of obtaining closed form solutions. Nevertheless, one is assured that the supply and demand functions thus derived do correspond to those that are obtained through the maximization of profits subject to some production function with the usual regularity properties. Many additional factors, such as imperfection of markets and technical change, may also be conveniently introduced in a straightforward way. Alternatively, given an arbitrary system of supply and demand functions, one can verify their consistency with profit maximization subject to a production function constraint by checking whether the system is integrable into a normalized profit function.

In addition, it should be emphasized that the normalized profit function contains *all* the empirically relevant information. Supply and demand functions derived from a normalized profit function satisfy all the *a priori* restrictions imposed by the production function. Hence there is no loss in generality, but a gain in elegance and analytical convenience, if one starts out with a normalized profit function.

Finally, through the examples provided, it may be seen that a large number of complete systems that (1) approximate any arbitrary normalized profit (and hence production) function, (2) can attain any value of elasticity of substitution between any pairs of commodities, and (3) are econometrically convenient to estimate – meaning in most cases linear in parameters – are available. They offer greater flexibility than the supply and demand systems traditionally used in the literature. This greater flexibility may result in more realistic modeling of the economy or the firm by making indispensable restrictive assumptions introduced for the sake of obtaining closed form solutions.

The potentials for profit function (and revenue and cost functions) are by no means exhausted here. Directions for future research include (1) dynamic models, (2) incorporation of adjustment costs, (3) non-classical technologies, (4) profit maximization under uncertainty, and (5) departures from profit-maximizing behavior.

Part II

Functional Forms in Production Theory

Chapter II.1

A SURVEY OF FUNCTIONAL FORMS IN THE ECONOMIC ANALYSIS OF PRODUCTION

MELVYN FUSS, DANIEL McFADDEN, and YAIR MUNDLAK*

University of Toronto, University of California, Berkeley, and The Hebrew University of Jerusalem

1. The Context and Objectives of Production Analysis

1.1. *Introduction*

Empirical analysis of technology is carried out in many contexts, for many purposes. Each situation raises specific conditions and objectives which must be met in the specification of an econometric production model. This chapter surveys a variety of functional forms for production processes, and their cost and profit duals, and discusses some of the applications for which they are suited.

The diversity and extent of the subject of applied production theory makes a comprehensive survey impossible. We emphasize the structure of alternative functional forms, and the relationship between "exact" models of technology and econometric models incorporating stochastic specifications. However, we have not attempted to provide either a full catalog of properties of functional forms or a general procedure for introducing stochastic elements in production models. We focus on several basic issues of technology – scale, separability, and substitutability. We have not attempted to treat a number of other major issues,

*The financial support through NSF grant no. SOC-73-05374-A01 is gratefully acknowledged by Y. Mundlak.

such as technical change and aggregation, which are equally important in many applications.

The uses of production models can be classified in two ways. The first is the distinction between analytic studies of the production process (for example, a test of the constancy of returns to scale), versus estimation to provide predictions for specific applications (for example, a prediction of industrial demand for energy). The former alternative requires close attention to the structure and parameterization of the production model, while the latter is more concerned with the robustness of the model and its extrapolative plausibility.

The second division is between macroeconomic analysis of aggregative production relationships and microeconomic treatment of industry, firm, and establishment technologies. Issues of aggregation over commodities, economic units, and technologies, and questions of proper parameterization of distribution, technical change, and growth effects have dominated the literature on aggregate production functions. Questions of compatibility with physical production processes and firm behavior have been important in the analysis of microeconomic production relations. Statistical issues in the estimation of technological relationships have concentrated on the stochastic nature of aggregate quantity and price indices, as determined by their definition and measurement, and on the stochastic specification of microeconomic firm behavior.

In a survey of functional forms, it is important to keep in mind the fact that these forms have been constructed for a variety of applications. One cannot expect to find a single "best" parametric production function for all purposes; to the contrary, many of these functional forms are well-suited for specific applications but poorly-suited for use as general purpose characterizations of technology.

1.2. *Objectives of Production Analysis*

Historically, emphasis has been placed on a number of different aspects of technology, depending on the objectives of analysis. We list below some of the major objectives of production studies which have motivated the development of functional forms:

(1) *Distribution* (the income shares of factors of production): Most attention has centered on the aggregate shares of capital and labor. Distribution issues also arise at the microeconomic level in problems

such as the incidence of tax and subsidy programs. Distribution parameters are of great importance in evaluating the growth process.

(2) *Scale* (the existence of constant returns to scale, or the presence of decreasing or increasing returns): Scale has aggregate implications for long-run growth, and for the structure of industry, which is also related to the question of the logical consistency of the neoclassical assumption of profit maximization. Microeconomic issues which focus on the supply and financing of public services often center on the technological question of the existence of increasing returns to scale.

(3) *Substitution* (the degree of substitutability of factors of production): Substitutability is a critical issue in the behavior of distributive shares when factors proportions change. It plays an important role in determining the incidence of taxes; and also the behavior of relative factor prices, and therefore product prices, in the process of growth.

(4) *Separability* (decomposition of production relationships into nested or additive components): Separability is an extremely important structural property in a production model which often permits econometric analysis to be carried out in terms of subsets of the total set of possible variables, in stages, or with consistent aggregates of variables. Separability is of direct economic interest, implying uniform or invariant behavior of certain economic quantities, and allowing decentralization in decision-making. It is also of critical interest in the specification of functional forms, influencing generality and simplicity, and becomes an important subject for empirical tests. (Because of its pivotal role in functional form specification, separability is discussed in detail in Section 6 of this survey.)

(5) *Technical change* (modification of the technological structure over time): Of interest are disembodied technical change (innovations which require no specific capital); technical change embodied in factors of production (usually capital, but potentially other factors such as skilled labor); factor-augmenting change which increases the effective quality of inputs; augmentation of other technological characteristics such as scale-augmenting change (increasing the scale level at which decreasing returns set in) or substitution-augmenting change (increasing the substitutability of inputs); and endogenous technical change (learning-by-doing; innovation and induced technical change).

In addition, a number of auxiliary topics have been the subject of econometric investigation, with attendant problems of functional specification.

1. *Technological flexibility* (the robustness of the technology in adapt-

ing to changing environments): Of interest is the degree to which flexibility is incorporated in the adopted technology, and its tradeoff against static efficiency.

2. *Efficiency* (operation on or inside the technology frontier): Relative efficiency of different economic units (firms, industries, nations) is of interest, as is the efficiency of the same unit in alternative economic environments.

3. *Homotheticity* (the presence of expansion paths with scale which are rays through the origin): Homothetic production functions will display unchanging distributive shares with changes in scale, ceteris paribus. In contrast, heterotheticity will yield changing factor intensities with changes in scale.

4. *Consistent aggregation* (the problem of specifying technological structures that are invariant with respect to aggregation over commodities or economic units): This problem is most critical in studies which want to ensure microeconomic compatibility of aggregate analysis, or want to obtain simple aggregate forecasts from microeconomic estimates.

In surveying various forms, one should keep in mind the alternative objectives listed above.

2. Criteria for the Design of Functional Forms

2.1. *Maintained Hypotheses*

In addition to the obvious criterion that a functional form should relate to the objectives of an analysis, there are a few general principles which should be adopted in modelling production. The first concerns the role of maintained hypotheses.

Any study in production economics (and, for that matter, in econometrics in general) takes place against the background of a series of maintained hypotheses which are not themselves tested as part of the analysis, but are assumed true. The most fundamental of these maintained hypotheses are basic axioms on the nature of technology (e.g., "the production possibility set is closed"), which are widely accepted because they are believed to be true, or at least irrefutable with existing data. Second come technological and behavioral assumptions which are not widely held to be universal truths, but may be widely accepted as plausible for the problem at hand (e.g., "convex technology" or "cost minimizing behavior"). Next come assumptions made to facilitate the analysis (e.g., "independent normal errors", "intermediate inputs separable from primary inputs"), which are believed to be harmless approximations to reality. Finally, there may be maintained hypotheses, such as the assumption of a specific parametric functional form, or of the constancy of some unobserved prices or quantities, which are accepted only for convenience or tractability. The analyst may then argue that his results are robust or insensitive with respect to these hypotheses, justifying their imposition on grounds of usefulness and lack of negative consequence rather than on grounds of plausibility.

The outcome of a specific test of hypothesis depends in general on *both* the validity of the hypothesis under examination and the validity of the maintained hypotheses. Consequently, a test performed in the presence of an implausible maintained hypothesis may not be convincing; the result may be a consequence of the validity of the maintained hypothesis rather than of the primary hypothesis in which one is interested. This suggests the general principle that *one should not attempt to test a hypothesis in the presence of maintained hypotheses that have less commonly accepted validity.* For example, it would be inappropriate to test a basic assumption such as convexity of the technology by examining the sign of the estimated elasticity of substitution when a C.E.S. production function is imposed as a maintained hypothesis, since a rejection is more likely to be interpreted as a failure of the C.E.S. specification than of convexity. An implication of this principle is the need for general, flexible functional forms, embodying few maintained hypotheses, to be used in tests of the fundamental hypotheses of production theory. Given the qualitative, non-parametric nature of the fundamental axioms, this suggests further that the more relevant tests will be non-parametric, rather than based on parametric functional forms, even very general ones. While non-parametric approaches to the study of production relationships have received some attention in economics [Farrell (1957), Hanoch and Rothschild (1972)], these methods have been exploited less systematically for tests of basic hypotheses than have parametric forms [e.g. Berndt-Christensen (1973a)]. Analyses of the latter type inevitably are subject to the criticism that a rejection of a hypothesis may be a result of the parametric specification rather than falseness of the hypothesis. This criticism must be balanced, however, against the observation that non-parametric tests have not yet been developed for some multivariate production hypotheses.

For most analyses, the econometrician has a choice of several starting points for the specification of functional forms. This book emphasizes the equivalence of production, cost, and profit functions as characterizations of technology under appropriate conditions (including competitive markets). It is also possible to specify a production model directly in terms of demand and supply functions, expressed either in prices or quantities, or even in terms of differential or difference equations for these demand and supply functions. Under appropriate integrability conditions, these systems can then be solved to obtain the implied production, cost, or profit functions. This survey will emphasize functional forms for production, cost, and profit functions, but will not attempt to survey specifications of technology which are formulated directly in terms of demand and supply functions or their derivatives.

2.2. *Criteria for Choosing Functional Forms*

Within the framework of the maintained hypotheses imposed on a particular problem or class of problems, a wide variety of compatible functional forms will usually be available. We list some of the criteria which may be used to select among them:

1. *Parsimony in parameters*: The functional form should contain no more parameters than are necessary for consistency with the maintained hypotheses. Excess parameters exacerbate problems of multicollinearity, which tend to be severe in any case in many applications due to market substitution which causes prices, and hence quantities, to be highly correlated. Further, when the sample is small, excess parameters mean a loss of degrees of freedom, a particular problem in aggregate analysis.

2. *Ease of interpretation*: Excessively complex or parameter-rich functional forms may contain implausible implications which are hidden from easy detection. Further, complex transformations may make it cumbersome to compute and assess economic effects of interest; for example, elasticities of substitution. Thus, ceteris paribus, it is better to choose functional forms in which the parameters have an intrinsic and intuitive economic interpretation, and in which functional structure is clear.

3. *Computational ease*: Historically, systematic multivariate empirical analysis has been confined to linear (in parameters) statistical models for computational reasons. While current computational technology makes

direct estimation of non-linear forms feasible, it remains the case that linear-in-parameters systems have a computation cost advantage, and have, in addition, the advantage of a more fully-developed statistical theory. The tradeoff between the computational requirements of a functional form and the thoroughness of empirical analysis should be weighed carefully in the choice of a model.

4. *Interpolative robustness*: Within the range of observed data, the chosen functional form should be well-behaved, displaying consistency with maintained hypotheses such as positive marginal products or convexity. If these properties must be checked numerically, then the form should admit convenient computational procedures for this purpose.

5. *Extrapolative robustness*: The functional form should be compatible with maintained hypotheses *outside* the range of observed data. This is a particularly important criterion for forecasting applications.

3. Dual Transformation, Cost, and Profit Functions – Maintained Hypotheses on the Technology and Its Representations[1]

In this section, we summarize the commonly imposed maintained hypotheses for production, cost, and profit functions. Much of the development of specific functional forms has concentrated on questions of consistency with these hypotheses. More detailed discussions of the relationships among these properties are given in Part I of this volume.

3.1. *Production Possibility and Input Requirement Sets*

The basic notion to be introduced is that of a technology. Let $\mathbf{v}$,$\mathbf{y}$ be vectors of inputs and outputs, respectively. The *production possibility set* $\mathbf{Y}$ is the set of all feasible input–output combinations, i.e., $\mathbf{Y} = \{\mathbf{v},\mathbf{y}:\mathbf{v}$ can yield $\mathbf{y}\}$. For each $\mathbf{y}$ occurring in some input–output vector in $\mathbf{Y}$ we can define the *input requirement set* $\mathbf{V}(\mathbf{y})$, containing all the input bundles which can produce $\mathbf{y}$, i.e., $\mathbf{V}(\mathbf{y}) = \{\mathbf{v}:(\mathbf{v},\mathbf{y}) \in \mathbf{Y}\}$. It is convenient to describe the maintained hypotheses on the technology in terms of the properties of $\mathbf{V}(\mathbf{y})$.

[1]This section is intended as a summary in order to make the chapter self-contained. A more detailed description of the characteristics of the representations of technology can be found in Chapter I.1 of this volume.

The properties of $\mathbf{V}(\mathbf{y})$ are assumed to be:

1.1 *Location.* $\mathbf{V}(\mathbf{y})$ is a non-empty subset of the non-negative orthant $\mathbf{R}^n$, denoted by $\mathbf{\Omega}_n$. It is possible that some factors will not be utilized. However, the only output that can be obtained with no inputs at all is the zero output. It is therefore required that $\mathbf{V}(\mathbf{0}) = \mathbf{\Omega}_n$ and $\mathbf{y} > \mathbf{0}$ imply $\mathbf{0} \notin \mathbf{V}(\mathbf{y})$.

1.2 *Closure.* The analysis is greatly simplified when $\mathbf{V}(\mathbf{y})$ is assumed to be closed. That is, if a sequence of points $\{\mathbf{v}^n\}$ in $\mathbf{V}(\mathbf{y})$ converges, the limiting point also belongs to $\mathbf{V}(\mathbf{y})$. This means that $\mathbf{V}(\mathbf{y})$ contains all its limit points, and assures that the efficiency frontier of $\mathbf{V}(\mathbf{y})$ belongs to $\mathbf{V}(\mathbf{y})$.

1.3 *Monotonicity.* If a given output can be produced by the input-mix $\mathbf{v}$ it can also be produced by a larger input: if $\mathbf{v} \in \mathbf{V}(\mathbf{y})$ and $\mathbf{v}' \geq \mathbf{v}$ then $\mathbf{v}' \in \mathbf{V}(\mathbf{y})$. Similarly, the inputs required to produce a given output can certainly produce a smaller output. If $\mathbf{y} \geq \mathbf{y}'$ then $\mathbf{V}(\mathbf{y}) \subset \mathbf{V}(\mathbf{y}')$. These conditions imply that, unless $\mathbf{Y}$ is bounded and the boundary belongs to $\mathbf{Y}$, there is no input-mix that can produce every $\mathbf{y}$ in $\mathbf{Y}$.

1.4 *Convexity.* $\mathbf{V}(\mathbf{y})$ is convex.

3.2. *Production and Distance Functions*

Suppose we restrict $\mathbf{y}$ to a single element y. Then, using the notion of the input requirement set, the *production function* for y can be defined by

$$f(\mathbf{v}) = \max_{y} \{y : \mathbf{v} \in \mathbf{V}(y)\}.$$

When $\mathbf{V}(y)$ has properties (1.1) through (1.4), $f(\mathbf{v})$ has the following properties (Diewert (1971)):

2.1 *Domain.* $f(\mathbf{v})$ is a real-valued function of $\mathbf{v}$ defined for every $\mathbf{v} \in \mathbf{\Omega}_n$ and it is finite if $\mathbf{v}$ is finite; $f(\mathbf{0}) = 0$.

2.2 *Monotonicity.* An increase in inputs cannot decrease production:

$$\mathbf{v} \geq \mathbf{v}' \Rightarrow f(\mathbf{v}) \geq f(\mathbf{v}').$$

2.3 *Continuity.* f is continuous from above: every sequence $\{\mathbf{v}^n\} \subset \mathbf{\Omega}_n$ such that $f(\mathbf{v}^n) \geq y^0$, $y^0 = f(\mathbf{v}^0)$ and $\mathbf{v}^n \to \mathbf{v}^0$ implies $\lim_{n\to\infty} f(\mathbf{v}^n) = y^0$. Of course, this is a weaker property than continuity, which is almost universally imposed on the production function in empirical work.

2.4 *Concavity.* f is quasi-concave over $\mathbf{\Omega}_n$: the set $\{\mathbf{v}: f(\mathbf{v}) \geq y, \mathbf{v} \in \mathbf{\Omega}_n\}$ is convex for every $y \geq 0$. This property insures diminishing marginal rates of substitution.

In addition, twice differentiability of f is commonly imposed in empirical work.

When $\mathbf{y}$ contains more than one element, efficient production of $\mathbf{y}$ can be described in terms of the *distance*[2] *function*

$$D(\mathbf{y},\mathbf{v}) = \max\left\{\lambda > 0 \,\middle|\, \frac{1}{\lambda}\mathbf{v} \in \mathbf{V}(\mathbf{y})\right\},$$

for $(\mathbf{v},\mathbf{y}) \in \mathbf{Y}$ and $\mathbf{v}$ strictly positive; the frontier satisfies $D(\mathbf{y},\mathbf{v}) = 1$.

Alternatively we can define the transformation function as the maximum amount of y_1 which can be produced given the amounts of the other commodities $\hat{\mathbf{y}} = (y_2, \ldots, y_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$, i.e.,

$$F(\hat{\mathbf{y}},\mathbf{v}) = \max_{y_1}\{y_1 : (y_1, \hat{\mathbf{y}}, \mathbf{v}) \in \mathbf{Y}\}.$$

The transformation function is assumed to have the following properties [Diewert (1974a)]:

2.1.1 *Domain.* F is an extended real-valued function defined and bounded from above for every $(\hat{\mathbf{y}},\mathbf{v}) \in \mathbf{\Omega}_{n+m-1}$. Also,

$$F(\mathbf{0},\mathbf{0}) = 0.$$

2.1.2 *Monotonicity.* F is non-increasing in $\hat{\mathbf{y}}$ and non-decreasing in $\mathbf{v}$.

2.1.3 *Continuity.* F is continuous from above.

2.1.4 *Concavity.* F is a concave function.

[2]For a detailed description of distance functions, see Chapter I.1.

The distance function and transformation function have a simple relationship:

$$D(\mathbf{y},\mathbf{v}) = \max\{\lambda > 0 | y_1 = F(\hat{\mathbf{y}},\mathbf{v}/\lambda)\},$$

and $y_1 = F(\hat{\mathbf{y}},\mathbf{v})$ is the solution to the equation

$$D(y_1,\hat{\mathbf{y}},\mathbf{v}) = 1.$$

Then, $D(F(\hat{\mathbf{y}},\mathbf{v}),\hat{\mathbf{y}},\mathbf{v}) \equiv 1$ is an identity, as is $y_1 \equiv F(\hat{\mathbf{y}},\mathbf{v}/D(\mathbf{y},\mathbf{v}))$. Using the properties of distance functions derived in Chapters I.1 and I.3, the reader can use these identities to deduce the properties of the transformation function.

3.3. *The Cost Function*

In general, economic models involving production need, in addition to the production function or transformation function, rules of behavior. The selection of the optimal input mix for some $\mathbf{y} \in \mathbf{Y}$ and some set of exogenous input prices $\mathbf{r}$ normally assumes cost minimizing behavior. Cost minimization for all $\mathbf{r} \in \mathbf{\Omega}_n^*$, where $\mathbf{\Omega}_n^*$ is the strictly positive orthant, and $\mathbf{y} \in \mathbf{Y}$ is described by the cost function

$$C(\mathbf{y},\mathbf{r}) = \min\{\mathbf{r}\cdot\mathbf{v}:\mathbf{v} \in \mathbf{V}(\mathbf{y})\}.$$

If the input markets are not competitive, a cost function can still be defined by this formula, with the prices $\mathbf{r}$ interpreted as shadow or imputed prices.

If $\mathbf{V}(\mathbf{y})$ possesses Properties (1.1) through (1.4) then $C(\mathbf{y},\mathbf{r})$ has Properties (3.1)–(3.5) listed below:

3.1 *Domain.* $C(\mathbf{y},\mathbf{r})$ is a positive real-valued function defined for all positive prices $\mathbf{r}$ and all positive producible outputs; $C(\mathbf{0},\mathbf{r}) = 0$.

3.2 *Monotonicity.* $C(\mathbf{y},\mathbf{r})$ is a non-decreasing function in output and tends to infinity as output tends to infinity. It is also non-decreasing in prices.

3.3 *Continuity.* $C(\mathbf{y},\mathbf{r})$ is continuous from below in $\mathbf{y}$ and continuous in $\mathbf{r}$.

3.4 *Concavity.* $C(\mathbf{y},\mathbf{r})$ is a concave function in $\mathbf{r}$.

3.5 *Homogeneity.* $C(\mathbf{y},\mathbf{r})$ is linear homogeneous in $\mathbf{r}$.

Empirical work usually assumes in addition:

3.6 *Differentiability.* In most empirical applications, $C(\mathbf{y},\mathbf{r})$ is to be twice differentiable in $\mathbf{r}$.

Under 3.6, the cost function possesses the important derivative property

(a) $\dfrac{\partial C}{\partial r_i} = v_i$ (Shephard's Lemma);

from which it follows that

(b) $\dfrac{\partial^2 C}{\partial r_i \partial r_j} = \dfrac{\partial^2 C}{\partial r_j \partial r_i}$ or $\dfrac{\partial v_i}{\partial r_j} = \dfrac{\partial v_j}{\partial r_i}$ (Symmetry).

Property (a) can be used to generate systems of factor demand functions. Property (b) is of use in reducing the number of parameters to be estimated, thus conserving degrees of freedom and possibly eliminating multicollinearity problems.

3.4. *The Profit Function*

Cost minimization can be construed as the first stage of a two-stage procedure. The second stage, given an exogenous output price vector $\mathbf{s}$, is the selection of $\mathbf{y}$ to maximize profit. Profit maximization for all $\mathbf{r} \in \mathbf{\Omega}_n^*$, $\mathbf{s} \in \mathbf{\Omega}_m^*$ is described by the profit function

$$\Pi(\mathbf{s},\mathbf{r}) = \max\{\mathbf{s}\cdot\mathbf{y} - \mathbf{r}\cdot\mathbf{v} : (\mathbf{v},\mathbf{y}) \in \mathbf{Y})\},$$

or

$$\Pi(\mathbf{p}) = \max\{\mathbf{p}\cdot\mathbf{x} : \mathbf{x} \in \mathbf{Y}\},$$

where $\mathbf{x}$ is a net output vector (>0 for outputs, and <0 for inputs) and $\mathbf{p} = (\mathbf{r},\mathbf{s}) \in \mathbf{\Omega}_{m+n}^*$.

If $F(\hat{\mathbf{y}},\mathbf{v})$ possesses Properties (2.1.1) through (2.1.4) then $\Pi(\mathbf{p})$ has Properties (4.1)–(4.5):

4.1 *Domain.* $\Pi(\mathbf{p})$ is a non-negative extended real-valued function defined for all positive prices $\mathbf{p}$.

4.2 *Monotonicity.* $\Pi(\mathbf{p})$ is a non-decreasing function of output prices and non-increasing function of input prices.

4.3 *Continuity.* $\Pi(\mathbf{p})$ is continuous in $\mathbf{p}$.

4.4 *Convexity.* $\Pi(\mathbf{p})$ is a convex function in $\mathbf{p}$.

4.5 *Homogeneity.* $\Pi(\mathbf{p})$ is linear homogeneous in $\mathbf{p}$.

Again, empirical analysis normally assumes, in addition,

4.6 *Differentiability.* Most empirical applications assume $\Pi(\mathbf{p})$ is twice differentiable. As was the case with the cost function, the profit function possesses two important corresponding derivative properties

(a) $$\frac{\partial \Pi(\mathbf{p})}{\partial p_i} = x_i, \qquad i = 1,\dots,m+n \quad \text{(Hotelling's Lemma)},$$

(b) $$\frac{\partial^2 \Pi}{\partial p_i \partial p_j} = \frac{\partial^2 \Pi}{\partial p_j \partial p_i} \Rightarrow \frac{\partial x_i}{\partial p_j} = \frac{\partial x_j}{\partial p_i} \qquad \text{(Symmetry)}.$$

4. A General Approach – Forms Linear-in-Parameters

4.1. *Parameterization of Economic Effects*

The main body of econometric and statistical technique requires models whose form is specified up to a finite vector of unknown parameters. This leads to the consideration of specific parametric production models which allow identification of particular economic effects, such as distribution and scale, while imposing no more maintained hypotheses than necessary on other aspects of technology. To a large extent we will be concerned with flexible representations of technologies, since flexibility is the issue which has led econometricians to seek alternatives to the first parametric production function, the Cobb–Douglas form [see Douglas and Cobb (1928)].

The objective of flexibility can be used to classify functional forms. Following Hanoch (1975a), we can specify the number of parameters required for representation of the economic effects discussed in Section 1. Consider an n input, one output production function $y = f(v_1,\dots,v_n)$, with partial derivatives $f_i = \partial f/\partial v_i$ and $f_{ij} = \partial^2 f/\partial v_i \partial v_j$. Economic effects such as scale, distribution, and substitutability can in general be quantified in terms of the production function and its first and second derivatives. Consider the following classification of these effects:

Economic effect	Formula	Number of distinct effects
Output level	$y = f(\mathbf{v})$	1
Returns to scale	$\mu = \left(\sum_{i=1}^{n} v_i f_i\right) \Big/ f$	1
Distributive share	$s_i = v_i f_i \Big/ \sum_{j=1}^{n} v_j f_j$	$n - 1$
Own "price" elasticity	$\epsilon_i = v_i f_{ii} / f_i$	n
Elasticity of substitution	$\sigma_{ij} = \dfrac{-f_{ii}/f_i^2 + 2(f_{ij}/f_i f_j) - f_{jj}/f_j^2}{1/v_i f_i + 1/v_j f_j}$	$\dfrac{n(n-1)}{2}$

This table contains $(n+1)(n+2)/2$ distinct economic effects. These effects characterize the usual comparative statics properties of a production function at a point.[3] These formulae can be inverted to determine the function value and the first and second partial derivatives at a point in terms of economic effects,

$$f = y,$$
$$f_i = \mu y s_i / v_i,$$
$$f_{ii} = \mu y s_i \epsilon_i / v_i^2,$$
$$f_{ij} = [\sigma_{ij}(s_i + s_j) + \epsilon_i s_i + \epsilon_j s_j] \mu y / 2 v_i v_j, \qquad i \neq j.$$

Hence, a necessary and sufficient condition for a functional form to reproduce comparative statics effects *at a point* without imposing restrictions across these effects is that it have $(n+1)(n+2)/2$ distinct parameters, such as would be provided by a Taylor's expansion to second-order.

[3]Exogenous technical change could be included by adding a variable t to the exogenous variables included in f. Then, $n+2$ economic effects would be added: the rate of technical change, $T = f_t/f$, the acceleration of technical change, $\dot{T} = (f_{tt}/f) - (f_t/f)^2$, and the rates of change of marginal products, $\dot{m}_i/m_i = f_{it}/f_i$. There would then be a total of $(n+2)(n+3)/2$ effects.

The development above in terms of a production function could equally well have been carried out in terms of a cost or profit function. Since the latter functions have $n+1$ arguments, compared to the n arguments of the production function, they may appear to permit a larger number of distinct effects involving first and second partial derivatives. However, the homogeneity properties of these functions reduce the number of independent parameters to $(n+1)(n+2)/2$, as before. For example, consider the cost function $C = C(y,\mathbf{r})$. Since C is homogeneous of degree one in $\mathbf{r}$, Euler's Theorem implies

$$\sum_{i=1}^{n} r_i C_i(y,\mathbf{r}) = C(y,\mathbf{r}),$$

$$\sum_{i=1}^{n} r_i C_{ij}(y,\mathbf{r}) = 0, \qquad j = 1,\ldots,n,$$

and

$$\sum_{i=1}^{n} r_i C_{yi}(y,\mathbf{r}) = C_y(y,\mathbf{r}),$$

where $C_i = \partial C/\partial r_i$, $C_{ij} = \partial^2 C/\partial r_i \partial r_j$, $C_y = \partial C/\partial y$, and $C_{yi} = \partial^2 C/\partial y \partial r_i$. These provide $n+2$ restrictions, known as the adding-up condition, the Cournot aggregation conditions, and the Engel aggregation condition, respectively. The number of distinct derivative conditions is therefore $(n+2)(n+3)/2 - (n+2) = (n+1)(n+2)/2$, as in the case of the production function.[4]

4.2. *Linear-in-Parameters Approximations*

Most of the flexible functional forms developed in the econometric literature can be viewed as linear-in-parameters expansions which approximate an arbitrary function. In general, such an expansion can be written in the form

$$f^*(\mathbf{x}) \approx f(\mathbf{x}) \equiv \sum_{i=1}^{N} a_i h^i(\mathbf{x}), \qquad (1)$$

where f^* is the true function, f is the approximating functional form, the

[4] This argument can be applied to an n-input linear homogeneous production function to show that it has $n(n+1)/2$ distinct economic effects.

a_i are parameters, the h^i are known functions, and $\mathbf{x}$ is a vector of independent variables. In production applications, $\mathbf{x}$ may be input quantities or prices, or transformations of these variables (e.g., a log transformation). If $N = (n+1)(n+2)/2$ and a determinental condition (a non-singular Wronskian)[5] is satisfied at a point $\mathbf{x}^*$, then parameter values a_i can be found for which this expansion approximates the value of $f(\mathbf{x})$ and its first and second partial derivatives in a neighborhood of $\mathbf{x}^*$. We term an expansion with this property a *parsimonious flexible* form.

A common method of generating parsimonious flexible forms is by use of a Taylor's series expansion to second-order about a point $\mathbf{x}^*$. In this case, the known functions and corresponding parameters have the values

$$\begin{aligned} &h^0(\mathbf{x}) = 1, && a_0 = f^*(\mathbf{x}^*), \\ &h^i(\mathbf{x}) = x_i - x_i^*, && a_i = f_i^*(\mathbf{x}^*), \quad i = 1,\dots,n, \\ &h^{ij}(\mathbf{x}) = (1/2)(x_i - x_i^*)(x_j - x_j^*), && a_{ij} = f_{ij}^*(\mathbf{x}^*), \quad i,j = 1,\dots,n. \end{aligned}$$

[For notational simplicity, the second-order terms in (1) have been reindexed in terms of i and j.]

A problem which arises when we consider parsimonious flexible functional forms as approximations to true functions is the accuracy of the approximation. If a flexible form is calibrated to provide a second-order approximation at a point, then the approximation is of this order only in a small neighborhood of this point. In other regions of interest, the form may be a poor approximation to the true function, and may even fail to satisfy basic properties of the true function such as mono-

[5]The Wronskian is the determinant

$$\begin{vmatrix} h^0(\mathbf{x}^*) & h^1(\mathbf{x}^*) & h^N(\mathbf{x}^*) \\ \partial h^0(\mathbf{x}^*)/\partial x_1 & \partial h^1(\mathbf{x}^*)/\partial x_1 & \partial h^N(\mathbf{x}^*)/\partial x_1 \\ \vdots & \vdots & \vdots \\ \partial h^0(\mathbf{x}^*)/\partial x_n & \vdots & \partial h^N(\mathbf{x}^*)/\partial x_n \\ \partial^2 h^0(\mathbf{x}^*)/\partial x_1^2 & \vdots & \partial^2 h^N(\mathbf{x}^*)/\partial x_1^2 \\ \partial^2 h^0(\mathbf{x}^*)/\partial x_1 \partial x_2 & \vdots & \partial^2 h^N(\mathbf{x}^*)/\partial x_1 \partial x_2 \\ \vdots & \vdots & \vdots \\ \partial^2 h^0(\mathbf{x}^*)/\partial x_n^2 & \cdot & \partial^2 h^N(\mathbf{x}^*)/\partial x_n^2 \end{vmatrix}$$

When this determinant is non-zero, the coefficients a_i in (1) can be chosen so that the approximation to f has first- and second-order derivatives at $\mathbf{x}^*$ equal to those of f at $\mathbf{x}^*$.

tonicity or convexity.[6] Further, the qualitative implications of the calibrated approximation may depend on the point of approximation; this is true, for example, of separability, which involves properties of the true function beyond second-order (see Section 6). The economic effects of interest in comparative statics, while unrestricted at the point of approximation, can be strongly and perhaps implausibly related at different points in the domain of the expansion.

If a parsimonious flexible form is fitted to observations over an extensive domain, as is normally the case in econometric production analysis, then the fitted form will not in general be a second-order approximation to the true function at any chosen point. As a result, the comparative statics effects deduced from the approximation will bear a complex and perhaps misleading relationship to the corresponding effects for the true function. In particular, multivariate fits to the approximate function and its derivatives may fail to satisfy restrictions on parameters across equations, even when the true function satisfies properties implying these restrictions. This could lead the analyst to conclude incorrectly that the true function fails to satisfy the properties in question. For example, tests of "profit-maximizing behavior" based on symmetry restrictions across equations may be rejected in the system of fitted functions even if the property holds in the true system. Note that this conclusion depends critically on the assumption that the expansion is being fitted to data over a large domain; a second-order approximation at a point will satisfy symmetry restrictions across equations when the true system does.

A simple example may help to clarify the issues raised in the preceding paragraphs. Suppose a true one-input production function is $y = e^v$, exhibiting increasing returns at an increasing rate (for $v > 1$), $\mu = v$, and a positive own-price elasticity, $\epsilon = v$. Suppose we approximate this production function with an expansion in logarithms, $\log y = a_1 + a_2 \log v + a_3(\log v)^2$. The estimated returns to scale and own-price elasticity from the expansion are $\hat{\mu} = a_2 + 2a_3 \log v$ and $\hat{\epsilon} = (2a_3/\hat{\mu}) + \hat{\mu} - 1$, respectively. Suppose $\log v$ is normally distributed with mean $\log m$ and variance σ^2. Then, an ordinary least-squares fit of the parameters in the expansion converges in probability to $a_1 =$

[6]Some expansions, such as the Translog function discussed below, can never except in trivial cases satisfy monotonicity or convexity conditions over the entire positive orthant. Hence, it is important in using these expansions to test for the satisfaction of maintained hypotheses in regions of interest. In Appendix A.4 of this volume, Lau provides computational methods for verifying convexity.

$me^{\sigma^2/2}[1 - \log m + \frac{1}{2}(\log m)^2 - \sigma^2)]$, $a_2 = me^{\sigma^2/2}(1 - \log m)$, and $a_3 = (m/2)e^{\sigma^2/2}$. Alternatively, a second-order approximation to the true function at a point $v = m$ satisfies these formulae with $\sigma^2 = 0$.) Let $\hat{y}$, $\hat{\mu}$, and $\hat{\epsilon}$ denote the economic effects measured from the fitted expansion. Then, for example,

$$\frac{\hat{\mu}}{\mu} = e^{\sigma^2/2}\left(1 + \log\frac{v}{m}\right)\bigg/\frac{v}{m},$$

which attains a maximum of $e^{\sigma^2/2}$ at $v = m$. Hence, a second-order expansion at a point will underestimate the returns to scale effect except at the point. A fit to data for which $\log v$ is normal with mean $\log m$ and variance σ^2 yields an overestimate of returns to scale at the data mean.

Table 1 indicates the accuracy of the approximation to y, μ, and ϵ for three alternative expansions. In each case, the approximation is good (say, within 10 percent) only in a narrow range, and is particularly poor for small v where the expansions fail to satisfy monotonicity. The effect of fitting the expansion to log normal data with $m = 10$, $\sigma^2 = 1$ is a

TABLE 1

	Second-order fit $m = 1$			Second-order fit $m = 10$			Data fit when $\log v$ has mean $\log m = \log 10$, var $\sigma^2 = 1$		
v	$\frac{\log \hat{y}}{\log y}$	$\frac{\hat{\mu}}{\mu}$	$\frac{\hat{\epsilon}}{\epsilon}$	$\frac{\log \hat{y}}{\log y}$	$\frac{\hat{\mu}}{\mu}$	$\frac{\hat{\epsilon}}{\epsilon}$	$\frac{\log \hat{y}}{\log y}$	$\frac{\hat{\mu}}{\mu}$	$\frac{\hat{\epsilon}}{\epsilon}$
0.3	1.74	−0.68	−20.36	121.38	−83.55	−88.22	172.64	−137.75	−142.42
0.4	1.26	0.21	27.57	74.04	−55.47	−59.10	101.47	−91.46	−95.08
0.5	1.09	0.61	5.13	49.83	−39.91	−42.92	65.67	−65.81	−68.81
0.9	1.00	0.99	1.12	16.57	−15.64	−17.54	18.16	−25.79	−27.69
1.0	1.00	1.00	1.00	13.48	−13.03	−14.79	13.99	−21.48	−23.24
1.1	1.00	1.00	0.92	11.17	−10.98	−12.64	10.92	−18.10	−19.76
2.0	0.97	0.85	0.64	3.43	−3.05	−4.37	1.53	−5.02	−6.34
3.0	0.90	0.70	0.53	1.74	−0.68	−2.65	0.11	−1.12	−3.09
4.0	0.84	0.60	0.45	1.26	0.21	2.95	0.01	0.35	3.08
5.0	0.78	0.52	0.40	1.09	0.61	1.07	0.16	1.01	1.46
9.0	0.62	0.36	0.28	1.00	0.99	1.01	0.73	1.64	1.65
10.0	0.60	0.33	0.26	1.00	1.00	1.00	0.82	1.65	1.65
11.0	0.57	0.31	0.24	1.00	1.00	0.99	0.90	1.64	1.63
20.0	0.42	0.20	0.16	0.97	0.85	0.83	1.18	1.40	1.38
30.0	0.34	0.15	0.12	0.90	0.70	0.68	1.21	1.15	1.14
40.0	0.29	0.12	0.10	0.84	0.60	0.58	1.17	0.98	0.97
50.0	0.25	0.10	0.08	0.78	0.52	0.51	1.12	0.86	0.85
90.0	0.17	0.06	0.05	0.62	0.36	0.35	0.94	0.59	0.58

substantial overestimate of μ and ϵ in the range $3.68 \leqq v \leqq 27.18$, which contains 68 percent of the data. This example suggests that fitted expansions can be relatively non-robust with respect to the point of approximation or range of data available, and that considerable caution should be used in utilizing the models for extrapolative prediction or the testing of basic hypotheses on production structure.

In principle, the difficulty in obtaining accurate approximations in the large can be overcome by introducing additional parameters. On a closed bounded domain, the Bernstein–Weierstrauss approximation theorem shows that a continuous function can be approximated uniformly by polynomials.[7] In practice, the number of parameters required in these theorems to guarantee a specified level of accuracy is too large for empirical purposes. A theory of approximation in the large for production functions which incorporates the qualitative properties of the true functions such as monotonicity and convexity might produce tighter bounds on the number of parameters required; however, this topic is beyond the scope of this survey.

[7]Suppose the domain of interest is defined – by translation, normalization, and extension if necessary – to be the closed bounded set $\mathbf{S} = \{(v_1,\ldots,v_n) \geqq 0 | v_1 + \cdots + v_n \leqq 1\}$. Consider the class of functions f which are uniformly Lipschitzian on $\mathbf{S}$ with constant M; i.e., $|f(\mathbf{v}) - f(\mathbf{v}')| \leqq M\|\mathbf{v} - \mathbf{v}'\|$. Define a multivariate Bernstein polynomial

$$B_N(\mathbf{v}) = \sum_{(k_1,\ldots,k_n)\in\mathbf{K}} f(k_1/N,\ldots,k_n/N)b_N(\mathbf{v};k_1,\ldots k_n),$$

where $\mathbf{K}$ is the set of integer vectors $(k_1,\ldots,k_n)$ with $(k_1/N,\ldots,k_n/N) \in \mathbf{S}$, and

$$b_N(\mathbf{v};k_1,\ldots,k_n) = \{N!/(k_1!\cdots k_n!(N - k_1 - \cdots - k_n)!)\}v_1^{k_1}\cdots v_n^{k_n}(1 - v_1 - \cdots - v_n)^{N-k_1-\cdots-k_n}.$$

Given $\epsilon > 0$, if $N \geqq nM^2/\epsilon^2$, then $|f(v) - B_N(v)| \leqq \epsilon$ uniformly on the cube. To establish this result, define $\mathbf{K}_1 = \{k \in \mathbf{K} | \|\mathbf{v} - \mathbf{k}/N\| \leqq (n/4N)^{1/2}\}$ and $\mathbf{K}_2 = \mathbf{K}\backslash\mathbf{K}_1$. Then

$$\begin{aligned} |f(\mathbf{v}) - B_N(\mathbf{v})| &\leqq \sum_{k\in K_1} \left|f(\mathbf{v}) - f\left(\frac{\mathbf{k}}{N}\right)\right| b_N(\mathbf{v};\mathbf{k}) + \sum_{k\in K_2} M \frac{\|\mathbf{v} - \mathbf{k}/N\|^2}{\|\mathbf{v} - \mathbf{k}/N\|} b_N(\mathbf{v};\mathbf{k}) \\ &\leqq (n/4N)^{1/2} M \sum_{k\in K} b_N(\mathbf{v};\mathbf{k}) + M(n/4N)^{-1/2} \sum_{k\in K} \|\mathbf{v} - \mathbf{k}/N\|^2 b_N(\mathbf{v};\mathbf{k}) \\ &\leqq M(n/4N)^{1/2} + M(n/4N)^{-1/2} n \sum_{i=1}^{n} \frac{v_i(1 - v_i)}{N} \\ &\leqq M\left(\frac{n}{N}\right)^{1/2} \leqq \epsilon, \end{aligned}$$

where the second and third inequalities follow from properties of the multinomial distribution.

4.3. Common Linear-in-Parameters Forms

Table 2 provides, in summary form, a list of the most commonly used linear-in-parameters functional forms and their approximation characteristics. The historic Cobb–Douglas function, while not originally proposed as an approximation, can be viewed as a first-order expansion in $\log v_i$ about $v_i = 1$. This form allows free assignment of the output level, returns to scale, and distributive shares effects at a point of approximation, but allows no flexibility with respect to the substitution and own-price elasticity effects. The CES function adds one substitution parameter to the (linear homogeneous) Cobb–Douglas case. We have included this functional form in the table, even though it is not linear-in-parameters unless the substitution parameter ρ is known, because it is the basis for several linear-in-parameter expansions.

The concept of linear-in-parameters functional forms and the property of second-order approximation at a point are due to Diewert (1971), who introduced the generalized linear and generalized Leontief systems. This development was followed by the introduction of the translog functional form by Christensen, Jorgenson, and Lau (1971). A direct generalization of the Cobb–Douglas function, the translog form has been widely used as a framework for analysis of structural properties of production.

All the forms in Table 2 with the exception of the Quadratic have restrictions implying linear homogeneity, and under this restriction have $n(n+1)/2$ parameters, as required for a parsimonious flexible linear homogeneous function. In the absence of homogeneity restrictions, the forms having $(n+1)(n+2)/2$ parameters are Generalized Leontief, Translog, and Quadratic. With the exception of Generalized Cobb–Douglas and Generalized Concave forms, the functions in Table 2 can be interpreted as Taylor's expansions about a point. In this interpretation, first proposed explicitly by Lau (1974), Cobb–Douglas is a first-order expansion of $\log y$ in powers of $\log x_i$, and Translog is a second-order expansion. CES is a first-order expansion of y^ρ in powers of x_i^ρ. Generalized Leontief and Quadratic are second-order expansions of y in powers of $\sqrt{x_i}$ and x_i, respectively.

Generally, the forms in Table 2, or analogous forms that could be obtained using other series of functions as a basis for expansions, will provide equally satisfactory representations of an arbitrary production function at a point. Choice between them should be based on their quality as approximations to the true functions over the full domain of

TABLE 2
Common linear-in-parameters functional forms.[a]

Functional form	Formula	Restrictions
Cobb–Douglas [Douglas–Cobb (1928)]	$\log y = a_0 + \sum_{i=1}^{n} a_i \log x_i$	$\sum_{i=1}^{n} a_i = 1$ for linear homogeneity
CES [Arrow et al. (1961)][b]	$y^{\rho} = a_0 + \sum_{i=1}^{n} a_i x_i^{\rho}$	$a_0 = 0$ for linear homogeneity
Generalized Leontief/Linear [Diewert (1971)]	$y = a_0 + \sum_{i=1}^{n} a_i \sqrt{x_i} + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \sqrt{(x_i x_j)}$	$a_i = 0,\ i = 0,\ldots,n$ for linear homogeneity
Translog [Christensen–Jorgenson–Lau (1971)]	$\log y = a_0 + \sum_{i=1}^{n} a_i \log x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} (\log x_i)(\log x_j)$	$\sum_{i=1}^{n} a_i = 1$ and $\sum a_{ij} = 0$ for linear homogeneity
Generalized Cobb–Douglas [Diewert (1973b)]	$\log y = a_0 + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \log(x_i + x_j)/2$	$\sum_i \sum_j a_{ij} = 1$ for linear homogeneity
Quadratic [Lau (1974)]	$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$	
Generalized Concave [McFadden (Chapter II.2)]	$y = \sum_{i=1}^{n} \sum_{j=1}^{n} x_j \phi^{ij}(x_i/x_j) a_{ij}$	ϕ^{ij} a known concave[c] function

[a] $\mathbf{x}$ is a vector of inputs or prices, depending on whether a direct or dual form is being considered. y is output, cost, or profit. Except for the case of self-dual forms [see Hanoch (Chapter I.2)], these formulae defined in terms of prices yield different technologies than are obtained from these formulae defined in terms of quantities. Linear homogeneous functions can also be represented by these forms by expressing all variables in relative terms.

[b] If

$$a_0 = 0 \quad \text{and} \quad \sum_{i=1}^{n} a_i = 1$$

in the CES formula, then a first-order expansion in ρ,

$$\log y = \sum_{i=1}^{n} a_i \log x_i + \frac{\rho}{2}\left[\sum_{i=1}^{n} a_i (\log x_i)^2 - \left(\sum_{i=1}^{n} a_i \log x_i\right)^2\right],$$

provides a linear-in-parameters translog form with second-order terms

$$a_{ii} = \frac{\rho}{2} a_i (1 - a_i) \quad \text{and} \quad a_{ij} = \frac{\rho}{2} a_i a_j, \quad i \neq j.$$

This expansion was first suggested by Kmenta (1967), and has been utilized by Sargan (1971) and Griliches and Ringstad (1971). An alternative derivation of this expansion from the translog function, based on the imposition of a common AES between input pairs, has been given by Denny and Fuss (1977).

[c] For a profit function, the ϕ^{ij} are convex functions.

interest, to the extent that this can be assessed *a priori*, and on the ease with which hypotheses of interest can be stated as restrictions on parameters.

5. Special Non-Linear Forms

5.1. *Elasticities of Substitution*

The flexible forms discussed in Section 4 can be viewed as extensions of simple functional forms where the extensions are constrained to remain linear-in-parameters. For example, Diewert's Generalized Leontief cost function is just such an extension of the cost function dual to a Leontief fixed coefficient production function. The Translog extends the Cobb–Douglas function and the Quadratic extends a linear function under the same linear-in-parameters constraint. While linearity is retained, it is at the cost of introducing a large number of parameters into the analysis. The variants of simple functional forms surveyed in this section are characterized by non-linearity in parameters. This fact makes them less useful in general for econometric estimation than those forms surveyed in Section 4. However, in some cases, non-linearity is compensated for by parsimony in parameters. An example discussed in this section is Hanoch's CRESH–CDE form for use in the study of factor substitution.

Most of the forms surveyed in this section were devised to generalize, using as a few additional parameters as possible, two restrictive features of the maintained hypotheses concerning substitution effects of the original ACMS function. First, in the two-factor case, the elasticity of substitution is constrained to be constant, and there is no apparent technological justification for this restriction. Second, extension of the CES function to more than two factors requires, with unimportant exceptions, the imposition of the maintained hypothesis that all partial AES are equal and constant [Uzawa (1962)]. In the multiple factor case, it is not clear that the AES will be the desired concept of the elasticity of substitution (ES). The attempts to apply this concept to the case of more than two inputs have produced various definitions [McFadden (1963)]. As indicated by Mundlak (1968b), those definitions differ in two major respects: (1) the variables which are held constant in the underlying economic experiment and (2) the number of variables which are involved in the operation. If we denote $\hat{v} = \mathrm{d}\ln v$, and assume that all derivatives are evaluated at an equilibrium point, we can distinguish between one-factor one-price ES (OOES), $\hat{v}_j/\hat{r}_i$ (the AES is of this form), two-factor one-price ES (TOES), $(\hat{v}_j - \hat{v}_i)/\hat{r}_i$, and two-factor two-price ES (TTES), $(\hat{v}_j - \hat{v}_i)/(\hat{r}_i - \hat{r}_j)$. The last is the "usual" definition of ES. Each of these concepts can be evaluated at constant output, cost, or marginal cost. Each of these alternatives corresponds to a different factor demand curve, where the prices not involved in the operation are held constant. However, it is also possible to hold constant the quantities of the factors which are not involved in the operation. In one extreme "short-run" case we have the direct ES (DES), which is a TTES with all factors other than those involved in the operation held constant. In the extreme "long-run" case, we have the shadow ES (SES), in which all quantities are allowed to vary. We can also have mixed situations in which the quantities of some factors and the prices of other factors are held constant. Detailed discussions of various definitions of the ES are given in McFadden (1963), Hanoch (Chapter II.3), and Lau (Chapter I.3). All these forms collapse to a common definition in a two-factor linear homogeneous production function. This is due to the singularity of the Hessian matrix, and therefore it cannot be used as an indication that any of the above expressions is a generalization of the two-factor measure [Mundlak (1968b, p. 231)].

In summary, once we depart from the two-input case we confront the following problems in attempting to develop production functions from the point of view of the elasticity of substitution:

(a) There is no unique natural generalization of the two factor definition of the ES. The different definitions involve different combinations of the elements of the underlying Hessian matrix. It is therefore reasonable to deal with the Hessian elements directly. The AES comes close to this approach. Other than that, it has no particular advantage over the others and perhaps the reference to it as an elasticity of substitution is misleading. It is simply proportional to the cross elasticity in the constant output factor demand function. We conclude that the selection of a particular definition should depend on the question asked.

(b) The choice of an ES does *not* imply constancy of the elasticity; this is an added hypothesis which may not hold in reality. As a result, there is no direct relationship between the concept of the ES to be used and the algebraic form of the production function.

Non-linear forms have been analyzed primarily in terms of the AES, and in the pages which follow we will maintain the classification in terms of the AES. However, we note that for the reasons above, it might also be useful to pursue a classification similar to that in Section 4.

5.2. *Variants of the Cobb–Douglas Function*

(1) *Variable elasticity of substitution* (*VES*) *production function* [Revankar (1971)]. This function was devised to relax the assumption of constant AES in the two-factor case. It takes the form

$$y = \alpha_0 v_1^{\alpha_1}(v_2 + \gamma_1 v_1)^{\alpha_2},$$

and has an AES $= 1 + \beta(v_1/v_2)$ where β is a function of the production function parameters. It is considered a variant of the CD form since the AES varies around one for $\beta \neq 0$ and variations in relative inputs.

(2) *Constant marginal share* (*CMS*) *function* [*Bruno* (1968)]. This function is explicitly a generalization of the CD form. It can be expressed as

$$y = \alpha_0 v_1^{\alpha_1} v_2^{\alpha_2} - \gamma v_2,$$

and has an AES $= 1 - (\gamma\alpha_1/\alpha_2)(v_2/y)$.

(3) *Transcendental production function* [Halter, Carter, and Hocking (1957)]. This function has the form

$$y = \alpha_0 v_1^{1-\alpha} v_2^{\alpha} e^{\gamma_1 v_1 + \gamma_2 v_2},$$

and has an AES $= (1 - \alpha + \gamma_1 v_1)(\alpha + \gamma_2 v_2)/((1-\alpha)(\alpha + \gamma_2 v_2)^2 + \alpha(1 - \alpha + \gamma_1 v_1)^2)$ which reduces to unity when $\gamma_1 = \gamma_2 = 0$.

5.3. *Variants of the CES Function*

Most of the variants of the CES function can be seen as the result of attempting to eliminate the assumption contained in the multifactor CES formulation, namely, equality of all partial AES [Uzawa (1962), McFadden (1963)]. One extension which relaxes this restriction is the nested CES function [Sato (1967), see also McFadden (Chapter IV.1)]. This form has not been used extensively in empirical work due to its complex nature when extended to more than three factors [however, see Mundlak and Razin (1969, 1971)]. Another avenue for CES-like extensions is the class of implicitly additive forms introduced by Hanoch (1975a). The direct form is

$$F(\mathbf{v},y) = \sum_{i=1}^{n} F^i(v_i,y) \equiv 1, \tag{3}$$

where F^i are functions with properties imposed to insure that the implied explicit production function satisfies the maintained hypotheses of Sections 3.1 and 3.2. The dual, or indirect form, is

$$G(\mathbf{r}/c,y) = \sum_i G^i(r_i/c,y) \equiv 1,^8 \tag{4}$$

where G^i are functions with properties imposed to insure that the implied cost function c satisfies the maintained hypotheses of Section 3.3. This class of functions contains as special cases the direct forms which have constant ratios of AES; such as the one derived by Mukerji (1963) and Gorman (1965) [and used by Dhrymes and Kurz (1964)] and the CRESH form developed by Hanoch (1971). The indirect form contains as special cases functions which exhibit constant differences in AES such as the CDE form of Hanoch (1971). The Mukerji form uses the transformation $F^i(v_i,y) = D_i(v_i^{d_i}/y^h)$, while the CRESH form uses the transformation $F^i(v_i,y) = D_i(v_i/y^h)^{d_i}$. The CDE form uses the transformation $G^i(r_i/c,y) = D_i(y^h r_i/c)^{d_i}$. In these transformations, D_i, d_i, and h are parameters. Detailed discussion of these transformations and extensions can be found in Hanoch (1975a).

Estimating equations for the indirectly additive class contain a small number of parameters when compared with the general linear expansions of Section 4. For example, consider the CRES (constant ratio elasticity of substitution – non-homothetic) form introduced by Hanoch (1975a),

$$\sum_i D_i y^{-e_i d_i} v_i^{d_i} \equiv 1, \tag{5}$$

where D_i, d_i, and e_i are parameters. Using the first-order conditions for profit maximization, one obtains the set of equations

$$\log v_i = A_i - a_i \log(r_i/r_1) + h_i \log y + (a_i/a_1) \log v_1, \qquad i = 1,\ldots,n,$$

where

$$a_i = 1/(1-d_i), \quad A_i = \log(D_i d_i/D_1 d_1)^{a_i}, \quad h_i = a_i(e_1 d_1 - e_i d_i),$$

and

$$\mathrm{AES}_{i,k}/\mathrm{AES}_{j,k} = a_i/a_j.$$

[8]For extensions of the indirectly additive class to the multiple output case see, Hanoch (Chapter II.3).

The above set of equations is non-linear in a_i and simultaneous in v_i but could be estimated using non-linear simultaneous equations procedures currently available. Note that there are $3n-2$ parameters to estimate in the system of equations (5) as compared with $(n+2)(n+1)/2$ for the second order linear-in-parameters approximations. The cost (in addition to the non-linearity) is the maintained hypothesis of implicit separability which is reflected in the fact that only two input prices appear as exogeneous variables in each demand equation. This is an example of the importance of separability assumptions for functional specification. It is to this issue that we now turn.

6. Separability: Functional Implications and Tests

6.1. *Basic Concepts*

Separability has various implications. It allows decentralization in decision-making or equivalently, optimization by stages. This opens up the possibility of consistent multi-stage estimation which may be the only feasible procedure when large numbers of inputs and outputs are involved; specifically, when applying the relatively simple concept of a production function to complex organizations. Historically, separability has played an important role in the specification of functional forms. The Cobb–Douglas and CES functions are explicitly strongly separable. Hanoch's (1971) CRESH–CDE class of functions is implicitly strongly separable. Sato's (1967) nested CES specification is strongly separable with respect to the highest level partition and then strongly separable within each sub-aggregate.

To define separability, we first denote the set of n inputs by $\mathbf{N}=\{1,\ldots,n\}$. A partition $\mathbf{S}$ of $\mathbf{N}$ is given by $\{\mathbf{N}_1,\ldots,\mathbf{N}_S\}$ where $\mathbf{N}=\mathbf{N}_1\cup\mathbf{N}_2\cdots\cup\mathbf{N}_S$, and $\mathbf{N}_r\cap\mathbf{N}_t$ is empty for $r\neq t$. Separability is characterized by the independence of the marginal rate of substitution between a pair of inputs from changes in the level of another input, i.e.,

$$\frac{\partial(f_i/f_j)}{\partial v_k}=0, \tag{6}$$

or $f_jf_{ik}-f_if_{jk}=0$. We say that f is *strongly separable* (SS) with respect to the partition $\mathbf{S}$ if (6) holds for all $i\in\mathbf{N}_r$, $j\in\mathbf{N}_t$, and $k\notin\mathbf{N}_r\cup\mathbf{N}_t$. The function is *weakly separable* (WS) with respect to the partition **S** if (6) holds for all $i,j \in \mathbf{N}_r$, and $k \notin \mathbf{N}_r$. Note that these properties may hold at a point or globally.

Goldman and Uzawa (1964) showed that a function $f(\mathbf{x})$ is globally SS with respect to the partition **S** ($S > 2$) if and only if $f(\mathbf{x}) = F\{\sum_{t=1}^{S} f^t(\mathbf{x}^t)\}$ where F is monotone increasing and $f^t(\mathbf{x}^t)$ is some function of $\mathbf{x}^t$. The function is globally WS with respect to the partition **S** if and only if it is of the form

$$f(\mathbf{x}) = G\{g^1(\mathbf{x}^1),\dots,g^S(\mathbf{x}^S)\}. \tag{7}$$

Berndt and Christensen (1973b) related separability to AES and obtained the result that any strictly quasi-concave homothetic production function $f(\mathbf{v})$ is WS with respect to the partition *S at a point* if and only if $\text{AES}_{ik} = \text{AES}_{jk}$ at that point for all $i,j \in \mathbf{N}_r$, $k \notin \mathbf{N}_r$. Similarly, the function is SS at a point if and only if $\text{AES}_{ik} = \text{AES}_{jk}$ for all $i \in \mathbf{N}_r$, $j \in \mathbf{N}_t$, $k \notin \mathbf{N}_r \cup \mathbf{N}_t$. Furthermore, if $n = S$, then all AES_{ik}, $i \neq k$, are equal. If this function is globally SS for any input combination then $f(\mathbf{v}) = F(\sum_{i=1}^{n} \alpha_i v_i^{\rho})$, a homothetic transformation of a CES function.

Finally, Berndt and Christensen showed that if $f(\mathbf{v})$ is homothetically separable then the dual cost function $C(y,\mathbf{r})$ is weakly separable so that

$$C_j C_{ik} - C_i C_{jk} = 0 \tag{8}$$

holds as well as (6).

In proving these theorems Berndt and Christensen use a result obtained by Lau (Chapter I.3) to the effect that the cost function is WS(SS) with respect to the partition **S** in input prices if and only if $f(\mathbf{v})$ is homothetically WS(SS) with respect to the partition **S** in input quantities.

The role of homogeneity of f in the Berndt–Christensen results is analyzed by Russell (1975), who extends the results to the case of non-homothetic production functions.

Separability results comparable to those obtained by Berndt and Christensen are developed in terms of cost and profit functions by McFadden (Chapters I.1 and IV.1) and Lau (Chapter I.3).

One important application of separability is in the derivation of value-added functions. If the gross output production function is weakly separable in primary inputs then a net output or value-added function can be defined and used for analysis. This issue is pursued by Bruno (Chapter III.1) and Denny and May (Chapter III.3).

6.2. *Separability in Forms Linear-in-Parameters*

Since the separability constraints (6) and (8) depend on second-order partial derivatives, functional forms linear-in-parameters must be at least of the second-order in the variables to contain separability as a testable implication. For example, the Cobb–Douglas function, which is of the first-order in logarithms, maintains separability since $(\partial^2 \log f)/(\partial \log v_i)(\partial \log v_j) = 0$ for all i,j, thus satisfying (6). The class of second-order approximation functions then will be the linear in parameters class necessary in general to test separability. Separability tests of production structures using the translog specification can be found in Berndt and Christensen (1973a, 1974), Berndt and Wood (1975), Denny and Fuss (1977), and Denny and May (Chapter III.3). Similar testing of the structure of utility functions appears in Christensen, Jorgenson, and Lau (1975) and Jorgenson and Lau (1975a). An alternative approach to testing separability, in the framework of the multistage Sato function, appears in Mundlak and Razin (1971).

The above tests fall into two categories. The first category is that of "exact" tests. These tests result from the imposition of the null hypothesis of separability for all possible values of the exogenous variables. The second category consists of "approximate" tests, where the null hypothesis is imposed only at a point of approximation, utilizing the notion of the function as a second-order Taylor series expansion. Berndt–Christensen and Berndt–Wood use the exact tests, Denny–Fuss and Denny–May use the approximate ones, while Christensen et al., and Jorgenson and Lau use both (under the terminology "intrinsic" and "explicit"). Exact tests would seem to be preferable if no additional constraints are imposed, since a single reject/non-reject decision is globally applicable. Unfortunately, with second-order expansions this is not the case. Blackorby et al. (1977a) and Denny–Fuss (1977) have shown that the restriction of global weak separability implies either strong separability within the partitioned sub-aggregates, or strong separability between aggregates. For example, suppose G in (7) is translog. Then either each $g^i(\mathbf{x}^i)$ is Cobb–Douglas in $\mathbf{x}^i$ or G is Cobb–Douglas in g^i. These results can also be found in Jorgenson and Lau (1975) for the case of utility functions. We are left with a tradeoff between tests which impose extraneous restrictions and those which depend on the data point chosen as the point of approximation. While the issue remains unresolved, one possible procedure is to explore higher-order expansions [Lau (1977)], which unfortunately requires the introduction of a large number of additional parameters. Another approach is to explore forms non-linear-in-parameters, to which we now turn.

6.3. *Separability in Forms Non-Linear-in-Parameters*

We begin by illustrating a procedure suggested by Mundlak (1973a) for generating non-separable functions which contain less than the $(n+1) \times (n+2)/2$ independent parameters of the second-order approximations. To sketch the approach to this problem, let

$$y = f(\mathbf{v}) = (g * h)(\mathbf{v}), \tag{9}$$

where $f(\mathbf{v})$ is the production function, g and h are two arbitrary functions, and $*$ is an arbitrary operator; i.e., addition, multiplication, exponentiation, or composition.[9] It can be shown that g and h can both be separable while $f(\mathbf{v})$ itself is not separable.

To illustrate the use of this approach, let $*$ be addition so that (9) becomes

$$f(\mathbf{v}) = g(\mathbf{v}) + h(\mathbf{v}). \tag{10}$$

Then to evaluate (6) we can write

$$\begin{aligned} f_i f_{jk} - f_j f_{ik} &= (g_i + h_i)(g_{jk} + h_{jk}) - (g_j + h_j)(g_{ik} + h_{ik}) \\ &= (h_i h_{jk} - h_j h_{ik}) + (g_i g_{jk} - g_j g_{ik}) \\ &\quad + (h_i g_{jk} - h_j g_{ik}) + (g_i h_{jk} - g_j h_{ik}). \end{aligned} \tag{11}$$

We now note that h and g can both be separable so that the first two terms on the r.h.s. of (11) vanish. Furthermore, we can select one of the functions to be linear. For instance, let $g_i \neq 0$ for at least one i and $g_{ir} = 0$ for all i and r. Thus, if g is linear and h is separable we get

$$f_i f_{jk} - f_j f_{ik} = g_i h_{jk} - g_j h_{ik}. \tag{12}$$

For f to be non-separable with respect to i and j it is sufficient that (12) differs from zero. For instance, we can assume h to be a CD with α_j being the output elasticity with respect to the jth factor, so that $h_{ik} = \alpha_i \alpha_k (h/v_i v_k)$. Then (7) becomes

$$g_i h_{jk} - g_j h_{ik} = \frac{\alpha_k}{v_k} h \left(g_i \frac{\alpha_j}{v_j} - g_j \frac{\alpha_i}{v_i} \right). \tag{13}$$

[9]By composition $g * h$, we shall mean that h becomes an argument of g; i.e., $(g * h)(\mathbf{v}) \equiv g(\mathbf{v}, h(\mathbf{v}))$.

(13) is equal to 0 for all i and j if and only if

$$\frac{g_j\alpha_i}{v_i} = \frac{g_i\alpha_j}{v_j}, \tag{14}$$

which is impossible except at a point.

Note that (14) could be used as an approximate test of separability at a point. In contrast to the translog function, the maintained hypotheses involve only $2n$ parameters.

We can further illustrate the above procedure by using it to generate a second-order approximation form which can be used to test separability among outputs within the class of exact tests.

Let $C(\mathbf{y},\mathbf{r})$ be a cost function dual to the distance function $f(\mathbf{y},\mathbf{v}) = 0$ where $\mathbf{y}$, $\mathbf{v}$, $\mathbf{r}$ are output, input, and input price vectors, respectively. Suppose

$$C(\mathbf{y},\mathbf{r}) = (g * h)(\mathbf{y},\mathbf{r}) \equiv g(\mathbf{y},h(\mathbf{r})), \tag{15}$$

where $*$ is a "composite function" operator and $h(\mathbf{r})$ is a vector consisting of elements $h_{ij}(\mathbf{r})$, $i,j = 1,\dots,n$. Let

$$h_{ij}(\mathbf{r}) = \sum_r \sum_s \alpha_{ij,kl}(r_k r_l)^{1/2} \quad \text{and} \quad g(\mathbf{y},h) = \sum_i \sum_j h_{ij}(y_i y_j)^{1/2}. \tag{16}$$

The resultant function is

$$C(\mathbf{y},\mathbf{r}) = \sum_i \sum_j \sum_k \sum_l \alpha_{ij,kl}(y_i y_j r_k r_l)^{1/2}, \tag{17}$$

which was analyzed by Hall (1973) and implemented empirically by Burgess (1976). Separability of the form $f^1(\mathbf{y}) = f^2(\mathbf{v})$ can be tested by imposing the restrictions $\alpha_{ij,kl} = a_{ij}a_{kl}$ [Hall (1973)] which results in an exact test. We arrived at the above form by combining two generalized Leontief specifications. Of course (17) still contains a large number of parameters, limiting its usefulness empirically. Nevertheless, this method of combining functions may prove useful for achieving a particular property with an efficient use of parameters.[10]

[10]The two-stage nested functional form developed in Fuss (1970) [see also Fuss (1977b) and Fuss and McFadden (Chapter II.4)] combines two generalized Leontief cost functions using a composite function rule much like that employed by Hall. This construction provides exact tests (in the sense used in the text) of the flexibility of the underlying technology.

7. Econometric Estimation of Production Parameters

The functional forms set out in Sections 4 and 5 characterize systematic relationships between economic variables, but take no account of the random effects which enter the determination of measurement of these variables. In application the stochastic specification is an intrinsic part of the specification of the production model. It should be emphasized that a specification of the model should be guided by the visualization of the *true* process and this is determined by nature and not by the econometrician. Hence the object is not to choose a specification that justifies a particular statistical procedure but on the contrary, to provide a general framework which allows for discrimination between various alternatives, as well as to examine the "robustness" of procedures dictated by the various alternatives.

Relations between *measured* production variables will in general contain stochastic components introduced at four levels: (1) the technology of the production unit, (2) the environment of each firm, particularly the market environment, (3) the behavior of the production unit, and (4) the process of observation, which often involves aggregation over commodities, production units, and time; direct errors in measurement; and incomplete observation. We discuss in turn each source of error.

Variations in technology from one production unit to the next may arise from specific or unit effects *known* to the production unit but not to the econometrician; such as management efficiency, availability and quality of specific factor inputs, and the presence of non-market inputs. They may also arise from effects which are *unknown* to the production unit at the time decisions are made. Examples are effects due to breakdown, weather, random variations in factor efficiency, and variations in quality control. The importance of the distinction between these two sources of variation [Mundlak and Hoch (1965)] is that effects known to the production unit enter the process of optimization and will be transmitted to the chosen input levels, whereas these chosen levels cannot depend on the realized values of random effects which are unknown at the time input decisions are made.[11] The statistical implications of this distinction are that observed factor inputs will be

[11]This is to some extent an oversimplification, because if production is performed by stages the error of one stage becomes a known error of higher stages [Mundlak (1963)].

endogenous if random effects are known to the production unit, and potentially exogenous if they are not.

The environment of a production unit includes a description of the markets in which input purchases and output sales must be made; the information available to the production unit at the time it makes decisions on market conditions; levels of non-market inputs; and the degree of organizational pressure or slack; as well as more general information on societal pressures on production unit decisions. For example, firms may face competitive input markets, and may purchase inputs of unknown quality in these markets at known prices, with the result that prices per efficiency unit of input are uncertain. Alternately, firms may find it necessary to contract for purchases or sales in some markets before other markets open, making relative prices uncertain. For instance, the purchase of durable inputs precedes the knowledge of all future prices of outputs and related inputs. If some markets are non-competitive, then stochastic components in demand or supply for non-competitive commodities will influence firm decisions and the resulting prevailing prices and quantities. In some cases it may be important to distinguish between stochastic effects on market equilibrium which are known to the firm, and thus part of its decision function, and those which are unknown to the firm. The knowledge need not be perfect for the argument to hold. The former will make observed prices endogenous; the latter makes them potentially exogenous.

Production unit behavior introduces stochastic components via deviations from idealized behavior patterns, as for example, failure in profit maximization to achieve exactly the desired marginal products of inputs. Such errors may arise from the finite computational ability of firms, from explicit calculations of computation costs versus expected gains, from satisficing behavior, or from firm objective functions which differ from those postulated in a maintained hypothesis by the econometrician. We note that some of these effects may introduce systematic biases into behavioral responses, and into the resulting observation. For example, Mundlak and Vulcani (1973) consider firm utility maximization, with utility depending not only on profit, but also on other variables like uncertainty and the leisure component in a production plan. It then follows that classical first-order conditions for profit maximization misspecify the true behavioral conditions, and therefore going from direct estimation of the production function to estimation of a system containing erroneous first-order conditions can be expected to worsen the quality of estimates [Mundlak (1973b)]. This caveat applies quite generally to the use of indirect forms or behavioral equations in estimating technological parameters; these forms require maintained hypotheses, such as profit maximization, in addition to those required by the basic specification of the technology. If these hypotheses prove to be false, then inferences on technology conditional on such hypotheses will be negated unless the estimation procedures can be shown to be robust. One such robust procedure, a direct estimation of the production function with prices serving as instrumental variables, is examined below. The robustness follows from the fact that even under a broader formulation, profit is considered to be an important argument in the utility function of the firm, and, ceteris paribus, prices have the same effect on quantities as in the neoclassical theory.

Broadening the framework of the analysis by allowing the utility function to include other variables in addition to profit leads to a duality relationship between technology and what may be referred to as a profit-like function – that is, a function which behaves like a profit function but whose arguments are some combination of actual prices and "prices" of the other variables that enter into the utility function. We can refer to the outcome of such combinations as pseudo prices. The profit-like function is the dual of the true production function. The use of profit rather than a profit-like function in empirical analysis can be considered as an approximation, the quality of which is to a large extent an empirical question. If however, it turns out that in a particular situation the approximation cannot be justified, the question is what information can be derived by working under the assumption that the system behaves *as if* the first-order conditions for profit maximization were met. Basically, this is a question of tracing the consequences of specification error in some equations on the model as a whole. If the technology is more stable than behavior, it may still be identified through the use of the first-order conditions for profit maximization. If on the other hand behavior is more stable, we derive behavioral equations which behave like reduced-form equations of a structure that is not fully identified.

In addition to stochastic components introduced in the technology and behavior of the production unit, there are observation errors introduced in the process of measurement of variables by the econometrician. First, classical measurement errors may occur in the process of soliciting, recording, and processing data. Second, a variety of sources of error, which can be lumped under the term aggregation errors, occur because of an inexact or ambiguous correspondence between ideal and practical definitions of variables. Further, "ideal" aggregation is determined by the true functional form, which is itself to be determined in the analysis. Thus, any given practical procedure of aggregation may lead to different kinds of aggregation error for alternative "true" production functions. Consequently, other things being equal, the aggregation error may influence the selection of a functional form in favor of the form for which the error is minimal. The aggregation problem occurs in various phases of the analysis. Aggregation over detailed commodity classifications (e.g., labor services distinguished by individual) to relatively homogeneous categories (e.g., labor services of stenographers) introduces errors. In the case of broad commodity classes, such as "capital" and "labor", these errors may be sufficiently major to influence the interpretation of the "technology". Aggregation over production units or through time may be dictated by the feasibility of data collection, or in the case of macroeconomic relationships, may be an objective of the analysis. Third, errors may arise because variables which are difficult or impossible to measure exactly are replaced by proxies, as for example the use of an average mortgage interest rate for a firm as a proxy for the actual interest rates on mortgages on specific structures.

In view of the complexity of the stochastic structure of production systems it should be clear that there is no simple universal estimation procedure. There are several alternatives whose merits depend on the relative strength of the various error components. In order to characterize these alternatives we note that the production function and the set of equations describing the first-order conditions for profit maximization constitute a complete system. The reduced-form of the system gives the product supply and factor demand equations. The profit function is an identity in the reduced-form equations.[12]

The main approaches to estimation are:

(1) direct estimation of the production function,

(2) estimation of the first-order equations,

(3) estimation of the reduced-form equations,

(4) estimation of the dual functions.

The selection of a particular approach depends not only on the stochastic specification but also on the functional form. However, in

[12]In this general discussion, we assume that all the variables are determined without any constraint on the maximization. Thus, the reduced form equations are long-run behavioral equations. If some variables are fixed, the reduced-form equations will include such constraints and thus result in short-run equations [Mundlak (1963)].

order to review the main points which have appeared in the literature dealing with the stochastic part of the model, we follow an example which assumes a very simple functional form – a Cobb–Douglas with one input only. We carefully specify and allow for the main sources of variations that have been discussed above and trace their effects on the various estimators considered. The discussion is oriented toward a cross-section analysis of firms. Some comments are also made on the possibilities which exist when there are repeated observations on firms. The specifications are listed as maintained hypotheses, and are not necessarily intended to represent an order of plausibility.

Example. An econometrician observes data on labor input (L), output (Y), and wage rate measured in output units (w) for a cross-section of firms, indexed $i = 1,...,T$. He wishes to estimate the elasticity of output with respect to labor input. The following maintained hypotheses are imposed.

7.1. *Technology*

7.1.1. *Variables*

Maintained Hypothesis 1. The technological possibilities of each firm are completely defined by two variables, the single variable input labor and a single output. There are no other variables such as capital, raw materials, knowledge, secondary outputs, etc., which vary systematically across the sample and enter the determination of technological possibilities.

7.1.2. *Functional Form*

Maintained Hypothesis 2. Each firm has the same technological possibilities, except for random effects due to (1) specific environment, management efficiency, and local labor quality, which will be referred to as the *firm effect*, and (2) breakdowns, weather, random variations in worker efficiency, which will be referred to as the *non-systematic error.* The technological possibilities have the Cobb–Douglas functional form

$$Y_* = AL_*^{\beta} e^{\epsilon+\lambda}, \qquad (18)$$

where A and β are parameters, Y_* and L_* are the "true" values of output and labor input, ϵ is the firm effect and λ is the non-systematic error. These errors are normalized so that $E\epsilon = E\lambda = 0$.

[Note: Alternative specifications might be: (1) a production function other than Cobb–Douglas, or (2) firm-to-firm variation in parameters, such as β, or a more general variation in production possibilities across firms.]

7.2. *Environment*

7.2.1. *Market Structure*

Maintained Hypothesis 3. The firm faces competitive input and output markets. The relative price in these markets varies across firms, and is non-stochastic and fixed in repeated samples.

[Note: An alternative specification might be a non-competitive input or output market with relative prices endogenous and depending on firm behavior.]

7.2.2. *Information Available to the Firm*

Maintained Hypothesis 4. At the time the firm must choose its labor input, it knows the true production function, *except* for the non-systematic error λ about which the firm forms expectations. The firm measures its "true" input and output levels without error. In particular, the firm has no ambiguity about the "quality" of input or output. The *firm* measures the relative price of labor in terms of output with a random error, $\tilde{w} = w_* e^{\xi}$, where w_* is the "true" real wage, ξ is the error, and $\tilde{w}$ is the relative price of labor seen by the firm. The source of the random error ξ may be uncertainty about price at the time the relative price is measured; e.g., the firm may measure the money wage without error and forecast the output price with error, so that the ratio of the money wage to the output price, or real wage, is measured with error. As a first approximation it is convenient to assume that $E(\xi) = 0$. In the present context this is a very restrictive assumption for it indicates that the log forecast price is on the average equal to the log true price. Therefore, eventually we shall trace the consequences of the elimination of this assumption.

[Note: Alternative and supplementary specifications might be: (1) that the firm is uncertain about its true production function, (2) that the firm makes errors in measuring the amount of "true" labor in the labor quantity it observes because of an unknown "quality" factor, or (3) that the firm exhibits some systematic bias in measuring the relative price of labor.]

7.3. *Firm Behavior*

7.3.1. *Market Posture of the Firm*

Maintained Hypothesis 5. The firm attempts to maximize competitive profit, given the information available to it and the *point* expectation that $\lambda = 0$, by a choice of the labor input. The quantity sold and actual profit are determined by the actual value of λ.

Analysis: With the point expectation $\lambda = 0$, the firm "sees" the production function

$$Y_* = Ae^{\epsilon}L_*^{\beta}, \tag{19}$$

and relative input price $\tilde{w}$. It then "sees" the profit (measured in output units)

$$\pi = Ae^{\epsilon}L_*^{\beta} - \tilde{w}L_*. \tag{20}$$

The firm chooses L_* to maximize (20), setting the marginal product of labor equal to the real wage that it "sees",

$$\partial Y_*/\partial L_* = \beta Ae^{\epsilon}L_*^{\beta-1} = \tilde{w}. \tag{21}$$

Errors in optimization can be subsumed in the random error ξ in forecasting the real wage. In this case, ξ may be subject to a firm effect, but this in turn makes the assumption of $E(\xi) = 0$ even more restrictive. As indicated, we return to this question later. From (21),

$$L_* = (\tilde{w}/\beta Ae^{\epsilon})^{-1/(1-\beta)}, \tag{22}$$

$$Y_* = Ae^{\epsilon+\lambda}L_*^{\beta} = (Ae^{\epsilon})^{1/(1-\beta)}\beta^{\beta/(1-\beta)}\tilde{w}^{-\beta/(1-\beta)}e^{\lambda}, \tag{23}$$

where L_* is the "true" input, Y_* is the "true" output. The firm's "expected" output is given by (23) with $\lambda = 0$. The profit which the firm would receive from the "true" input–output combination if $\tilde{w}$ were the "true" relative price is

$$\tilde{\Pi} = Y_* - \tilde{w}L_* = \{Ae^{\epsilon}(\beta/\tilde{w})^{\beta}\}^{1/(1-\beta)}(e^{\lambda} - \beta). \tag{24}$$

"Expected" profit for the firm is given by (24) with $\lambda = 0$. Finally, the profit the firm actually receives from the "true" input–output combination with the true relative price $w_* = \tilde{w}e^{-\xi}$ is

$$\Pi_* = Y_* - w_* L_* = \{Ae^{\epsilon}(\beta/\tilde{w})^{\beta}\}^{1/(1-\beta)}(e^{\lambda} - \beta e^{-\xi}). \tag{25}$$

As a consequence of the forementioned hypotheses, true input, output, and profit satisfy (22), (23), and (25).

[Note: Alternative specifications of firm behavior are: (1) non-competitive behavior rules (even in the face of competitive markets), (2) objectives other than profit maximization (e.g., sales maximization, managerial tastes), (3) alternative models of expectation formation, particularly where the firm has some prior beliefs on the likelihood of various λ and ξ, and (4) treatment of risk aversion and a "utility" function of profits.]

7.4. *Observed Data*

7.4.1. *Relation of Observed and "True" Series*

Maintained Hypothesis 6. The econometrician observes the "true" relative wage, labor input, and output with error (but without systematic bias); specifically $w = w_* e^{\tau}$, $Y = Y_* e^{\eta}$, and $L = L_* e^{\nu}$, where w, Y, and L are the observed quantities and τ, η, ν are random measurement errors with $E(\tau) = E(\eta) = E(\nu) = 0$.

7.4.2. *Relation Between Observations*

Maintained Hypothesis 7. Errors are statistically independent in different firms.

[Note: an alternative specification might be: (1) that ϵ follows some geographical pattern and therefore is not distributed independently over firms, (2) that the non-systematic error λ is correlated between firms, or (3) that the error in forecasting output price ξ is correlated between firms because of common output demand fluctuations.]

Maintained Hypothesis 8. Errors are homoscedastic; i.e., $\epsilon, \lambda, \xi, \eta, \nu$ have variances which do not vary across firms.

It will be necessary to make several further technical specifications in order to reach conclusions on the properties of estimators. These will be introduced as they are needed.

Taking equations (19), (22), (23), (25), plus the definitions $\tilde{w} = we^{\xi-\tau}$, $Y = Y_* e^{\eta}$, and $L = L_* e^{\nu}$, we can summarize the relations holding among the observed variables,

$$Y = AL^{\beta} e^{\epsilon+\lambda+\eta-\beta\nu}, \tag{26}$$

$$L = (we^{\xi-\tau-\epsilon}/\beta A)^{-1/(1-\beta)} e^{\nu}, \tag{27}$$

$$Y = (Ae^{\epsilon})^{1/(1-\beta)} (we^{\xi-\tau})^{-\beta/(1-\beta)} e^{\lambda+\eta} \beta^{\beta/(1-\beta)}, \tag{28}$$

$$\Pi \equiv Y - wL = (Ae^{\epsilon})^{1/(1-\beta)} (we^{\xi-\tau}/\beta)^{-\beta/(1-\beta)} \{e^{\lambda+\eta} - \beta e^{\nu-\xi+\tau}\}. \tag{29}$$

Taking logs, (26)–(29) become

$$y = \alpha + \beta l + \overbrace{\{\epsilon + \lambda + \eta - \beta\nu\}}^{u_1} = \alpha + \beta l + u_1, \tag{30}$$

$$l = \delta - \frac{1}{1-\beta}\,\omega + \overbrace{\left\{\frac{\epsilon + \tau - \xi}{1-\beta} + \nu\right\}}^{u_2} = \delta - \frac{1}{1-\beta}\,\omega + u_2, \tag{31}$$

$$y = \gamma - \frac{\beta}{1-\beta}\,\omega + \overbrace{\left\{\frac{\epsilon + \beta\tau - \beta\xi}{1-\beta} + \lambda + \eta\right\}}^{u_3} = \gamma - \frac{\beta}{1-\beta}\,\omega + u_3 \tag{32}$$

$$\begin{aligned}\pi &\doteq \theta - \frac{\beta}{1-\beta}\,\omega + \overbrace{\left\{\frac{\epsilon + \lambda + \eta - \beta\nu}{1-\beta} - \frac{\beta}{(1-\beta)^2}\frac{\tau^2}{2} - \frac{\beta(1+\beta)}{(1-\beta)^3}\frac{\tau^3}{6}\right\}}^{u_4} \\ &= \theta - \frac{\beta}{1-\beta}\,\omega + u_4,\end{aligned} \tag{33}$$

where $y = \log Y$, $l = \log L$, $\pi = \log \Pi$, $\omega = \log w$, $\alpha = \log A$, $\delta = (1/(1-\beta)) \log(\beta A)$, $\gamma = (1/(1-\beta)) \log A + (\beta/(1-\beta)) \log \beta$, and $\theta = \gamma + \log(1-\beta)$, and where we have approximated the non-linear error in (33) by a Taylor's expansion,

$$\log\left\{\frac{e^{\lambda+\eta} - \beta e^{\nu-\xi+\tau}}{1-\beta}\right\} \cong \frac{1}{1-\beta}\{\lambda + \eta - \beta\nu + \beta\xi - \beta\tau\} - \frac{\beta}{(1-\beta)^2}\frac{\tau^2}{2} - \frac{\beta(1+\beta)}{(1-\beta)^3}\frac{\tau^3}{6} + \mathcal{O}(\lambda^2, \eta^2, \nu^2, \xi^2, \tau^4).$$

The system of equations then contains the production function (30) and the first-order conditions (31). Since we deal with one input only equation (31) is also a reduced-form equation and as such it is referred to as the labor demand equation. The second reduced-form equation (32), is the supply function. Equation (33) is an approximation of the profit function.

The direct estimation of the production elasticities from the first-order conditions, termed by Klein (1953) the factor share estimate, is derived from the following relation:

$$\log \frac{wL}{Y} = \omega + l - y = \log \beta + \overbrace{(\tau - \xi + \nu - \lambda - \eta)}^{u_5}. \quad (34)$$

The first-order conditions are widely used in estimating the parameters of more complex production functions. In this case the right-hand side of

TABLE 3

			Error structure			
No.	Name	Expression	ϵ	$(\lambda + \eta)$	ν	$(\tau - \xi)$
(30)	Production	$y = \alpha + \beta l + u_1$	1	1	$-\beta$	0
(31)	First-order or labor demand	$l = \delta - c\omega + u_2$	c	0	1	c
(32)	Supply	$y = \gamma - c\beta\omega + u_3$	c	1	0	$c\beta$
(33)	Profit[a]–approximation	$\pi \doteq \theta - c\beta\omega + u_4$	c	c	$-c\beta$	0
(34)	Factor share	$\omega + l - y = \log \beta + u_5$	0	-1	1	1
(35)	First-order transformed	$\omega = (1/c)(\delta - l) + u_6$	1	0	$1/c$	1

where

$c = 1/(1-\beta)$
ϵ = firm effect in the production function
λ = non-systematic error in the production function
η = measurement error of output
ν = measurement error of input
τ = measurement error of real wages
ξ = forecasting error of real wages

Since λ and η have the same coefficients in the various equations, the two are combined here; τ and $-\xi$ are similarly combined. The question of identifying these components is of no major concern to us and will therefore be disregarded.

[a]Add $c_2\tau^2 + c_3\tau^3$ to error term – see discussion above for details.

the equation consists of either quantities or prices. It is therefore of interest to compare these two alternatives in the present simple formulation. Such a comparison is also useful when wage is measured with an error. If such an error is more serious than the error of measuring inputs, then it may be desirable to estimate β not from (31), but rather from

$$\omega = (1-\beta)\delta - (1-\beta)l + \overbrace{\{\epsilon + \tau - \xi + (1-\beta)v\}}^{u_6}. \tag{35}$$

We refer to this as the transformed or inverted first-order condition.

In what follows it might be convenient to refer to the summary table, Table 3. The panel on the right-hand side titled "Error structure" should be read as follows: $u_1 = \epsilon + (\lambda + \eta) - \beta\nu + 0(\tau - \xi)$, and similarly for the other terms.

In order to evaluate the estimators we have to further specify the moments of the random errors. It is reasonable to assume that most of the error components are independent. The analysis begins by allowing for some non-zero covariances, as described in:

Maintained Hypothesis 9. Let $(\epsilon,\lambda,\eta,\nu,\tau,\xi) = (\cdot)$, then

$$E(\cdot) = \mathbf{0}$$
$$V(\cdot) = \begin{matrix} \sigma_\epsilon^2 & \sigma_{\epsilon\lambda} & 0 & 0 & 0 & \sigma_{\epsilon\xi} \\ & \sigma_\lambda^2 & 0 & 0 & 0 & 0 \\ & & \sigma_\eta^2 & 0 & 0 & 0 \\ & & & \sigma_\nu^2 & \sigma_{\nu\tau} & 0 \\ & & & & \sigma_\tau^2 & 0 \\ & & & & & \sigma_\xi^2 \end{matrix}$$

For some parts of the analysis it is also required that the first five moments of $(\cdot)$ exist.

Finally it is assumed that $0 \leq \text{plim}(\omega - \bar{\omega})^2 < \infty$.

The problems involved in estimation are related to the fact that the right-hand-side variables in (30)–(32) are stochastic and correlated with residuals, implying that estimates obtained by least squares (LS) will be inconsistent. We look at this more closely, and ask under what stochastic structures each estimate will be consistent. Define the sample moment about means

$$s_{yl} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(l_i - \bar{l}),$$

with corresponding notation for other moments. Table 4 presents the various estimators to be considered and their errors.

TABLE 4
Alternative estimators of β.

Estimator	Source	Error: $b_{3j}-\beta$
$b_{30}=\frac{s_{yl}}{s_{ll}}$	PF direct – by LS	$\frac{s_{l1}}{s_{ll}}$
$b_{31}=\frac{s_{\omega\omega}+s_{l\omega}}{s_{l\omega}}$	RF – ILS from labor demand	$\frac{-(1-\beta)^2 s_{\omega 2}}{s_{\omega\omega}-(1-\beta)s_{\omega 2}}$
$b_{32}=\frac{s_{y\omega}}{s_{y\omega}-s_{\omega\omega}}$	RF – ILS from product supply	$\frac{-(1-\beta)^2 s_{\omega 3}}{s_{\omega\omega}-(1-\beta)s_{\omega 3}}$
$b_{33}=\frac{s_{\pi\omega}}{s_{\pi\omega}-s_{\omega\omega}}$	Dual – ILS from approximated profit function	$\frac{-(1-\beta)^2 s_{\omega 4}}{s_{\omega\omega}-(1-\beta)s_{\omega 4}}$
$b_{34}=\exp(\bar{\omega}+\bar{l}-\bar{y})$	Factor share	$\exp(\bar{u}_5)$
$b_{35}=\frac{s_{\omega l}+s_{ll}}{s_{ll}}$	FOC – ω as "dependent variable"	$\frac{s_{l6}}{s_{ll}}$
$b_{36}=\frac{s_{y\omega}}{s_{l\omega}}$	PF direct – ω as instrumental variable	$\frac{(1-\beta)s_{\omega 1}}{(1-\beta)s_{\omega 2}-s_{\omega\omega}}$

LS = Least Squares
ILS = Indirect LS
PF = Production Function
RF = Reduced Form
FOC = First-Order Condition
$s_{\omega j}=\text{cov}(\omega u_j)$
$\bar{y}$ = sample arithmetic average of y, and similarly for other variables

Under Hypothesis 9, the probability limits of the errors of the various estimators are

$$\text{plim}(b_{30}-\beta)=\left(\frac{\sigma_\epsilon^2+\sigma_{\epsilon\lambda}-\sigma_{\epsilon\xi}}{1-\beta}-\beta\sigma_\nu^2\right)\Big/\sigma_l^2,$$

$$\text{plim}(b_{31}-\beta)=\left(\frac{\sigma_\tau^2}{1-\beta}+\sigma_{\nu\tau}\right)\Big/\sigma_\omega^2,$$

$$\text{plim}(b_{32}-\beta)=\frac{\beta\sigma_\tau^2}{1-\beta}\Big/\sigma_\omega^2,$$

$$\text{plim}(b_{33}-\beta)=\left(\frac{\beta^2}{2(1-\beta)^3}E\tau^3+\frac{\beta^2(1+\beta)}{6(1-\beta)^4}E\tau^4+\cdots\right)\Big/\sigma_\omega^2,$$

$$\text{plim}(b_{34}/\beta)=1,$$

$$\text{plim}(b_{35} - \beta) = \frac{\sigma_\epsilon^2 - 2\sigma_{\epsilon\xi} + \sigma_\xi^2}{1 - \beta} + \sigma_{\nu\tau} + (1 - \beta)\sigma_\nu^2,$$

$$\text{plim}(b_{36} - \beta) = (1 - \beta)\beta\sigma_{\nu\tau}/\sigma_\omega^2.$$

Using these expressions we can now review the relative merits of the various estimators.

(1) Direct estimation of the production function by LS: b_{30} is a consistent estimator if the labor input is measured without error ($\sigma_\nu^2 = 0$) and there are no firm effects ($\sigma_\epsilon^2 = 0$). The specific or firm effect was first introduced in the classical paper by Marschak and Andrews (1944). As we have seen, this effect is taken into account in maximization, and consequently the input cannot be considered as exogeneous. Solutions to this problem in the framework of direct estimation of the PF were discussed by Hoch (1958, 1962), Mundlak (1961, 1963), and Mundlak and Hoch (1965). In essence, there are two basic solutions. These can be stated in a more general form that will also apply to a multiple-input technology: (i) Impose enough restrictions on the covariances of the various error terms so as to identify the production elasticities. The resultant estimate can be considered as an instrumental variable estimator where $y - l$ is the instrument; $y - l$ varies with ω and the errors which appear in line (34) of Table 3. That is, $y - l$ is obtained as a difference of the reduced-form equations, (31) minus (32). However, if λ, ν, and η are present, then the estimate obtained with $y - l$ as an instrumental variable will not be consistent. In the terminology of Mundlak and Hoch (1965), this estimator overcomes the transmitted error (ϵ) but it is susceptible to the non-transmitted errors (λ, η, and ν in our case). At this point, we digress briefly to the case of more than one input, and consider a suggestion made by Cavallo (1976). For each of the inputs there will be a first-order equation of the form of (31). Consequently, a difference between two inputs provides an instrumental variable whose systematic part consists of the price difference of the two variables and whose error part consists of the difference in measurement and forecasting errors of the two inputs, a term which does not appear in the error of the production function. Thus, the performance of such instrumental variables is independent of the relative strength of the transmitted versus non-transmitted errors. If there are k inputs, there are $k - 1$ such instrumental variables and there will be a need for one more such variable. It should be noted that if there are serious errors in the measurement of inputs, such instrumental variables

will not yield a consistent estimator. (ii) The foregoing discussion is pertinent primarily to a strictly cross-section analysis. When repeated observations in time are available for each firm, covariance analysis can be applied to eliminate the firm effect, so that for the within firm variations, $\sigma_\epsilon^2 = 0$. Such an estimator is susceptible only to error in measurement of the input. Further, if the measurement error is also subject to a firm effect, then covariance analysis will solve this problem as well.

(2) Reduced-form estimates: The reduced-form estimates, b_{31} and b_{32}, require some wage rate variations among firms. Such estimates are consistent if the wage rates are measured without error. If this condition is not met, the degree of inconsistency depends on β and on the ratio $\sigma_\tau^2/\sigma_\omega^2$.

(3) Profit function: Since this equation approximates an identity in the reduced-form equations, the estimator, b_{33}, requires similar but somewhat weaker conditions for consistency than the reduced form estimators, namely that moments of third- and higher-orders be zero.

(4) Direct estimation of the production function with wages as instrumental variables: Under the present assumptions this estimator is consistent, provided measurement errors in wages and inputs are independent.

(5) Factor share: b_{34} provides a consistent estimator of β, and if the errors are log normally distributed, it is possible also to adjust the estimator so as to obtain a minimum variance unbiased estimate [Bradu and Mundlak (1970)].

(6) First-order conditions with price as a dependent variable: The consistency of this estimator requires that the input be measured without error, $\sigma_\nu^2 = 0$, that there are no firm effects in the production function ($\sigma_\epsilon^2 = 0$) and no error in optimization and price forecasting ($\sigma_\xi^2 = 0$). These are strong assumptions indeed.

Under the assumption of profit maximization the factor share estimator seems to be the simplest and easiest to compute and at the same time its consistency depends on fewer assumptions than some of the alternative estimators. This result is a direct consequence of the simplifying assumptions that the error components u_5 have zero expectations. At the purely technical level, the importance of this assumption stems from the fact that there is no other coefficient beside β to absorb deviations from such an assumption.

In the discussion of the specification of the model we have cast doubt on the general validity of the assumption $E(\xi)=0$. Indeed, the early studies of Cobb and Douglas were largely motivated by the desire to test the hypothesis that factors are paid according to their marginal productivity. It is therefore inadequate to impose such a hypothesis as a constraint as is done in the factor share estimator [Mundlak (1963)]. This point holds equally well for more complex functional forms, the coefficients of which are estimated from the first-order conditions.

The relaxation of the assumption $E(\xi)=0$ also has an effect on the reduced-form estimators. In evaluating the probability limit of these estimators there will be another term, $\text{plim}(\sum \omega^* \xi/n)$, and this term need not vanish even if the two components ω^* and ξ are independent. The only estimators that are not affected by this term are the direct estimates of the production function as discussed under (1) above or by using prices as instrumental variables.

We can now summarize the discussion by listing the consequences of the various error components on the alternative estimators:

(1) Firm effect in the production function (ϵ) results in inconsistency of the direct LS fit of the production function and of the transformed first-order condition estimator, b_{35}.

(2) Non-systematic error in the production function (λ) does not lead to inconsistency.

(3) Measurement errors, as is well-known, lead to inconsistency only if they occur in the independent variables of the regression. Measurement errors of the real wage, $\sigma_\tau^2 \neq 0$, lead to inconsistency of the reduced-form estimates, b_{31}, b_{32}, and possibly b_{33}. If this error is serious, it can be avoided by estimating the transformed first-order condition, b_{35}. The latter is sensitive to measurement error in input, $\sigma_\nu^2 \neq 0$, as is also the case with direct LS fit of the production function.

(4) Non-systematic errors of optimization effect only the transformed first-order condition equation.

(5) Systematic errors of optimization, which also include errors in wage forecasting, result in inconsistency of the reduced-form equations as well as the factor share estimate.

The estimator which seems to be most robust with respect to avoiding bias due to the various stochastic components is the direct estimate of the production function with real wages as an instrumental variable.

Consider now the case where there is variation in l and ω, so all the estimators b_{30} to b_{36} are defined. Suppose labor inputs and wages are measured without error ($\sigma_\nu^2 = \sigma_\tau^2 = 0$). Which of the estimators is "best"? A partial answer for large samples can be obtained by comparing asymptotic variances. A tiresome computation for the case of normally distributed errors yields

$$\begin{aligned}\text{plim}\, n(b_{30} - \text{plim}\, b_{30})^2 = {} & (1-\beta)^2\{\sigma_\epsilon^2 + \sigma_\lambda^2 + \sigma_\eta^2 + 2\sigma_{\epsilon\eta})\sigma_\omega^2/(\sigma_\omega^2 + \sigma_\epsilon^2 \\ & + \sigma_\xi^2 - 2\sigma_{\epsilon\xi})^2\} + (1-\beta)^2\{(\sigma_\epsilon^2 - 2\sigma_{\epsilon\xi} \\ & + \sigma_\xi^2)(\sigma_\lambda^2 + \sigma_\eta^2) + (3\sigma_\epsilon^4 - 6\sigma_\epsilon^2\sigma_{\epsilon\xi} + \sigma_\xi^2\sigma_\epsilon^2 \\ & + 2\sigma_{\epsilon\xi}^2\}/(\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\xi^2 - 2\sigma_{\epsilon\xi})^2, \qquad (37)\end{aligned}$$

$$\text{plim}\, n(b_{31} - \beta)^2 = (1-\beta)^2\{\sigma_\epsilon^2 + \sigma_\xi^2 - 2\sigma_{\epsilon\xi}\}/\sigma_\omega^2,$$

$$\begin{aligned}\text{plim}\, n(b_{32} - \beta)^2 = {} & (1-\beta)^2\{\sigma_\epsilon^2 + \beta^2\sigma_\xi^2 - 2\beta\sigma_{\epsilon\xi} \\ & + (1-\beta)^2(\sigma_\lambda^2 + \sigma_\eta^2)\}/\sigma_\omega^2,\end{aligned}$$

$$\text{plim}\, n(b_{33} - \beta)^2 = (1-\beta)^2\{\sigma_\epsilon^2 + \sigma_\lambda^2 + \sigma_\eta^2 + 2\sigma_{\epsilon\lambda}\}/\sigma_\omega^2,$$

$$\text{plim}\, n(b_{34} - \beta)^2 = \beta^2\{\sigma_\xi^2 + \sigma_\lambda^2 + \sigma_\eta^2\},$$

$$\begin{aligned}\text{plim}\, n(b_{35} - \text{plim}\, b_{35})^2 = {} & (\sigma_\epsilon^2 + \sigma_\xi^2 - 2\sigma_{\epsilon\xi})\{\sigma_\omega^2 + 2(\sigma_\epsilon^2 + \\ & \sigma_\xi^2 - 2\sigma_{\epsilon\xi})\}/(1-\beta)^2\sigma_l^2,\end{aligned}$$

$$\text{plim}\, n(b_{36} - \beta)^2 = \frac{(1-\beta)^2(\sigma_\epsilon^2 + \sigma_\lambda^2 + 2\sigma_{\epsilon\lambda}) + 3\beta(1-\beta)\sigma_{\epsilon\xi} + 4\beta^2\sigma_\xi^2}{\sigma_\omega^2 + \sigma_\epsilon^2 + \sigma_\xi^2 - 2\sigma_{\epsilon\xi}}.$$

When $\sigma_\epsilon^2 = 0$, so that b_{30} is consistent, (37) reduces to

$$\text{plim}\, n(b_{30} - \beta)^2 = (1-\beta)^2(\sigma_\lambda^2 + \sigma_\eta^2)/(\sigma_\omega^2 + \sigma_\xi^2).$$

In this case, the relative efficiency of the estimators depends on relative variances. For example, if optimization errors (σ_ξ^2) are large, then b_{31} and b_{32} are undesirable and b_{30} will tend to be most efficient. If σ_ξ^2 is low, then b_{31} will tend to be most efficient and b_{32} will be more efficient than b_{33}. When σ_ω^2 is low relative to σ_ξ^2, b_{30} will be most efficient. For β near one, the estimators b_{30} to b_{33} will be relatively efficient, while β near zero will make b_{34} most efficient. When σ_ϵ^2 is large, b_{34} will tend to be most efficient. All these conclusions, it should be noted, are shown only for large samples. While it is dangerous to over-generalize from specialized small-sample results, there seems to be a tendency for direct ordinary least-squares estimators such as b_{30} to be the best estimators in small samples more often than one would guess from the extrapolation of asymptotic results; i.e., direct LS estimators seem to be somewhat more robust than their competitors in small samples. The exact small sample distributions of the estimators b_{31} to b_{33} can be derived for the example above, and are found to have tails that behave like Cauchy distributions. Consequently, the mean and variance of b_{31} to b_{33} are not defined in finite samples, and the probability of estimates of β which are far from the true value are rather large. Thus, the estimators in this example tend to confirm the generalization regarding the relative small sample robustness of direct ordinary least-squares estimators.

Before concluding the discussion it should be pointed out that we have made repeated reference to the instrumental variable estimator. Such an estimator overcomes difficulties caused by measurement errors and lack of independence between the explanatory variables and the error terms. In general there are several difficulties with the use of this method of which the user should be aware. First, instrumental variable estimates are not as efficient (have larger variance) as the direct LS estimator. This problem can be reduced by a proper selection of instrumental variables, which leads us to the second problem – that of finding such variables. Instrumental variables should be uncorrelated with the error terms in the equation and at the same time be correlated with the explanatory variables. The larger is the latter correlation (properly defined when there are more than one variable) – the smaller is the variance of the estimator. Third, in small samples, instrumental variables estimators usually have distributions with "fat" tails, tending to produce extreme values. Thus, one may buy consistency at the cost of a less accurate estimator in a small sample.

In the foregoing discussion we considered the use of the real wage as an instrumental variable. This generalizes to the use of real factor prices in the case of more than one input and the use of product price ratios in the case of a multiproduct production function. We have also mentioned the use of some linear combinations of the quantities as instrumental variables which can eliminate some of the errors. All these are variables which come from the model. It is also possible to use variables which come from outside the model [Berndt and Christensen (1973a)].

We have discussed in the context of the example above the difficulties encountered when there is insufficient variation in the independent variables in a direct or indirect estimation of production parameters. In the case of multiple inputs, this problem reappears as that of *multicollinearity*, or high correlation among the independent variables so that there is insufficient cross-variation to allocate with precision the contribution of separate variables to the determination of the dependent variable. This problem is particularly acute for time series analysis, and in functional forms where the independent variables appear as "substitutes". Considerable success in dealing with multicollinearity has been achieved in production applications by considering the production function as part of a complete economic model. In the example above, one may view (30)–(33) as equations in a simultaneous system. When non-redundant sets of equations are estimated jointly, they can provide more efficient estimates than any one equation considered above. The effectiveness of the analysis of complete systems to reduce multicollinearity and increase precision is most evident in estimation of general linear-in-parameters forms such as the Diewert or translog systems [e.g., Burgess (1975) and Woodland (1975)].

We can summarize our conclusions including the inferences that can be drawn from the example. The relative desirability of estimation of the production function, its dual profit or cost function, factor demand or supply equations, or their inverse first-order conditions, depends primarily on the stochastic structure of the data. In the general case, these equations together constitute a simultaneous system, and the most efficient estimators are obtained by estimation of the complete system. When the source of stochastic errors is confined to technological effects not observed by production units, then direct estimation of the production function is a good procedure, although multicollinearity will be a problem for many data sets. In principle, estimation of factor demand equations in which the independent variables are prices is a good procedure, being consistent in the presence of stochastic components which make direct estimation of the production function inconsistent.

However, if multicollinearity constitutes a problem in the direct estimation, it is likely to remain so in the estimation of the factor demand equations. This problem is overcome in part by estimation of the first-order conditions, which under the separability conditions frequently imposed in empirical analysis have fewer variables than the demand functions. The use of the first-order equations represents, like the direct estimation of the production function, a limited information approach which does not use all the constraints of the system.

It should be noted that several caveats apply in the use of dual profit or cost functions and their derivative demand and supply functions, as well as first-order conditions. First, the construction of these functions require maintained hypotheses on market environment and behavior which may not be necessary for direct estimation of the production function. Failure of one of these maintained hypotheses may result in a model which does not have the postulated structural relationship to the underlying technological parameters. For example, if markets are not competitive, or if firms fail to maximize profits, non-technological factors are introduced into the "as if" technology reconstructed under a competitive profit maximization assumption. Second, there may be insufficient variation in factor prices to allow accurate estimation of production parameters. Mundlak (1968a) has noted that variation in production quantities in many data sets is much greater than variation in prices. This is presumably due to random effects on technology, environment, or firm behavior. In some cases, this may mean that more accurate estimates can be obtained by direct estimation, even in the likelihood of an introduction of bias. On the other hand, McFadden (Chapter IV.1) has found in a data set on establishments substantial price variation at the plant level in inputs which have "national" markets, due to transportation costs, timing of purchase, volume of contracts, and local conditions. This suggests that indirect methods may be quite satisfactory when accurate establishment price data are available, but may perform poorly when more general market price indices are used.

Third, in the analysis of firms facing non-competitive markets, or of industry or macroeconomic production aggregates, prices are not exogenous. Valid estimation requires information of the remainder of the system, with simultaneous estimation; or the use of instrumental variables methods. As noted in the example, the small sample advantages of ordinary least-squares regression estimates over instrumental methods may suggest use of instrumental estimators only for large data sets.

8. Overview of Empirical Analysis

The empirical literature which utilizes Cobb–Douglas and CES production functions has been surveyed by Walters (1963), Nerlove (1967), and Bridge (1971). The outstanding example of the use of the Cobb–Douglas cost function is Nerlove's (1963) study of electricity supply. The Cobb–Douglas profit function has been used by Lau and Yotopoulos (1971) to analyze efficiency in Indian agriculture. An example of the use of a CES cost function can be found in Chapter IV.1 of this volume by McFadden. Recent estimates of Cobb–Douglas and CES production functions can be found in Griliches and Ringstad (1971).

In the past several years most of the empirical literature has been devoted to attempts to implement the flexible functional forms discussed in Section 4. Generalized Leontief cost functions have been estimated for Sweden by Parks (1971), for Canada by Woodland (1975), and for

Norway by Frenger in Chapter V.2 of this volume. Fuss, Chapter IV.4, estimated a two-stage nested variant of this function for the U.S. steam-electric generation industry. Translog production functions have been applied to U.S. manufacturing by Berndt and Christensen (1973a, 1974) and to aggregate U.S. activity by Christensen, Jorgenson, and Lau (1973) and Burgess (1975). Examples of the estimation of translog cost functions are papers by Burgess (1975), Denny and Pinto (Chapter V.1), Berndt and Wood (1975), and Fuss (1977a). Translog profit functions have been utilized by Christensen, Jorgenson, and Lau (1973), and Hudson and Jorgenson (1974). Finally, the quadratic profit function is the functional form used in Cowing's study of the regulatory constraint, Chapter IV.5 of this volume.

9. Conclusion

This chapter has stressed the importance of economic and statistical criteria for the choice of functional forms in the estimation of production relationships. We have pointed out that linear-in-parameters forms provide a flexible, general purpose approach to functional specification, and that the linear-in-parameters approach can be utilized to tailor functional forms to specific applications. However, we have also emphasized the use of non-linear functional forms in applications where economy and ease of interpretation of parameters is important, as in studies of elasticities of substitution. The critical role of separability as an economic assumption, and as a tool in the construction of functional forms, has been stressed. Finally, we have used a simple example to illustrate the implications of alternative sources of stochastic error for the choice of functional form and estimation method.

We emphasize in conclusion that the primary interest in specific functional forms lies in their empirical application, and that the choice of a functional form should be based on an integrated consideration of the economic problem and likely stochastic structure of the observed data.

Chapter II.2

THE GENERAL LINEAR PROFIT FUNCTION

Daniel McFADDEN*

University of California, Berkeley

1. Introduction

The classical competitive firm is assumed to face exogenously determined technological possibilities and choose variable inputs and outputs to maximize profits at exogenous competitive market prices. This behavior can be summarized in a *restricted profit function* specifying maximum profit as a function of the exogenous variables, market prices and parameters specifying technological possibilities. By varying the interpretation of commodities and parameters, one can formulate as special cases of this general model the problems of cost minimization, revenue maximization, intertemporal operation of the firm and operation of the firm under uncertainty. In Chapter I.1, the author has given a detailed discussion of properties and possible applications of the restricted profit function.

The practical advantage of formulating a model of the competitive firm in terms of a restricted profit function lies in the computationally simple relationship between this function and the derived demand and supply functions which form the basis for comparative analysis or econometric estimation; namely, the net supply functions can be computed as partial derivatives of the restricted profit function with respect to market prices. Judicious choice of a functional form for the

*The concepts of a linear-in-parameters cost function and the second-order approximation property are due to E. Diewert, and I am indebted to him and to M. Fuss for many useful discussions. This research was supported by NSF Grant GS-2345.

restricted profit function can yield net supply systems which embody economic phenomena of interest and which are convenient for statistical analysis. This chapter introduces a class of *general linear profit functions* which should provide useful functional forms from the standpoint of both these criteria. These functional forms have the properties:

(a) They are linear in the underlying parameters of the production process, making it possible to estimate the net supply system by multivariate linear regression techniques and formulate economic hypotheses as linear restrictions on this system.
(b) They satisfy globally (i.e., for all positive market prices) the criteria for a function to be the restricted profit function associated with some technology.
(c) They can approximate a large class of restricted profit functions (e.g., those satisfying a gross substitutes property[1]) up to the second order at any specified argument, thus agreeing on net supply quantities and price elasticities at this argument.

An additional advantage of these functional forms is that aggregation over firms with common technologies "carries past" the unknown parameters, permitting a simple theory of aggregation and estimation from aggregate data. The general linear profit function is an extension of the generalized Leontief cost function introduced by Diewert (1971), and can reduce to his cost function in the case of cost minimization for fixed output.

2. The Basic Model

Consider a firm facing competitive markets in N commodities, indexed $n = 1,\ldots,N$, with a commodity price vector $\mathbf{p} = (p_1,\ldots,p_N)$. A production plan for the firm is a vector $\mathbf{x} = (x_1,\ldots,x_N)$ with x_n interpreted as the net supply (or, for compactness, netput) of commodity n, negative if the commodity is an input and positive if it is an output. The profit associated with a production plan $\mathbf{x}$ is $\pi = \mathbf{p}\cdot\mathbf{x} = p_1x_1 + \cdots + p_nx_n$. The technological possibilities of the firm can be described by a set $\mathbf{T}$ of possible production plans. This set will in general depend on variables

[1]A technology has the gross substitutes property if the optimal net supply of each commodity is non-increasing in the price of every other commodity.

exogenous to the firm, as for example the state of technical progress, fixed outputs in the case of cost minimization, and fixed capital inputs in the case of short-run profit maximization. To simplify notation, we leave to the reader the task of introducing this dependence explicitly in the formulae below.

Define the production possibility set $\mathbf{T}$ to be *regular* if it is non-empty and closed and satisfies the free disposal property that $\mathbf{x} \in \mathbf{T}$ and $\mathbf{x}' \leqq \mathbf{x}$ implies $\mathbf{x}' \in \mathbf{T}$. Define $\mathbf{T}$ to be *asymptotically irreversible* (or semi-bounded) if there is a bound on the vectors of production plans $\mathbf{x}^0,\mathbf{x}^1,\ldots,\mathbf{x}^n \in \mathbf{T}$ satisfying $\sum_{i=0}^{N} \mathbf{x}^i = 0$. This condition excludes the possibility of a "perpetual motion" production process of unbounded "amplitude", and will hold if there are some non-producible commodities which are essential inputs to production.

The *restricted profit function* of the firm with technology $\mathbf{T}$ is

$$\pi = \Pi(\mathbf{p}) = \sup_{\mathbf{x}\in\mathbf{T}} \mathbf{p}\cdot\mathbf{x}, \tag{1}$$

and gives the least upper bound (possibly $+\infty$) on the level of profits attainable at price vector $\mathbf{p}$. Let $(\mathbf{dom}\,\Pi)$ denote the set of price vectors for which $\Pi(\mathbf{p})$ is finite.

An extended real-valued function $Q:\mathbf{E}^n \to [-\infty,+\infty]$ is said to be of *type* **RP** if it satisfies

(1) the set $(\mathbf{dom}\,Q)$ on which Q is finite is a convex cone with a non-empty interior which is contained in the non-negative orthant of $\mathbf{E}^n$; and
(2) Q is a convex conical closed[2] function on $(\mathbf{dom}\,Q)$.

A basic duality between production possibility sets and restricted profit functions is established in the following theorem, proved in Chapter I.1, Lemmas 11, 23.

Theorem 1. If $\mathbf{T}$ is a regular asymptotically irreversible production possibility set, then the restricted profit function Π defined by (1) is of type **RP**. Alternatively, if Π is a function of type **RP**, then the set

$$\mathbf{T}^* = \{\mathbf{x}\in\mathbf{E}^N \mid \mathbf{p}\cdot\mathbf{x} \leqq \Pi(\mathbf{p}) \quad \text{for} \quad \mathbf{p}\in\mathbf{E}^N\} \tag{2}$$

is a regular asymptotically irreversible convex production possibility

[2] A function Q is *convex* if $Q(\mathbf{p})$, $Q(\mathbf{p}') < +\infty$, $0 < \theta < 1$ implies $Q(\theta\mathbf{p} + (1-\theta)\mathbf{p}') \leqq \theta Q(\mathbf{p}) + (1-\theta)Q(\mathbf{p}')$; *conical* if $Q(\lambda\mathbf{p}) = \lambda Q(\mathbf{p})$ for $\lambda > 0$; and *closed* if the set $(\mathbf{epi}\,Q) = \{(\mathbf{p},q) \mid q \geqq Q(\mathbf{p})\}$ is closed.

set. In particular, if the function Π in (2) is the restricted profit function of a regular asymptotically irreversible production possibility set **T**, then **T*** is the closed convex hull of **T**. The mappings (1) and (2) are mutually inverse between the family of regular asymptotically irreversible convex production possibility sets and the family of functions of type **RP**; e.g., applying the mapping (2) to a function Π of type **RP** and then applying the mapping (1) to the resulting set **T*** returns the function Π.

A second basic property of the restricted profit function is the *derivative property*, proved in Chapter I.1, Lemmas 17–19.

Theorem 2. Consider a function Π of type **RP**. Π is differentiable at $\mathbf{p}'$ in the interior of **dom** Π if and only if in the technology **T*** given by (2), there exists a unique vector $\mathbf{x}' \in \mathbf{T}^*$ at which $\mathbf{p}' \cdot \mathbf{x}$ is maximized on **T***, in which case $\Pi_{\mathbf{p}}(\mathbf{p}') = \mathbf{x}'$.

In analyzing the general linear profit function below, we shall use on a function $Q\colon \mathbf{E}^n \to [-\infty,+\infty]$ of type **RP** the condition **C2** that Q be twice continuously differentiable with a Hessian of rank $n-1$ on the interior of (**dom** Q), and the condition **FP** that (**dom** Q) contain the positive orthant. Condition **C2** implies the dual technology of Q given by equation (2) is strictly convex (as viewed from the positive orthant of $\mathbf{E}^n$) with a "specific curvature" which is bounded positive. Condition **FP** implies that as the scale of production becomes large, the set of possible activities in the dual technology shrinks to the set of disposal activities, i.e., the asymptotic cone of the dual technology is the non-positive orthant. These duality implications are discussed in detail in Chapter I.1, Lemma 12, Theorem 26.

3. General Linear Profit Functions

A function $\Pi(\mathbf{p};\boldsymbol{\alpha})$ which is linear in a vector of underlying parameters $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_n)$ can be written in the form

$$\Pi(\mathbf{p};\boldsymbol{\alpha}) = \sum_{m=1}^{M} \alpha_m Q^m(\mathbf{p}). \tag{3}$$

where Q^m is a numerical function. Further, we can always standardize the parameter specification so that $\boldsymbol{\alpha}$ is restricted to be non-negative (i.e., we can first write each bivalent parameter as the difference of its positive and negative parts, and then re-define the Q function associated with each negative parameter to absorb its sign). This convention will be imposed hereafter in discussion of equation (3) unless explicitly assumed otherwise.

If the function $\Pi(\mathbf{p};\boldsymbol{\alpha})$ in equation (3) is of type **RP** for all non-negative $\boldsymbol{\alpha}$, then clearly each function Q^m is of type **RP** and $\bigcap_{m=1}^{M}(\mathbf{dom}\ Q^m)$ has a non-empty interior.[3] Conversely, it is an elementary property of convex functions that if each function Q^m is of type **RP** and if $\bigcap_{m=1}^{M}(\mathbf{dom}\ Q^m)$ has a non-empty interior, then Π given by equation (3) for any non-negative vector $\boldsymbol{\alpha}$ is of type **RP**.

A function Π in equation (3) which is of type **RP** for all non-negative $\boldsymbol{\alpha}$ will be termed a *general linear profit form*. This form can be specialized for econometric purposes by choosing specific numerical functions Q^m. To aid computation and interpretation it is convenient to take each function Q^m to depend on a small subset of the commodity prices. If each Q^m depends on a single price, then it is linear and the resulting linear profit function in equation (3) is dual to a pure fixed coefficients Leontief technology. The next case with the Q^m depending on pairs of commodity prices yields a variety of useful functional forms corresponding to a fairly broad class of technologies. Rewrite equation (3) by indexing over pairs of commodity prices as

$$\Pi(\mathbf{p};\boldsymbol{\alpha}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{ij} p_i Q^{ij}(p_j/p_i), \tag{4}$$

where the α_{ij} are non-negative parameters for $i \neq j$ with $\alpha_{ij} = \alpha_{ji}$, the Q^{ij} are closed convex functions of a positive real variable (implying $p_i Q^{ij}(p_j/p_i)$ is of type **RP** and satisfies condition **FP**), and the diagonal parameters α_{ii} are unrestricted in sign. One case of equation (4) is a version of Diewert's (1971) generalized Leontief form,

$$\Pi(\mathbf{p};\boldsymbol{\alpha}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{ij}(-(p_i p_j)^{1/2}); \tag{5}$$

others can be obtained by substituting for Q^{ij} in equation (4) some combination of the standard numerical forms for convex functions of one variable given in the first column of Table 1. When the functions Q^{ij} in (4) are differentiable, Theorem 2 implies the existence of an optimal

[3]Taking $\alpha_m > 0$, $\alpha_l = 0$ for $l \neq m$ implies $\Pi(\mathbf{p};\boldsymbol{\alpha}) = Q^m(\mathbf{p})$, so that Q^m is of type **RP**. Taking $\boldsymbol{\alpha}$ strictly positive implies $\mathbf{dom}\ \Pi(\cdot;\boldsymbol{\alpha}) = \bigcap_{m=1}^{M}(\mathbf{dom}\ Q^m)$.

TABLE 1

Convex functions of a positive real variable (θ and δ are numerical parameters).

	$Q^{ij}(r)$	Q^{ij}_r	$Q^{ij} - rQ^{ij}_r$	Q^{ij}_{rr}	Dual technology $\mathbf{T}^{ij}$ consisting of the points $\mathbf{x} \in \mathbf{E}^N$ with $x_k = 0$ for $k \neq i,j$ and x_i, x_j below, plus points obtained by free disposal
(1)	r^θ $\theta > 1$ or $\theta < 0$	$\theta r^{\theta-1}$	$(1-\theta)r^\theta$	$\theta(\theta-1)r^{\theta-2}$	$\left(\frac{x_i}{1-\theta}\right)^{1-\theta}\left(\frac{x_j}{\theta}\right)^\theta = 1$ $\frac{x_i}{1-\theta} \geq 0, \frac{x_j}{\theta} \geq 0$
(2)	$-r^\theta$ $0 < \theta < 1$	$-\theta r^{\theta-1}$	$-(1-\theta)r^\theta$	$-\theta(\theta-1)r^{\theta-2}$	$\left(\frac{-x_i}{1-\theta}\right)^{1-\theta}\left(\frac{-x_j}{\theta}\right)^\theta = 1$ $x_i, x_j \leq 0$
(3)	$[1-\delta+\delta r^{1+\sigma}]^{1/(1+\sigma)}$ $\sigma > 0$ $0 < \delta < 1$	$\delta r^\sigma (Q^{ij})^{-\sigma}$	$(1-\delta)(Q^{ij})^{-\sigma}$	$\delta(1-\delta)\sigma r^{\sigma-1}(Q^{ij})^{-2\sigma-1}$	$[(1-\delta)^{-1/\sigma} x_i^{1+1/\sigma} + \delta^{1/\sigma} x_j^{1+1/\sigma}] = 1$ $x_i, x_j \geq 0$

(4)	$-[1-\delta+\delta r^{1-\sigma}]^{1/(1-\sigma)}$ $\sigma > 0,\ \sigma \neq 1$ $0 < \delta < 1$	$-\delta r^{-\sigma}(-Q^{ij})^{\sigma}$	$-(1-\delta)(-Q^{ij})^{\sigma}$	$-\delta(1-\delta)\sigma r^{-\sigma-1}(-Q^{ij})^{2\sigma-1}$	$[(1-\delta)^{1/\sigma}(-x_i)^{1-1/\sigma}+\delta^{1/\sigma}(-x_j)^{1-1/\sigma}]=1$ $x_i, x_j \leq 0$
(5)	$e^{\theta r}$ $\theta \neq 0$	$\theta e^{\theta r}$	$(1-r\theta)e^{\theta r}$	$\theta^2 e^{\theta r}$	$x_i = \left(\frac{x_j}{\theta}\right)\left[1-\log\left(\frac{x_j}{\theta}\right)\right]$ $\frac{x_j}{\theta} > 0$
(6)	$-\log(\theta+r)$ $\theta \geq 0$	$-(\theta+r)^{-1}$	$-\log(\theta+r)+\frac{r}{\theta+r}$	$(\theta+r)^{-2}$	$x_i = \log(-x_j)+(1+\theta x_j)$ $-\theta^{-1} \leq x_j < 0$
(7)	$-\frac{r}{\theta+r}$ $\theta > 0$	$-\frac{\theta}{(\theta+r)^2}$	$-\frac{r^2}{(\theta+r)^2}$	$\frac{2\theta}{(\theta+r)^3}$	$x_j = \frac{x_i}{\theta}(-1+(-x_i)^{-1/2})^2$ $-1 \leq x_i \leq 0$
(8)	$\left(r+\frac{\theta}{r}\right)$ $\theta > 0$	$1-\frac{\theta}{r^2}$	$2\frac{\theta}{r}$	$\frac{\theta}{r^3}$	$x_j = 1-\frac{x_i^2}{4\theta}$ $x_i \geq 0$

production plan $\hat{\mathbf{x}}(\mathbf{p})$ for positive p satisfying

$$\hat{x}_k(\mathbf{p}) = \frac{\partial \Pi}{\partial p_k} = \sum_{j=1}^{N} \alpha_{kj} \left[Q^{kj}\left(\frac{p_j}{p_k}\right) - \frac{p_j}{p_k} Q_r^{kj}\left(\frac{p_j}{p_k}\right) + Q_r^{jk}\left(\frac{p_k}{p_j}\right) \right], \tag{6}$$

and

$$\frac{\partial \hat{x}_k(\mathbf{p})}{\partial p_l} = \frac{\partial^2 \Pi}{\partial p_k \partial p_l} = -\alpha_{kl} \left[\frac{p_l}{p_k^2} Q_{rr}^{kl}\left(\frac{p_l}{p_k}\right) + \frac{p_k}{p_l^2} Q_{rr}^{lk}\left(\frac{p_k}{p_l}\right) \right], \tag{7}$$

for $k \neq l$, where Q_r^{ij} and Q_{rr}^{ij} denote the first and second derivatives, respectively, of the function Q^{ij}. Since the expressions in brackets in equations (6) and (7) are numerical functions, these formulae allow application of multivariate regression analysis to estimate the net supply system.

A technology is said to have the gross substitutes (**GS**) property if the optimal net supply of each commodity k is non-increasing in the price of every other commodity. This property corresponds to the "normal" case where all outputs are substitutes (e.g., the quantity of one falls when the price and quantity of a second rises), all inputs are non-regressive in the production of outputs (e.g., each input quantity rises when the price and quantity of an output rise), and all inputs are strong substitutes (e.g., an increase in the price of one input leads to substitution of a second input which is sufficient to offset the tendency of an input price increase to reduce output quantity, and thus input quantities[4]). When the restricted profit function Π of the technology has the differentiability property **C2**, the gross substitutes property can be defined as the condition $\partial^2 \Pi / \partial p_k \partial p_l \leqq 0$ for $k \neq l$. A profit function of two prices must satisfy **GS**, and a sum of functions satisfying **GS** must again have this property. Thus, the linear profit form (4) has property **GS**; this is also clear from the sign of the cross-price effects in equation (7).

Consider an arbitrary function $\Phi(\mathbf{p})$ of type **RP** with property **C2** at a price vector $\mathbf{p}^*$ in the interior of (**dom** Φ). A general linear profit form $\Pi(\mathbf{p};\boldsymbol{\alpha})$ from equation (3) is said to have the *second-order approximation property* to Φ at $\mathbf{p}^*$ if there exists a non-negative parameter vector α^* such that the first and second derivatives of Π and Φ agree at $\mathbf{p}^*$ [i.e., $\Pi(\mathbf{p}^*;\boldsymbol{\alpha}^*) = \Phi(\mathbf{p}^*)$, $\Pi_k(\mathbf{p}^*;\boldsymbol{\alpha}^*) = \Phi_k(\mathbf{p}^*)$, and $\Pi_{kl}(\mathbf{p}^*;\boldsymbol{\alpha}^*) = \Phi_{kl}(\mathbf{p}^*)$]. The following result establishes that the general linear profit form (4) is

[4]Consider for example a technology **T** satisfying $x_2, x_3 \leqq 0$ and $x_1 \leqq [(-x_2)^{1-1/\sigma} + (-x_3)^{1-1/\sigma}]^{\mu/(1-1/\sigma)}$ for $\sigma > 0$, $\sigma \neq 1$, and $0 < \mu < 1$ (i.e., a CES production function with an elasticity of substitution σ, homogeneous of degree μ). Its restricted profit function is $\Pi = (1-\mu)\mu^{\mu/(1-\mu)} p_1^{1/(1-\mu)} (p_2^{1-\sigma} + p_3^{1-\sigma})^{-(\mu/(1-\mu))(1/(1-\sigma))}$. Π has the property **GS** for $\sigma \geqq 1/(1-\mu)$.

robust in the sense that locally it can mimic the net supply system of any restricted profit function with the gross substitutes property.

Lemma 1. Consider a general linear profit function Π satisfying equation (4) such that Q^{ij} has property **C2** and Q^{ij}_{rr} positive. If $\Phi(\mathbf{p})$ is any function of type **RP**, $\mathbf{p}^*$ is a vector in the interior of **dom** Φ, and Φ satisfies conditions **C2** and **GS** at $\mathbf{p}^*$, then Π has the second-order approximation property to Φ at $\mathbf{p}^*$.

Proof: From equation (7), one can choose α^*_{kl} for $k \neq l$ such that $\Pi_{kl}(\mathbf{p}^*,\boldsymbol{\alpha}^*) = \Phi_{kl}(\mathbf{p}^*)$. Then from equation (6), one can choose α^*_{kk} such that $\Pi_k(\mathbf{p}^*,\boldsymbol{\alpha}^*) = \Phi_k(\mathbf{p}^*)$. Since both Φ and Π are conical, it follows that $\Pi_{kk}(\mathbf{p}^*,\boldsymbol{\alpha}^*) = \Phi_{kk}(\mathbf{p}^*)$ and $\Pi(\mathbf{p}^*,\boldsymbol{\alpha}^*) = \Phi(\mathbf{p}^*)$. Q.E.D.

Note that the general linear profit form (4) has $N(N+1)/2$ independent parameters. This equals the number of independent conditions which must be met to obtain the second-order approximation property. In this sense, (4) is a "parameter-efficient" form among those with the approximation property.

It is clear that linear profit forms with Q-functions of more than two prices can be introduced which need not have the **GS** property; one possible form for econometric purposes will be introduced later. However, the following result shows that it is fruitless to seek a linear profit form which has the second-order approximation property to each function of type **RP** and which is itself of type **RP** over its entire domain of definition.

Lemma 2. Given any linear profit form Π in equation (3) with specified M and Q^m, there exists a function Φ of type **RP** satisfying **C2** at $\mathbf{p}^*$ in the interior of (**dom** Φ) such that Π does not have the second-order approximation property to Φ at $\mathbf{p}^*$ for $N > 2$.

Proof: Let $\mathbf{H}^*$ denote the $N-1$ matrix of derivatives $\Phi_{ij}(\mathbf{p}^*)$ for $i,j = 2,\ldots,N$, and let $\mathbf{H}^m$ denote the corresponding matrix for Q^m. For the second-order approximation property to hold, $\mathbf{H}^*$ must lie in the convex cone spanned by the $\mathbf{H}^m$. Now $\mathbf{H}^*$ can be any positive semidefinite matrix [e.g., the function

$$\Phi(\mathbf{p}) = \frac{1}{2p_1} \sum_{i,j=2}^{N} p_i p_j H^*_{ij}$$

is of type **RP** and returns this matrix]. Representing an $(N-1)$-square symmetric matrix as a point in $\mathbf{E}^{N(N-1)/2}$, the cone of positive semidefinite matrices is not polyhedral (e.g., the 2×2 submatrix

$$\begin{bmatrix} 1 & \alpha \\ \alpha & \beta \end{bmatrix}$$

is positive semidefinite on the set $\beta \geqq \alpha^2$ bounded by a parabola). Hence, $\mathbf{H}^*$ can be chosen to lie in an extreme ray of the cone which does not contain an $\mathbf{H}^m$. Q.E.D.

In view of this result, we must either restrict the class of profit functions we wish to approximate by a linear profit form, or else relax the conditions we have imposed on the linear profit form. We next give a very general result of the first type. Unfortunately, the argument is not constructive and thus does not provide a way of generating linear forms for econometric purposes.

If a function Φ of type **RP** has property **C2** and the matrix $\mathbf{H}(\mathbf{p}) = (\Phi_{ij}(\mathbf{p}))$ for $i,j = 2,\dots,N$ is non-singular, then the dual technology at $\hat{\mathbf{x}}(\mathbf{p})$ is bounded by a surface which can be described by a twice continuously differentiable concave function $x_1 = f(\mathbf{x}_*)$ of $\mathbf{x}_* = (x_2,\dots,x_n)$ with a non-singular Hessian matrix $f_{\mathbf{xx}}(\hat{\mathbf{x}}_*(\mathbf{p})) = [-p_1\mathbf{H}(\mathbf{p})]^{-1}$ (Chapter I.1, Theorem 26). Define an index $\rho(\mathbf{p}) =$ (minimum root of $\mathbf{H}(\mathbf{p})$)/(maximum root of $\mathbf{H}(\mathbf{p})$). Then $\rho(\mathbf{p})$ is a measure of the "relative definiteness" of the matrix $\mathbf{H}$, or equivalently of the relative curvature of the surface of the technology.[5]

Lemma 3. Given $\epsilon > 0$, there exists a linear profit form Π in equation (3) with specified M and Q^m (depending in general on ϵ) such that if Φ is any function of type **RP** satisfying **C2** at $\mathbf{p}^*$ in the interior of (**dom** Φ) and if $\rho(\mathbf{p}^*) \geqq \epsilon$ for this function, then Π has the second-order approximation property to Φ at $\mathbf{p}^*$.

Proof: As in Lemma 2, denote symmetric $(N-1)$ matrices $\mathbf{H}$ as points in $\mathbf{E}^{N(N-1)/2}$. Define the set

$$\mathbf{A}_\theta = \left\{ \mathbf{H} \in \mathbf{E}^{N(N-1)/2} \,\middle|\, \sum_{i=2}^{N} H_{ii} = 1 \text{ and } \min_{\mathbf{q} \neq 0} (\mathbf{q}'\mathbf{H}\mathbf{q})/\mathbf{q}'\mathbf{q} \geqq \theta \right\}.$$

[5]Note that $\rho(\mathbf{p}) =$ (minimum root of $-f_{\mathbf{xx}}(\hat{\mathbf{x}}_*(\mathbf{p}))$)/(maximum root of $-f_{\mathbf{xx}}(\hat{\mathbf{x}}_*(\mathbf{p}))$), where $\mathbf{x}_* = (x_2,\dots,x_N)$. Let $\mathbf{G}$ be a matrix with $\mathbf{G}' = \mathbf{G}^{-1}$ such that $\mathbf{G}'\mathbf{H}\mathbf{G}$ is diagonal. Define a set of "composite" commodities $\mathbf{y}_* = \mathbf{G}'\mathbf{x}_*$ and corresponding prices $\mathbf{q}_* = \mathbf{G}'\mathbf{p}_*$. Then ρ is the ratio of the smallest to the largest own price effects of the composite commodities.

Then $\mathbf{A}_\theta$ is non-empty, closed, bounded, and convex for $0 \leqq \theta < 1/(N-1)$, and for $\theta > 0$, $\mathbf{A}_\theta$ is contained in the relative interior of $\mathbf{A}_0$. Hence, for $\theta = (\text{Min}(\epsilon,.5))/(N-1)$, there exists a convex polytope with vertices $\mathbf{H}^1,\ldots,\mathbf{H}^J$ which contains $\mathbf{A}_\theta$ and is contained in $\mathbf{A}_0$. Define $Q^m = (1/2p_1)\mathbf{p}_*'\mathbf{H}^m\mathbf{p}_*$ for $m = 1,\ldots,J$, where $\mathbf{p}_* = (p_2,\ldots,p_N)$; $Q^m = p_{m-J}$ for $m = J+1,\ldots,J+N$; and $Q^m = -p_{m-J-N}$ for $m = J+N+1,\ldots,J+2N$, with $M = J+2N$. Given Φ and $\mathbf{H}(\mathbf{p}^*)$, one has

$$\epsilon \leqq \rho(\mathbf{p}^*) \leqq [\text{Min}(\mathbf{q}'\mathbf{H}(\mathbf{p}^*)\mathbf{q}/\mathbf{q}'\mathbf{q})] \Big/ \left[\sum_{i=2}^{N} H_{ii}(\mathbf{p}^*)/(N-1)\right].$$

Hence, $\mathbf{H}(\mathbf{p}^*)$ is contained in the convex cone spanned by $\mathbf{H}^1,\ldots,\mathbf{H}^J$. Choosing $(\alpha_1,\ldots,\alpha_J)$ to equate the second partials of Π and Φ, and then choosing $(\alpha_{J+1},\ldots,\alpha_M)$ to equate the first partials, yields the desired conclusion. Q.E.D.

A function Φ of type **RP** with property **C2** at a vector $\mathbf{p}$ is said to have a *dominant own price effect* with numeraire commodity 1 if

$$p_i\Phi_{ii}(\mathbf{p}) \geqq \sum_{\substack{j=2 \\ j\neq i}}^{N} p_j|\Phi_{ij}(\mathbf{p})| \quad \text{for} \quad i = 2,\ldots,N.$$

From homogeneity, satisfaction of this condition requires that commodity 1 be a gross substitute for every other commodity. However, some patterns of gross complements among the remaining commodities are possible. If Φ has property **GS**, then it has the dominant own-price effect property. The next result provides a constructive proof that the class of profit functions with the dominant own-price effect property can be approximated to second order by a linear profit form:[6]

Lemma 4. Suppose good 1 is a specified numeraire commodity and $\mathbf{p}^*$ is a specified positive vector, and consider the linear profit form

$$\Pi(\mathbf{p};\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \sum_{i=1}^{N} \alpha_i p_i + \frac{1}{2p_1}\sum_{i=2}^{N}\sum_{j=2}^{N}\left[\beta_{ij}\left(\frac{p_i}{p_i^*}+\frac{p_j}{p_j^*}\right)^2 + \gamma_{ij}\left(\frac{p_i}{p_i^*}-\frac{p_j}{p_j^*}\right)^2\right], \tag{8}$$

with β_{ij} and γ_{ij} non-negative and symmetric in $ij, \gamma_{ii} = 0$, and $\beta_{ij}\cdot\gamma_{ij} =$

[6]The restricted profit function given in footnote 4 has the dominant own-price effect property if $\sigma \geqq 1/2(1-\mu)$, or if $\sigma < 1/2(1-\mu)$ and $(1-2\sigma(1-\mu))^{-1} \geqq (p_2/p_3)^{1-\sigma} \geqq 1-2\sigma(1-\mu)$. In particular, this property holds for all positive prices in the limiting case $\sigma = 1$.

0. If Φ is any function of type **RP** satisfying **C2** and the dominant own price effect property at $\mathbf{p}^*$ in the interior of (**dom** Φ), then (8) has the second-order approximation property to Φ at $\mathbf{p}^*$.

Proof: Differentiating (8), we obtain

$$\Pi_{kj} = 2(\beta_{kl} - \gamma_{kl})/p_1 p_k^* p_l^* \quad \text{for} \quad k \neq l, \tag{9}$$

$$\Pi_{kk} = 2\beta_{kk}/p_1 p_k^{*2} + 2\sum_{j=2}^{N}(\beta_{kj} + \gamma_{kj})/p_1 p_k^{*2}. \tag{10}$$

Condition (9) with $\Pi_{kl} = \Phi_{kl}(\mathbf{p}^*)$ and $\beta_{kl}\cdot\gamma_{kl} = 0$ determines β_{kl},γ_{kl} for $k \neq l$. Substituting these values in (10) yields

$$\begin{aligned} \beta_{kk}/p_1^* p_k^* &= p_k^* \Phi_{kk}(\mathbf{p}^*) - \frac{2}{p_1^* p_k^*}\sum_{j=2}^{N}(\beta_{kj} + \gamma_{kj}) \\ &= p_k^* \Phi_{kk}(\mathbf{p}^*) - \sum_{j=2}^{N} p_j^* |\Phi_{kj}(\mathbf{p}^*)| \geqq 0, \end{aligned}$$

by the dominant own price effect property. Choose the α_i to equate the first partials of Π and $\Phi(\mathbf{p}^*)$. This establishes the desired result. Q.E.D.

By weakening the requirement that a general linear profit form be finite for all positive prices, one can obtain the second-order approximation property to a broad class of restricted profit functions, as shown in the following result. This is in effect an unrestricted *local* approximation theorem.

Lemma 5. Consider the linear profit form (4) with specified twice continuously differentiable Q^{ij} having Q^{ij}_{rr} non-zero. Suppose Φ is of type **RP**, has property **C2** at $\mathbf{p}^*$ in the interior of **dom** Φ, and has $\mathbf{H}(\mathbf{p}) = (\Phi_{ij}(\mathbf{p}))$, $i,j = 2,\ldots,N$ non-singular at $\mathbf{p}^*$. Then there exists a parameter vector $\boldsymbol{\alpha}^*$ in (4), not necessarily non-negative, and a closed cone $\mathbf{P}^*$ containing $\mathbf{p}^*$ in its interior such that the function $\Pi^*(\mathbf{p})$, defined to equal $+\infty$ for $\mathbf{p} \notin \mathbf{P}^*$ and to equal (4) for $\boldsymbol{\alpha}^*$ and $\mathbf{p} \in \mathbf{P}^*$, is of type **RP** and has the second-order approximation property to Φ at $\mathbf{p}^*$.

Proof: From the proof of Lemma 1, $\boldsymbol{\alpha}^*$ can be chosen so that $\Pi(\mathbf{p};\boldsymbol{\alpha}^*)$ has the second-order approximation property to Φ at $\mathbf{p}^*$. Since $\mathbf{H}(\mathbf{p}^*)$ is positive definite, it follows from continuity that $\Pi(\mathbf{p};\boldsymbol{\alpha}^*)$ is convex for $\mathbf{p}$ in a convex neighborhood of $\mathbf{p}^*$; let $\mathbf{P}^*$ be the smallest closed convex cone containing this neighborhood. It is then immediate that $\Pi^*(\mathbf{p})$ is of type **RP**. Q.E.D.

Note that this result does not require the Q-functions to be convex. Thus for the purposes of this type of approximation, one could take equation (4) to have a wide variety of functional forms, such as (5) with unrestricted signs, or the Christensen–Jorgenson–Lau (1973) "translog" function. The difficulty with using Lemma 5 as a justification for choice of simple functional forms for econometric analysis without regard for global properties of the Q-functions is that the domain $\mathbf{P}^*$ cannot be determined *a priori* and may not include all observations. Use of a fitted equation (4) without restriction of domain may be inconsistent even locally with competitive profit maximization. An ex post consistency check for this inclusion is typically highly non-linear and computationally forbidding. However, Appendix A.4 by Lau has established feasible methods of testing the convexity of a function at each data point.

4. The Dual Technology of the General Linear Profit Function

The lemmas above giving second-order approximation properties of the general linear profit function to an arbitrary function of type **RP** can be interpreted dually as establishing that an arbitrary convex technology can be mimicked locally by the dual technology of the linear profit form. Beyond this conclusion, it is useful to establish some of the global properties of the dual linear technology.

We noted earlier that when the linear profit form is linear in prices, it is dual to a Leontief fixed coefficients technology. All the forms we consider yield this technology as a special case. More generally, we can from Theorem 2 express the dual technology of the general linear profit form (3) as a sum (see Chapter I.1, Table 5),

$$\mathbf{T} = \sum_{m=1}^{M} \alpha_m \mathbf{T}^m, \tag{11}$$

where

$$\mathbf{T}^m = \{\mathbf{x} | \mathbf{p} \cdot \mathbf{x} \leqq Q^m(\mathbf{p}) \text{ for all } \mathbf{p}\}. \tag{12}$$

When the Q-functions are simple forms, the sets $\mathbf{T}^m$ can often be characterized explicitly. The last column of Table 1 lists the dual technologies corresponding to a variety of two-price Q-functions. The three-price functional form (8) has the Q-function $(p_i/p_i^* + p_j/p_j^*)^2/2p_1$ dual to a technology $\mathbf{T}_+^{ij}$ with $x_k \leqq 0$ for $k \neq i,j$, and $\text{Max}\,[p_i^* x_i, p_j^* x_j] \leqq \sqrt{(-2x_1)}$; and the Q-function $(p_i/p_i^* - p_j/p_j^*)^2/2p_1$ dual to a technology T_-^{ij} with $x_k \leqq 0$ for $k \neq i,j$, $p_i^* x_i + p_j^* x_j \leqq 0$, and $\text{Max}[p_i^* x_i, p_j^* x_j] \leqq \sqrt{(-2x_1)}$. The structure (11) of the dual technology has a direct economic interpretation of non-jointness of the component technologies $\mathbf{T}^m$, implying that one can "decentralize" the optimization decisions in these components. In Chapter II.4, several examples are given in which this structure arises naturally for a multiple production unit firm.

The technological structure of equations (11) and (12) can also be characterized by "transformation" or gauge functions for the technologies. Let $\mathbf{e}$ denote a vector of ones, and for functions Q^m in the general linear profit form (3), define $\mathbf{x}^{*^m} = Q_{\mathbf{p}}^m(\mathbf{e}) - \mathbf{e}$, where $Q_{\mathbf{p}}^m$ is the vector of partial derivatives of Q^m, or more generally any optimal net supply vector for Q^m at the price vector $\mathbf{e}$. Then $\mathbf{x}^{*^m}$ is an interior point of $\mathbf{T}^m$. Define

$$F^m(\mathbf{x}) = \text{Inf}\{\lambda > 0 | \mathbf{p}\cdot\mathbf{x} \leqq \lambda(Q^m(\mathbf{p}) - \mathbf{p}\cdot\mathbf{x}^{*^m}) \text{ for all } \mathbf{p}\}. \tag{13}$$

Then $\mathbf{x} \in \mathbf{T}^m$ if and only if $F^m(\mathbf{x} - \mathbf{x}^{*m}) \leqq 1$ (Chapter I.1, Theorem 24). Assume $\boldsymbol{\alpha}$ strictly positive and define

$$F(\mathbf{x}) = \text{Inf}\left\{\underset{m}{\text{Max}}\, \frac{1}{\alpha_m} F^m(\mathbf{x}^m) \middle| \sum_{m=1}^{M} \alpha_m \mathbf{x}^m = \mathbf{x}\right\}. \tag{14}$$

Then $\mathbf{x} \in \mathbf{T}$ if and only if $F(\mathbf{x} - \sum_{m=1}^{M} \alpha_m \mathbf{x}^{*^m}) \leqq 1$ (Chapter I.1, Corollary 29). In the special case where the Q^m functions are separable, depending on disjoint subsets of commodity prices, the transformation function (14) has a corresponding separable structure. Linear profit forms chosen for econometric purposes usually contain this structure as a special case that can be tested as a linear hypothesis.

For the special two-price linear form (5) with $Q^{ij}(\mathbf{p}) = -(p_i p_j)^{1/2}$, an ingenious argument of Diewert (1971) provides an analytic characterization of the dual technology. Define $x_i^* = \max(0, -\alpha_{ii})$, let $\mathbf{A}$ denote the matrix of parameters α_{ij}, and for a vector $\mathbf{x}$ let $\hat{\mathbf{x}}$ denote a diagonal matrix constructed from the components of $\mathbf{x}$. For $\mathbf{x} \ll \mathbf{x}^*$, define $f(\mathbf{x})$ to be the reciprocal of the Frobenius root of the non-negative matrix

$$(\widehat{\mathbf{x}^* - \mathbf{x}})^{-1/2}(\hat{\mathbf{x}}^* + \mathbf{A})(\widehat{\mathbf{x}^* - \mathbf{x}})^{-1/2}.$$

Then f is a concave function, and $\mathbf{x} \in \mathbf{T}$ if and only if $f(\mathbf{x}) \geqq 1$.[7] Unfortunately this type construction does not seem to carry over to other two-price functional forms.

Finally, in the case of two commodities, the linear profit form (4) has a simple geometric interpretation: The function $Q^{12}(p_2/p_1)$ is dual to a "one input–one output" production function or an "isoquant" in the negative quadrant. The technology (11) is defined by shifting this surface by a scale factor and then shifting the axes. In Diewert's form (5) the surface is a translated rectangular hyperbola.

5. Applications of the Linear Profit Function

Our interest in the general linear profit function is based on its linear-in-parameters form, which allows estimation of the net supply system by linear regression methods. We now suggest several ways in which this structure can be exploited. The first comment concerns constant returns technologies.

(1) Use of the derivative property to obtain the net supply system under the assumption that the restricted profit function is differentiable with a Hessian of full rank implies that the dual technology is strictly convex. This condition is inconsistent with the assumption of a constant returns technology, and more fundamentally the specification of a constant returns technology and competitive profit maximization is insufficient to determine a net supply function describing the behavior of the firm. An obvious and reasonable way to obtain a definite net supply is to assume that the firm at each point in time treats as fixed some durable inputs which are essential to production and maximizes profit in variable goods, and then adjusts durable inputs over time subject to, say, equity constraints. The "per durable input unit" technology can then be strictly convex, and the formulae (3) and (6) specify a "per unit" net

[7]The argument is based on equation (2), which implies $\mathbf{x} \in \mathbf{T}$ if and only if

$$f(\mathbf{x})^{-1} = \sup\left\{\left(\sum_{i,j} p_i^{1/2} p_j^{1/2} a_{ij} + \mathbf{p}\cdot\mathbf{x}^*\right) \Big/ \sum_i p_i(x_i^* - x_i)\right\} \leqq 1.$$

Defining $q_i = p_i^{1/2}(x_i^* - x_i)^{-1/2}$, and $\hat{\mathbf{x}} = \mathrm{diag}(x_i)$, this formula becomes

$$f(\mathbf{x})^{-1} = \sup\{\mathbf{q}'(\widehat{\mathbf{x}^* - \mathbf{x}})^{-1/2}(\hat{\mathbf{x}}^* + \mathbf{A})(\widehat{\mathbf{x}^* - \mathbf{x}})^{-1/2}\mathbf{q}/\mathbf{q}'\mathbf{q}\},$$

and the result follows from the theory of non-negative matrices.

supply system, with the value of Π interpreted as the implicit rate of return on the durable input.

(2) Our second comment concerns aggregation over firms $f = 1,...,F$ which have a common technology characterized by a parameter vector $\boldsymbol{\alpha}$, but face differing price vectors $\mathbf{p}_f$, and $\Pi(\mathbf{p}_f;\boldsymbol{\alpha})$ denotes the restricted profit function of firm f. Then aggregate profit equals $\Sigma_{f=1}^{F} \Pi(\mathbf{p}_f;\boldsymbol{\alpha})$, and aggregate net supply is given by the corresponding sum of price derivatives. In the linear-in-parameter form (3) for the profit function, this aggregation "carries past" the parameters, preserving the linear structure,

$$\sum_{f=1}^{F} \Pi(\mathbf{p}_f;\boldsymbol{\alpha}) = \sum_{m=1}^{M} \alpha_m \left[\sum_{f=1}^{F} Q^m(\mathbf{p}_f)\right]. \tag{15}$$

The parameters of this problem could then be estimated from aggregate net supply data and disaggregated price data. The system has an obvious application when detailed price data is available but disclosure rules prevent the release of detailed quantity data. In practice, something less than completely disaggregate price data may be sufficient to compute the expressions $[\Sigma_{f=1}^{F} Q^m(\mathbf{p}_f)]$. For example, with the Diewert specification (5) of the linear profit form involving terms $p_i Q^{ij} = -(p_i p_j)^{1/2}$, if the mean μ_i of p_{fi} and covariance σ_{ij} of $p_{fi}p_{fj}$ across firms are reported, and if one can make the maintained hypothesis that $\mathbf{p}_f$ is multivariate log-normally distributed and the number of firms in the aggregate is large, then one has

$$\frac{1}{F}\sum_{f=1}^{F} [-(p_{fi}p_{fj})^{1/2}] \cong -(\mu_i\mu_j)^{1/2} \frac{(1+\sigma_{ij}/\mu_i\mu_j)^{1/4}}{(1+\sigma_{ii}/\mu_i^2)^{1/8}(1+\sigma_{jj}/\mu_j^2)^{1/8}}. \tag{16}$$

Then a series of observations on aggregate net supply and means and covariances of prices within each aggregate observation would be sufficient to estimate the model.

Another interpretation of the system (15) and (16) can be given for a single firm facing uncertainty, with f denoting the state of nature. The corresponding net supply system expresses expected net supplies as linear-in-parameters functions of the means and covariances of prices.

A second class of aggregation problems occurs when firms $f = 1,...,F$ face common commodity prices, but have different technologies with non-measured "local" factors. If the restricted profit function of each firm is of the general linear form, with the parameter vector $\boldsymbol{\alpha}_f$ differing across firms, but the Q-functions common to all firms, then aggregate net supplies can be interpreted as coming from a "representative" technology of the same form with a parameter vector $\boldsymbol{\alpha} = (1/F)\sum_{f=1}^{F} \boldsymbol{\alpha}_f$. [See Klein (1952–53).]

(3) Our third comment concerns tests of restrictions on the technology. In a number of cases, these can be formulated as linear restrictions on the parameter vector $\boldsymbol{\alpha}$, and thus tested using standard linear statistical theory. For example, suppose goods 1 and 2 are outputs, the remaining commodities are inputs and we wish to test the non-jointness of production of the two commodities. This hypothesis implies that the net supply of good 1 cannot be affected by the price of good 2, i.e., $\alpha_{12} = 0$ in the two-price linear profit form (4).

(4) Our fourth comment concerns the introduction of exogenous variables from the technology into the restricted profit function when it is assumed to have the linear profit form. Important cases include cost minimization for fixed output and profit maximization in a subset of commodities with the remaining commodity levels fixed. These variables will typically enter in a non-linear way (except in the case of constant returns). However, one can introduce a linear-in-parameter form jointly over the variable commodity prices and exogenous variables, and establish second-order approximation properties for this form in both sets of variables. For example, if the profit function has arguments $\mathbf{p} = (p_1,\ldots,p_N)$ and $\mathbf{z} = (z_{N+1},\ldots,z_L)$, and the underlying technology is convex in $(\mathbf{x},\mathbf{z})$, then the profit function is concave in $\mathbf{z}$ and one might consider a linear form,

$$\Pi(\mathbf{p},\mathbf{z}) = \sum_{i,j=1}^{N} \alpha_{ij}(-(p_i p_j)^{1/2}) + \sum_{i=1}^{N} \sum_{j=N+1}^{L} \beta_{ij} p_i z_j + p_1 \sum_{\substack{i,j=N+1 \\ i\neq j}}^{L} \gamma_{ij}(z_i z_j)^{1/2} + p_1 \sum_{i=N+1}^{N} \delta_i(-z_i^2), \tag{17}$$

in $L(L+1)/2 + (L-N)$ parameters with α_{ij} and γ_{ij} non-negative and symmetric, δ_i non-negative. This form has the second-order approximation to any function $\Phi(\mathbf{p},\mathbf{z})$ which is of type **RP** in $\mathbf{p}$, is concave in $\mathbf{z}$, is twice continuously differentiable jointly in $(\mathbf{p},\mathbf{z})$, has the property **GS** in $\mathbf{p}$, and has a dual **GS** property in $\mathbf{z}$ (i.e., the marginal product $\partial\Pi/\partial z_i$ is non-decreasing in all other z_j).

Our last comment concerns the construction of "nested" functional forms for the restricted profit function which can be interpreted as arising from a two-stage decision process (ex ante and ex post) of the firm. Suppose the linear profit form (3) summarizes the result of ex post optimization, with the α_m which are fixed ex post being the ex ante decision variables. Suppose these ex ante decision variables are described parametrically by a second linear profit form,

$$\psi(q_1,\dots,q_m) = \sum_{l=1}^{L} \beta_l R^l(q_1,\dots,q_m),$$

with β non-negative, R^l non-decreasing in $\mathbf{q}$, and

$$\alpha_m = \partial\psi/\partial q_m. \tag{18}$$

Then, the optimal ex ante profit maximum is found to equal $\psi(Q^i(\mathbf{p}),\dots,Q^m(\mathbf{p})) \equiv \Phi(\mathbf{p})$, which is of type **RP**, and the optimal net supply vector resulting from the two-stage optimization is found to satisfy

$$\hat{x}_i(\mathbf{p}) = \sum_{l=1}^{L} \beta_l \sum_{m=1}^{M} R^l_m(Q^1(\mathbf{p}),\dots,Q^M(\mathbf{p}))Q^m_i(\mathbf{p}), \tag{19}$$

and

$$\hat{\alpha}_m(\mathbf{p}) = \sum_{l=1}^{L} \beta_l R^l_m(Q^1(\mathbf{p}),\dots,Q^M(\mathbf{p})). \tag{20}$$

This structure is linear in the underlying production parameters β_l. A detailed discussion of ex ante–ex post production structures and their estimation by nested linear profit forms is given in Chapter II.4.

Chapter II.3

POLAR FUNCTIONS WITH CONSTANT TWO FACTORS – ONE PRICE ELASTICITIES*

GIORA HANOCH

The Hebrew University of Jerusalem

1. Introduction

If more than two variable factors are involved in a production process, the degree of substitutability between factors, measured by the Elasticity of Substitution (ES), may de defined in a variety of ways. Mundlak (1968b) has shown that the different concepts of ES are different combinations (constrained or unconstrained) of elements of the underlying Hessian matrix. Following his classification, we distinguish between one factor-one price ES concepts (such as Allen–Uzawa partial ES, denoted here by A_{ij}), two factors–one price ES (TOES), and two factors–two prices ES (TTES) (such as Hicks' Direct ES) (DES_{ij}), or McFadden's Shadow ES (SES_{ij}).[1]

*This article comprises a revision of parts of two earlier papers: Section 6 of "Generation of New Production Functions Through Duality" and "Production with Constant Two Factors-One Price Elasticities of Substitution", Discussion Papers nos. 118 and 117, respectively, Harvard Institute of Economic Research, April 1970. It also draws from a more recent article in *Econometrica* [Hanoch (1975a)]. I am thankful to Zvi Griliches, who encouraged and supported this research. I have benefited from discussions with Kenneth Arrow, Erwin Diewert, Melvyn Fuss, Dale Jorgenson, Lawrence Lau, Daniel McFadden, Michael Rothschild, and Christopher Sims. I am indebted to the National Science Foundation for financial assistance (Grant No. 2762X), and to Harvard University, where I visited in 1969–70 and in 1973–74, while on leave from the Hebrew University, Jerusalem.

[1]See Allen (1938), Hicks (1963), McFadden (1963), and Mundlak (1968b).

Mundlak argued correctly, that the choice of the relevant ES measure is independent of assumptions regarding constancy of any particular measure, and generally of the choice of functional forms and methods of estimation of the production relation. However, traditionally the search for econometrically convenient functional forms with a limited set of parameters was attempted by imposing constancy on some ES concept. This yielded disappointing results for both the one factor–one price and the TTES concepts for more than two variable factors. Uzawa (1962) and McFadden (1963) have shown, that constant A_{ij}, DES_{ij}, or SES_{ij} yield functional forms which are too restrictive, and generally unacceptable, for more than two or three factors. For example, constant A_{ij} implies that all A_{ij} are equal for pairs of factors within the same group, and $A_{ij} = 1$ for factors belonging to different groups. (For two factors, all ES concepts coincide, and constant A_{ij} yield the well-known two-factor CES model.[2])

This chapter presents and analyzes several useful polar pairs of functional forms for production functions and joint production frontiers, the common feature of which is the *constancy of* some *TOES* concept.[3] All these functions are true generalizations of the CES model, which yield, in many cases, less restrictive but manageable estimation equations for factor–demand and output–supply relations under competitive markets.

As shown below, the two TOES concepts held constant in these models are simply related to the more basic ES concepts A_{ij}. The first, R^k_{ij}, equals the ratio A_{ik}/A_{jk}, the constancy of which yields the family of Constant-Ratio-ES (CRES) functions defined by Gorman (1965), and the specific subfamily analyzed in Hanoch (1975a). Two noted special cases of CRES are: (1) the homothetic case of CRESH, defined and analyzed in Hanoch (1971); and (2) the non-homothetic Mukerji (1963) function, used also by Dhrymes and Kurz (1964).

The second TOES concept, D^k_{ij}, equals the difference $A_{ik} - A_{jk}$. The major focus of the present analysis are functional forms with constant R_{ij}, and their *polar* functions, which turn out to yield constant D_{ij}, and to have some additional desirable properties.

The family of implicitly additive models with a single output, given in

[2]See Arrow et al. (1961). The statement refers to ES defined for constant output. See Mundlak (1968b, p. 231).

[3]On polar functions, see Chapter I.2. Another common feature of these models is their *Implicit Additivity*, as defined in Hanoch (1975a), where constancy of TOES concepts is shown to be equivalent to implicit additivity.

Hanoch (1975a), yields many well-known special cases of polar pairs of production functions, when equality of certain parameters is imposed.

In joint production situations with multiple outputs, the ES between outputs is defined in an analogous manner to A_{ij}, R_{ij}^k, and D_{ij}^k, substituting maximum revenue (at fixed inputs) for minimum costs (at fixed outputs). To each constant TOES production function model, there corresponds a similar constant TOES factor requirement function (with modified parameter restrictions to assure convexity instead of concavity). Equating two such functions to each other, clearly yields a frontier which exhibits separability of outputs from inputs, and constant R_{ij}^k or D_{ij}^k for both outputs and inputs.

The concept of *Elasticity of Transformation*[4] (T_{ij}) is defined as a generalization of A_{ij} to the situation of competitive profit maximization with multiple variable outputs and inputs. Again, we define two quantities–one price ET (TOET),

$$RT_{ij}^k = T_{ik}/T_{jk} \quad \text{and} \quad DT_{ij}^k = T_{ik} - T_{jk}.$$

Finally, polar pairs of production relations with *constant TOET* are presented, generalizing the single-output non-homothetic CRES and CDE models, through the profit-polar transformation suggested in Hanoch (1975a).

Section 2 below defines and interprets various ES and ET concepts used here. Section 3 summarizes, in the most part, previous results concerning CRES and CDE models, for production functions with a single output, and their corresponding various special cases. Finally, Section 4 includes generalizations of both the CDE and the CRES models to joint production relations with many outputs, under constant ratios or differences of elasticities of substitution or transformation.

2. Elasticities of Substitution and Transformation

Let y_j and x_i denote output and input quantities, respectively, with p_j and w_i the corresponding prices. Assume first that a firm produces efficiently a single output y, minimizing costs $\Sigma\, x_i w_i$, under competitive factor markets (exogenous w_i). The production function $y = f(x_1,x_2,\ldots,x_n) = f(\mathbf{x})$ is assumed to be a strictly increasing, twice continuously differentiable, quasi-concave function, possessing a unique

[4]For the definition of T_{ij}, see Diewert (1973a).

dual cost function $C(\mathbf{w};y)$, which is twice continuously differentiable, concave and linear-homogeneous in $\mathbf{w}$, and increasing with y from 0 to ∞.[5]

Following the notation in Mundlak (1968b), let

$$\hat{z}_i = \mathrm{d}\log z_i = \frac{1}{z_i}\mathrm{d}z_i.$$

The elasticity of demand for factor x_i with respect to w_j, at constant output, is

$$E_{ij} = \left.\frac{\hat{x}_i}{\hat{w}_j}\right|_y = \frac{w_j}{x_i}K_{ij} = s_j A_{ij}, \tag{1}$$

where K_{ij} is the element of the inverse bordered Hessian matrix,

$$[K] = \begin{bmatrix} 0 & f_1 \dots f_n \\ f_1 & f_{11} \dots f_{1n} \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ f_n & f_{n1} \dots f_{nn} \end{bmatrix}^{-1},$$

and s_j is the optimal share of factor x_j in total variable costs, or the elasticity of C with respect to w_j,

$$s_j = \frac{w_j x_j}{C} = \left.\frac{\hat{C}}{\hat{w}_j}\right|_y \quad \text{since} \quad \frac{\partial C}{\partial w_j} = x_j.$$[6]

A useful interpretation of Allen's A_{ij} is as follows:

$$A_{ij} = \frac{E_{ij}}{s_j} = \frac{\hat{x}_i/\hat{w}_j}{\hat{C}/\hat{w}_j} = \left.\frac{\hat{x}_i}{\hat{C}}\right|_{(\mathrm{d}w_j)}, \tag{2}$$

namely, A_{ij} is the elasticity of x_i with respect to C, for a change in another price w_j (output and all w_i constant). It is also the demand cross-elasticity E_{ij}, "normalized" by the relative change in C. Also,

$$A_{ij} = A_{ji} = \left.\frac{\hat{x}_j}{\hat{C}}\right|_{(\mathrm{d}w_i)},$$

due to the symmetry of $[K]$.

Generalizing this particular interpretation to two factors–one price

[5]See Chapter I.2 for specifications of the conditions for the general (non-differentiable) case, and for references with respect to this particular case.
[6]See Shephard (1953, p. 11).

elasticities,[7] we get

$$D_{ij}^k = \frac{\widehat{(x_i/x_j)}}{\hat{C}}\bigg|_{(dw_k)} = \frac{\hat{x}_i}{\hat{C}}\bigg|_{(dw_k)} - \frac{\hat{x}_j}{\hat{C}}\bigg|_{(dw_k)} = A_{ik} - A_{jk}, \tag{3}$$

that is, D_{ij}^k is the elasticity of the factors ratio x_i/x_j with respect to costs C, for a change in another price w_k. It is also the cross-elasticity of *relative* demand,

$$\frac{\widehat{(x_i/x_j)}}{\hat{w}_k}\bigg|_y,$$

normalized by $\hat{C}/\hat{w}_k = s_k$.

Another TOES concept is defined as follows:

$$R_{ij}^k = \frac{\hat{x}_i}{\hat{x}_j}\bigg|_{(dw_k)} = \frac{\hat{x}_i/\hat{C}}{\hat{x}_j/\hat{C}}\bigg|_{(dw_k)} = \frac{A_{ik}}{A_{jk}}; \tag{4}$$

hence R_{ij}^k is the elasticity of x_i with respect to x_j, for a change in another factor's price w_k, output and other prices w_i $(i \neq k)$ constant. Also, by equation (1),

$$R_{ij}^k = \frac{\hat{x}_i/\hat{w}_k}{\hat{x}_j/\hat{w}_k}\bigg|_y = \frac{E_{ik}}{E_{jk}}.$$

Thus, both D_{ij}^k and R_{ij}^k defined in equations (3) and (4) are subject to relatively simple and intuitive economic interpretations. In addition these TOES concepts are free from a basic flaw common to all TTES concepts, as pointed out by Mundlak; namely, that the relative magnitude of two price changes, $\hat{w}_i$ and $\hat{w}_j$, has to be restricted. Any particular such restriction (equivalent to picking a particular directional change in the prices space) is essentially arbitrary, and yields a different TTES concept.[8]

The polar pairs of CDE and CRES models presented here, have constant D_{ij}^k and R_{ij}^k, respectively – with the additional (necessary) property, that both are independent of k, for any i,j $(i \neq k \neq j)$.

For the case of multiple outputs $\mathbf{y} = (y_1,...,y_m)$, assume the joint production frontier to be given by $F(\mathbf{y};\mathbf{x}) \equiv 0$, where F is increasing in $\mathbf{y}$, decreasing in $\mathbf{x}$, continuously twice differentiable and convex in $(\mathbf{y};\mathbf{x})$,

[7] The "pure" TOES concept defined by Mundlak (1968b, p. 229) is $H_{ij}^k = \widehat{(x_i/x_j)}/\hat{w}_k|_y$, hence $H_{ij}^k = s_k D_{ij}^k$, analogous to $E_{ij} = s_j A_{ij}$.

[8] See Mundlak (1968b, Sect. 3.3).

with $F(\mathbf{0};\mathbf{0})=0$.[9] A competitive firm maximizes profits $(\sum y_j p_j - \sum x_i w_i)$, under exogenous (positive) prices $(\mathbf{p};\mathbf{w})$. The unique dual profit function $\pi(\mathbf{p};\mathbf{w})$, is increasing in $\mathbf{p}$ and decreasing in $\mathbf{w}$, and is also twice continuously differentiable; π is non-negative, positive linear homogeneous in $(\mathbf{p};\mathbf{w})$, and convex over the domain where π is finite.[10] The partial derivatives of $\pi(\mathbf{p};\mathbf{w})$ yield the factor–demands and output–supplies, as follows:[11]

$$\frac{\partial \pi}{\partial w_i} = -x_i^* \quad \text{and} \quad \frac{\partial \pi}{\partial p_j} = y_j^*.$$

The ES between inputs may be generalized to this joint-production case, by simply substituting in their definition the fixed vector of outputs $\mathbf{y}$ for the single output y. Analogously, the corresponding ES between outputs are defined so that maximization of revenue at fixed input quantities $\mathbf{x}$ is substituted for minimization of costs at fixed $\mathbf{y}$, with the revenue function $R(\mathbf{p};\mathbf{x})$ being linear homogeneous and convex in $\mathbf{p}$. At any point $(\mathbf{y}^*,\mathbf{x}^*)$ which yields maximum profits π at given prices $(\mathbf{p};\mathbf{w})$, costs are minimized, subject to fixed $\mathbf{y}^*$; revenue maximized for fixed $\mathbf{x}^*$, and π satisfies $\pi = R(\mathbf{p};\mathbf{x}^*) - C(\mathbf{w};\mathbf{y}^*)$. The output–output ES are then defined as

$$A_{ik} = \frac{\hat{y}_i/\hat{p}_k}{\hat{R}/\hat{p}_k} = \frac{\hat{y}_i}{\hat{R}}\bigg|_{(\mathrm{d}p_k)},$$

$$D_{ij}^k = \frac{(\widehat{y_i/y_j})}{\hat{R}}\bigg|_{(\mathrm{d}p_k)} = A_{ik} - A_{jk},$$

and

$$R_{ij}^k = \frac{\hat{y}_i}{\hat{y}_j}\bigg|_{(\mathrm{d}p_k)} = \frac{A_{ik}}{A_{jk}},$$

where i,j,k refer to outputs, and $\mathbf{x}^*$ constant.

Generalized separable CRES and CDE models presented in Section 4, exhibit constant TOES between inputs as well as between outputs.

[9] $F(\mathbf{y},\mathbf{x})$ as given may not satisfy these conditions, although some transformation $F^* = h(F)$ does. In this case, choose $F^*(\mathbf{y},\mathbf{x}) = 0$ to represent the production frontier. The differentiability requirements restrict this relative to the general (non-differentiable) case of duality discussed in Chapter I.2.

[10] See McFadden Chapter I.1 and Diewert (1973a) for proofs of this version of the duality theorem and a more complete specification of the properties of π and F.

[11] See Diewert (1973a) for this extension of Shephard's Lemma, and Hanoch in Chapter I.2.

A one quantity–one price Elasticity of Transformation (ET) is the following generalization of A_{ij}: $T_{ij} = -\pi\pi_{ij}/\pi_i\pi_j$, where $i,j = 1,\ldots,m+n$, and subscripts of π denote partial derivatives.[12] In analogy to equation (2), we may interpret T_{ij} as (minus) the elasticity of a variable quantity with respect to profits π, for a change in one price (holding constant other prices, including the variable's own price); that is, for inputs x_i and x_j:

$$-T_{ij} = \frac{\hat{x}_i}{\hat{\pi}}\bigg|_{(dw_j)} = \frac{\hat{x}_j}{\hat{\pi}}\bigg|_{(dw_i)},^{13}$$

for outputs y_i and y_j:

$$-T_{ij} = \frac{\hat{x}_i}{\hat{\pi}}\bigg|_{(dp_j)} = \frac{\hat{y}_j}{\hat{\pi}}\bigg|_{(dp_i)}, \tag{5}$$

and for output y_j and input x_i:

$$-T_{ij} = \frac{\hat{y}_j}{\hat{\pi}}\bigg|_{(dw_i)} = \frac{\hat{x}_i}{\hat{\pi}}\bigg|_{(dp_j)}.$$

Equations (5) are easily derived from the definition of T_{ij} above, noting that the elasticity of π with respect to a price is the corresponding share in profits,

$$\frac{\hat{\pi}}{\hat{w}_i} = -\frac{x_i w_i}{\pi} = \bar{s}_i, \qquad \frac{\hat{\pi}}{\hat{p}_j} = \frac{y_j p_j}{\pi} = \bar{s}_j, \qquad \sum_{k=1}^{m+n} \bar{s}_k = 1,$$

where $\bar{s}_k$ is negative for inputs and positive for outputs.

Extending the analogy, define two concepts of two quantities–one price ET (TOET), in analogy to equation (3),

$$DT_{ij}^k = \frac{\widehat{(\pi_j/\pi_i)}}{\hat{\pi}}\bigg|_{(dw_k)\text{ or }(dp_k)} = T_{ik} - T_{jk}, \tag{6}$$

and, in analogy to equation (4),

$$RT_{ij}^k = \frac{\hat{\pi}_i}{\hat{\pi}_j}\bigg|_{(dw_k)\text{ or }(dp_k)} = \frac{T_{ik}}{T_{jk}}. \tag{7}$$

CDET and CRET models, which generalize the corresponding single-output CDE and CRES models, are production frontiers with constant

[12]See Diewert (1973a). This is in analogy to $A_{ij} = CC_{ij}/C_iC_j$ with respect to the cost function.

[13]The last equality follows from the symmetry of $[\pi_{ij}]$, in the continuously twice differentiable case. Note that $\hat{\pi}/\hat{w}_i < 0$.

DT^k_{ij} and RT^k_{ij}, respectively, which are again independent of k ($i \neq k \neq j$; $n+m>3$).

3. A Summary of CRES and CDE (Implicitly Additive) Models

This section presents without proofs functional forms with constant R^k_{ij} or D^k_{ij} between inputs, for production functions with one output. These are discussed in more detail in Gorman (1965), Hanoch (1971, 1975a) and others. The polar functions are derived through the cost-polar transformation (Chapter I.2, Theorem 3). Profit-polar transforms (Theorem 5) may also be derived, if the production functions are restricted to be concave, but are omitted. Note, however, that in the homothetic cases, the cost-polar and the profit-polar functions have identical isoquant-maps (see Chapter I.2). Observe first, that constant A_{ij}/A_{kl} ($i \neq j$; $k \neq l$) for all pairs of factors is equivalent to constant $R^k_{ij} = A_{ik}/A_{jk}$ for all $i \neq k \neq j$, since

$$\frac{A_{ij}}{A_{kl}} = \frac{A_{ij}}{A_{kj}}.\frac{A_{jk}}{A_{lk}} = R^j_{ik}\cdot R^k_{jl}.$$

Gorman (1965) proved, that the general form of a CRES production function is the function $y(\mathbf{x})$ defined implicitly by the equation[14]

$$\sum D_i(y)x_i^{d_i(y)} \equiv 1, \tag{8}$$

where

$$d_i(y) = 1 - \frac{1}{\theta(y)a_i},$$

assuming that a_i, $D_i(y)$ and $\theta(y)$ preserve the required conditions on $y(\mathbf{x})$;[15] if $d_i(y)=0$, $\log x_i$ is substituted for $x_i^{d_i(y)}$. The ES are given by

$$A_{ij} = \theta(y)\frac{a_i a_j}{\sum s_k a_k}, \tag{9}$$

where

[14]The function $\phi(y)$ appearing on the right-hand side in Gorman (1965) may be absorbed in the functions $D_i(y)$, as in equation (8).

[15]Namely, the conditions for yielding a positive, finite, quasi-concave and increasing function $y = f(\mathbf{x})$.

$$s_k = \frac{x_k^* w_k}{\sum x_l^* w_l}$$

is the (variable) share of costs at optimal factor combinations. Thus, $R_{ij}^k = a_i/a_j$ is constant everywhere, and is *independent of k*.

The specific CRES and CDE models. Any econometric application of CRES production functions requires specification of the general functions $D_i(y)$ and $\theta(y)$ appearing in equation (8).

The specific form of CRES presented in Hanoch (1975a) is sufficiently flexible and generalizes many other well-known functions. It is derived by specifying $D_i(y) = D_i y^{-e_i d_i}$ and $\theta(y) \equiv 1$, where D_i (the "distribution parameters") and e_i (the "expansion parameters") are positive constants.[16]

Applying the polar transformation[17] yields another model defined through the dual cost function, in which the ES have *constant differences*, $D_{ij}^k = d_j - d_i$.

The following defines implicitly a function $f(\mathbf{z})$ of n variables z:

$$\sum D_i f^{-e_i d_i} z_i^{d_i} \equiv 1, \tag{10}$$

where $D_i > 0$, $e_i > 0$, $d_i < 1$, and all d_i of same sign; i.e., either $0 < d_i < 1$, or $d_i \leq 0$, $i = 1,\ldots,n$. $\text{Log}[f^{-e_i} z_j]$ replaces $[f^{-e_j d_j} z_j^{d_j}]$ if $d_j = 0$.[18]

The CRES production function is defined by

$$y = f(\mathbf{x}). \tag{11}$$

The polar *CDE cost function* C^* is defined implicitly by applying the cost-polar transformation,

$$\frac{1}{y} = f\left(\frac{1}{C^*}\mathbf{w}\right),$$

which is in the form of "reciprocal indirect production function", defined in Chapter I.2, and $f(\quad)$ is defined in equation (10).

[16]The parameter restrictions given here may be relaxed, if $f(\mathbf{z})$ defined in equation (10) is to be valid only *locally*, rather than for all $\mathbf{x} \gg \mathbf{0}$ (globally). In particular, d_1 (say) may satisfy $d_1 > 1$ $(a_1 < 0)$, if $n \geq 3$ and Z_1 is large enough, relative to other Z_i. See Hanoch (1971, 1975a) for specification and analysis of the weaker local conditions.

[17]I.e., substituting $1/y$ and $\mathbf{w}/C^*$ for y and $\mathbf{x}$, respectively. See Hanoch's Theorem 3 in Chapter I.2.

[18]The conditions for *local* validity are weaker, allowing one d_i to be larger than 1, and some D_i and d_i to be of different sign (if $D_i d_i$ are all of the same sign, and $\max_i D_i > 0$). See footnote 16.

More specifically, the equation defining $C^*(\mathbf{w},y)$ implicitly is

$$\sum D_i y^{e_i d_i}(w_i/C^*)^{d_i} \equiv 1, \tag{12}$$

where $\log[y^{e_j}(w_j/C^*)]$ replaces $[y^{e_j d_j}(w_j/C^*)^{d_j}]$ for $d_j = 0$.

The ES corresponding to equations (11) and (12) are given by $(i \neq j; i,j = 1,\dots,n)$[19]

$$A_{ij} = \frac{a_i a_j}{\sum s_k a_k}, \qquad R^k_{ij} = \frac{a_i}{a_j} = \frac{1-d_j}{1-d_i}, \tag{13}$$

$$A^*_{ij} = \frac{1}{a_i} + \frac{1}{a_j} - \sum s^*_k \cdot \frac{1}{a_k}, \qquad D^{k*}_{ij} = \frac{1}{a_i} - \frac{1}{a_j} = d_j - d_i, \tag{14}$$

where $a_i = 1/(1-d_i) > 0$, and the corresponding cost shares are given by

$$s_k = \frac{D_k d_k y^{-e_k d_k} x_k^{d_k}}{\sum_j D_j d_j y^{-e_j d_j} x_j^{d_j}}, \qquad s^*_k = \frac{D_k d_k y^{e_k d_k} w_k^{d_k}}{\sum_j D_j d_j y^{e_j d_j} w_j^{d_j}}, \tag{15}$$

where x_k are the cost minimizing quantities for output y, and y is defined as a function of $\mathbf{x}$ or $\mathbf{w}$, respectively, in equations (11) or (12).

As shown in Hanoch (1975a), this pair of polar functions may be estimated by log-linear equations, and yield various well-known models as special cases, by assuming equality restrictions on various parameters. Thus, many special cases are testable within the more general framework of these models.

The following summarizes some of these special cases. [For more details, see Hanoch (1975a).]

The homogeneous and homothetic cases (CRESH and HCDE). If $e_i = e = 1/\mu$, all i, both production functions are *homogeneous* of degree μ.

Let a function $H(\mathbf{z})$ be defined as

$$\sum D_i (z_i/H)^{d_i} \equiv 1, \tag{16}$$

specializing equation (10) for this case, with $H = f^e$. $H(\mathbf{z})$ is clearly linear homogeneous. The CRESH homogeneous production function is

$$y = [H(\mathbf{x})]^{\mu}, \tag{17}$$

[19] See Hanoch (1971, 1975a).

and its polar CDEH function is defined by its cost function,

$$C^* = y^{1/\mu} H(\mathbf{w}), \tag{18}$$

where equation (18) implies that the polar production function is also homogeneous of degree μ.[20] Replacing $y^{1/\mu}$ by any function $h(y)$, where h is strictly increasing in y from 0 to ∞ gives the general *homothetic* CRES and CDE models.

If the expansion parameters e_i are not equal to each other, the functions defined in equations (11) and (12) are non-homothetic. A sufficient condition for *concavity* of the production functions for all $\mathbf{x} \gg \mathbf{0}$ is $e_i > 1$, $i = 1,...,n$, for both models.

Explicitly additive models. If the products $\{e_i d_i\}$ are constant,

$$e_i d_i = d/\mu, \qquad i = 1,...n,$$

the corresponding production and cost functions are *explicitly additive.*

The CRES case:

$$y = \left(\sum D_i x_i^{d_i}\right)^{\mu/d}, \tag{19}$$

is the Mukerji (1963) and Dhrymes–Kurz (1964) production function, analyzed also by Hanoch (1971).

The polar CDE function exhibits "indirect explicit additivity", with the "indirect production function" given by

$$y = \left(\sum D_i (w_i/C^*)^{d_i}\right)^{-\mu/d}. \tag{20}$$

Both equations (19) and (20) yield constant TOES, as given in equations (13) and (14), and have simplified log-linear estimation equations as compared with equations (11) and (12). However, as shown in Hanoch (1975a), their usefulness is limited, due to the built-in restrictions connecting expansion behavior to substitution behavior in any explicitly additive (direct or indirect) models.

Direct and indirect addilog. The special cases, $d = \mu$ in equation (19) and $d = -\mu$ in equation (20), yield the functional forms applied in consumer demand analysis by Houthakker (1965).

[20]See Chapter I.2.

"*Direct Addilog*":

$$y = \sum D_i x_i^{d_i}, \qquad 0 < d_i < 1, \tag{21}$$

and

"*Indirect Addilog*":

$$y = \sum D_i (w_i / C^*)^{d_i}, \qquad d_i < 0. \tag{22}$$

These functions, however, are not polar to each other. Since $e_i > 0$, all i, d_i must be positive in equation (21) and negative in equation (22), for these equations to yield valid (positive monotone and quasi-concave) production functions, respectively.[21]

Non-homothetic CES. If e_i are not all equal, but $d_i = d$, all i, then both equations (11) and (12) correspond to the non-homothetic CES model, with constant and equal ES,

$$A_{ij} = \frac{1}{1-d} = a \quad \text{and} \quad A_{ij}^* = 1 - d = \frac{1}{a} = a^*, \tag{23}$$

by equations (13) and (14). In this case, the polar pair of production functions $f(\mathbf{x})$ and $f^*(\mathbf{x})$ may both be expressed (implicitly) by similar forms in the direct mode, namely,

$$\sum D_i (f^{-e_i} x_i)^d \equiv 1 \quad \text{and} \quad \sum D_i^* (f^{*-e_i} x_i)^{d^*} \equiv 1, \tag{24}$$

where

$$D_i^* = D_i^{1/(1-d)} \quad \text{and} \quad d^* = -d/(1-d).$$

Thus, these functions are "self-polar", in the weak sense defined by Houthakker (1965), having the same functional form but different parameters for the polar function. But the subfamily with $0 < a < 1$ has significantly distinct features from its polar sub-family with $a^* = 1/a > 1$. If $a < 1$, all factors are "absolutely essential", since the isoquant surfaces do not intersect any of the axes; whereas if $a^* > 1$, the isoquant surfaces intersect all axes, and no factor is essential.

[21]These restrictions are ignored in the discussions of consumer-demand applications, since utility is ordinal, and may assume negative values.

Limiting CES cases: Linear and Leontief production functions (non-homothetic). If $d = 1$ in equation (24) [or $d_i = 1$, all i, in equation (11)], the ES are the limiting cases: $a = \infty$ and $a^* = 0$ in equation (23). The first case corresponds to the (non-homothetic) production function,

$$\sum D_i y^{-e_i} x_i \equiv 1, \tag{25}$$

with *linear isoquant-surfaces* for each y. The polar model exhibits a linear cost function,

$$C^* = \sum D_i y^{e_i} w_i, \tag{26}$$

and the polar direct production function corresponding to equation (26) is the (non-homothetic) *Leontief fixed coefficients production function,*

$$y = f^*(\mathbf{x}) = \min_i \{(x_i/D_i)^{1/e_i}\}. \tag{27}$$

The optimal quantities (for given $y,\mathbf{w}$) under equation (27) then satisfy

$$(x_i^*/D_i)^{1/e_i} = y \quad \text{or} \quad x_i^* = D_i y^{e_i}, \qquad i = 1,\dots,n,$$

and thus $\mathbf{x}^*$ is independent of factor prices $\mathbf{w}$, with $C^* = \Sigma\, x_i^* w_i$, as in equation (26).

Homogeneous CES. If both $e_i = e = 1/\mu$ and $d_i = d$ in equation (11), the well-known homogeneous CES function is obtained:[22]

$$y = f(\mathbf{x}) = \left(\sum D_i x_i^d\right)^{\mu/d}, \tag{28}$$

where equation (28) is clearly a special case of both equation (19) (explicitly additive) and equation (24) (CES). The cost function dual to equation (28) is

$$C = y^{1/\mu} \sum D_i^{1/(1-d)} w_i^{-d/(1-d)}. \tag{29}$$

Applying the cost-polar transformation to equation (29) gives the polar production function,

$$y = f^*(\mathbf{x}) = \left(\sum D_i^* x_i^{d^*}\right)^{\mu/d^*},$$

[22]See McFadden (1963) and Uzawa (1962) for this n factor case, and the derivation of the cost function. See also Hanoch (1975a).

as in equation (24), with

$$A^*_{ij} = a^* = \frac{1}{a} = \frac{1}{A_{ij}}, \qquad i \neq j.$$

The *homogeneous* limiting cases corresponding to $d = 1$ in equation (28) are the

Linear production function:

$$y = \left(\sum D_i x_i\right)^{\mu},$$

and the

Leontief production function:

$$y = f^*(\mathbf{x}) = [\min_i \{x_i / D_i\}]^{\mu},$$

both homogeneous of degree μ.

Cobb–Douglas. If $d_i = 0$, all i, in equation (11), the (direct) function is

$$\sum D_i \log(x_i / y^{e_i}) \equiv 1,$$

which yields, after some manipulations,

$$y = f(\mathbf{x}) = A \Pi x_i^{b_i \mu}, \tag{30}$$

where

$$b_i = \frac{D_i}{\sum D_k}, \quad \mu = \frac{\sum D_k}{\sum e_k D_k}, \quad A = \exp\left\{-\frac{1}{\sum e_k D_k}\right\},$$

with $\sum b_k = 1$.

This is the well-known homogeneous Cobb–Douglas production function, with ES all equal to 1, and degree of homogeneity μ. Its corresponding (dual) cost function is

$$C = (y/A)^{1/\mu} \Pi[(w_i/b_i)^{b_i}] = A^* y^{1/\mu} \Pi(w_i^{b_i}), \tag{31}$$

where

$$A^* = A^{-1/\mu} \Pi(b_i^{-b_i}).$$

Applying the polar transformation to equation (31) gives the original function (except for the scale constant),

$$y = f^*(\mathbf{x}) = (A^*)^{\mu} \Pi (x_i^{b_i \mu}).$$

This production function is thus "self-polar" (in the strong sense), and is exactly self-polar (i.e., $f \equiv f^*$) if $(A^*)^{\mu} = A$, or equivalently if

$$A = (\Pi b_i^{b_i \mu})^{-1/2}.$$

[The function (30) could also be obtained as a limiting case of the homogeneous CES case (28), for $d \to 0$.]

The *non-homothetic Cobb–Douglas* function is defined as a special case of the general CRES function (8), i.e.,

$$\sum D_i(y) \log x_i \equiv 1. \tag{32}$$

However, the previous discussion shows that the specialized CRES model in equation (11) yields a homogeneous function in this case even if e_i are not equal for all i.

The models summarized here, as well as all their special cases, could be generalized somewhat by substituting an arbitrary increasing function $h(y)$ for y. This will generalize any *homogeneous* case to a *homothetic* function, and will allow some additional flexibility in estimating the non-homothetic models as well, if the functional form of $h(y)$ is specified with a small number of unknown parameters. For example, if $h(y) = e^{y/y_0}$, the average cost function in the homogeneous cases will be U-shaped, with a minimum at y_0. Substituting y/y_0 for $\log y$ in the estimation equations will then preserve their linear properties, and allow estimation of y_0.[23]

4. Multiproduct Production Frontiers with Constant TOES or TOET

This section presents several generalizations of the single-output production functions presented above, for joint-production with multiple outputs. Some of these frontiers exhibit CRES and CDE for elasticities of substitution between inputs (at constant outputs), and between outputs (at constant inputs); others have constant TOET, i.e., constant $RT_{ij}^k = T_{ik}/T_{jk}$ (CRET), or constant $DT_{ij}^k = T_{ik} - T_{jk}$ (CDET).

Special cases of these, which are generalizations to m outputs of the

[23]See Hanoch (1971).

CES and the Cobb–Douglas production functions, are given at the end. Additional generalizations or specializations may be derived along similar lines. Proofs of monotonicity and convexity conditions, as well as details regarding demand and supply relations and estimation equations are similar in nature to the corresponding single output cases, and are generally omitted.

Separable frontiers with constant TOES: constant TOES between inputs. A production frontier $F(\mathbf{y},\mathbf{x}) \equiv 0$, where $\mathbf{y}$ is a vector of m outputs, is denoted *separable* (between outputs and inputs), if $\mathbf{y}$ is weakly separable in F, i.e., there exists a function $g(\mathbf{y})$ such that

$$F(\mathbf{y},\mathbf{x}) = F[g(\mathbf{y}),\mathbf{x}] \equiv 0. \tag{33}$$

with $\partial F/\partial g > 0$. This implies strong separability as well,[24] since $g(\mathbf{y}) = f(\mathbf{x})$ may be solved from equation (33), applying the implicit function theorem.

An immediate implication of this, is that the aggregate $y = g(\mathbf{y})$ may be substituted for the single output y in all the CRES and CDE models presented in Section 3. The cost function is defined for a fixed vector $\mathbf{y}$, hence a fixed scalar $y = g(\mathbf{y})$, and the cost-polar transformation is applicable in complete analogy to the one-output case. Specifying an arbitrary (but valid) $g(\mathbf{y})$ thus generates the production models of Section 3, with constant R^k_{ij} or D^k_{ij} *between inputs.*

Constant TOES between outputs. The function $g(\mathbf{y})$ may itself exhibit constant R^k_{ij} or D^k_{ij} *between outputs* (given constant input quantities $\mathbf{x}$), if it is of similar form to the models discussed above, with appropriate modifications.

Specifically, the aggregate $x = f(\mathbf{x})$, in the separable frontier $g(\mathbf{y}) = f(\mathbf{x})$, is substituted for the single input x in the *revenue function* $R(\mathbf{p},x)$. The revenue function is analogous to the cost function $C(\mathbf{w},y)$, with the exception that R is *convex* (and linear homogeneous) in $\mathbf{p}$, whereas $C(\mathbf{w},y)$ is concave in $\mathbf{w}$. Similarly, $g(\mathbf{y})$ must be quasi-convex, and the transformation surfaces $g(\mathbf{y}) = g_0$ are concave [whereas $f(\mathbf{x})$ is quasi-concave, with convex isoquant surfaces].

The functional form (10) is modified to yield the function $g(\mathbf{y})$ with CRES between outputs as follows:

[24]Weak and strong separability are known to be identical concepts with respect to a partition into two groups. See Goldman and Uzawa (1964).

$$\sum b_j g^{-h_j b_j} y_j^{b_j} \equiv 1, \tag{34}$$

where

$$b_j > 1, \qquad B_j, h_j > 0, \qquad j = 1,\dots,m.$$

The ES between outputs are

$$A_{ij} = \frac{a_i a_j}{\sum a_k s_k} < 0 \quad \text{and} \quad R_{ij}^k = \frac{a_i}{a_j} > 0, \tag{35}$$

where

$$a_j = \frac{1}{1 - b_j} < 0,$$

and s_k are shares of revenues,

$$s_k = \frac{y_k p_k}{\sum y_j p_j} > 0,$$

and A_{ij} are therefore negative (corresponding to concave transformation surfaces).

The corresponding *polar transformation* function $g^*(\mathbf{y})$ exhibits, by analogy, CDE between outputs.

The polar revenue function $R^*(\mathbf{p},x)$ [dual to $g^*(\mathbf{y})$] is defined implicitly by

$$\sum B_j x^{h_j b_j} (p_j / R^*)^{b_j} \equiv 1, \qquad b_j > 1, \tag{36}$$

with

$$A_{ij}^* = -b_i - b_j + \left(\sum s_k^* b_k - 1\right) = \frac{1}{a_i} + \frac{1}{a_j} - \sum s_k^* \left(\frac{1}{a_k}\right),$$

and

$$D_{ij}^{k*} = A_{ik}^* - A_{jk}^* = \frac{1}{a_i} - \frac{1}{a_j} = b_j - b_i,$$

in analogy to equation (14). Note, that the CDE models (12) or (36) allow some A_{ij}^* (but not all) to be of opposite sign to A_{ij}, if more than two inputs or outputs exist, and thus allow cases of *complementarity* between outputs ($A_{ij}^* > 0$) or inputs ($A_{ij}^* < 0$), if $|1/a_i + 1/a_j|$ is small relative to the weighted mean $|\sum s_k^*(1/a_k)|$. Thus, CDE models are generally more flexible in this respect, than their polar CRES models.

Reductions of the general CRES and CDE models to various special cases, by imposing equality restrictions on various parameters, are entirely analogous to all the cases discussed in Section 3, and thus yield a variety of special functional forms for output transformation functions $g(\mathbf{y})$ and their corresponding dual revenue functions $R(\mathbf{p},x)$.

Constant TOES between inputs and between outputs. If $n > 2$ and $m > 2$, the models with CRES and CDE for outputs and for inputs may be combined to yield a polar pair of CRES/CDE frontiers for both inputs and outputs. The following three equations define (simultaneously) the CRES frontier:

$$\sum D_i f^{-e_i d_i} x_i^{d_i} \equiv 1, \qquad d_i < 1,$$
$$\sum B_j g^{-h_j b_j} y_j^{b_j} \equiv 1, \qquad b_j > 1, \tag{37}$$

and

$$f(\mathbf{x}) = g(\mathbf{y}).$$

The polar CDE frontier applies both the cost (for inputs) and the revenue (for outputs) polar transformations to equation (37), to give

$$\sum D_i f^{e_i d_i} (w_i/C^*)^{d_i} \equiv 1,$$
$$\sum B_j g^{h_j b_j} (p_j/R^*)^{b_j} \equiv 1, \tag{38}$$

and

$$f(\mathbf{w}/C^*) = g(\mathbf{p}/R^*).$$

It follows, that the direct form of the polar frontier [equation (38)] is also separable, of the general form $f^*(\mathbf{x}) = g^*(\mathbf{y})$. Various special cases of either f or g or both are analogous to the single-output cases of Section 3.

The profit-polar frontier. The condition for existence of a *profit function* dual to the frontier [equation (37)], is that the feasible set $\mathbf{T} = \{(\mathbf{y},\mathbf{x}): f(\mathbf{x}) \geq g(\mathbf{y})\}$ is *convex*. A sufficient (but not necessary) condition for this is $\max_j h_j \leq \min_i e_i$. For example, if all $e_i > 1$, the function $f(\mathbf{x})$ in equation (37) is concave; if all $h_j < 1$, $g(\mathbf{y})$ is convex, implying convexity of the feasible set $\mathbf{T}$, and of the frontier function $F = g(\mathbf{y}) - f(\mathbf{x})$.

If the profit function exists, the *profit polar* transformation may be applied to equation (37), in accordance with Chapter I.2, Theorem 5. That is, a polar profit function $\pi^*(\mathbf{p},\mathbf{w})$ is defined by the following pair of equations, after elimination of the common parameter f,

$$\sum D_i f^{-e_i d_i}(w_i/\pi^*)^{d_i} \equiv 1,$$
$$\sum B_j f^{-h_j b_j}(p_j/\pi^*)^{b_j} \equiv 1. \tag{39}$$

However, the (direct) frontier corresponding to equation (39) is generally not separable, not homothetic, and *does not* in general exhibit constant TOES.

Constant TOES separable homogeneous and homothetic frontiers. Specializing the case given by equation (37) to equality of the expansion parameters, yields a *homogeneous* frontier.[25] That is: if $h_j = h$, all j, and $e_i = e$, all i, the frontier is homogeneous of degree $\mu = h/e$.

If $\mu < 1$, the frontier is convex, the profit function exists, and the profit-polar transformation is defined, and yields a *separable* special case of equation (39).

However, as shown in Chapter I.2, the cost-polar and the profit-polar transformations yield the same function $f(\mathbf{x})$ (except for a constant proportionality factor) in the homogeneous case. By analogy, the revenue-polar and profit-polar transformation are the same. Consequently, in this special case equations (39) and (38) are virtually identical, with the CDE property $D^k_{ij} = d_j - d_i$ between inputs, and $D^k_{ij} = b_j - b_i$ between outputs. These results may be generalized somewhat, using a similar reasoning. If $f(\mathbf{x}) = h[g(\mathbf{y})]$ is substituted for $g(\mathbf{y})$ in equation (37), where $h(g)$ is a positive and strictly increasing function, then the frontiers (37), (38) and (39) are all *homothetic and separable.* The two alternative polar transformations [i.e., the cost/revenue-polar of equation (38) and the profit-polar of equation (39)] give rise to frontiers with identical maps of input-isoquants and output-transformation surfaces, all with constant TOES. (Detailed proofs of these brief statements are left as an exercise to the interested reader).

A CRET non-homothetic, separable frontier. A non-homothetic separable frontier is derived from equation (37) under the special *explicitly*

[25] A frontier $F(\mathbf{y},\mathbf{x}) \equiv 0$ is *μ homogeneous*, if it satisfies $F(\lambda^\mu \mathbf{y},\lambda \mathbf{x}) \equiv 0$ for all $\lambda > 0$. See Hanoch (1970).

additive case, $h_j b_j = e_i d_i = c > 0$ $(j = 1,...,m; i = 1,...,n)$, in analogy to case (19) for the one-output model. Applying these restrictions to equation (37), this yields

$$f^c = \sum D_i x_i^{d_i}, \quad g^c = \sum B_j y_j^{b_j}, \quad f = g.$$

Hence the frontier is

$$\sum B_j y_j^{b_j} - \sum D_i x_i^{d_i} \equiv 0, \tag{40}$$

where sufficient parameter restrictions for validity are again $B_j > 0$, $D_i > 0$, $b_j > 1$, $0 \leq d_i < 1$ *or* $d_i \leq 0$, all i,j (with $\log x_i$ replacing $x_i^{d_i}$ if $d_i = 0$).

The frontier (40) is directly additive separable, but is non-homothetic in either outputs (unless all $b_j = b$) or inputs (unless all $d_i = d$).

Derivation of the ET in an analogous manner to the single-output case[26] yields

$$T_{ij} = -\frac{a_i a_j}{\sum \delta_k a_k}, \qquad i,j = 1,...,m+n, \tag{41}$$

where $a_i = 1/(1-d_i) > 0$ for inputs, $a_j = 1/(1-b_j) < 0$ for outputs. δ_k are the profit shares, given by $\delta_k = -w_k x_k^*/\pi < 0$ for inputs, and $\delta_k = p_k y_k^*/\pi > 0$ for outputs, where $(\mathbf{y}^*, \mathbf{x}^*)$ are optimal input and output quantities, respectively, yielding positive maximum profits

$$\pi = \sum p_j y_j^* - \sum w_i x_i^* > 0.$$

Note that

$$\sum^{m+n} \delta_k = 1 \quad \text{and} \quad \delta_k a_k < 0, \qquad \text{all } k.$$

Thus, T_{ij} of equation (41) are positive between outputs or inputs, and negative between an output y_j and an input x_i, under the sufficient parameter restrictions given above. Since this is a special case of equation (37), it also exhibits CRES for ES between inputs (under $\mathbf{y}$ constant) and between outputs (under $\mathbf{x}$ constant). This specializes, of course, to the Mukerji function of equation (19), in the case of a single output, with $B_1 = 1$ and $b_1 = d$.

[26]We omit the details of this derivation, which may be worked out by the reader.

The polar frontier: non-separable, non-homothetic CDET. By analogy, a *CDET Frontier* is obtained through the profit-polar transformation, such that the polar unit-profit frontier has the same form as equation (40) in the price variables $(\mathbf{p};\mathbf{w})$. The profit function $\bar{\pi}(\mathbf{p};\mathbf{w})$ is then given implicitly by

$$\sum B_j(p_j/\bar{\pi})^{b_j} - \sum D_i(w_i/\bar{\pi})^{d_i} = 0. \tag{42}$$

Equation (42) yields a valid profit function, and hence a valid production frontier, under exactly the same parameter restrictions as in equation (41).

The ET for this frontier are given, in analogy to previous results, by

$$T_{ij} = -\alpha_i - \alpha_j + \sum \delta_l \alpha_l, \tag{43}$$

where $\alpha_i = 1 - b_i < 0$ for outputs, $\alpha_i = 1 - d_i > 0$ for inputs, and δ_l are profit shares, with $\delta_l \alpha_l < 0$, and $\sum \delta_l = 1$. It follows immediately, that T_{ij} between pairs of inputs are negative, in sharp contrast to the CRET frontier. This reflects dominance of expansion effects over substitution effects, for this particular frontier (under the sufficient conditions $b_j > 1 > d_i$). For output–output and output–input pairs, T_{ij} may be of either sign, depending on the relative magnitudes of α_i, α_j, and $\bar{\alpha} = \sum \delta_l \alpha_l$.

The polar frontier corresponding to equation (42) is not directly separable, and is non-homothetic in either outputs or inputs. As a result, the corresponding short-run ES, for fixed outputs or inputs, do not exhibit CDES nor any other simple relation to one another or to the corresponding T_{ij}.

This CDET frontier yields convenient estimation equations for relative demands or supplies, under competitive profit maximization, in analogy to the results corresponding to equation (20). That is, (y_j/y_1) is log-linear in the two variables, $(p_j/\bar{\pi})$ and $(p_1/\bar{\pi})$, $j \geq 2$; and (x_i/y_1) is log-linear in $(w_i/\bar{\pi})$, and $(p_1/\bar{\pi})$. Alternatively, $\bar{\pi}$ could be eliminated, to yield $(m + n - 2)$ equations, each including two quantity-ratios and two price-ratios.[27]

The generalized CES–CET frontier. The homogeneous separable models discussed above have *constant ES* between inputs as well as between

[27] See Hanoch (1975a) for the one-output analogous case.

outputs, under the special case: $d_i = d$, $b_j = b$, $e_i = e$ and $h_j = h$, for all i,j, in equation (37).

Denoting $\mu = h/e$, this frontier is given as follows:

$$\left(\sum B_j y_j^b\right)^{1/b} = \left(\sum D_i x_i^d\right)^{\mu/d}, \tag{44}$$

where the ES are $A_{ij} = 1/(1-d) > 0$ for inputs, and $A_{ij} = 1/(1-b) > 0$ for outputs, and the frontier is homogeneous of degree μ.

By analogy to the CES production function, the polar frontier has a similar form, but different parameters [see equation (46)]. The ES of the polar frontier are given by $A^*_{ij} = 1-d$ between inputs, and $A^*_{ij} = 1-b$ between outputs. (The polar frontier is also μ homogeneous.)

A homothetic generalization of equation (44), with the same CES property, is given by

$$\left(\sum B_j y_j^b\right)^{1/b} = h\left[\left(\sum D_i x_i^d\right)^{1/d}\right], \tag{45}$$

where h is an arbitrary positive increasing function.

In the homogeneous CES case, however, *the elasticities of transformation are also constant* (hence the name CES–CET). This may be shown directly, by spelling out the explicit form of the profit function corresponding to equation (44),[28]

$$\pi = \left(\sum B^*_j p_j^{b^*}\right)^{1/b^*(1-\mu)} \left(\sum D^*_i w_i^{d^*}\right)^{-\mu/d^*(1-\mu)}, \tag{46}$$

where

$$B^*_j = [(1-\mu)^{1-\mu}\mu^{\mu}]^{b/(b-1)} B_j^{-1/(b-1)},$$
$$D^*_i = D_i^{1/(1-d)},$$
$$d^* = -d/(1-d),$$
$$b^* = b/(b-1).$$

Computing the ET for this frontier by direct differentiation of the profit function (46), we get

$$T_{ij} = \frac{b}{b-1}(1-\mu) - 1,$$

for all pairs of outputs;

[28]The proof is omitted. See McFadden (1963) and Chapter I.1.

$$T_{ij} = \frac{d}{1-d} \cdot \frac{1-\mu}{\mu} - 1,$$

for all pairs of inputs; and

$$T_{ij} = -1,$$

for all input–output pairs (as in all separable, homogeneous frontiers).

A special case of this CET frontier is obtained if $d > 0$ and $\mu = d/b$, which is readily seen to be also *a special case of the additive CRET frontier* (40), derived by putting $b_i = b$, $d_i = d$ in (40). The ET are then given by

$$T_{ij} = \frac{b-1}{1-d},$$

for inputs, and their reciprocal

$$T_{kl} = \frac{1-d}{b-1},$$

for outputs, and are all positive.

The generalized Cobb–Douglas frontier. A further specialization of the homogeneous CET frontier (44) may be called the *Generalized Cobb–Douglas* frontier, which is the limiting case of equation (44) if $d \to 0$, $b = 2$. The frontier is given by

$$\sum B_j y_j^2 = \Pi x_i^{2\mu D_i}, \tag{47}$$

where $\Sigma D_i = 1$, $B_j, D_i > 0$. It yields *unitary* ES, $A_{ij} = 1$ between inputs, and $A_{ij} = -1$ between outputs. The constant ET are given by $T_{ij} = -1$ between inputs, $T_{ij} = 2(1-\mu) - 1$ between outputs, and $T_{ij} = -1$ between an output and an input.

Finally, the special case of equation (47), with $B_j = B = \mu^{\mu}(1-\mu)^{1-\mu}\Pi_i D_i^{D_i}$, all j, may be shown to be exactly self-polar, yielding a unit-profit frontier $B \Sigma p_j^2 = \Pi w_i^{2\mu D_i}$, which is exactly of the same functional form as equation (47) with $B_j = B$.[29]

The well-known CES and Cobb–Douglas production functions are obviously special cases of equations (44) and (47), respectively, in the presence of one single output.

[29]On self-duality (self-polarity) in the context of consumer demand, see Houthakker (1965) and Samuelson (1969b). The frontier (47) is viewed as a generalization of Cobb–Douglas, due to its self-polarity, as well as due to its unitary ES.

Chapter II.4

FLEXIBILITY VERSUS EFFICIENCY IN *EX ANTE* PLANT DESIGN

MELVYN FUSS and DANIEL McFADDEN*

University of Toronto and University of California, Berkeley

1. Introduction

Operating flexibility is an important attribute of fixed plant and equipment utilized in a production process, and a factor in the economics of capital–equipment design. The cost of adding flexibility is usually a loss in economic efficiency relative to a "best practice" design for a specific static operating environment. Consequently, the flexibility–efficiency margin is an intrinsic part of the economic calculus of the firm, and a factor to be weighed in econometric analysis of a firm's technological possibilities and behavior. In this paper we develop a model of the firm that incorporates in the plant-design decision a recognition of the possibilities for a tradeoff between flexibility and efficiency. We also provide an algorithm for generating econometric net supply systems within which this phenomenon can be studied empirically.

The following examples illustrate the role of the flexibility–efficiency decision in production processes, and indicate possible applications of the model developed in this paper.

Example 1. Electric utilities are required to meet a system demand that varies over time in a known cycle, and they do so by constructing a mix

*We are indebted to Kenneth Arrow, Peter Diamond, Erwin Diewert, Terence Gorman, and Zvi Griliches for useful comments, but absolve them of responsibility for errors. This chapter was written when the first author was an Assistant Professor at Harvard University; and the second a visiting scholar at M.I.T. It first appeared as Harvard Institute of Economics Research Discussion Paper no. 190 (1971).

of base-load and peak-load generating units. Base-load units have higher capital cost, but lower operating cost, and represent the lower cost technology for providing continuous output. The optimal mix of generating units balances the economic efficiency of supplying the average demand against the flexibility of response to demand variation.

Example 2. In areas where variation in oil, gas, and coal prices causes the least-cost fuel type to vary over time, electric utilities often install boilers in thermal power plants which can be converted to use any of these fuels. These boilers increase capital and maintenance costs and thus result in inefficient production if only one type of fuel is used throughout the lifetime of the plant.

Example 3. In the construction of commercial and industrial buildings, heating and electrical systems are often installed with conversion features to facilitate future expansion or remodeling. The increase in initial cost is justified by uncertainty about eventual use of the structure.

Example 4. Most manufacturing and distribution processes require inventories of various goods due to uncertainties in demand. Firms increase flexibility in meeting demand variations by increasing inventories. However, the increase in inventory carrying cost lowers the economic efficiency of meeting average demand. Similarly, choice of a flexible design as a response to demand uncertainty can be seen in the number of product lines (models, brands) carried in retail stores and the number of commodities produced by multiple-output manufacturers.

Example 5. In multiple-stage production processes, the early stages may be designed so that their output can be tailored to alternative specifications in the latter stages. For example, an automobile chassis is designed so that it can be used in several model types – station wagons, sedans, etc. Consequently, it may not be the least-cost chassis for a single model type – say, station wagons. Flexibility in the production of model types is achieved at the expense of a loss of efficiency in the production of a single model type. Agriculture provides several other examples of this phenomenon: corn may be planted in a pattern that provides maximum yield when chopped for silage; in a second pattern that provides maximum grain yield; or in a third flexible pattern so that it can be either chopped or grown to maturity, depending on the corn price at harvest, with some sacrifice in yield. A similar effect occurs in the selection of breeds of pigs and chickens.

2. Historical Background

Economic theory has traditionally recognized that in the *ex ante* design of a plant, a firm can be expected to weight static efficiency versus dynamic flexibility, the latter being emphasized when the plant is expected to face a variable or uncertain environment. The following classic geometric argument was given by Stigler (1939). Remarkably, this argument has not been quantified and adapted to empirical analysis in the general case.

Consider the textbook *ex ante–ex post* cost curves illustrated in Figure 1(a). The firm can choose *ex ante* a plant design yielding one of the *ex post* average cost curves $EPAC_1$, $EPAC_2$, etc. The curve $EAAC$ is the *ex ante* envelope of such *ex post* curves, and each design is efficient (i.e., tangent to the *ex ante* envelope) at a single output. The curves $EPMC_1$, $EPMC_2$, etc. are *ex post* marginal cost curves, and $EAMC$ is the marginal curve for the *ex ante* envelope. At an output price $p = 4$, anticipated with certainty by the firm, the design $EPAC_1$ will be chosen. Figures 1(b) and 1(c) illustrate the same behavior in terms of total cost and profit. Turn now to the case in which variation in output

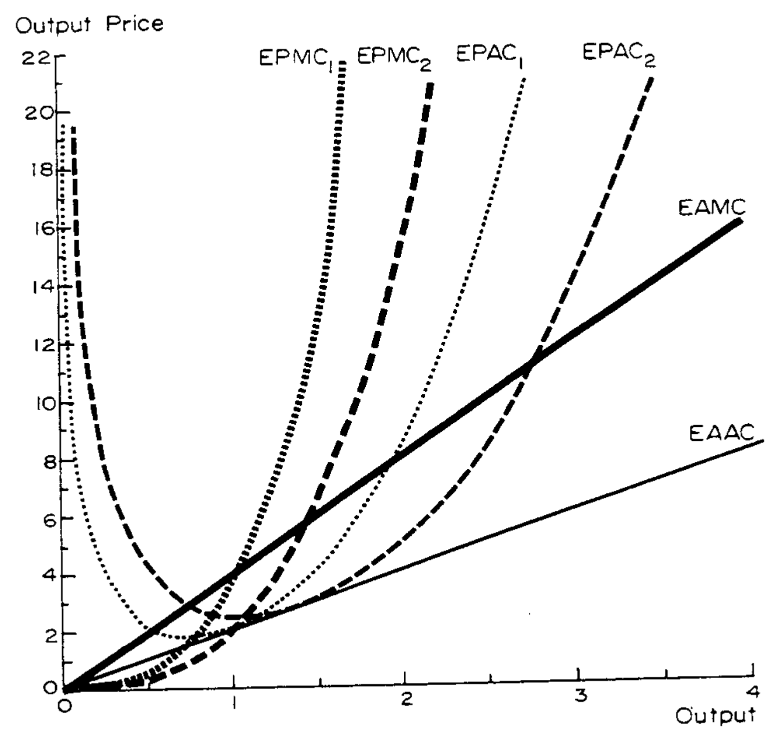


FIGURE 1(a)

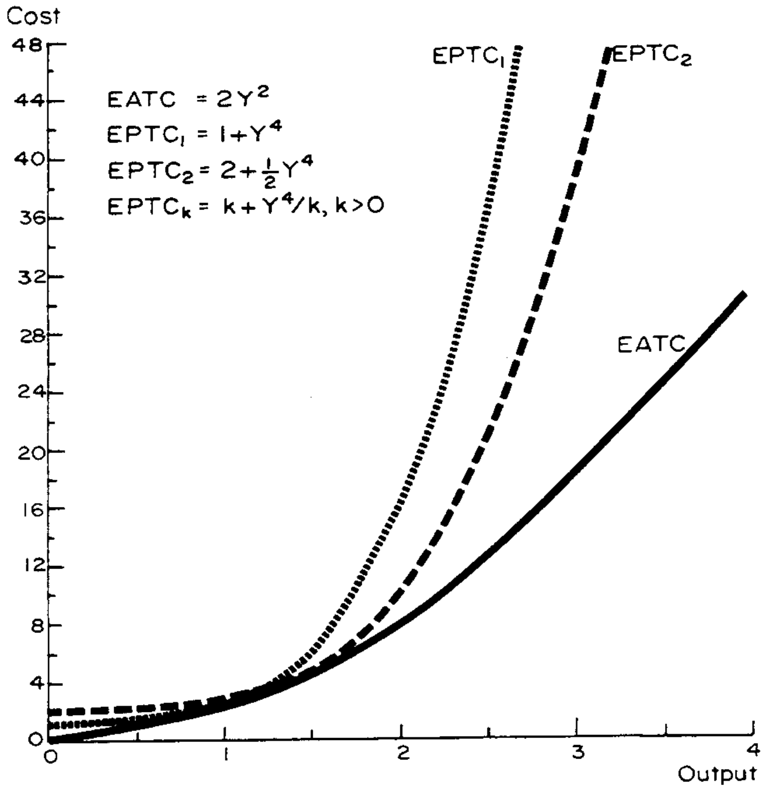


FIGURE 1(b)

price is anticipated, either because of uncertainty or because of intertemporal demand variation for the services of a durable plant. As illustrated in Figure 2(a), there may be a flexible plant design with *ex post* average cost curve $EPAC_3$ which is not efficient at any single output, but which may be a least-cost design given output price variability. For example, if output prices $p = 0$ and $p = 8$ are each anticipated with probability one-half, then this design can be seen in Figure 2(c) to yield higher expected profit than either of the statically efficient designs $EPAC_1$ or $EPAC_2$; i.e., the expected profit 3.0 from the third *ex post* technology, given by the abscissa of chord 3 at the expected price $p = 4$ exceeds the expected profit 2.77 from *ex post* technology one or 2.8 from *ex post* technology two (given by the mid-points of chords 1 and 2, respectively). It follows from the general property of convexity of profit as a function of output price that the expected profit from each of these technologies exceeds the corresponding profit associated with the expected output price.

One extreme case of the *ex ante–ex post* production structure illus-

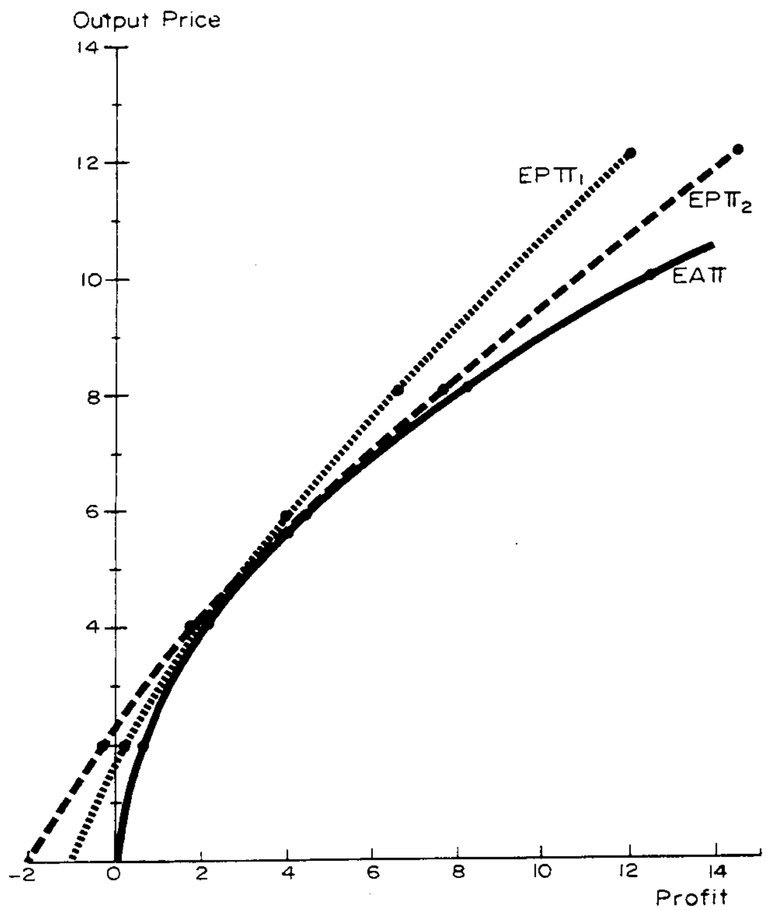


FIGURE 1(c)

trated in Figures 1 and 2 is the "putty–putty" model in which the envelope curve *EAAC* is itself a possible *ex post* design. Clearly, in this case *EAAC* will be the optimal design, achieving both maximum efficiency and flexibility. A second extreme case, illustrated in Figure 3, is the strict "putty–clay" model in which an *ex ante* design that achieves static efficiency at some output fixes the quantities of both capital and variable inputs in the *ex post* technology, and the only source of flexibility is free disposal of output. A less rigid "putty–clay" model in which variable inputs are required *ex post* in fixed proportion to capital services, but free disposal of capital services is possible, is illustrated in Figure 4. The class of intermediate cases between the "putty–putty" and "putty–clay" models, as illustrated in Figure 1, will be called "putty–semiputty" models. Note that the possibility of substituting flexibility for efficiency in the "putty–semiputty" model may be

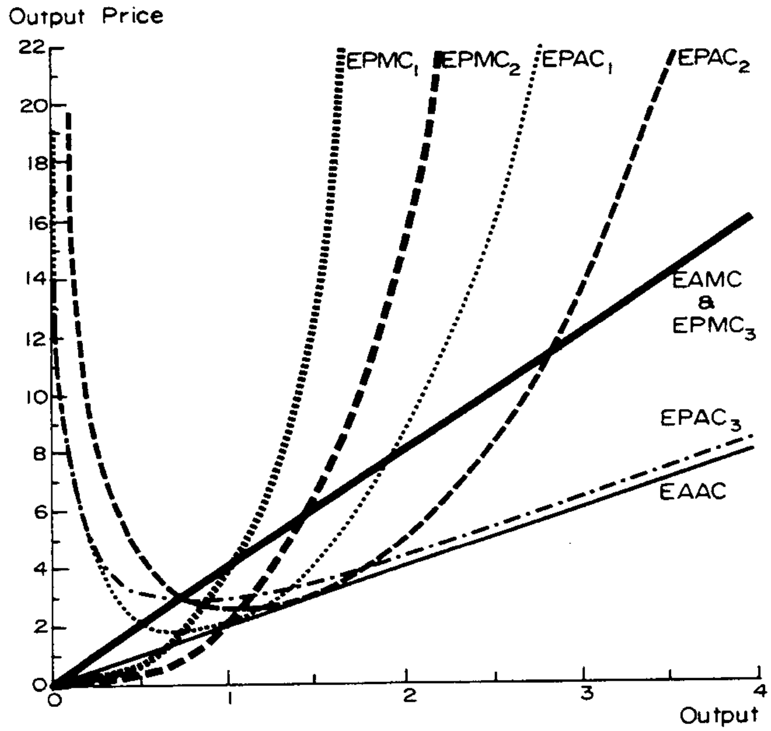


FIGURE 2(a)

present (Figure 2) or absent (Figure 1). Finally, note that the designs available to the firm may offer alternative types of flexibility as well as a simple flexibility–efficiency tradeoff. Figure 5 illustrates a case in which the *ex ante* options are an *ex post* technology that is flexible "upward" (output > 1) and one that is flexible "downward" (output < 1).[1]

The possibility of this flexibility–efficiency tradeoff in plant design has been largely ignored in econometric estimation of production functions,

[1]One footnote on the geometry of *ex ante–ex post* cost curves is in order. We first remind the reader of the geometric property established in the famous Wong–Viner footnote: The mutual tangency of an *ex ante* envelope and *ex post* average cost curve occurs at the output at which the *ex ante* and *ex post* marginal cost curves intersect (for example, unit output in Figure 1), and identifies the *ex ante* optimal *ex post* technology for the production of this static output. However, this output will not coincide with minimum *ex post* average cost unless the mutual tangency is horizontal. Next, note that when free disposal of output is possible, the total cost curve is monotone non-decreasing, implying that the elasticity of *ex post* average cost with respect to output is at least minus one; i.e., the negatively sloped leg of an average cost curve cannot rise more rapidly than a rectangular hyperbola. Thus, Figure 3 illustrates the most extreme possible case, and a diagram such as Figure 6 is impossible unless disposal of output is costly and *ex post* marginal cost is negative. This minor point is missed in many textbooks.

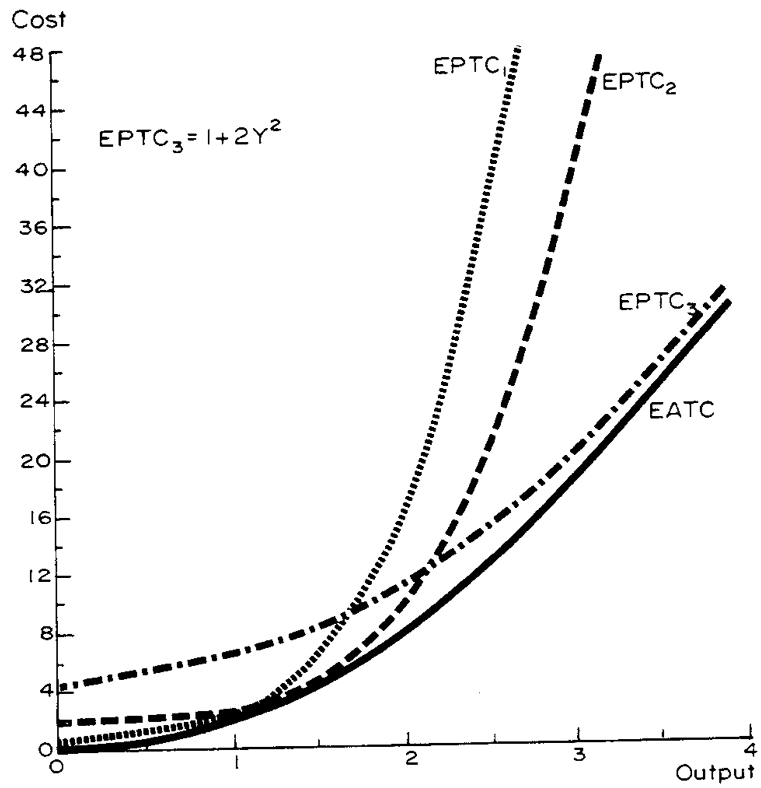


FIGURE 2(b)

probably due to the difficulty of quantifying the effect. In the "putty–clay" model, an attempt has been made by Attiyeh (1967) to obtain estimates when *ex post* factor price ratios may deviate from the factor price ratio for which the *ex post* design is efficient. However, all studies that have come to our attention assume that except for random errors, observed operating points lie on a locus of efficient production plans – an *ex ante* frontier for a cross-section study, or a particular *ex post* frontier for a time-series study. These frontiers are sometimes linked by an implicit "putty–putty" assumption, particularly in combined cross-section time-series studies. On the other hand, it should be clear that if the flexibility–efficiency tradeoff in plant design is present, the meaning of econometric production functions is altered substantially. This point can be developed most easily using Figure 7, which illustrates a family of *ex post* unit isoquants I_1, I_2, etc. and an envelope of efficient points E, conventionally interpreted as the *ex ante* unit isoquant. If firms anticipate variations in relative input prices, they may choose an *ex post* technology like I_3 or I_4. Observed operating points – say, v_1, v_2,... – will

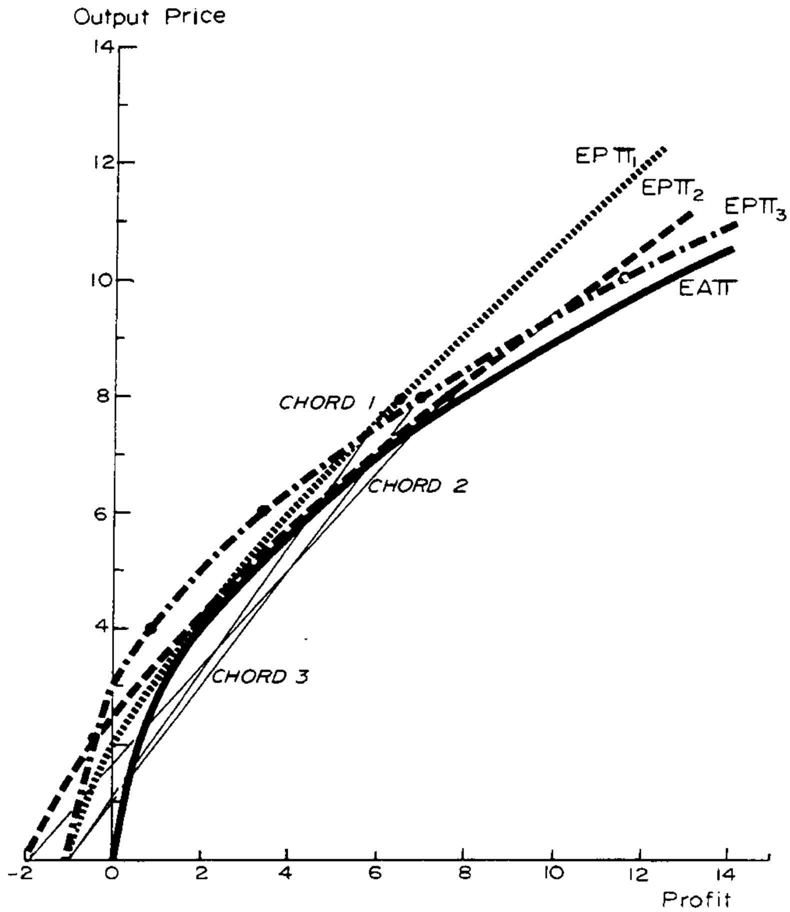


FIGURE 2(c)

lie on these curves. Suppose an econometric production function is fitted to these points. First, note that the unit isoquant of this function will typically underestimate substantially the efficiency of the "best practice" *ex ante* envelope *E*. This will be the case even if the estimated isoquant is taken to be the southwest boundary of the convex hull of the points v_1, v_2,... in order to reduce "optimization errors of the firm". Second, note that the points v_1, v_2,... may show considerable dispersion about the estimated isoquant, suggesting large measurement and optimization errors that are not in fact present. Third, note that the curvature of the estimated isoquant will bear no simple relation to the curvature of the envelope *E*. For example, in Figure 7 the utilization of data points like $\dot{v}_1$, v_2, v_3, v_4 will produce an estimated isoquant with substantial curvature, whereas the envelope *E* has zero curvature.

The conclusion we draw from these observations is that, in the

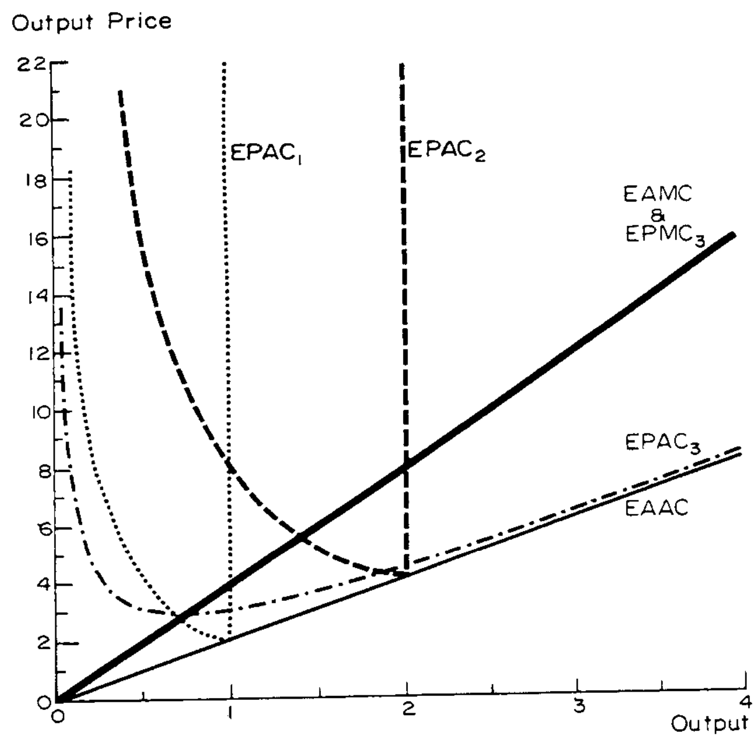


FIGURE 3(a)

presence of a significant flexibility–efficiency tradeoff, conventional econometric production functions provide very little information on the structure of the *ex ante* "best practice" envelope *E*, and may indeed provide misinformation. More fundamentally, we conclude that the concept of a static "best practice" envelope *E* characterizing the *ex ante* technology is inadequate, and in environments where firms face considerable uncertainty and intertemporal variation, irrelevant. It is in this case impossible to define meaningful isoquants, either theoretically or empirically, in a static picture of "one-period" production possibilities in which the flexibility–efficiency tradeoff has no explicit representation. The most satisfactory procedure would seem to be to abandon the elusive concept of the static *ex ante* isoquant and seek a quantification of *ex ante* design possibilities in which the total *ex post* operation of the plant is considered in the flexibility–efficiency decision.

Reliance on the concept of a static isoquant can also give rise to misleading conclusions in analyzing applied problems. For example, in the field of public utility regulation, consider the controversial Averch–Johnson (1962) thesis that a firm under a lax regulatory constraint may

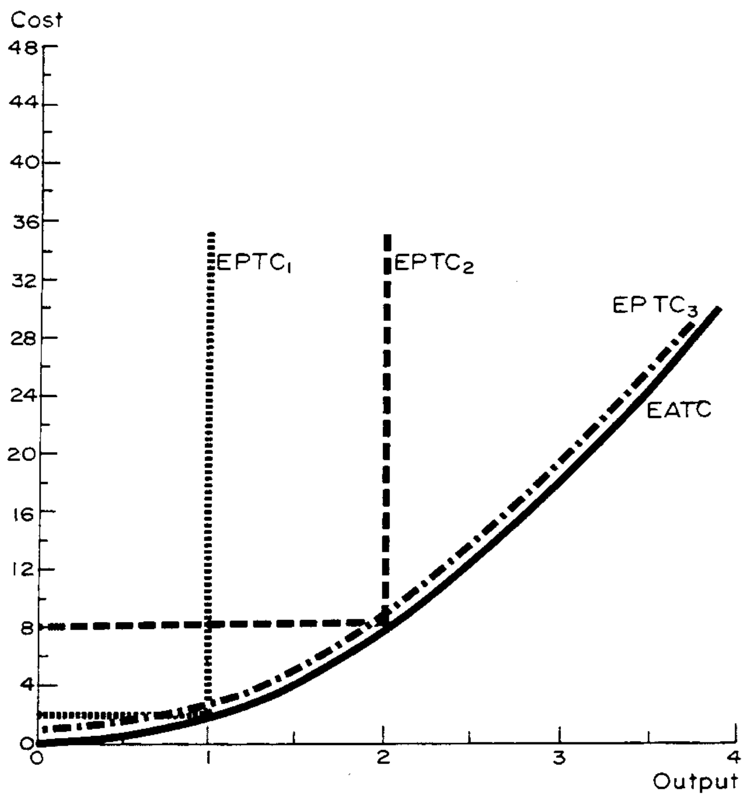


FIGURE 3(b)

overcapitalize and therefore operate inefficiently. Their analysis is based on the static isoquant. This isoquant is irrelevant for the majority of regulated industries, whose factors of production include a large percentage of durable capital with a long lifetime. Faced with intertemporal variation and uncertainty, cost-minimizing firms of this type will trade static efficiency for flexibility and, misleadingly, may appear inefficient under the Averch–Johnson analysis. The effect of regulatory constraint on investment and similar problems should be analyzed within the more general model presented in this paper when data are drawn from samples characterized by substantial shifts from "normal" values. An obvious example is post-1973 energy data.

3. A Model of the Firm with an *Ex Ante–Ex Post* Technology

We now present a model of the firm in which an *ex ante* decision is made on plant design, based on expectations about the environment to

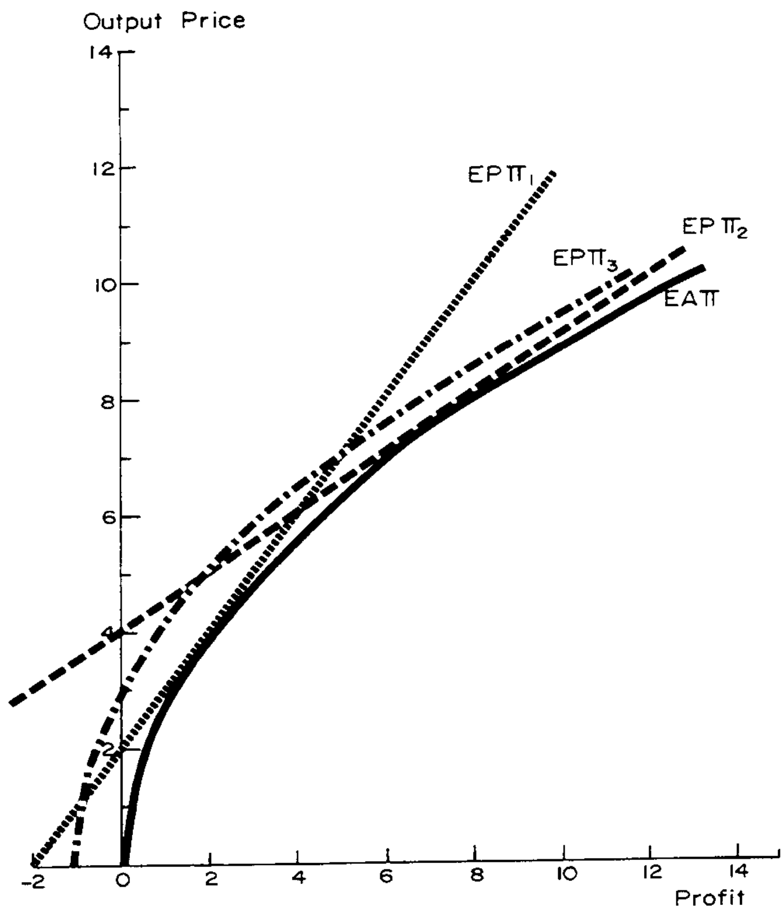


FIGURE 3(c)

be faced *ex post*; and then the resulting *ex post* technology is operated, taking into account the environment which actually prevails. We assume initially a simple choice structure in which a single *ex ante* design decision is made, followed by *ex post* operating decisions that are conditional on the design chosen. This structure will be appropriate for the case of a firm considering the construction of a durable plant that faces intertemporal variability, but no uncertainty. Alternately, it will be appropriate for the formally equivalent case of a firm that faces an *ex ante* design decision under uncertainty, and an *ex post* operating decision with no intertemporal variation.[2] It is possible to show that

[2]This formal equivalence is analogous to the equivalence that exists in consumer theory between the state preference analysis and the intertemporal allocation under certainty analysis when commodities in either different states of nature or different time periods are considered to be distinct.

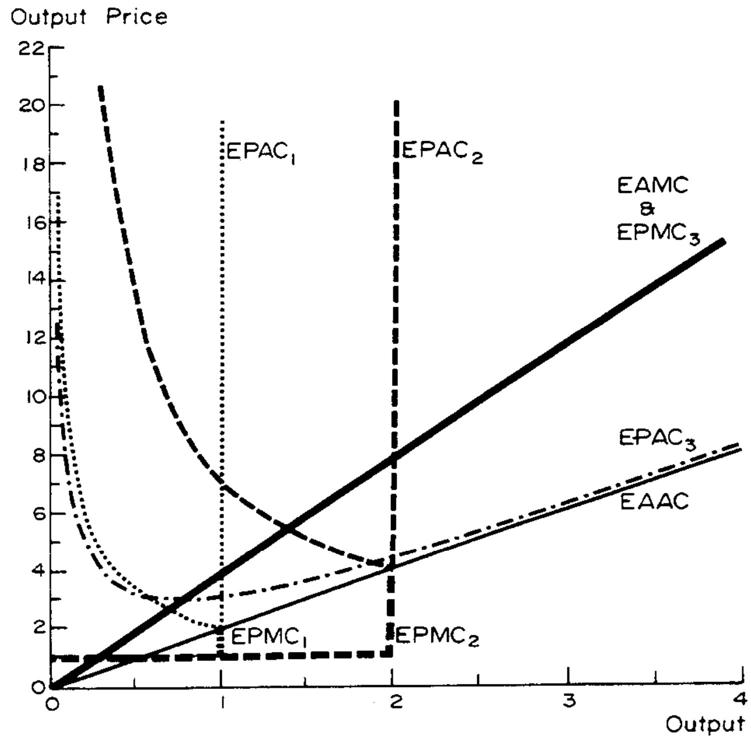


FIGURE 4(a)

more complex multiple-level choice structures (for example, involving construction decisions over time under uncertainty) can be formulated as dynamic programming problems in which the state equation corresponds to the two-stage choice structure described here.

We consider a firm that faces a set **S** of states of the future. A state s in **S** will have one of the following interpretations:

(a) The firm faces an intertemporal future without uncertainty, with $s \in \mathbf{S}$ denoting a chronological time and **S** denoting the set of times in which the plant under consideration might operate. **S** may be a set of discrete times or a continuum and may extend over a finite life-time or the entire future.

(b) The firm faces a one-period future with uncertainty, with $s \in \mathbf{S}$ denoting a state of nature, observed with the *ex ante* design decision but before the *ex post* operating decision. The set **S** of possible states of nature may be finite or infinite.

(c) When both intertemporal variation and uncertainty are present, $s \in \mathbf{S}$ may index both chronological time and state of nature. This case is

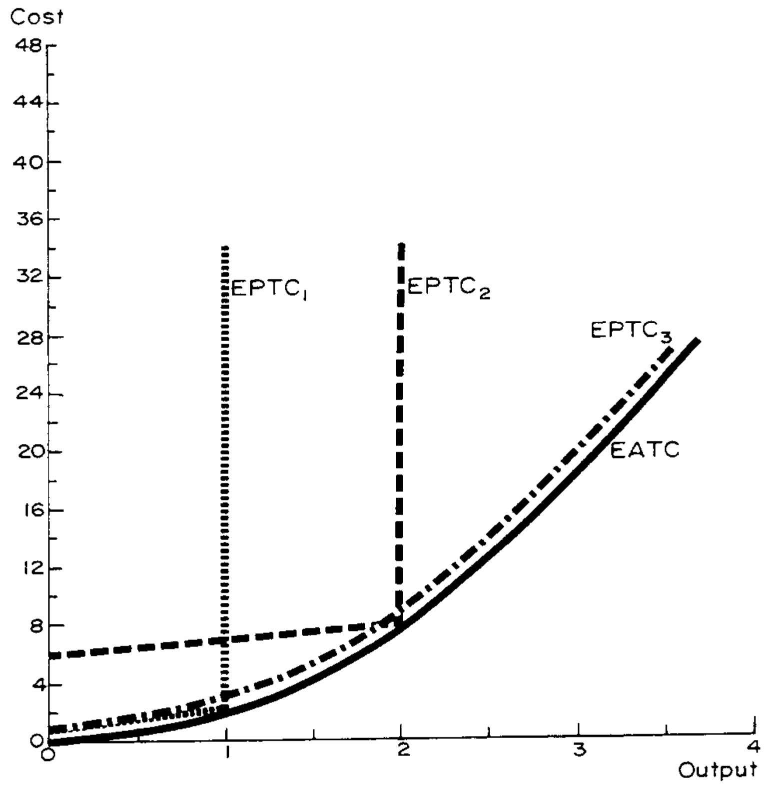


FIGURE 4(b)

formally equivalent to the previous ones, provided the assumption can be made that the firm receives no information in the course of operation that would induce a third-stage choice (and hence make necessary *ex ante* consideration of the strategic possibilities in this reconsideration).

The appropriateness of these three descriptions of the set of future states depends on the way producers form expectations. If a producer has a myopic time horizon but is uncertain about which state of nature will occur in the next period, description (b) is appropriate. If a producer has a planning horizon extending over several time periods but expects that one time sequence of states will occur with probability one and all others with probability zero, description (a) is appropriate. Description (c) applies when a producer forms expectations over a multiple-period time horizon and is uncertain about which state of nature will occur in each future time period, *provided* that the technological possibilities in each time period are independent of operating decisions in prior time periods (given the *ex ante* design) and that the producer does not revise

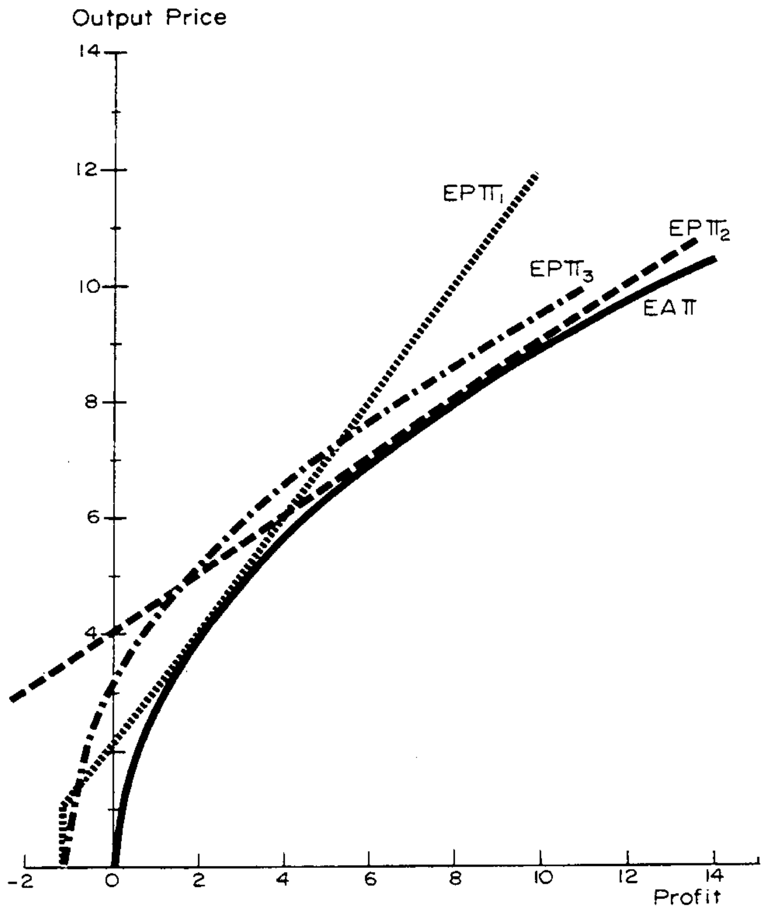


FIGURE 4(c)

his expectations in light of operating experience as time goes on. (This case is discussed in further detail in Section 6.)

Ex ante, the firm has available a set **B** of possible plant designs, with $\mathbf{b} = (\mathbf{a},\mathbf{K}) \in \mathbf{B}$ specifying an abstract vector **a** that describes plant layout, management organization, and exogenous variables influencing *ex post* operation, and a vector $\mathbf{K} = (K_1,...,K_J)$ of inputs of structures and fixed capital equipment.

Ex post, the firm faces competitive markets for N commodities, indexed $n = 1,...,N$, in each future state s, and the plant under consideration will supply a net output (netput) vector $\mathbf{x}_s = (x_{1s},...,x_{Ns})$ to these markets. A component x_{ns} is positive (negative) if commodity n is an output (input). The vector $\mathbf{x} = (x_s : s \in \mathbf{S})$ is termed an *ex post production plan*. We emphasize that the commodities in the *ex post* production plan are those for which competitive spot markets exist, and hence are identified as "variable" netputs in the usual terminology.[3]

Choice of a plant design $\mathbf{b} = (\mathbf{a},\mathbf{K}) \in \mathbf{B}$ determines an *ex post variable technology* $\mathbf{V}(\mathbf{b})$, the set of *ex post* production plans that are possible for a plant with design $\mathbf{b}$. Define a corresponding *ex post total technology* $T(\mathbf{a},\mathbf{K}) = \{(\mathbf{a},\mathbf{K},\mathbf{x}) | \mathbf{x} \in \mathbf{V}(\mathbf{a},\mathbf{K})\}$ for $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$. The set of all designs and *ex post* production plans available defines an *ex ante envelope technology* $\mathbf{T}^{ea} = \{(\mathbf{a},\mathbf{K},\mathbf{x}) | (\mathbf{a},\mathbf{K}) \in \mathbf{B},\ \mathbf{x} \in \mathbf{V}(\mathbf{a},\mathbf{K})\}$.

For the classical case of a firm that faces a non-varying future with certainty so that flexibility is not a factor, the relation of the *ex ante* envelope technology and *ex post* total technology is analogous to that of the *ex ante* and *ex post* total cost curves of Figure 1(b): The optimal design will choose an *ex post* technology that will be operated at a point of "mutual tangency" of this technology and the *ex ante* envelope technology. We show below that this geometric property continues to hold in the general case where the firm's environment induces it to trade flexibility for efficiency. Note that the *ex ante* envelope technology contains explicit information about the tradeoff between flexibility and efficiency. We conclude that this technology is the appropriate generalization of the production structure underlying the classical *ex ante* total cost curves of Figures 1 and 2.

The competitive spot market prices faced by the firm in a state s are denoted by an N-vector $\tilde{\mathbf{p}}_s = (\tilde{p}_{1s},\ldots,\tilde{p}_{Ns})$. Then, $\tilde{\pi}_s = \tilde{\mathbf{p}}_s \cdot \mathbf{x}_s = \sum_{n=1}^{N} \tilde{p}_{ns} \cdot x_{ns}$

[3] A commodity which is to be delivered in some future date (or state) s may be traded in a *futures* market and/or in a *spot* market. Trade in a futures market occurs at the present time, when *ex ante* decisions are being made. Trade in a spot market occurs at the date at which the commodity is to be delivered. The price of a commodity in a futures market may be expressed in present currency units (the *forward* price), with trade interpreted as an exchange of present currency units for a contract to deliver the commodity at date s. It may also be expressed in currency units at date s (the *future* price), with trade interpreted as an exchange of contracts to deliver the commodity in state s and to deliver currency in state s. The commodity price in the spot market (the *spot* price) is denominated in units of currency at date s. In the presence of a full set of futures markets for the "variable" commodities, spot markets will be redundant as long as no agent gains "new" information in the interval between the openings of the future and spot markets. Then, spot prices will equal future prices, and will be related to forward prices by a discount factor determined in a competitive bond market. In the absence of formal future markets other than the bond market, this formulation can be assumed to continue to hold provided firms are hypothesized to have point expectations of spot prices, or to be risk neutral so that spot prices have the interpretation of certainty equivalents. However, a preferable model formulation in this case is to introduce firm expectations explicitly in the description of future states S, as in the case (c) above. To avoid ambiguities in interpretation, we assume hereafter that formal futures markets do exist for variable commodities, and that spot prices equal future prices.

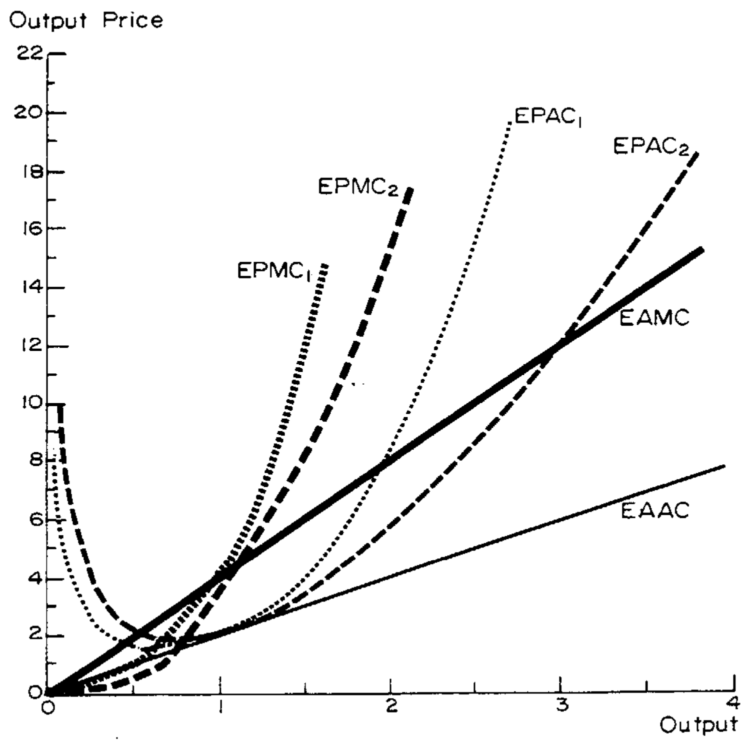


FIGURE 5(a)

is the variable profit in state s associated with the netput vector $\mathbf{x}_s$. We shall assume that the firm weighs (discounts) variable profit in state s by a factor δ_s, and has the objective for a given *ex post* variable technology $\mathbf{V}(\mathbf{b})$ of maximization of the "discounted sum" of variable profits over *ex post* production plans in $\mathbf{V}(\mathbf{b})$. For the model with intertemporal production and no uncertainty, δ_s is the discount rate from time s to reference time 0 established in a competitive bond market, $\tilde{\mathbf{p}}_s$ is a vector of spot prices, and the *ex post* objective of the firm is maximization of present value of variable profit. In the model with uncertainty and no intertemporal variation, δ_s is the probability that state s will occur, and the *ex post* objective of the firm is maximization of expected value of variable profit.[4]

The "discounted sum", or present value, of variable profit for an *ex post* production plan $\mathbf{x}$ can be written as the sum $\pi = \sum_{s \in \mathbf{S}} \delta_s \tilde{\mathbf{p}}_s \cdot \mathbf{x}_s =$

[4]In the model of intertemporal production $\delta_s \cdot \tilde{p}_{is}$ is a forward price and the forward prices are the variables with respect to which the firm maximizes present value. In the single-period uncertainty model, $\delta_s \cdot \tilde{p}_{is}$ is a certainty-equivalent price and the existence of competitive contingency bond markets implies that the firm will maximize expected value of profit, using certainty–equivalent prices as the exogenous variables.

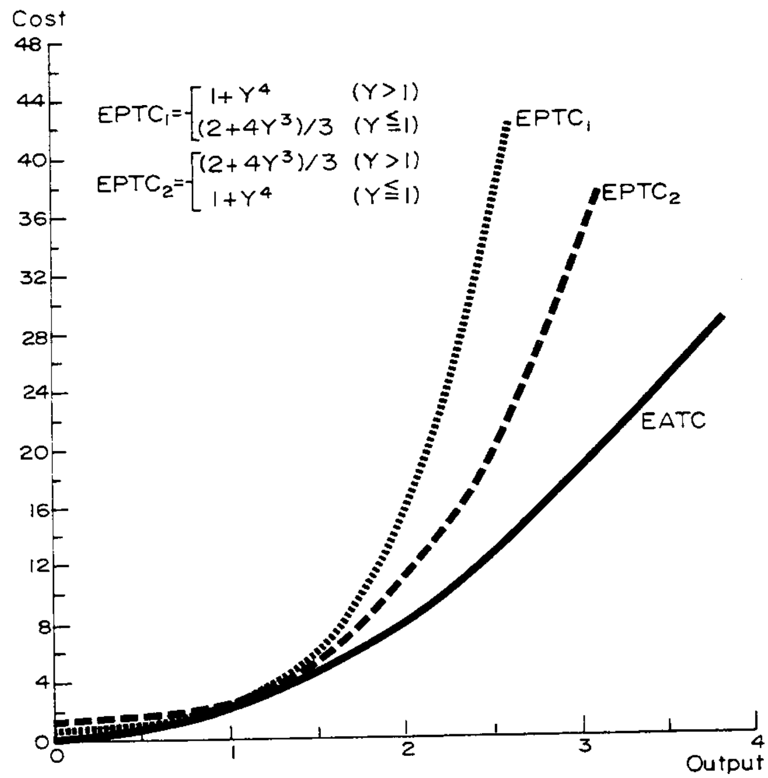


FIGURE 5(b)

$\sum_{s\in S}\delta_s\bar{\pi}_s$ when **S** is a finite set, and as an integral $\pi = \int_S \delta_s \bar{\mathbf{p}}_s\cdot\mathbf{x}_s\, d\mu(s)$ in the general case of a measure space $(\mathbf{S},\mathcal{S},\mu)$ of future states. Define an N-vector $\mathbf{p}_s = \delta_s\bar{\mathbf{p}}_s$ of forward prices for state s and a *forward price vector* $\mathbf{p} = (\mathbf{p}_s\!:\! s \in \mathbf{S})$.[5] The present value of variable profit, hereafter termed *intertemporal variable profit*, can then be written in the case of finite **S** as the inner product of the vectors **p** and **x**, $\pi = \mathbf{p}\cdot\mathbf{x} = \sum_{s\in S}\mathbf{p}_s\cdot\mathbf{x}_s$. We shall carry over this inner product notation to the general case $\pi = \mathbf{p}\cdot\mathbf{x} \equiv \int_S \mathbf{p}_s\cdot\mathbf{x}_s\, d\mu(s)$.

We emphasize that the intertemporal variable profit $\pi = \mathbf{p}\cdot\mathbf{x}$ of an *ex post* production plan includes the present value of the quasi-rents accruing to the fixed inputs of structure and equipment. We can define the fixed cost of the firm by assuming that the capital inputs $\mathbf{K} = (K_1,\ldots,K_J)$ can be purchased in competitive markets at prices given by a

[5]In the following sections we shall use the language of intertemporal model since it is the most familiar one. From the preceding footnote it should be clear that only a slight change of language is required to render the analysis applicable to the other cases described in the paper.

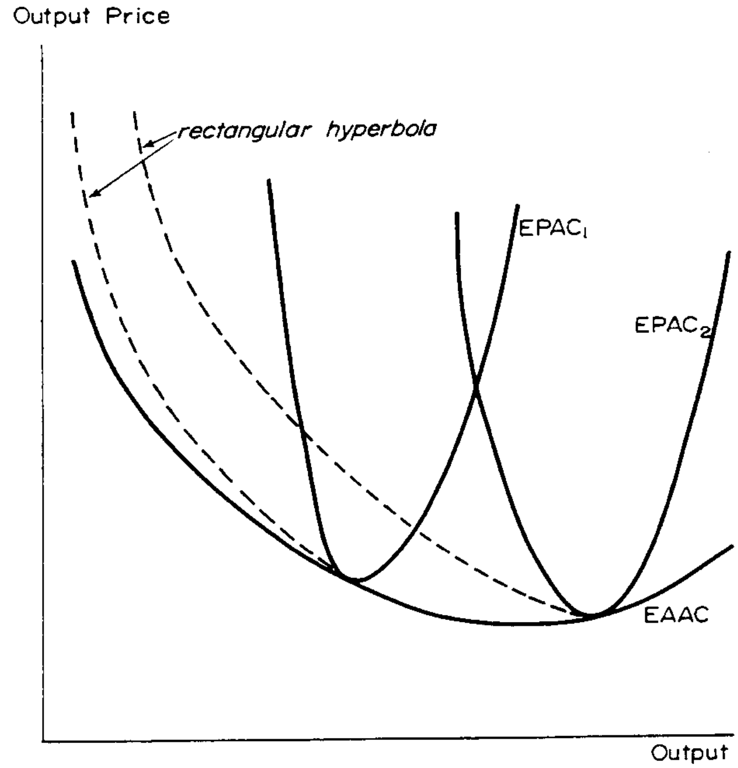


FIGURE 6

vector $\mathbf{r} = (r_1,\dots,r_J)$. We employ the convention that the signs of quantities and prices of these durable inputs are defined so that $K_j \leqq 0$ and $r_j \geqq 0$ for each j, and $\mathbf{r}\cdot\mathbf{K}$ is fixed cost with a negative sign. Then, the intertemporal total profit associated with a plant design $(\mathbf{a},\mathbf{K})$ and *ex post* production plan $\mathbf{x}$ is $\mathbf{p}\cdot\mathbf{x} + \mathbf{r}\cdot\mathbf{K}$.

The optimizing behavior of the competitive firm can now be summarized. Given an *ex post* variable technology $\mathbf{V}(\mathbf{b})$ with $\mathbf{b} \in \mathbf{B}$, maximization of intertemporal variable profit for a forward price vector $\mathbf{p}$ yields an *intertemporal variable profit function,*

$$\Pi(\mathbf{b},\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{x} \mid \mathbf{x} \in \mathbf{V}(\mathbf{b})\}. \tag{1}$$

The function Π is finite on a convex cone of prices for each design $\mathbf{b}$, and is a convex, conical (i.e., positively linear homogeneous), closed function of $\mathbf{p}$ on this cone (see Chapter II.2). Analogously, one may define an *ex post intertemporal total profit function,*

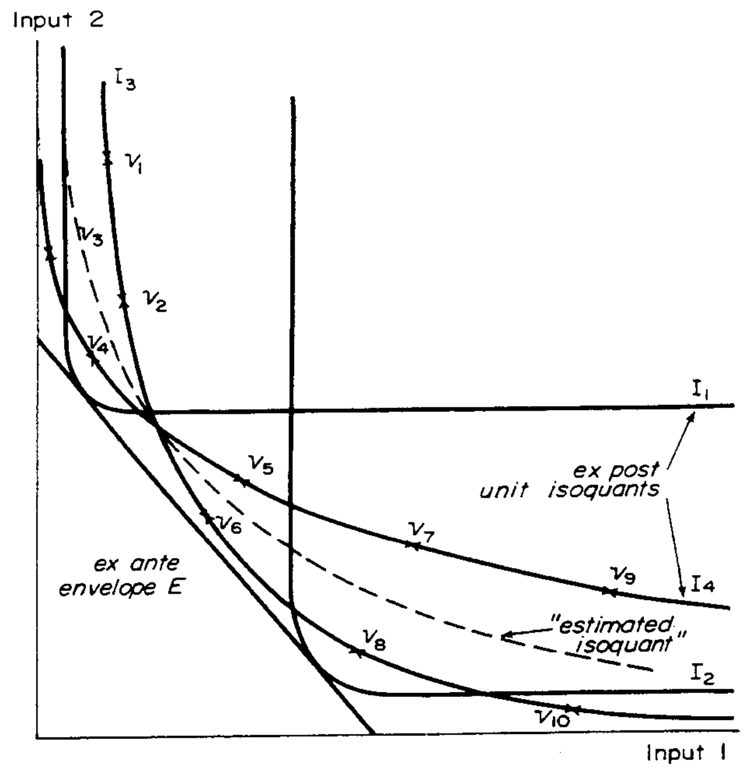


FIGURE 7

$$\boldsymbol{\phi}(\mathbf{a},\mathbf{K},\mathbf{r},\mathbf{p}) = \Pi(\mathbf{a},\mathbf{K},\mathbf{p}) + \mathbf{r}\cdot\mathbf{K}. \tag{2}$$

Ex ante maximization of total profit over possible *ex ante* designs for a forward price vector **p** and durables price vector **r** yields an *ex ante envelope profit function*,

$$\begin{aligned}\Phi(\mathbf{r},\mathbf{p}) &= \sup\{\mathbf{r}\cdot\mathbf{K} + \mathbf{p}\cdot\mathbf{x} \mid (\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea}\}\\ &= \sup\{\mathbf{r}\cdot\mathbf{K} + \Pi(\mathbf{a},\mathbf{K},\mathbf{p}) \mid (\mathbf{a},\mathbf{K}) \in \mathbf{B}\}\\ &= \sup\{\boldsymbol{\phi}(\mathbf{a},\mathbf{K},\mathbf{r},\mathbf{p}) \mid (\mathbf{a},\mathbf{K}) \in \mathbf{B}\}.\end{aligned} \tag{3}$$

In summary, for each possible *ex post* total technology and set of competitive markets, one can (normally) find an optimal *ex post* production plan and associated level of intertemporal total profit. The firm chooses *ex ante* a design that maximizes this profit, and then chooses *ex post* the actual optimal production plan given the *ex ante* design and the information available at the time the *ex post* decision is made.

4. Functional Forms for the *Ex Ante–Ex Post* Production Structure

The preceding section describes the two-stage (or two-level) decision procedure that is contained in our model of *ex ante–ex post* technological structures. In this section we will provide an intuitive and non-rigorous introduction to our procedure for generating quantitative *ex ante–ex post* technologies, and will illustrate (with the aid of examples) the use of this algorithm in the formation of econometric models. The mathematical justification of the procedure is contained in Section 5.

4.1. *The Algorithm*

The algorithm is an extension to the *ex ante–ex post* analysis of a result that is called "the derivative property of the restricted profit function" in Chapter II.2. This result states that the vector of partial derivatives of this function with respect to commodity prices, when it exists, equals a unique profit-maximizing netput bundle. That is,

$$\pi(\mathbf{q};\boldsymbol{\alpha}) = \sum_i q_i \cdot \hat{x}_i(\mathbf{q};\boldsymbol{\alpha}), \tag{4}$$

and

$$\frac{\partial \pi(\mathbf{q};\boldsymbol{\alpha})}{\partial q_i} = \hat{x}_i(\mathbf{q};\boldsymbol{\alpha}), \tag{5}$$

where $\pi(\mathbf{q};\boldsymbol{\alpha})$ is a restricted profit function, $\mathbf{q} = \{q_i\}$ is a vector of commodity prices, $\hat{\mathbf{x}} = \{\hat{x}_i\}$ is the profit-maximizing netput bundle, and $\boldsymbol{\alpha}$ is the vector of production parameters.

[Comparing (5) and the derivative of (4) with respect to q_k, one obtains the condition $\sum_i q_i \cdot \partial \hat{x}_i(\mathbf{q};\boldsymbol{\alpha})/\partial q_k = 0$.] The duality theorem for restricted profit functions (see McFadden's Chapter I.1) implies the existence of a unique technology $X(\boldsymbol{\alpha})$ satisfying $X(\boldsymbol{\alpha}) = \{\mathbf{x} | \mathbf{q}\cdot\mathbf{x} \leqq \pi(\mathbf{q},\boldsymbol{\alpha}) \text{ for all } \mathbf{q}\}$ and $\pi(\mathbf{q},\boldsymbol{\alpha}) = \sup\{\mathbf{q}\cdot\mathbf{x} | \mathbf{x} \in X(\boldsymbol{\alpha})\}$. Then, $\hat{\mathbf{x}}(\mathbf{q},\boldsymbol{\alpha}) \in \mathbf{X}(\boldsymbol{\alpha})$ and $\pi(\mathbf{q},\boldsymbol{\alpha}) \leqq \mathbf{q}\cdot\hat{\mathbf{x}}(\mathbf{q}',\boldsymbol{\alpha})$, with equality if $\mathbf{q} = \mathbf{q}'$. Using property (5), it is possible to generate single-level netput supply systems for econometric estimation of the underlying production parameters, starting from choice of an appropriate functional form for the restricted profit function (see, for example, Chapter II.2). The value of this approach is that it can provide closed functional forms for the netput supply system. The associated technology need not be described by a closed functional form. We emphasize that the algorithm does *not* provide a constructive process for obtaining profit functions from functional forms for the technology; it is, in fact, most useful in complex settings where it is impossible to obtain closed functional forms simultaneously for the profit function and the technology.

A similar algorithm holds for the two-stage, *ex ante–ex post* production structure. Suppose the *ex ante* envelope profit function can be written in the form

$$\Phi(\mathbf{r},\mathbf{p};\boldsymbol{\alpha}) = \sum_{i=1}^{L} Q_i(\mathbf{p})\hat{a}_i(\mathbf{r},\mathbf{p};\boldsymbol{\alpha}) + \sum_{j=1}^{J} r_j\hat{K}_j(\mathbf{r},\mathbf{p};\boldsymbol{\alpha}), \tag{6}$$

where $\mathbf{p}$ is the vector of forward prices, the $\hat{a}_i$ and $\hat{K}_j$ are the parameters, variable *ex ante* and fixed *ex post*, which specify the optimal *ex post* technology (or, equivalently, the optimal *ex ante* design), and $\boldsymbol{\alpha}$ is a vector of underlying *ex ante* parameters. The $Q_i(\mathbf{p})$ are functions of the vector of forward prices and play a role analogous to the prices q_i in the single-level formulation. The parameters of the optimal *ex post* technology satisfy

$$\partial\Phi/\partial Q_i = \hat{a}_i(\mathbf{r},\mathbf{p};\boldsymbol{\alpha}), \qquad \partial\Phi/\partial r_j = \hat{K}_j(\mathbf{r},\mathbf{p};\boldsymbol{\alpha}), \tag{7}$$

and the intertemporal profit-maximizing variable netputs satisfy

$$\begin{aligned}\partial\Phi/\partial p_{is} &= \sum_{k=1}^{L} (\partial\Phi/\partial Q_k)(\partial Q_k/\partial p_{is}) \\ &= \sum_{k=1}^{L} \hat{a}_k(\mathbf{r},\mathbf{p};\boldsymbol{\alpha})(\partial Q_k/\partial p_{is}) = x_{is}(\mathbf{r},\mathbf{p};\boldsymbol{\alpha}).\end{aligned} \tag{8}$$

(These relationships will be derived more rigorously in Section 5.) Writing Φ as a "nested" function $\Phi(\mathbf{r},\mathbf{p};\boldsymbol{\alpha}) = \psi(Q_1(\mathbf{p}),\dots,Q_L(\mathbf{p}),r_1,\dots,r_J;\boldsymbol{\alpha})$, and choosing appropriate convex conical closed functional forms for ψ and the Q_i, one can use equations (7) and (8) to generate estimable netput systems for the *ex ante–ex post* production model.

We will now proceed, by means of a series of examples, to show that a number of interesting cases can be analyzed using the above format.

Example 1: Cobb–Douglas Production with Durable Capital

Consider the classical model of a firm that has an *ex post* production function in each state s and produces a single output Y_s from a single variable input L_s and a single input K, fixed *ex post*. Then, the choice of K is the single *ex ante* design decision.

Assume the technology to be "separable across states" in the sense that the production possibilities in any state s are independent of operating decisions made for the remaining states. Let $Y_s = f(K,L_s;s)$ be a functional form defining this technology. A corresponding profit function, $\pi_s = \Pi_s(K,p_{Ys},p_{Ls})$, will specify optimal operation of the plant *ex post*, given forward prices p_{Ys}, p_{Ls} for the variable commodities in this state. One method of obtaining Π_s is to choose a functional form for the production function f and solve the *ex post* optimization problem explicitly. This procedure is satisfactory for some common production functions (e.g., Cobb–Douglas), but for others such as the C.E.S., an explicit closed-form solution for the *ex post* profit function is, in general, impossible. An alternative procedure is to specify a profit function directly, using duality theory to ensure that it is the solution to the optimization problem for some implicitly defined technology.[6] We illustrate this procedure for the Cobb–Douglas case:

The profit function

$$\pi_s = \Pi_s(K,p_{Ys},p_{Ls}) = A_{0s}K^{\alpha}p_{Ys}^{1+\beta}p_{Ls}^{-\beta} \tag{9}$$

is dual to the Cobb–Douglas production function

$$Y_s = (1+\beta)\beta^{-\beta/(1+\beta)}A_{0s}^{1/(1+\beta)}K^{\alpha/(1+\beta)}L_s^{\beta/(1+\beta)}, \tag{10}$$

where A_{0s} is an efficiency index for state s that incorporates all depreciation effects.

The *ex ante* design problem of the firm is to choose K to maximize

$$\pi = \int_S \Pi_s(K,p_{Ys},p_{Ls})\mathrm{d}\mu(s) - r\cdot K, \tag{11}$$

where r is the purchase price of capital. Now π can be written

$$\begin{aligned}\pi &= K^{\alpha}\int_S A_{0s}p_{Ys}^{1+\beta}p_{Ls}^{-\beta}\,\mathrm{d}\mu(s) - r\cdot K\\ &= a_1\cdot Q_1(\mathbf{p}) + r\cdot(-K),\end{aligned} \tag{12}$$

where $(a_1,-K)$ are the *ex post* (fixed) parameters contained in the set $\mathbf{B} = \{(a_1,-K)|a_1 = K^{\alpha},K \geqq 0\}$, and

[6]The conditions on Π_s for this construction are that it be a convex, positively linear homogeneous closed function of (p_{Ys},p_{Ls}). The specification that Y is an output and L is an input requires that Π_s be increasing in p_{Ys} and decreasing in p_{Ls}. The specification that K is an input (with a positive marginal product) requires that Π_s be increasing in K. The specification that the implicitly defined technology be convex (i.e., display generalized diminishing returns) requires that Π_s be concave in K.

$$\mathbf{p} = (\mathbf{p}_Y, \mathbf{p}_L) \quad \text{for} \quad \mathbf{p}_Y = \{p_{Ys}: s \in \mathbf{S}\}, \quad \mathbf{p}_L = \{p_{Ls}: s \in \mathbf{S}\},$$

$$Q_1(\mathbf{p}) = \int_{\mathbf{S}} A_{0s} p_{Ys}^{1+\beta} p_{Ls}^{-\beta} \, \mathrm{d}\mu(s).$$

Defining

$$\psi(Q_1, r) = \sup\{Q_1 a_1 + r(-K) | (a_1, -K) \in \mathbf{B}\}, \tag{13}$$

for $Q_1 > 0$, $r > 0$, and solving the maximization problem we have

$$\psi(Q_1, r) = \alpha^{\alpha/(1+\alpha)}(1-\alpha) r^{-\alpha/(1-\alpha)} Q_1^{1/(1-\alpha)} \tag{14}$$

$$= \hat{a}_1 Q_1(\mathbf{p}) + r(-\hat{K}), \tag{15}$$

where $\hat{a}_1$, $\hat{K}$ are the optimal *ex post* parameters. Then,

$$\frac{\partial \psi}{\partial Q_1} = \hat{a}_1 = \alpha^{\alpha/(1-\alpha)} Q_1^{\alpha/(1-\alpha)} r^{-\alpha/(1-\alpha)}$$

$$= \alpha^{\alpha/(1-\alpha)} \left[\int_{\mathbf{S}} A_{0s} p_{Ys}^{1+\beta} p_{Ls}^{-\beta} \, \mathrm{d}\mu(s) \right]^{\alpha/(1-\alpha)} r^{-\alpha/(1-\alpha)}, \tag{16}$$

$$\frac{\partial \psi}{\partial r} = -\hat{K} = -\alpha^{1/(1-\alpha)} Q_1^{1/(1-\alpha)} r^{-1/(1-\alpha)}$$

$$= -\alpha^{1/(1-\alpha)} \left[\int_{\mathbf{S}} A_{0s} p_{Ys}^{1+\beta} p_{Ls}^{-\beta} \, \mathrm{d}\mu(s) \right]^{1/(1-\alpha)} r^{-1/(1-\alpha)}. \tag{17}$$

Equations (16) and (17) specify the optimal *ex post* parameter vector $\hat{\mathbf{b}} = (\hat{a}_1, -\hat{K})$ in terms of the price vector $\mathbf{p}$ and the underlying *ex ante* parameters α, β, and $A_{0s}, s \in \mathbf{S}$. This system of equations is nonlinear in the underlying parameters and difficult to implement empirically. However, it illustrates the relationship between ψ and the *ex post* parameters that will be used in a later example to obtain a more tractable system.

Example 2: *Activity Analysis with Capacity Constraints*

Suppose a firm has available *ex ante* a finite set $\mathbf{T}^{ea}$ of possible activities $(K^j, \mathbf{x}^j)$, $j = 1, \ldots, J$, with K^j specifying the capital input per unity of capacity required by activity j, with a negative sign, and $\mathbf{x}^j$ specifying the *ex post* production plan (per-unit capacity) for activity j. Let $\mathbf{a} = (a_1, \ldots, a_J)$ be a design vector specifying the capacities in an *ex post* technology; i.e., activity j can be operated *ex post* at any intensity level up to a_j. A set $\mathbf{B}$ of possible vectors $\mathbf{a} = (a_1, \ldots, a_J)$ specifies the designs that are available *ex ante*. In the "putty–putty" case, positive capacity can be provided *ex post* for all activities (i.e., $\mathbf{B}$ is a rectangle). In the "putty–clay" case, the activities are mutually exclusive *ex post*, and, at most, one capacity can be positive (i.e., **B** is the set of vertices of a simplex). The *ex post* technology is

$$V(\mathbf{a}) = \left\{ \sum_{j=1}^{J} (a_{sj} x_s^j : s \in \mathbf{S}) | 0 \leqq a_{sj} \leqq a_j \right\},$$

where a_j is the chosen capacity of activity j, fixed *ex post* and a_{sj} is an *ex post* intensity of operation. One can also write

$$V(\mathbf{a}) = \sum_{j=1}^{J} a_j V^j, \tag{18}$$

where $V^j = \{\mathbf{x} | x = \gamma_s^j \cdot x_s^j, 0 \leqq \gamma_s^j \leqq 1\}$, implying $a_{sj} = \gamma_s^j a_j$. The intertemporal variable profit function is then

$$\pi(\mathbf{a},\mathbf{p}) = \sum_{j=1}^{J} a_j \pi^j(\mathbf{p}), \tag{19}$$

where

$$\begin{aligned} \pi^j(\mathbf{p}) &= \int_S \max(p_s \cdot x_s^j, 0) \mathrm{d}\mu(s) \\ &= \sup\{\mathbf{p} \cdot \mathbf{x} | \mathbf{x} \in V^j\}. \end{aligned} \tag{20}$$

The intertemporal total profit function then equals $\sum_{j=1}^{J} a_j [\pi^j(\mathbf{p}) + r \cdot K^j]$, and the *ex ante* envelope profit function satisfies

$$\Phi(r,\mathbf{p}) = \sup_{\mathbf{a} \in \mathbf{B}} \sum_{j=1}^{J} a_j [\pi^j(\mathbf{p}) + r \cdot K^j]. \tag{21}$$

We can define a profit function for the design set **B**,

$$\psi(Q_1,\ldots,Q_j) = \sup_{\mathbf{a} \in \mathbf{B}} \sum_{j=1}^{J} a_j \cdot Q_j, \tag{22}$$

where $Q_j = \pi^j(\mathbf{p}) + r \cdot K^j$.

Then, the *ex ante* envelope profit function has the form

$$\Phi = \psi(Q_1,\ldots,Q_J) = \sum_{j=1}^{J} \theta_j \cdot Q_j, \tag{23}$$

where θ_j satisfies the maximization problem. For example, in a putty–clay case, where the maximum designed capacity of any activity is unity,

$$\Phi = \psi(Q_1,\ldots,Q_J) = \sum_j \theta_j Q_j = \max_j Q_j = \max_j (\pi^j(\mathbf{p}) + r \cdot K^j). \tag{24}$$

Example 3: *Base Load Versus Peaking Capacity in an Electricity Generating System*

An electric utility is required to meet an output demand that has a known expected daily cycle (normalized in this example so that anticipated peak demand is equal to one), and some random variation. Let t denote time of day, scaled so that $0 \leqq t \leqq 1$, and $l(t)$ denote instantaneous output demand as a fraction of capacity (termed the system load factor) at time t. Figure 8 illustrates a typical expected load curve. The total daily output of the system, expressed as a fraction of maximum daily output, or average load factor, equals $\int_0^1 l(t)dt$, the area under the load curve in Figure 8.

For any load factor l, define $f(l)$ to be the fraction of the day for which the system load factor $l(t)$ is at least l, as illustrated in Figure 8. Clearly, f is a non-increasing function of l, with $f(0) = 1$ and $f(1) \geqq 0$. Further, the average load factor of the system equals $\int_0^1 f(l)dl$. A typical "time-at-load" function f is illustrated in Figure 9(a).

The utility can provide capacity by constructing base-load or peaking plants or by contracting to purchase power from an electricity grid. The

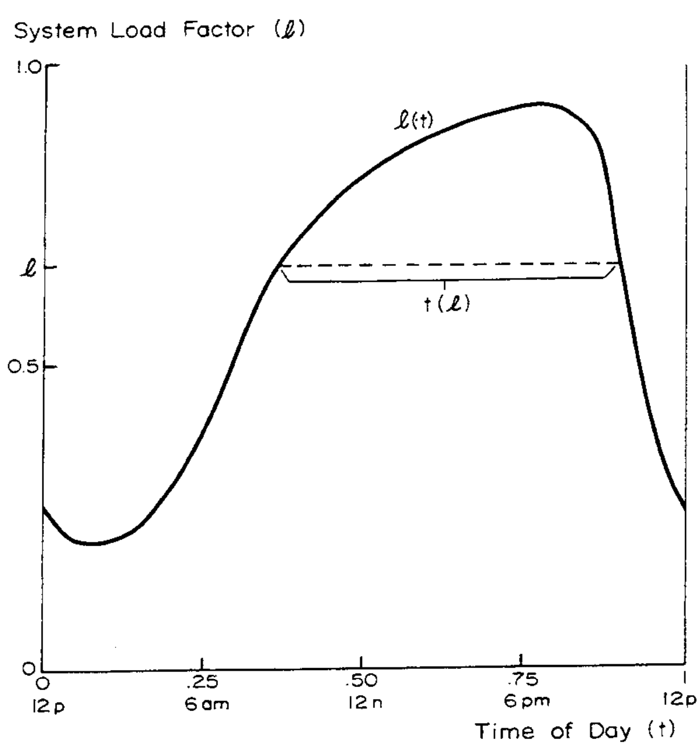


FIGURE 8

following notation is employed for the costs of these production modes:

	Base-load plant	Peaking plant	Grid purchase
Initial capital cost per unit of capacity ($/kw)	k_b	k_p	0
Present value of operating cost ($/kw)	c_b	c_p	c_g

The *ex ante* parameters of the supply process are subsumed in the unit costs, for differences in these costs reflect basic underlying differences in the possible methods of "producing" electricity.

For simplicity, we assume that these costs are constant, independent of plant scale and of the load curve. We also assume that the utility has no discretionary control over sales to the electricity grid.

Of the three production modes, base-load plant has the highest capital cost and the lowest total cost of providing continuous capacity output, while grid purchase is the most expensive mode for providing continuous output: $k_b > k_p$ and $k_b + c_b < k_p + c_p < c_g$. Suppose that *ex post*, the utility has a proportion α of its system capacity in base-load plant and a proportion β in combined base-load and peaking plant. In optimal operation, demand will be met first by base-load plant, second by peaking plant (for load factors above α), and last by purchased power (for load factors above β). Present value of total cost per unit of system capacity then equals[7]

$$C = \alpha k_b + (\beta - \alpha) k_p + c_p \int_0^\alpha f(l)\mathrm{d}l \pm c_p \int_\alpha^\beta f(l)\mathrm{d}l + c_g \int_\beta^1 f(l)\mathrm{d}l. \tag{25}$$

[7]More realistic assumptions on unit costs would be that there are some increasing returns to plant scale, particularly in small spatially concentrated utilities, and that marginal future operating costs for base-load and peaking plants depend on plant load factors. (There are generally decreasing marginal costs with load in a given unit, but increasing marginal costs in the system as successively less efficient units are operated at capacity.) The construction of the formula (25) is illustrated by derivation of the variable cost of base-load plant operation. Let $c_b(\lambda)$ equal the marginal cost of operating a unit capacity base-load plant at load factor λ. Then, $\gamma(\lambda) = \int_0^\lambda c_b(\lambda')\mathrm{d}\lambda'$ is the total variable cost of operating this plant at load factor λ. Suppose the utility has a system load curve $l(t)$, a proportion α of base-load capacity, and the policy of operating base-load capacity first. (Provided the marginal cost of peaking-plant operation exceeds that of base-load plant operation at any respective plant-load factors, this policy is always optimal.) Then the base-load plant is operated with load curve $\lambda(t) = \min((l(t)/\alpha), 1)$. Retaining the assump-

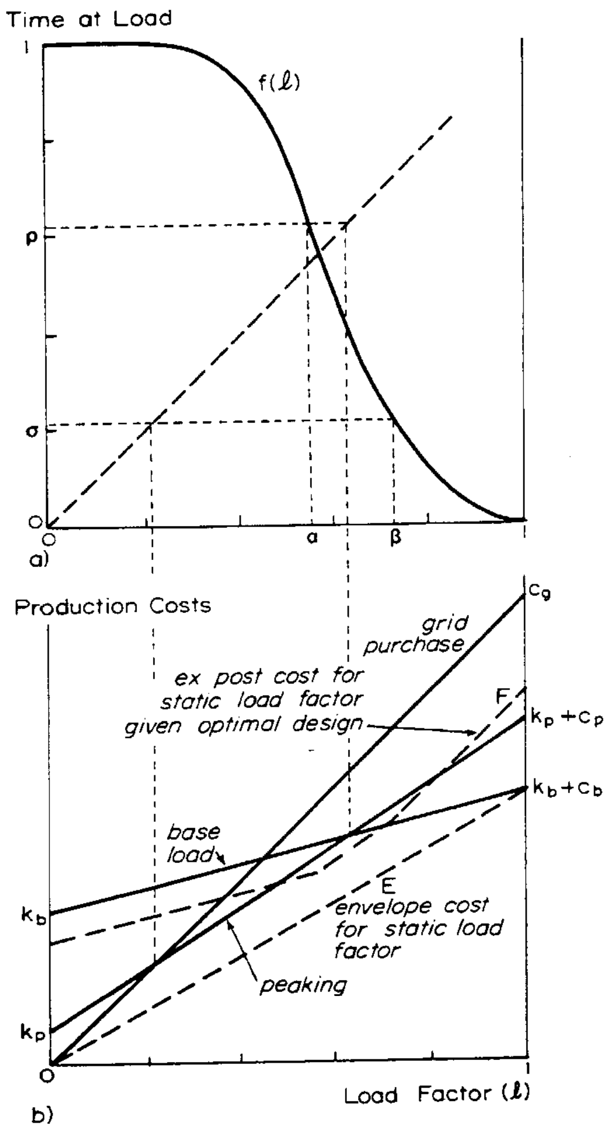


FIGURE 9

tion that marginal cost is independent of plant scale, the total variable cost of operating the base-load plant at load factor $\lambda(t)$, divided by system capacity, equals $\alpha\gamma(\lambda(t))$. Hence, the variable cost of base-load plant operation is

$$\int_0^1 \alpha\gamma(\lambda(t))\mathrm{d}t = \alpha \int_{t=0}^{1}\int_{\lambda=0}^{\min((l(t)/\alpha),1)} c_b(\lambda)\mathrm{d}\lambda\,\mathrm{d}t = \alpha \int_{\lambda=0}^{1}\int_{t=t_1(\alpha\lambda)}^{t_2(\alpha\lambda)} \mathrm{d}t\, c_b(\lambda)\mathrm{d}\lambda$$

$$= \alpha \int_{\lambda=0}^{1} f(\alpha\lambda)c_b(\lambda)\mathrm{d}\lambda = \int_0^{\alpha} f(l)c_b(l/\alpha)\mathrm{d}l,$$

where the second equality is obtained by interchanging the order of integration (see Figure 8), and the third equality follows from the definition of $f(l)$. When $c_b(l/\alpha)$ is constant, the corresponding term in (25) is obtained.

Define $\rho = (k_b - k_p)/(c_p - c_b)$ and $\sigma = k_p/(c_g - c_p)$. Note from Figure 9(b) that ρ is the load factor at which unit capacity base-load and peaking plants would yield the same total cost. Similarly, σ is the load factor at which purchased power and a unit capacity peaking plant yield the same total cost. Consider, now, the *ex ante* minimization of C in the design parameters α and β. Note that $\partial C/\partial\alpha = (c_p - c_b)(\rho - f(\alpha))$ and $\partial C/\partial\beta = (c_g - c_p)(\sigma - f(\beta))$. We confine our attention to the case $\sigma \leqq \rho$ illustrated in Figure 9(b) in which there is a range of load factors for which a unit capacity peaking plant will yield lower total cost than either a unit capacity base-load plant or purchased power. Then the *ex ante* minimum occurs for α and β satisfying $f(\alpha) = \rho$, $f(\beta) = \sigma$. The determination of these quantities is illustrated in Figure 9(a).[8]

The total cost function (25) is nonlinear in the design parameters α, β. However, we can reparameterize *ex ante* design possibilities to make total cost linear in design parameters. Define **B** to be the set of vectors $\mathbf{a} = (a_1,\ldots,a_5) = (\alpha, \beta - \alpha, \int_0^\alpha f(l)\mathrm{d}l, \int_\alpha^\beta f(l)\mathrm{d}l, \int_\beta^1 f(l)\mathrm{d}l)$ for $0 \leqq \alpha \leqq \beta \leqq 1$, and define $\psi(Q_1,\ldots,Q_5) = \min\{\sum_i Q_i a_i | \mathbf{a} \in \mathbf{B}\}$. Then, $C = a_1k_b + a_2k_p + a_3c_b + a_4c_p + a_5c_g$, and the *ex ante* envelope cost function equals $\psi(k_b,k_p,c_b,c_p,c_g)$. Further, ψ has the property that its derivative with respect to Q_i, evaluated at (k_b,k_p,c_b,c_p,c_g), equals the optimal value of the corresponding design parameter $\hat{a}_i$.

It is of interest to compare the cost curves generated by this concrete model with the classic curves in Figures 1 and 2. A system with unit capacity, all base-load, has the total cost curve $k_b + c_b l$ in Figure 9(b). A least-cost unit capacity system to produce an output θ uniform in time has base-load capacity θ, zero peaking capacity, and purchased power capacity $1 - \theta$. The line E is the envelope of such least-cost unit capacity systems for various θ. The curve F in this figure is the total cost curve for the optimal system with the "time-at-load" curve illustrated in Figure 9(a). Note that there is no single output uniform in time for which this system remains optimal. Note, further, that a shift in the load curve that increases variability while keeping total output constant will generally decrease the proportion of optimal base-load capacity and increase the proportion of peaking capacity.

The reader may find it useful to verify these conclusions for an example. Suppose the utility has the system "time-at-load" function illustrated in Figure 10(b), which yields an average load u and a variance

[8]In the case $\sigma > \rho$, no peaking capacity will be constructed ($\alpha = \beta$), and optimal α satisfies $f(\alpha) = k_b/(c_g - c_b)$.

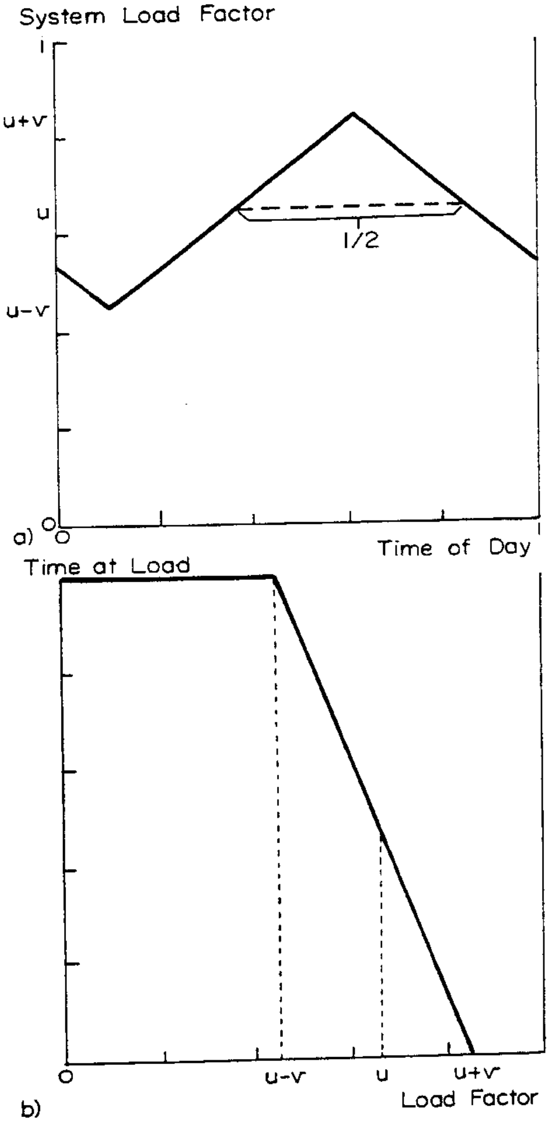


FIGURE 10

in load $v^2/3$. Minimization of $\Sigma\, Q_i a_i$ over $\mathbf{a} \in \mathbf{B}$ yields

$$\psi(Q_1,\ldots,Q_5) = (Q_1 + Q_3)u - Q_2^2/(Q_5 - Q_4)], \\ + v[Q_1 - (Q_1 - Q_2)^2/(Q_4 - Q_3) \qquad (26)$$

in the case $(Q_1 - Q_2)/(Q_4 - Q_3) > Q_2/(Q_5 - Q_4)$, and an optimal *ex ante* design $\hat{\alpha} = u + v - 2v(k_b - k_p)/(c_p - c_b)$ and $\hat{\beta} = u + v - 2vk_p/(c_g - c_p)$. Note that an increase in the variance of the load with average load fixed will always increase peaking capacity. Under normal conditions, unit costs satisfy $\sigma < \rho < 1/2$, implying that base load capacity will increase and grid purchases will decrease as load variance rises.

Example 4: A Nested Ex Ante–Ex Post Functional Form for the Two-Level Technology

The model presented below was developed by the authors as an extension of Diewert's generalized Leontief cost function. It has been analyzed extensively in Fuss (1970, 1977b), and its parameters estimated by Fuss in Chapter IV.4, using electricity generation data, for the case of intertemporal variation with no uncertainty. We summarize the model's basic characteristics. The reader is referred to the cited works for a more detailed description.

For this example the "*a,b*" notation is reversed from that in the rest of the paper in order to retain consistency with the cited literature. Thus, for this example *only*, **b** refers to a vector of *ex post* fixed design variables (rather than **a**), and **a** refers to a vector of characteristics of the underlying technology.

Suppose a producer expects to use n variable factors to produce one unit of output in each future time period t and state of nature u_t in period t, where factor prices vary with u_t in each period and the u_t in different periods are statistically independent. A future state is then a vector $\mathbf{s} = (t,u_t : t = 1,2,\ldots)$. Suppose the *ex post* variable cost function in state s is Diewert's second-order approximation to an arbitrary unit cost function,

$$C_t\left(\tilde{p}_{1tu_t},\ldots,\tilde{p}_{ntu_t}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\left(\tilde{p}_{itu_t}\tilde{p}_{jtu_t}\right)^{1/2}, \tag{27}$$

where b_{ij} is an *ex post* parameter and $\tilde{p}_{itu_t}$ is a future price. The present value of expected cost is

$$C = \sum_t \delta^t E_{u_t}(C_t) = \sum_i \sum_j b_{ij}Q_{ij}, \tag{28}$$

where

$$Q_{ij} = \sum_t \delta^t E_{u_t}\left(\tilde{p}_{itu_t}\tilde{p}_{jtu_t}\right)^{1/2} = \sum_t E_{u_t}\left(p_{itu_t}p_{jtu_t}\right)^{1/2},$$

and p_{itu_t} is a forward price. Note that C is linear in the *ex post* parameters since Q_{ij} is a function of prices alone and is analogous to the terms Q_i appearing in the previous examples. It has been shown in Fuss (1977b) that if we specify as a functional form for the *ex post* parameters

$$b_{ij} = \left(\frac{1}{d_{ij}}\right) \sum_{k,l=1,\dots,n} (d_{kl}d_{ij})^{1/2} a_{ijkl}, \tag{29}$$

where d_{ij}, d_{kl} are arbitrary non-negative "design" variables and a_{ijkl} are fixed *ex ante* parameters, then C is minimized when

$$\hat{b}_{ij} = \frac{1}{Q_{ij}} \sum_k \sum_l (Q_{kl}Q_{ij})^{1/2} a_{ijkl}. \tag{30}$$

Note that the optimal *ex post* parameters are linear in the *ex ante* parameters.

We can now describe this structure in the format developed in the previous examples. The *ex ante* envelope cost function is

$$C(\{Q_{ij}\} | i,j = 1,\dots,n) = \min_{\mathbf{B}} \left\{ \sum_{i,j} Q_{ij} \cdot b_{ij} \right\}, \tag{31}$$

where the design parameters are contained in the set

$$\mathbf{B} = \left[\left\{ b_{ij} = \frac{1}{d_{ij}} \sum_{k,l} (d_{kl}d_{ij})^{1/2} a_{ijkl} \right\} \middle| d_{ij}, d_{kl} \geqq 0 \right].$$

The solution to this minimization problem is

$$\begin{aligned} C(\{Q_{ij}\}) &= \sum_{ij} Q_{ij} \cdot \hat{b}_{ij} \\ &= \sum_{ij} \sum_{kl} (Q_{ij}Q_{kl})^{1/2} a_{ijkl}, \end{aligned} \tag{32}$$

where $\hat{b}_{ij}$ is given by (30). Note that C is linear in the *ex ante* parameters and is therefore amenable to estimation using standard regression techniques.

Using the theory outlined at the beginning of this section, we could have begun with the *ex ante* envelope cost function $C = \sum_{i,j,k,l} (Q_{ij}Q_{kl})^{1/2} a_{ijkl}$ and derived the optimal *ex post* parameters and netput bundles. That is,

$$\hat{b}_{ij} = \frac{\partial C}{\partial Q_{ij}} = \frac{1}{Q_{ij}} \sum_{kl} (Q_{kl}Q_{ij})^{1/2} a_{ijkl},$$

and

$$\hat{x}_{itu_t} = \frac{\partial C}{\partial [p_{itu_t}]} = \sum_j \frac{\partial C}{Q_{ij}} \cdot \frac{\partial Q_{ij}}{\partial [p_{itu_t}]} = \sum_j \hat{b}_{ij} \cdot \frac{\partial Q_{ij}}{\partial [p_{itu_t}]}.$$

If the p_{it} are assumed known with certainty, then

$$\hat{x}_{it} = \sum_j \hat{b}_{ij} \left(\frac{p_{jt}}{p_{it}}\right)^{1/2}, \tag{33}$$

since

$$\frac{\partial Q_{ij}}{\partial p_{it}} = \left(\frac{p_{jt}}{p_{it}}\right)^{1/2}.$$

Then, using (30),

$$\hat{x}_{it} = \sum_{j,k,l} \left[\left(\frac{Q_{kl}}{Q_{ij}}\right)^{1/2} \left(\frac{p_{jt}}{p_{it}}\right)^{1/2}\right] a_{ijkl}, \tag{34}$$

which is the expression for the *ex post* optimal inputs found in Fuss (1977b).

The structure above does not identify explicitly inputs of capital equipment whose levels are set *ex ante*. With some added notation, this can be done as follows: Suppose, of the n "variable" factors above, the first J are identified as inputs of capital equipment. Define the set of indices $N = \{(i,j) | i = j \leqq J \text{ or } J < i,j \leqq n\}$. Define $Q_{jj} = r$, the price of capital good j, for $j = 1,\dots,J$, and define Q_{ij} as above for $i,j > J$. Define the *ex ante* envelope cost function

$$C = \sum_{ij\in N} \sum_{kl\in N} (Q_{ij}Q_{kl})^{1/2} a_{ijkl}.$$

Then, the optimal *ex post* parameters satisfy $\partial C / \partial Q_{jj} = \hat{K}_j$ for $j = 1,\dots,J$; $\partial C / \partial Q_{ij} = \hat{b}_{ij}$ for $i,j > J$; and $\partial C / \partial p_{itu_t} = \hat{x}_{itu_t}$ for $i > J$.

We shall now illustrate that this functional form can be used to specify the flexibility–efficiency tradeoff by considering a simplified two-factor example with one operating period.

Suppose the forward prices p_{1u}, p_{2u} satisfy $E_u(p_{iu}^{1/2}) = \mu$, $Q_{12} = E_u(p_{1u}p_{2u})^{1/2} = \mu^2$, and $Q_{ii} = E_u(p_{iu}) = \mu^2 + \sigma^2 \equiv \mu^2\beta^2$. Then, the two price series are uncorrelated, and each has a variance $\sigma^2 = \mu^2(\beta^2 - 1)$. (Thus, $\beta \geqq 1$ is an increasing index of price variability.) For further simplicity, assume the underlying *ex ante* parameters a_{ijkl} take the values in the following table:

		Index ij			
		11	12	21	22
Index kl	11	0	1/2	1/2	A
	12	1/2	A	A	1/2
	21	1/2	A	A	1/2
	22	A	1/2	1/2	0

where A is a non-negative scalar. From equation (30), $\hat{b}_{11} = \hat{b}_{22} = A + \beta^{-1}$ and $\hat{b}_{12} = \hat{b}_{21} = A + \beta$. For a given vector of forward prices (p_{1u}, p_{2u}), the input demands are

$$\hat{x}_{1u} = \hat{b}_{11} + \hat{b}_{12}(p_{2u}/p_{1u})^{1/2},$$
$$\hat{x}_{2u} = \hat{b}_{22} + \hat{b}_{21}(p_{1u}/p_{2u})^{1/2}. \tag{35}$$

Evaluated at the "mean" price vector $(p_{1s}, p_{2s}) = (\mu, \mu)$,

$$\hat{x}_{1u} = \hat{x}_{2u} = 2A + \beta^{-1} + \beta. \tag{36}$$

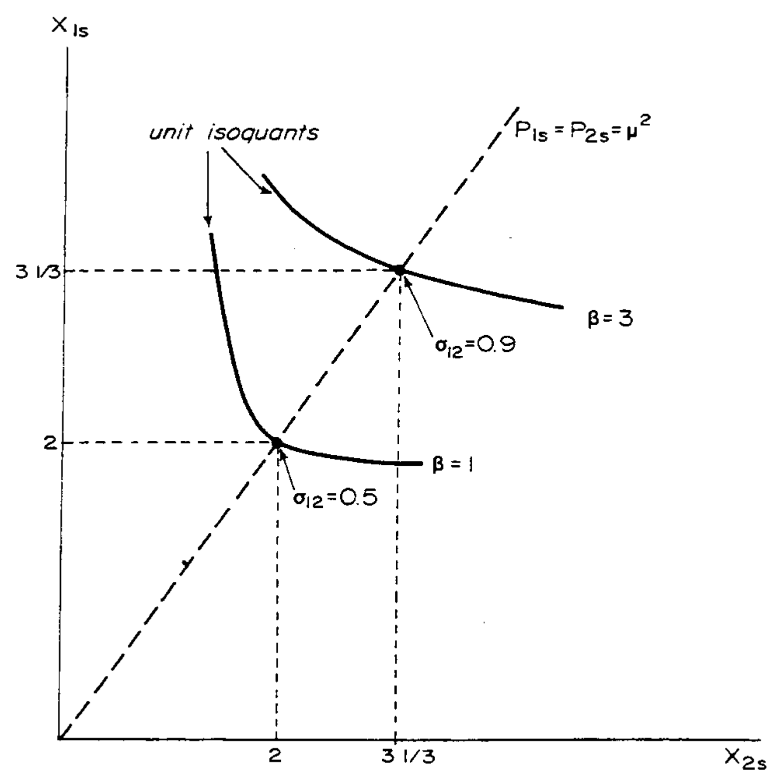


FIGURE. 11

The *ex post* elasticity of factor substitution between inputs 1 and 2 is given by

$$\text{ES} = (p_{1u}\hat{x}_{1u} + p_{2u}\hat{x}_{2u})(\partial\hat{x}_{1u}/\partial p_{2u})/\hat{x}_{1u}\hat{x}_{2u}. \tag{37}$$

Evaluated at $(p_1,p_2) = (E_u p_{1u}, E_u p_{2u}) = (\mu^2\beta^2, \mu^2\beta^2)$, this elasticity is

$$\text{ES} = (A+\beta)/(2A+\beta+\beta^{-1}). \tag{38}$$

As $\beta \geqq 1$ is increased (price variability rises), ES increases (flexibility rises) and $\hat{x}_{1u}$ evaluated at the price vector $(E_u p_{1u}, E_u p_{2u})$ also increases (static efficiency falls). Figure 11 illustrates unit isoquants of the *ex post* technology for levels of β in the case $A = 0$.

Define as an index of static efficiency the ratio e of the input level (36) at $\beta = 1$ to this input level at the value of β corresponding to a given level of price variability. Figure 12 illustrates the tradeoff between *ex post* flexibility measured by the elasticity of substitution and this index of static efficiency. Note that as A increases, the possibility of substituting flexibility for efficiency falls.

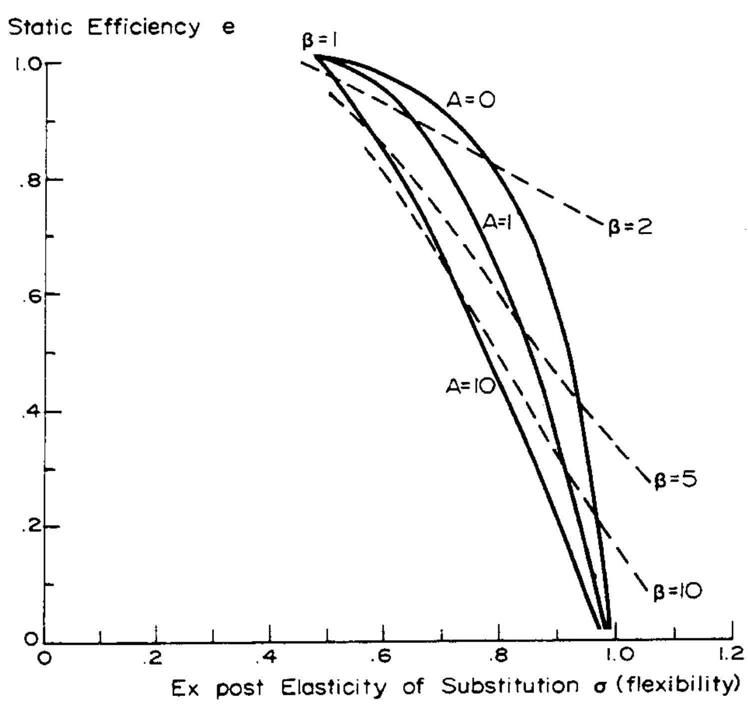


FIGURE 12

5. Derivation of the Two-Level Structure of Technology

In this section we use the model of an *ex ante–ex post* technology described in Section 3 and develop analytically the algorithm for generating econometric netput supply systems illustrated in Section 4. These examples have the common feature that *ex ante* design possibilities can be parameterized so that the *ex post* intertemporal total profit function is *linear* in the *ex ante* design variables (*ex post* parameters). It is this property and the duality of technologies and profit functions that we exploit to develop the desired algorithm. We define, on one hand, a family of *ex ante–ex post* technologies with the linear structure above and, on the other hand, a family of "nested" profit functions. The operations of profit maximization and of construction of the set of production plans consistent with profit maximization are shown to define one-to-one mutually inverse mappings between these families. Then, each "nested" profit function characterizes some *ex ante–ex post* technology. The algorithm is then to choose an appropriate functional form for a nested-profit function and use the derivative property to obtain expressions for the optimal *ex ante* design and optimal *ex post* netput supply vector. Properties of the associated technology are obtained implicitly from the profit function, using the duality mappings. We consider, first, the case in which the set of future states and the vector of *ex ante* design variables are finite.

5.1. *The Finite Case*

Suppose there are a finite number of future states, indexed $s = 1,...,S$. Suppose that an *ex ante* plant design $\mathbf{b} = (\mathbf{a},\mathbf{K})$ is composed of finite vectors $\mathbf{a} = (a_1,...,a_L)$ and $\mathbf{K} = (K_1,...,K_J)$. A *technological structure* is defined by an *ex ante* envelope technology $\mathbf{T}^{ea}$, a non-empty subset of $\mathbf{E}^{L+J+N\cdot S}$, and the following associated sets: the set of *ex ante* designs $\mathbf{B} = \{(\mathbf{a},\mathbf{K}) \in \mathbf{E}^{L+J} | \exists \mathbf{x} \ni (\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea}\}$, the *ex post* variable technology $\mathbf{V}(\mathbf{a},\mathbf{K}) = \{\mathbf{x} \in \mathbf{E}^{N\cdot S} | (\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea}\}$ defined for $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$, an *ex ante* netput possibility set $\mathbf{W} = \{(\mathbf{K},\mathbf{x}) \in \mathbf{E}^{J+N\cdot S} | \exists \mathbf{a} \ni (\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea}\}$, and the *normal cone* of $\mathbf{W}$, the set $\mathbf{F} \subseteq \mathbf{E}^{J+N\cdot S}$ of normals to hyperplanes whose lower half-spaces contain $\mathbf{W}$; i.e., $\mathbf{F}$ is the set of $(\mathbf{r},\mathbf{p}) \in \mathbf{E}^{J+N\cdot S}$ such that $\mathbf{r}\cdot\mathbf{K} + \mathbf{p}\cdot\mathbf{x}$ is bounded above for $(\mathbf{K},\mathbf{x}) \in \mathbf{W}$.

We term a technological structure *regular* if the following conditions hold: (1) the sets $\mathbf{B}$, $\mathbf{W}$, and $\mathbf{V}(\mathbf{a},\mathbf{K})$ for $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$ are closed, (2) capital inputs are non-positive and exhibit free disposal; i.e., $\mathbf{K} \leqq \mathbf{0}$ for all $(\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea}$, and $(\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea}$, $\mathbf{K}' \leqq \mathbf{K}$ implies $(\mathbf{a},\mathbf{K}',\mathbf{x}) \in \mathbf{T}^{ea}$, and (3) the normal cone $\mathbf{F}$ has a non-empty interior (denoted by $\mathbf{F}_0$). Note that $\mathbf{F}$ is the set of prices that yield finite maximum profits in the *ex ante* envelope technology. Condition (2), above, states our accounting convention for capital inputs. The free disposal assumption implies that r_j is non-negative for $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$, with $r_j K_j \leqq 0$ for all $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$. The condition that $\mathbf{F}_0$ be non-empty holds if and only if the set $\mathbf{W}$ is semibounded (see Chapter I.1). A sufficient condition for $\mathbf{F}_0$ non-empty is that there be some commodities that are essential inputs to production and cannot themselves be produced in the technology. It is *not* necessary for $\mathbf{W}$ to be convex in order to have $\mathbf{F}_0$ non-empty; however, $\mathbf{W}$ cannot exhibit indefinitely increasing returns to the extent that "average products" are unbounded with unbounded scale. When $\mathbf{T}^{ea}$ is convex, then $\mathbf{W}$, $\mathbf{B}$, and $\mathbf{V}(\mathbf{a},\mathbf{K})$ are convex and we say the technological structure is *convex*.

Suppose the technological structure is such that the *ex post* variable technology is linear in the parameter vector $\mathbf{a}$ and not explicitly dependent on $\mathbf{K}$. (An implicit dependence on $\mathbf{K}$ results from the relation of $\mathbf{a}$ and $\mathbf{K}$ in the set $\mathbf{B}$.) Then we can write

$$\mathbf{V}(\mathbf{a},\mathbf{K}) = \sum_{l=1}^{L} a_l \mathbf{V}^l = \left\{ \sum_{l=1}^{L} a_l \mathbf{x}^l \middle| \mathbf{x}^l \in \mathbf{V}^l \right\}, \tag{39}$$

where the $\mathbf{V}^l$ are non-empty subsets of $\mathbf{E}^{N \cdot S}$. We term the technological structure *design linear* if it satisfies equation (39) and if for each $l = 1,\ldots,L$, $\mathbf{V}^l$ is a closed set and either a_l is non-negative for all $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$ or $\mathbf{V}^l$ is a singleton.

A *regular profit structure* is defined by: (1) a convex cone $\mathbf{F}$ of price vectors $(\mathbf{r},\mathbf{p}) \in \mathbf{E}^{J+N \cdot S}$ such that its interior $\mathbf{F}_0$ is non-empty and r_j is non-negative for all $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$, $j = 1,\ldots,J$; (2) a non-empty closed set $\mathbf{B}$ of *ex ante* design variables $(\mathbf{a},\mathbf{K}) \in \mathbf{E}^{L+J}$ such that $\mathbf{K} \leqq \mathbf{0}$ for $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$; (3) a convex conical closed function of $\mathbf{p}$, $\Pi(\mathbf{a},\mathbf{K},\mathbf{p})$, defined for $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$ and $\mathbf{p}$ in the set $\mathbf{F}^p = \{\mathbf{p} \in \mathbf{E}^{N \cdot S} | \exists \mathbf{r} \ni (\mathbf{r},\mathbf{p}) \in \mathbf{F}\}$; and (4) a convex conical closed function $\Phi(\mathbf{r},\mathbf{p}) = \sup\{\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) + \mathbf{r} \cdot \mathbf{K} | (\mathbf{a},\mathbf{K}) \in \mathbf{B}\}$ defined for $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$. The function Π is interpreted as an intertemporal variable profit function, while Φ is interpreted as an *ex ante* envelope profit function. We term a regular profit structure *convex* if the set $\mathbf{B}$ is convex.

A regular profit structure is termed *design linear* if the intertemporal variable profit function can be written

$$\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) = \sum_{l=1}^{L} a_l Q^l(\mathbf{p}), \tag{40}$$

where for each $l = 1,\dots,L$, Q^l is a convex [resp., linear] conical closed function of $\mathbf{p}$ when a_l is non-negative [resp., bivalent] for all $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$.

Define a *nested profit form* by (1) a convex cone $\mathbf{F}$ of vectors $(\mathbf{r},\mathbf{p}) \in \mathbf{E}^{J+N\cdot S}$ with a non-empty interior and with r_j non-negative for $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$, $j = 1,\dots,J$; (2) a convex cone $\mathbf{H}$ of vectors $(\mathbf{r},\mathbf{q}) \in \mathbf{E}^{J+L}$; (3) a convex conical closed function $\psi(\mathbf{r},\mathbf{q})$ on $\mathbf{H}$ which is non-increasing in r_j for all $(\mathbf{r},\mathbf{q}) \in \mathbf{H}$ and which is non-decreasing [resp., non-monotone] in q_l for l in a set of indices $\mathbf{L}_+$ [resp., $\mathbf{L}_0$]; and (4) a vector of functions $\mathbf{Q}(\mathbf{p}) = (Q^1(\mathbf{p}),\dots,Q^L(\mathbf{p}))$ on $\mathbf{F}^p$ such that $(\mathbf{r},\mathbf{q}) \in \mathbf{H}$ if and only if there exists $\mathbf{p} \in \mathbf{F}^p$ with $\mathbf{q} = \mathbf{Q}(\mathbf{p})$ and $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$, and such that $Q^l(\mathbf{p})$ is a convex [resp., linear] conical closed function for $l \in \mathbf{L}_+$ [resp., $\mathbf{L}_0$]. The following result links design linear profit structures and nested profit forms.

Lemma 1. Consider a regular, design linear profit structure satisfying (40), and define

$$\mathbf{H} = \{(\mathbf{r},\mathbf{q}) \in \mathbf{E}^{J+L} | \mathbf{q} = \mathbf{Q}(\mathbf{p}) \text{ for some } \mathbf{p} \in \mathbf{F}^p \text{ with } (\mathbf{r},\mathbf{p}) \in \mathbf{F}\}, \quad (41)$$

$$\psi(\mathbf{r},\mathbf{q}) = \sup\{\mathbf{r}\cdot\mathbf{K} + \mathbf{q}\cdot\mathbf{a} | (\mathbf{a},\mathbf{K}) \in \mathbf{B}\} \text{ for } (\mathbf{r},\mathbf{q}) \in \mathbf{H}. \quad (42)$$

Then, $\mathbf{F}$, $\mathbf{H}$, ψ, $\mathbf{Q}$ define a nested profit form, and for $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$,

$$\Phi(\mathbf{r},\mathbf{p}) = \psi(\mathbf{r},\mathbf{Q}(\mathbf{p})). \quad (43)$$

Conversely, given a nested profit form $\mathbf{F}$, $\mathbf{H}$, ψ, $\mathbf{Q}$, define

$$\mathbf{B} = \{(\mathbf{a},\mathbf{K}) \in \mathbf{E}^{L+J} | \mathbf{q}\cdot\mathbf{a} + \mathbf{r}\cdot\mathbf{K} \leqq \psi(\mathbf{r},\mathbf{q}) \text{ for all } (\mathbf{r},\mathbf{q}) \in \mathbf{H}\}, \quad (44)$$

$$\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) = \sum_{l=1}^{L} a_l Q^l(\mathbf{p}) \quad \text{for} \quad (\mathbf{a},\mathbf{K}) \in \mathbf{B}, \quad (45)$$

and Φ satisfying (43). Then, $\mathbf{B}$, Π, Φ, $\mathbf{F}$ define a regular, convex, design linear profit structure satisfying (41) and (42).

Proof: $\mathbf{F}$ non-empty implies $\mathbf{H}$ non-empty. From the definition of $\mathbf{H}$, the right-hand side of (42) is bounded above by $\Phi(\mathbf{r},\mathbf{p})$ for some $\mathbf{p}$ with $\mathbf{q} = \mathbf{Q}(\mathbf{p})$. Hence, ψ exists on $\mathbf{H}$, and is convex conical closed by McFadden, Appendix A.3, Lemma 12.3. If $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$, then $(\mathbf{r},\mathbf{q}) \in \mathbf{H}$ for $\mathbf{q} = \mathbf{Q}(\mathbf{p})$, and $\Phi(\mathbf{r},\mathbf{p}) = \sup\{\mathbf{r}\cdot\mathbf{K} + \Pi(\mathbf{a},\mathbf{K},\mathbf{p}) | (\mathbf{a},\mathbf{K}) \in \mathbf{B}\} = \sup\{\mathbf{r}\cdot\mathbf{K} + \mathbf{a}\cdot\mathbf{Q}(\mathbf{p}) | (\mathbf{a},\mathbf{K}) \in \mathbf{B}\} = \psi(\mathbf{r},\mathbf{Q}(\mathbf{p}))$. The function $\mathbf{r}\cdot\mathbf{K} + \mathbf{a}\cdot\mathbf{Q}(\mathbf{p})$ is convex conical closed for each $(\mathbf{a},\mathbf{k}) \in \mathbf{B}$. The supremum of an arbitrary family of convex conical closed functions on $\mathbf{F}$ is again a convex conical closed function.

Next, suppose a nested profit form **F**, **H**, ψ, **Q** is given. The fundamental duality theorem for restricted profit functions (Chapter I.1, Lemma 11, Lemma 23, Theorem 24) establishes that the set **B** given by (44) is non-empty, closed, and convex and satisfies (42). The function Π defined by (45) is a sum of convex conical closed functions $a_l\mathbf{Q}^l(\mathbf{p})$, and hence also has these properties. By (42), $\psi(\mathbf{r},\mathbf{Q}(\mathbf{p})) = \sup\{\mathbf{r}\cdot\mathbf{K} + \mathbf{a}\cdot\mathbf{Q}(\mathbf{p})|(\mathbf{a},\mathbf{K}) \in \mathbf{B}\} = \sup\{\mathbf{r}\cdot\mathbf{K} + \Pi(\mathbf{a},\mathbf{K},\mathbf{p})|(\mathbf{a},\mathbf{K}) \in \mathbf{B}\}$. Hence, by the same argument as in the previous paragraph, $\psi(\mathbf{r},\mathbf{Q}(\mathbf{p}))$ is convex conical closed, and Φ defined by (43), Π, **F**, **B** define a regular convex design linear profit structure. Q.E.D.

The next result links profit structures and technological structures.

Lemma 2. Consider a regular (convex) technological structure and define

$$\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{x}|\mathbf{x} \in \mathbf{V}(\mathbf{a},\mathbf{K})\} \quad \text{for} \quad \mathbf{p} \in \mathbf{F}^p, \quad (\mathbf{a},\mathbf{K}) \in \mathbf{B}, \tag{1}$$

$$\Phi(\mathbf{r},\mathbf{p}) = \sup\{\mathbf{r}\cdot\mathbf{K} + \mathbf{p}\cdot\mathbf{x}|(\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea} \quad \text{for} \quad (\mathbf{r},\mathbf{p}) \in \mathbf{F}. \tag{3}$$

Then, **F**, **B**, Π, Φ define a regular (convex) profit structure. Conversely, given a regular (convex) profit structure, define

$$\mathbf{V}(\mathbf{a},\mathbf{K}) = \{\mathbf{x} \in \mathbf{E}^{N\cdot S}|\mathbf{p}\cdot\mathbf{x} \leqq \Pi(\mathbf{a},\mathbf{K},\mathbf{p}) \quad \text{for all} \quad \mathbf{p} \in \mathbf{F}^p\} \quad \text{for} \quad (\mathbf{a},\mathbf{K}) \in \mathbf{B}, \tag{46}$$

$$\mathbf{T}^{ea} = \{(\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{E}^{L+J+N\cdot S}|\mathbf{x} \in \mathbf{V}(\mathbf{a},\mathbf{K}), (\mathbf{a},\mathbf{K}) \in \mathbf{B}\}, \tag{47}$$

$$\mathbf{W} = \{(\mathbf{K},\mathbf{x}) \in \mathbf{E}^{J+N\cdot S}|\exists \mathbf{a} \ni (\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{T}^{ea}\} \tag{48a}$$

$$= \{(\mathbf{K},\mathbf{x}) \in \mathbf{E}^{J+N\cdot S}|\mathbf{r}\cdot\mathbf{K} + \mathbf{p}\cdot\mathbf{x} \leqq \Phi(\mathbf{r},\mathbf{p}) \text{ for all } (\mathbf{r},\mathbf{p}) \in \mathbf{F}\}. \tag{48b}$$

Then, the sets **V(a,K)** and **W** are convex, and $\mathbf{T}^{ea}$, **B**, **V(a,K)**, **W**, **F** define a regular (convex) technological structure such that (1) and (3) hold.

Proof: The fundamental duality theorem for restricted profit functions (Chapter I.1, Lemma 11, Lemma 23, Theorem 24) establishes the dual properties of (1) and (46), or the dual properties of $\Phi(\mathbf{r},\mathbf{p}) = \sup\{\mathbf{r}\cdot\mathbf{K} + \mathbf{p}\cdot\mathbf{x}|(\mathbf{K},\mathbf{x}) \in \mathbf{W}\}$, equivalent to (3) and (48). The properties of $\mathbf{T}^{ea}$ defined by (47) follow from this equivalence. Q.E.D.

The next result relates design linear technological and profit structures:

Lemma 3. Suppose a regular (convex) design linear technological structure satisfies (39). Then the regular (convex) profit structure given by (1) and (3) is design linear, satisfying (40) with

$$Q^l(\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{x} | \mathbf{x} \in \mathbf{V}^l\}, \tag{49}$$

for $\mathbf{p} \in \mathbf{F}^p$, $l = 1,...,L$. Conversely, suppose a regular (convex) design linear profit structure satisfies (40). Then, the regular (convex) technological structure given by (46)–(48) is design linear, satisfying (39) with

$$\mathbf{V}^l = \{\mathbf{x} \in \mathbf{E}^{N\cdot S} | \mathbf{p}\cdot\mathbf{x} \leqq Q^l(\mathbf{p}) \text{ for } \mathbf{p} \in \mathbf{F}^p\}, \tag{50}$$

for $l = 1,...,L$. The $\mathbf{V}^l$ are convex, and satisfy (1), (3), (39), and (49).

Proof: Let $\mathbf{L}_+$ (resp., $\mathbf{L}_0$) denote the set of indices $l = 1,...,L$, such that a_l is non-negative (resp., bivalent) for all $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$.

For a design linear technological structure,

$$\begin{aligned} \Pi(\mathbf{a},\mathbf{K},\mathbf{p}) &= \sup\left\{\mathbf{p}\cdot\sum_{l=1}^{L} a_l\mathbf{x}^l | \mathbf{x}^l \in \mathbf{V}^l\right\} \\ &= \sum_{l\in\mathbf{L}_+} a_l \sup\{\mathbf{p}\cdot\mathbf{x}^l | \mathbf{x}^l \in \mathbf{V}^l\} \\ &\quad + \sum_{l\in\mathbf{L}_0} a_l\{\mathbf{p}\cdot\mathbf{x}^l | \mathbf{V}^l = \{\mathbf{x}^l\}\} = \sum_{l=1}^{L} a_l Q^l(\mathbf{p}), \end{aligned}$$

and (49) holds. The fundamental duality theorem applied to each $\mathbf{V}^l$ and Q^l pair and Lemma 2 establish the first conclusion.

For a design linear profit structure, the fundamental duality theorem applied to each Q^l and $\mathbf{V}^l$ pair establishes (49), $\mathbf{V}^l$ convex, and

$$\sum_{l=1}^{L} a_l Q^l(\mathbf{p}) = \sup\left\{\sum_{l=1}^{L} a_l \mathbf{p}\cdot\mathbf{x}^l | \mathbf{x}^l \in \mathbf{V}^l\right\} = \sup\left\{\mathbf{p}\cdot\mathbf{x} | \mathbf{x} \in \sum_{l=1}^{L} a_l\mathbf{V}^l\right\}.$$

The duality theorem applied to this function then establishes

$$\mathbf{V}(\mathbf{a},\mathbf{K}) = \sum_{l=1}^{L} a_l\mathbf{V}^l. \qquad \text{Q.E.D.}$$

Using the three lemmas above, we can now give the basic result which establishes that starting from the choice of any nested profit form, one can derive a netput supply system that is associated with an underlying *ex ante–ex post* technology:

Theorem 4. Consider any nested profit form **F**, **H**, ψ, **Q**. Then:

(i) There exists a regular convex design linear technological structure **F**, **B**, $\mathbf{T}^{ea}$, **W**, $\mathbf{V}(\mathbf{a},\mathbf{K}) = \sum_{l=1}^{L} a_l \mathbf{V}^l$, where **B** satisfies (44), the $\mathbf{V}^l$ satisfy (50), $\mathbf{T}^{ea}$ satisfies (47), and **W** satisfies (48a).

(ii) The technological structure in (i) yields an intertemporal variable profit function Π satisfying (40), (45), (1), and (46), and an *ex ante* envelope profit function Φ satisfying (43), (3), and (48b). The technological structure in (i) is the only regular convex design linear technological structure satisfying all these conditions.

(iii) Recall that $\mathbf{F}^p = \{\mathbf{p} \in \mathbf{E}^{NS} | (\mathbf{r},\mathbf{p}) \in \mathbf{F}\}$, define $\mathbf{F}^r(\mathbf{p}) = \{\mathbf{r} \in \mathbf{E}^J | (\mathbf{r},\mathbf{p}) \in \mathbf{F}\}$, and let $\mathbf{F}_0^p$ and $\mathbf{F}_0^r(\mathbf{p})$ denote the respective interiors of these sets. The partial profit functions Q^l are differentiable almost everywhere in $\mathbf{F}_0^p$. Hence, $\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) = \sum a_l Q^l(\mathbf{p})$ is differentiable almost everywhere in $\mathbf{F}_0^p$. When $\partial \Pi / \partial \mathbf{p} = \Pi_{\mathbf{p}}(\mathbf{a},\mathbf{K},\mathbf{p}) = \sum a_l Q_{\mathbf{p}}^l(\mathbf{p})$ exists, it equals the (unique) optimal *ex post* netput supply vector $\tilde{\mathbf{x}}(\mathbf{a},\mathbf{p}) = \sum_{l=1}^{L} a_l \hat{\mathbf{x}}^l(\mathbf{p})$ for this *ex ante* design and forward price vector. More generally, the subdifferential of Π with respect to **p** (see Chapter I.1) exists for all $\mathbf{p} \in \mathbf{F}_0^p$ and $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$, and each extreme vector in this subdifferential is an optimal netput supply vector in *any* regular design linear technological structure (not necessarily convex) satisfying (1), (3), (41), (42), (43) for this nested profit form.

(iv) For each $\mathbf{p} \in \mathbf{F}^p$, $\Phi(\mathbf{r},\mathbf{p}) = \psi(\mathbf{r},\mathbf{Q}(\mathbf{p}))$ is differentiable in **r** for almost all $\mathbf{r} \in \mathbf{F}_0^r(\mathbf{p})$, and $\Phi_{\mathbf{r}}(\mathbf{r},\mathbf{p}) = \psi_{\mathbf{r}}(\mathbf{r},\mathbf{Q}(\mathbf{p})) = \hat{\mathbf{K}}(\mathbf{r},\mathbf{p})$, where $\hat{\mathbf{K}}(\mathbf{r},\mathbf{p})$ is the (unique) optimal capital equipment netput vector at $(\mathbf{r},\mathbf{p})$. [A generalization analogous to that in (iii) holds for the subdifferential of Φ with respect to **r**.]

(v) For almost all $(\mathbf{r},\mathbf{p}) \in \mathbf{F}_0$, Φ and the Q^l, $l = 1,\ldots,L$, are differentiable in **p**, and $\hat{\mathbf{x}}(\mathbf{r},\mathbf{p}) = \Phi_{\mathbf{p}}(\mathbf{r},\mathbf{p}) = \hat{\mathbf{a}}(\mathbf{r},\mathbf{p}) \cdot \mathbf{Q}_{\mathbf{p}}(\mathbf{p})$ for any $\hat{\mathbf{a}}(\mathbf{r},\mathbf{p})$ in the subgradient of ψ with respect to **q** evaluated at $(\mathbf{r},\mathbf{Q}(\mathbf{p}))$, where $\hat{\mathbf{x}}(\mathbf{r},\mathbf{p})$ is the (unique) optimal netput supply vector for the forward price vector **p** and an optimal *ex ante* design for $(\mathbf{r},\mathbf{p})$. Each extreme vector $\hat{\mathbf{a}}(\mathbf{r},\mathbf{p})$ in the subgradient of ψ with respect to **q** is an optimal design in *any* regular design linear technological structure (not necessarily convex) satisfying (1), (3), (41), (42), (43) for this nested profit form. If ψ is differentiable with respect to $\mathbf{q}$ [and this is true for almost all $(\mathbf{r},\mathbf{q})$ in the interior of $\mathbf{H}$] at $(\mathbf{r},\mathbf{Q}(\mathbf{p}))$, then $\hat{\mathbf{a}}(\mathbf{r},\mathbf{q}) = \psi_{\mathbf{q}}(\mathbf{r},\mathbf{Q}(\mathbf{p}))$ is the (unique) optimal design.

Proof: Conclusions (i) and (ii) follow from Lemmas 1–3. The derivative properties (iii)–(v) are corollaries of Chapter I.1, Lemmas 17–19. Q.E.D.

It is convenient to summarize the formulae implied by Theorem 4 in the case that all the derivatives taken exist:

$$Q^l(\mathbf{p}) = \mathbf{p}\cdot\hat{\mathbf{x}}^l(\mathbf{p}), \tag{51}$$

$$\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) = \sum_{l=1}^{L} a_l Q^l(\mathbf{p}) = \mathbf{p}\cdot\tilde{\mathbf{x}}(\mathbf{a},\mathbf{p}) = \sum_{l=1}^{L} a_l \mathbf{p}\cdot\hat{\mathbf{x}}^l(\mathbf{p}), \tag{52}$$

$$\tilde{\mathbf{x}}_{ns}(\mathbf{a},\mathbf{p}) = \Pi_{p_{ns}}(\mathbf{a},\mathbf{K},\mathbf{p}) = \sum_{l=1}^{L} a_l Q^l_{p_{ns}}(\mathbf{p}), \tag{53}$$

$$\begin{aligned}\Phi(\mathbf{r},\mathbf{p}) = \psi(\mathbf{r},\mathbf{Q}(\mathbf{p})) &= \mathbf{r}\cdot\hat{\mathbf{K}}(\mathbf{r},\mathbf{p}) + \mathbf{p}\cdot\hat{\mathbf{x}}(\mathbf{r},\mathbf{p}) \\ &= \mathbf{r}\cdot\hat{\mathbf{K}}(\mathbf{r},\mathbf{p}) + \sum_{l=1}^{L} \hat{a}_l(\mathbf{r},\mathbf{p})\mathbf{p}\cdot\hat{\mathbf{x}}^l(\mathbf{p}),\end{aligned} \tag{54}$$

$$\hat{a}_l(\mathbf{r},\mathbf{p}) = \psi_{q_l}(\mathbf{r},\mathbf{Q}(\mathbf{p})), \tag{55}$$

$$\hat{x}_{ns}(\mathbf{r},\mathbf{p}) = \Phi_{p_{ns}}(\mathbf{r},\mathbf{p}) = \sum_{l=1}^{L} \hat{a}_l(\mathbf{r},\mathbf{p}) Q^l_{p_{ns}}(\mathbf{p}). \tag{56}$$

In econometric applications, choice of a convenient nested profit form provides a basis for empirical analysis, with formulae (51)–(56) yielding the netput supply equations (as functions of observed prices and underlying production parameters) and Theorem 4 ensuring consistency with some regular convex design linear technological structure.

5.2. *The General Case*

The results stated above for the finite case continue to hold more generally when the number of future states is infinite (as in a continuous time intertemporal model or an uncertainty model with a continuum of states of nature) or when the vector describing an *ex ante* design is infinite (as in the activity analysis model of Example 2 with an infinite set of activities). For simplicity, we confine our attention to the generalization in which the number of future states may be infinite, retaining the earlier assumption that an *ex ante* design can be described by a finite vector. This case is of some practical importance in the generation of functional forms for econometric purposes, as it allows use of the methods of calculus and differential equations.

Consider a measure space $(\mathbf{S},\mathscr{S},\mu)$ of future states, where $\mathbf{S}$ is the set of states, $\mathscr{S}$ is a σ-field of subsets of $\mathbf{S}$, and μ is a measure (a non-negative countably additive set function) defined on $\mathscr{S}$. (In applications, $\mathbf{S}$ might be an interval in the real line or a rectangle in finite-dimensional space with Lebesque measure, or a set of integers with counting measure. In the uncertainty model, μ may be the firm's subjective probability of occurrence of states of nature.)

An *ex post* production plan $\mathbf{x}$ or forward price vector $\mathbf{p}$ is a function from $\mathbf{S}$ into $\mathbf{E}^N$, hereafter assumed to be μ-measurable, i.e., $\{s \in \mathbf{S}|\mathbf{x}_s \in \mathbf{R}\} \in \mathscr{S}$ for each Borel set $\mathbf{R} \subseteq \mathbf{E}^N$. Consider a pair $(\mathbf{X},\mathbf{P})$ of linear spaces of μ-measurable functions from $\mathbf{S}$ into $\mathbf{E}^N$ such that the bilinear functional $\mathbf{p}\cdot\mathbf{x} = \int_S \mathbf{p}_s\cdot\mathbf{x}_s \, d\mu(s)$ is defined for $\mathbf{p}\in\mathbf{P},\mathbf{x}\in\mathbf{X}$. Assume the pair $(\mathbf{X},\mathbf{P})$ is separated; i.e., $\mathbf{p}\cdot\mathbf{x} = 0$ for all $\mathbf{p}\in\mathbf{P}$ implies $\mathbf{x} = \mathbf{0}$ and $\mathbf{p}\cdot\mathbf{x} = 0$ for all $\mathbf{x}\in\mathbf{X}$ implies $\mathbf{p} = \mathbf{0}$. We identify $\mathbf{X}$ as the space of *ex post* production plans, $\mathbf{P}$ as the space of forward price vectors. In applications, $\mathbf{P}$ is usually taken to be a topological space and $\mathbf{X}$ to be its adjoint. Examples of particular interest are (1) the finite case with $\mathbf{X} = \mathbf{P} = \mathbf{E}^{N\cdot S}$; (2) a case often used in problems involving uncertainty with $\mathbf{X} = \mathbf{P} = L_2(\mathbf{S},\mathscr{S},\mu,\mathbf{E}^N)$, the Hilbert space of functions from $\mathbf{S}$ into $\mathbf{E}^N$; and (3) a case occurring in intertemporal economics with the Banach spaces $\mathbf{P} = L_1(\mathbf{S},\mathscr{S},\mu,\mathbf{E}^N)$ and $\mathbf{X} = L_\infty(\mathbf{S},\mathscr{S},\mu,\mathbf{E}^N)$. We can assume, without seriously restricting potential applications, that $\mathbf{P}$ is a normed linear space and $\mathbf{X}$ is the Banach space of continuous linear functionals on $\mathbf{X}$. We shall use the weak* topology ($\mathbf{P}$-topology) on $\mathbf{X}$, denoted by $w(\mathbf{X},\mathbf{P})$, which is the weakest topology on $\mathbf{X}$ in which every functional in $\mathbf{P}$ is continuous. A generalized sequence $\mathbf{x}^d,\mathbf{d}\in\mathbf{D}$, converges to $\mathbf{x}$ in the weak* topology of $\mathbf{X}$ if and only if $\mathbf{p}\cdot\mathbf{x}^d$ converges to $\mathbf{p}\cdot\mathbf{x}$ for each $\mathbf{p}\in\mathbf{P}$. We shall use a mathematical result [Kelley and Namioka (1963, 18.6)] stating that $\mathbf{T}\subseteq\mathbf{X}$ weak* closed and bounded in norm (i.e., if $\|\mathbf{p}\|$ is the norm on $\mathbf{P}$, then $\sup_{\|\mathbf{p}\|=1} \sup_{x\in T} \mathbf{p}\cdot\mathbf{x} < +\infty$) implies $\mathbf{T}$ weak* compact.

Recall that an *ex ante* plant design is described by a vector $\mathbf{b} = (\mathbf{a},\mathbf{K})$, where $\mathbf{a}\in\mathbf{E}^L$ is an abstract design vector and $\mathbf{K}\in\mathbf{E}^J$ is a capital equipment vector with a corresponding price vector $\mathbf{r}\in\mathbf{E}^J$. Analogously to the finite case, a *technological structure* is defined by an *ex ante* envelope technology $\mathbf{T}^{ea}$ which is a non-empty set of vectors $(\mathbf{a},\mathbf{K},\mathbf{x})$ in $\mathbf{E}^{L+J} \times \mathbf{X}$, and the associated sets $\mathbf{B} \subseteq \mathbf{E}^{L+J}$, $\mathbf{V}(\mathbf{a},\mathbf{K}) \subseteq \mathbf{X}$ for $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$, $\mathbf{W} \subseteq \mathbf{E}^J \times \mathbf{X}$, and the normal cone of $\mathbf{W}$, defined as the convex set $\mathbf{F}$ of vectors $(\mathbf{r},\mathbf{p}) \in \mathbf{E}^J \times \mathbf{P}$ such that $\mathbf{r}\cdot\mathbf{K} + \mathbf{p}\cdot\mathbf{x}$ is bounded above on $\mathbf{W}$.

A technological structure is *strongly regular* if the following conditions hold: (1) the set $\mathbf{B}$ is closed, the sets $\mathbf{W}$ and $\mathbf{V}(\mathbf{a},\mathbf{K})$ for $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$ are weak* closed, and $\mathbf{V}(\mathbf{a},\mathbf{K})$ is bounded in norm; (2) capital netputs are non-positive and exhibit free disposal; and (3) for each $\mathbf{p} \in \mathbf{P}$, there exists $\delta > 0$ and $\mathbf{r} \in \mathbf{E}^J$ such that $\mathbf{r}$ is in the interior of $\mathbf{F}^r(\mathbf{p}')$ for $\mathbf{p}' \in \mathbf{P}, \|\mathbf{p}' - \mathbf{p}\| \leqq \delta$. Several comments on this definition are in order. The requirement that the *ex post* variable technology be bounded is a new condition not imposed in the finite case. Note that it is not consistent with free disposal in the *ex post* variable technology. However, it can normally be made to hold in applications by truncating the technology, carrying out the analysis below, and then reintroducing the omitted disposal activities. This condition is not essential for many of the following results; however, it greatly simplifies the mathematical arguments. Condition (3) is equivalent to the requirement that $\mathbf{F}$ have a non-empty interior in the norm topology of $\mathbf{E}^J \times \mathbf{P}$. A sufficient condition for (3) to hold is that the "average product" of capital go to zero when capital inputs are unbounded; i.e., $\lim_{|\mathbf{K}|\to\infty} \sup_{x,a}\{\|\mathbf{x}\|_x/|\mathbf{K}| \mid \mathbf{x} \in \mathbf{V}(\mathbf{a},\mathbf{K}), (\mathbf{a},\mathbf{K}) \in \mathbf{B}\} = 0$, where $\|\mathbf{x}\|_x$ is the norm of the Banach space $\mathbf{X}$ and $|\mathbf{K}|$ is the Euclidean norm on $\mathbf{E}^J$.

A technological structure is *convex* if $\mathbf{T}^{ea}$ is convex. It is *design linear* when the *ex post* variable technology has the form

$$\mathbf{V}(\mathbf{a},\mathbf{K}) = \sum_{l=1}^{L} a_l \mathbf{V}^l,$$

where the $\mathbf{V}^l$ are non-empty, weak* closed, and bounded in norm; and for each $l = 1,\ldots,L$, either a_l is non-negative for all $(\mathbf{a},\mathbf{K}) \in \mathbf{B}$ or $\mathbf{V}^l$ is a singleton.

Define a *strong nested profit form* by (1) a convex cone $\mathbf{F}$ of vectors $(\mathbf{r},\mathbf{p}) \in \mathbf{E}^J \times \mathbf{P}$ with r_j non-negative for $(\mathbf{r},\mathbf{p}) \in \mathbf{F}, j = 1,\ldots,J$, and such that for each $\mathbf{p} \in \mathbf{P}$ there exists $\delta > 0$ and $\mathbf{r} \in \mathbf{E}^J$ such that $\mathbf{r}$ is in the interior of $\mathbf{F}^r(\mathbf{p}')$ for $\mathbf{p}' \in \mathbf{P}, \|\mathbf{p} - \mathbf{p}'\| \leqq \delta$; (2) a convex cone $\mathbf{H}$ of vectors $(\mathbf{r},\mathbf{q}) \in \mathbf{E}^{J+L}$; (3) a convex conical closed function $\psi(\mathbf{r},\mathbf{q})$ on $\mathbf{H}$ which is non-increasing in r_j for all $(\mathbf{r},\mathbf{q}) \in \mathbf{H}$ with $r_j \geqq 0$ and which is non-decreasing [resp., non-monotone] in q_l for l in a set of indices $\mathbf{L}_+$ [resp., $\mathbf{L}_0$]; and (4) a vector of functions $\mathbf{Q}(\mathbf{p}) = (Q^1(\mathbf{p}),\ldots,Q^L(\mathbf{p}))$ on $\mathbf{P}$ such that $(\mathbf{r},\mathbf{q}) \in \mathbf{H}$ if and only if there exists $\mathbf{p} \in \mathbf{P}$ with $\mathbf{q} = \mathbf{Q}(\mathbf{p})$ and $(\mathbf{r},\mathbf{p}) \in \mathbf{F}$, and such that $Q^l(\mathbf{p})$ is a convex [resp., linear] conical uniformly Lipschitz function for $l \in \mathbf{L}_+$ [resp., $\mathbf{L}_0$].

The following result extends the conclusions of Theorem 4 to the general case where **S** need not be finite:

Theorem 5. Suppose one is given a strong nested profit form **F**, **H**, ψ, **Q**, $\mathbf{L}_+$, $\mathbf{L}_0$. Define

$$\mathbf{B} = \{(\mathbf{a},\mathbf{K}) \in \mathbf{E}^{L+J} | \mathbf{r}\cdot\mathbf{K} + \mathbf{q}\cdot\mathbf{a} \leqq \psi(\mathbf{r},\mathbf{q}) \text{ for all } (\mathbf{r},\mathbf{q}) \in \mathbf{H}\}, \tag{57}$$

$$\mathbf{V}^l = \{\mathbf{x} \in \mathbf{X} | \mathbf{p}\cdot\mathbf{x} \leqq Q^l(\mathbf{p}) \text{ for all } \mathbf{p} \in \mathbf{P}\}, \tag{58}$$

$$\mathbf{T}^{ea} = \left\{(\mathbf{a},\mathbf{K},\mathbf{x}) \in \mathbf{E}^{L+J} \times \mathbf{X} | \mathbf{x} \in \sum_{l=1}^{L} a_l \mathbf{V}^l \text{ and } (\mathbf{a},\mathbf{K}) \in \mathbf{B}\right\}. \tag{59}$$

Then (57)–(59) define a strongly regular convex design linear technological structure. This structure satisfies

$$Q^l(\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{x} | \mathbf{x} \in \mathbf{V}^l\}, \tag{60}$$

$$\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) = \sum_{l=1}^{L} a_l Q^l(\mathbf{p}) \quad \text{for} \quad \mathbf{p} \in \mathbf{P}, \tag{61}$$

$$\psi(\mathbf{r},\mathbf{q}) = \sup\{\mathbf{r}\cdot\mathbf{K} + \mathbf{q}\cdot\mathbf{a} | (\mathbf{a},\mathbf{K}) \in \mathbf{B}\} \quad \text{for} \quad (\mathbf{r},\mathbf{q}) \in \mathbf{H}, \tag{62}$$

$$\Phi(\mathbf{r},\mathbf{p}) = \psi(\mathbf{r},\mathbf{Q}(\mathbf{p})) \quad \text{for} \quad (\mathbf{r},\mathbf{p}) \in \mathbf{F}, \tag{63}$$

and is the only strongly regular convex design linear technological structure satisfying (60)–(63).

Proof: Except for the duality conditions (58) and (60), the statements of this theorem follow from the arguments of Lemmas 1–3. We first show that the $\mathbf{V}^l$ defined by (58) are non-empty, convex, bounded in norm, and weak* closed. We consider $l \in \mathbf{L}_+$; the remaining case is left to the reader. Define $\mathbf{epi}\, Q^l = \{(\mathbf{p},\mathbf{q}) \in \mathbf{P} \times \mathbf{E} | q \geqq Q^l(\mathbf{p})\}$ and $\mathbf{G}^l = \{(\mathbf{x},\xi) \in \mathbf{X} \times \mathbf{E} | -\mathbf{p}\cdot\mathbf{x} + \xi q \geqq 0 \text{ for } (\mathbf{p},\mathbf{q}) \in \mathbf{epi}\, Q^l\}$. Since Q^l is Lipschitz and conical, $\mathbf{epi}\, Q^l$ is a closed (in norm) convex cone with $(0,-1) \notin \mathbf{epi}\, Q^l$, $\mathbf{G}^l$ and $\mathbf{epi}\, Q^l$ are polar cones, and $\mathbf{G}^l$ is a weak* closed convex cone. $\mathbf{G}^l$ contains non-zero vectors [Dunford and Schwartz (1958, V.9.8)], and $\mathbf{0} \neq (\mathbf{x},\xi) \in \mathbf{G}^l$ implies $\xi > 0$. Then, $\mathbf{V}^l = \{\mathbf{x} \in \mathbf{X} | (\mathbf{x},1) \in \mathbf{G}^l\}$ satisfies (58) and is non-empty, convex, and weak* closed. Since Q^l satisfies $|Q^l(\mathbf{p})| \leqq m\|\mathbf{p}\|$ for some $m > 0, \mathbf{x} \in \mathbf{V}^l$ satisfies $|\mathbf{p}\cdot\mathbf{x}| \leqq m\|\mathbf{p}\|$ for all $\mathbf{p} \in \mathbf{P}$, or $\|\mathbf{x}\| \leqq m$. Hence, $\mathbf{V}^l$ is bounded in norm. Equation (60) and the uniqueness statement of the theorem follow from the polarity of the cones **epi** Q^l and $\mathbf{G}^l$. Q.E.D.

As in the finite case, one may define the subdifferential of Q^l at $\mathbf{p}$ to be the set of points $\mathbf{x} \in \mathbf{X}$ satisfying $\mathbf{p}' \cdot \mathbf{x} \leqq Q^l(\mathbf{p} + \mathbf{p}') - Q^l(\mathbf{p})$ for $\mathbf{p}' \in \mathbf{P}$. The assumption that Q^l is convex conical and uniformly Lipschitz implies the subdifferential is always non-empty and equals the set of $\mathbf{x} \in \mathbf{V}^l$ maximizing $\mathbf{p} \cdot \mathbf{x}'$ on $\mathbf{V}^l$. The almost everywhere differentiability of Q^l in the finite case does not carry over to the general infinite-dimensional case; however, the following partial generalization holds. First, several definitions: A subspace $\mathbf{P}_0$ of $\mathbf{P}$ is *separable* if it contains a countable dense subset. Sufficient conditions for $\mathbf{P}_0$ to be separable are (i) that it be the closed space spanned by a countable subset of $\mathbf{P}$ [Dunford and Schwartz (1958, II.1.5)] or (ii) that $\mathbf{S}$ be a compact metric space and $\mathbf{P}_0$ be a closed space of continuous functions on $\mathbf{S}$ [Dunford and Schwartz (1958, V.7.12)]. $\mathbf{P}$ is the *direct sum* of subspaces $\mathbf{P}_0$ and $\mathbf{P}_1$, written $\mathbf{P} = \mathbf{P}_0 \oplus \mathbf{P}_1$, if each $\mathbf{p} \in \mathbf{P}$ has a unique representation $\mathbf{p} = \mathbf{p}_0 + \mathbf{p}_1, \mathbf{p}_0 \in \mathbf{P}_0, \mathbf{p}_1 \in \mathbf{P}_1$. Define $\mathbf{X}_0 = \{\mathbf{x} \in \mathbf{X} | \mathbf{p} \cdot \mathbf{x} = 0 \text{ for } \mathbf{p} \in \mathbf{P}_0\}$ and the quotient space $\mathbf{X}/\mathbf{X}_0$. A function $\mathbf{Q}:\mathbf{P}_0 \to \mathbf{E}^L$ is *differentiable* at $\mathbf{p}_0 \in \mathbf{P}_0$ if there exists a unique $\mathbf{x}_0^l \in \mathbf{X}/\mathbf{X}_0$ such that $\mathbf{p}' \cdot \mathbf{x}_0^l = \lim_{\theta \to 0^+} [\tilde{Q}^l(\mathbf{p}_0 + \mathbf{p}') - \tilde{Q}^l(\mathbf{p}_0)]/\theta$ for all $\mathbf{p}' \in \mathbf{P}_0$, $l = 1,\ldots,L$.

Lemma 6. Suppose $\mathbf{P} = \mathbf{P}_0 \oplus \mathbf{P}_1$, where $\mathbf{P}_0$ is a closed separable subspace of $\mathbf{P}$. For each $\mathbf{p}_1 \in \mathbf{P}_1$, the convex conical uniformly Lipschitz functions $\mathbf{Q}(\mathbf{p}_0 + \mathbf{p}_1) = (Q^1(\mathbf{p}_0 + \mathbf{p}_1),\ldots,Q^L(\mathbf{p}_0 + \mathbf{p}_1))$, considered as functions of $\mathbf{p}_0 \in \mathbf{P}_0$ are differentiable on a dense subset of $\mathbf{P}_0$.

Proof: Define $q^l:\mathbf{P}_0 \to \mathbf{E}$ by $q^l(\mathbf{p}_0) = Q^l(\mathbf{p}_0 + \mathbf{p}_1)$ for $l \in L$, and define $\mathbf{q}(\mathbf{p}_0) = \sum_{l=1}^{L} q^l(\mathbf{p}_0)$. Consider the tangent functional $\tau^l(\mathbf{p}_0,\mathbf{p}') = \lim_{\theta \to 0^+}(q^l(\mathbf{p}_0 + \theta\mathbf{p}') - q^l(\mathbf{p}_0))/\theta, \mathbf{p}' \in \mathbf{P}_0$, and note that the tangent functional $\tau(\mathbf{p}_0,\mathbf{p}')$ of $q(\mathbf{p}_0)$ satisfies $\tau(\mathbf{p}_0,\mathbf{p}') = \sum_{l=1}^{L} \tau^l(\mathbf{p}_0,\mathbf{p}')$. Note that $\mathbf{q}$ is uniformly Lipschitz on $\mathbf{P}_0$, and that $\tau^l(\mathbf{p}_0,\mathbf{p}') \geqq -\tau^l(\mathbf{p}_0, -\mathbf{p}')$. The proof of V.9.8 in Dunford and Schwartz (1958) establishes that for $\mathbf{p}_0$ in a dense subset $\mathbf{P}_0'$ of $\mathbf{P}_0, \tau(\mathbf{p}_0,\mathbf{p}') = -\tau(\mathbf{p}_0, -\mathbf{p}')$ for $\mathbf{p}' \in \mathbf{P}_0$. From the inequality above, this implies $\tau^l(\mathbf{p}_0,\mathbf{p}') = -\tau^l(\mathbf{p}_0, -\mathbf{p}')$ for $\mathbf{p}_0 \in \mathbf{P}_0', \mathbf{p}' \in \mathbf{P}_0, l = 1,\ldots,L$. Then, $\tau^l(\mathbf{p}_0,\mathbf{p}') = \mathbf{p}' \cdot \mathbf{x}_0$, where $\mathbf{x}_0$ is in the subdifferential of $Q^l(\mathbf{p})$ at $\mathbf{p} = \mathbf{p}_0 + \mathbf{p}_1$, implying $\mathbf{x}_0$ is unique in $\mathbf{X}/\mathbf{X}_0$. Q.E.D.

6. Separable Technology Across States

The abstract model of *ex ante–ex post* firm behavior presented in Sections 3 and 5 requires no specific assumptions on the structure of *ex post* technologies across future states. However, each of the examples in Section 4 is based on a technology structure that is "separable across states". While we wish to avoid imposing unnecessary structure that precludes possible applications, it is important to explore the implications of this conventional structural assumption of separability ("non-joint" production) across future states.[9] This condition reduces substantially the "dimension" of the intertemporal profit function, making it an important property in empirical applications, where this function must be given an explicit form.

An *ex post* variable technology $V(\mathbf{b}), \mathbf{b} \in \mathbf{B}$, is *separable across states* if for each future state $s \in \mathbf{S}$ there exists a set $\mathbf{V}_s(\mathbf{b})$ of N-vectors of netputs $\mathbf{x}_s$, termed the *ex post s*-technology, such that

$$\mathbf{V}(\mathbf{b}) = \{\mathbf{x} \in X | \mathbf{x}_s \in \mathbf{V}_s(\mathbf{b}) \text{ for } s \in \mathbf{S}\}.$$

When $\mathbf{S}$ is finite, $\mathbf{V}(\mathbf{b})$ is just the Cartesian product of the s-technologies. The important characteristic of this structure is that, given the *ex post* technology, the set of possible netput vectors in a state s is independent of the operating points chosen in other states. For the model with uncertainty and one period of operation, this condition will always be imposed (recall that decisions made before the state of nature is known are described in $\mathbf{b}$). An example illustrates application of this condition in the model with intertemporal variation and no uncertainty. Suppose a firm chooses *ex ante* an input level K of one producer durable, and in each future period $s = 1,...,L$ chooses a labor input L_s and output Y_s satisfying a production function $Y_s \leqq (Kd_s)^\alpha L_s^\beta$, where d_s is a depreciation effect. If d_s is exogenous, resulting from weathering depreciation, then this technology is separable across states. Alternately, if d_s depends on output levels in previous periods because of wear and tear depreciation, this separable structure does not hold.[10]

[9]A quick accounting of problems employing intertemporal production yields the following list of phenomena related to intertemporal structure: disembodied technical change, weathering depreciation of durable equipment, wear and tear depreciation of durable equipment, learning-by-doing in the plant, endogenous construction rate and scrapping decisions, expansion and modernization decisions. The first two phenomena are consistent with intertemporal separability, the remainder are not.

[10]In a form of the separability assumption that appears in the economic growth literature, inputs in one period yield outputs in the following period in two-period

For each *ex post* s-technology $\mathbf{T}_s(\mathbf{b})$ and forward price vector $\mathbf{p} = (\mathbf{p}_s : s \in \mathbf{S})$, define an *s-future profit function*,

$$\Pi_s(\mathbf{b},\mathbf{p}_s) = \sup\{\mathbf{p}_s \cdot \mathbf{x}_s | \mathbf{x}_s \in \mathbf{T}_s(\mathbf{b})\}. \tag{64}$$

The future profit function is then given by the sum $\Pi(\mathbf{b},\mathbf{p}) = \sum_{s \in S} \Pi_s(\mathbf{b},\mathbf{p}_s) = \sum_{s \in S} \delta_s \Pi_s(\mathbf{b},\tilde{\mathbf{p}}_s)$ when S is finite, and more generally by the integral $\Pi(\mathbf{b},\mathbf{p}) = \int_S \Pi_s(\mathbf{b},\mathbf{p}_s) d\mu(s)$. Conversely, additive separability of the profit function across states implies that the *ex post* technology is, in effect, separable across states; the precise implication for S finite is

$$\begin{aligned} \text{convex hull } (\mathbf{V}(\mathbf{b})) &= \{\mathbf{x} | \mathbf{p} \cdot \mathbf{x} \leqq \Pi(\mathbf{b},\mathbf{p}) \text{ for all } \mathbf{p}\} \\ &= \{\mathbf{x} | \textstyle\sum_{s \in S} \mathbf{p}_s \cdot \mathbf{x}_s \leqq \sum_{s \in S} \Pi_s(\mathbf{b},\mathbf{p}_s) \text{ for all } \mathbf{p}\} \\ &= \times_{s \in S} \{\mathbf{x}_s | \mathbf{p}_s \cdot \mathbf{x}_s \leqq \Pi_s(\mathbf{b},\mathbf{p}_s) \text{ for all } \mathbf{p}_s\} \\ &= \times_{s \in S} \text{ convex hull } (V_s(\mathbf{b})). \end{aligned}$$

In applications, the s-technologies are frequently assumed to vary in a simple pattern over states. A first example in the intertemporal variation model is the "one-hoss-shay" technology, with $\mathbf{T}_s(\mathbf{b}) = \mathbf{T}_1(\mathbf{b})$ for $s = 1,\ldots,L$. A second example giving a uniform s-technology across states is the single-period model in which the firm faces uncertain market prices, but a certain technology $\mathbf{T}_1(\mathbf{b})$. The profit function then satisfies $\Pi(\mathbf{b},\mathbf{p}) = \int_S \Pi_1(\mathbf{b},\mathbf{p}_s) d\mu(s)$. The following transformation of this second example will be useful in applications: Let $\mathbf{p}_.$ denote an N-vector of prices and define $\tilde{S}(\mathbf{p}_.) = \{s \in S | \tilde{\mathbf{p}}_s \leqq \mathbf{p}_.\}$ and $G(\mathbf{p}_.) = \int_{\tilde{S}(\mathbf{p}_.)} \delta_s \, d\mu(s)$. Then, G is the distribution function of current price vectors, and $\Pi(\mathbf{b},\mathbf{p}) = \int_{E^N} \Pi_1(\mathbf{b},\mathbf{p}_.) dG(\mathbf{p}_.)$ is the expected value of current profit. Since the profit function $\Pi_1(\mathbf{b},\mathbf{p}_.)$ is convex in $\mathbf{p}_.$, we have the implication $\int_{E^N} \Pi_1(\mathbf{b},\mathbf{p}_.) dG(\mathbf{p}_.) \geqq \Pi_1(\mathbf{b}, \int_{E^N} \mathbf{p}_. \, dG(\mathbf{p}_.))$, with equality holding if the $\mathbf{p}_s$ are proportional in all states. First, we conclude that when relative prices are not affected by the state of nature, the *ex post* operation of the plant is reduced to the problem of maximizing current profit at expected prices. Second, that given the certain *ex post* technology $\mathbf{T}_1(\mathbf{b})$, the firm cannot lose and may gain from increased uncertainty. To make

technologies. This recursive technology structure assumes the existence of markets for all intermediate goods, including rental and second-hand markets for producer durables. This structure can be included in the separable across-states assumption in our analysis by treating each state s as made up of two subperiods, with the second subperiod of state s coinciding chronologically with the first subperiod of the successive state. The markets for inputs and outputs at the same chronological time are then treated as distinct markets (with arbitrage).

this statement more precise, we consider a comparative argument in which commodities have a current price vector $\mathbf{p}_.$ with distribution $G(\mathbf{p}_.)$ in the first case and a current price vector $\mathbf{p}_. + \mathbf{p}'_.$ in the second, and $\mathbf{p}'_.$ introduces a "pure spread" in the price distribution; i.e., $\mathbf{p}'_.$ has a conditional distribution $H(\mathbf{p}'_.|\mathbf{p}_.)$ with $\int_{\mathbf{E}^N} \mathbf{p}'_.\,\mathrm{d}H(\mathbf{p}'_.|\mathbf{p}_.) = 0$. Then, $\int_{\mathbf{E}^N}(\int_{\mathbf{E}^N} \Pi_1(\mathbf{b},\mathbf{p}_. + \mathbf{p}'_.)\mathrm{d}H(\mathbf{p}'_.|\mathbf{p}_.))\mathrm{d}G(\mathbf{p}_.) \geqq \int_{\mathbf{E}^N} \Pi_1(\mathbf{b},\mathbf{p}_. + \int_{\mathbf{E}^N} \mathbf{p}'_.\,\mathrm{d}H(\mathbf{p}'_.|\mathbf{p}_.))\mathrm{d}G(\mathbf{p}_.) = \int_{\mathbf{E}^N} \Pi_1(\mathbf{b},\mathbf{p}_.)\mathrm{d}G(\mathbf{p}_.)$, and expected profits from the "spread" prices $\mathbf{p}_. + \mathbf{p}'_.$ are at least as high as those from the prices $\mathbf{p}_.$. An intuitive justification of this conclusion can be given using Figure 13. Suppose an industry with identical firms faces an uncertain demand for industry output Y, each of the curves D^1 and D^2 occurring with probability one-half, and yielding an expected output price $\bar{p}_.$. Supply at the expected output price will be greater than (less than) the expected value of output in the case of a convex industry supply in Figure 13(a) [a concave industry supply in Figure 13(b)]. However, because higher equilibrium prices always induce higher supply, and hence higher profits, the expected value of profit will be unambiguously higher than profit at the expected output price in either case (a) or (b).

A structure of the s-technologies across states only slightly less simple than the uniform examples considered above is the case of exogenous commodity augmentation in which $V_s(\mathbf{b}) = \{(x_1A_{1s},\ldots,x_NA_{Ns})|(x_1,\ldots,x_N) \in V_1(\mathbf{b})\}$, the A_{ns} being exogenous non-negative numbers. In the model with intertemporal production, the A_{ns} may represent the effects of weathering depreciation and disembodied technical change. In the single-period model with uncertainty, the A_{ns} reflect the quality of commodities in various states of nature. In the concrete example with a single durable input K and production of output Y_1 in state 1 from input L_1 satisfying $Y_1 \leqq K^\alpha L_1^\beta$, this structure gives a production function $Y_s \leqq K^\alpha L_s^\beta(A_{Ls}^\beta A_{Ys}^{-1})$. The commodity-augmenting structure implies that the variable profit function can be written $\Pi(\mathbf{b},\mathbf{p}) = \int_S \Pi_1(K,p_{1s}/A_{1s},\ldots,p_{Ns}/A_{Ns})\mathrm{d}\mu(s)$, with p_{ns}/A_{ns} interpreted as an efficiency forward price.

A final comment is in order on the differentiability of $\Pi(\mathbf{a},\mathbf{K},\mathbf{p})$ in the general case of a measure space $(\mathbf{S},\mathcal{S},\mu)$ of future states when the technology is separable across states. If each s-technology $V_s(\mathbf{a},\mathbf{K})$ is strictly convex in $\mathbf{E}^N$, then $\Pi_s(\mathbf{a},\mathbf{K},\mathbf{p}_s)$ is differentiable in $\mathbf{p}_s$ on $\mathbf{E}^N$. Hence, $\mathbf{p}'\cdot\hat{\mathbf{x}} = \int_S(\partial\Pi_s/\partial\mathbf{p}_s)\cdot\mathbf{p}'_s\mathrm{d}\mu(s)$ for $\mathbf{p}' \in \mathbf{P}$, $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_s : s \in \mathbf{S})$, with $\hat{\mathbf{x}}_s$ the unique optimal netput vector in $\mathbf{V}_s(\mathbf{a},\mathbf{K})$ at price vector $\mathbf{p}_s$. Hence, $\Pi(\mathbf{a},\mathbf{K},\mathbf{p})$ is differentiable in $\mathbf{p}$ on $\mathbf{P}$.

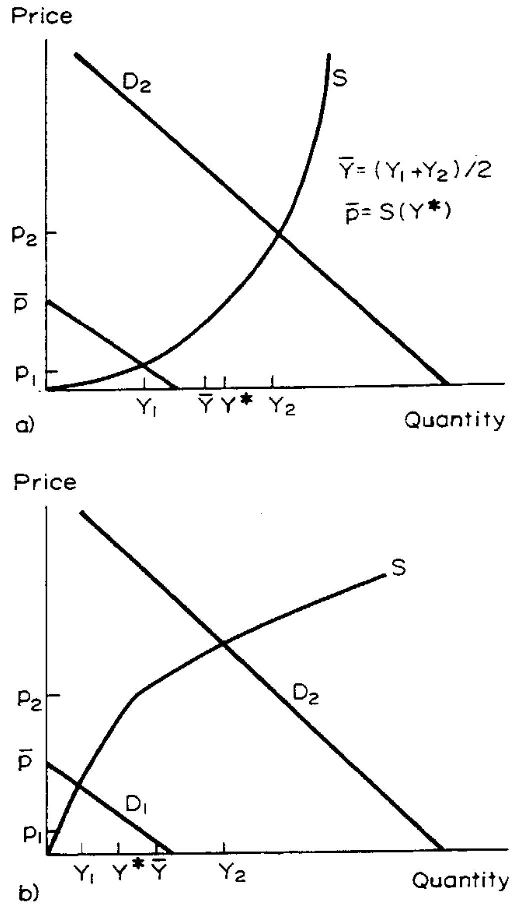


FIGURE 13

7. A General Linear-in-Parameters *Ex Ante–Ex Post* Technology

The algorithm introduced in Section 5 provides a general procedure for generating two-level technologies. Example 4 in Section 4 illustrates a construction for the case of cost minimization that is linear in the underlying production parameters, making it particularly convenient for statistical analysis. We now present a generalization of this family of nested forms to the profit-maximization case, and show that this generalization is robust in the sense that locally it can mimic the net supply behavior of a broad class of two-level technologies. For simplicity, we assume in this analysis that the set of states **S** is finite.

Consider a nested profit form as defined in Section 5, described by **F**,

H, ψ, and $\mathbf{Q} = (Q^1,\ldots,Q^L)$. Suppose ψ has a linear-in-parameters form

$$\psi(\mathbf{r},\mathbf{q}) = \sum_{g=1}^{G} b_g R^g(\mathbf{r},\mathbf{g}), \tag{65}$$

where b_g is non-negative and R^g is non-decreasing in $\mathbf{q}$ for $(\mathbf{r},\mathbf{q}) \in \mathbf{H}$. Then, Theorem 4 holds, implying equations (51)–(56) can be used to specify the net supply system. Rewriting (53)–(56) for the functional form (65),

$$\Phi(\mathbf{r},\mathbf{p}) = \psi(\mathbf{r},Q(\mathbf{p})) = \sum_{g=1}^{G} b_g R^g(\mathbf{r},Q(\mathbf{p})), \tag{66}$$

$$\hat{a}_l(\mathbf{r},\mathbf{p}) = \sum_{g=1}^{G} b_g R^g_{q_l}(\mathbf{r},Q(\mathbf{p})), \tag{67}$$

$$\hat{x}_{ns}(\mathbf{r},\mathbf{p}) = \sum_{l=1}^{L} \hat{a}_l(\mathbf{r},\mathbf{p}) Q^l_{p_{ns}}(\mathbf{p}) = \sum_{g=1}^{G} b_g \left[\sum_{l=1}^{L} R^g_{q_l}(\mathbf{r},Q(\mathbf{p}))Q^l_{p_{ns}}(\mathbf{p})\right], \tag{68}$$

Note that the system (66)–(68) is linear in the underlying technological parameters b_g.

Adapting the functional form of Example 4, let **N**, **J**, and **S** denote the sets of indices $\{1,\ldots,N\}$, $\{1,\ldots,J\}$, and $\{1,\ldots,S\}$, respectively, and define

$$Q^{ijst}(\mathbf{p}) = -(p_{is}p_{jt})^{1/2}, \tag{69}$$

$$\tilde{Q}^{ij}(\mathbf{p}) = -\sum_{s\in\mathbf{S}} (p_{is}p_{js})^{1/2}, \tag{70}$$

for $i,j \in \mathbf{N}$ and $s,t \in \mathbf{S}$. Suppose ψ has a linear-in-parameters form

$$\begin{aligned}\psi(\mathbf{r},\mathbf{q}) = &\sum_{i,j,k,l\in\mathbf{N}} b^1_{ij,kl}[-(q_{ij}q_{kl})^{1/2}] + 2\sum_{\substack{i,j\in\mathbf{N}\\ k\in\mathbf{J}}} b^2_{ij,k}[-(-q_{ij}r_k)^{1/2}]\\ &+ \sum_{\substack{ijst\in\mathbf{L}\\ kluv\in\mathbf{L}}} b^3_{ijst,kluv}[-(q_{ijst}q_{kluv})^{1/2}]\\ &+2\sum_{\substack{ijst\in\mathbf{L}\\ k\in\mathbf{J}}} b^4_{ijst,k}[-(-q_{ijst}r_k)^{1/2}] + \sum_{k,l\in\mathbf{J}} b^5_{kl}[-(r_kr_l)^{1/2}],\end{aligned} \tag{71}$$

where $\mathbf{L} = \{(ijst)'|i,j \in \mathbf{N}; s,t \in \mathbf{S}\}$ and the underlying b parameters satisfy the following conditions:

(1) $b^1_{ij,kl}$ symmetric under permutation of i and j, of (ij) and (kl), and of combinations of these permutations; and non-negative unless $(ij) = (kl)$ for some permutation of i and j;

(2) $b^2_{ij,k}$ symmetric under permutation of i and j; and non-negative;
(3) $b^3_{ijst,kluv}$ symmetric under permutation of i and j, of s and t, or $(ijst)$ and $(kluv)$, and of combinations of these permutations; and non-negative unless $(ijst) = (kluv)$ for some permutation of i and j, and of s and t;
(4) $b^4_{ijst,k}$ symmetric under permutation of i and j, or s and t, of combinations of these permutations; and non-negative;
(5) b^5_{kl} symmetric under permutation of k and l, and non-negative unless $k = l$.

The set $\mathbf{H}$ is defined so that $\mathbf{q}$ is non-positive and ψ is non-decreasing in $\mathbf{q}$ at $(\mathbf{r},\mathbf{q}) \in \mathbf{H}$; this implies a constraint on the range of the "diagonal" b parameters. One may readily verify that this system meets the requirements of a nested profit form. Writing out the net supply system:

$$\hat{a}_{ij}(\mathbf{r},\mathbf{p}) = -\sum_{k,l\in\mathbf{N}} b^1_{ij,kl}(\tilde{Q}^{kl}(\mathbf{p})/\tilde{Q}^{ij}(\mathbf{p}))^{1/2} - \sum_{k\in\mathbf{J}} b^2_{ij,k}(-r_k/\tilde{Q}^{ij}(\mathbf{p}))^{1/2}, \tag{72}$$

$$\hat{a}_{ijst}(\mathbf{r},\mathbf{p}) = -\sum_{kluv\in\mathbf{L}} b^3_{ijst,kluv}(Q^{kluv}(\mathbf{p})/Q^{ijst}(\mathbf{p}))^{1/2} - \sum_{k\in\mathbf{J}} b^4_{ijst,k}(-r_k/Q^{ijst}(\mathbf{p}))^{1/2}, \tag{73}$$

$$\hat{K}_k(\mathbf{r},\mathbf{p}) = -\sum_{l\in\mathbf{J}} b^5_{kl}(r_l/r_k)^{1/2} - \sum_{i,j\in\mathbf{N}} b^2_{ij,k}(-\tilde{Q}^{ij}(\mathbf{p})/r_k)^{1/2} - \sum_{ijst\in\mathbf{L}} b^4_{ijst,k}(-Q^{ijst}(\mathbf{p})/r_k)^{1/2}, \tag{74}$$

$$\hat{x}_{is}(\mathbf{r},\mathbf{p}) = -\sum_{j\in\mathbf{N}} \hat{a}_{ij}(p_{js}/p_{is})^{1/2} - \sum_{\substack{j\in\mathbf{N}\\ t\in\mathbf{S}}} \hat{a}_{ijst}(p_{jt}/p_{is})^{1/2}, \tag{75}$$

$$\begin{aligned}\hat{x}_{is}(\mathbf{r},\mathbf{p}) = &\sum_{j,k,l\in\mathbf{N}} b^1_{ij,kl}(p_{js}/p_{is})^{1/2}(\tilde{Q}^{kl}(\mathbf{p})/\tilde{Q}^{ij}(\mathbf{p}))^{1/2} \\ &+ \sum_{\substack{j\in\mathbf{N}\\ k\in\mathbf{J}}} b^2_{ij,k}(p_{js}/p_{is})^{1/2}(-r_k/\tilde{Q}^{ij}(\mathbf{p}))^{1/2} \\ &+ \sum_{\substack{j\in\mathbf{N}\\ t\in\mathbf{S}\\ kluv\in\mathbf{L}}} b^3_{ijst,kluv}(p_{jt}/p_{is})^{1/2}(Q^{kluv}(\mathbf{p})/Q^{ijst}(\mathbf{p}))^{1/2} \\ &+ \sum_{\substack{j\in\mathbf{N}\\ t\in\mathbf{S}\\ k\in\mathbf{J}}} b^4_{ijst,k}(p_{jt}/p_{is})^{1/2}(-r_k/Q^{ijst}(\mathbf{p}))^{1/2}.\end{aligned} \tag{76}$$

Given data on a cross-section of firms operated in each future state, the system (74), (76) can be estimated by multivariate regression methods. It is of interest to state explicitly in terms of the b parameters some of the important hypotheses imposed on the *ex ante–ex post* technology:

(1) Separable across states *ex post* variable technology. This hypothesis holds if and only if the *ex post* variable profit function is additively separable across states, or $\hat{a}_{ijst} = 0$ for $s \neq t$. This is equivalent to the linear hypothesis $b^3_{ijst,kluv} = 0$ unless $s = t$ and $u = v$ and $b^4_{ijst,k} = 0$ unless $s = t$.

(2) Separable and uniform across states *ex post* variable technology. This hypothesis holds if and only if (1) above holds and $\hat{a}_{ijss} = \hat{a}_{ij11}$. This is equivalent to the linear hypothesis $b^3_{ijst,kluv} = 0 = b^4_{ijst,k}$, since the equality above holds only if $\hat{a}_{ij11} \equiv 0$.

(3) Putty–clay hypothesis. This condition holds if and only if $\hat{\hat{a}}_{ij} = 0$ for $i \neq j$ and $\hat{a}_{ijst} = 0$ for $i \neq j$ or $s \neq t$. This is equivalent to the linear hypothesis $b^1_{ij,kl} = 0$ unless $i = j$, $k = l$; $b^2_{ij,k} = 0$ unless $i = j$; $b^3_{ijst,kluv} = 0$ unless $i = j$, $k = l$, $s = t$, $u = v$; and $b^4_{ijst,k} = 0$ unless $i = j$, $s = t$.

(4) Non-jointness of net supplies (is) and (jt) in the *ex post* variable technology. If $s \neq t$, this hypothesis is equivalent to the linear hypothesis $b^3_{ijst,kluv} = 0$ for $kluv \in \mathbf{L}$ and $b^4_{ijst,k} = 0$ for $k \in \mathbf{J}$. If $s = t$, this hypothesis is equivalent to $b^1_{ij,kl} = 0$ for $k,l \in \mathbf{N}$, $b^2_{ij,k} = 0$ for $k \in \mathbf{J}$, $b^3_{ijss,kluv} = 0$ for $kluv \in \mathbf{L}$, and $b^4_{ijss,k} = 0$ for $k \in \mathbf{J}$.

Finally, we note that the nested profit form (69)–(71) has the following approximation property, established in Chapter II.2: Consider any nested profit form which is twice continuously differentiable in the neighborhood of a point and which has at this point the *gross substitutes* property that the mixed second partial derivatives of ψ and the Q^m are non-positive. Then there exist b parameters in the form (69)–(71) such that this form agrees with the nested profit function through second-order partials at this point. Hence, the system (69)–(71) can mimic locally the net supplies and elasticities of any *ex ante–ex post* technology yielding a nested profit form with the gross substitutes property. The procedures of McFadden in Chapter II.2 can be used to establish a stronger result. Suppose $\Pi(\mathbf{a},\mathbf{K},\mathbf{p})$ and $\Phi(\mathbf{r},\mathbf{p})$ are the profit functions associated with an arbitrary *ex ante–ex post* technology. Suppose they are differentiable at a point $(\bar{\mathbf{r}},\bar{\mathbf{p}})$, yielding optimal quantities $(\bar{\mathbf{a}},\bar{\mathbf{K}},\bar{\mathbf{x}})$. Suppose that Φ and Π are differentiable at this point, and have the following gross substitute properties:

$$\begin{aligned}&\partial^2\Pi/\partial a_i\partial a_j \geqq 0, && \partial^2\Pi/\partial K_i\partial K_j \geqq 0,\\ &\partial^2\Pi/\partial a_i\partial K_j \geqq 0, && \partial^2\Pi/\partial p_i\partial p_j \leqq 0,\\ &\partial^2\Phi/\partial r_i\partial r_j \leqq 0, && \partial^2\Phi/\partial p_i\partial p_j \leqq 0,\\ &\partial^2\Phi/\partial p_i\partial r_j \leqq 0.\end{aligned}$$

Suppose Π is concave in $\mathbf{a}$, independent of $\mathbf{K}$, and define

$$\Pi(\mathbf{a},\mathbf{K},\mathbf{p}) \doteq \Pi(\bar{\mathbf{a}},\bar{\mathbf{K}},\bar{\mathbf{p}}) + \sum_{l=1}^{L} \alpha_l \Pi_{a_l}(\bar{\mathbf{a}},\bar{\mathbf{K}},\mathbf{p}) + 1/2 \sum_{l,m=1}^{L} \beta_{lm}\Pi_{a_l a_m}(\bar{\mathbf{a}},\bar{\mathbf{K}},\mathbf{p}),$$

where

$$\alpha_l = a_l - \bar{a}_l, \qquad \beta_{lm} = (a_l - \bar{a}_l)(a_m - \bar{a}_m).$$

This function is then linear in the parameters α_l,β_{lm}. Define $\mathbf{B}$ to be the set of parameter vectors $(1,(\alpha_l), (\beta_{lm}))$ corresponding to $(\mathbf{a},\mathbf{K})$ in the domain of Π, and define $\psi(q_0,(q_l), (q_{lm}))$ to be its "profit function". Then the "linearized" Π and the function ψ define a nested profit form that can, by the procedure outlined previously, be approximated to second order by the system (69)–(71). Thus, the tests suggested above should be relatively robust for deviations of the true *ex ante–ex post* production structure from that implied by the linear-in-parameters form. Because the multivariate model estimated in these tests utilizes much more information on the structure of the production process than would parallel non-parametric statistics, it should yield substantially more powerful tests.

8. Concluding Remarks

The preceding chapters in this volume have demonstrated that the theory of duality is a concept that is extremely useful in the estimation of production parameters when production is specified in terms of a single-level decision rule. In this chapter we have extended the application of duality theory to the more complicated two-stage (*ex ante–ex post*) description of technology. This extension allowed us to pursue interesting phenomena, such as the role of uncertainty in the design decision, which cannot be analyzed within the more limited framework.

The main purpose of this chapter was to demonstrate a means of generating functional forms that are capable of empirically describing the *ex ante–ex post* structure. Econometricians who estimate production functions require these functional forms in order to take into account the dynamic efficiency of flexible techniques in a world of durable inputs and uncertain outcomes.

Appendix A.1

DEFINITE QUADRATIC FORMS SUBJECT TO CONSTRAINTS

DANIEL McFADDEN

University of California, Berkeley

1. Conditions for a Matrix to be Positive Definite

Let $\mathbf{A}$ denote a real symmetric matrix of order n. Let σ denote a subvector $(i_1,\ldots,i_r)$ of the vector of integers $(1,\ldots,n)$, ordered so that $i_1 < i_2 < \cdots < i_r$. Where the order r of σ is not clear from the context, we write σ_r. Let $\mathbf{S}_r$ denote the set of subvectors σ of order r; $\mathbf{S}_r$ contains ${}_n\mathbf{C}_r$ elements. For $r = 0,\ldots,n-1$, let $\mathbf{A}_{\sigma_r}$ denote the matrix formed from the rows and columns of $\mathbf{A}$ which are not contained in σ_r; i.e., the matrix formed by deleting the rows and columns in σ_r. The determinant $|\mathbf{A}_{\sigma_r}|$ is termed an $(n-r)$th order principal minor of $\mathbf{A}$.

We assume the following basic matrix results to be known:

(1) Associated with the symmetric matrix $\mathbf{A}$ are n (not necessarily distinct) real characteristic values, given by the roots of the polynomial $P(\lambda) = |\mathbf{A} - \lambda\mathbf{I}|$.

(2) The characteristic polynomial has the expansion

$$P(\lambda) = (-\lambda)^n + k_{n-1}(-\lambda)^{n-1} + \cdots + k_1(-\lambda) + k_0,$$

where

$$k_j = \sum_{\sigma_j \in S_j} |\mathbf{A}_{\sigma_j}|$$

is the sum of the $(n-j)$th order principal minors of $\mathbf{A}$.

(3) There exists a matrix $\mathbf{T}$ such that $\mathbf{T'T} = \mathbf{I}$ and $\mathbf{T'AT} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the characteristic values of $\mathbf{A}$.

(4) The determinant of a matrix equals the product of its characteristic values, and the determinant of a product of matrices equals the product of the determinants.

(5) If $\mathbf{A}$ is positive definite (i.e., $\mathbf{x'x} = 1$ implies $\mathbf{x'Ax} > 0$), then $\mathbf{A}_{\sigma_r}$ is positive definite for each σ_r.

(6) $\mathbf{A}$ is positive definite if and only if all the characteristic values of $\mathbf{A}$ are positive.

The following result relates positive definiteness to properties of the principal minors.

Lemma 1. The following conditions are equivalent:

(i) $\mathbf{A}$ is positive definite.
(ii) All principal minors of $\mathbf{A}$ are positive; i.e., $|\mathbf{A}_\sigma| > 0$ for all $\sigma \in \mathbf{S}_r$ $r = 0,\ldots,n-1$.
(iii) The sum of the principle minors of order $n - r$ is positive for $r = 0,\ldots,n-1$; i.e., $\sum_{s \in S_r} |\mathbf{A}_\sigma| > 0$.
(iv) For each $r = 0,\ldots,n-1$, $|\mathbf{A}_{\sigma_r}| \geqq 0$ and for at least one $\sigma \in \mathbf{S}_r$, $|\mathbf{A}_\sigma| > 0$.
(v) There exists at least one nested sequence of positive principal minors; i.e., there exist $\sigma_0 \subseteq \sigma_1 \subseteq \cdots \subseteq \sigma_{n-1}$ such that $|\mathbf{A}_{\sigma_r}| > 0$.

Proof: The equivalence of (i) and (ii) follows from the basic matrix results (4), (5), and (6). Clearly, (ii) implies (iv) implies (iii) and (ii) implies (v). The proof is completed by showing (iii) implies (i) and (v) implies (i).

Suppose (iii) holds. Then, the coefficient k_{n-j} in the characteristic polynomial $|\mathbf{A} - \lambda\mathbf{I}| = (-\lambda)^n + k_{n-1}(-\lambda)^{n-1} + \cdots + k_0$ equals $\sum_{\sigma \in S_{n-j}} |\mathbf{A}_\sigma|$. Since the characteristic values of $\mathbf{A}$ are all real and the k_{n-j} are all positive, Descarte's rule of signs implies the roots are all positive. Basic result (6) then implies (i).

Suppose (v) holds. We shall show that (i) holds by induction on the order of the matrix n. The result holds trivially for $n = 1$. Suppose that it has been proved for matrices of order up through $n - 1$, and consider a matrix $\mathbf{A}$ of order n with property (v). Assume without loss of generality that the rows and columns of $\mathbf{A}$ have been permuted so that the nested sequence of positive minors is formed by successively deleting the first (remaining) row and column of $|\mathbf{A}|$. Write $\mathbf{A}$ as a partitioned matrix

$$\mathbf{A} = \left[\begin{array}{c|c} a_{11} & \mathbf{a}' \\ \hline \mathbf{a} & \mathbf{A}_\sigma \end{array}\right],$$

where $\mathbf{a}' = (a_{12},\ldots,a_{1n})$ and $\sigma = (1)$. Expanding $|\mathbf{A}|$ about elements of the first row and the first column yields

$$|\mathbf{A}| = a_{11}|\mathbf{A}_\sigma| + \sum_{i=2}^{n}(-1)^i a_{1i}\sum_{j=2}^{n}(-1)^{j-1}a_{j1}|\mathbf{A}_\sigma^{ij}|,$$

where $|\mathbf{A}_\sigma^{ij}|$ is the minor formed from $|\mathbf{A}_\sigma|$ by deleting row i and column j. Since $|\mathbf{A}_\sigma| \neq 0$, $\mathbf{A}_\sigma^{-1}$ exists and its ijth element is $(\mathbf{A}_\sigma^{-1})_{ij} = (-1)^{i+j}|\mathbf{A}_\sigma^{ij}|/|\mathbf{A}_\sigma|$. Hence,

$$|\mathbf{A}| = a_{11}|\mathbf{A}_\sigma| - |\mathbf{A}_\sigma|\mathbf{a}'\mathbf{A}_\sigma^{-1}\mathbf{a}.$$

Since (v) implies (i) for matrices of order $n-1$, $\mathbf{A}_\sigma$ is positive definite, implying $\mathbf{A}_\sigma^{-1}$ positive definite. Now consider a vector $(x_0,\mathbf{x}')$ with $\mathbf{x}'$ of order $n-1$, $x_0^2 + \mathbf{x}'\mathbf{x} = 1$, and consider the quadratic form

$$Q = (x_0,\mathbf{x}')\mathbf{A}\begin{pmatrix} x_0 \\ \mathbf{x} \end{pmatrix} = x_0^2 a_{11} + 2x_0\mathbf{a}'\mathbf{x} + \mathbf{x}'\mathbf{A}_\sigma\mathbf{x}.$$

Using the expression obtained above to eliminate a_{11},

$$\begin{aligned} Q &= x_0^2|\mathbf{A}|/|\mathbf{A}_\sigma| + x_0^2\mathbf{a}'\mathbf{A}_\sigma^{-1}\mathbf{a} + 2x_0\mathbf{a}'\mathbf{A}_\sigma^{-1}\mathbf{A}_\sigma\mathbf{x} + \mathbf{x}'\mathbf{A}_\sigma\mathbf{A}_\sigma^{-1}\mathbf{A}_\sigma\mathbf{x} \\ &= x_0^2|\mathbf{A}|/|\mathbf{A}_\sigma| + (x_0\mathbf{a} + \mathbf{A}_\sigma\mathbf{x})'\mathbf{A}_\sigma^{-1}(x_0\mathbf{a} + \mathbf{A}_\sigma\mathbf{x}). \end{aligned}$$

Since $\mathbf{A}_\sigma^{-1}$ is positive definite, the second term in Q is non-negative, and is positive if $x_0 = 0$, while the first term is non-negative and is positive if $x_0 \neq 0$. Hence $Q > 0$ and $\mathbf{A}$ is positive definite. By induction, (v) implies (i). Q.E.D.

An example shows that it is essential that the positive principal minors in condition (v) of the lemma be nested. The matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

is not positive definite, but has a positive principal minor of each order.

2. Conditions for a Matrix to be Positive Definite Subject to Constraint

Let $\mathbf{A}$ denote a real symmetric matrix of order n, and $\mathbf{B}$ denote a real $m \times n$ matrix of rank m, with $m < n$. We say $\mathbf{A}$ is positive definite subject to constraint $\mathbf{B}$ if $\mathbf{x}'\mathbf{x} = 1$ and $\mathbf{Bx} = \mathbf{0}$ implies $\mathbf{x}'\mathbf{Ax} > 0$.

> *Lemma* 2. $\mathbf{A}$ is positive definite subject to constraint $\mathbf{B}$ if and only if there exists $\lambda_0 \geqq 0$ such that $\mathbf{A} + \lambda \mathbf{B}'\mathbf{B}$ is positive definite for $\lambda \geqq \lambda_0$. (*Note*: This lemma does not require that $\mathbf{B}$ be of full rank.)

Proof: Suppose $\mathbf{A} + \lambda \mathbf{B}'\mathbf{B}$ is positive definite. Then $\mathbf{x}'\mathbf{x} = 1$ implies $\mathbf{x}'(\mathbf{A} + \lambda \mathbf{B}'\mathbf{B})\mathbf{x} > 0$. If $\mathbf{Bx} = \mathbf{0}$, then $\mathbf{x}'\mathbf{Ax} > 0$ and $\mathbf{A}$ is positive definite subject to constraint $\mathbf{B}$.

Suppose there exists a sequence $\lambda_k \to +\infty$ such that $\mathbf{A} + \lambda_k \mathbf{B}'\mathbf{B}$ fails to be positive definite; i.e., there exists $\mathbf{x}_k$ such that $\mathbf{x}_k'\mathbf{x}_k = 1$ and $\mathbf{x}_k'(\mathbf{A} + \lambda_k \mathbf{B}'\mathbf{B})\mathbf{x}_k \leqq 0$. The $\mathbf{x}_k$ have a subsequence converging to $\mathbf{x}_*$. Since $\mathbf{x}_k'\mathbf{A}\mathbf{x}_k$ is bounded, $\mathbf{x}_k'\mathbf{B}'\mathbf{B}\mathbf{x}_k \geqq 0$, and $\lambda_k \to +\infty$, we have $\lim_k \mathbf{B}\mathbf{x}_k = \mathbf{B}\mathbf{x}_* = \mathbf{0}$ and $\mathbf{x}_*'\mathbf{A}\mathbf{x}_* \leqq \limsup -\lambda_k \mathbf{x}_k'\mathbf{B}'\mathbf{B}\mathbf{x}_k \leqq 0$. But $\mathbf{B}\mathbf{x}_* = \mathbf{0}$ and $\mathbf{x}_*'\mathbf{A}\mathbf{x}_* \leqq 0$ imply $\mathbf{A}$ is not positive definite subject to constraint $\mathbf{B}$. Q.E.D.

In examining conditions for $\mathbf{A}$ to be positive definite subject to constraint $\mathbf{B}$, we shall utilize the symmetric $n+m$ bordered matrix

$$\mathbf{H} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B}' \\ \hline \mathbf{B} & \mathbf{0} \end{array}\right].$$

The relevance of this matrix is established by the following argument: A necessary and sufficient condition for $\mathbf{A}$ to be positive definite subject to constraint $\mathbf{B}$ is that one have a positive solution to

$$\operatorname{Min}\{\mathbf{x}'\mathbf{Ax} \mid \mathbf{x}'\mathbf{x} = 1 \text{ and } \mathbf{Bx} = \mathbf{0}\}. \tag{1}$$

The following lemma establishes that the solution of (1) can be found by examining the critical points of the Lagrangian

$$\mathbf{L} = \tfrac{1}{2}\mathbf{x}'\mathbf{Ax} + \tfrac{1}{2}\lambda(1 - \mathbf{x}'\mathbf{x}) + \mathbf{p}'\mathbf{Bx}. \tag{2}$$

> *Lemma* 3. If $\mathbf{x}$ solves (1), then there exist $\mathbf{p}$ and λ such that $\lambda = \mathbf{x}'\mathbf{Ax}$, and

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B}' \\ \hline \mathbf{B} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{p} \end{array}\right] = \lambda \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{0} \end{array}\right]. \tag{3}$$

Conversely, for any λ solving

$$\left|\begin{array}{c:c} \mathbf{A}-\lambda\mathbf{I} & \mathbf{B}' \\ \hdashline \mathbf{B} & \mathbf{0} \end{array}\right| = 0, \tag{4}$$

there exists $\mathbf{x}$, $\mathbf{p}$ such that $\mathbf{x}'\mathbf{x} = 1$, (3) holds, and $\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda$.

Proof: We give a proof of this lemma involving only elementary matrix manipulations. A shorter and more elegant proof could alternately be given by first making an orthogonal change of basis.[1]

We solve (1) by elimination of the constraint. By hypothesis, $\mathbf{B}$ is of rank m and can be partitioned

$$\mathbf{B} = [\mathbf{B}_1 \quad \mathbf{B}_2],$$

where $\mathbf{B}_1$ is $m \times m$ and non-singular and $\mathbf{B}_2$ is $m \times (n-m)$. Partition $\mathbf{x}$ and $\mathbf{A}$ commensurately,

$$\mathbf{x}' = (\mathbf{x}_1' \quad \mathbf{x}_2') \text{ and } \mathbf{A} = \left[\begin{array}{c:c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hdashline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right].$$

If $\mathbf{B}\mathbf{x} = \mathbf{B}_1\mathbf{x}_1 + \mathbf{B}_2\mathbf{x}_2 = 0$, then $\mathbf{x}_1 = -\mathbf{B}_1^{-1}\mathbf{B}_2\mathbf{x}_2$ and $\mathbf{x}_1'\mathbf{x}_1 + \mathbf{x}_2'\mathbf{x}_2 = \mathbf{x}_2'[\mathbf{I}_2 + \mathbf{B}_2'(\mathbf{B}_1')^{-1}(\mathbf{B}_1)^{-1}\mathbf{B}_2]\mathbf{x}_2$. Let $\mathbf{C}_2 = [\mathbf{I}_2 + \mathbf{B}_2'(\mathbf{B}_1\mathbf{B}_1')^{-1}\mathbf{B}_2]^{-1/2}$ and $\mathbf{z}_2 = \mathbf{C}_2^{-1}\mathbf{x}_2$. Then $\mathbf{x}'\mathbf{x} = \mathbf{z}_2'\mathbf{z}_2$. Also,

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{z}_2'\mathbf{C}_2[\mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{11}\mathbf{B}_1^{-1}\mathbf{B}_2 - \mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{B}_1^{-1}\mathbf{B}_2 + \mathbf{A}_{22}]\mathbf{C}_2\mathbf{z}_2.$$

The minimum of this expression subject to $\mathbf{x}'\mathbf{x} = \mathbf{z}_2'\mathbf{z}_2 = 1$ is attained for a characteristic vector $\hat{\mathbf{z}}_2$ of this matrix giving its minimum characteristic value λ. Then

$$\mathbf{C}_2[\mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{11}\mathbf{B}_1^{-1}\mathbf{B}_2 - \mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{B}_1^{-1}\mathbf{B}_2 + \mathbf{A}_{22}]\mathbf{C}_2\hat{\mathbf{z}}_2 = \lambda\hat{\mathbf{z}}_2,$$

or

$$[\mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{11}\mathbf{B}_1^{-1}\mathbf{B}_2 - \mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{12} - \mathbf{A}_{21}\mathbf{B}_1^{-1}\mathbf{B}_2 + \mathbf{A}_{22}]\hat{\mathbf{x}}_2 = \lambda[\mathbf{I}_2 + \mathbf{B}_2'(\mathbf{B}_1\mathbf{B}_1')^{-1}\mathbf{B}_2]\hat{\mathbf{x}}_2,$$

where $\hat{\mathbf{x}}_2 = \mathbf{C}_2\hat{\mathbf{z}}_2$. For the above value of λ, we wish to show that there exists a $\mathbf{p}$ such that (3) is satisfied. A solution must satisfy

$$\begin{aligned} \mathbf{A}_{11}\hat{\mathbf{x}}_1 + \mathbf{A}_{12}\hat{\mathbf{x}}_2 + \mathbf{B}_1'\mathbf{p} &= \lambda\hat{\mathbf{x}}_1, \\ \mathbf{A}_{21}\hat{\mathbf{x}}_1 + \mathbf{A}_{22}\hat{\mathbf{x}}_2 + \mathbf{B}_2'\mathbf{p} &= \lambda\hat{\mathbf{x}}_2. \end{aligned} \tag{5}$$

[1] There exists an orthogonal $n \times n$ matrix $\mathbf{S}$ such that $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{S}' = [\mathbf{I}_m \vdots \mathbf{0}_{n-m}]$, where $\mathbf{I}_m$ is the identity matrix and $\mathbf{0}_{n-m}$ is an $m \times (n-m)$ matrix of zeroes. Let $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}\mathbf{S}'$. Then, it is straightforward to establish Lemma 3 for the problem of minimizing $\mathbf{x}'\tilde{\mathbf{A}}\mathbf{x}$ subject to $\mathbf{x}'\mathbf{x} = 1$ and $\tilde{\mathbf{B}}\mathbf{x} = 0$, and then to show that (3) and (4) are invariant under an orthogonal change of bias. This argument was suggested by S. Cosslett.

Solving for $\mathbf{p}$ from the first set of equations yields

$$\begin{aligned}\mathbf{p} &= (\mathbf{B}_1')^{-1}[\lambda\hat{\mathbf{x}}_1 - \mathbf{A}_{11}\hat{\mathbf{x}}_1 - \mathbf{A}_{12}\hat{\mathbf{x}}_2] \\ &= (\mathbf{B}_1')^{-1}[-\lambda\mathbf{B}_1^{-1}\mathbf{B}_2\hat{\mathbf{x}}_2 + \mathbf{A}_{11}\mathbf{B}_1^{-1}\mathbf{B}_2\hat{\mathbf{x}}_2 - \mathbf{A}_{12}\hat{\mathbf{x}}_2].\end{aligned}$$

Substituting this expression into the second set of equations,

$$\begin{aligned}\mathbf{A}_{21}\hat{\mathbf{x}}_1 + \mathbf{A}_{22}\hat{\mathbf{x}}_2 + \mathbf{B}_2'\mathbf{p} - \lambda\hat{\mathbf{x}}_2 &= -\mathbf{A}_{21}\mathbf{B}_1^{-1}\mathbf{B}_2\hat{\mathbf{x}}_2 + \mathbf{A}_{22}\hat{\mathbf{x}}_2 - \lambda\mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{B}_1^{-1}\mathbf{B}_2\hat{\mathbf{x}}_2 \\ &\quad + \mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{11}\mathbf{B}_1^{-1}\mathbf{B}_2\hat{\mathbf{x}}_2 - \mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{12}\hat{\mathbf{x}}_2 - \lambda\hat{\mathbf{x}}_2 \\ &= -\lambda[\mathbf{I} + \mathbf{B}_2'(\mathbf{B}_1\mathbf{B}_1')^{-1}\mathbf{B}_2]\hat{\mathbf{x}}_2 \\ &\quad + [\mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{11}\mathbf{B}_1^{-1}\mathbf{B}_2 - \mathbf{A}_{21}\mathbf{B}_1^{-1}\mathbf{B}_2 \\ &\quad - \mathbf{B}_2'(\mathbf{B}_1')^{-1}\mathbf{A}_{12} + \mathbf{A}_{22}]\hat{\mathbf{x}}_2 = \mathbf{0},\end{aligned}$$

by the characteristic value property. Hence, the system (5) has a consistent solution for $\mathbf{p}$. Premultiplying the last equation by $\hat{\mathbf{x}}_2$, and noting that $\hat{\mathbf{x}}_2'[\mathbf{I} + \mathbf{B}_2'(\mathbf{B}_1'\mathbf{B}_1)^{-1}\mathbf{B}_2]\hat{\mathbf{x}}_2 = 1$ and $\hat{\mathbf{x}}_1 = -\mathbf{B}_1^{-1}\mathbf{B}_2\hat{\mathbf{x}}_2$, we obtain $\lambda = \hat{\mathbf{x}}'\mathbf{A}\hat{\mathbf{x}}$. Hence, (3) holds.

Suppose λ is a root of the polynomial (4). Then equation (3) holds for some $\mathbf{x}$ and $\mathbf{p}$ not both identically zero. If $\mathbf{x} = \mathbf{0}$, one has $\mathbf{B}'\mathbf{p} = \mathbf{0}$, contradicting the assumption that $\mathbf{B}$ is of rank m. Hence $\mathbf{x} \neq \mathbf{0}$ and we can normalize $\mathbf{x}'\mathbf{x} = 1$. Then $\mathbf{x}'(\mathbf{A}\mathbf{x} + \mathbf{B}'\mathbf{p}) = \lambda\mathbf{x}'\mathbf{x}$ and $\mathbf{B}\mathbf{x} = \mathbf{0}$ implies $\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda\mathbf{x}'\mathbf{x} = \lambda$. Q.E.D.

As before, let σ be a subvector of $(1,\ldots,n)$, and $\mathbf{A}_\sigma$ be the matrix formed from the rows and columns of $\mathbf{A}$ not contained in σ. For the $m \times n$ matrix $\mathbf{B}$, let $\mathbf{B}_\sigma$ be the submatrix formed by deleting columns from σ. Hence, if $\sigma = \sigma_r$ contains r elements, then $\mathbf{A}_\sigma$ is a square matrix of order $n - r$ and $\mathbf{B}_\sigma$ is $m \times (n - r)$. The following result relates positive definiteness subject to constraint to properties of the principal minors.

Lemma 4. If $\mathbf{A}$ is an $n \times n$ symmetric matrix and $\mathbf{B}$ is an $m \times n$ matrix of rank $m < n$, then the following conditions are equivalent:

(i) $\mathbf{A}$ is positive definite subject to constraint $\mathbf{B}$.
(ii) For $r = 0,\ldots,n - m$,

$$(-1)^m \sum_{\sigma \in \mathbf{S}_r} \begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}_\sigma' \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix} > 0.$$

(iii) For $r = 0,\ldots,n - m$, and $\sigma \in \mathbf{S}_r$,

$$(-1)^m \begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}_\sigma' \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix} \geqq 0,$$

with strict inequality holding for at least one $\sigma \in \mathbf{S}_r$.

(iv) There exists at least one nested sequence of principle minors of order $2m$ through $n+m$ formed by deleting symmetric rows and columns from the first n rows and columns, which have the sign of $(-1)^m$; i.e., there exist $\sigma_0 \subseteq \sigma_1 \subseteq \cdots \subseteq \sigma_{n-m}$ such that

$$(-1)^m \begin{vmatrix} \mathbf{A}_{\sigma_r} & \mathbf{B}'_{\sigma_r} \\ \mathbf{B}_{\sigma_r} & \mathbf{0} \end{vmatrix} > 0.$$

Further, note that (iv) implies $|\mathbf{B}_{\sigma_{n-m}}| \neq 0$, while if (i) holds and $|\mathbf{B}_{\sigma_{n-m}}| \neq 0$ for some σ_{n-m}, then any nested sequence starting from σ_{n-m} satisfies (iv).

Proof: We first show that (i) and (ii) are equivalent. By Lemma 3, (i) holds if and only if the roots of the polynomial in (4) are all positive. Expand this polynomial in powers of $(-\lambda)$,

$$\begin{vmatrix} \mathbf{A}-\lambda\mathbf{I} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0} \end{vmatrix} = \sum_{j=0}^{n+m} k_j(-\lambda)^j. \tag{6}$$

The argument below establishes that

$$k_j = 0 \quad \text{for} \quad j > n-m,$$

and

$$k_j = \sum_{\sigma \in \mathbf{S}_j} \begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix} \quad \text{for} \quad j \leqq n-m.$$

A term in a full expansion of the determinant (6) is the product of $n-l$ elements from the northwest submatrix, l elements from each of the northeast and southwest submatrices, and $m-l$ elements from the southeast submatrix. This term can be non-zero only if $m-l=0$, implying at most $n-m$ elements are taken from the northwest submatrix. Hence, $k_j = 0$ for $j > n-m$. Consider the collection of all terms in a full expansion of (6) which contribute to k_j for some $j \leqq n-m$. Each such term is the product of factors $(-\lambda)$ taken from diagonal elements of the northwest submatrix in (6), corresponding to columns in σ_j; multiplied by a term in the expansion of the determinant formed by deleting the σ_j rows and columns from (6); i.e., a term in the expansion of

$$\begin{vmatrix} \mathbf{A}_{\sigma_j} & \mathbf{B}'_{\sigma_j} \\ \mathbf{B}_{\sigma_j} & \mathbf{0} \end{vmatrix}.$$

Then, k_j equals the sum over $\sigma \in \mathbf{S}_j$ of terms of the form

$$\begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix},$$

as was to be demonstrated.

For $\sigma \in \mathbf{S}_{n-m}$,

$$\begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix} = (-1)^{m^2} \begin{vmatrix} \mathbf{B}'_\sigma & \mathbf{A}_\sigma \\ \mathbf{0} & \mathbf{B}_\sigma \end{vmatrix} = (-1)^m |\mathbf{B}_\sigma|^2.$$

Since $\mathbf{B}$ is of rank m, $|\mathbf{B}_\sigma| \neq 0$ for some $\sigma \in \mathbf{S}_{n-m}$, implying $(-1)^m k_{n-m} > 0$. By Descartes' rule of signs, the roots of the polynomial (6), which are real since the matrix is symmetric, are all positive if and only if the coefficients $k_0,\dots,k_{n-m}$ are of uniform sign. Hence, (i) and (ii) are equivalent.

It is trivial that (iii) implies (ii). We next show that (i) implies (iii). If A is positive definite subject to constraint B, then for any $\sigma \in \mathbf{S}_r$, $r = 0,\dots,n-m$, $\mathbf{A}_\sigma$ is positive definite subject to constraint $\mathbf{B}_\sigma$. (This can be seen by setting the components of $\mathbf{x}$ outside of σ equal to zero in the definition $\mathbf{x}'\mathbf{x} = 1$, $\mathbf{B}\mathbf{x} = \mathbf{0}$ implies $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$.) By Lemma 2, $\mathbf{A}_\sigma + \lambda \mathbf{B}'_\sigma \mathbf{B}_\sigma$ is positive definite for λ sufficiently large. Further, if $\mathbf{B}_\sigma$ is of rank m, then by Lemma 3,

$$\begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix} \neq 0.$$

Consider the matrices

$$\mathbf{C}_\sigma = \begin{vmatrix} \mathbf{A}_\sigma & \lambda \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & -\mathbf{I}_m \end{vmatrix} \text{ and } \mathbf{E}_\sigma = \begin{vmatrix} \mathbf{I}_\sigma & \mathbf{0} \\ \mathbf{B}_\sigma & \mathbf{I}_m \end{vmatrix}.$$

Then

$$\mathbf{C}_\sigma \mathbf{E}_\sigma = \begin{vmatrix} \mathbf{A}_\sigma + \lambda \mathbf{B}'_\sigma \mathbf{B}_\sigma & \lambda \mathbf{B}'_\sigma \\ \mathbf{0} & -\mathbf{I}_m \end{vmatrix},$$

$$|\mathbf{E}_\sigma| = 1 \quad \text{and} \quad |\mathbf{C}_\sigma|\,|\mathbf{E}_\sigma| = |\mathbf{C}_\sigma \mathbf{E}_\sigma| = (-1)^m |\mathbf{A}_\sigma + \lambda \mathbf{B}'_\sigma \mathbf{B}_\sigma|.$$

Consider the limit

$$\lim_{\lambda \to +\infty} \lambda^{-m}(-1)^m |\mathbf{C}_\sigma| = \lim_{\lambda \to +\infty} (-1)^m \begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & -(1/\lambda)\mathbf{I}_m \end{vmatrix} = (-1)^m \begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix}.$$

Since $(-1)^m |\mathbf{C}_\sigma| > 0$, one obtains in the limit

$$(-1)^m \begin{vmatrix} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \mathbf{B}_\sigma & \mathbf{0} \end{vmatrix} \geqq 0. \tag{7}$$

Further, if $\mathbf{B}_\sigma$ is of rank m, the inequality in (7) is strict. Since $\mathbf{B}_\sigma$ must be of full rank for some $\sigma \in \mathbf{S}_r$, this establishes that (iii) holds.

Next we show that (i) and (iv) are equivalent. Suppose (iv) holds.

Then for $\sigma = \sigma_{n-m}$,

$$0 < (-1)^m \left|\begin{array}{c:c} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \hdashline \mathbf{B}_\sigma & \mathbf{0} \end{array}\right| = |\mathbf{B}_\sigma|^2,$$

implying $\mathbf{B}'_\sigma\mathbf{B}_\sigma$ positive definite. Hence, for λ sufficiently large, $\mathbf{A}_{\sigma_{n-m}} + \lambda\mathbf{B}'_{\sigma_{n-m}}\mathbf{B}_{\sigma_{n-m}}$ is positive definite, implying the existence of a nested sequence of positive principle minors of $\mathbf{A}_{\sigma_{n-m}} + \mathbf{B}'_{\sigma_{n-m}}\mathbf{B}_{\sigma_{n-m}}$. Further, (iv) implies the existence of a nested sequence $\sigma_0 \subseteq \cdots \subseteq \sigma_{n-m}$, such that for $\sigma = \sigma_r$, $r = 0,\ldots,n-m$,

$$(-1)^m \left|\begin{array}{c:c} \mathbf{A}_\sigma & \mathbf{B}'_\sigma \\ \hdashline \mathbf{B}_\sigma & \mathbf{0} \end{array}\right| > 0. \tag{8}$$

Since $\lim_{\lambda\to\infty}(-\lambda)^{-m}|\mathbf{C}_\sigma|$ equals the left-hand side of (8), we conclude that for λ sufficiently large, $(-1)^m|\mathbf{C}_\sigma| = |\mathbf{A}_\sigma + \lambda\mathbf{B}'_\sigma\mathbf{B}_\sigma|$ is positive for $r = 0,\ldots,n-m$. Hence, since $(\mathbf{A} + \lambda\mathbf{B}'\mathbf{B})_\sigma = \mathbf{A}_\sigma + \lambda\mathbf{B}'_\sigma\mathbf{B}_\sigma$, we have established the existence of a full nested sequence of positive principal minors of $\mathbf{A} + \lambda\mathbf{B}'\mathbf{B}$. By Lemma 1, $\mathbf{A} + \lambda\mathbf{B}'\mathbf{B}$ is then positive definite, and by Lemma 2, condition (i) holds.

Next suppose (i) holds, and choose σ_{n-m} such that $|\mathbf{B}_{\sigma_{n-m}}| \neq 0$. (This can be done since $\mathbf{B}$ is of rank m.) Choose any σ_r such that $\sigma_0 \subseteq \sigma_1 \subseteq \cdots \subseteq \sigma_{n-m}$ for $r = 0,\ldots,n-m-1$. The proof above that (i) implies (iii) establishes that

$$(-1)^m \left|\begin{array}{c:c} \mathbf{A}_{\sigma_r} & \mathbf{B}'_{\sigma_r} \\ \hdashline \mathbf{B}_{\sigma_r} & \mathbf{0} \end{array}\right| > 0,$$

for $r = 0,\ldots,n-m$. Hence, (iv) holds.

The note at the end of the lemma follows from the fact that

$$(-1)^m \left|\begin{array}{c:c} \mathbf{A}_{\sigma_{n-m}} & \mathbf{B}'_{\sigma_{n-m}} \\ \hdashline \mathbf{B}_{\sigma_{n-m}} & \mathbf{0} \end{array}\right| = |\mathbf{B}_{\sigma_{n-m}}|^2,$$

and from the construction used to establish that (i) implies (iii) and (iv). Q.E.D.

Appendix A.2

NECESSARY AND SUFFICIENT CONDITIONS FOR THE CLASSICAL PROGRAMMING PROBLEM

DANIEL McFADDEN

University of California, Berkeley

We consider the following classical programming problem:

$$\begin{aligned}&\text{Min } F(x_1,\ldots,x_n),\\ &\text{s.t. } g^i(x_1,\ldots,x_n)+b_i=0 \quad \text{for} \quad i=1,\ldots,m<n.\end{aligned} \tag{CPP}$$

Letting

$$\underset{1\times n}{\mathbf{x}'} = (x_1,\ldots,x_n),$$

and defining

$$\underset{1\times m}{\mathbf{G}(\mathbf{x})} = (g^1(x),\ldots,g^m(x)),$$

this problem is written in matrix notation as

$$\text{Min } F(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{G}(\mathbf{x})+\mathbf{b}'=0.$$

A point $\bar{\mathbf{x}}$ is a *local solution* of CPP if $\mathbf{G}(\bar{\mathbf{x}})+\mathbf{b}'=\mathbf{0}$ and there exists a neighborhood N of $\bar{\mathbf{x}}$ such that $\mathbf{x}\in \mathbf{N}$ and $\mathbf{G}(\mathbf{x})+\mathbf{b}'=\mathbf{0}$ implies $F(\mathbf{x})\geqq F(\bar{\mathbf{x}})$.

We associate with this problem the Lagrangian

$$L(\mathbf{x},\mathbf{p}) = F(\mathbf{x}) + [\underset{1\times m}{\mathbf{G}(\mathbf{x})}+\mathbf{b}']\underset{m\times 1}{\mathbf{p}},$$

where $\mathbf{p}'=(p_1,\ldots,p_m)$ is a vector of Lagrangian multipliers. We assume hereafter that the functions F, $g^1,\ldots,g^m$ are twice continuously differentiable, and define

$$\underset{n\times 1}{F_{\mathbf{x}}}=\begin{bmatrix}\partial F/\partial x_1\\ \vdots\\ \partial F/\partial x_n\end{bmatrix},$$

$$\underset{n\times m}{\mathbf{G}_{\mathbf{x}}}=\begin{bmatrix}\partial g^1/\partial x_1\cdots\partial g^m/\partial x_1\\ \vdots \qquad \vdots\\ \partial g^1/\partial x_n\cdots\partial g^m/\partial x_n\end{bmatrix},$$

and

$$F_{\mathbf{xx}}=\begin{bmatrix}\partial^2F/\partial x_1^2 & \cdots\partial^2F/\partial x_1\partial x_n\\ \vdots & \vdots\\ \partial^2F/\partial x_n\partial x_1 & \cdots\partial^2F/\partial x_n^2\end{bmatrix}.$$

Note that

$$L_{\mathbf{x}}=F_{\mathbf{x}}+G_{\mathbf{x}}\mathbf{p} \quad\text{and}\quad L_{\mathbf{p}}=\mathbf{G}(\mathbf{x})+\mathbf{b}',$$

and that

$$L_{\mathbf{xx}}=F_{\mathbf{xx}}+\sum_{i=1}^{m}g_{\mathbf{xx}}^{i}p_i.$$

A *Lagrangian critical point* is a vector

$$(\underset{n\times 1}{\bar{\mathbf{x}}},\underset{m\times 1}{\bar{\mathbf{p}}}),$$

such that

$$L_{\mathbf{x}}(\bar{\mathbf{x}},\bar{\mathbf{p}})=0 \quad\text{and}\quad L_{\mathbf{p}}(\bar{\mathbf{x}},\bar{\mathbf{p}})=0. \tag{LCP}$$

The first result of classical optimization theory establishes that under mild non-degeneracy conditions on the constraints, each local solution of CPP will correspond to a Lagrangian critical point. We say CPP is *strongly non-degenerate* at a point $\mathbf{x}$ if

$$\operatorname{rank}\mathbf{G}_{\mathbf{x}}(\mathbf{x})=m. \tag{SND}$$

We say CPP is *weakly non-degenerate* at a point $\mathbf{x}$ if

$$\operatorname{rank}\underset{n\times m}{\mathbf{G}_{\mathbf{x}}(\mathbf{x})}=\operatorname{rank}[\underset{n\times m}{\mathbf{G}_{\mathbf{x}}(\mathbf{x})}\vdots\underset{n\times 1}{F_{\mathbf{x}}(\mathbf{x})}]. \tag{WND}$$

Condition (SND) will hold in most practical programming problems, and can be made to hold in any CPP by an arbitrarily small perturbation of the constraints. Hence, we concentrate our attention on problems satisfying this condition.

Theorem 1. Suppose $\bar{\mathbf{x}}$ is a local solution of CPP, and SND holds at $\bar{\mathbf{x}}$. Then there exists $\bar{\mathbf{p}}$ such that $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is a LCP.

The proof of this theorem can be found in many textbooks, and will not be repeated here. See, for example, Intriligator (1971), pp. 31–33).

The next result establishes that at a local solution $\bar{\mathbf{x}}$ of CPP, SND implies WND, and WND holds if and only if $\bar{\mathbf{x}}$ corresponds to a LCP.

Theorem 2. Suppose $\bar{\mathbf{x}}$ is a local solution of CPP. If SND holds at $\bar{\mathbf{x}}$, then WND holds at $\bar{\mathbf{x}}$. There exists $\bar{\mathbf{p}}$ such that $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is a LCP if and only if WND holds at $\bar{\mathbf{x}}$.

Proof: If SND holds at $\bar{\mathbf{x}}$, then Theorem 1 implies the existence of $\bar{\mathbf{p}}$ such that $F_x(\bar{\mathbf{x}}) = -\mathbf{G}_x(\bar{\mathbf{x}})\bar{\mathbf{p}}$. But this implies rank $\mathbf{G}_x(\bar{\mathbf{x}}) = \text{rank}[\mathbf{G}_x(\bar{\mathbf{x}}) \vdots F_x(\bar{\mathbf{x}})]$, and WND holds.

The theorem of the alternative for the solution of linear equations states that there exists $\bar{\mathbf{p}}$ such that $F_x(\bar{\mathbf{x}}) = -\mathbf{G}_x(\bar{\mathbf{x}})\bar{\mathbf{p}}$ if and only if condition WND holds. But this is precisely the condition needed for the existence of $\bar{\mathbf{p}}$ such that $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is a LCP, since $L_p(\bar{\mathbf{x}},\bar{\mathbf{p}}) = \mathbf{G}(\bar{\mathbf{x}}) + \mathbf{b}' = \mathbf{0}$ is satisfied by assumption. Q.E.D.

The next result establishes that at a strongly non-degenerate local solution of CPP, the Hessian matrix of the Lagrangian, L_{xx}, is positive semidefinite subject to constraint.

Theorem 3. Suppose $\bar{\mathbf{x}}$ is a local solution of CPP and SND holds at $\bar{\mathbf{x}}$. Then, $\mathbf{z}'\mathbf{z} = 1$ and $\mathbf{z}'\mathbf{G}_x(\bar{\mathbf{x}}) = 0$ imply $\mathbf{z}'L_{xx}(\bar{\mathbf{x}},\bar{\mathbf{p}})\mathbf{z} \geqq 0$, where $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is the LCP whose existence is established by Theorem 1.

Proof: As in the proof of Theorem 1, SND implies that the system of equations

$$\mathbf{G}(\mathbf{v},\mathbf{w}) + \mathbf{b}' = \mathbf{0},$$

where

$$\mathbf{x}' = (\underset{m\times 1}{\mathbf{v}'}, \underset{(n-m)\times 1}{\mathbf{w}'})$$

is a partition of $\mathbf{x}$ such that $\mathbf{G}_\mathbf{v}(\bar{\mathbf{v}},\bar{\mathbf{w}})$ is non-singular, has a solution

$$\mathbf{v} = \underset{m\times 1}{\mathbf{h}(\mathbf{w})}$$

in a neighborhood of $\bar{\mathbf{w}}$ satisfying $\bar{\mathbf{v}} = \mathbf{h}(\bar{\mathbf{w}})$ and $\mathbf{G}(\mathbf{h}(\mathbf{w}),\mathbf{w}) + \mathbf{b}' \equiv \mathbf{0}$. Define $f(\mathbf{w}) \equiv F(\mathbf{h}(\mathbf{w}),\mathbf{w})$. Then $\bar{\mathbf{w}}$ is an unconstrained local minimum of $f(\mathbf{w})$. Let

$$\underset{(n-m)\times 1}{\mathbf{y}}$$

be a vector satisfying $\mathbf{y}'\mathbf{y} = 1$, and consider $f(\bar{\mathbf{w}} + \theta\mathbf{y})$ as a function of a scalar θ. A Taylor's expansion in θ yields

$$f(\bar{\mathbf{w}} + \theta\mathbf{y}) - f(\bar{\mathbf{w}}) = \theta\mathbf{y}'f_\mathbf{w}(\bar{\mathbf{w}}) + \frac{\theta^2}{2}\mathbf{y}'f_{\mathbf{ww}}(\bar{\mathbf{w}} + \hat{\theta}\mathbf{y})\mathbf{y},$$

where $\hat{\theta}$ is in the interval between 0 and θ. Since $0 \leqq f(\bar{\mathbf{w}} + \theta\mathbf{y}) - f(\bar{\mathbf{w}})$ for θ sufficiently small, we obtain the necessary conditions $f_\mathbf{w}(\bar{\mathbf{w}}) = \mathbf{0}$ and $\mathbf{y}'f_{\mathbf{ww}}(\bar{\mathbf{w}})\mathbf{y} \geqq 0$.

Differentiating the identity $g^i(\mathbf{h}(\mathbf{w}),\mathbf{w}) + b_i \equiv 0$, we obtain

$$\underset{(n-m)\times m}{\mathbf{h}'_\mathbf{w}}\ \underset{m\times 1}{g^i_\mathbf{v}} + \underset{(n-m)\times 1}{g^i_\mathbf{w}} \equiv \sum_{j=1}^{m} \underset{(n-m)\times 1}{h^j_\mathbf{w}}\ \underset{1\times 1}{g^i_{v_j}} + \underset{(n-m)\times 1}{g^i_\mathbf{w}} \equiv \mathbf{0},$$

and

$$0 \equiv \sum_{j=1}^{m} h^j_{\mathbf{ww}} g^i_{v_j} + \sum_{k=1}^{m}\sum_{j=1}^{m} h^j_\mathbf{w}(h^k_\mathbf{w})' g^i_{v_j v_k} + \sum_{j=1}^{m} [h^j_\mathbf{w} g^i_{v_j\mathbf{w}} + g^i_{\mathbf{w}v_j}(h^j_\mathbf{w})'] + g^i_{\mathbf{ww}}. \tag{i}$$

Differentiating $f(\mathbf{w}) \equiv F(\mathbf{h}(\mathbf{w}),\mathbf{w})$, we obtain

$$f_\mathbf{w} \equiv \mathbf{h}'_\mathbf{w} F_\mathbf{v} + F_\mathbf{w} \equiv \sum_{j=1}^{m} h^j_\mathbf{w} F_{v_j} + F_\mathbf{w},$$

and

$$f_{\mathbf{ww}} = \sum_{j=1}^{m} h^j_{\mathbf{ww}} F_{v_j} + \sum_{k=1}^{m}\sum_{j=1}^{m} h^j_\mathbf{w}(h^k_\mathbf{w})' F_{v_j v_k} + \sum_{j=1}^{m} [h^j_\mathbf{w} F_{v_j\mathbf{w}} + F_{\mathbf{w}v_j}(h^j_\mathbf{w})'] + F_{\mathbf{ww}}. \tag{0}$$

Let $\bar{\mathbf{p}}$ be the vector of Lagrangian multipliers given in the proof of Theorem 1, $\bar{\mathbf{p}} = -G_\mathbf{v}(\bar{\mathbf{v}},\bar{\mathbf{w}})^{-1}F_\mathbf{v}(\bar{\mathbf{v}},\bar{\mathbf{w}})$. Multiply each equation (i) by $\bar{p}_i$ and add it to equation (0) to obtain

$$f_{\mathbf{ww}} = \sum_{j=1}^{m} h^{j}_{\mathbf{ww}}\left[F_{v_j} + \sum_{i=1}^{m} \bar{p}_i g^{i}_{v_j}\right] + [\underset{(n-m)\times m}{\mathbf{h}'_{\mathbf{w}}} \,\vdots\, \underset{(n-m)^2}{\mathbf{I}_{\mathbf{w}}}] \left[\begin{array}{c|c} L_{\mathbf{vv}} & L_{\mathbf{vw}} \\ \hline L_{\mathbf{wv}} & L_{\mathbf{ww}} \end{array}\right]\left[\begin{array}{c} \mathbf{h}_{\mathbf{w}} \\ \hline \mathbf{I}_{\mathbf{w}} \end{array}\right].$$

But

$$L_{v_j}(\bar{\mathbf{v}},\bar{\mathbf{w}}) = F_{v_j} + \sum_{i=1}^{m} \bar{p}_i g^{i}_{v_j} = 0,$$

and we have

$$\mathbf{y}' f_{\mathbf{ww}}\mathbf{y} = \mathbf{y}'[\mathbf{h}'_{\mathbf{w}} \vdots \mathbf{I}_{\mathbf{w}}] \left[\begin{array}{c|c} L_{\mathbf{vv}} & L_{\mathbf{vw}} \\ \hline L_{\mathbf{wv}} & L_{\mathbf{ww}} \end{array}\right]\left[\begin{array}{c} \mathbf{h}_{\mathbf{w}} \\ \hline \mathbf{I}_{\mathbf{w}} \end{array}\right] \mathbf{y} \geqq 0.$$

But $\mathbf{z}' = (\mathbf{u}',\mathbf{y}')$ satisfying $\mathbf{z}'\mathbf{G}_{\mathbf{x}}(\bar{\mathbf{x}}) = 0$, or $\mathbf{u}'\mathbf{G}_{\mathbf{v}}(\bar{\mathbf{v}},\bar{\mathbf{w}}) + \mathbf{y}'\mathbf{G}_{\mathbf{w}}(\bar{\mathbf{v}},\bar{\mathbf{w}}) = \mathbf{0}$, implies $\mathbf{u}' = -\mathbf{y}'\mathbf{G}_{\mathbf{w}}\mathbf{G}_{\mathbf{v}}^{-1} = \mathbf{y}'\mathbf{h}'_{\mathbf{w}}$, and hence

$$(\mathbf{u}',\mathbf{y}') \left[\begin{array}{c|c} L_{\mathbf{vv}} & L_{\mathbf{vw}} \\ \hline L_{\mathbf{wv}} & L_{\mathbf{ww}} \end{array}\right]\left[\begin{array}{c} \mathbf{u} \\ \hline \mathbf{y} \end{array}\right] \equiv \mathbf{y}' f_{\mathbf{ww}}\mathbf{y} \geqq 0.$$

This proves that $L_{\mathbf{xx}}(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is positive semidefinite subject to the constraint $\mathbf{G}_{\mathbf{x}}(\bar{\mathbf{x}})$; see Section 2 of Appendix A.1. Q.E.D.

The next result shows that a sufficient condition for a LCP $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ to yield a local solution to CPP is that SND hold at $\bar{\mathbf{x}}$ and $L_{\mathbf{xx}}(\bar{\mathbf{x}},\bar{\mathbf{p}})$ be positive definite subject to the constraint $\mathbf{G}_{\mathbf{x}}(\bar{\mathbf{x}})$

Theorem 4. Suppose $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is a LCP, and suppose SND holds at $\bar{\mathbf{x}}$. If $\mathbf{z}'\mathbf{z} = 1$ and $\mathbf{z}'G_{\mathbf{x}}(\bar{\mathbf{x}}) = \mathbf{0}$ imply $\mathbf{z}'L_{\mathbf{xx}}(\bar{\mathbf{x}},\bar{\mathbf{p}})\mathbf{z} > 0$, then $\bar{\mathbf{x}}$ is a local solution of CPP.

Proof: We give a proof by contradiction. Suppose there exist

$$\underset{n\times 1}{\mathbf{z}_k}$$

and θ_k such that $\mathbf{z}'_k\mathbf{z}_k = 1$, $\theta_k \to 0$, $\mathbf{G}(\bar{\mathbf{x}} + \theta_k\mathbf{z}_k) + \mathbf{b}' = 0$, and $F(\bar{\mathbf{x}} + \theta_k\mathbf{z}_k) \leqq F(\bar{\mathbf{x}})$. Without loss of generality, we can assume $\mathbf{z}_k \to \mathbf{z}$. Consider $F(\bar{\mathbf{x}} + \theta\mathbf{z}_k)$ and $g^i(\bar{\mathbf{x}} + \theta\mathbf{z}_k)$ as functions of θ. Taylor's expansions of these functions around $\theta = 0$ yield

$$F(\bar{\mathbf{x}} + \theta_k\mathbf{z}_k) - F(\bar{\mathbf{x}}) = \theta_k\mathbf{z}'_kF_{\mathbf{x}}(\bar{\mathbf{x}}) + \frac{\theta_k^2}{2}\mathbf{z}'_kF_{\mathbf{xx}}(\bar{\mathbf{x}} + \theta_k^0\mathbf{z}_k)\mathbf{z}_k, \tag{0}$$

$$g^i(\bar{\mathbf{x}} + \theta_k\mathbf{z}_k) + b_i = \theta_k\mathbf{z}'_kg^i_{\mathbf{x}}(\bar{\mathbf{x}}) + \frac{\theta_k^2}{2}\mathbf{z}'_kg^i_{\mathbf{xx}}(\bar{\mathbf{x}} + \theta_k^i\mathbf{z}_k)\mathbf{z}_k, \tag{i}$$

where θ_k^0, θ_k^i are in the interval between 0 and θ_k, and we have used $g^i(\bar{\mathbf{x}})+b_i=0=L_{p_i}(\bar{\mathbf{x}},\bar{\mathbf{p}})$. Multiplying equation (i) above by $\bar{p}_i$ for each i and adding it to equation (0) yields, since $g^i(\bar{\mathbf{x}}+\theta_k\mathbf{z}_k)+b_i=0$,

$$0 \geqq F(\bar{\mathbf{x}}+\theta_k\mathbf{z}_k)-F(\bar{\mathbf{x}})$$
$$= \theta_k\mathbf{z}_k'L_{\mathbf{x}}(\bar{\mathbf{x}},\bar{\mathbf{p}})+\frac{\theta_k^2}{2}\mathbf{z}_k'\left[F_{\mathbf{xx}}(\bar{\mathbf{x}}+\theta_k^0\mathbf{z}_k)+\sum_{i=1}^{m}g_{\mathbf{xx}}^i(\bar{\mathbf{x}}+\theta_k^i\mathbf{z}_k)\bar{p}_i\right]\mathbf{z}_k,$$

or dividing by θ_k^2 and taking the limit $\mathbf{z}_k\to\mathbf{z}$, $\theta_k\to 0$,

$$0 \geqq \mathbf{z}'L_{\mathbf{xx}}(\bar{\mathbf{x}},\bar{\mathbf{p}})\mathbf{z}.$$

But dividing (i) by θ_k and taking the limit $\mathbf{z}_k\to\mathbf{z}$, $\theta_k\to 0$ yields, since $g^i(\bar{\mathbf{x}}+\theta_k\mathbf{z}_k)+b_i=0$, $\mathbf{z}'g_x^i(\bar{\mathbf{x}})=\mathbf{0}$. Then $\mathbf{z}'\mathbf{z}=1$, $\mathbf{z}'\mathbf{G}_{\mathbf{x}}(\bar{\mathbf{x}})=0$, and $\mathbf{z}'L_{\mathbf{xx}}(\bar{\mathbf{x}},\bar{\mathbf{p}})\mathbf{z}\leqq 0$ contradicts the hypothesis. Q.E.D.

Lemma 4 in Appendix A.1 provides the following reformulation:

Theorem 5. Suppose $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is a LCP, and suppose in the matrix

$$\left[\begin{array}{c|c} \underset{n\times m}{L_{\mathbf{xx}}(\bar{\mathbf{x}},\bar{\mathbf{p}})} & \underset{n\times m}{\mathbf{G}_{\mathbf{x}}(\bar{\mathbf{x}})} \\ \hline \underset{m\times n}{\mathbf{G}_{\mathbf{x}}(\bar{\mathbf{x}})'} & \underset{m\times m}{\mathbf{0}} \end{array}\right],$$

there exists a nested sequence of principal minors with the sign $(-1)^m$ formed by deleting r symmetric rows and columns from the first n, for $r=0,\ldots,n-m$. Then SND holds and $\bar{\mathbf{x}}$ is a local solution of CPP.

Proof: For $r=n-m$, the principal minor above is, except for sign, the squared determinant of a $m\times m$ submatrix of $\mathbf{G}_{\mathbf{x}}(\bar{\mathbf{x}})$, establishing that SND holds. Lemma 4 in Appendix A.1 establishes the result. Q.E.D.

A series of examples demonstrate the role of the assumptions in Theorems 1–4.

Example 1. Min $x_1+x_2^2$ subject to $x_1^2=0$. The minimum is at $\bar{x}_1=\bar{x}_2=0$ where $F_{\mathbf{x}}(\bar{\mathbf{x}})'=(1,0)$ and $g_{\mathbf{x}}^1(\bar{\mathbf{x}})'=(0,0)$. Then rank $g_{\mathbf{x}}^1(\bar{\mathbf{x}})<$ rank $(g_{\mathbf{x}}^1(\bar{\mathbf{x}})\ F_{\mathbf{x}}(\bar{\mathbf{x}}))$ and WND fails, so the Lagrangian method cannot be applied.

Example 2. Min $-x_1^2+x_2^2$ subject to $x_1^2=0$. The minimum is at $\bar{x}_1=\bar{x}_2=0$ where

$$F_x(\bar{\mathbf{x}})' = (0,0) = g_x^1(\bar{\mathbf{x}})',$$

and WND holds. Then $(\bar{\mathbf{x}},\bar{p})$ is a LCP for any $\bar{p}$, and

$$L_{xx}(\bar{\mathbf{x}},\bar{\mathbf{p}}) = \begin{bmatrix} 2(\bar{p}-1) & 0 \\ 0 & 2 \end{bmatrix}.$$

For $\bar{p} < 1$, $L_{xx}(\bar{\mathbf{x}},\bar{p})$ is not positive semidefinite subject to $g_x^1(\bar{\mathbf{x}})$. Hence, WND cannot replace SND in Theorem 3.

Example 3. Min $-x_1^2+x_2^2$ subject to $-x_1^2+x_2^2=0$. A minimum is at $\bar{x}_1=\bar{x}_2=1$, where $g_x^1(\bar{\mathbf{x}})' = (-2,2)$ and SND is satisfied. Then $(\bar{\mathbf{x}},\bar{p})$ with $\bar{p}=-1$ is a LCP and $L_{xx}(\bar{x},\bar{p})$ is the zero matrix. Then L_{xx} is positive semidefinite, but not positive definite, subject to constraint.

Example 4. Min $x_1^2+2x_1x_2+x_2^2+x_3^2$ subject to $x_1-x_2=0$. The minimum occurs at $\bar{\mathbf{x}}=\mathbf{0}$, where

$$g_x(\bar{\mathbf{x}})' = (1,-1,0),$$

$$F_x(\bar{\mathbf{x}})' = (0,0,0),$$

and

$$F_{xx}(\bar{\mathbf{x}}) = L_{xx}(\bar{\mathbf{x}}) = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Then SND holds and $\bar{p}=0$. $L_{xx}(\bar{\mathbf{x}})$ is positive semidefinite, and subject to the constraint $g_x(\bar{\mathbf{x}})'\mathbf{z}=0=z_1=z_2$ is positive definite.

A final comment regards the maximization problem

$$\text{Max } F(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{G}(\mathbf{x})+\mathbf{b}'=\mathbf{0}. \tag{CPP2}$$

This is equivalent to the minimization problem

$$\text{Min } [-F(\mathbf{x})] \quad \text{s.t.} \quad -\mathbf{G}(\mathbf{x})-\mathbf{b}'=\mathbf{0}, \tag{*}$$

to which Theorems 1 and 5 apply. Hence, defining $L^*(\mathbf{x},\mathbf{p}) = F(\mathbf{x})+[\mathbf{G}(\mathbf{x})+\mathbf{b}']\mathbf{p}$, we have the following result:

Theorem 6. Suppose $\bar{\mathbf{x}}$ is a local solution of CPP2 and SND holds at $\bar{\mathbf{x}}$. Then there exists $\bar{\mathbf{p}}$ such that $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is a LCP. Alternately, suppose $(\bar{\mathbf{x}},\bar{\mathbf{p}})$ is a LCP, and suppose in the matrix

$$\left[\begin{array}{c|c} \underset{n\times n}{L^*_{xx}(\bar{\mathbf{x}},\bar{\mathbf{p}})} & \underset{n\times m}{\mathbf{G}_x(\bar{\mathbf{x}})} \\ \hline \underset{m\times n}{\mathbf{G}_x(\bar{\mathbf{x}})'} & \underset{m\times m}{\mathbf{0}} \end{array}\right],$$

there exists a nested sequence of principal minors with the sign $(-1)^{n-r}$ formed by deleting r symmetric rows and columns from the first n, for $r = 0,\ldots,n-m$. Then SND holds and $\bar{\mathbf{x}}$ is a local solution of CPP2.

Proof: The proof of Theorem 1 applies without modification to the first part of this theorem. To prove the second part, we apply Theorem 5 to the minimization problem (*), obtaining the sufficient condition for $\bar{\mathbf{x}}$ to be a local solution that

$$\left|\begin{array}{c|c} -L^*_{\mathbf{xx}}(\bar{\mathbf{x}},\bar{\mathbf{p}}) & -\mathbf{G_x}(\bar{\mathbf{x}}) \\ \hline -\mathbf{G_x}(\bar{\mathbf{x}})' & \mathbf{0} \end{array}\right|,$$

have a nested sequence of principal minors of sign $(-1)^m$. Reversing the sign of all rows in the principal minor formed by deleting r rows and columns multiplies the determinant by the factor $(-1)^{n+m-r}$; it is then required to have the sign $(-1)^{n-r+2m} = (-1)^{n-r}$. But these are just the principal minors considered in the statement of the theorem, giving the desired result. Q.E.D.

Appendix A.3

CONVEX ANALYSIS

DANIEL McFADDEN

University of California, Berkeley

1. Introduction

This appendix lists mathematical properties of convex sets and convex conjugate functions used in developing the theory of cost and profit functions. Familiarity with the basic concepts of analysis (e.g., open and closed sets in Euclidean space, interior and closure of sets, compactness, continuity of real-valued functions) is assumed at the level of Rosenlicht (1968) or Bartle (1964). Brief introductions to the theory of convex sets can be found in Karlin (1959) or Mangasarian (1969). More advanced surveys are in Busemann (1958), Fan (1959), Fenchel (1953), Grunbaum (1967), Klee (1963, 1969), and Valentine (1964). The definitive reference work on the topic is Rockafellar (1970).

2. Notation

The results in this appendix will be stated for sets and functions in an N-dimensional Euclidean space $\mathbf{E}^N$. Subsets of $\mathbf{E}^N$ are denoted by boldface Roman caps (e.g., **W**, **X**, **Y**, **Z**), and points in $\mathbf{E}^N$ by lower case, boldface Roman letters (e.g., **w**, **x**, **y**, **z**). Real numbers are denoted by lower case Greek letters (e.g., α, β, σ, θ). Real-valued functions on $\mathbf{E}^N$ are denoted by Roman caps (e.g., F, G, H). The interior and closure of a set $\mathbf{Y}$ are denoted by $\operatorname{int}\mathbf{Y}$ and $\bar{\mathbf{Y}}$, respectively. The algebraic sum of non-empty sets $\mathbf{Y}$, $\mathbf{Z}$ is defined by $\mathbf{Y}+\mathbf{Z}=\{\mathbf{y}+\mathbf{z}\,|\,\mathbf{y}\in\mathbf{Y},\mathbf{z}\in\mathbf{Z}\}$. The set of points in $\mathbf{Y}$, but not in $\mathbf{Z}$, is denoted by $\mathbf{Y}\setminus\mathbf{Z}$.

3. Hyperplanes

A real-valued linear function P on $\mathbf{E}^N$ can be represented by a vector $\mathbf{p} \in \mathbf{E}^N$, with the value of P at $\mathbf{u} \in \mathbf{E}^N$ given by the inner product $P(\mathbf{u}) = \mathbf{p}\cdot\mathbf{u}$. A *hyperplane* is a set $\mathbf{H}(\mathbf{p},\alpha) = \{\mathbf{v} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{v} = \alpha\}$, where $\mathbf{p} \in \mathbf{E}^N$ is non-zero. Note that a hyperplane is a level set of a non-identically zero real-valued linear functional. The sets $\mathbf{H}^-(\mathbf{p},\alpha) = \{\mathbf{v} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{v} \leqq \alpha\}$ and $\mathbf{H}^+(\mathbf{p},\alpha) = \{\mathbf{v} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{v} \geqq \alpha\}$ are termed the *closed half-spaces* determined by the hyperplane $\mathbf{H}(\mathbf{p},\alpha)$, and $\mathbf{p}$ is termed the *normal* to $\mathbf{H}(\mathbf{p},\alpha)$. A hyperplane $\mathbf{H}(\mathbf{p},\alpha)$ is a *barrier* to a non-empty set $\mathbf{Y}$ if $\mathbf{Y}$ is contained in a closed half-space determined by $\mathbf{H}(\mathbf{p},\alpha)$, and $\mathbf{H}(\mathbf{p},\alpha)$ *supports* $\mathbf{Y}$ if it is a barrier to $\mathbf{Y}$ and intersects $\bar{\mathbf{Y}}$.

4. Convex Sets

A set $\mathbf{Y}$ is *convex* if $\mathbf{u},\mathbf{v} \in \mathbf{Y}, 0 < \theta < 1$ implies $\theta\mathbf{u} + (1-\theta)\mathbf{v} \in \mathbf{Y}$. The *(closed) convex hull* of a set $\mathbf{X}$ is the intersection of all the (closed) convex sets containing $\mathbf{X}$. The convex hull of $\mathbf{X}$ is denoted by $[\mathbf{X}]$. The closed convex hull of $\mathbf{X}$ equals the closure of the convex hull of $\mathbf{X}$, and is denoted by $\overline{[\mathbf{X}]}$.

5. Affine Subspaces

A set $\mathbf{F}$ is a *flat* (or affine subspace) if $\mathbf{u},\mathbf{v} \in \mathbf{F}$ implies $\theta\mathbf{u} + (1-\theta)\mathbf{v} \in \mathbf{F}$ for all real θ. Note that $\mathbf{E}^N$ itself is a flat; that points, lines, and hyperplanes in $\mathbf{E}^N$ are flats; that an arbitrary intersection of flats is a flat; that all flats are closed and convex; and that a non-empty flat is a translation of a linear subspace of $\mathbf{E}^N$. The *affine hull* of a set $\mathbf{X}$ is the intersection of all flats containing $\mathbf{X}$, and is denoted by aff $\mathbf{X}$. The *relative interior* of $\mathbf{X}$, denoted by intr $\mathbf{X}$, is the interior of $\mathbf{X}$ in the relative topology of $\mathbf{X}$ as a subset of aff $\mathbf{X}$; i.e., the set of points in $\mathbf{X}$ which are not in the closure of (aff $\mathbf{X})\setminus\mathbf{X}$.

6. Separation of Sets

Non-empty sets $\mathbf{Y},\mathbf{Z}$ have the *separation property* if there exists a hyperplane $\mathbf{H}(\mathbf{p},\alpha)$ such that $\mathbf{Y} \subseteq \mathbf{H}^-(\mathbf{p},\alpha)$ and $\mathbf{Z} \subseteq \mathbf{H}^+(\mathbf{p},\alpha)$. They have the *strong separation property* if there exist parallel hyperplanes $\mathbf{H}(\mathbf{p},\alpha)$ and $\mathbf{H}(\mathbf{p},\beta)$ such that $\alpha < \beta$, $\mathbf{Y} \subseteq \mathbf{H}^-(\mathbf{p},\alpha)$, $\mathbf{Z} \subseteq \mathbf{H}^+(\mathbf{p},\beta)$.

7. Cones

A set $\mathbf{K}$ is a *cone* with vertex at the origin if $\mathbf{v} \in \mathbf{K}$ and $\theta > 0$ imply $\theta\mathbf{v} \in \mathbf{K}$. $\mathbf{K}$ is a cone with vertex at $\mathbf{v}$ if $\mathbf{K} - \{\mathbf{v}\}$ is a cone with vertex at the origin. Hereafter, all cones are defined with vertex at the origin unless specified otherwise. The *(closed) cone with vertex* v *spanned by* a set **Y** is defined as the intersection of all (closed) cones with vertex v which contain **Y**, and denoted by $\mathbf{K}_v\mathbf{Y}$ (for the closed cone, $\bar{\mathbf{K}}_v\mathbf{Y}$). If v is the origin, the subscript is omitted. The *asymptotic cone* (recession cone) of a set **Y**, denoted by **AY**, is defined as follows: $\mathbf{v} \in \mathbf{AY}$ if and only if there exist a sequence $\mathbf{v}^k \in \mathbf{Y}$ and a sequence of positive real numbers θ_k such that $\theta_k \to 0$ and $\theta_k \mathbf{v}^k \to \mathbf{v}$. A cone **K** is *pointed* if $\mathbf{v}^i \in \bar{\mathbf{K}}$, $i = 1,\dots,m$, for finite m and $\sum_{i=1}^{m} \mathbf{v}^i = \mathbf{0}$ implies $\mathbf{v}^i = \mathbf{0}$, $i = 1,\dots,m$. A set **Y** is *semi-bounded* if **AY** is pointed.

8. Polar and Normal Cones

The *polar cone* of a set **Y** is the set $\mathbf{PY} = \{\mathbf{p} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{v} \leqq 0$ for all $\mathbf{v} \in \mathbf{Y}\}$. The *normal cone at* v of a set **Y** is the set of all $\mathbf{p} \in \mathbf{E}^N, \mathbf{p} \neq \mathbf{0}$, such that $\mathbf{Y} \subseteq \mathbf{H}^-(\mathbf{p},\mathbf{p}\cdot\mathbf{v})$, and is denoted by $\mathbf{N}(\mathbf{Y},\mathbf{v})$. The *normal cone* of **Y** is the union of $\mathbf{N}(\mathbf{y},\mathbf{v})$ for all $\mathbf{v} \in \mathbf{E}^N$, and is denoted by **NY**. Clearly, $\mathbf{p} \in \mathbf{NY}$ if and only if $\sup\{\mathbf{p}\cdot\mathbf{y} | \mathbf{y} \in \mathbf{Y}\} < +\infty$.

9. Convex Functions

A non-empty set **X** and real-valued function F with domain **X** are denoted by $\langle F,\mathbf{X}\rangle$. We say $\langle F,\mathbf{X}\rangle$ is *convex* if **X** is convex, and $\mathbf{u}, \mathbf{v} \in \mathbf{X}$, $0 < \theta < 1$ implies $F(\theta\mathbf{u} + (1-\theta)\mathbf{v}) \leqq \theta F(\mathbf{u}) + (1-\theta)F(\mathbf{v})$. We say $\langle F,\mathbf{X}\rangle$ is *positively linear homogeneous* if **X** is a cone and $\mathbf{v} \in \mathbf{X}$, $\theta > 0$ implies $F(\theta\mathbf{v}) = \theta F(\mathbf{v})$. We say $\langle F,\mathbf{X}\rangle$ is *closed* if the following conditions hold: (1) $\mathbf{v} \in \bar{\mathbf{X}} \setminus \mathbf{X}$ implies $\lim_{\mathbf{u}\in\mathbf{X},\mathbf{u}\to\mathbf{v}} \inf F(\mathbf{u}) = +\infty$; and (2) $\mathbf{v} \in \mathbf{X}$ implies $\lim_{\mathbf{u}\in\mathbf{X},\mathbf{u}\to\mathbf{v}} \inf F(\mathbf{u}) = F(\mathbf{v})$.[1] The *support function* $\langle G^{\mathbf{Y}},\mathbf{NY}\rangle$ of a non-empty set **Y** is a real-valued function defined by $G^{\mathbf{Y}}(\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{y} | \mathbf{y} \in \mathbf{Y}\}$ for $\mathbf{p} \in \mathbf{NY}$.

10. Properties of Convex Sets

A list of well-known properties of non-empty sets **Y**,**Z** follows. For completeness, references to proofs, or abbreviated or shortened proofs, are given for the non-trivial results.

10.1 $\mathbf{Y} \subseteq [\mathbf{Y}] \subseteq [\bar{\mathbf{Y}}] \subseteq \overline{[\mathbf{Y}]}$.

[1]The notation $\lim_{\mathbf{u}\in\mathbf{X},\mathbf{u}\to\mathbf{v}} \inf F(\mathbf{u}) = F(\mathbf{v})$ is defined as follows: Given $\epsilon > 0$, there exists an open ball **N** about v with radius at most ϵ such that

$$\epsilon > \Big|F(\mathbf{v}) - \inf_{\substack{\mathbf{u}\in\mathbf{N}\cap\mathbf{X}\\ \mathbf{u}\neq\mathbf{v}}} F(\mathbf{u})\Big|.$$

10.2 $\mathbf{Y}$ convex implies intr $\mathbf{Y} \neq \emptyset$ and the following sets convex: int $\mathbf{Y}$, intr $\mathbf{Y}$, $\bar{\mathbf{Y}}$, $\mathbf{AY}$, $\mathbf{K_vY}$ for any $\mathbf{v} \in \mathbf{E}^N$.
Proof: Fenchel (1953, Ch. II, Results 11, 15, 17, 22), Rockafellar (1970, Thm 6.2).

10.3 $\mathbf{Y}$ bounded implies $\overline{[\mathbf{Y}]}$ bounded and $\mathbf{AY} = \{\mathbf{0}\}$.

10.4 $\mathbf{Y}$ compact implies $[\mathbf{Y}]$ compact.
Proof: Grunbaum (1967, 2.3.5).

10.5 $\mathbf{Y},\mathbf{Z}$ convex implies $\mathbf{Y} \cap \mathbf{Z}$ and $\mathbf{Y} + \mathbf{Z}$ convex.

10.6 $\mathbf{Y}$ a convex cone implies $\mathbf{Y} = \mathbf{Y} + \mathbf{Y}$.

10.7 $\mathbf{0} \in \mathbf{Y},\mathbf{Z}$ implies $\mathbf{KY} \subseteq \mathbf{K}(\mathbf{Y}+\mathbf{Z}) \subseteq \mathbf{KY} + \mathbf{KZ}$.

10.8 $\mathbf{0} \in \mathbf{Y},\mathbf{Z}$ and $\mathbf{Y},\mathbf{Z}$ convex implies $\mathbf{K}(\mathbf{Y}+\mathbf{Z}) = \mathbf{KY} + \mathbf{KZ}$.
Proof: Use 10.6 and 10.7.

10.9 $\mathbf{Y}$ closed and convex implies $\mathbf{Y} = \mathbf{Y} + \mathbf{AY}$.
Proof: Winter (forthcoming), Rockafellar (1970, Thm. 8.3). If $\mathbf{v} \in \mathbf{Y} + \mathbf{AY}$, then there exist $\mathbf{u} \in \mathbf{Y}$, $\mathbf{w} \in \mathbf{AY}$ with $\mathbf{v} = \mathbf{u} + \mathbf{w}$. Then there exist $\mathbf{w}^k \in \mathbf{Y}$, $\theta_k \geqq 0$ with $\theta_k \mathbf{w}^k \to \mathbf{w}$ and $\theta_k \to 0$, implying $(1-\theta_k)\mathbf{u} + \theta_k \mathbf{w}^k \in \mathbf{Y}$ for k large, and $(1-\theta_k)\mathbf{u} + \theta_k \mathbf{w}^k \to \mathbf{v} \in \mathbf{Y}$.

10.10 Every $\mathbf{v} \in [\mathbf{Y}]$ is expressible in the form $\mathbf{v} = \sum_{i=0}^{N} \theta_i \mathbf{v}^i$ with $\theta_i \geqq 0$, $\sum_{i=0}^{N} \theta_i = 1$, and $\mathbf{v}^i \in \mathbf{Y}$.

10.11 Convex sets $\mathbf{Y},\mathbf{Z}$ have the separation property if and only if one or both of the following are true: (1) (intr $\mathbf{Y}$) $\cap$ (intr $\mathbf{Z}$) $= \phi$, or (2) $\mathbf{Y} \cup \mathbf{Z}$ lies in a hyperplane.
Proof: Klee (1969, Thm. 2.1), Grunbaum (1967, 2.2.2), Rockafellar (1970, Thm. 11.3).

10.12 For $\mathbf{Y}$ convex, $\mathbf{Y}$ and $\{\mathbf{z}\}$ have the separation property if and only if $\mathbf{z} \notin$ int $\mathbf{Y}$. There exists a hyperplane $H(\mathbf{p},\alpha)$ with $\mathbf{z} \in H(\mathbf{p},\alpha)$, $\mathbf{Y} \subseteq \mathbf{H}^-(\mathbf{p},\alpha)$, and (intr $\mathbf{Y}$) $\cap \mathbf{H}(\mathbf{p},\alpha) = \emptyset$ if and only if $\mathbf{z} \notin$ intr $\mathbf{Y}$.
Proof: Klee (1969, Thm. 1.1), Rockafellar (1970, Thm. 11.3).

10.13 $\mathbf{Y},\mathbf{Z}$ convex and disjoint, $\mathbf{Y}$ closed, $\mathbf{Z}$ compact implies $\mathbf{Y},\mathbf{Z}$ have the strong separation property.
Proof: Grunbaum (1967, 2.2.1), Rockafellar (1970, Corol. 11.4.2).

10.14 $\mathbf{Y}$ convex, $\mathbf{Z}$ a non-empty flat, $\mathbf{Z} \cap$ intr $\mathbf{Y} = \emptyset$ implies the existence of a hyperplane $\mathbf{H}(\mathbf{p},\alpha)$ such that $\mathbf{Y} \subseteq \mathbf{H}^-(\mathbf{p},\alpha)$, $\mathbf{Z} \subseteq \mathbf{H}(\mathbf{p},\alpha)$, and $\mathbf{H}(\mathbf{p},\alpha) \cap$ intr $\mathbf{Y} = \emptyset$.
Proof: Rockafellar (1970, Thm. 11.2).

10.15 $\mathbf{Y} \subseteq \mathbf{H}^-(\mathbf{p},\alpha)$ implies $\mathbf{p} \in \mathbf{NY}$.

10.16 int $\mathbf{P}(\mathbf{AY}) \subseteq \mathbf{NY} \subseteq \mathbf{P}(\mathbf{AY})$.
Proof: Rockafellar (1970, Corol. 14.2.1). If $\mathbf{p} \notin \mathbf{P}(\mathbf{AY})$, then $\mathbf{p}\cdot\mathbf{v} > 0$ for some $\mathbf{v} \in \mathbf{AY}$. There exist $\mathbf{v}^j \in \mathbf{Y}$, $\theta_j \geqq 0$ such that

$|\mathbf{v}^j| \to +\infty$ and $\theta_j \mathbf{v}^j \to \mathbf{v}$, implying $\mathbf{p}\cdot\mathbf{v}^j \to +\infty$, and hence $\mathbf{p} \notin \mathbf{NY}$. If $\mathbf{p} \notin \mathbf{NY}$, then there exist $\mathbf{v}^j \in \mathbf{Y}$ with $\mathbf{p}\cdot\mathbf{v}^j \to +\infty$. Then, $\mathbf{v}^j/|\mathbf{v}^j|$ has a subsequence converging to $\mathbf{v} \in \mathbf{AY}$ with $\mathbf{p}\cdot\mathbf{v} \geqq 0$. Then, $(\mathbf{p} + \epsilon\mathbf{v})\cdot\mathbf{v} > 0$ for all $\epsilon > 0$, implying $\mathbf{p} + \epsilon\mathbf{v} \notin \mathbf{P(AY)}$ and hence $\mathbf{p} \notin \operatorname{int} \mathbf{P(AY)}$.

10.17 $\mathbf{Y} \subseteq \mathbf{Z}$ implies $\mathbf{PZ} \subseteq \mathbf{PY}$.

10.18 $\mathbf{PY}$ is closed and convex, and $\mathbf{NY}$ is convex.

10.19 $\mathbf{Y} \subseteq \mathbf{P(PY)} = \overline{[\mathbf{KY}]}$.

10.20 $\mathbf{0} \in \mathbf{Y},\mathbf{Z}$ implies $\mathbf{P(Y + Z)} = (\mathbf{PY}) \cap (\mathbf{PZ})$.

10.21 $\mathbf{Y} \cap \mathbf{PY} \subseteq \{\mathbf{0}\}$.

10.22 $\mathbf{AY} = \overline{\mathbf{AY}} = \mathbf{A}\bar{\mathbf{Y}}$ and $NY = N\bar{Y}$.

11. Semi-Bounded Sets

The next series of results give properties of semibounded sets. Most of these properties can be obtained as consequences of theorems of Fenchel (1953), Grunbaum (1967), or Rockafellar (1970); however, we shall give direct proofs which are somewhat simpler.

11.1. *Lemma.* If $\mathbf{Y}$ is a closed pointed cone, then (1) there exists a positive scalar μ such that for any finite set of points $\mathbf{v}^i \in \mathbf{Y}$, $i = 1,\ldots,m$, it follows that $|\mathbf{v}^i| \leqq \mu|\sum_{i=1}^m \mathbf{v}^i|$; and (2) $[\mathbf{Y}]$ is a closed pointed cone.

Proof: Rockafellar (1970, Corol. 9.1.2). We first establish the existence of a μ depending in general on m for which (1) is valid. Suppose, for fixed m, no μ with the required property exists. Then, there exist $\mathbf{v}^{ij} \in \mathbf{Y}$ such that $\sum_{i=1}^m \mathbf{v}^{ij} = \mathbf{u}^j$ with $|\mathbf{u}^j| \leqq 1$ and $\lambda_j = 1/\max_i|\mathbf{v}^{ij}| \to 0$. Hence, there is a subsequence of j (retain notation) such that $\lambda_j \mathbf{v}^{ij} \to \mathbf{w}^i \in \mathbf{Y}$ for $i = 1,\ldots,m$ and $\sum_{i=1}^m \mathbf{w}^i = \mathbf{0}$. Since $1 \leqq \sum_{i=1}^m \lambda_j|\mathbf{v}^{ij}| \leqq m$, at least one $\mathbf{w}^i \neq \mathbf{0}$. This contradicts the hypothesis that $\mathbf{Y}$ is pointed.

We next prove that $[\mathbf{Y}]$ is closed. If $\mathbf{v}^i \in [\mathbf{Y}]$, $\mathbf{v}^i \to \mathbf{v}$, then there exist $\mathbf{v}^{ij} \in \mathbf{Y}$, $j = 0,\ldots,N$, such that $\mathbf{v}^i = \sum_{j=0}^N \mathbf{v}^{ij}$. By the result just proved, the $\mathbf{v}^{ij}$ are bounded, and hence there is a subsequence of i (retain notation) such that $\mathbf{v}^{ij} \to \mathbf{u}^j \in \mathbf{Y}$. Then, $\mathbf{v} = \sum_{j=0}^N \mathbf{u}^j \in [\mathbf{Y}]$, and $[\mathbf{Y}]$ is closed.

If $[\mathbf{Y}]$ were not pointed, then there would exist non-zero $\mathbf{v}^i \in [\mathbf{Y}]$ with $\sum_{i=0}^m \mathbf{v}^i = 0$ (since $[\mathbf{Y}]$ is closed). Since each $\mathbf{v}^i = \sum_{j=0}^N \mathbf{v}^{ij}$ for some $\mathbf{v}^{ij} \in \mathbf{Y}$, the implication $\sum_{i=0}^m \sum_{j=0}^N \mathbf{v}^{ij} = \mathbf{0}$ would be obtained, contradicting the hypothesis that $\mathbf{Y}$ is pointed. Hence, (2) is verified.

By (2), the condition (1) proved for fixed m can be applied to $[\mathbf{Y}]$, establishing μ such that $|\mathbf{v}^i| \leqq \mu|\sum_{i=1}^2 \mathbf{v}^i|$ for $\mathbf{v}^i \in [\mathbf{Y}]$. But the sum of any finite sequence $\mathbf{u}^j \in [\mathbf{Y}]$ can be written as $\sum_{j=1}^m \mathbf{u}^j = \mathbf{u}^k + \mathbf{w}$ with $\mathbf{w} =$

$\sum_{j=1,j\neq k}^{m}\mathbf{u}^j \in [\mathbf{Y}]$, implying $|\mathbf{u}^k| \leqq \mu|\sum_{j=1}^{m}\mathbf{u}^j|$ for $k = 1,\ldots,m$. This verifies (1). Q.E.D.

11.2. *Lemma.* If $\mathbf{Y}$ is non-empty and semi-bounded, then (1) given a positive scalar λ, there exists a positive scalar μ such that for any finite set of points $\mathbf{v}^i \in \mathbf{Y}$, and scalars $\theta_i \geqq 0$, $i = 1,\ldots,m$, with $|\sum_{i=1}^{m}\theta_i\mathbf{v}^i| \leqq \lambda$, it follows that $|\theta_i\mathbf{v}^i| \leqq \mu$; (2) $\mathbf{A}[\mathbf{Y}] = \overline{[\mathbf{AY}]}$; (3) $[\mathbf{Y}]$ is semi-bounded; and (4) if $\mathbf{Y}$ is closed, then $[\mathbf{Y}] + \mathbf{A}[\mathbf{Y}] = \overline{[\mathbf{Y}]}$.

Proof: The argument parallels that of 11.1. If (1) is violated for fixed m, there exist $\mathbf{v}^{ij} \in \mathbf{Y}$ and $\theta_{ij} \geqq 0$ such that $\sum_{i=1}^{m}\theta_{ij}\mathbf{v}^{ij} = \mathbf{u}^j$ with $|\mathbf{u}^j| \leqq \lambda$ and $\lambda_j = 1/\max_i|\theta_{ij}\mathbf{v}^{ij}| \to 0$, implying the existence of limit points $\mathbf{w}^i \in \mathbf{AY}$ of the $\lambda_j\theta_{ij}\mathbf{v}^{ij}$ such that $\sum_{i=1}^{m}\mathbf{w}^i = \mathbf{0}$ and $\sum_{i=1}^{m}|\mathbf{w}^i| \geqq 1$, contradicting the hypothesis that $\mathbf{Y}$ is semi-bounded.

We next show $\mathbf{A}[\mathbf{Y}] \subseteq \overline{[\mathbf{AY}]}$. If $\mathbf{v} \in \mathbf{A}[\mathbf{Y}]$, there exist $\mathbf{u}^j \in [\mathbf{Y}]$, $\theta_j \geqq 0$, $\mathbf{v}^{ij} \in \mathbf{Y}$, $\lambda_{ij} \geqq 0$ such that $\theta_j \to 0$, $\theta_j\mathbf{u}^j \to \mathbf{v}$, $\sum_{i=0}^{N}\lambda_{ij} = 1$, and $\mathbf{u}^j = \sum_{i=0}^{N}\lambda_{ij}\mathbf{v}^{ij}$. By 11.2(1) for fixed m, $\theta_j\lambda_{ij}\mathbf{v}^{ij}$ is bounded in j, and hence there is a subsequence of j (retain notation) such that $\theta_j\lambda_{ij}\mathbf{v}^{ij} \to \mathbf{v}^i \in \mathbf{AY}$. Then, $\mathbf{v} = \sum_{i=0}^{N}\mathbf{v}^i \in [\mathbf{AY}]$. By 10.2, $\mathbf{A}[\mathbf{Y}]$ is convex. Then, $\mathbf{AY} \subseteq \mathbf{A}[\mathbf{Y}]$ implies $[\mathbf{AY}] \subseteq \mathbf{A}[\mathbf{Y}]$. Hence, (2) is verified.

$\mathbf{Y}$ semi-bounded implies $\mathbf{AY}$ pointed, which implies $[\mathbf{AY}]$ pointed by 11.1(2), which implies in turn $\mathbf{A}[\mathbf{Y}]$ pointed by 11.2(1). Hence, $[\mathbf{Y}]$ is semi-bounded, proving (3). Applying 11.2(1) for fixed m to the set $[\mathbf{Y}]$, we obtain for given λ a scalar μ such that $\mathbf{v}^i \in [\mathbf{Y}]$, $\theta_i \geqq 0$, and $|\theta_1\mathbf{v}^1 + \theta_2\mathbf{v}^2| \leqq \lambda$ implies $|\theta_i\mathbf{v}^i| \leqq \mu$. For any finite set of $\mathbf{v}^i \in [\mathbf{Y}]$ and $\theta_i \geqq 0$, $i = 1,\ldots,m$ with $|\sum_{i=1}^{m}\theta_i\mathbf{v}^i| \leqq \lambda$, define $\alpha = \sum_{i=1,i\neq k}^{m}\theta_i$. Without loss, assume $\alpha > 0$ and define $\mathbf{u} = \alpha^{-1}\sum_{i=1,i\neq k}^{m}\theta_i\mathbf{v}^i \in [\mathbf{Y}]$. Then, $\sum_{i=1}^{m}\theta_i\mathbf{v}^i = \theta_k\mathbf{v}^k + \alpha\mathbf{u}$, implying $|\theta_k\mathbf{v}^k| \leqq \mu$. This verifies 11.2(1) for all m.

We next prove (4). By 10.2 and 10.5, $[\mathbf{Y}] + \mathbf{A}[\mathbf{Y}]$ is convex and contains $[\mathbf{Y}]$, and by 10.9 and 10.22, $[\mathbf{Y}] + \mathbf{A}[\mathbf{Y}] \subseteq \overline{[\mathbf{Y}]} + \mathbf{A}\overline{[\mathbf{Y}]} = \overline{[\mathbf{Y}]}$. Hence, it is sufficient to show $[\mathbf{Y}] + \mathbf{A}[\mathbf{Y}]$ closed. Since $\overline{[\mathbf{Y}]}$ is semi-bounded by (3), $[\mathbf{Y}] + \mathbf{A}[\mathbf{Y}]$ is semi-bounded. Suppose $\mathbf{v}^j \in [\mathbf{Y}] + \mathbf{A}[\mathbf{Y}]$, $\mathbf{v}^j \to \mathbf{v}$. Then, there exist $\mathbf{u}^j \in [\mathbf{Y}]$, $\mathbf{w}^j \in \mathbf{A}[\mathbf{Y}]$, $\mathbf{u}^{ij} \in \mathbf{Y}$, $\theta_{ij} \geqq 0$ such that $\mathbf{v}^j = \mathbf{u}^j + \mathbf{w}^j$, $\sum_{i=0}^{N}\theta_{ij} = 1$, and $\mathbf{u}^j = \sum_{i=0}^{N}\theta_{ij}\,\mathbf{u}^{ij}$. By (1), $\mathbf{w}^j$ and $\theta_{ij}\mathbf{u}^{ij}$ are bounded. Hence there exists a subsequence of j (retain notation) such that $\mathbf{w}^j \to \mathbf{w} \in \mathbf{A}[\mathbf{Y}]$, $\theta_{ij}\mathbf{u}^{ij} \to \mathbf{x}^i$, and $\theta_{ij} \to \theta_i$ with $\sum_{i=0}^{N}\theta_i = 1$. Let $\mathbf{I}$ denote the set of i indices with $\theta_i > 0$, and $\mathbf{J}$ denote the set of remaining indices, and let $y^i = \mathbf{x}^i/\theta_i$ for $i \in \mathbf{I}$. Then, $\mathbf{y}^i \in \mathbf{Y}$, $i \in \mathbf{I}$, and $\mathbf{y} = \sum_I\theta_i\mathbf{y}^i \in [\mathbf{Y}]$. For $i \in \mathbf{J}$, $\mathbf{x}^i \in \mathbf{AY}$. Hence, $\mathbf{u} = \mathbf{w} + \sum_J\mathbf{x}^i \in \mathbf{A}[\mathbf{Y}]$, implying $\mathbf{v} = \mathbf{y} + \mathbf{u} \in [\mathbf{Y}] + \mathbf{A}[\mathbf{Y}]$. This verifies (4). Q.E.D.

11.3. *Lemma.* If $\mathbf{Y}$ is a closed cone, then the following conditions are equivalent: (1) $\mathbf{Y}$ is pointed; (2) $[\mathbf{Y}] \cap [-\mathbf{Y}] = \{\mathbf{0}\}$; (3) int $\mathbf{PY} \neq \emptyset$; and $\mathbf{p} \in \text{int}\, \mathbf{PY}$, $\mathbf{v} \in \mathbf{Y}$, $\mathbf{v} \neq \mathbf{0}$ implies $\mathbf{p}\cdot\mathbf{v} < 0$; (4) there exists $\mathbf{p} \in \mathbf{E}^N$ such that $\mathbf{v} \in \mathbf{Y}$, $\mathbf{v} \neq \mathbf{0}$, implies $\mathbf{p}\cdot\mathbf{v} < 0$.

Proof: (4)→(3). If $\mathbf{v} \in \mathbf{Y}$, $|\mathbf{v}| = 1$, then $\mathbf{p}\cdot\mathbf{v} \leqq -\alpha < 0$. For $\mathbf{q} \in \mathbf{E}^N$, $|\mathbf{q}| \leqq \alpha/2$, $(\mathbf{p}+\mathbf{q})\cdot\mathbf{v} \leqq -\alpha + |\mathbf{q}\cdot\mathbf{v}| \leqq -\alpha/2$, implying $\mathbf{p}+\mathbf{q} \in \mathbf{PY}$, and $\mathbf{p} \in \text{int}\, \mathbf{PY}$. The second part of (3) is a trivial consequence of the first part.

(3)→(2). $\mathbf{v} \in [\mathbf{Y}] \cap [-\mathbf{Y}]$ implies $\mathbf{q}\cdot\mathbf{v} = 0$ for all $\mathbf{q} \in \mathbf{PY}$. If $\mathbf{p} \in \text{int}\, \mathbf{PY}$, then $\mathbf{p} + \alpha\mathbf{v} \in \mathbf{PY}$ for α small positive, implying $\mathbf{v}\cdot\mathbf{v} = 0$, or $\mathbf{v} = \mathbf{0}$.

(2)→(1). If $\mathbf{Y}$ is not pointed there exist $\mathbf{v}^i \in \mathbf{Y}$ such that $\mathbf{v}^0 \neq \mathbf{0}$ and $\sum_{i=0}^m \mathbf{v}^i = \mathbf{0}$, implying $-\mathbf{v}^0 = \sum_{i=1}^m \mathbf{v}^i \in [\mathbf{Y}]$ and contradicting (2).

(1)→(4). Define the set $\mathbf{Z} = \{\mathbf{v} \in \mathbf{Y} \mid |\mathbf{v}| \geqq 1\}$. Then, $\mathbf{Z}$ is closed and $\mathbf{AZ} = \mathbf{Y}$, implying $\mathbf{Z}$ semi-bounded. By 11.2 (4) and 10.9, $[\mathbf{Z}] + \mathbf{A}[\mathbf{Z}] = [\mathbf{Z}]$ is closed. If $\mathbf{0} \in [\mathbf{Z}]$, then there exist $\mathbf{v}^i \in \mathbf{Z}$, $\theta_i \geqq 0$ such that $\sum_{i=0}^N \theta_i = 1$ and $\sum_{i=0}^N \theta_i \mathbf{v}^i = \mathbf{0}$, contradicting (1). Hence, $\mathbf{0} \notin [\mathbf{Z}]$ and by 10.13 there exists $\mathbf{p}$ and $\alpha > 0$ such that $\mathbf{p}\cdot\mathbf{v} \leqq -\alpha$ for all $\mathbf{v} \in [\mathbf{Z}]$. Then, $\mathbf{p}$ satisfies (4). Q.E.D.

11.4. *Lemma.* For $\mathbf{Y},\mathbf{Z}$ non-empty, the following conditions hold: (1) $\mathbf{AY} \subseteq \mathbf{A}(\mathbf{Y} \cup \mathbf{Z}) = (\mathbf{AY}) \cup (\mathbf{AZ}) \subseteq \mathbf{A}(\mathbf{Y} + \mathbf{Z})$; (2) $\mathbf{Y},\mathbf{Z}$ convex implies $\mathbf{AY} + \mathbf{AZ} \subseteq \mathbf{A}(\mathbf{Y} + \mathbf{Z})$; (3) $\mathbf{Y} \cup \mathbf{Z}$ semi-bounded implies $\mathbf{A}(\mathbf{Y} + \mathbf{Z}) \subseteq \mathbf{AY} + \mathbf{AZ} \subseteq \mathbf{A}[\mathbf{Y} \cup \mathbf{Z}]$.

Proof: (1) $\mathbf{v} \in \mathbf{AY}$ implies there exist $\mathbf{v}^i \in \mathbf{Y}$, $\theta_i \geqq 0$ such that $\theta_i \to 0$, $\theta_i \mathbf{v}^i \to \mathbf{v}$. Take any $\mathbf{w} \in \mathbf{Z}$. Then $\mathbf{v}^i + \mathbf{w} \in \mathbf{Y} + \mathbf{Z}$, and $\theta_i(\mathbf{v}^i + \mathbf{w}) \to \mathbf{v} \in \mathbf{A}(\mathbf{Y} + \mathbf{Z})$. The remaining conditions are immediate.

(2) Using (1) and 10.6, $\mathbf{AY} + \mathbf{AZ} \subseteq \mathbf{A}(\mathbf{Y}+\mathbf{Z}) + \mathbf{A}(\mathbf{Y}+\mathbf{Z}) = \mathbf{A}(\mathbf{Y}+\mathbf{Z})$.

(3) If $\mathbf{v} \in \mathbf{A}(\mathbf{Y}+\mathbf{Z})$, then there exist $\mathbf{u}^j \in \mathbf{Y}$, $\mathbf{w}^j \in \mathbf{Z}$, $\theta_j \geqq 0$ such that $\theta_j \to 0$ and $\theta_j(\mathbf{u}^j + \mathbf{w}^j) \to \mathbf{v}$. By 11.2(1), $\theta_j \mathbf{u}^j$ and $\theta_j \mathbf{w}^j$ are bounded, and there exists a subsequence (retain notation) with $\theta_j \mathbf{u}^j \to \mathbf{u} \in \mathbf{AY}$ and $\theta_j \mathbf{w}^j \to \mathbf{w} \in \mathbf{AZ}$, implying $\mathbf{v} = \mathbf{u} + \mathbf{w} \in \mathbf{AY} + \mathbf{AZ}$. Finally, $\mathbf{v} = \mathbf{u} + \mathbf{w}$, $\mathbf{u} \in \mathbf{AY}$, $\mathbf{w} \in \mathbf{AZ}$ imply $\mathbf{v} \in \mathbf{A}[\mathbf{Y} \cup \mathbf{Z}]$ immediately if $\mathbf{u} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$, and if $\mathbf{u}$, $\mathbf{w} \neq \mathbf{0}$, imply the existence of $\mathbf{u}^j \in \mathbf{Y}$, $\mathbf{w}^j \in \mathbf{Z}$, $\theta_j, \lambda_j \geqq 0$ such that $\theta_j \mathbf{u}^j \to \mathbf{u}$, $\lambda_j \mathbf{w}^j \to \mathbf{w}$, $\theta_j \to 0$, $\lambda_j \to 0$. Then, $\mathbf{v}^j = (\theta_j + \lambda_j)^{-1}(\theta_j \mathbf{u}^j + \lambda_j \mathbf{w}^j) \in [\mathbf{Y} \cup \mathbf{Z}]$ for j large, and $(\theta_j + \lambda_j)\mathbf{v}^j \to \mathbf{v} \in \mathbf{A}[\mathbf{Y} \cup \mathbf{Z}]$. Q.E.D.

11.5. *Lemma.* $\mathbf{Y},\mathbf{Z}$ non-empty and closed, and $\mathbf{Y} \cup \mathbf{Z}$ semi-bounded implies $\mathbf{Y} + \mathbf{Z}$ closed and semi-bounded.

Proof: $\mathbf{Y} + \mathbf{Z}$ is semi-bounded by 11.4(3). If $\mathbf{v}^i \in \mathbf{Y} + \mathbf{Z}$, $\mathbf{v}^i \to \mathbf{v}$, then there exist $\mathbf{u}^i \in \mathbf{Y}$, $\mathbf{w}^i \in \mathbf{Z}$ such that $\mathbf{v}^i = \mathbf{u}^i + \mathbf{w}^i$. By 11.2(1), $\mathbf{u}^i, \mathbf{w}^i$ are bounded, and there exists a subsequence (retain notation) with $\mathbf{u}^i \to \mathbf{u} \in \mathbf{Y}$, $\mathbf{w}^i \to \mathbf{w} \in \mathbf{Z}$, implying $\mathbf{v} = \mathbf{u} + \mathbf{w} \in \mathbf{Y} + \mathbf{Z}$. Q.E.D.

11.6. *Lemma.* If **Y** is non-empty, convex, and semi-bounded, then $\emptyset \neq \text{int}\,\mathbf{AY}$ if and only if **NY** is pointed.

Proof: Since $\mathbf{AY} = \mathbf{P}(\mathbf{N\bar{Y}})$, 11.3 implies the result. Q.E.D.

12. Properties of Convex Functions

The next series of results give properties of convex functions, particularly support functions. General treatments of this topic can be found in Fenchel (1953, Ch. III), Karlin (1959, 7.5), and Rockafellar (1970, Sects. 10, 13, 25).

12.1. *Lemma.* If $\langle F,\mathbf{X}\rangle$ is convex, then (1) F is continuous on $\text{intr}\,\mathbf{X}$, and is uniformly Lipschitzian on any compact subset **Y** of $\text{intr}\,\mathbf{X}$ (i.e., given **Y**, there exists μ such that $|F(\mathbf{x}) - F(\mathbf{y})| \leqq \mu|\mathbf{x}-\mathbf{y}|$ for all $\mathbf{x},\mathbf{y} \in \mathbf{Y}$.

(2) If $\text{int}\,\mathbf{X} \neq \emptyset$, then F possesses a first and second differential in a set $\mathbf{Y} \subseteq \text{int}\,\mathbf{X}$, with $(\text{int}\,\mathbf{X})\setminus\mathbf{Y}$ a set of Lebesgue measure zero. The vector of first-order partial derivatives of F, denoted by $\mathbf{F}'$ and termed the *gradient*, is continuous in **Y**. At each point in **Y**, the matrix of second-order partial derivatives of F, denoted by $\mathbf{F}''$ and termed the *Hessian*, is symmetric (i.e., derivatives are independent of the order of differentiation) and the quadratic form $Q(\mathbf{v},\mathbf{F}'')$ of any vector **v** and the matrix $\mathbf{F}''$ is non-negative (i.e., $\mathbf{F}''$ is a non-negative definite matrix). Further, $F(\mathbf{z}) = F(\mathbf{x}) + \mathbf{F}'(\mathbf{x})(\mathbf{z}-\mathbf{x}) + \frac{1}{2}Q(\mathbf{z}-\mathbf{x},\mathbf{F}''(\mathbf{x})) + i(|\mathbf{z}-\mathbf{x}|^2)$ for $\mathbf{x} \in \mathbf{Y}$, $\mathbf{z} \in \text{int}\,\mathbf{X}$, where $i(\alpha)$ is a term satisfying $\lim_{\alpha\to 0^+} i(\alpha)/\alpha = 0$.

Proof: For (1) see Fenchel (1953, Ch. III, Results 21, 23, 34), Popoviciu (1945), or Rockafellar (1970, Thms. 10.1 and 10.4). For (2) see Reidemeister (1921), Alexandrov (1939), Rockafellar (1970, Thm. 25.5), and Busemann and Feller (1935–36). Q.E.D.

12.2. *Definition.* If $\langle F,\mathbf{X}\rangle$ is convex and closed, define $\langle H,\mathbf{Y}\rangle$ by $\mathbf{y} \in \mathbf{Y}$ if and only if $\sup_{\mathbf{x}\in\mathbf{X}}\{\mathbf{y}\cdot\mathbf{x} - F(\mathbf{x})\} < +\infty$, and $H(\mathbf{y}) = \sup_{\mathbf{x}\in\mathbf{X}}\{\mathbf{y}\cdot\mathbf{x} - F(\mathbf{x})\}$ for $\mathbf{y} \in \mathbf{Y}$. $\langle H,\mathbf{Y}\rangle$ is termed the *conjugate dual* of $\langle F,\mathbf{X}\rangle$, and will be denoted by $\langle H,\mathbf{Y}\rangle = D\langle F,\mathbf{X}\rangle$. The following theorem is due to Fenchel, and is proved in Karlin (1959, 7.5.2 and 7.5.3), or Rockafellar (1970, Thm. 12.2).

12.3. *Theorem.* If $\langle F,\mathbf{X}\rangle$ is convex and closed, then the conjugate dual $\langle H,\mathbf{Y}\rangle = D\langle F,\mathbf{X}\rangle$ has $\mathbf{Y} \neq \emptyset$ and is convex and closed. Further, $D\langle H,\mathbf{Y}\rangle = \langle F,\mathbf{X}\rangle$.

12.4. *Lemma.* If $\mathbf{Y}$ is non-empty, closed, and semi-bounded, then the support function $\langle G^{\mathbf{Y}},\mathbf{NY}\rangle$ is convex, closed, and positively linear homogeneous, and $\overline{[\mathbf{Y}]} = \{\mathbf{y} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{y} \leqq G^{\mathbf{Y}}(\mathbf{p}) \text{ for all } \mathbf{p} \in \mathbf{NY}\}$.

Proof: Rockafellar (1970, Thm. 13.2). Define $\langle F,\mathbf{X}\rangle$ with $\mathbf{X} = \overline{[\mathbf{Y}]}$ and $F(\mathbf{x}) = \mathbf{0}$ for $\mathbf{x} \in \mathbf{X}$. Then, $\langle G,\mathbf{V}\rangle = D\langle F,\mathbf{X}\rangle$ is defined for $\mathbf{p} \in \mathbf{V}$ if and only if $\sup_{\mathbf{x}\in\mathbf{X}} \mathbf{p}\cdot\mathbf{x} = \sup_{\mathbf{y}\in\mathbf{Y}} \mathbf{p}\cdot\mathbf{y} < +\infty$, and $G(\mathbf{p}) = \sup_{\mathbf{y}\in\mathbf{Y}} \mathbf{p}\cdot\mathbf{y}$ for $\mathbf{p} \in \mathbf{V}$. Hence, $\mathbf{V} = \mathbf{NY}$ and $G = G^{\mathbf{Y}}$. By 12.3, $\langle G^{\mathbf{Y}},\mathbf{NY}\rangle$ is convex and closed. Since $\mathbf{NY}$ is a cone, the positive linear homogeneity of $\langle G^{\mathbf{Y}},\mathbf{NY}\rangle$ is a consequence of the definition of $G^{\mathbf{Y}}$. Finally, by 12.3, $D\langle G^{\mathbf{Y}},\mathbf{NY}\rangle = \langle F,\mathbf{X}\rangle$ and $\mathbf{x} \in \mathbf{X}$ if and only if $\sup_{\mathbf{p}\in\mathbf{NY}}\{\mathbf{p}\cdot\mathbf{x} - G^{\mathbf{Y}}(\mathbf{p})\} < +\infty$. Using the positive linear homogeneity, $\mathbf{x} \in \mathbf{X}$ if and only if $\mathbf{p}\cdot\mathbf{x} \leqq G^{\mathbf{Y}}(\mathbf{p})$ for all $\mathbf{p} \in \mathbf{NY}$. Q.E.D.

12.5. *Lemma.* If $\langle F,\mathbf{X}\rangle$ is convex, closed, and positively linear homogeneous, and $\operatorname{int}\mathbf{X} \neq \emptyset$, then $\mathbf{Y} = \{\mathbf{y} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{y} \leqq F(\mathbf{p}) \text{ for all } \mathbf{p} \in \mathbf{X}\}$ is non-empty, convex, closed, and semi-bounded, and $\langle F,\mathbf{X}\rangle$ is the support function of $\mathbf{Y}$.

Proof: By 12.3 and the homogeneity argument used in the proof of 12.4, the conjugate dual of $\langle F,\mathbf{X}\rangle$ is $\langle G,\mathbf{Y}\rangle$, with $\mathbf{Y}$ the set given in the statement of this lemma, and $G(\mathbf{y}) = 0$ for $\mathbf{y} \in \mathbf{Y}$. Hence, by 12.3, $\mathbf{Y}$ is non-empty and convex. The closedness of $\mathbf{Y}$ is immediate from its definition. Since $\mathbf{p} \in \mathbf{X}$ implies $\mathbf{p}\cdot\mathbf{y} \leqq F(\mathbf{p})$ for all $\mathbf{y} \in \mathbf{Y}$, $\mathbf{X} \subseteq \mathbf{NY} \subseteq \mathbf{P}(\mathbf{AY})$, implying $\mathbf{AY} \subseteq \mathbf{PX}$. Since $\operatorname{int}\mathbf{X} \neq \emptyset$, $\mathbf{PX}$ is pointed by 11.3 (3), and hence $\mathbf{Y}$ is semi-bounded. By the argument of 12.4, $\langle G^{\mathbf{Y}},\mathbf{NY}\rangle = D\langle G,\mathbf{Y}\rangle$, implying $\langle G^{\mathbf{Y}},\mathbf{NY}\rangle = \langle F,\mathbf{X}\rangle$ by 12.3. Q.E.D.

12.6. *Definition.* A set $\mathbf{X}$ is a *polytope* if it is the convex hull of a finite set of points. $\mathbf{X}$ is *boundedly polyhedral* if its intersection with any polytope is a polytope.

12.7. *Lemma.* If $\langle F,\mathbf{X}\rangle$ is convex and closed, and $\mathbf{X}$ is boundedly polyhedral, then $\langle F,\mathbf{X}\rangle$ is continuous; i.e., $\mathbf{v} \in \mathbf{X}$ implies $\lim_{\mathbf{u}\in\mathbf{X},\mathbf{u}\to\mathbf{v}} F(\mathbf{u}) = F(\mathbf{v})$.

Proof: Gale, Klee, and Rockafellar (1968, Thm. 2) establish $\lim_{\mathbf{u}\in\mathbf{X},\mathbf{u}\to\mathbf{v}} \sup F(\mathbf{u}) = F(\mathbf{v})$. Since $\langle F,\mathbf{X}\rangle$ is closed, the result follows. Q.E.D.

13. Properties of Maximand Correspondences

We now establish properties of maximands of $\mathbf{p}\cdot\mathbf{y}$ for $\mathbf{y}$ in a closed, semi-bounded set $\mathbf{Y}$. In this and succeeding sections, we shall deal with pairs of Euclidean spaces $\mathbf{E}^N$ and $\mathbf{E}^M$. The spaces in which sets lie will be clear from the context.

13.1. *Definition.* A mapping $\mathbf{\Phi}$ from a non-empty set $\mathbf{Z} \subseteq \mathbf{E}^M$ into *subsets* of $\mathbf{E}^N$ (i.e., $\mathbf{\Phi}(\mathbf{z}) \subseteq \mathbf{E}^N$ for each $\mathbf{z} \in \mathbf{Z}$) is termed a *set-valued function* and denoted by $\langle \mathbf{\Phi},\mathbf{Z}\rangle$. If $\mathbf{\Phi}(\mathbf{z})$ is non-empty for all $\mathbf{z} \in \mathbf{Z}$, then $\langle \mathbf{\Phi},\mathbf{Z}\rangle$ is termed a *correspondence.* We say a set-valued function or correspondence $\langle \mathbf{\Phi},\mathbf{Z}\rangle$ is *convex-valued* (or closed-, compact-, or semi-bounded-valued) if $\mathbf{\Phi}(\mathbf{z})$ is convex (or closed, compact, or semi-bounded) for each $\mathbf{z} \in \mathbf{Z}$. If $\bigcup_{\mathbf{z}\in\mathbf{Z}} \mathbf{\Phi}(\mathbf{z})$ is bounded (or semi-bounded), we say $\langle \mathbf{\Phi},\mathbf{Z}\rangle$ has *bounded range* (or semi-bounded range).

13.2. *Definition.* A correspondence $\langle \mathbf{\Phi},\mathbf{Z}\rangle$, $\mathbf{Z} \subseteq \mathbf{E}^M$, $\mathbf{\Phi}(\mathbf{z}) \subseteq \mathbf{E}^N$, is *upper hemicontinuous* if $\mathbf{z}^j \in \mathbf{Z}$, $\mathbf{z}^j \to \mathbf{z} \in \mathbf{Z}$, $\mathbf{y}^j \in \mathbf{\Phi}(\mathbf{z}^j)$, $\mathbf{y}^j \to \mathbf{y}$ implies $\mathbf{y} \in \mathbf{\Phi}(\mathbf{z})$. $\langle \mathbf{\Phi},\mathbf{Z}\rangle$ is *lower hemicontinuous* if $\mathbf{z}^j \in \mathbf{Z}$, $\mathbf{z}^j \to \mathbf{z} \in \mathbf{Z}$, $\mathbf{y} \in \mathbf{\Phi}(\mathbf{z})$ implies there exist $\mathbf{y}^j \in \mathbf{\Phi}(\mathbf{z}^j)$ with $\mathbf{y}^j \to \mathbf{y}$. $\langle \mathbf{\Phi},\mathbf{Z}\rangle$ is *strongly upper hemicontinuous* if it is upper hemicontinuous and $\mathbf{z}^j \in \mathbf{Z}$, $\mathbf{z}^j \to \mathbf{z} \in \mathbf{Z}$, $\mathbf{y}^j \in \mathbf{\Phi}(\mathbf{z}^j)$, $\theta_j \geqq 0$, $\theta_j \to 0$, $\theta_j\mathbf{y}^j \to \mathbf{y}$ implies $\mathbf{y} \in \mathbf{A\Phi}(\mathbf{z})$. $\langle \mathbf{\Phi},\mathbf{Z}\rangle$ is *(strongly) continuous* if it is lower hemicontinuous and (strongly) upper hemicontinuous.

13.3 *Lemma.* A correspondence $\langle \mathbf{\Phi},\mathbf{Z}\rangle$ has the following properties: (1) if it is upper hemicontinuous, then it is closed-valued; (2) if it is continuous and convex-valued, then it is strongly continuous; (3) suppose it is lower hemicontinuous and convex-valued, with $\operatorname{int}\mathbf{\Phi}(\mathbf{z}) \neq \emptyset$ for $\mathbf{z} \in \mathbf{Z}$. Then for any sequence $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$ and any compact set $\mathbf{R} \subseteq \operatorname{int}\mathbf{\Phi}(\mathbf{z}^0)$, there exists k_0 such that $\mathbf{R} \subseteq \operatorname{int}\mathbf{\Phi}(\mathbf{z}^k)$ for $k \geqq k_0$.

Proof: The first proposition is an immediate consequence of the definition of upper hemicontinuity. To show the second, suppose $\mathbf{z}^j \in \mathbf{Z}$, $\mathbf{z}^j \to \mathbf{z} \in \mathbf{Z}$, $\mathbf{y}^j \in \mathbf{\Phi}(\mathbf{z}^j)$, $\theta_j \geqq 0$ such that $\theta_j \to 0$ and $\theta_j\mathbf{y}^j \to \mathbf{y}$. Take any $\mathbf{v} \in \mathbf{\Phi}(\mathbf{z})$. By lower hemicontinuity, there exist $\mathbf{v}^j \in \mathbf{\Phi}(\mathbf{z}^j)$, $\mathbf{v}^j \to \mathbf{v}$. For any $\lambda > 0$, $0 \leqq \theta_j\lambda < 1$ for j large, and $\theta_j\lambda\mathbf{y}^j + (1-\theta_j\lambda)\mathbf{v}^j \in \mathbf{\Phi}(\mathbf{z}^j)$ with $\theta_j\lambda\mathbf{y}^j + (1-\theta_j\lambda)\mathbf{v}^j \to \lambda\mathbf{y} + \mathbf{v} \in \mathbf{\Phi}(\mathbf{z})$, by upper hemicontinuity. Hence, $\mathbf{y} \in \mathbf{A\Phi}(\mathbf{z})$.

To prove (3), first consider a sequence $\mathbf{z}^k \in \mathbf{Z}$, $\mathbf{z}^k \to \mathbf{z}^0 \in \mathbf{Z}$, and a vector $\mathbf{y}^* \in \operatorname{int}\mathbf{\Phi}(\mathbf{z}^0)$. Let $\mathbf{e}^j$ denote the jth unit vector in $\mathbf{E}^N$, and $\mathbf{e}_N^0$ denote a vector of ones in $\mathbf{E}^N$. For a positive scalar δ, define $\mathbf{y}^0 = \mathbf{y}^* - \delta\mathbf{e}_N^0$ and

$\mathbf{y}^j = \mathbf{y}^* + \delta\mathbf{e}^j$, $j = 1,\ldots,N$. For δ sufficiently small, $\mathbf{y}^j \in \operatorname{int}\mathbf{\Phi}(\mathbf{z}^0)$, $j = 0,\ldots,N$, and $\mathbf{y}^* = (\sum_{j=0}^N \mathbf{y}^j)/(N+1)$. By the lower hemicontinuity of $\mathbf{\Phi}$, there exist $\mathbf{y}^{jk} \in \operatorname{int}\mathbf{\Phi}(\mathbf{z}^k)$ with $\mathbf{y}^{jk} \to \mathbf{y}^j$, $j = 0,\ldots,N$. Treating the $\mathbf{y}$ vectors as column vectors, form $(N+1)$-dimensional matrices $\mathbf{A}_k$ and $\mathbf{A}_0$ and $(N+1)$-dimensional column vectors $\mathbf{a}$ and $\boldsymbol{\theta}$ satisfying

$$\mathbf{A}_k = \left[\begin{array}{cccc}\mathbf{y}^{0k} & \mathbf{y}^{1k} & \cdots & \mathbf{y}^{Nk} \\ \hline 1 & 1 & \cdots & 1\end{array}\right], \qquad \mathbf{A}_0 = \left[\begin{array}{cccc}\mathbf{y}^{0} & \mathbf{y}^{1} & \cdots & \mathbf{y}^{N} \\ \hline 1 & 1 & \cdots & 1\end{array}\right],$$

$$\mathbf{a} = \left[\begin{array}{c}\mathbf{y}^* \\ \hline 1\end{array}\right], \qquad \boldsymbol{\theta} = \begin{bmatrix}\theta_0 \\ \vdots \\ \theta_N\end{bmatrix}.$$

Then, $\mathbf{A}_0$ is non-singular, and $\mathbf{A}_0\boldsymbol{\theta} = \mathbf{a}$ has a unique solution $\boldsymbol{\theta}_0 = (N+1)^{-1}\,\mathbf{e}^0_{N+1}$. Since $\mathbf{A}_k \to \mathbf{A}_0$, it follows that $\mathbf{A}_k$ is non-singular for k large, and $\mathbf{A}_k^{-1} \to \mathbf{A}_0^{-1}$. Define $\boldsymbol{\theta}_k = \mathbf{A}_k^{-1}\mathbf{a}$. Then, $\boldsymbol{\theta}_k \to \boldsymbol{\theta}_0$, implying $\boldsymbol{\theta}_k$ non-negative for all sufficiently large k. Then, for all sufficiently large k, the polytope with vertices $\mathbf{y}^* - (\delta/2)\mathbf{e}^0_N$ and $\mathbf{y}^* + (\delta/2)\mathbf{e}^j$ for $j = 1,\ldots,N$, contains $\mathbf{y}^*$ in its interior and is contained in $\mathbf{\Phi}(\mathbf{z}^k)$. Therefore, for each $\mathbf{y}^* \in \operatorname{int}\mathbf{\Phi}(\mathbf{z}^0)$, there exists an open neighborhood $\mathbf{N}_{\mathbf{y}^*}$ of $\mathbf{y}^*$ and an index $k_{\mathbf{y}^*}$ such that $\mathbf{N}_{\mathbf{y}^*} \subseteq \operatorname{int}\mathbf{\Phi}(\mathbf{z}^k)$ for $k = 0$ and $k \geqq k_{\mathbf{y}^*}$. If $\mathbf{R}$ is a compact subset of $\operatorname{int}\mathbf{\Phi}(\mathbf{z}^0)$, then the neighborhoods $\mathbf{N}_{\mathbf{y}^*}$ for $\mathbf{y}^* \in \mathbf{R}$ cover $\mathbf{R}$, implying the existence of a finite subcovering of $\mathbf{R}$. Then, for $k \geqq k_{\mathbf{y}^*}$ for each $\mathbf{y}^*$ in the finite subcovering, $\mathbf{R} \subseteq \operatorname{int}\mathbf{\Phi}(\mathbf{z}^k)$. Q.E.D.

13.4. *Definition.* For a non-empty, closed, semi-bounded set $\mathbf{Y} \subseteq \mathbf{E}^N$, define $\mathbf{p}^{\mathbf{Y}} = \{\mathbf{p} \in \mathbf{NY} | G^{\mathbf{Y}}(\mathbf{p}) = \mathbf{p}\cdot\mathbf{y}$ for some $\mathbf{y} \in \mathbf{Y}\}$ and $\mathbf{\Phi}^{\mathbf{Y}}(\mathbf{p}) = \{\mathbf{y} \in \mathbf{Y} | G^{\mathbf{Y}}(\mathbf{p}) = \mathbf{p}\cdot\mathbf{y}\}$ for $\mathbf{p} \in \mathbf{P}^{\mathbf{Y}}$. We term $\langle\mathbf{\Phi}^{\mathbf{Y}},\mathbf{P}^{\mathbf{Y}}\rangle$ the *maximand correspondence* of $\mathbf{Y}$.

13.5. *Lemma.* If $\mathbf{Y} \subseteq \mathbf{E}^N$ is non-empty, closed, and semi-bounded, then (1) $\mathbf{Q} \equiv \operatorname{int}\mathbf{NY} \subseteq \mathbf{P}^{\mathbf{Y}}$; (2) $\langle\mathbf{\Phi}^{\mathbf{Y}},\mathbf{P}^{\mathbf{Y}}\rangle$ is closed-valued and upper hemicontinuous; (3) $\langle\mathbf{\Phi}^{\mathbf{Y}},\mathbf{Q}\rangle$ is compact-valued; (4) for any compact non-empty set $\mathbf{R} \subseteq \mathbf{Q}$, $\langle\mathbf{\Phi}^{\mathbf{Y}},\mathbf{R}\rangle$ has bounded range; and (5) if $\mathbf{Y}$ is convex, then $\langle\mathbf{\Phi}^{\mathbf{Y}},\mathbf{P}^{\mathbf{Y}}\rangle$ is convex-valued, and if $\mathbf{p} \in \mathbf{Q}$, $\mathbf{y} \in \mathbf{\Phi}^{\mathbf{Y}}(\mathbf{p})$, $\mathbf{y} \notin \operatorname{intr}\mathbf{\Phi}^{\mathbf{Y}}(\mathbf{p})$, then there exists $(\mathbf{y}_j,\mathbf{p}^j) \to (\mathbf{y},\mathbf{p})$ with $\mathbf{y}^j \in \mathbf{\Phi}^{\mathbf{Y}}(\mathbf{p}^j)$ and $\mathbf{p}^j \in \mathbf{Q}$ not proportional to $\mathbf{p}$.

Proof: We first show that for any non-empty compact set $\mathbf{R} \subseteq \mathbf{Q}$, $\bigcup_{\mathbf{z}\in\mathbf{R}}\mathbf{\Phi}(\mathbf{z})$ is non-empty and bounded. Taking $\mathbf{R} = \{\mathbf{p}\}$, this will verify (1). Suppose there exist $\mathbf{R} \subseteq \mathbf{Q}$, $\mathbf{R} \neq \boldsymbol{\emptyset}$, $\mathbf{R}$ compact, $\mathbf{p}^j \in \mathbf{R}$, $\mathbf{y}^j \in \mathbf{Y}$ such that $|\mathbf{y}^j| \to +\infty$ and $\mathbf{p}^j\cdot\mathbf{y}^j - G^{\mathbf{Y}}(\mathbf{p}^j) \to 0$. By the compactness of $\mathbf{R}$ and the continuity of $G^{\mathbf{Y}}$, we can assume $\mathbf{p}^j \to \mathbf{p} \in \mathbf{R}$ and $\mathbf{p}^j\cdot\mathbf{y}^j \to G^{\mathbf{Y}}(\mathbf{p})$. Then, there exists a subsequence of j (retain notation) such that $\mathbf{y}^j/|\mathbf{y}^j| \to \mathbf{y} \in \mathbf{AY}$, $\mathbf{y} \neq \mathbf{0}$, and $\mathbf{p}\cdot\mathbf{y} = 0$, contradicting $\mathbf{p} \in \mathbf{Q}$ by 11.3(3). Hence, for $\mathbf{p} \in \mathbf{R}$, an optimizing sequence $\mathbf{y}^j \in \mathbf{Y}$ with $\mathbf{p}\cdot\mathbf{y}^j \to G^{\mathbf{Y}}(\mathbf{p})$ is bounded, and has a subsequence converging to $\mathbf{y} \in \mathbf{Y}$ with $\mathbf{p}\cdot\mathbf{y} = G^{\mathbf{Y}}(\mathbf{p})$. This proves $\bigcup_{\mathbf{z}\in\mathbf{R}} \Phi(\mathbf{z})$ non-empty and bounded, verifying (1) and (4). If $\mathbf{p}^j \in \mathbf{P}^{\mathbf{Y}}$, $\mathbf{p}^j \to \mathbf{p} \in \mathbf{P}^{\mathbf{Y}}$, $\mathbf{y}^j \in \Phi^{\mathbf{Y}}(\mathbf{p}^j)$, $\mathbf{y}^j \to \mathbf{y}$, then $\mathbf{p}^j\cdot\mathbf{y}^j \geqq \mathbf{p}^j\cdot\mathbf{w}$ for $\mathbf{w} \in \mathbf{Y}$, implying in the limit $\mathbf{p}\cdot\mathbf{y} \geqq \mathbf{p}\cdot\mathbf{w}$ for $\mathbf{w} \in \mathbf{Y}$, and hence $\mathbf{y} \in \Phi^{\mathbf{Y}}(\mathbf{p})$. Hence, $\langle \Phi^{\mathbf{Y}}, \mathbf{P}^{\mathbf{Y}} \rangle$ is upper hemicontinuous, and therefore closed-valued. This verifies (2). By (2) and (4), $\langle \Phi^{\mathbf{Y}}, \mathbf{Q} \rangle$ is compact-valued, verifying (3).

Suppose in (5) that $\mathbf{Y}$ is convex. Then, $\mathbf{y}, \mathbf{z} \in \Phi^{\mathbf{Y}}(\mathbf{p})$ implies $\mathbf{p}\cdot(\theta\mathbf{y} + (1-\theta)\mathbf{z}) \geqq \mathbf{p}\cdot\mathbf{w}$ for $\mathbf{w} \in \mathbf{Y}$, $0 < \theta < 1$, and hence $\Phi^{\mathbf{Y}}(\mathbf{p})$ convex. Suppose $\mathbf{p} \in \mathbf{Q}$, $\mathbf{y} \in \Phi^{\mathbf{Y}}(\mathbf{p})$, $\mathbf{y} \notin \operatorname{intr} \Phi^{\mathbf{Y}}(\mathbf{p})$. We first show $\mathbf{y} \notin \operatorname{intr} \mathbf{Y}$. If $\Phi^{\mathbf{Y}}(\mathbf{p}) = \mathbf{Y}$, this result is immediate. Alternately, there exists $\mathbf{w} \in \mathbf{Y}$, $\mathbf{w} \notin \Phi^{\mathbf{Y}}(\mathbf{p})$, and hence $\mathbf{p}\cdot\mathbf{y} > \mathbf{p}\cdot\mathbf{w}$. If $\mathbf{y} \in \operatorname{intr} \mathbf{Y}$, then for small $\theta > 0$, $(1+\theta)\mathbf{y} - \theta\mathbf{w} \in \mathbf{Y}$, implying the contradictory inequality $\mathbf{p}\cdot\mathbf{y} \leqq \mathbf{p}\cdot\mathbf{w}$. Two cases will be distinguished:

Case 1. $\Phi^{\mathbf{Y}}(\mathbf{p}) \cap (\operatorname{intr} \mathbf{Y}) \neq \emptyset$. By 10.12, there exists $\mathbf{q} \neq \mathbf{0}$ such that $\mathbf{Y} \subseteq \mathbf{H}^-(\mathbf{q}, \mathbf{q}\cdot\mathbf{y})$ and $(\operatorname{intr} \mathbf{Y}) \cap \mathbf{H}(\mathbf{q}, \mathbf{q}\cdot\mathbf{y}) = \phi$. This implies $\mathbf{q}$ not proportional to $\mathbf{p}$. Defining $\mathbf{p}^j = (1 - j^{-1})\mathbf{p} + j^{-1}\mathbf{q}$ and $\mathbf{y}^j = \mathbf{y}$, the sequence $(\mathbf{y}^j, \mathbf{p}^j) \to (\mathbf{y}, \mathbf{p})$ satisfies (5).

Case 2. $\Phi^{\mathbf{Y}}(\mathbf{p}) \cap \operatorname{intr} \mathbf{Y} = \emptyset$. Since $\mathbf{y} \notin \operatorname{intr} \Phi^{\mathbf{Y}}(\mathbf{p})$, there exists a sequence $\mathbf{z}^j$ in the flat spanned by $\Phi^{\mathbf{Y}}(\mathbf{p})$ such that $\mathbf{z}^j \notin \Phi^{\mathbf{Y}}(\mathbf{p})$ and $\mathbf{z}^j \to \mathbf{y}$. Then, in particular, $\mathbf{z}^j$ is in the flat spanned by $\mathbf{Y}$ and is in the hyperplane $\mathbf{H}(\mathbf{p}, \mathbf{p}\cdot\mathbf{y})$. Choose $\mathbf{w} \in \operatorname{intr} \mathbf{Y}$. Then, there exists θ_j, $0 < \theta_j < 1$, such that $\mathbf{y}^j = \theta_j\mathbf{z}^j + (1-\theta_j)\mathbf{w} \in \mathbf{Y}$ and $\mathbf{y}^j \notin \operatorname{intr} \mathbf{Y}$. If a subsequence had $\theta_j \to \theta < 1$, then $\mathbf{y}^j \to \theta\mathbf{y} + (1-\theta)\mathbf{w} \in \operatorname{intr} \mathbf{Y}$, contradicting the closedness of $\mathbf{Y} \setminus (\operatorname{intr} \mathbf{Y})$. Hence, $\theta_j \to 1$ and $\mathbf{y}^j \to \mathbf{y}$. By 10.12, for each $\mathbf{y}^j$ there exists $\mathbf{p}^j$ such that $|\mathbf{p}^j| = 1$ and $\mathbf{p}^j\cdot\mathbf{y}^j \geqq \mathbf{p}^j\cdot\mathbf{w}$ for all $\mathbf{w} \in \mathbf{Y}$. Since $\mathbf{y}^j \notin \Phi^{\mathbf{Y}}(\mathbf{p})$, $\mathbf{p}^j$ is not proportional to $\mathbf{p}$. Take a subsequence of j with $\mathbf{p}^j$ converging to a vector $\mathbf{q}$. If $\mathbf{q}$ is a positive multiple of $\mathbf{p}$, then after the $\mathbf{p}^j$ are rescaled by dividing by this quantity, $(\mathbf{y}^j, \mathbf{p}^j) \to (\mathbf{y}, \mathbf{p})$ is the required sequence. If $\mathbf{q}$ is a negative multiple of $\mathbf{p}$, then $\mathbf{Y} \subseteq \mathbf{H}(\mathbf{p}, \mathbf{p}\cdot\mathbf{y})$, contradicting this case. If $\mathbf{q}$ is not proportional to $\mathbf{p}$, then $\mathbf{y}^j = \mathbf{y}$ and $\mathbf{p}^j = (1 - j^{-1})\mathbf{p} + j^{-1}\mathbf{q}$ gives the required sequence. This verifies (5). Q.E.D.

13.6. *Remark.* The *interior* of the set $\mathbf{P}^{\mathbf{Y}}$ on which the support function $G^{\mathbf{Y}}$ achieves a maximum is a non-empty convex set when $\mathbf{Y}$ is closed and semi-bounded, and its closure is also convex; i.e., we have the string of inclusions

$$\emptyset \neq \operatorname{int} \mathbf{P}(\mathbf{AY}) \subseteq \mathbf{P}^{\mathbf{Y}} \subseteq \mathbf{NY} \subseteq \mathbf{P}(\mathbf{AY}),$$

with $\mathbf{P}(\mathbf{AY})$ closed and convex. Nevertheless, $\mathbf{P}^{\mathbf{Y}}$ itself, need not be convex. One example has been given by Winter (forthcoming); a second follows: define the set

$$\mathbf{Y} = \{\mathbf{y} \in \mathbf{E}^3 | y_2, y_3 \leqq 0, y_1 \leqq (y_2 + y_3)^{-1} < 0\}.$$

This set is closed and convex. For $\mathbf{p}^1 = (0,0,1)$, a maximum of $\mathbf{p}^1 \cdot \mathbf{y}$ for $\mathbf{y} \in \mathbf{Y}$ is attained at $\mathbf{y} = (-1,-1,0)$. For $\mathbf{p}^2 = (0,1,0)$, a maximum is attained at $\mathbf{y} = (-1,0,-1)$. But for $\mathbf{p} = \mathbf{p}^1 + \mathbf{p}^2 = (0,1,1)$, the supremum of $\mathbf{p} \cdot \mathbf{y}$ for $\mathbf{y} \in \mathbf{Y}$, equal to zero, is approached by $y_2 + y_3 = y_1^{-1}$ and $y_1 \to \infty$, but is not achieved.

13.7. *Definition.* For a convex, closed, positively linear homogeneous real-valued function $\langle F, \mathbf{X} \rangle$, define a set-valued function $\langle \mathbf{\Gamma}, \operatorname{intr} \mathbf{X} \rangle$, $\mathbf{\Gamma}(\mathbf{x}) \subseteq \mathbf{E}^N$, by $\mathbf{y} \in \mathbf{\Gamma}(\mathbf{x})$ if and only if for all $\mathbf{z} \in \mathbf{X}$,

$$(\mathbf{z} - \mathbf{x}) \cdot \mathbf{y} \leqq \lim_{\theta \to 0^+} \inf (F((1-\theta)\mathbf{x} + \theta \mathbf{z}) - F(\mathbf{x}))/\theta.$$

$\langle \mathbf{\Gamma}, \operatorname{intr} \mathbf{X} \rangle$ is termed the ***sub-differential*** of $\langle F, \mathbf{X} \rangle$. Note that if F is differentiable at $\mathbf{x} \in \operatorname{int} \mathbf{X}$, then $\lim_{\theta \to 0}(F((1-\theta)\mathbf{x} + \theta \mathbf{z}) - F(\mathbf{x}))/\theta$ exists for all $\mathbf{z} \in \mathbf{E}^N$, and $\mathbf{y}$ is unique and equals the gradient (i.e., the vector of partial derivatives of F).

13.8. *Lemma.* If $\langle F, \mathbf{X} \rangle$ is convex, closed, and positively linear homogeneous, then (1) the sub-differential $\langle \mathbf{\Gamma}, \operatorname{intr} \mathbf{X} \rangle$ is a convex-valued upper hemicontinuous correspondence, with $\mathbf{y} \in \mathbf{\Gamma}(\mathbf{x})$, $\mathbf{x} \in \operatorname{intr} \mathbf{X}$, if and only if $F(\mathbf{x}) = \mathbf{y} \cdot \mathbf{x}$ and $F(\mathbf{z}) \geqq \mathbf{y} \cdot \mathbf{z}$ for all $\mathbf{z} \in \mathbf{X}$; (2) if $\operatorname{int} \mathbf{X} \neq \emptyset$, $\langle \mathbf{\Gamma}, \operatorname{int} \mathbf{X} \rangle$ is compact-valued; (3) F is differentiable at $\mathbf{x} \in \operatorname{int} \mathbf{X}$ with a vector of partial derivatives $\mathbf{y}$ if and only if $\mathbf{\Gamma}(\mathbf{x}) = \{\mathbf{y}\}$; and (4) if $\mathbf{x} \in \operatorname{int} \mathbf{X} \neq \emptyset$ and $\mathbf{Y} = \{\mathbf{w} \in \mathbf{E}^N | \mathbf{x} \cdot \mathbf{w} \leqq F(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}\}$, then $\mathbf{y} \in \mathbf{\Gamma}(\mathbf{x})$ if and only if $\mathbf{y} \in \mathbf{Y}$ and $\mathbf{x} \cdot \mathbf{y} \geqq \mathbf{x} \cdot \mathbf{w}$ for all $\mathbf{w} \in \mathbf{Y}$.

Proof: See also Rockafellar (1970, Sect. 23). (a) Consider the set $\mathbf{A} = \{(\mathbf{x}, \xi) \in \mathbf{E}^{N+1} | \mathbf{x} \in \mathbf{X},\ \xi \geqq F(\mathbf{x})\}$. One can easily show that $\mathbf{A}$ is convex and closed, and that for any $\mathbf{x} \in \operatorname{intr} \mathbf{X}$, $(\mathbf{x}, F(\mathbf{x})) \notin \operatorname{intr} \mathbf{A}$. By 10.12, there exists $(\mathbf{y}, -\eta) \in \mathbf{E}^{N+1}$, $(\mathbf{y}, -\eta) \neq 0$, such that $\mathbf{y} \cdot \mathbf{x} - \eta F(\mathbf{x}) \geqq \mathbf{y} \cdot \mathbf{z} - \eta \xi$ for all $(\mathbf{z}, \xi) \in \mathbf{A}$, with the inequality strict for $(\mathbf{z}, \xi) \in \operatorname{intr} \mathbf{A}$. Taking $\mathbf{z} = \mathbf{x}$ and $\xi > F(\mathbf{x})$ yields $(\mathbf{z}, \xi) \in \operatorname{intr} \mathbf{A}$ and implies $\eta > 0$. Normalize $\eta = 1$. For $\mathbf{z} \in \mathbf{X}$, $0 < \theta < 1$, the inequality yields $\mathbf{y} \cdot (\theta \mathbf{z} + (1-\theta)\mathbf{x}) - \mathbf{y} \cdot \mathbf{x} = \theta \mathbf{y} \cdot (\mathbf{z} - \mathbf{x}) \leqq F(\theta \mathbf{z} + (1-\theta)\mathbf{x}) - F(\mathbf{x})$. Letting $\theta \to 0^+$, this implies $\mathbf{y} \in \mathbf{\Gamma}(\mathbf{x})$. Hence, $\langle \mathbf{\Gamma}, \operatorname{intr} \mathbf{X} \rangle$ is a correspondence.

(b) We next show that $(F(\theta \mathbf{z} + (1-\theta)\mathbf{x}) - F(\mathbf{x}))/\theta$ is a non-decreasing function of positive θ for $\mathbf{z} \in \mathbf{X}$, $\mathbf{x} \in \operatorname{intr} \mathbf{X}$. Suppose $0 < \theta_1 < \theta_2$ and let

$\alpha = \theta_1/\theta_2$. By the convexity of F, $F(\alpha(\theta_2\mathbf{z} + (1-\theta_2)\mathbf{x}) + (1-\alpha)\mathbf{x}) = F(\theta_1\mathbf{z} + (1-\theta_1)\mathbf{x}) \leqq \alpha F(\theta_2\mathbf{z} + (1-\theta_2)\mathbf{x}) + (1-\alpha)F(\mathbf{x})$, or $(F(\theta_1\mathbf{z} + (1-\theta_1)\mathbf{x}) - F(\mathbf{x}))/\theta_1 \leqq (F(\theta_2\mathbf{z} + (1-\theta_2)\mathbf{x}) - F(\mathbf{x}))/\theta_2$.

(c) Suppose $\mathbf{x} \in \operatorname{intr} \mathbf{X}$, $\mathbf{y} \in \Gamma(\mathbf{x})$. For any $\mathbf{v} \in \mathbf{E}^N$ with $\mathbf{z} = \mathbf{x} + \mathbf{v} \in \mathbf{X}$, paragraph (b) and the definition of Γ imply $\mathbf{v}\cdot\mathbf{y} \leqq F(\mathbf{v}+\mathbf{x}) - F(\mathbf{x})$. Taking $\mathbf{v} = -\mathbf{x}/2$, homogeneity implies $\mathbf{x}\cdot\mathbf{y} \geqq F(\mathbf{x})$. Taking $\mathbf{v} \in \mathbf{X}$, convexity implies $\mathbf{v}\cdot\mathbf{y} \leqq F(\mathbf{v})$. Hence, we have established $\mathbf{y} \in \Gamma(\mathbf{x})$ if and only if $\mathbf{x}\cdot\mathbf{y} = F(\mathbf{x})$ and $\mathbf{z}\cdot\mathbf{y} \leqq F(\mathbf{z})$ for all $\mathbf{z} \in \mathbf{X}$. Since F is continuous on intr $\mathbf{X}$ by 12.1, the results that $\langle \Gamma, \operatorname{intr} \mathbf{X}\rangle$ is convex-valued and upper hemicontinuous follow immediately from this characterization, verifying (1).

(d) In (2), (3), (4), int $\mathbf{X} \neq \phi$ is assumed. Then, by 12.5, $\mathbf{Y}$ defined in (4) is non-empty, convex, closed, and semi-bounded, and $\langle F, \mathbf{X}\rangle$ is the support function of $\mathbf{Y}$. From the characterization of $\langle \Gamma, \operatorname{intr} \mathbf{X}\rangle$ established in (c), (4) holds and $\Gamma(\mathbf{x}) = \Phi^{\mathbf{Y}}(\mathbf{x})$ for $\mathbf{x} \in \operatorname{int} \mathbf{X}$. By 13.5, $\langle \Gamma, \operatorname{int} \mathbf{X}\rangle$ is compact-valued, verifying (2). We noted previously that if F is differentiable at $\mathbf{x} \in \operatorname{int} \mathbf{X}$, then the definition of Γ implies the "only if" implication in (3). The converse implication in (3) is a consequence of the definition of differentiability – a detailed argument is given by Fenchel (1953, Ch. 3, result 32) or Rockafellar (1970, Thm. 25.1). Q.E.D.

13.9. *Corollary.* If $\mathbf{Y}$ is non-empty, closed, and semi-bounded, $\langle G^{\mathbf{Y}}, \mathbf{NY}\rangle$ is the support function of $\mathbf{Y}$, $\langle \Gamma, \operatorname{int} \mathbf{NY}\rangle$ is the subdifferential of $G^{\mathbf{Y}}$, and $\langle \Phi^{\mathbf{Y}}, \operatorname{int} \mathbf{NY}\rangle$ is the maximand correspondence of $\mathbf{Y}$, then $\Gamma(\mathbf{p}) = [\Phi^{\mathbf{Y}}(\mathbf{p})]$ for $\mathbf{p} \in \operatorname{int} \mathbf{NY}$.

Proof: $\mathbf{p} \in \operatorname{int} \mathbf{NY}$ and $\mathbf{y} \in \Phi^{\mathbf{Y}}(\mathbf{p}) \Rightarrow \mathbf{p}\cdot\mathbf{y} \geqq \mathbf{p}\cdot\mathbf{w}$ for all $\mathbf{w} \in \mathbf{Y} \Rightarrow \mathbf{p}\cdot\mathbf{y} \geqq \mathbf{p}\cdot\mathbf{w}$ for all $\mathbf{w} \in [\mathbf{Y}] \Rightarrow \mathbf{y} \in \Phi^{[\mathbf{Y}]}(\mathbf{p})$. Alternately, $\mathbf{y} \in \Phi^{[\mathbf{Y}]}(\mathbf{p}) \Rightarrow$ [by 11.2(4)] $\mathbf{y} = \mathbf{u} + \mathbf{v}$ with $\mathbf{u} \in [\mathbf{Y}]$, $\mathbf{v} \in \mathbf{A}[\mathbf{Y}]$. But $\mathbf{v} \neq \mathbf{0} \Rightarrow \mathbf{p}\cdot\mathbf{v} < 0$ for $\mathbf{p} \in \operatorname{int} \mathbf{NY}$. Hence, $\mathbf{v} = \mathbf{0}$ and $\mathbf{y} \in [\mathbf{Y}] \Rightarrow \mathbf{y} = \sum_{i=0}^{m} \theta_i \mathbf{y}^i$ with $\mathbf{y}^i \in \mathbf{Y}$, $\theta_i \geqq 0$, $\sum_{i=0}^{m} \theta_i = 1$, and $\mathbf{p}\cdot\mathbf{y} \geqq \mathbf{p}\cdot\mathbf{w}$ for all $\mathbf{w} \in [\mathbf{Y}] \Rightarrow \mathbf{y}^i \in \Phi^{\mathbf{Y}}(\mathbf{p}) \Rightarrow \mathbf{y} \in [\Phi^{\mathbf{Y}}(\mathbf{p})]$. Hence, $[\Phi^{\mathbf{Y}}(\mathbf{p})] = \Phi^{[\mathbf{Y}]}(\mathbf{p}) = \Gamma(\mathbf{p})$ by 12.4 and 13.8. Q.E.D.

14. Exposed Sets

The next series of results establish relationships between a closed semi-bounded set and the set of all its maximands.

14.1. *Lemma.* If $\mathbf{Y}$ is non-empty, convex, closed, and semi-bounded, $\mathbf{Z}$ is non-empty, convex, and compact, and $\mathbf{Y} \cap \mathbf{Z} = \emptyset$, then there exists an open set $\mathbf{W} \subseteq \operatorname{int} \mathbf{NY}$ and scalars α, β such that $\mathbf{w}\cdot\mathbf{y} \leqq \alpha < \beta \leqq \mathbf{w}\cdot\mathbf{z}$ for all $\mathbf{w} \in \mathbf{W}$, $\mathbf{y} \in \mathbf{Y}$, $\mathbf{z} \in \mathbf{Z}$. In particular, $\mathbf{w} \in \mathbf{W}$ may be chosen so that $\mathbf{w}\cdot\mathbf{y}$ is maximized over $\mathbf{y} \in \mathbf{Y}$ at a unique point.

Proof: By 10.13, there exist $\mathbf{p}$, α_1, β_1 such that $\mathbf{p}\cdot\mathbf{y} \leqq \alpha_1 - \beta_1 < \alpha_1 + \beta_1 \leqq \mathbf{p}\cdot\mathbf{z}$ for $\mathbf{y} \in \mathbf{Y}$, $\mathbf{z} \in \mathbf{Z}$. Since $\mathbf{Z}$ is compact, $\text{int}\,\mathbf{NY} \neq \emptyset$, and $\langle G^{\mathbf{Y}},\mathbf{NY}\rangle$ is closed and convex, we can choose $\mathbf{q} \in \text{int}\,\mathbf{NY}$ such that $|\mathbf{q}-\mathbf{p}| < \beta_1/2(1+\max_{\mathbf{Z}}|\mathbf{z}|)$ and $G^{\mathbf{Y}}(\mathbf{q}) - G^{\mathbf{Y}}(\mathbf{p}) < \beta_1/2$, implying $\mathbf{q}\cdot\mathbf{y} \leqq \alpha_1 - \beta_1 < \alpha_1 + \beta_1 \leqq \mathbf{q}\cdot\mathbf{z}$ for $\mathbf{y} \in \mathbf{Y}$, $\mathbf{z} \in \mathbf{Z}$, $\beta_2 = \beta_1/2$. Then, a small open neighborhood $\mathbf{W}$ of $\mathbf{q}$ is contained in $\text{int}\,\mathbf{NY}$, and by the continuity of $G^{\mathbf{Y}}$ and the compactness of $\mathbf{Z}$ can be taken so that the strict separation is preserved. Since $G^{\mathbf{Y}}$ is differentiable almost everywhere in $\text{int}\,\mathbf{NY}$ (Lemma 12.1) and $\mathbf{w}\cdot\mathbf{y}$ achieves a unique maximum on $\mathbf{Y}$ if $G^{\mathbf{Y}}$ is differentiable at $\mathbf{w}$, the last conclusion is immediate. Q.E.D.

14.2. *Definition.* $\mathbf{X}$ is an *exposed set* of a non-empty, closed, semi-bounded set $\mathbf{Y}$ if $\mathbf{X}$ is the intersection of $\mathbf{Y}$ and a supporting hyperplane, i.e., $\Phi^{\mathbf{Y}}(\mathbf{p}) = \mathbf{X}$ for some $\mathbf{p} \in \mathbf{NY}$. If $\mathbf{X} = \{\mathbf{y}\}$, $\mathbf{y}$ is termed an *exposed point*. Let $\mathbf{Y}^{\#}$ denote the set of exposed points of $\mathbf{Y}$.

14.3. *Lemma.* If $\mathbf{Y}$ is non-empty, convex, closed, and semi-bounded, then $\mathbf{Y} = \overline{[\mathbf{Y}^{\#}]} + \mathbf{AY}$.

Proof: By 10.9, $\mathbf{Z} \equiv \overline{[\mathbf{Y}^{\#}]} + \mathbf{AY} \subseteq \mathbf{Y}$, implying $\mathbf{AZ} = \mathbf{AY}$, and hence $\mathbf{P}(\mathbf{AZ}) = \mathbf{P}(\mathbf{AY})$. If $\mathbf{y} \in \mathbf{Y}$, $\mathbf{y} \notin \mathbf{Z}$, then by 14.1 there exists an open set $\mathbf{W} \subseteq \text{int}\,\mathbf{NZ} = \text{int}\,\mathbf{NY}$ such that $\mathbf{p}\cdot\mathbf{z} \leqq \alpha < \beta \leqq \mathbf{p}\cdot\mathbf{y}$ for all $\mathbf{z} \in \mathbf{Z}$, $\mathbf{p} \in \mathbf{W}$. Choose $\mathbf{p} \in \mathbf{W}$ such that $G^{\mathbf{Y}}(\mathbf{p})$ is differentiable. Then, there exists a unique $\mathbf{v} \in \mathbf{Y}$ such that $G^{\mathbf{Y}}(\mathbf{p}) = \mathbf{p}\cdot\mathbf{v} \leqq \mathbf{p}\cdot\mathbf{y}$. But then $\mathbf{v} \in \mathbf{Y}^{\#} \subseteq \mathbf{Z}$, contradicting the inequality $\mathbf{p}\cdot\mathbf{z} \leqq \alpha$. Hence, $\mathbf{Z} = \mathbf{Y}$. Q.E.D.

15. Conjugate Correspondences

Thus far, we have investigated properties of the support function of a single set $\mathbf{Y}$. We now list properties of a family of support functions corresponding to a parametric family of sets $\mathbf{Y}$. This topic does not seem to have been investigated in the mathematical literature, although the work of Rockafellar (1970) on perturbations is closely related. Throughout this section, we shall consider a non-empty set of parameters $\mathbf{V} \subseteq \mathbf{E}^M$ and a mapping from $\mathbf{v} \in \mathbf{V}$ into non-empty subsets $\mathbf{Y}(\mathbf{v}) \subseteq \mathbf{E}^N$. Then, $\langle \mathbf{Y},\mathbf{V}\rangle$ is a correspondence.

15.1. *Definition.* For a correspondence $\langle \mathbf{Y},\mathbf{V}\rangle$ mapping $\mathbf{v} \in \mathbf{V} \subseteq \mathbf{E}^M$ into $\mathbf{Y}(\mathbf{v}) \subseteq \mathbf{E}^N$ which has $\mathbf{Y}(\mathbf{v})$ closed and semi-bounded for each $\mathbf{v} \in \mathbf{V}$, let $\mathbf{AY}(\mathbf{v})$ and $\mathbf{NY}(\mathbf{v})$ denote the asymptotic cone and normal cone of $\mathbf{Y}(\mathbf{v})$, respectively. Then, $\langle \mathbf{AY},\mathbf{V}\rangle$ and $\langle \mathbf{NY},\mathbf{V}\rangle$ are correspondences. Define the sets

$$\mathbf{D} = \{(\mathbf{v},\mathbf{p}) \in \mathbf{E}^{M+N} | \mathbf{v} \in \mathbf{V}, \mathbf{p} \in \mathbf{NY}(\mathbf{v})\},$$

$$\mathbf{D}^0 = \{(\mathbf{v},\mathbf{p}) \in \mathbf{E}^{M+N} | \mathbf{v} \in \mathbf{V}, \mathbf{p} \in \text{int}\, \mathbf{NY}(\mathbf{v})\}.$$

On the domain $\mathbf{D}$, define the support function G by

$$G(\mathbf{v},\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{y} | \mathbf{y} \in \mathbf{Y}(\mathbf{v})\}.$$

On the domain $\mathbf{D}^0$, let $\mathbf{\Gamma}(\mathbf{v},\mathbf{p})$ denote the sub-differential of G, and let

$$\mathbf{\Phi}(\mathbf{v},\mathbf{p}) = \{\mathbf{y} \in \mathbf{Y}(\mathbf{v}) | \mathbf{p}\cdot\mathbf{y} \geqq \mathbf{p}\cdot\mathbf{w} \text{ for all } \mathbf{w} \in \mathbf{Y}(\mathbf{v})\}$$

denote the maximand correspondence. The abbreviated notation $\langle G,\mathbf{D}\rangle$,$\langle \mathbf{\Gamma},\mathbf{C}^0\rangle$, $\langle \mathbf{\Phi},\mathbf{D}^0\rangle$ will also be used for these mappings.

15.2. *Lemma.* If $\langle \mathbf{Y},\mathbf{V}\rangle$ is a strongly continuous semi-bounded-valued correspondence, then (1) $\langle \mathbf{AY},\mathbf{V}\rangle$ is a upper hemicontinuous correspondence; (2) $\langle \mathbf{NY},\mathbf{V}\rangle$ is a lower hemicontinuous correspondence; and (3) if $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v} \in \mathbf{V}$, $\mathbf{R} \subseteq \text{int}\, \mathbf{NY}(\mathbf{v})$ is non-empty and compact, then there exists j_0 such that

$$\mathbf{R} \subseteq \text{int}\, \mathbf{NY}(\mathbf{v}^j) \quad \text{for} \quad j \geqq j_0,$$

and

$$\mathbf{Z} = \bigcup_{j \geqq j_0} \bigcup_{\mathbf{p} \in \mathbf{R}} \mathbf{\Phi}(\mathbf{v}^j,\mathbf{p})$$

is bounded.

Proof: If $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v} \in \mathbf{V}$, $\mathbf{y}^j \in \mathbf{AY}(\mathbf{v}^j)$, $\mathbf{y}^j \to \mathbf{y}$, then there exist $\mathbf{w}^j \in \mathbf{Y}(\mathbf{v}^j)$ and $\theta_j \geqq 0$ such that $\theta_j < j^{-1}$ and $|\theta_j \mathbf{w}^j - \mathbf{y}^j| < j^{-1}$, implying $\theta_j \mathbf{w}^j \to \mathbf{y}$. Then, $\mathbf{y} \in \mathbf{AY}(\mathbf{v})$ by strong continuity. This verifies (1).

Consider $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v} \in \mathbf{V}$, $\mathbf{R} \subseteq \text{int}\, \mathbf{NY}(\mathbf{v})$, $\mathbf{R}$ non-empty and compact. Suppose for an infinite subsequence of j, there exists $\mathbf{p}^j \in \mathbf{R}$ with $\mathbf{p}^j \notin \text{int}\, \mathbf{NY}(\mathbf{v}^j)$, implying $\mathbf{p}^j\cdot\mathbf{y}^j \geqq 0$ for some $\mathbf{y}^j \in \mathbf{AY}(\mathbf{v}^j)$ with $|\mathbf{y}^j| = 1$. Then, there exists a subsequence of $(\mathbf{p}^j,\mathbf{y}^j)$ converging to $(\mathbf{p},\mathbf{y})$ with $\mathbf{p} \in \mathbf{R}$, $\mathbf{y} \in \mathbf{AY}(\mathbf{v})$ by (1), and $\mathbf{p}\cdot\mathbf{y} \geqq 0$, contradicting the definition of $\mathbf{R}$. Hence, there exists j_0 such that $\mathbf{R} \subseteq \text{int}\, \mathbf{NY}(\mathbf{v}^j)$ for $j \geqq j_0$. Next suppose there exists $\mathbf{p}^j \in \mathbf{R}$, $\mathbf{y}^j \in \mathbf{\Phi}(\mathbf{v}^j,\mathbf{p}^j)$ for $j \geqq j_0$ with $\mathbf{y}^j$ unbounded. Then there exists a subsequence of j with $\mathbf{y}^j/|\mathbf{y}^j|$ converging to $\mathbf{u} \in \mathbf{AY}(\mathbf{v})$ by strong continuity and $\mathbf{p}^j$ converging to $\mathbf{p} \in \mathbf{R}$. But for any $\mathbf{w} \in \mathbf{Y}(\mathbf{v})$, by lower hemicontinuity there exists $\mathbf{w}^j \in \mathbf{Y}(\mathbf{v}^j)$ with $\mathbf{w}^j \to \mathbf{w}$, and $\mathbf{p}^j\cdot\mathbf{y}^j = G(\mathbf{v}^j,\mathbf{p}^j) \geqq$

$\mathbf{p}^j \cdot \mathbf{w}^j \to \mathbf{p} \cdot \mathbf{w}$, implying $\mathbf{p}^j \cdot \mathbf{y}^j / |\mathbf{y}^j| \to \mathbf{p} \cdot \mathbf{u} \geqq 0$ and contradicting the definition of $\mathbf{R}$. This verifies (3).

Finally, suppose $(\mathbf{v},\mathbf{p}) \in \mathbf{D}$, $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v}$. There exist $\mathbf{p}^i \in \operatorname{int} \mathbf{NY}(\mathbf{v})$, $\mathbf{p}^i \to \mathbf{p}$. $\{\mathbf{p},\mathbf{p}^i\}$ is a compact subset of $\operatorname{int} \mathbf{NY}(\mathbf{v})$, and thus by (3) there exists j_0 such that $\mathbf{p}^i \in \mathbf{NY}(\mathbf{v}^{j_0+i})$, verifying (2). Q.E.D.

15.3. *Lemma.* If $\langle \mathbf{Y},\mathbf{V} \rangle$ is a strongly continuous semi-bounded-valued correspondence, then (1) for fixed $\mathbf{v} \in \mathbf{V}$, the support function $\langle G,\mathbf{NY}(\mathbf{v}) \rangle$ is a convex, closed, positively linear homogeneous function of $\mathbf{p} \in \mathbf{NY}(\mathbf{v})$; (2) $\langle G,\mathbf{D}^0 \rangle$ is continuous [i.e., $G(\mathbf{v},\mathbf{p})$ is continuous jointly in $\mathbf{v}$ and $\mathbf{p}$ at each $(\mathbf{v},\mathbf{p}) \in \mathbf{D}^0$]; (3) $\langle G,\mathbf{D} \rangle$ is lower hemicontinuous; i.e., if $(\mathbf{v},\mathbf{p}) \in \mathbf{D}$, $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v}$, then

$$G(\mathbf{v},\mathbf{p}) = \liminf_{\substack{\mathbf{p}^j \in \mathbf{NY}(\mathbf{v}^j) \\ \mathbf{p}^j \to \mathbf{p}}} G(\mathbf{v}^j,\mathbf{p}^j);$$

(4) If $(\mathbf{v}^j,\mathbf{p}^j) \in \mathbf{D}$, $(\mathbf{v}^j,\mathbf{p}^j) \to (\mathbf{v},\mathbf{p})$ with $\mathbf{v} \in \mathbf{V}$, $\mathbf{p} \notin \mathbf{NY}(\mathbf{v})$, then $\lim_j G(\mathbf{v}^j,\mathbf{p}^j) = +\infty$.

Proof: Result 12.4 implies (1). We next establish an inequality used to prove (3) and (4). Suppose $(\mathbf{v}^j,\mathbf{p}^j) \in \mathbf{D}$ is a sequence converging to $(\mathbf{v},\mathbf{p})$, $\mathbf{v} \in \mathbf{V}$, for which the limit of $G(\mathbf{v}^j,\mathbf{p}^j)$, possibly infinite, exists. Take $\mathbf{y}^i \in \mathbf{Y}(\mathbf{v})$ with $\mathbf{p} \cdot \mathbf{y}^i \to G(\mathbf{v},\mathbf{p})$. By the lower hemicontinuity of $\mathbf{Y}$, there exists a subsequence j_i and points $\mathbf{w}^i \in \mathbf{Y}(\mathbf{v}^{j_i})$ such that $|\mathbf{w}^i - \mathbf{y}^i| < i^{-1}$ and $|\mathbf{p}^{j_i} \cdot \mathbf{w}^i - \mathbf{p} \cdot \mathbf{y}^i| < i^{-1}$. Then,

$$G(\mathbf{v}^{j_i},\mathbf{p}^{j_i}) \geqq \mathbf{p}^{j_i} \cdot \mathbf{w}^i,$$

and

$$G(\mathbf{v},\mathbf{p}) = \lim_i \mathbf{p} \cdot \mathbf{y}^i \leqq \lim_j G(\mathbf{v}^j,\mathbf{p}^j).$$

If $\mathbf{p} \notin \mathbf{NY}(\mathbf{v})$, then $G(\mathbf{v},\mathbf{p}) = +\infty$, and (4) holds. If $\mathbf{p} \in \mathbf{NY}(\mathbf{v})$, then $G(\mathbf{v},\mathbf{p}) \leqq \liminf_{\substack{(\mathbf{v}^j,\mathbf{p}^j) \in \mathbf{D} \\ (\mathbf{v}^j,\mathbf{p}^j) \to (\mathbf{v},\mathbf{p})}} G(\mathbf{v}^j,\mathbf{p}^j)$.

Next suppose $(\mathbf{v}^j,\mathbf{p}^j) \in \mathbf{D}$, $(\mathbf{v}^j,\mathbf{p}^j) \to (\mathbf{v},\mathbf{p}) \in \mathbf{D}^0$. For some j_0, the set $\mathbf{Z} = \{\mathbf{p},\mathbf{p}^{j_0},\mathbf{p}^{j_0+1},\ldots\}$ is a compact subset of $\operatorname{int} \mathbf{NY}(\mathbf{v})$. By 15.2(3) and 13.5, $\mathbf{Z} \subseteq \operatorname{int} \mathbf{NY}(\mathbf{v}^j)$ and there exists a bounded sequence $\mathbf{y}^j \in \mathbf{\Phi}(\mathbf{v}^j,\mathbf{p}^j)$ for j large. Then, there exists a subsequence of j with $\mathbf{y}^j$ converging to $\mathbf{y} \in \mathbf{Y}(\mathbf{v})$ by upper hemicontinuity. Then, retaining the same notation for this subsequence, $\lim_j G(\mathbf{v}^j,\mathbf{p}^j) = \lim_j \mathbf{p}^j \cdot \mathbf{y}^j = \mathbf{p} \cdot \mathbf{y} \leqq G(\mathbf{v},\mathbf{p})$. Since the opposite inequality was shown to hold above, this verifies (2).

Finally, suppose $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v} \in \mathbf{V}$, $\mathbf{p} \in \mathbf{NY}(\mathbf{v})$. By the closedness and convexity of G in $\mathbf{p}$ for fixed $\mathbf{v}$, there exists a sequence $\mathbf{p}^i \in \operatorname{int} \mathbf{NY}(\mathbf{v})$ such that $\mathbf{p}^i \to \mathbf{p}$ and $G(\mathbf{v},\mathbf{p}) = \lim_i G(\mathbf{v},\mathbf{p}^i)$. By (2) proved above, for each i there exists j_i such that $\mathbf{p}^i \in \mathbf{NY}(\mathbf{v}^j)$ and $|G(\mathbf{v},\mathbf{p}^i) - G(\mathbf{v}^j,\mathbf{p}^i)| < i^{-1}$ for $j \geqq j_i$. Take $q^j = \mathbf{p}^i$ for $j_i \leqq j < j_{i+1}$. Then, $\lim_j G(\mathbf{v}^j,q^j) = G(\mathbf{v},\mathbf{p})$, and (3) holds. Q.E.D.

15.4. *Lemma.* If $\langle \mathbf{Y},\mathbf{V}\rangle$ is a strongly continuous, semi-bounded-valued correspondence, then the maximand correspondence $\langle \mathbf{\Phi},\mathbf{D}^0\rangle$ is upper hemicontinuous.

Proof: Suppose $(\mathbf{v}^j,\mathbf{p}^j) \in \mathbf{D}^0$, $(\mathbf{v}^j,\mathbf{p}^j) \to (\mathbf{v},\mathbf{p}) \in \mathbf{D}^0$, $\mathbf{y}^j \in \mathbf{\Phi}(\mathbf{v}^j,\mathbf{p}^j)$, $\mathbf{y}^j \to \mathbf{y}$. For any $\mathbf{w} \in \mathbf{Y}(\mathbf{v})$, by lower hemicontinuity there exist $\mathbf{w}^j \in \mathbf{Y}(\mathbf{v}^j)$, $\mathbf{w}^j \to \mathbf{w}$. Then, $\mathbf{p}^j \cdot \mathbf{w}^j \leqq \mathbf{p}^j \cdot \mathbf{y}^j$, implying in the limit that $\mathbf{p}\cdot\mathbf{w} \leqq \mathbf{p}\cdot\mathbf{y}$. Since $\mathbf{y} \in \mathbf{Y}(\mathbf{v})$ by upper hemicontinuity, this implies $\mathbf{y} \in \mathbf{\Phi}(\mathbf{v},\mathbf{p})$. Q.E.D.

15.5. *Lemma.* Suppose $\mathbf{V} \subseteq \mathbf{E}^M$ is non-empty, $\langle \mathbf{K},\mathbf{V}\rangle$ is a lower hemicontinuous correspondence mapping $\mathbf{v} \in \mathbf{V}$ into $\mathbf{K}(\mathbf{v}) \subseteq \mathbf{E}^N$, with $\mathbf{K}(\mathbf{v})$ a pointed convex cone and $\operatorname{int} \mathbf{K}(\mathbf{v}) \neq \emptyset$ for each $\mathbf{v} \in \mathbf{V}$. Define

$$\mathbf{D} = \{(\mathbf{v},\mathbf{p}) \in \mathbf{E}^{M+N} | \mathbf{v} \in \mathbf{V}, \mathbf{p} \in \mathbf{K}(\mathbf{v})\},$$

$$\mathbf{D}^0 = \{(\mathbf{v},\mathbf{p}) \in \mathbf{D} | \mathbf{p} \in \operatorname{int} \mathbf{K}(\mathbf{v})\},$$

and suppose $\langle \mathbf{F},\mathbf{D}\rangle$ is a closed real-valued function with $\langle F,\mathbf{D}^0\rangle$ continuous (for definitions, see 15.3) such that for each fixed $\mathbf{v} \in \mathbf{V}$, $\langle F,\mathbf{K}(\mathbf{v})\rangle$ is convex, closed, and positively linear homogeneous (as a function of $\mathbf{p}$). Then, the correspondence $\langle \mathbf{Y},\mathbf{V}\rangle$ defined by $\mathbf{Y}(\mathbf{v}) = \{\mathbf{Y} \in \mathbf{E}^N | \mathbf{p}\cdot\mathbf{y} \leqq F(\mathbf{v},\mathbf{p}) \text{ for all } \mathbf{p} \in \mathbf{K}(\mathbf{v})\}$ is strongly continuous, convex-valued, semi-bounded-valued, with $\operatorname{int} \mathbf{AY}(\mathbf{v}) \neq \emptyset$.

Proof: By 12.5, $\mathbf{Y}(\mathbf{v})$ is non-empty, closed, convex, and semi-bounded. By 11.6 and 10.9, $\operatorname{int} \mathbf{Y}(\mathbf{v}) \neq \emptyset$. Suppose $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v} \in \mathbf{V}$, $\mathbf{y}^j \in \mathbf{Y}(\mathbf{v}^j)$, $\mathbf{y}^j \to \mathbf{y}$. If $\mathbf{p} \in \operatorname{int} \mathbf{K}(\mathbf{v})$, then by lower hemicontinuity there exist $\mathbf{p}^j \in \mathbf{K}(\mathbf{v}^j)$ with $\mathbf{p}^j \to \mathbf{p}$, implying $\mathbf{p}^j \cdot \mathbf{y}^j \leqq F(\mathbf{v}^j,\mathbf{p}^j)$. Taking the limit, $\mathbf{p}\cdot\mathbf{y} \leqq F(\mathbf{v},\mathbf{p})$ by the continuity of $\langle F,\mathbf{D}^0\rangle$. For any $\mathbf{p} \in \mathbf{K}(\mathbf{v})$, there exist $\mathbf{p}^i \in \operatorname{int} \mathbf{K}(\mathbf{v})$, $\mathbf{p}^i \to \mathbf{p}$, and $F(\mathbf{v},\mathbf{p}^i) \to F(\mathbf{v},\mathbf{p})$ by the closedness of $\langle F,\mathbf{D}\rangle$. Hence, $\mathbf{p}^i \cdot \mathbf{y} \leqq F(\mathbf{v},\mathbf{p}^i)$ implies in the limit $\mathbf{p}\cdot\mathbf{y} \leqq F(\mathbf{v},\mathbf{p})$. Therefore, $\mathbf{y} \in \mathbf{Y}(\mathbf{v})$, implying $\langle \mathbf{Y},\mathbf{V}\rangle$ upper hemicontinuous.

Next suppose $\mathbf{v}^j \in \mathbf{V}$, $\mathbf{v}^j \to \mathbf{v} \in \mathbf{V}$, $\mathbf{y} \in \operatorname{int} \mathbf{Y}(\mathbf{v})$. Then, there exists $\alpha > 0$ such that $\mathbf{y} + \alpha\mathbf{p} \in \mathbf{Y}(\mathbf{v})$ for all $\mathbf{p} \in \mathbf{E}^N$ with $\mathbf{p}\cdot\mathbf{p} = 1$. Hence, $\mathbf{p}\cdot(\mathbf{y} + \alpha\mathbf{p}) \leqq F(\mathbf{v},\mathbf{p})$, or $\mathbf{p}\cdot\mathbf{y} \leqq F(\mathbf{v},\mathbf{p}) - \alpha$, for all $\mathbf{p} \in \mathbf{K}(\mathbf{v})$ with $\mathbf{p}\cdot\mathbf{p} = 1$. Suppose that for an infinite subsequence of j, $\mathbf{y} \notin \mathbf{Y}(\mathbf{v}^j)$. Then, there exists $\mathbf{p}^j \in \operatorname{int} \mathbf{K}(\mathbf{v}^j)$

such that $\mathbf{p}^j \cdot \mathbf{p}^j = 1$ and $\mathbf{p}^j \cdot \mathbf{y} > F(\mathbf{v}^j, \mathbf{p}^j)$. Extract a subsequence (retain notation) with $\mathbf{p}^j \to \mathbf{p}$. Then, the closedness of $\langle F, \mathbf{D} \rangle$ implies in the limit the inequality $\mathbf{p} \cdot \mathbf{y} \geqq F(\mathbf{v}, \mathbf{p})$. This condition also implies $\mathbf{p} \in \mathbf{K}(\mathbf{v})$. But then the inequality $\mathbf{p} \cdot \mathbf{y} \leqq F(\mathbf{v}, \mathbf{p}) - \alpha$ is contradicted. Therefore, $\mathbf{y} \in \operatorname{int} \mathbf{Y}(\mathbf{v})$ implies $\mathbf{y} \in \mathbf{Y}(\mathbf{v}^j)$ for j large. Finally, for any $\mathbf{y} \in \mathbf{Y}(\mathbf{v})$, there exist $\mathbf{y}^i \in \operatorname{int} \mathbf{Y}(\mathbf{v})$, $\mathbf{y}^i \to \mathbf{y}$. Then, there exists a subsequence j_i of j such that $\mathbf{y}^i \in \mathbf{Y}(\mathbf{v}^{j_i})$. Hence, $\langle \mathbf{Y}, \mathbf{V} \rangle$ is lower hemicontinuous, and therefore continuous. By 13.3, it is strongly continuous. Q.E.D.

15.6. *Lemma.* If $\langle \mathbf{Y}, \mathbf{V} \rangle$ is a lower hemicontinuous convex-valued correspondence, then there exists a continuous function $\mathbf{y}^*: \mathbf{V} \to \mathbf{E}^N$ such that $\mathbf{y}^*(\mathbf{v}) \in \operatorname{intr} \mathbf{Y}(\mathbf{v})$ for each $\mathbf{v} \in \mathbf{V}$.

Proof: $\langle \mathbf{Y}, \mathbf{V} \rangle$ lower hemicontinuous and convex implies $\langle \operatorname{intr} \mathbf{Y}, \mathbf{V} \rangle$ lower hemicontinuous. Then a theorem of Michael [see Parthasarathy (1971, Thm. 1.1)] establishes the existence of a continuous function $\mathbf{y}^*$ with $\mathbf{y}^*(\mathbf{v}) \in \mathbf{Y}(\mathbf{v})$. Q.E.D.

16. Differential Properties

In this section, we examine the relation between the curvature of the surface of a closed convex set $\mathbf{Y}$ and the curvature of its support function.

16.1. *Definition.* Consider a non-empty, closed, convex, semi-bounded set $\mathbf{Y} \subseteq \mathbf{E}^N$ and assume without loss (by translation, if necessary) $\mathbf{0} \in \mathbf{Y}$. Let $\mathbf{T}^{\mathbf{Y}}$ denote the cone spanned by $\mathbf{Y}$, and let $\mathbf{S}^{\mathbf{Y}} = \mathbf{N}\mathbf{Y}$. Define the real-valued function $\langle H^{\mathbf{Y}}, \mathbf{T}^{\mathbf{Y}} \rangle$ by $H^{\mathbf{Y}}(\mathbf{y}) = \inf\{\lambda > 0 | (1/\lambda) y \in \mathbf{Y}\}$ for $\mathbf{y} \in \mathbf{T}^{\mathbf{Y}}$. $\langle H^{\mathbf{Y}}, \mathbf{T}^{\mathbf{Y}} \rangle$ is termed the *gauge function* of $\mathbf{Y}$.

16.2. *Lemma.* If $\mathbf{Y} \subseteq \mathbf{E}^N$ is non-empty, convex, closed, and semi-bounded, and $\mathbf{0} \in \mathbf{Y}$, then (1) $\langle H^{\mathbf{Y}}, \mathbf{T}^{\mathbf{Y}} \rangle$ is convex, closed, and positively linear homogeneous; (2) for $\mathbf{y} \in \mathbf{T}^{\mathbf{Y}}$, $\lambda \geqq H^{\mathbf{Y}}(\mathbf{y}) > 0$ implies $\mathbf{y}/\lambda \in \mathbf{Y}$, $H^{\mathbf{Y}}(\mathbf{y}) > \lambda > 0$ implies $\mathbf{y}/\lambda \notin \mathbf{Y}$, and $H^{\mathbf{Y}}(\mathbf{y}) = 0$ implies $\mathbf{y} \in \mathbf{A}\mathbf{Y} \subseteq \mathbf{Y}$; (3) the sub-differential $\langle \mathbf{\Lambda}^{\mathbf{Y}}, \operatorname{intr} \mathbf{T}^{\mathbf{Y}} \rangle$ of $\langle H^{\mathbf{Y}}, \mathbf{T}^{\mathbf{Y}} \rangle$ is an upper hemicontinuous correspondence; and (4) if $\mathbf{0} \in \operatorname{intr} \mathbf{Y}$, then $\mathbf{T}^{\mathbf{Y}} = \operatorname{intr} \mathbf{T}^{\mathbf{Y}}$ is linear subspace.

Proof: It is an immediate consequence of the definition of $\langle H^{\mathbf{Y}}, \mathbf{T}^{\mathbf{Y}} \rangle$ that this function exists, is non-negative and positively linear homogeneous, and satisfies (2). For $\mathbf{y}$, $\mathbf{z} \in \mathbf{T}^{\mathbf{Y}}$, consider any $\lambda, \mu > 0$ such that $\mathbf{y}/\lambda \in \mathbf{Y}, \mathbf{z}/\mu \in \mathbf{Y}$, and let $\theta = \lambda/(\lambda + \mu)$. Then $\theta(\mathbf{y}/\lambda) + (1 - \theta)(\mathbf{z}/\mu) =$

$(\mathbf{y}+\mathbf{z})/(\lambda+\mu)\in\mathbf{Y}$ by convexity, implying $H^{\mathbf{Y}}(\mathbf{y}+\mathbf{z})\leqq\lambda+\mu$. Taking $\lambda\to H^{\mathbf{Y}}(\mathbf{y})$, $\mu\to H^{\mathbf{Y}}(\mathbf{z})$ establishes convexity of $H^{\mathbf{Y}}$. The following argument establishes $H^{\mathbf{Y}}$ closed. Suppose $\mathbf{y}^i\in\mathbf{T}^{\mathbf{Y}}$, $\mathbf{y}^i\to\mathbf{y}\neq\mathbf{0}$. For any $\epsilon>0$, there exist λ_i such that either $\mathbf{y}^i\in\mathbf{AY}$ and $\lambda_i=0$, or $\mathbf{y}^i\notin\mathbf{AY}$ and $\mathbf{y}^i/\lambda_i\notin\mathbf{Y}$, $\mathbf{y}^i/(\lambda_i+\epsilon)\in\mathbf{Y}$. If $\liminf\lambda_i=\lambda$ is finite, then $\mathbf{y}/(\lambda+\epsilon)\in\mathbf{Y}$, implying $H^{\mathbf{Y}}(\mathbf{y})\leqq\lambda+\epsilon\leqq\epsilon+\liminf H^{\mathbf{Y}}(\mathbf{y}^i)$. If $\mathbf{y}\notin\mathbf{T}^{\mathbf{Y}}$, then $+\infty=\liminf\lambda_i$, implying $\liminf H^{\mathbf{Y}}(\mathbf{y}^i)=+\infty$. At $\mathbf{y}=\mathbf{0}$, $H^{\mathbf{Y}}(\mathbf{0})=0\leqq H^{\mathbf{Y}}(\mathbf{y}^i)$. Hence, $H^{\mathbf{Y}}$ is closed, and (1) holds. Result (3) follows from 13.8, and (4) from the definition of the relative interior of a set. Q.E.D.

16.3. *Lemma.* The gauge function $\langle H^{\mathbf{Y}},\mathbf{T}^{\mathbf{Y}}\rangle$ and the support function $\langle G^{\mathbf{Y}},\mathbf{S}^{\mathbf{Y}}\rangle$ of a non-empty, closed, convex, semi-bounded set $\mathbf{Y}\subseteq\mathbf{E}^N$ with $\mathbf{0}\in\operatorname{intr}\mathbf{Y}$ are related by: (1) $\mathbf{p}\cdot\mathbf{y}\leqq G^{\mathbf{Y}}(\mathbf{p})H^{\mathbf{Y}}(\mathbf{y})$ for all $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}$, $\mathbf{y}\in\mathbf{T}^{\mathbf{Y}}$. If $\mathbf{p}\in\operatorname{intr}\mathbf{S}^{\mathbf{Y}}$ and $\mathbf{y}\in\operatorname{intr}\mathbf{T}^{\mathbf{Y}}$, then for $G^{\mathbf{Y}}(\mathbf{p})>0$, equality holds if and only if $\mathbf{p}/G^{\mathbf{Y}}(\mathbf{p})\in\mathbf{\Lambda}^{\mathbf{Y}}(\mathbf{y})$, and for $H^{\mathbf{Y}}(\mathbf{y})>0$, equality holds if and only if $\mathbf{y}/H^{\mathbf{Y}}(\mathbf{y})\in\mathbf{\Gamma}^{\mathbf{Y}}(\mathbf{p})$. (2) $G^{\mathbf{Y}}(\mathbf{p})=\inf\{\lambda>0|\mathbf{p}\cdot\mathbf{y}\leqq\lambda H^{\mathbf{Y}}(\mathbf{y})$ for all $\mathbf{y}\in\mathbf{T}^{\mathbf{Y}}\}$ for $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}$. (3) $H^{\mathbf{Y}}(\mathbf{y})=\inf\{\lambda>0|\mathbf{p}\cdot\mathbf{y}\leqq\lambda G^{\mathbf{Y}}(\mathbf{p})$ for all $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}\}$ for $\mathbf{y}\in\mathbf{T}^{\mathbf{Y}}$. (4) For $\mathbf{p}\in\operatorname{intr}\mathbf{S}^{\mathbf{Y}}$, $\mathbf{y}\in\operatorname{intr}\mathbf{T}^{\mathbf{Y}}$, $\mathbf{p}\in\mathbf{\Lambda}^{\mathbf{Y}}(\mathbf{y})$ if and only if $H^{\mathbf{Y}}(\mathbf{y})=\mathbf{p}\cdot\mathbf{y}$ and $G^{\mathbf{Y}}(\mathbf{p})=1$, and $\mathbf{y}\in\mathbf{\Gamma}^{\mathbf{Y}}(\mathbf{p})$ if and only if $G^{\mathbf{Y}}(\mathbf{p})=\mathbf{p}\cdot\mathbf{y}$ and $H^{\mathbf{Y}}(\mathbf{y})=1$.

Remark: $\langle H^{\mathbf{Y}},T^{\mathbf{Y}}\rangle$ and $\langle G^{\mathbf{Y}},\mathbf{S}^{\mathbf{Y}}\rangle$ are termed *polar reciprocal functions*, and the sets $\mathbf{Y}=\{\mathbf{y}\in\mathbf{T}^{\mathbf{Y}}|H^{\mathbf{Y}}(\mathbf{y})\leqq1\}$ and $\mathbf{P}=\{\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}|G^{\mathbf{Y}}(\mathbf{p})\leqq1\}$ are termed *polar reciprocal sets*.

Proof: If $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}$, $\{\lambda>0|\mathbf{p}\cdot\mathbf{y}\leqq\lambda H^{\mathbf{Y}}(\mathbf{y})$ for all $\mathbf{y}\in\mathbf{T}^{\mathbf{Y}}\}=\{\lambda>0|\mathbf{p}\cdot(\mathbf{y}/H^{\mathbf{Y}}(\mathbf{y}))\leqq\lambda$ for all $\mathbf{y}\in\mathbf{T}^{\mathbf{Y}}$ with $H^{\mathbf{Y}}(\mathbf{y})>0\}=\{\lambda>0|\mathbf{p}\cdot\mathbf{y}\leqq\lambda$ for all $\mathbf{y}\in\mathbf{Y}\}$, where the first equality follows from 16.2 (2) and the second equality follows from the definition of $H^{\mathbf{Y}}$ and $\mathbf{AY}=\mathbf{PS}^{\mathbf{Y}}$. Hence, (2) holds. Letting $\lambda\to G^{\mathbf{Y}}(\mathbf{p})$ in the condition $\mathbf{p}\cdot\mathbf{y}\leqq\lambda H^{\mathbf{Y}}(\mathbf{y})$ in (2) verifies the inequality in (1). To show (3), note that $H^{\mathbf{Y}}(\mathbf{y})\leqq1$ implies $\mathbf{y}\in\mathbf{Y}=\{\mathbf{v}\in\mathbf{T}^{\mathbf{Y}}|\mathbf{p}\cdot\mathbf{v}\leqq G^{\mathbf{Y}}(\mathbf{p})$ for all $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}\}$, and hence $\inf\{\lambda>0|\mathbf{p}\cdot\mathbf{y}\leqq\lambda G^{\mathbf{Y}}(\mathbf{p})$ for all $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}\}\leqq1$, and that $H^{\mathbf{Y}}(\mathbf{y})>1$ implies $\mathbf{p}\cdot\mathbf{y}>G^{\mathbf{Y}}(\mathbf{p})$ for some $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}$, and hence $\inf\{\lambda>0|\mathbf{p}\cdot\mathbf{y}\leqq\lambda G^{\mathbf{Y}}(\mathbf{p})$ for all $\mathbf{p}\in\mathbf{S}^{\mathbf{Y}}\}>1$. Then, homogeneity implies (3). Next, the "if and only if" conditions in (1) will be verified.

(a) Assume $G^{\mathbf{Y}}(\mathbf{p})>0$. If $\mathbf{p}\in G^{\mathbf{Y}}(\mathbf{p})\mathbf{\Lambda}^{\mathbf{Y}}(\mathbf{y})$, then $\mathbf{p}/G^{\mathbf{Y}}(\mathbf{p})\in\mathbf{\Lambda}^{\mathbf{Y}}(\mathbf{y})$ implies $\mathbf{p}\cdot\mathbf{y}/G^{\mathbf{Y}}(\mathbf{p})=H^{\mathbf{Y}}(\mathbf{y})$ by 13.8 (1). Conversely, if $\mathbf{p}\cdot\mathbf{y}=G^{\mathbf{Y}}(\mathbf{p})H^{\mathbf{Y}}(\mathbf{y})$, then $\mathbf{y}\cdot\mathbf{p}/G^{\mathbf{Y}}(\mathbf{p})=H^{\mathbf{Y}}(\mathbf{y})$ and $\mathbf{z}\cdot\mathbf{p}/G^{\mathbf{Y}}(\mathbf{p})\leqq H^{\mathbf{Y}}(\mathbf{z})$ for all $\mathbf{z}\in\mathbf{T}^{\mathbf{Y}}$, implying $\mathbf{p}/G^{\mathbf{Y}}(\mathbf{p})\in\mathbf{\Lambda}^{\mathbf{Y}}(\mathbf{y})$ by 13.8 (1). (b) Assume $H^{\mathbf{Y}}(\mathbf{y})>0$. By an argument symmetric to that in (a), $\mathbf{y}/H^{\mathbf{Y}}(\mathbf{y})\in\mathbf{\Gamma}^{\mathbf{Y}}(\mathbf{p})$ if and only if $\mathbf{p}\cdot\mathbf{y}=G^{\mathbf{Y}}(\mathbf{p})H^{\mathbf{Y}}(\mathbf{y})$.

Result (4) follows from (1) and 13.8. Q.E.D.

16.4. *Definition.* A convex, closed, positively linear homogeneous function $\langle F,\mathbf{X}\rangle$ is *exposed* at $\mathbf{x}\in \text{intr}\,\mathbf{X}$ if $(\mathbf{x},F(\mathbf{x}))$ is a point in an exposed ray in the set $\mathbf{A}=\{(\mathbf{x},\xi)\in \mathbf{E}^{N+1}|\mathbf{x}\in\mathbf{X},\ \ \xi\geqq F(\mathbf{x})\}$, or, equivalently, there exists $\mathbf{y}\in\mathbf{\Gamma}(\mathbf{x})$, where $\langle\mathbf{\Gamma},\text{intr}\,\mathbf{X}\rangle$ is the sub-differential of $\langle F,\mathbf{X}\rangle$, such that $F(\mathbf{z})-F(\mathbf{x})>\mathbf{y}\cdot(\mathbf{z}-\mathbf{x})$ for all $\mathbf{z}\in\mathbf{X}$, $\mathbf{z}$ not proportional to $\mathbf{x}$. $\langle F,\mathbf{X}\rangle$ is *strictly quasiconvex* at $\mathbf{x}\in\text{intr}\,\mathbf{X}$ if $F(\theta\mathbf{x}+(1-\theta)\mathbf{z})<\theta F(\mathbf{x})+(1-\theta)F(\mathbf{z})$ for $0<\theta<1$ and $\mathbf{z}\in\mathbf{X},\mathbf{z}$ not proportional to $\mathbf{x}$. When $\text{int}\,\mathbf{X}\neq\emptyset$, $\langle F,\mathbf{X}\rangle$ is *strictly differentially quasiconvex* at $\mathbf{x}\in\text{int}\,\mathbf{X}$ if $\langle F,\mathbf{X}\rangle$ has a first and second differential at $\mathbf{x}$ and the quadratic form $Q(\mathbf{v},\mathbf{F}''(\mathbf{x}))$ in the Hessian matrix $\mathbf{F}''(\mathbf{x})$ is positive for $\mathbf{v}$ not proportional to $\mathbf{x}$ (i.e., $\mathbf{F}''(\mathbf{x})$ is non-negative definite and of rank $N-1$).

16.5. *Lemma.* Consider a convex, closed, positively linear homogeneous function $\langle F,\mathbf{X}\rangle$ with sub-differential $\langle\mathbf{\Gamma},\text{intr}\,\mathbf{X}\rangle$. (1) For $\mathbf{x}\in\text{intr}\,\mathbf{X}$, $F(\mathbf{z})-F(\mathbf{x})>\mathbf{y}\cdot(\mathbf{z}-\mathbf{x})$ for all $\mathbf{y}\in\text{intr}\,\mathbf{\Gamma}(\mathbf{x})$, $\mathbf{z}\in\mathbf{X}$, $\mathbf{z}$ not proportional to $\mathbf{x}$, if and only if $\langle F,\mathbf{X}\rangle$ is exposed at $\mathbf{x}$. (2) For $\mathbf{x}\in\text{intr}\,\mathbf{X}$, $F(\mathbf{z})-F(\mathbf{x})>\mathbf{y}\cdot(\mathbf{z}-\mathbf{x})$ for all $\mathbf{y}\in\mathbf{\Gamma}(\mathbf{x})$, $\mathbf{z}\in\mathbf{X}$, $\mathbf{z}$ not proportional to $\mathbf{x}$, if and only if $\langle F,\mathbf{X}\rangle$ is strictly quasiconvex at $\mathbf{x}$. For $\mathbf{x}\in\text{intr}\,\mathbf{X}$, $\langle F,\mathbf{X}\rangle$ strictly quasiconvex at $\mathbf{x}$ implies $\langle F,\mathbf{X}\rangle$ exposed at $\mathbf{x}$. (3) If $\text{int}\,\mathbf{X}\neq\emptyset$ and $\langle F,\mathbf{X}\rangle$ is strictly differentially quasiconvex at $\mathbf{x}\in\text{int}\,\mathbf{X}$, then $\langle F,\mathbf{X}\rangle$ is strictly quasiconvex at $\mathbf{x}$. (4) If $\text{int}\,\mathbf{X}\neq\emptyset$, $\langle F,\mathbf{X}\rangle$ possesses continuous first and second differentials in a neighborhood of $\mathbf{x}\in\text{int}\,\mathbf{X}$, and $\langle F,\mathbf{X}\rangle$ is strictly quasiconvex at $\mathbf{x}$, then for any neighborhood $\mathbf{Z}$ of $\mathbf{x}$, there exists a neighborhood $\mathbf{W}$ contained in $\mathbf{Z}$ such that $\langle F,\mathbf{X}\rangle$ is strictly differentially quasiconvex on $\mathbf{W}$.

Proof: (1) The "only if" condition follows from the definition of $\langle F,\mathbf{X}\rangle$ exposed at $\mathbf{x}$. To show the "if" condition, suppose $\langle F,\mathbf{X}\rangle$ exposed at $\mathbf{x}$. Then, $F(\mathbf{z})-F(\mathbf{x})>\mathbf{y}\cdot(\mathbf{z}-\mathbf{x})$ for $\mathbf{z}\in\mathbf{X}$, $\mathbf{z}$ not proportional to $\mathbf{x}$, and some $\mathbf{y}\in\mathbf{\Gamma}(\mathbf{x})$. Suppose that for some $\mathbf{v}\in\text{intr}\,\mathbf{\Gamma}(\mathbf{x})$ and $\mathbf{z}\in\mathbf{X}$, $\mathbf{z}$ not proportional to $\mathbf{x}$, $F(\mathbf{z})-F(\mathbf{x})=\mathbf{v}\cdot(\mathbf{z}-\mathbf{x})$. Then, $(1+\theta)\mathbf{v}-\theta\mathbf{y}\in\mathbf{\Gamma}(\mathbf{x})$ for θ small positive, implying $F(\mathbf{z})-F(\mathbf{x})<((1+\theta)\mathbf{v}-\theta\mathbf{y})\cdot(\mathbf{z}-\mathbf{x})$ and contradicting the definition of $\mathbf{\Gamma}$.

(2) If $\langle F,\mathbf{X}\rangle$ is strictly quasiconvex at $\mathbf{x}$, but there exists $\mathbf{y}\in\mathbf{\Gamma}(\mathbf{x})$, $\mathbf{z}\in\mathbf{X}$, $\mathbf{z}$ not proportional to $\mathbf{x}$ such that $F(\mathbf{z})-F(\mathbf{x})=\mathbf{y}\cdot(\mathbf{z}-\mathbf{x})$, then $F((\mathbf{z}+\mathbf{x})/2)-F(\mathbf{x})<(F(\mathbf{z})-F(\mathbf{x}))/2=\mathbf{y}\cdot((\mathbf{z}-\mathbf{x})/2)$, contradicting the definition of $\mathbf{\Gamma}$. Hence, the "if" implication in (2) holds, and (1) then implies the last result in (2). Next, the "only if" implication will be established. Suppose $\langle F,\mathbf{X}\rangle$ is not strictly quasiconvex at $\mathbf{x}\in\text{intr}\,\mathbf{X}$. Then, there exists $\mathbf{z}\in\mathbf{X},\mathbf{z}$ not proportional to $\mathbf{x}$, $\alpha\in(0,1)$ such that $F(\alpha\mathbf{z}+(1-\alpha)\mathbf{x})=$

$\alpha F(\mathbf{z})+(1-\alpha)F(\mathbf{x})$. From the proof of 13.8, paragraph (b), we have $(F(\theta\mathbf{u}+(1-\theta)\mathbf{v})-F(\mathbf{v}))/\theta$ non-decreasing in $\theta>0$ for $\mathbf{u},\mathbf{v}\in\mathbf{X}$. For $\gamma\in(0,1)$, $F(\gamma\mathbf{z}+(1-\gamma)\mathbf{x})\leqq\gamma F(\mathbf{z})+(1-\gamma)F(\mathbf{x})$ by convexity. If $\alpha<\gamma<1$ and $\theta=\gamma/\alpha>1$, then

$$\frac{F(\theta(\alpha\mathbf{z}+(1-\alpha)\mathbf{x})+(1-\theta)\mathbf{x})-F(\mathbf{x})}{\theta}\geqq F(\alpha\mathbf{z}+(1-\alpha)\mathbf{x})-F(\mathbf{x}) = \alpha[F(\mathbf{z})-F(\mathbf{x})],$$

or

$$F(\gamma\mathbf{z}+(1-\gamma)\mathbf{x})\geqq\gamma F(\mathbf{z})+(1-\gamma)F(\mathbf{x}).$$

If $0<\gamma<\alpha$ and $\theta=(1-\gamma)/(1-\alpha)$, then

$$\frac{F(\theta(\alpha\mathbf{z}+(1-\alpha)\mathbf{x})+(1-\theta)\mathbf{z})-F(\mathbf{z})}{\theta}\geqq F(\alpha\mathbf{z}+(1-\alpha)\mathbf{x})-F(\mathbf{x}) = (1-\alpha)[F(\mathbf{x})-F(\mathbf{z})],$$

or

$$F(\gamma\mathbf{z}+(1-\gamma)\mathbf{x})\geqq\gamma F(\mathbf{z})+(1-\gamma)F(\mathbf{x}).$$

Hence,

$$F(\gamma\mathbf{z}+(1-\gamma)\mathbf{x})=\gamma F(\mathbf{z})+(1-\gamma)F(\mathbf{x}),$$

for $\gamma\in(0,1)$.

Set $\mathbf{A}=\{(\mathbf{x},\xi)\in\mathbf{E}^{N+1}|\mathbf{x}\in\mathbf{X},\ \xi\geqq F(\mathbf{x})\}$ and the flat $\mathbf{B}=\{(\theta\mathbf{z}+(1-\theta)\mathbf{x},\ \theta F(\mathbf{z})+(1-\theta)F(\mathbf{x}))|\theta \text{ real}\}$. Then, $\mathbf{A}$ is convex and $\mathbf{B}\cap\operatorname{intr}\mathbf{A}=\emptyset$, implying by 10.14 the existence of a hyperplane $\mathbf{H}((\mathbf{y},-\eta),\alpha)$ with $\mathbf{B}\subseteq\mathbf{H}((\mathbf{y},-\eta),\alpha)$, $\mathbf{A}\subseteq\mathbf{H}^{-}((\mathbf{y},-\eta),\alpha)$, and $\operatorname{intr}\mathbf{A}\cap\mathbf{H}((\mathbf{y},-\eta),\alpha)=\emptyset$. Using the argument of paragraph (a) of the proof of 13.8, we can normalize $\eta=1$ and obtain the implication $\mathbf{y}\in\mathbf{\Gamma}(\mathbf{x})$. But $(\mathbf{z},F(\mathbf{z}))\in\mathbf{B}$ implies $F(\mathbf{z})-F(\mathbf{x})=\mathbf{y}\cdot(\mathbf{z}-\mathbf{x})$, contradicting the hypothesis in (2).

(3) As in the proof of (2), if $\langle F,\mathbf{X}\rangle$ is not strictly quasiconvex, then there exists $\mathbf{z}\in\mathbf{X}$, $\mathbf{z}$ not proportional to $\mathbf{x}$, such that $F(\theta\mathbf{z}+(1-\theta)\mathbf{x})=\theta F(\mathbf{z})+(1-\theta)F(\mathbf{x})$ for all $\theta\in(0,1)$. Letting $\mathbf{v}=\theta\mathbf{z}+(1-\theta)\mathbf{x}$, F has a second-order expansion given in 12.1, $F(\mathbf{v})=F(\mathbf{x})+\mathbf{y}\cdot(\mathbf{v}-\mathbf{x})+\frac{1}{2}\theta^2Q(\mathbf{z}-\mathbf{x},\mathbf{F}''(\mathbf{x}))+i(\theta^2|\mathbf{z}-\mathbf{x}|^2)$, where $\mathbf{y}=\mathbf{F}'(\mathbf{x})$ and $\mathbf{\Gamma}(\mathbf{x})=\{\mathbf{y}\}$. For $\theta>0$ small, $\frac{1}{2}\theta^2Q(\mathbf{z}-\mathbf{x},\mathbf{F}''(\mathbf{x}))+i(\theta^2|\mathbf{z}-\mathbf{x}|^2)>0$, implying $F(\mathbf{v})-F(\mathbf{x})>\mathbf{y}\cdot(\mathbf{v}-\mathbf{x})$. But this contradicts the supposition, and $\langle F,\mathbf{X}\rangle$ is strictly quasiconvex at $\mathbf{x}$.

(4) Consider any neighborhood $\mathbf{U}$ of $\mathbf{x},\mathbf{U}\subseteq\mathbf{X}$, in which F is twice continuously differentiable. Define $\tilde{\mathbf{x}}=(x_1,\ldots,x_{N-1})$ and $\tilde{F}(\tilde{\mathbf{z}})=F(\tilde{\mathbf{z}},x_N)$

for $\tilde{\mathbf{z}} \in \mathbf{E}^{N-1}$ such that $(\tilde{\mathbf{z}}, x_N) \in \mathbf{U}$. Since F is strictly quasiconvex at $\mathbf{x}$, $\tilde{F}(\tilde{\mathbf{x}}+\tilde{\mathbf{y}}) - F(\tilde{\mathbf{x}}) - \tilde{\mathbf{F}}'(\tilde{\mathbf{x}})\tilde{\mathbf{y}} > 0$ for $(\tilde{\mathbf{x}}+\tilde{\mathbf{y}}, x_N) \in \mathbf{U}$, $\tilde{\mathbf{y}} \neq 0$. By continuity, there exists a neighborhood $\tilde{\mathbf{U}}$ of $\tilde{\mathbf{x}}$ and $\alpha > 0$ such that $|\tilde{\mathbf{y}}| \leqq \alpha$ and $\tilde{\mathbf{z}} \in \tilde{\mathbf{U}}$ implies $(\tilde{\mathbf{z}}+\tilde{\mathbf{y}}, x_N) \in \mathbf{U}$ and $\tilde{F}(\tilde{\mathbf{z}}+\tilde{\mathbf{y}}) - \tilde{F}(\tilde{\mathbf{z}}) - \tilde{\mathbf{F}}'(\tilde{\mathbf{z}})\tilde{\mathbf{y}} > 0$ for $\tilde{\mathbf{y}} \neq \mathbf{0}$. Then, $\tilde{F}$ is strictly convex on $\tilde{\mathbf{U}}$, and a theorem of Bernstein and Toupin (1962) establishes that the hessian $\tilde{\mathbf{F}}''$ is positive definite on an open dense subset of $\tilde{\mathbf{U}}$. Let $\tilde{\mathbf{W}} \subseteq \tilde{\mathbf{U}}$ be a neighborhood on which $\tilde{\mathbf{F}}''$ is positive definite, and define $\mathbf{W} = \{(\tilde{\mathbf{y}}, y_N) \in \mathbf{U} | (x_N/y_N)\tilde{\mathbf{y}} \in \tilde{\mathbf{W}}\}$. The hessian matrix of F on $\mathbf{W}$ is

$$\left[\begin{array}{c|c} \dfrac{x_N}{y_N}\tilde{\mathbf{F}}''\left(\dfrac{x_N}{y_N}\tilde{\mathbf{y}}\right) & -\dfrac{x_N}{y_N^2}\tilde{\mathbf{F}}''\left(\dfrac{x_N}{y_N}\tilde{\mathbf{y}}\right)\tilde{\mathbf{y}} \\ \hline -\dfrac{x_N}{y_N^2}\tilde{\mathbf{y}}'\tilde{\mathbf{F}}''\left(\dfrac{x_N}{y_N}\tilde{\mathbf{y}}\right) & \dfrac{x_N}{y_N^3}\tilde{\mathbf{y}}'\tilde{\mathbf{F}}''\left(\dfrac{x_N}{y_N}\tilde{\mathbf{y}}\right)\tilde{\mathbf{y}} \end{array}\right]$$

By construction, this matrix is of rank $N-1$. Q.E.D.

16.6. *Definition*. Suppose $\mathbf{Y} \subseteq \mathbf{E}^N$ is closed, convex, and semi-bounded, with $\mathbf{0} \in \operatorname{int} \mathbf{Y}$, and define $\mathbf{Y}^* = \{\mathbf{y} \in \mathbf{E}^N | \mathbf{y} \in \mathbf{\Phi}(\mathbf{p})$ for some $\mathbf{p} \in \mathbf{S}\}$, where $\mathbf{S} = \operatorname{int} \mathbf{NY}$ and $\langle \mathbf{\Phi}, \mathbf{S}\rangle$ is the maximand correspondence of $\mathbf{Y}$. Note that $\mathbf{0} \in \operatorname{int} \mathbf{Y}$ implies $\mathbf{p}\cdot\mathbf{y} = G(\mathbf{p}) > 0$ for $\mathbf{p} \in \mathbf{S}$, $\mathbf{y} \in \mathbf{\Phi}(\mathbf{p})$, and hence, by 16.3 (1), $H(\mathbf{y}) = 1$.

16.7. *Lemma*. Suppose $\mathbf{Y} \subseteq \mathbf{E}^N$ is closed, convex, and semi-bounded, with $\mathbf{0} \in \operatorname{int} \mathbf{Y}$. Let $\langle G, \mathbf{NY}\rangle$ denote the support function of $\mathbf{Y}$, and $\langle \mathbf{\Gamma}, \mathbf{S}\rangle$ denote its sub-differential. Let $\langle H, \mathbf{E}^N\rangle$ denote the gauge function of $\mathbf{Y}$, and $\langle \mathbf{\Lambda}, \mathbf{E}^N\rangle$ denote its sub-differential. Then, the following conditions hold: (1) If $\mathbf{y} \in \mathbf{E}^N$, then $\mathbf{\Lambda}(\mathbf{y}) \subseteq \mathbf{NY}$. If $\mathbf{y} \in \mathbf{Y}^*$, then $\operatorname{intr} \mathbf{\Lambda}(\mathbf{y}) \subseteq \mathbf{S}$. If $\mathbf{y} \in \mathbf{Y}^*$ and $\langle H, \mathbf{E}^N\rangle$ is strictly quasiconvex at $\mathbf{y}$, then $\mathbf{\Lambda}(\mathbf{y}) \subseteq \mathbf{S}$. (2) If $\mathbf{p} \in \mathbf{S}$, then $\mathbf{\Gamma}(\mathbf{p}) \subseteq \mathbf{Y}^*$. (3) $\langle H, \mathbf{E}^N\rangle$ differentiable at $\mathbf{y} \in \mathbf{Y}^*$ implies $\langle G, \mathbf{NY}\rangle$ exposed at $\mathbf{p} \in \mathbf{\Lambda}(\mathbf{y}) \subseteq \mathbf{S}$. (4) $\langle G, \mathbf{NY}\rangle$ differentiable at $\mathbf{p} \in \mathbf{S}$ implies $\langle H, \mathbf{E}^N\rangle$ exposed at $\mathbf{y} \in \mathbf{\Gamma}(\mathbf{p})$. (5) $\langle H, \mathbf{E}^N\rangle$ exposed at $\mathbf{y} \in \mathbf{Y}^*$ implies $\langle G, \mathbf{NY}\rangle$ differentiable at $\mathbf{p} \in \operatorname{intr} \mathbf{\Lambda}(\mathbf{y})$. (6) $\langle G, \mathbf{NY}\rangle$ exposed at $\mathbf{p} \in \mathbf{S}$ implies $\langle H, \mathbf{E}^N\rangle$ differentiable at $\mathbf{y} \in \operatorname{intr} \mathbf{\Gamma}(\mathbf{p})$. (7) $\langle H, \mathbf{E}^N\rangle$ strictly quasiconvex at $\mathbf{y} \in \mathbf{Y}^*$ implies $\langle G, \mathbf{NY}\rangle$ differentiable at $\mathbf{p} \in \mathbf{\Lambda}(\mathbf{y})$. (8) $\langle G, \mathbf{NY}\rangle$ strictly quasiconvex at $\mathbf{p} \in \mathbf{S}$ implies $\langle H, \mathbf{E}^N\rangle$ differentiable at $\mathbf{y} \in \mathbf{\Gamma}(\mathbf{p})$. (9) $\langle H, \mathbf{E}^N\rangle$ differentiable at all $\mathbf{y} \in \mathbf{\Gamma}(\mathbf{p})$ for some $\mathbf{p} \in \mathbf{S}$ implies $\langle G, \mathbf{NY}\rangle$ strictly quasiconvex at $\mathbf{p}$. (10) $\langle G, \mathbf{NY}\rangle$ differentiable at all $\mathbf{p} \in \mathbf{\Lambda}(\mathbf{y})$ for some $\mathbf{y} \in \mathbf{Y}^*$ implies $\langle H, \mathbf{E}^N\rangle$ strictly quasiconvex at $\mathbf{y}$. (11) If $\langle H, \mathbf{E}^N\rangle$ possesses a continuous first and second differential in a neighborhood of $\mathbf{y} \in \mathbf{Y}^*$, and is strictly differentially quasiconvex at $\mathbf{y}$, then $\langle G,\mathbf{NY}\rangle$ possesses a continuous first and second differential in a neighborhood of $\mathbf{p} \in \Lambda(\mathbf{y})$, and is strictly differentially quasiconvex at $\mathbf{p}$. (12) If $\langle G,\mathbf{NY}\rangle$ possesses a continuous first and second differential in a neighborhood of $\mathbf{p} \in \mathbf{S}$, and is strictly differentially quasiconvex at $\mathbf{p}$, then $\langle H,\mathbf{E}^N\rangle$ possesses a continuous first and second differential in a neighborhood of $\mathbf{y} \in \Gamma(\mathbf{p})$, and is strictly differentially quasiconvex at $\mathbf{y}$.

Proof: (1) By 13.8, $\mathbf{p} \in \Lambda(\mathbf{y})$ implies $\mathbf{p}\cdot\mathbf{y} = H(\mathbf{y})$ and $\mathbf{p}\cdot\mathbf{z} \leqq H(\mathbf{z})$ for all $\mathbf{z} \in \mathbf{E}^N$. Hence, $G(\mathbf{p}) = \sup\{\mathbf{p}\cdot\mathbf{z}|\mathbf{z} \in \mathbf{Y}\} = 1$ and $\mathbf{p} \in \mathbf{NY}$. If $\mathbf{y} \in \mathbf{Y}^*$, then $\mathbf{y} \in \Gamma(\mathbf{q})$ for some $\mathbf{q} \in \mathbf{S}$ by 13.9. Since $H(\mathbf{y}) = 1$, we can scale $\mathbf{q}$ so that $\mathbf{q}\cdot\mathbf{y} = G(\mathbf{q}) = 1$, and hence $\mathbf{q} \in \Lambda(\mathbf{y})$, by 16.3 (1). If there exists $\mathbf{p} \in \operatorname{intr} \Lambda(\mathbf{y})$, $\mathbf{p} \notin \mathbf{S}$, then there exists $\mathbf{z} \in \mathbf{AY}$, $\mathbf{z} \neq \mathbf{0}$, such that $\mathbf{p}\cdot\mathbf{z} = 0$. Further, $(1+\theta)\mathbf{p} - \theta\mathbf{q} \in \Lambda(\mathbf{y})$ for θ small positive and $((1+\theta)\mathbf{p} - \theta\mathbf{q})\cdot\mathbf{z} > 0$, contradicting $\Lambda(\mathbf{y}) \subseteq \mathbf{NY}$. Hence, $\operatorname{intr} \Lambda(\mathbf{y}) \subseteq \mathbf{S}$. If H is strictly quasiconvex at $\mathbf{y} \in \mathbf{Y}^*$, then by 16.5 (2), $H(\mathbf{z}) - H(\mathbf{y}) > \mathbf{p}\cdot(\mathbf{z}-\mathbf{y})$ for $\mathbf{z} \in \mathbf{E}^N$, $\mathbf{z}$ not proportional to $\mathbf{y}$, $\mathbf{p} \in \Lambda(\mathbf{y})$. Taking $\mathbf{z} = \mathbf{y} + \mathbf{v}$ with $\mathbf{v} \in \mathbf{AY}$, $\mathbf{V} \neq \mathbf{0}$ implies $\mathbf{z} \in \mathbf{Y}$ by 10.9. For some $\mathbf{q} \in \mathbf{S}$, $\mathbf{q}\cdot\mathbf{y} = G(\mathbf{y}) > 0$, implying $\mathbf{v}$, and hence $\mathbf{z}$, not proportional to $\mathbf{y}$. Therefore, $0 \geqq H(\mathbf{z}) - H(\mathbf{y}) > \mathbf{p}\cdot\mathbf{v}$. But $\mathbf{p}\cdot\mathbf{v} < 0$ for all $\mathbf{v} \in \mathbf{AY}$, $\mathbf{v} \neq \mathbf{0}$, implies $\mathbf{p} \in \mathbf{S}$.

(2) If $\mathbf{p} \in \mathbf{S}$, then $\Gamma(\mathbf{p}) = \Phi(\mathbf{p}) \subseteq \mathbf{Y}^*$ by 13.9.

(3) $\langle H,\mathbf{E}^N\rangle$ differentiable at $\mathbf{y} \in \mathbf{Y}^*$ implies $\operatorname{intr} \Lambda(\mathbf{y}) = \Lambda(\mathbf{y}) = \{\mathbf{p}\} \subseteq \mathbf{S}$. For $\mathbf{q} \in \mathbf{S}$, $\mathbf{q}$ not proportional to $\mathbf{p}$, 16.3(1) implies $\mathbf{q}\cdot\mathbf{y} < G(\mathbf{q})H(\mathbf{y})$, and hence $(\mathbf{q}-\mathbf{p})\cdot\mathbf{y}/H(\mathbf{y}) < G(\mathbf{q}) - G(\mathbf{p})$. But this is the condition for $\langle G,\mathbf{NY}\rangle$ to be exposed at $\mathbf{p}$.

(4) $\langle G,\mathbf{NY}\rangle$ differentiable at $\mathbf{p} \in \mathbf{S}$ implies $\Gamma(\mathbf{p}) = \{\mathbf{y}\} \subseteq \mathbf{Y}^*$ and $\mathbf{p}\cdot\mathbf{z} < G(\mathbf{p})H(\mathbf{z})$ for $\mathbf{z} \in \mathbf{E}^N$, $\mathbf{z}$ not proportional to $\mathbf{y}$. Hence, $(\mathbf{z}-\mathbf{y})\cdot\mathbf{p}/G(\mathbf{p}) < H(\mathbf{z}) - H(\mathbf{y})$, and $\langle H,\mathbf{E}^N\rangle$ is exposed at $\mathbf{y}$.

(5) $\langle H,\mathbf{E}^N\rangle$ exposed at $\mathbf{y} \in \mathbf{Y}^*$ implies that for $\mathbf{p} \in \operatorname{intr} \Lambda(\mathbf{y})$, $H(\mathbf{z}) - H(\mathbf{y}) > \mathbf{p}\cdot(\mathbf{z}-\mathbf{y})$ for $\mathbf{z}$ not proportional to $\mathbf{y}$. Then, $\mathbf{p}\cdot\mathbf{y} = H(\mathbf{y}) = 1$ and $\mathbf{p}\cdot\mathbf{z} < 1$ for $\mathbf{z} \in \mathbf{Y}$, $\mathbf{z} \neq \mathbf{y}$, implying $\Gamma(\mathbf{p}) = \Phi(\mathbf{p}) = \{\mathbf{y}\}$. Hence, $\langle G,\mathbf{NY}\rangle$ is differentiable at $\mathbf{p}$.

(6) $\langle G,\mathbf{NY}\rangle$ exposed at $\mathbf{p} \in \mathbf{S}$ implies that $G(\mathbf{q}) - G(\mathbf{p}) > \mathbf{y}\cdot(\mathbf{q}-\mathbf{p})$ for $\mathbf{q} \in \mathbf{NY}$, $\mathbf{q}$ not proportional to $\mathbf{p}$, $\mathbf{y} \in \operatorname{intr} \Gamma(\mathbf{p})$, and hence $\mathbf{y} \in \mathbf{Y}^*$ and $G(\mathbf{q}) > \mathbf{y}\cdot\mathbf{q}$. If $\mathbf{q} \in \Lambda(\mathbf{y})$, then $H(\mathbf{z}) - H(\mathbf{y}) \geqq \mathbf{q}\cdot(\mathbf{z}-\mathbf{y})$, implying $G(\mathbf{q}) = \mathbf{q}\cdot\mathbf{y} \geqq \mathbf{q}\cdot\mathbf{z}$, for all $\mathbf{z} \in \mathbf{Y}$. Then, $\mathbf{q}$ must be proportional to $\mathbf{p}$, and since $H(\mathbf{y}) = \mathbf{q}\cdot\mathbf{y}$, $\mathbf{q} = \mathbf{p}/G(\mathbf{p})$. Hence, $\Lambda(\mathbf{y}) = \{\mathbf{p}/G(\mathbf{p})\}$ and $\langle H,\mathbf{E}^N\rangle$ is differentiable at $\mathbf{y}$.

(7) $\langle H,\mathbf{E}^N\rangle$ strictly quasiconvex at $\mathbf{y}\in\mathbf{Y}^*$ implies, by 16.5(2) and (1) above, that $\Lambda(\mathbf{y})\subseteq\mathbf{S}$ and the proof of (5) above holds for every $\mathbf{p}\in\Lambda(\mathbf{y})$.

(8) By 16.5 (2), the proof of (6) above holds for every $\mathbf{y}\in\Gamma(\mathbf{p})$.

(9) $\langle H,\mathbf{E}^N\rangle$ differentiable at $\mathbf{y}\in\Gamma(\mathbf{p})$, $\mathbf{p}\in\mathbf{S}$ implies, by 16.3(4), that intr $\Lambda(\mathbf{y})=\Lambda(\mathbf{y})=\{\mathbf{p}/G(\mathbf{p})\}\subseteq\mathbf{S}$. Then, $\mathbf{q}\in\mathbf{S}$, $\mathbf{q}$ not proportional to $\mathbf{p}$, and $\mathbf{y}\in\Gamma(\mathbf{p})$ implies $H(\mathbf{y})=1$ and $\mathbf{q}\cdot\mathbf{y}<G(\mathbf{q})$. Hence, $(\mathbf{q}-\mathbf{p})\cdot\mathbf{y}<G(\mathbf{q})-G(\mathbf{p})$ for all $\mathbf{y}\in\Gamma(\mathbf{p})$, and by 16.5(2), $\langle G,\mathbf{NY}\rangle$ is strictly quasiconvex at $\mathbf{p}$.

(10) The proof is the same as that for (9), with the appropriate interchange in notation.

(11) If $\Lambda(\mathbf{y})=\{\mathbf{p}\}$, then $G(\mathbf{p})=\mathbf{p}\cdot\mathbf{y}=1=\mathrm{Max}\{\mathbf{p}\cdot\mathbf{z}|H(\mathbf{z})\leqq 1\}$. Since $H(\mathbf{0})=0<1$, the Kuhn–Tucker theorem [Karlin (1959, Thm. 7.1.1)] can be applied to establish the existence of $\lambda\geqq 0$ such that $\mathbf{p}\cdot\mathbf{z}+\lambda(1-H(\mathbf{z}))\leqq\mathbf{p}\cdot\mathbf{y}$ for all $\mathbf{z}\in\mathbf{E}^N$. Using differentiability, this implies $\mathbf{p}=\lambda\mathbf{H}'(\mathbf{y})$ and $1=\mathbf{p}\cdot\mathbf{y}=\lambda\mathbf{y}\cdot\mathbf{H}'(\mathbf{y})=\lambda$. Hence the system of $N+1$ equations $\mathbf{q}=\lambda\mathbf{H}'(\mathbf{z})$ and $1=H(\mathbf{z})$ has a solution at $\mathbf{q}=\mathbf{p}$, $\lambda=1$, $\mathbf{z}=\mathbf{y}$. Further, the Jacobian matrix of this system exists in a neighborhood of $(\mathbf{q},\lambda,\mathbf{z})=(\mathbf{p},1,\mathbf{y})$, and has the form

$$\mathbf{J}(\lambda,\mathbf{z})=\left[\begin{array}{c|c}\lambda\mathbf{H}''(\mathbf{z}) & \mathbf{H}'(\mathbf{z})\\ \hline (\mathbf{H}'(\mathbf{z}))^T & 0\end{array}\right],$$

where the transpose of any column vector $\mathbf{v}$ is denoted by $\mathbf{v}^T$. By hypothesis, $Q(\mathbf{v},\mathbf{H}''(\mathbf{y}))$ is positive for $\mathbf{v}$ not proportional to $\mathbf{y}$, and hence is positive for non-zero $\mathbf{v}$ satisfying $\mathbf{v}\cdot\mathbf{H}'(\mathbf{y})=0$. Therefore, $\mathbf{J}(1,\mathbf{y})$ is non-singular (Appendix A.1, Lemma 4). Hence, by continuity of the first and second differentials of H, $\mathbf{J}(\lambda,\mathbf{z})$ is non-singular in a neighborhood of $(1,\mathbf{y})$. The implicit function theorem [Bartle (1964, 21.11)] establishes the existence of a neighborhood of $\mathbf{p}$ and continuously differentiable functions $\lambda=L(\mathbf{q})$, $\mathbf{z}=\mathbf{Z}(\mathbf{q})$ defined on this neighborhood such that $\mathbf{q}\equiv L(\mathbf{q})\mathbf{H}'(\mathbf{Z}(\mathbf{q}))$, $1\equiv H(\mathbf{Z}(\mathbf{q}))$, and $L(\mathbf{p})=1$, $\mathbf{Z}(\mathbf{p})=\mathbf{y}$. The Kuhn–Tucker theorem then implies $G(\mathbf{q})=\mathrm{Max}\{\mathbf{q}\cdot\mathbf{z}|H(\mathbf{z})\leqq 1\}=\mathbf{q}\cdot\mathbf{Z}(\mathbf{q})$ on the given neighborhood of $\mathbf{p}$. From the identity $1\equiv H(\mathbf{Z}(\mathbf{q}))$, the implication $\mathbf{0}\equiv\mathbf{Z}'(\mathbf{q})\mathbf{H}'(\mathbf{Z}(\mathbf{q}))$, where $\mathbf{Z}'(\mathbf{q})$ is the $N\times N$ matrix of partial derivatives of $\mathbf{Z}(\mathbf{q})$, is obtained by differentiation. Hence, $\mathbf{G}'(\mathbf{q})=\mathbf{Z}(\mathbf{q})+\mathbf{Z}'(\mathbf{q})\mathbf{q}=\mathbf{Z}(\mathbf{q})+L(\mathbf{q})\mathbf{Z}'(\mathbf{q})\mathbf{H}'(\mathbf{Z}(\mathbf{q}))=\mathbf{Z}(\mathbf{q})$ exists and is continuously differentiable, implying that the first and second differentials of G exist and are continuous in the given neighborhood of $\mathbf{p}$. Finally, $\mathbf{G}''(\mathbf{q})=\mathbf{Z}'(\mathbf{q})$ satisfies $\mathbf{Z}'(\mathbf{q})\mathbf{q}=\mathbf{0}$ and the matrix equation

$$\mathbf{J}(L(\mathbf{q}),\mathbf{Z}(\mathbf{q}))\left[\begin{array}{c}\mathbf{Z}'(\mathbf{q})\\ \hline (\mathbf{L}'(\mathbf{q}))^T\end{array}\right]=\left[\begin{array}{c}\mathbf{I}_{N\times N}\\ \hline \mathbf{0}^T_{N\times 1}\end{array}\right],$$

where $\mathbf{I}_{N\times N}$ is the N-dimensional identity matrix, $\mathbf{0}_{N\times 1}$ is the N-dimensional column vector of zeros. Since the right-hand side of this equation is of rank N, the matrix

$$\left[\frac{\mathbf{Z}'(\mathbf{q})}{(\mathbf{L}'(\mathbf{q}))^T}\right]$$

must be of rank N, and hence $\mathbf{Z}'(\mathbf{q})$ must be at least of rank $N-1$. Since G is convex, $\mathbf{Z}'(\mathbf{q})$ is non-negative definite, implying the quadratic form $Q(\mathbf{r},\mathbf{Z}'(\mathbf{q}))$ positive for r not proportional to q. This establishes $\langle G,\mathbf{NY}\rangle$ strictly differentially quasiconvex on a neighborhood of $\mathbf{p}$.

(12) Without loss, assume $G(\mathbf{p})=1$. If $\mathbf{\Gamma}(\mathbf{p})=\{\mathbf{y}\}$, then from 16.3, $H(\mathbf{y})=1=\inf\{\lambda>0|\mathbf{q}\cdot\mathbf{y}\leqq\lambda G(\mathbf{q})$ for all $\mathbf{q}\in\mathbf{S}\}=\text{Max}\{\mathbf{q}\cdot\mathbf{y}|\mathbf{q}\in\mathbf{S},\ G(\mathbf{q})=1\}=\mathbf{p}\cdot\mathbf{y}$. As in the proof of (11), the Kuhn–Tucker theorem can be applied to establish the existence of λ such that $\mathbf{y}=\lambda\mathbf{G}'(\mathbf{p})$ and $1=\mathbf{p}\cdot\mathbf{y}=\lambda\mathbf{p}\cdot G'(\mathbf{p})=\lambda$. The result then follows by an argument completely symmetric with the proof of (11). Q.E.D.

Appendix A.4

TESTING AND IMPOSING MONOTICITY, CONVEXITY AND QUASI-CONVEXITY CONSTRAINTS

LAWRENCE J. LAU*

Stanford University

1. Introduction

1.1. *Statement of the Problem*

In many areas of economic analysis, functions are frequently assumed to be monotonic, convex or quasi-convex.[1] Production, profit, and utility functions are obvious examples. The natural questions that arise are first, whether one can test the hypotheses of monotonicity, convexity, or quasi-convexity of these functions statistically; and second, whether one

*The author wishes to thank T. Amemiya, E.R. Berndt, W.E. Diewert, A.S. Goldberger, Z. Griliches, G. Hanoch, J.A. Hausman, D.W. Jorgenson, D.L. McFadden, F.C. Nold, I. Olkin, and W.H. Yang for helpful discussions. This research was supported in part by the National Science Foundation, through Grant GS-40104 at the Institute of Mathematical Studies in the Social Sciences, Stanford University and the John Simon Guggenheim Memorial Foundation. It was essentially completed during the author's tenure as a Visiting Scholar in the Department of Economics, Massachusetts Institute of Technology, 1973–74. He retains sole responsibility for any remaining errors. An earlier version of this paper was presented at the North American Winter Meeting of the Econometric Society, New York, December 1973, and issued as Technical Report no. 123, "Econometrics of monotonicity, convexity and quasi-convexity", Institute of Mathematical Studies in the Social Sciences, Stanford University, 1974.

[1]Since the negative of a convex function is concave, and the negative of a quasi-convex function is quasi-concave, all the statements made about convexity and quasi-convexity apply to concavity and quasi-concavity as well. Concavity and quasi-concavity will not be separately considered.

can estimate these functions subject to monotonicity, convexity or quasi-convexity constraints.

In the past, empirical estimation of the parameters of these functions have been limited to those of rather simple algebraic form for which the constraints of monotonicity, convexity or quasi-convexity are either automatically or readily satisfied or can be easily imposed. For example, the linear function is always convex (and concave); the Cobb–Douglas production function estimated by the factor shares method is always monotonic and concave;[2] and, more generally, estimated Cobb–Douglas production functions are automatically quasi-concave if they satisfy the monotonicity conditions. Thus there has not been any pressing need for the development of techniques to test or impose the hypotheses of monotonicity, convexity or quasi-convexity.

However, two recent developments in empirical economic analysis have made it necessary to confront the dual problems of hypothesis testing and constrained estimation. First, partly because of advances in computational technology, partly because of substantial improvements in the quality and quantity of economic data, and partly because of a general dissatisfaction with the restrictive implications of the simple functional forms, there is a proliferation of new algebraic forms in empirical work in recent years. We shall name only a few which are capable of providing a second-order numerical approximation to an arbitrary function.[3] There is the "Transcendental Logarithmic Function", proposed by Christensen, Jorgenson, and Lau (1971, 1973, 1975),

$$\ln F(\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}' \ln \mathbf{x} + \tfrac{1}{2} \ln \mathbf{x}'\mathbf{B} \ln \mathbf{x},$$

where $\ln \mathbf{x} \equiv [\ln x_1 \ln x_2 \cdots \ln x_m]'$; the generalized version of the "Generalized Linear Function" proposed by Diewert (1971),

$$F(\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}^{1/2} + \tfrac{1}{2}\mathbf{x}^{1/2'}\mathbf{B}\mathbf{x}^{1/2},$$

where $\mathbf{x}^{1/2} \equiv [x_1^{1/2} x_2^{1/2} \cdots x_m^{1/2}]'$; and the "Quadratic Function",

$$F(\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x} + \tfrac{1}{2}\mathbf{x}'\mathbf{B}\mathbf{x},$$

where $\mathbf{x} \equiv [x_1 x_2 \cdots x_m]'$.[4]

These functions may be employed as production, profit (and normal-

[2]The factor shares method was first proposed by Klein (1953).

[3]For a definition of a second-order numerical approximation, see Lau (1974).

[4]It is a remarkable fact that Heady and Dillon (1961) have proposed a two-factor version of the transcendental logarithmic function (p. 205), the generalized linear function (pp. 91–92 and p. 206) and the quadratic function (pp. 88–91 and p. 205); and made empirical

ized profit), or utility functions. For an arbitrary set of parameters, these functions do not necessarily satisfy monotonicity, convexity or quasi-convexity conditions, either locally or globally. Hence there is a need to test or maintain these hypotheses.

Second, the increasing use of duality principles which rely heavily on convexity assumptions in empirical applications makes it mandatory that the estimated dual functions be monotonic, convex or quasi-convex.[5] While one may be willing to entertain the possibility that the production function may not be convex, the normalized profit function is always convex if the output and input markets are competitive under the assumption of profit maximization.[6] Thus a non-convex normalized profit function is inconsistent with profit maximization – the basic behavioral postulate of the theory of production. Likewise an indirect utility function is always quasi-convex by virtue of its being a maximum subject to a linear constraint. Thus, a non-quasi-convex indirect utility function is inconsistent with utility maximization – the basic behavioral postulate of the theory of consumer demand. Moreover, if the estimated normalized profit function is non-convex, or the indirect utility function is non-quasi-convex, the own and cross-price supply and demand elasticities will not have the theoretically expected signs and magnitudes. Thus one should at least test the hypotheses of monotonicity, convexity, or quasi-convexity; and if one does not reject these hypotheses, impose the corresponding constraints on the estimators so as to obtain economically meaningful estimates in practical applications.

1.2. *Historical Review*

There are two principal approaches to the testing of monotonicity, convexity and quasi-convexity – the parametric approach and the non-parametric approach.

There is little previous work in the parametric approach. Judge and

---

application with the latter two functions. For other scholars who have independently proposed the transcendental logarithmic functions, see the references listed in Christensen, Jorgenson and Lau (1973). Lau (1974) appears to be the first to propose the quadratic function as a normalized profit function.

[5]For an excellent survey of applications of duality theory, see Diewert (1974a); see also the comments by Lau (1974). Jorgenson and Lau (1974a, 1974b) give an exhaustive treatment of the role of convexity in production theory.

[6]McFadden (1966) appears to be the first person to emphasize this point. The concept of a normalized profit function is introduced by Lau (1969c). See also Jorgenson and Lau (1974a, 1974b).

Takayama (1966) analyze the question of inequality constraints in regression analysis and propose an estimator based on the solution of a constrained quadratic programming problem. Liew (1976) gives another algorithm for the computation of inequality constrained least-squares estimators.

Hudson (1969) proposes a method for fitting a polynomial in x such that it is convex in a closed interval. However, the computations are not fully worked out for all cases; the function to be estimated is restricted to be defined on a subset of R; and there is no distribution theory for the estimators.

In a previous version of this paper, a method for testing and maintaining the hypothesis of convexity of an estimated function based on the eigenvalue decomposition was proposed. The method based on the eigenvalue decomposition, however, does not reduce to an unconstrained minimization problem because the requirement of orthonormality of the eigenvectors makes necessary the construction of a set of orthonormal vectors and hence a great deal of computation, although it involves no conceptual difficulties. The present method based on the Cholesky factorization requires much less computation.

The previous work in the non-parametric approach is somewhat more numerous. Hildreth (1954) is the pioneer of this approach: assuming no algebraic functional form, he proposes to estimate the values of a function $F(x)$ at given values of x such that $F(x)$ is concave. This approach is extended by Dent (1973), who proposes to approximate $F(x)$ by polygonal segmentation.

Hanoch and Rothschild (1972) provide algorithms for testing monotonicity and quasi-convexity without assuming a specific algebraic form of the function. Afriat (1967, 1968, 1972) and Diewert (1974d) give alternative methods for estimating non-parametric functions which satisfy the assumptions of monotonicity, convexity, and homogeneity.

1.3. *Proposed Solution*

The basic technique consists of a transformation of parameters constrained to be non-negative into the squares of arbitrary real parameters and may be referred to as the "method of squaring". This technique appears to have been introduced by Valentine (1937), in connection with the solution of problems of calculus of variations subject to inequality constraints.[7] Thus, if a parameter B is required to be non-negative, it can be transformed into the parameter B^{*2}, and the estimation problem becomes that of choosing a B^* such that an appropriate sum of squares of residuals is minimized.

Further variations of this theme are possible. For example, one can substitute for any positive parameter B by e^{B^*}; any parameter lying between zero and one by the substitution $B = \frac{1}{2}(1 + \sin B^*)$,[8] or alternatively by

$$B = 1/(1 + e^{-B^*}).^9$$

By transformations similar to these one can suitably restrict a parameter to be within any prescribed interval. The advantage of this transformation is that it reduces the likelihood maximization problem to an *unconstrained* nonlinear least-squares problem.

The method of squaring thus provides a straightforward solution to the problem of monotonicity or for that matter any inequality constraint. The solution to the problem of convexity and quasi-convexity makes use of the properties of the Hessian matrices of convex and quasi-convex functions. A twice differentiable real-valued function is convex on an open convex set if and only if the Hessian is positive semidefinite everywhere on the open convex set. A twice differentiable real-valued function is quasi-convex if and only if the Hessian is positive semidefinite on any set of vectors $\mathbf{y}$ such that $\nabla F(\mathbf{x}) \cdot \mathbf{y} = 0$, where $\nabla F(\mathbf{x})$ is the non-zero gradient of $F(\mathbf{x})$. The task of this paper is to transform these conditions on the parameters into simple non-negativity conditions by a suitable reparametrization. The basis of the reparametrization is the Cholesky factorization of real symmetric matrices. Through this factorization, the determinantal conditions of positive semidefiniteness are transformed into non-negativity constraints. Once more, the method of squaring may be employed to convert the likelihood maximization problem into an unconstrained nonlinear least-squares problem.

Our method has been employed by Jorgenson and Lau (forthcoming) in the analysis of production and by Barten and Geyskens (1975) in the analysis of consumer demand. Jorgenson and Lau (1975b) have developed an alternative procedure for testing and imposing quasi-

[7]See Valentine (1937, pp. 407–409).

[8]Since the sine function is periodic, $\sin B^* = \sin(B^* + 2n\pi)$, where n is an integer. Thus any $B^* + 2n\pi$, n being an arbitrary integer, results in the same B.

[9]Alternatively, one may use the transformation $B = \sin^2 B^*$. The possibilities are limitless.

convexity constraints and extended it for testing monotonicity of correspondences not necessarily derivable from a single function.

2. Hessian Matrices of Convex and Quasi-Convex Functions

2.1. *Introduction*

In this section we discuss the properties of the Hessian matrices of twice differentiable real-valued convex and quasi-convex functions. We show that these Hessian matrices may indeed be represented by positive semidefinite matrices, which as we shall show in Section 3 have convenient factorization properties which facilitate the solution of the problems of testing and constrained estimation under the hypothesis of convexity or quasi-convexity.

In Section 2.2, the Hessian matrix of a convex function is characterized, and in Section 2.3, the Hessian matrix of a quasi-convex function is characterized. In Section 2.4, we examine the Hessian matrices of approximating functions under the hypotheses of convexity and quasi-convexity.

2.2. *The Hessian of a Convex Function*

Theorem 2.1. A twice differentiable real-valued function defined on an open convex set C is convex if and only if the Hessian matrix is positive semidefinite everywhere on C.

This theorem is well known. A proof may be found in Rockafellar (1970).[10]

2.3. *The Hessian of a Quasi-convex Function*

Definition. A real-valued function $F(\mathbf{x})$ defined on a convex set C is *quasi-convex* if

$$F(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leqq \max(F(\mathbf{x}_1), F(\mathbf{x}_2)), \qquad 0 \leqq \lambda \leqq 1,$$

for all $\mathbf{x}_1$, $\mathbf{x}_2$ in C.

[10]See Rockafellar (1970, p. 27).

We specialize our attention to consider only the class of twice differentiable functions with everywhere non-zero first partial derivatives.

> *Theorem* 2.2. A twice differentiable real-valued function $F(\mathbf{x})$ defined on an open convex set C with everywhere non-zero first partial derivatives is quasi-convex if and only if for all $\mathbf{x}$ in C, $\mathbf{y}'\mathbf{H}(\mathbf{x})\mathbf{y} \geqq 0$ whenever $\nabla F(\mathbf{x})'\mathbf{y} = 0$, where $\mathbf{H}(\mathbf{x})$ and $\nabla F(\mathbf{x})$ are respectively the Hessian matrix and the gradient of the function $F(\mathbf{x})$.

A version of this theorem is proved by Katzner (1970).[11] We omit the proof. Diewert (1973b) has derived a similar theorem under the weaker condition that not all of the components of $\nabla F(\mathbf{x})$ are zero.

> *Theorem* 2.3. A necessary and sufficient condition that $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ whenever $\mathbf{a}'\mathbf{x} = 0$ is that $\mathbf{x}'(\mathbf{A} + \lambda\mathbf{a}\mathbf{a}')\mathbf{x} \geqq 0$ for all $\mathbf{x}$ for all sufficiently large positive scalar constants λ.

Proof: Necessity is proved by contradiction. Suppose the theorem is false, then there exists $\mathbf{x}$ such that

$$\mathbf{x}'(\mathbf{A} + \lambda\mathbf{a}\mathbf{a}')\mathbf{x} < 0,$$

for all sufficiently large positive scalar constant λ. If $\mathbf{a}'\mathbf{x} \neq 0$, this implies that $\mathbf{x}'\mathbf{A}\mathbf{x}$ is not only negative but unbounded, which is not possible. Thus, $\mathbf{a}'\mathbf{x} = 0$. However, this implies $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$, a contradiction.

Sufficiency follows in a straightforward manner. If $\mathbf{x}'(\mathbf{A} + \lambda\mathbf{a}\mathbf{a}')\mathbf{x} \geqq 0$ for all $\mathbf{x}$ for some positive scalar constant λ, then $\mathbf{x}'\mathbf{A}\mathbf{x} + \lambda\mathbf{x}'\mathbf{a}\mathbf{a}'\mathbf{x} \geqq 0$ for all $\mathbf{x}$. In particular, this holds whenever $\mathbf{a}'\mathbf{x} = 0$, in which case

$$\mathbf{x}'\mathbf{A}\mathbf{x} + \lambda\mathbf{x}'\mathbf{a}\mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0. \qquad \text{Q.E.D.}$$

The basic result here is due to Finsler (1936–37). Bellman (1970) gives another proof of Finsler's Theorem.[12] The proof here is somewhat different from Bellman's and is modified to apply to the case of positive semidefiniteness of a constrained quadratic form. For the present application $\mathbf{A}$ and $\mathbf{a}$ may be identified with the Hessian and gradient of $F(\mathbf{x})$, respectively.

[11] See Katzner (1970, pp. 210–211).
[12] See Bellman (1970, pp. 76–81).

Theorem 2.4. A necessary condition for a twice differentiable real-valued function $F(\mathbf{x})$ defined on an open convex set C to be quasi-convex is that all the ordered principal minors of the bordered Hessian matrix be non-positive for all $\mathbf{x}$ in C.

This theorem is due to Arrow and Enthoven (1961).[13] We omit the proof.

We now show that indeed if $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ is positive semidefinite for sufficiently large positive λ, then all the ordered principal minors of the matrix

$$\begin{bmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{bmatrix},$$

which has the interpretation of a bordered Hessian in the present application, are non-positive. Positive semidefiniteness of $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ implies that all the principal minors of $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ are non-negative. Consider the following matrix identity:[14]

$$\begin{bmatrix} -1 & \mathbf{a}' \\ \lambda\mathbf{a} & \mathbf{A} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{a}' \\ 0 & \\ 0 & \\ \vdots & \mathbf{I} \\ 0 & \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0\cdots 0 \\ \lambda\mathbf{a} & & \mathbf{a}+\lambda\mathbf{a}\mathbf{a}' \end{bmatrix}.$$

Taking determinants of both sides, we obtain,

$$-1|\mathbf{A}| + \lambda \begin{vmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{vmatrix} = -1|\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'| \leqq 0,$$

or

$$-|\mathbf{A}| + \lambda \begin{vmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{vmatrix} \leqq 0.$$

Now this inequality must hold for all sufficiently large positive λ (in fact it can be easily shown that if $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ is positive semidefinite for some positive $\bar{\lambda}$, then it is positive semidefinite for all $\lambda > \bar{\lambda}$). Thus, one must have

$$\begin{vmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{vmatrix} \leqq 0.$$

[13]See Arrow and Enthoven (1961, pp. 797–799).
[14]This construction follows Bellman (1970, pp. 78–80).

The same proof applies to any principal submatrix of

$$\begin{bmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{bmatrix}.$$

We note that if the domain of $F(\mathbf{x})$ is restricted to the positive orthant then negativity of all the ordered principal minors of the bordered Hessian matrix except the first one everywhere is sufficient for quasi-convexity.[15]

Thus we conclude that a necessary and sufficient condition for a twice differentiable real-valued function with everywhere non-zero first partial derivatives to be quasi-convex is that the matrix

$$\mathbf{H}(\mathbf{x}) + \lambda \nabla F(\mathbf{x}) \nabla F(\mathbf{x}')$$

be positive semidefinite for all $\mathbf{x}$ for all sufficiently large positive scalar constant λ.

For the purposes of testing the hypothesis of quasi-convexity, the following theorem is more useful:

> *Theorem* 2.5. A necessary and sufficient condition that there exists a vector $\mathbf{a}$ such that $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ whenever $\mathbf{a}'\mathbf{x} = 0$ is that the number of non-negative eigenvalues of $\mathbf{A}$, an $n \times n$ real symmetric matrix, must be greater than or equal to $(n-1)$.

Proof: *Necessity*. Suppose $\mathbf{A}$ has two negative eigenvalues. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the unit eigenvectors corresponding to these two negative eigenvalues. Consider any $\mathbf{a}$: if either $\mathbf{a}'\mathbf{v}_1$ or $\mathbf{a}'\mathbf{v}_2 = 0$, then let $\mathbf{x} = \mathbf{v}_1$ or $\mathbf{v}_2$ and $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{v}_i'\mathbf{A}\mathbf{v}_i = \rho_i < 0$; so assume $\mathbf{a}'\mathbf{v}_1 \times \mathbf{a}_1^* \neq 0$, $\mathbf{a}'\mathbf{v}_2 = \mathbf{a}_2^* \neq 0$. Let $\mathbf{x} = a_2^*\mathbf{v}_1 - a_1^*\mathbf{v}_2$, then $\mathbf{a}'\mathbf{x} = a_2^* a_1^* - a_1^* a_2^* = 0$. But $\mathbf{x}'\mathbf{A}\mathbf{x} = a_2^*\mathbf{v}_1'\mathbf{A}\mathbf{v}_1 a_2^* + a_1^*\mathbf{v}_2'\mathbf{A}\mathbf{v}_2 a_1^* - a_2^*\mathbf{v}_1'\mathbf{a}\mathbf{v}_2 a_1^* = \rho_1 a_2^{*2} + \rho_2 a_1^{*2} < 0 (\mathbf{v}_1'\mathbf{v}_2 = 0)$. Thus the number of non-negative eigenvalues of $\mathbf{A}$ must be greater than or equal to $(n-1)$.

Sufficiency. If the number of non-negative eigenvalues of $\mathbf{A}$ is $(n-1)$, and suppose $\mathbf{v}_1$ is the eigenvector corresponding to the one remaining negative eigenvalue, let $\mathbf{a} = \mathbf{v}_1$. Then whenever $\mathbf{a}'\mathbf{x} = \mathbf{v}_1'\mathbf{x} = 0$, $\mathbf{x}$ is a linear combination of the remaining $(n-1)$ eigenvectors with non-negative eigenvalues. Thus $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$. Q.E.D.

[15] See Arrow and Enthoven (1961, pp. 797–799).

The practical implication of this theorem for the quasi-convexity problem is that the Hessian of a twice differentiable quasi-convex function, $\mathbf{H}(\mathbf{x})$, must have greater than or equal to $(n-1)$ non-negative eigenvalues (or alternatively at most one negative eigenvalue) everywhere on the interior of its effective domain. In other words, in order that one can find a $\nabla F(\mathbf{x})$ such that $\mathbf{y}'\mathbf{H}(\mathbf{x})\mathbf{y} \geqq 0$ whenever $\nabla F(\mathbf{x})'\mathbf{y} = 0$, it is necessary and sufficient that $\mathbf{H}(\mathbf{x})$ has at least $(n-1)$ non-negative eigenvalues.

2.4. *Hessian Matrices of Functions Approximating Convex and Quasi-Convex Functions*

In general, economic theory itself does not provide sufficient restrictions on the functional relationships used in economic analysis so that these relationships may be represented by a single parametric class of algebraic functions.[16] The restrictions derived from economic theory are almost always of a more general type – monotonicity, convexity, quasi-convexity, homogeneity, etc., which may be satisfied by many different parametric classes of algebraic functions.

However, in applied econometrics, it is frequently necessary to estimate a function, for example, a production function, parametrically and a particular parametric form of an algebraic function must be specified. Such a functional form, inasmuch as it is not directly derivable from economic theory, should be looked upon as an approximation to the unknown underlying true function. Thus, the "translog", the generalized linear, and the quadratic functions may all be considered as alternative second-order numerical approximations to the same unknown, underlying true function.

The distinction between the approximating functions and the underlying true function is not so important, were it not for the fact that not all of the properties of the underlying true function will be inherited by the approximating function.[17] In general, even if the underlying true function is convex globally, the approximating function, while providing a good approximation, may not be globally convex or even locally convex itself.

However, if we restrict our attention to those second-order approximating functions which agree with the first and second derivatives

[16]By contrast, consider the inverse-square law of electric potential, or Boltzman's equation in statistical mechanics.

[17]See the discussion in Lau (1975).

at the point of approximation, then all approximating functions to underlying monotonic, convex, or quasi-convex functions will exhibit behavior similar to the function they are approximating at the point of approximation. In particular, the approximating function to a monotonic function will have all of its partial derivatives of one sign at the point of approximation. The approximating function to a convex function will have its Hessian matrix positive semidefinite at the point of approximation. The approximating function to a quasi-convex function will have a Hessian matrix that is positive semidefinite with respect to all vectors orthogonal to its gradient at the point of approximation.

It should be noted that convexity of the approximating function at the point of approximation does not in general guarantee that the approximating function itself will be globally convex. It also does not guarantee global convexity of the underlying true function that is being approximated. However, non-convexity of the approximating function at the point of approximation necessarily implies non-convexity of the underlying true function. Hence from the point of view of statistical inference, one may use local convexity of the approximating function as a basis for a test.[18]

Thus these local conditions are necessary conditions in the sense that if the underlying true function were to be monotonic, convex and quasi-convex, the approximating function must exhibit corresponding properties at the point of approximation. They are therefore ideally suited for hypothesis testing.

On the other hand, these properties at the point of approximation are all that one can expect to hold for an arbitrary approximating function if one wants to impose the constraints implied by these hypotheses, rather than to test them statistically.

As has been pointed out elsewhere, one has the choice of imposing these hypotheses either globally or locally on the approximating functions.[19] Convexity at the point of approximation, of course, does not in general guarantee that the approximating function itself will be globally convex with the exception of special cases such as approximation by a quadratic function. For many families of approximating functions such as the generalized linear function, it is possible to find restrictions on the parameters such that a member of the family is globally convex on its effective domain. However, these restrictions will

[18]One should bear in mind here that under classical procedures for statistical inference, one can attach a confidence level to rejections, but not to "acceptances".

[19]Lau (1974).

usually turn out to be more stringent than those implied by local convexity at the point of approximation. Moreover, not all families of approximating functions can be made globally convex without severely restricting the parameters. The transcendental logarithmic function, for example, can be made globally convex (for all **x**) only under rather restrictive assumptions such as a unitary elasticity of substitution between all pairs of commodities. It can be shown, however, that a sufficient condition for convexity on the set $\{\mathbf{x}|\mathbf{x} \geqq \mathbf{e}\}$, where $\mathbf{e} = [e\, e \cdots e]'$, is

$$\alpha_i \leqq 0 \quad \text{and} \quad B_{ij} \leqq 0 \quad \forall\, i \quad \text{and} \quad j.$$ [20]

Unfortunately similar conditions do not obtain when the effective domain properly contains $\mathbf{x} = [\mathbf{1}]$, a vector of units, in which case one has to be content with local convexity. It can also be shown that a necessary and sufficient condition for a generalized version of the generalized linear function to be globally convex on its effective domain (the non-negative orthant of R^n), is $\alpha_i \leqq 0$ and $B_{ij} \leqq 0$, $i \neq j$, $\forall i$ and j.[21] Finally, it can be shown that for the quadratic function a necessary and sufficient condition for global convexity (that is, convexity on all of R^n) is that **B** is positive semidefinite.

Thus, the problem of testing the hypotheses of monotonicity, convexity and quasi-convexity becomes that of testing non-negativity and positive semidefiniteness constraints; and the problem of constrained estimation becomes that of imposition of non-negativity and positive semidefiniteness restrictions. Non-negativity constraints suffice for global convexity of the generalized linear function. However, convexity of the quadratic function, the transcendental logarithmic function, and the generalized linear function (at $\mathbf{x} = [\mathbf{1}]$), requires positive semidefiniteness constraints.

[20] Sufficiency follows from consideration of the convexity conditions of functions such as $B_{ij} \ln x_i \ln x_j$ and the fact that a non-negative linear combination of convex functions is convex and that an increasing convex function of a convex function is convex.

[21] Sufficiency is trivial and follows from the following facts: (i) $-y_i^{1/2} y_j^{1/2}$, $i \neq j$, is a convex function; (ii) a non-negative linear combination of convex function is convex; and finally (iii) the sum of a convex function and a linear function is convex.

Necessity follows from consideration of the diagonal elements of the Hessian matrix, which are

$$F_{ii} = -y_i^{-3/2}\left(\alpha_i + \sum_{j \neq i} B_{ij} y_j^{1/2}\right), \quad \forall i.$$

In order for $F_{ii} \geqq 0$ for all y in the non-negative orthant of R^n, it is necessary that $\alpha_i \leqq 0$ and $B_{ij} \leqq 0$, $i \neq j$, $\forall i,j$. Since these conditions are also sufficient for global convexity, they are therefore both necessary and sufficient.

3. The Cholesky Factorizability of Semidefinite and Indefinite Matrices

3.1. *Introduction*

In the preceding section we have seen that the hypothesis of convexity implies that the Hessian matrix of the approximating function is positive semidefinite at the point of approximation. Similarly, the hypothesis of quasi-convexity implies that the matrix $(\mathbf{H} + \lambda\, \nabla F\, \nabla F')$ where $\mathbf{H}$ and ∇F are respectively the Hessian matrix and gradient of the approximating function is positive semidefinite for a sufficiently large λ at the point of approximation. Thus, the problem of testing convexity and quasi-convexity of the underlying true functions becomes that of testing whether a given real symmetric matrix is positive semidefinite.

One solution which naturally suggests itself is based on the eigenvalue decomposition of real symmetric matrices.[22] It is well known that any real symmetric matrix $\mathbf{B}$ can be written as

$$\mathbf{B} = \mathbf{P\Lambda P'},$$

where Λ is a diagonal matrix whose elements are the eigenvalues of the matrix $\mathbf{B}$ and $\mathbf{PP'} = \mathbf{I}$. Moreover, $\mathbf{B}$ is positive semidefinite if and only if $\Lambda_{ii} \geqq 0$, $\forall i$. Thus, in principle, given an estimator of $\mathbf{B}$ and its variance–covariance matrix, one can compute an estimator of Λ and the variance–covariance matrix of Λ, which may then be used to test the hypothesis that $\Lambda_{ii} \geqq 0$, $\forall i$. However, since Λ is in general a rather complicated function of $\mathbf{B}$, such computation is likely to be laborious. Further, if the estimate of Λ turns out not to be non-negative, and constrained estimation is necessary, that is, one needs to impose the constraint that $\Lambda_{ii} \geqq 0$, $\forall i$, the computational problem for the eigenvalue decomposition becomes quite complex because in the estimation of $\mathbf{B} = \mathbf{P\Lambda p}'$, it is necessary to impose not only the constraint that $\Lambda_{ii} \geqq 0$, $\forall i$, but also the orthonormality constraint of $\mathbf{PP'} = \mathbf{I}$. For $n > 2$ the computational burden becomes quite formidable.

For the practical reason, we introduce a different factorization of the matrix $\mathbf{B}$ – the Cholesky factorization – which avoids some of these difficulties.[23] It will be shown that every positive semidefinite matrix has

[22]This method was proposed in the cited earlier version of this paper and independently by Salvas-Bronsard et al. (1973).

[23]After this paper was substantially completed, Arthur Goldberger brought to my attention an article by Wiley, Schmidt and Bramble (1973) which also makes use of the Cholesky factorization in connection with imposing positive definiteness constraints.

a Cholesky factorization with non-negative Cholesky values.[24] Thus, the Hessian matrices of twice differentiable convex and quasi-convex functions may be characterized in terms of their Cholesky factorizations.

Although not all real symmetric matrices have Cholesky factorizations, it will be shown that the set of real symmetric matrices of order n which do not have Cholesky factorizations has measure zero in the set of real symmetric matrices of order n.

3.2. *Cholesky Factorization*

Definition. A square matrix **A** is a *unit lower triangular matrix* if

$$A_{ii} = 1, \qquad \forall i,$$
$$A_{ij} = 0, \qquad j > i, \quad \forall i,j.$$

A unit lower triangular matrix will be denoted **L**.

Definition. A square matrix **A** is a *unit upper triangular matrix* if

$$A_{ii} = 1, \qquad \forall i,$$
$$A_{ij} = 0, \qquad j < i, \quad \forall i,j.$$

A unit upper triangular matrix will be denoted by **R**. The transpose of a unit lower triangular matrix is of course a unit upper triangular matrix, and vice versa.

Definition. A real symmetric square matrix **A** is said to have a *Cholesky factorization* if there exists a unit lower triangular matrix **L** and a diagonal matrix **D** such that

$$\mathbf{A} = \mathbf{LDL}',$$

where $\mathbf{L}'$ denotes the transpose of **L**. The matrix **A** is also said to be *Cholesky factorizable.*

Definition. A square matrix **A** is an *upper triangular matrix* if

$$A_{ij} = 0, \qquad j < i, \quad \forall i,j.$$

[24]The concept of Cholesky factorization is due to Cholesky. It is discussed in Householder (1964, pp. 10–17) and Wilkinson (1965, pp. 229–230).

An upper triangular matrix will be denoted $\mathbf{U}$. We note that the product $\mathbf{DL}'$ is an upper triangular matrix. Thus, for any matrix $\mathbf{A}$ which has a Cholesky factorization, one may write equivalently

$$\mathbf{A} = \mathbf{LU}.$$

Lemma 3.1. The inverse of a unit lower triangular matrix is a unit lower triangular matrix.

Proof: The proof is by induction on the order of the matrix. The lemma is obviously true for $n = 1$. Assume that it is true for $n - 1$, we shall prove that it is true for n. An nth order unit lower triangular matrix may be written as

$$\mathbf{L}_n = \begin{bmatrix} \mathbf{L}_{n-1} & 0 \\ & 0 \\ & \cdot \\ & \cdot \\ \mathbf{l}' & 1 \end{bmatrix}$$

where $\mathbf{L}_{n-1}$ is a unit lower triangular matrix of order $n - 1$. The inverse may be directly computed as

$$\mathbf{L}_n^{-1} = \begin{bmatrix} \mathbf{L}_{n-1}^{-1} & 0 \\ & 0 \\ & \cdot \\ & \cdot \\ -\mathbf{l}'\mathbf{L}_{n-1}^{-1} & 0 \\ & 1 \end{bmatrix}$$

since $\mathbf{L}_{n-1}$ is of order $n - 1$, $\mathbf{L}_{n-1}^{-1}$ is unit lower triangular by hypothesis. Hence $\mathbf{L}_n^{-1}$ is also unit lower triangular and the lemma is proved. Q.E.D.

We shall now show how a Cholesky factorization may be accomplished by way of an example. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

and

$$\mathbf{LU} = \begin{bmatrix} 1 & 0 \\ L_{21} & 1 \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} \\ L_{21}U_{11} & L_{21}U_{12} + U_{22} \end{bmatrix}.$$

By equating the matrices element by element, we have

$$U_{11} = 1,$$
$$U_{12} = 1,$$
$$L_{21}U_{11} = 1,$$

and

$$L_{21}U_{12} + U_{22} = 1,$$

which gives

$$U_{11} = 1,$$
$$U_{12} = 1,$$
$$L_{21} = 1,$$

and

$$U_{22} = 0.$$

Thus

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{U} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

Now

$$\mathbf{U} = \mathbf{D}\mathbf{L}',$$
$$\mathbf{D} = \mathbf{U}\mathbf{L}'^{-1},$$
$$= \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus, the Cholesky factorization is given by

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

In actual fact, it is never necessary to calculate the inverse of **L** explicitly. Since **D** is diagonal, one can obtain the elements of **D** from the equations

$$U_{ii} = D_{ii}L_{ii}, \quad \forall i,$$

or

$$D_{ii} = U_{ii},$$

since

$$L_{ii} = 1, \quad \forall i.$$

Notice that in the example, the matrix $\mathbf{A}$ is singular. Hence non-singularity of a matrix is not a necessary condition for the possibility of Cholesky factorization. It is not true, however, that all real symmetric matrices have Cholesky factorizations. For example, let

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

then the equations for the elements of $\mathbf{L}$ and $\mathbf{U}$ become

$$\begin{aligned} U_{11} &= 0, \\ U_{12} &= 1, \\ L_{21}U_{11} &= 1, \\ L_{21}U_{12} + U_{22} &= 1. \end{aligned}$$

But the equations $U_{11} = 0$ and $L_{21}U_{11} = 1$ are inconsistent with each other. Hence there do not exist matrices $\mathbf{L}$ and $\mathbf{U}$ such that $\mathbf{A} = \mathbf{LU}$ and therefore the matrix $\mathbf{A}$ does not have a Cholesky factorization. Notice also that $\mathbf{A}$ is non-singular. Hence non-singularity is not a sufficient condition for the possibility of Cholesky factorization. Finally, we note that by a simultaneous permutation of rows and columns,

$$\mathbf{A}^* = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

$\mathbf{A}$ becomes Cholesky factorizable. But this is not true of all real symmetric matrices either, as demonstrated by the example

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which is non-singular but not Cholesky factorizable, even with simultaneous permutations of rows and columns. We shall show, however, that every positive semidefinite symmetric matrix has a Cholesky factorization in Section 3.3 and characterize all Cholesky factorizable real symmetric matrices in Section 3.4.

3.3. *Representation Theorem for Positive Semidefinite Matrices*

First we define positive definiteness and prove a couple of useful lemmas:

Definition. A real symmetric matrix $\mathbf{A}$ is positive definite if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$; it is positive semidefinite if $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ for all $\mathbf{x}$.

An equivalent definition of positive semidefiniteness is the following:

Definition. A real symmetric matrix $\mathbf{A}$ is positive semidefinite if and only if *all* the principal minors of $\mathbf{A}$ of all orders are non-negative.[25]

Lemma 3.2. If a diagonal element of a positive semidefinite matrix $\mathbf{A}$ is zero, the corresponding row and column must be identically zero.

Proof: Without loss of generality, we may take the matrix to be

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix},$$

where

$$\mathbf{A}_1 \equiv \begin{bmatrix} A_{12} \\ A_{13} \\ \vdots \\ A_{1n} \end{bmatrix}, \qquad \mathbf{A}_n \equiv \begin{bmatrix} A_{22}A_{23}\cdots A_{2n} \\ A_{23}A_{33}\cdots A_{3n} \\ \vdots \\ A_{2n}A_{3n}\cdots A_{nn} \end{bmatrix}.$$

Positive semidefiniteness of $\mathbf{A}$ implies that for all $\mathbf{x}, \mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$. Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ A_1 \end{bmatrix},$$

where $x_1 < 0$. Then

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1\mathbf{A}_1'] \begin{bmatrix} 0 & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \mathbf{A}_1 \end{bmatrix} \\ &= \mathbf{A}_1'\mathbf{A}_n\mathbf{A}_1 + 2x_1\mathbf{A}_1'\mathbf{A}_1. \end{aligned}$$

[25]Note that non-negativity of the *ordered* principal minors alone is not sufficient (although positivity is). For example, consider the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}.$$

If $A_1'A_1 > 0$, then by choosing x_1 sufficiently large in magnitude, $\mathbf{x}'\mathbf{A}\mathbf{x}$ will become negative. Thus in order for $\mathbf{A}$ to be positive semidefinite, one must have $\mathbf{A}_1'\mathbf{A}_1 = 0$ which implies that $\mathbf{A}_1 = \mathbf{0}$. Q.E.D.

Lemma 3.3. If a matrix $\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}$ is positive semidefinite, and $A_{11} \neq 0$ then $\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1'$ is also positive semidefinite.

Proof: $\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}$ positive semidefinite implies that

$$[x_1\mathbf{x}']\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}\begin{bmatrix} x_1 \\ \mathbf{x} \end{bmatrix} \geqq 0, \quad \forall x_1, \mathbf{x}.$$

Thus, $x_1A_{11}x_1 + \mathbf{x}'\mathbf{A}_1x_1 + x_1\mathbf{A}_1'\mathbf{x} + \mathbf{x}'\mathbf{A}_n\mathbf{x} \geqq 0, \forall x_1, \mathbf{x}$. One may choose

$$x_1 = -A_{11}^{-1}\mathbf{A}_1'\mathbf{x}.$$

The inequality becomes

$$A_{11}^{-1}\mathbf{x}'\mathbf{A}_1\mathbf{A}_1'\mathbf{x} - \mathbf{x}'\mathbf{A}_1A_{11}^{-1}\mathbf{A}_1'\mathbf{x} - \mathbf{x}'\mathbf{A}_1A_{11}^{-1}\mathbf{A}_1'\mathbf{x} + \mathbf{x}'\mathbf{A}_n\mathbf{x} \geqq 0,$$

or

$$\mathbf{x}'(\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1')\mathbf{x} \geqq 0.$$

Since this holds for all $\mathbf{x}$, the matrix

$$(\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1') \quad \text{is positive semidefinite.} \quad \text{Q.E.D.}$$

A corollary of this lemma is that if $\mathbf{A}$ is positive definite, then $\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1'$ is also positive definite.

Now we can prove the following representation theorem:

Theorem 3.1. Every positive semidefinite matrix $\mathbf{A}$ has a Cholesky factorization.

Proof: The proof is by induction on the order of the matrix. Clearly for $n = 1$,

$$\mathbf{A} = [A_{11}], \qquad A_{11} \geqq 0,$$

one can choose

$$\mathbf{L} = [1], \qquad \mathbf{D} = [A_{11}].$$

Now suppose that the theorem is true for positive semidefinite matrices or order less than or equal to $(n-1)$, we shall prove that it is true for positive semidefinite matrices of order n.

Let

$$\mathbf{A} = \begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}.$$

Since $\mathbf{A}$ is positive semidefinite, $A_{11} \geqq 0$. If $A_{11} = 0$, then by Lemma 3.2,

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \cdots 0 \\ 0 & \\ \vdots & \mathbf{A}_n \\ 0 & \end{bmatrix}.$$

$\mathbf{A}_n$ is a positive semidefinite matrix of order $(n-1)$ and hence has a Cholesky factorization. One may therefore choose

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \cdots 0 \\ 0 & & \\ 0 & & \\ \vdots & & \mathbf{L}_n \\ 0 & & \end{bmatrix} \begin{bmatrix} 0 & 0 \cdots 0 \\ 0 & \\ \vdots & \\ 0 & \mathbf{D}_n \\ 0 & \end{bmatrix} \begin{bmatrix} 1 & 0 \; 0 \cdots 0 \\ 0 & \\ 0 & \\ \vdots & \mathbf{L}_n' \\ 0 & \end{bmatrix},$$

where $\mathbf{A}_n = \mathbf{L}_n \mathbf{D}_n \mathbf{L}_n'$ is the Cholesky factorization of $\mathbf{A}_n$. If $A_{11} > 0$, consider the following matrix identity:

$$\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \cdots 0 \\ A_{11}^{-1}\mathbf{A}_1 & & \mathbf{I} \end{bmatrix} \begin{bmatrix} A_{11} & 0 & 0 \cdots 0 \\ 0 & & \\ 0 & & \\ \vdots & & \mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1' \\ 0 & & \end{bmatrix} \begin{bmatrix} 1 & A_{11}^{-1}\mathbf{A}_1' \\ 0 & \\ 0 & \\ \vdots & \mathbf{I} \\ 0 & \end{bmatrix},$$

which may be directly verified by computation. Now $\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1'$ is a positive semidefinite matrix of order $(n-1)$, by Lemma 3.3. Thus it has a Cholesky factorization,

$$\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1' = \mathbf{L}_n^* \mathbf{D}_n^* \mathbf{L}_n^{*\prime}.$$

Therefore $\mathbf{A}$ has a Cholesky factorization given by

$$\mathbf{A} = \begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0\cdots 0 \\ A_{11}^{-1}\mathbf{A}_1 & \mathbf{L}_n^* \end{bmatrix} \begin{bmatrix} A_{11} & 0 & 0\cdots 0 \\ 0 & & \\ 0 & & \\ \vdots & & \mathbf{D}_n^* \\ 0 & & \end{bmatrix} \begin{bmatrix} 1 & A_{11}^{-1}\mathbf{A}_1' \\ 0 & \\ 0 & \\ \vdots & \mathbf{L}_n^{*\prime} \\ 0 & \end{bmatrix}. \qquad \text{Q.E.D.}$$

The proof given here follows essentially the argument given by Householder (1964).[26] Note that this theorem also implies that all negative semidefinite matrices have Cholesky factorizations.

The elements of the diagonal matrix $\mathbf{D}$, that is, the D_{ii}'s will be referred to as *Cholesky values*. The following theorem establishes the properties of the Cholesky values of positive definite and semidefinite matrices.

Theorem 3.2. A real symmetric matrix $\mathbf{A}$ is positive definite (semidefinite) if and only if its Cholesky values are positive (nonnegative).

Proof: A positive definite matrix $\mathbf{A}$ can be written as $\mathbf{LDL}'$. Thus

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{LDL}'\mathbf{x} > 0, \quad \forall \mathbf{x}, \quad \mathbf{x} \neq 0.$$

Writing $\mathbf{y} = \mathbf{L}'\mathbf{x}$, we have

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{D}\mathbf{y} = \sum_i D_{ii} y_i^2 > 0, \quad \forall \mathbf{y}, \quad \mathbf{y} \neq 0.$$

Thus all $D_{ii} > 0, \forall i$. Conversely, if $D_{ii} > 0$, $\forall i, \mathbf{y}'\mathbf{D}\mathbf{y} > 0$, $\forall \mathbf{y}, \mathbf{y} \neq 0$. Buy $\mathbf{y} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{L}'^{-1}\mathbf{y} = \mathbf{0}$. Thus

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{D}\mathbf{y} > 0, \quad \forall \mathbf{x}, \quad \mathbf{x} \neq 0.$$

A similar proof goes through for a positive semidefinite matrix $\mathbf{A}$. Q.E.D.

It is useful to give a determinantal interpretation to the Cholesky values. Consider the following partitioned matrix identity:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{11} & 0 \\ 0 & \mathbf{D}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{11}' & \mathbf{L}_{21}' \\ 0 & \mathbf{L}_{22}' \end{bmatrix},$$

[26]See Householder (1964, pp. 12–13).

where $\mathbf{L}_{11}$ and $\mathbf{L}_{22}$ are unit lower triangular matrices, and $\mathbf{D}_{11}$ and $\mathbf{D}_{22}$ are diagonal matrices. By direct computation

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11}\mathbf{D}_{11}\mathbf{L}'_{11} & \mathbf{L}_{11}\mathbf{D}_{11}\mathbf{L}'_{21} \\ \mathbf{L}_{21}\mathbf{D}_{11}\mathbf{L}'_{11} & \mathbf{L}_{21}\mathbf{D}_{11}\mathbf{L}'_{21} + \mathbf{L}_{22}\mathbf{D}_{22}\mathbf{L}'_{22} \end{bmatrix}.$$

First, since the determinants of unit lower triangular matrices are identically one, we have

$$|\mathbf{A}| = \begin{vmatrix} \mathbf{D}_{11} & 0 \\ 0 & \mathbf{D}_{22} \end{vmatrix} = \prod_i D_{ii}.$$

The determinant of $\mathbf{A}$ is thus the product of all of its Cholesky values. Second, take any principal minor of $\mathbf{A}$, say $|\mathbf{A}_{11}|$,

$$|\mathbf{A}_{11}| = |\mathbf{L}_{11}||\mathbf{D}_{11}||\mathbf{L}'_{11}| = |\mathbf{D}_{11}| = \prod_i D_{ii},$$

where the product is taken over the Cholesky values corresponding to the $\mathbf{A}_{11}$ block. Thus, we have the following system of determinantal equalities:

$$|A_{11}| = D_{11},$$

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{vmatrix} = D_{11}D_{22},$$

$$\begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{12} & A_{22} & A_{23} \\ A_{13} & A_{23} & A_{33} \end{vmatrix} = D_{11}D_{22}D_{33},$$

$$\vdots$$

$$\begin{vmatrix} A_{11} & A_{12}\cdots A_{1n} \\ A_{12} & A_{22}\cdots A_{2n} \\ \vdots & \vdots \\ A_{1n} & A_{2n}\cdots A_{nn} \end{vmatrix} = D_{11}D_{22}D_{33}\cdots D_{nn}.$$

This interpretation is valid, however, if and only if a Cholesky factorization of the matrix $\mathbf{A}$ exists.

> *Theorem* 3.3. If $\mathbf{A}$ is positive definite, then the Cholesky factorization is unique.

Proof: The proof is by direct computation. Suppose $\mathbf{A} = \mathbf{LDL}' = \mathbf{L^*D^*L^{*\prime}}$, and $\mathbf{A}$ is positive definite. This implies that $D_{ii} > 0$ and

$D_{ii}^* > 0$, $\forall i$, by Theorem 3.2. By equating $\mathbf{LDL}'$ and $\mathbf{L^*D^*L^{*\prime}}$ element by element, it can be shown that indeed $\mathbf{L} = \mathbf{L}^*$ and $\mathbf{D} = \mathbf{D}^*$. Q.E.D.

Uniqueness is also discussed by Householder (1964) in terms of determinantal conditions.[27]

> *Theorem* 3.4. If a real symmetric matrix **A** has a Cholesky factorization, then the number of positive, negative, and zero Cholesky values is the same as the number of positive, negative, and zero eigenvalues.

This theorem follows from Sylvester's Law of Inertia, a proof of which may be found in Gantmacher (1959),[28] which implies conservation of the signature of a real symmetric matrix. From this theorem we can also deduce immediately that if **A** is positive definite, all the Cholesky values are positive; if **A** is positive semidefinite, all the Cholesky values are nonnegative.

We conclude that every positive semidefinite matrix **A** has a Cholesky factorization $\mathbf{LDL}'$ with all the elements of **D** non-negative. Thus, to check whether a real symmetric matrix is positive semidefinite, one needs only check its Cholesky values. And to impose the condition that a real symmetric matrix is positive semidefinite, one needs only constrain the Cholesky values to be non-negative.

3.4. *Representation Theorem for Arbitrary Real Symmetric Matrices*

We have shown that all semidefinite matrices are Cholesky factorizable in the previous subsection. However, semidefiniteness is by no means necessary for Cholesky factorizability as the last example in subsection 3.2 illustrates.

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

The Cholesky factorization of **A** is

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

[27]See Householder (1964, pp. 10–12).
[28]See Gantmacher (1959, pp. 296–298).

Since the Cholesky values are 1 and -1, $\mathbf{A}$ is an indefinite matrix. It is also clear that non-singularity of $\mathbf{A}$ is neither necessary (because of semidefinite matrices) nor sufficient – because of example

$$A=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix},$$

for Cholesky factorizability. The purpose of this subsection is to characterize the set of all real symmetric matrices which are Cholesky factorizable and to show that the set of all real symmetric matrices which are not Cholesky factorizable is a subset of measure zero in the set of all real symmetric matrices.

We first prove two simple but useful lemmas:

Lemma 3.4. A real symmetric matrix $\mathbf{A}$ with $A_{11}=0$ is Cholesky factorizable only if the first row and first column are identically zero.

Proof: Let

$$\mathbf{A}=\begin{bmatrix}0 & \mathbf{A}_1'\\ \mathbf{A}_1 & \mathbf{A}_n\end{bmatrix}=\mathbf{LDL}'.$$

The first row of $\mathbf{LDL}'$ may be directly computed as

$$\begin{aligned} D_{11} &= A_{11}=0,\\ L_{21}D_{11} &= A_{12},\\ &\vdots\\ L_{n1}D_{11} &= A_{1n}. \end{aligned}$$

Since $D_{11}=0$, $A_{1i}=0$, $i=2\cdots n$. Q.E.D.

Lemma 3.5. A real symmetric Cholesky factorizable square matrix is non-singular if and only if all the Cholesky values are non-zero.

Proof: Referring back to the determinantal interpretation of Cholesky values in Section 3.3, if one or more of the Cholesky values is zero, then the determinant of the matrix is zero, and therefore the matrix is singular. If the matrix is non-singular, then its determinant is non-zero and none of the Cholesky values can be zero. Q.E.D.

Our first theorem characterizes all real symmetric matrices which are Cholesky factorizable.

Theorem 3.5. Let a real symmetric matrix $\mathbf{A}$ be partitioned conformably as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix},$$

where $\mathbf{A}_1$ and $\mathbf{A}_n$ are both real symmetric square matrices. In addition, suppose that $\mathbf{A}_1$ is Cholesky factorizable and non-singular. Then $\mathbf{A}$ is Cholesky factorizable if and only if $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is Cholesky factorizable.

Proof: If $\mathbf{A}$ were Cholesky factorizable then there exist $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{L}_{21}$, $\mathbf{D}_1$, $\mathbf{D}_2$ such that

$$\begin{bmatrix} \mathbf{L}_1 & 0 \\ \mathbf{L}_{21} & \mathbf{L}_2 \end{bmatrix}\begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix}\begin{bmatrix} \mathbf{L}'_1 & \mathbf{L}'_{21} \\ 0 & \mathbf{L}'_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix},$$

which implies

$$\begin{aligned} \mathbf{L}_1\mathbf{D}_1\mathbf{L}'_1 &= \mathbf{A}_1, \\ \mathbf{L}_1\mathbf{D}_1\mathbf{L}'_{21} &= \mathbf{A}'_{1n}, \\ \mathbf{L}_{21}\mathbf{D}_1\mathbf{L}'_{21} + \mathbf{L}_2\mathbf{D}_2\mathbf{L}'_2 &= \mathbf{A}_n. \end{aligned}$$

But $\mathbf{A}_1$ is non-singular, then by Lemma 3.5, $\mathbf{D}^{-1}$ exists. Therefore, $\mathbf{L}'_{21} = \mathbf{D}_1^{-1}\mathbf{L}_1^{-1}\mathbf{A}'_{1n}$ and $\mathbf{L}_{21}\mathbf{D}_1\mathbf{L}'_{21} = \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$. Hence $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n} = \mathbf{L}_2\mathbf{D}_2\mathbf{L}'_2$ and is Cholesky factorizable.

Conversely, consider the matrix identity

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{A}_{1n}\mathbf{A}_1^{-1} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{A}_1^{-1}\mathbf{A}'_{1n} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

which may be verified by direct multiplication. If

$$\mathbf{A}_1 = \mathbf{L}_1\mathbf{D}_1\mathbf{L}'_1,$$

and

$$\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n} = \mathbf{L}_2\mathbf{D}_2\mathbf{L}'_2,$$

then the matrix identity becomes

$$\begin{aligned} &\begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{A}_{1n}\mathbf{A}_1^{-1} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{L}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_2 \end{bmatrix}\begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}\begin{bmatrix} \mathbf{L}'_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{L}'_2 \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{A}_1^{-1}\mathbf{A}'_{1n} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{L}_1 & 0 \\ \mathbf{A}_{1n}\mathbf{A}_1^{-1} & \mathbf{L}_2 \end{bmatrix}\begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}\begin{bmatrix} \mathbf{L}'_1 & \mathbf{A}_1^{-1}\mathbf{A}'_{1n} \\ \mathbf{0} & \mathbf{L}'_2 \end{bmatrix}, \end{aligned}$$

which implies that $\mathbf{A}$ is Cholesky factorizable. Q.E.D.

Corollary 5.1. Let the real symmetric matrix be

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix},$$

where $\mathbf{A}_n$ is a scalar, then $\mathbf{A}_1$ is Cholesky factorizable and nonsingular implies that $\mathbf{A}$ is Cholesky factorizable.

Proof: $A_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is a scalar matrix which is always Cholesky factorizable. Q.E.D.

This corollary implies that any 2×2 real symmetric matrix $\mathbf{A}$ is Cholesky factorizable if A_{11} is different from zero.

This theorem provides a constructive way of verifying whether a given real symmetric matrix $\mathbf{A}$ is Cholesky factorizable. If $A_{11} = 0$, we know that the first row and column must be identically zero for Cholesky factorizability. Then $\mathbf{A}$ is Cholesky factorizable if and only if $\mathbf{A}_n$, the submatrix of $\mathbf{A}$ with the first row and first column deleted is Cholesky factorizable. If $A_{11} \neq 0$, then $\mathbf{A}$ is Cholesky factorizable if and only if $\mathbf{A}_n - \mathbf{A}_{1n}A_{11}^{-1}\mathbf{A}'_{1n}$ is Cholesky factorizable. One can continue in this way until some $\mathbf{A}_n^* - \mathbf{A}_{1n}^* A_{11}^{*-1} \mathbf{A}_{1n}^{*\prime}$ becomes a scalar matrix or is shown to be not Cholesky factorizable.

Of special interest, of course, is the case in which the 1,1-th element of $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is zero. In that case, by Lemma 3.4, the first row and the first column of $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is identically zero. To be more specific let us write:

$$\mathbf{A}_1 = \begin{bmatrix} A_{11} & A_{12} \cdots A_{1m} \\ A_{12} & A_{22} \cdots A_{2m} \\ \vdots & \vdots \quad \vdots \\ A_{1m} & A_{2m} \cdots A_{mm} \end{bmatrix},$$

$$\mathbf{A}'_{1n} = \begin{bmatrix} A_{1,m+1} & A_{1,m+2} \cdots A_{1n} \\ \vdots & \vdots \quad \vdots \\ A_{m,m+1} & A_{m,m+2} \cdots A_{m,n} \end{bmatrix},$$

$$\mathbf{A}_n = \begin{bmatrix} A_{m+1,m+1} & A_{m+1,m+2} \cdots A_{m+1,n} \\ \vdots & \vdots \quad \vdots \\ A_{m+1,n} & A_{m+2,n} \cdots A_{nn} \end{bmatrix},$$

then the first row and column of $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ being zero implies

$$\begin{bmatrix} A_{m+1,m+1} \\ A_{m+1,m+2} \\ \vdots \\ A_{m+1,n} \end{bmatrix} = \mathbf{A}_{1n}\mathbf{A}_1^{-1} \begin{bmatrix} A_{1,m+1} \\ A_{2,m+1} \\ \vdots \\ A_{m,\,m+1} \end{bmatrix},$$

which in turn implies that

$$\begin{bmatrix} A_{m+1,m+2} \\ A_{m+1,m+2} \\ \vdots \\ A_{m+1,n} \end{bmatrix} = \sum_{i=1}^{m} \alpha_i \begin{bmatrix} A_{i,m+1} \\ A_{i,m+2} \\ \vdots \\ A_{i,n} \end{bmatrix},$$

that is, the first column of $\mathbf{A}_n$ can be expressed as a linear combination of the columns of $\mathbf{A}_{1n}$.

Next, we give a set of sufficient conditions for Cholesky factorizability which do not depend on semidefiniteness.

> *Theorem* 3.6. If the ordered principal submatrices of all orders up to $n-1$ of a real symmetric matrix $\mathbf{A}$ of order n are non-singular, then $\mathbf{A}$ is Cholesky factorizable.

Proof: The proof is by induction on the order of the matrix. For $n = 1$, the theorem is trivially true. For $n = 2$, if $A_{11} \neq 0$, $\mathbf{A}$ is Cholesky factorizable by Corollary 5.1. Now suppose the theorem is true for $(n-1)$, consider an nth order real symmetric matrix, partitioned into

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_{nn} \end{bmatrix},$$

where A_1 is $(n-1)\times(n-1)$ and all the ordered principal submatrices of $\mathbf{A}$ are non-singular. Since all the ordered principal submatrices of $\mathbf{A}_1$ are non-singular, $\mathbf{A}_1$ is Cholesky factorizable by hypothesis. But $\mathbf{A}_1$ is also non-singular, because it is a principal submatrix of $\mathbf{A}$. Then by Theorem 3.5, $\mathbf{A}$ is Cholesky factorizable. Q.E.D.

Finally, we want to show that although not all real symmetric matrices are Cholesky factorizable, those which are not constitute a subset of measure zero in the set of all real symmetric matrices.

Let A_n be the set of all real symmetric matrices of order n. Let A_n^0 be the subset of all such matrices which are singular. Let A_n^* be the subset of all such matrices which are non-singular. Then, by definition

$$\mathbf{A}_n^0 \bigcup \mathbf{A}_n^* = \mathbf{A}_n, \qquad \mathbf{A}_n^0 \bigcap \mathbf{A}_n^* = \phi.$$

Lemma 3.6. The set of singular matrices $\mathbf{A}_n^0$ in $\mathbf{A}_n$ is of measure zero.

Proof: Consider each element $\mathbf{A}$ of $\mathbf{A}_n$ as an element in $R^{n(n+1)/2}$. For each $\{A_{12}A_{13}\cdots A_{nn}\}$ in $R^{n(n+1)/2-1}$, the set of A_{11}'s such that

$$|\mathbf{A}| = A_{11}\begin{vmatrix} A_{22}\cdots A_{2n} \\ A_{23}\cdots A_{3n} \\ \vdots \quad \vdots \\ A_{2n}\cdots A_{nn}\end{vmatrix} + \begin{vmatrix} 0 & A_{12} & A_{13}\cdots A_{1n} \\ A_{12} & A_{22} & \cdots A_{2n} \\ \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots A_{nn}\end{vmatrix} = 0$$

is a set of measure zero. Hence by Fubini's theorem, the measure of the set of $\mathbf{A}$'s for which $|\mathbf{A}| = 0$ is of measure zero. Q.E.D.

Lemma 3.7. The set of matrices $\mathbf{A}_{n,k}^0$ in $\mathbf{A}_n$ for which the kth order principal submatrix, that is, $[A_{ij}; i,j = 1,\dots,k; k < n]$, is singular is of measure zero.

Proof: Again consider each element A of $\mathbf{A}_n$ as an element in $R^{n(n+1)/2}$. For each $\{A_{12}\,A_{13}\cdots A_{nn}\}$ in $R^{n(n+1)/2-1}$, the set of A_{11} is such that

$$|\mathbf{A}_{n,k}| = A_{11}\begin{vmatrix} A_{22}\cdots A_{2k} \\ A_{23}\cdots A_{3k} \\ \\ A_{2k}\cdots A_{kk}\end{vmatrix} + \begin{vmatrix} 0 & A_{12} & A_{13}\cdots A_{1k} \\ A_{12} & A_{22} & \cdots A_{2k} \\ \vdots & & \vdots \\ A_{1k} & A_{2k} & \cdots A_{kk}\end{vmatrix} = 0$$

is a set of measure zero. Hence by Fubini's theorem, the set of $\mathbf{A}$'s for which $|\mathbf{A}_{n,k}| = 0$ is of measure zero. Q.E.D.

Lemma 3.8. The set of matrices $\mathbf{A}_{n,n}^*$ for which the naturally ordered sequence of principal submatrices are all non-singular is of measure one in $\mathbf{A}_n$.

Proof: Any element of $\mathbf{A}_{n,n}^*$ cannot be contained in $\mathbf{A}_{n,k}^0$ for any k, otherwise at least one of the principal submatrices will be singular. Hence it must be contained in $\mathbf{A}_n - \bigcup_k \mathbf{A}_{n,k}^0$. But any element of $\mathbf{A}_n - \bigcup_k \mathbf{A}_{n,k}^0$ must have its naturally ordered principal submatrices all non-singular, and is hence contained in $A^*_{n,n}$. Thus,

$$A^*_{n,n} = A_n - \bigcup_k A^0_{n,k},$$

and

$$A^*_{n,n} \bigcap \left[\bigcup_k A^0_{n,k} \right] = \phi.$$

Therefore,

$$\mu(A^*_{n,n}) = \mu(A_n) - \mu\left(\bigcup_k A^0_{n,k} \right),$$

but the measure of a finite union of sets of measure zero is zero. Thus

$$\mu(A^*_{n,n}) = 1. \qquad \text{Q.E.D.}$$

Theorem 3.7. The set of real symmetric matrices of order n which are not Cholesky factorizable is of measure zero within the set of all real symmetric matrices of order n.

Proof:[29] The set of real symmetric matrices which are Cholesky factorizable contains the set of real symmetric matrices with all ordered principal submatrices which are non-singular, by Theorem 3.6. But the latter set has measure one by Lemma 3.8. Hence the set of Cholesky factorizable real symmetric matrices is of measure one, and its complement, the set of matrices which are not Cholesky factorizable, is of measure zero. Q.E.D.

4. Estimation

4.1. *Introduction*

The proposed method of estimation is maximum likelihood, which is known to have certain optimal properties. For the sake of expositional convenience, we shall focus our attention on the following model:[30]

$$Y_t = \alpha_0 + \alpha'\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{B}\mathbf{Z}_t + \epsilon_t, \quad t = 1,\dots,T,$$

[29]The proof of this theorem is due to Daniel McFadden.

[30]In most models of producer or consumer behavior, there will be more than one equation as well as parametric restrictions across equations. We abstract from these complications so as to keep the exposition simple.

where $Z_{ti} = g(X_{ti})$ with $g(\cdot)$ a known algebraic function of a single variable, $\mathbf{X}_t$ is a vector of independent variables, and $\boldsymbol{\epsilon}$ is distributed as $N(\mathbf{0},\sigma^2\mathbf{I})$, $\mathbf{B}$ is a real symmetric matrix and our objective is to test the hypotheses that the gradient is non-negative and/or the Hessian matrix is positive semidefinite and to find an estimator of $\boldsymbol{\alpha}$ and $\mathbf{B}$ that are consistent with a non-negative gradient and/or a positive semidefinite Hessian matrix at the point of approximation.

Since the Hessian matrix of any convex function has a Cholesky factorization, it is possible to transform the elements of the Hessian matrix in terms of the elements of its Cholesky factorization $\mathbf{L}$ and $\mathbf{D}$. Then a test of the convexity hypothesis consists of a simultaneous test that all the D_{ii}'s of the Cholesky factorization are non-negative. If constrained estimates are needed, they can be obtained by setting each D_{ii} equal to the square of a new parameter, say, D_{ii}^{*2}.

In terms of the parameters of the quadratic function, this implies

$$\begin{bmatrix} B_{11} & B_{12}\cdots B_{1n} \\ B_{12} & B_{22}\cdots B_{2n} \\ \vdots & \\ B_{1n} & B_{2n}\cdots B_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} D_{11} & L_{21}D_{11} & \cdots L_{n1}D_{11} \\ L_{21}D_{11} & L_{21}^2D_{11}+D_{22} & \cdots L_{21}L_{n1}D_{11}+L_{n2}D_{22} \\ \vdots & & \\ L_{n1}D_{11} & L_{21}L_{n1}D_{11}+L_{n2}D_{22} & \cdots L_{n1}^2D_{11}+L_{n2}^2D_{22}+\cdots+D_{nn} \end{bmatrix}.$$

We note that this condition is global, as the Hessian matrix is constant for the quadratic function, and is consistent with local convexity of the unknown, underlying true function of which the quadratic function is an approximation.

In terms of the parameters of the transcendental logarithmic function at $\mathbf{x} = [\mathbf{1}]$, a vector of units, this implies

$$\begin{bmatrix} B_{11}+\alpha_1(\alpha_1-1) & B_{12}+\alpha_1\alpha_2 & \cdots B_{1n}+\alpha_1\alpha_n \\ B_{12}+\alpha_1\alpha_2 & B_{22}+\alpha_2(\alpha_2-1) & \cdots B_{2n}+\alpha_2\alpha_n \\ B_{1n}+\alpha_1\alpha_n & B_{2n}+\alpha_2\alpha_n & \cdots B_{nn}+\alpha_n(\alpha_n-1) \end{bmatrix} = \mathbf{LDL}'.$$

Satisfaction of this equality with $D_{ii} \geqq 0$, $\forall i$, does not imply global convexity of the transcendental logarithmic function, but is consistent with convexity in a neighborhood of $\mathbf{x} = [\mathbf{1}]$, and with local convexity of the unknown, underlying true function of which the transcendental logarithmic function is an approximation.

And in terms of the parameters of the generalized linear function at $\mathbf{x} = [\mathbf{1}]$, a vector of units, this implies

$$\frac{1}{4}\begin{bmatrix} -\left(\alpha_1 + \sum\limits_{j\neq 1} B_{1j}\right) & B_{12} & B_{13}\cdots & B_{1n} \\ B_{12} & -\left(\alpha_2 + \sum\limits_{j\neq 2} B_{2j}\right) & \cdots & B_{2n} \\ \vdots & \vdots & & \vdots \\ B_{1n} & B_{2n} & & -\left(\alpha_n + \sum\limits_{j\neq n} B_{nj}\right) \end{bmatrix} = \mathbf{LDL}'.$$

Again this does not imply global convexity. Estimation of the original, untransformed model will be referred to as "Problem 0". For the purposes of hypotheses testing and constrained estimation, two additional, separate estimation problems may be distinguished. For the quadratic function, these are[31]

Problem 1:

$$Y_t = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{LDL}'\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\ldots,T.$$

In this case we seek parameters α_i's, L_{ij}'s $(j < i)$ and D_{ii}'s, without constraints.

Problem 2:

$$Y_t = \alpha_0 + \boldsymbol{\alpha}^{*2\prime}\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{LD}^{*2}\mathbf{L}'\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\ldots,T.$$

In the second case we seek parameters α_i^*'s, L_{ij}'s $(j < i)$ and D_{ii}^*'s, where

$$\alpha_i^{*2} \equiv \alpha_i,$$

and

$$D_{ii}^{*2} \equiv D_{ii}, \quad \forall i.$$

Again the problem is unconstrained. But the resultant estimates of α_i and D_{ii} from Problem 2 will be non-negative, hence $\hat{\alpha}_i \geqq 0$, $\forall i$, and $\hat{\mathbf{B}} \equiv \mathbf{LDL}' \equiv \mathbf{LD}^{*2}\mathbf{L}'$ will be positive semidefinite. Under our specification,

[31]Similar problems 1 and 2 may be set up for the "translog" and generalized linear functions. One should note that it is the Hessian matrix, which may be different from the matrix of second order coefficients, which should be set equal to $\mathbf{LDL}'$.

the likelihood maximization problem is equivalent to an unconstrained nonlinear least-squares problem.

Here, the computational advantage of the Cholesky factorization over the eigenvalue decomposition is most easily seen. If the Cholesky factorization $\mathbf{B} = \mathbf{LDL}'$ is used, then the number of independent unknown parameters of $\mathbf{B}$, $n(n+1)/2$, is precisely equal to the number of independent unknown parameters of $\mathbf{L}$ and $\mathbf{D}$, $n(n-1)/2 + n = n(n+1)/2$. Hence no additional constraints are required. On the other hand, if the eigenvalue decomposition $\mathbf{B} = \mathbf{P\Lambda P}'$ is used, the number of unknown parameters of $\mathbf{P}$ and $\mathbf{\Lambda}$ are (n^2+n). The parameters of $\mathbf{P}$ are subject to additional orthonormality restrictions such that $\mathbf{P'P} = \mathbf{I}$. One can either solve out for the parameters of $\mathbf{P}$ in terms of a minimal set of independent parameters or alternatively impose $\mathbf{P'P} = \mathbf{I}$ as $n(n+1)/2$ additional side conditions. In any event, substantial computations are involved. All these are avoided by using the Cholesky factorization.

The basic model for quasiconvexity is again

$$Y_t = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{B}\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\dots,T,$$

where $\boldsymbol{\epsilon}$ is $N(\mathbf{0},\sigma^2\mathbf{I})$ and $\nabla F \geqq 0$, $(H + \lambda\nabla F\nabla F')$ positive semidefinite for all sufficiently large positive scalar constant λ. In terms of the parameters of the quadratic function, at $\mathbf{x} = [\mathbf{0}]$, the vector of zeroes,

$$\mathbf{H}(\mathbf{x}) + \lambda\nabla F(\mathbf{x})\nabla F(\mathbf{x})' = \begin{bmatrix} B_{11} + \lambda\alpha_1^2 & B_{12} + \lambda\alpha_1\alpha_2 & B_{1n} + \lambda\alpha_1\alpha_n \\ B_{12} + \lambda\alpha_1\alpha_2 & B_{22} + \lambda\alpha_2^2 & B_{2n} + \lambda\alpha_2\alpha_n \\ \vdots & & \vdots \\ B_{1n} + \lambda\alpha_1\alpha_n & B_{2n} + \lambda\alpha_2\alpha_n & B_{nn} + \lambda\alpha_n^2 \end{bmatrix}.$$

In terms of the parameters of the transcendental logarithmic function, at $\mathbf{x} = [\mathbf{1}]$, a vector of units,

$$\mathbf{H}(\mathbf{x}) + \lambda\nabla F(\mathbf{x})\nabla F(\mathbf{x})' = \begin{bmatrix} B_{11} - \alpha_1 + \lambda\alpha_1^2 & B_{12} + \lambda\alpha_1\alpha_2 & \cdots B_{1n} + \lambda\alpha_1\alpha_n \\ B_{12} + \lambda\alpha_1\alpha_2 & B_{22} - \alpha_2 + \lambda\alpha_2^2 & \cdots B_{2n} + \lambda\alpha_2\alpha_n \\ \vdots & \vdots & \\ B_{1n} + \lambda\alpha_1\alpha_n & B_{2n} + \lambda\alpha_2\alpha_n & \cdots B_{nn} - \alpha_n + \lambda\alpha_n^2 \end{bmatrix}.$$

Here we have made use of the fact that a monotonic transformation (in this case ln) of a quasi-convex function is quasi-convex.

Besides the Problem 0′ referred to earlier, three additional separate estimation problems are distinguished for the quadratic function:[32]

Problem 1′:

$$Yt = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \mathbf{Z}_t'(\mathbf{LDL}' - \boldsymbol{\alpha}\boldsymbol{\alpha}')\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\ldots,T.$$

In this case, we seek parameters α_i's, L_{ij}'s $(j < i)$ and D_{ii}'s, without constraints.

Problem 2′:

$$Y_t = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \mathbf{Z}_t'(\mathbf{LDL}' - \bar{\lambda}\boldsymbol{\alpha}\boldsymbol{\alpha}')\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\ldots,T,$$

where $\bar{\lambda}$ is the largest positive scalar constant that the computer can recognize. In the second case we seek parameter α_i's, L_{ij}'s $(j < i)$ and D_{ii}'s, again without constraints.

Problem 3′:

$$Y_t = \alpha_0 + \boldsymbol{\alpha}^{*2\prime}\mathbf{Z}_t + \mathbf{Z}_t'(\mathbf{LD}^{*2}\mathbf{L}' - \lambda^{*2}\boldsymbol{\alpha}\boldsymbol{\alpha}')\mathbf{Z}_t + \boldsymbol{\epsilon}_t, \qquad t = 1,\ldots,T,$$

where $\alpha_i^{*^2} \equiv \alpha_i$ and $D_{ii}^{*^2} \equiv D_{ii}$, $\forall i$. We seek parameters α_i^*'s L_{ij}'s $(j < i)$, D_{ii}^*'s and λ^*. Again the problem is unconstrained. But the resultant estimates of α_i, D_{ii}, and λ will be non-negative. Hence $\hat{\alpha}_i \geqq 0$, and $(\hat{B} + \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}}')$ is positive semidefinite for a non-negative $\hat{\lambda}$. Under our specification the likelihood maximization problem is again equivalent to an unconstrained nonlinear least-squares problem.

4.2. *The Method of Maximum Likelihood*

The method of maximum likelihood is well known.[33] In order to contain a paper that is already too long, maximum likelihood equations will not be reproduced here. We do want to point out several possible approaches. First, if **x** is entirely exogenous and there is only one stochastic equation in the system, the nonlinear least-squares estimate is efficient for both Problems 1 and 2. Second, the estimator obtained by stopping after the first iteration of the Newton–Raphson process corresponds to the linearized maximum likelihood estimator proposed by Rothenberg and Leenders (1964), which is also an efficient estimator,

[32]Again, similar problems may be set up for other functional forms.
[33]See, for instance, Kendall and Stuart (1967, Vol. 2, pp. 35–74).

provided that the initial estimator is consistent. Since consistent estimators are easy to obtain – for example, they can be obtained in this case from the unconstrained ordinary least-squares estimator of $\boldsymbol{\alpha}$ and $\mathbf{B}$ – of Problem 0 – computationally the linearized maximum likelihood estimator is quite attractive. Third, in the case that there are other equations in the system, $\mathbf{x}$ being still entirely exogenous, full information maximum likelihood will yield an efficient estimator. The full information maximum likelihood procedure is equivalent to an iterative nonlinear weighted least-squares procedure. As before, ordinary least-squares applied to Problem 0 will provide a set of consistent estimates which may be used to initialize the Newton–Raphson process, yielding the linearized maximum likelihood estimator, which is known to be efficient, after the first iteration. Finally, in the case that some or all of $\mathbf{x}$ is endogenous, one can either do the full maximum likelihood calculation for the system, or alternatively, one can first apply Amemiya's (1974) procedure to obtain a set of consistent estimates, and then using these as initial estimators, compute the linearized full information maximum likelihood estimator. The latter estimator has the same asymptotic distribution properties as full information maximum likelihood estimator but is much easier to compute since convergence is no longer required.

On a more practical level the following observations may be relevant. First, it can be verified by direct inspection that the maximum likelihood estimator of $\mathbf{B}$ obtained from substitution of an estimator of $\mathbf{LDL}'$ is independent of the ordering of the variables. For example: $B_{11}=0$ implies $L_{21}=0$ and hence $B_{12}=0$; but likewise $B_{22}=0$ implies $L_{21}=0$ and hence $B_{12}=0$. On the other hand $L_{21}=0$ implies $B_{12}=0$ but has no effect on B_{11} or B_{22}. Thus the ordering of $\mathbf{x}$ imposes no special restrictions on the form that $\mathbf{B}$ can assume other than that it is Cholesky factorizable. Second, it is not in general possible to test or impose the hypothesis of strict convexity of $F(\mathbf{x})$ or positive definiteness of the Hessian, say $\mathbf{B}$. The difficulty lies in that the maximum likelihood problem in which $D_{ii}>0$, $\forall i$, has only a least upper bound, which necessarily cannot be attained. In actual implementation, however, if one is interested in imposing strict convexity, one may set $D_{ii}\equiv D_{ii}^{*2}+\bar{\epsilon}$, where $\bar{\epsilon}$ is the smallest possible positive number that a given computer can recognize. Third, it is clear that if the Hessian of a twice differentiable function is positive definite at some $\mathbf{x}$, by continuity it will be positive semidefinite in an open neighborhood of $\mathbf{x}$. Thus, one can expand the region on which the Hessian is positive semidefinite, and possibly to include all of the data points, by the choice of a sufficiently large $\bar{\epsilon}$.[34] Finally, a special word should be said with regard to estimation subject to quasi-convexity constraints. Recall that quasi-convexity implies that

$$\mathbf{H}(\mathbf{x}) + \lambda \nabla F(\mathbf{x}) \nabla F(\mathbf{x})'$$

is positive semidefinite. For Problem 1′, it is evident that λ can be set equal to one without loss of generality. However, if one or more of the constraints turns out to be violated, that is, the estimated $D_{ii} < 0$ for at least one i, then one should set $\lambda = \lambda^{*2}$ as is done in Problem 3′ and estimate λ^* as an additional parameter.[35]

4.3. *Efficiency and Asymptotic Distribution Theory*

First, it is well known that under mild regularity conditions the maximum likelihood estimators are asymptotically efficient with the asymptotic variance–covariance matrix given by $-E([\partial^2 \ln L/\partial \mathbf{l}^2])^{-1}$ where $\ln L$ is the logarithm of the likelihood function. This matrix may be consistently estimated by $-[\partial^2 \ln L/\partial \mathbf{l}^2]^{-1}_{\mathbf{l}=\mathbf{l}^*}$, where $\mathbf{l}^*$ is the maximum likelihood estimator of $\mathbf{l}$.[36] For our problem, the regularity conditions are satisfied. Thus for both Problem 1 and Problem 2, the estimated asymptotic variance–covariance matrix may be directly computed.

However, it can be shown that, asymptotically, the estimated variance–covariance matrices of Problem 1 and Problem 2 converge to the same matrix, $-E([\partial^2 \ln L/\partial \mathbf{l}^2])^{-1}$. This is because, as Rothenberg (1966) has shown, the use of inequality constraints in maximum likelihood estimation does not increase the efficiency of the estimators.[37] Thus, for

[34]Thus in cases involving functions such as the transcendental logarithmic production function, which can only be made locally convex at some specific $\mathbf{x} = \mathbf{x}_0$, one can use this technique to enlarge the region of convexity. This idea is due to Mr. Yoichi Okita. However, there does not seem to be any optimality properties associated with this procedure. This basis of the procedure is related to the fact that given any real symmetric matrix $\mathbf{B}$ and positive scalar λ, $\mathbf{B} + \lambda \mathbf{I}$ can be made positive definite for a sufficiently large λ.

[35]The reason λ^{*2} should be re-introduced at this point is to offset partially the loss of estimable parameters as a result of a binding positive semidefiniteness constraint. If λ^{*2} is not re-introduced, one will have, in effect, constrained $\mathbf{B}$ itself to be positive semidefinite. It is possible that $\partial^2 \ln L/\partial \mathbf{l}^2$, where $\mathbf{l}$ now includes λ^* may become singular, in which case a generalized inverse should be used. See Section 4.4 below, especially Lemma 4.1.

[36]As noted below, $\partial^2 \ln L/\partial \mathbf{l}^2$ may become singular if the positive semidefiniteness constraint becomes binding. In that case the variance–covariance matrix of the estimable parameters consists of appropriate parts of the generalized inverse of $\partial^2 \ln L/\partial \mathbf{l}^2$.

[37]See Rothenberg (1966, pp. 51–55).

the purposes of obtaining efficient estimators, monotonicity and convexity constraints may be ignored; and any consistent estimator of $-[\partial^2 \ln L/\partial \Gamma^2]^{-1}$ will do for the asymptotic variance–covariance matrix.

This is not to say that the monotonicity, convexity or quasi-convexity constraints are worthless. In most econometric applications, we require that the results are reasonable, that is, consistent with economic theory. Very often, the only way to ensure reasonableness is through the imposition of these constraints. Moreover, it is possible that the constrained estimators do better in finite samples. And a different criterion of optimality of the estimator, such as minimum expected mean squared error, may favor the constrained estimators over the unconstrained ones.

4.4. *Computational Notes*

Under our transformations the maximum likelihood problem becomes that of finding an unconstrained maximum for a nonlinear programming problem – in fact, a quadric programming problem. It is also equivalent to a nonlinear least-squares problem.

Eisenpress and Greenstadt (1966) and Eisenpress (1968) provide a maximum likelihood program that will accommodate both nonlinearities in variables and in parameters. Basically, a Marquardt–Levenberg algorithm is used, with the Newton–Raphson algorithm available as a special case. The latter is of particular interest because the first iteration of the Newton–Raphson algorithm produces the linearized maximum likelihood estimator proposed by Rothenberg and Leenders (1964).

Many other methods for unconstrained maximization are available. In general, however, a second derivative method should be used because the asymptotic variance–covariance matrix must be estimated by the negative of the Hessian matrix of the natural logarithm of the likelihood function at the point of convergence. Murray (1972) provides a comprehensive survey of alternative unconstrained maximization methods.[38]

A special feature of the positive semidefiniteness constraints as indicated by the following lemma imposes additional requirements on the algorithm:

[38]See Murray (1972, especially Chs. 3 and 4).

Lemma 4.1. The matrix product $\mathbf{LDL}'$ where $\mathbf{L}$ is unit lower triangular and $\mathbf{D}$ is diagonal is independent of the ith column of $\mathbf{L}$ if D_{ii} is zero.

Proof: The proof is by direct computation.

$$\mathbf{LDL}' = \begin{bmatrix} 1 & & & \\ L_{21} & 1 & & 0 \\ L_{31} & L_{32} & & \\ \vdots & & \ddots & \\ L_{n1} & L_{n2} & & 1 \end{bmatrix} \begin{bmatrix} 0 & D_{11} & & 0 \\ & D_{22} & & \\ & 0 & \ddots & \\ & & & D_{nn} \end{bmatrix} \begin{bmatrix} 1 & L_{21} & L_{31} & L_{n1} \\ & 1 & L_{32} & L_{n2} \\ & 0 & & \ddots \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} D_{11} & & & \cdots 0 \\ L_{21}D_{11} & D_{22} & & \cdots 0 \\ L_{31}D_{11} & L_{32}D_{22} & D_{33} & \cdots 0 \\ L_{n1}D_{11} & L_{n2}D_{22} & & \cdots D_{nn} \end{bmatrix} \begin{bmatrix} 1 & L_{21} & L_{31} & L_{n1} \\ & 1 & L_{32} & L_{n2} \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} D_{11} & L_{21}D_{11} & L_{31}D_{11} & \cdots L_{n1}D_{11} \\ L_{21}D_{11} & L_{21}^2D_{11} + D_{22} & L_{21}L_{31}D_{11} + L_{32}D_{22} & L_{21}L_{n1}D_{11} + L_{n2}D_{22} \\ \vdots & \vdots & \vdots & \\ L_{n1}D_{11} & L_{21}L_{n1}D_{11} + L_{n2}D_{22} & & L_{n1}^2D_{11} + L_{n2}^2D_{22} + \cdots + D_{nn} \end{bmatrix}$$

It is evident that if $D_{ii} = 0$, the values of $L_{ji}, \forall j$, do not matter at all, because all the L_{ji} entries are multiplied by D_{ii}. Q.E.D.

Lemma 4.1 implies that if one knows $D_{ii} = 0$, one can, without loss, replace the ith column of $\mathbf{L}$ by zeroes everywhere off the diagonal. By definition, $L_{ii} = 1$.

What this means is that if one of the positive semidefiniteness constraints turns out to be binding, the matrix $[\partial^2 \ln L/\partial \mathbf{l}^2]$ will always become singular. Hence the usual Newton–Raphson iterative method or any one of the second-derivative modified Newton methods fails. There are various ways to remedy this situation. The most elegant method is to continue the Newton–Raphson process using a generalized inverse of $[\partial^2 \ln L/\partial \mathbf{l}^2]$ until the process converges.[39] Conditions for convergence of the Newton–Raphson process using a generalized inverse are given by Ben-Israel (1966).

[39]For discussion of generalized inverses, see Rao and Mitra (1971). Rao and Mitra also propose the use of the generalized inverse when the information matrix becomes singular (pp. 201–203).

5. Testing of Hypotheses

5.1. *Introduction*

The basic statistical problem in testing the hypotheses of monotonicity, and/or convexity can be reduced to the following: the null hypothesis to be considered is of the type

$$H_0 : \boldsymbol{\alpha} \geqq 0,$$

and the alternative hypothesis is

$$H_a : \boldsymbol{\alpha} \ngeqq 0,$$

that is, at least one $\alpha_i < 0$. If α is one-dimensional, the usual one-tailed t-test will work. However, the usual likelihood ratio test procedure based on asymptotic distribution theory breaks down because under the null hypothesis there is no reduction in the dimensionality of the space of possible parameters of the likelihood function. An alternative test procedure is needed. We shall examine three alternatives: Bonferroni t-statistics; distribution of extreme values; and finally a test procedure based on the likelihood ratio. The Bonferroni t-statistics is recommended because of its easy computability and the ready availability of tables.

To test these hypotheses, one needs to use the estimates from Problem 1, that is, the unconstrained estimates. An alternative approach is possible: let $\mathbf{B}$ be a maximum likelihood estimate of B of the original untransformed model, that is, Problem 0. For convenience we shall stack the parameters of $\mathbf{B}$ into one vector, of dimension $n(n+1)/2$,

$$\boldsymbol{\beta} \equiv [B_{11}B_{12}\cdots B_{1n}B_{22}B_{23}\cdots B_{nn}]',$$

similarly,

$$\mathbf{l} \equiv [D_{11}L_{21}L_{31}\cdots L_{n1}D_{22}L_{32}\cdots D_{nn}]'.$$

Defining the matrix

$$\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{l}} = \begin{bmatrix} \frac{\partial \beta_1}{\partial l_1} & \frac{\partial \beta_2}{\partial l_1} \cdots \frac{\partial \beta_{n^*}}{\partial l_1} \\ \frac{\partial \beta_1}{\partial l_2} & \frac{\partial \beta_2}{\partial l_2} \cdots \frac{\partial \beta_{n^*}}{\partial l_2} \\ \vdots & \vdots \quad\quad \vdots \\ \frac{\partial \beta_1}{\partial l_{n^*}} & \frac{\partial \beta_2}{\partial l_{n^*}} \cdots \frac{\partial \beta_{n^*}}{\partial l_{n^*}} \end{bmatrix},$$

that is, the Jacobian of the transformation from $\boldsymbol{\beta}$ into $\mathbf{l}$, we have, at the point of maximum likelihood:

$$\left[\frac{\partial^2 \ln L}{\partial \mathbf{l} \partial \mathbf{l}'}\right] = \left[\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{l}}\right]\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right]\left[\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{l}}\right]'.$$

Thus an estimator of $V(\hat{\mathbf{l}})$, which is given by

$$-E\left(\left[\frac{\partial^2 \ln L}{\partial \mathbf{l} \partial \mathbf{l}'}\right]\right)^{-1},$$

the variance–covariance matrix of $\hat{\mathbf{l}}$, may be computed given a knowledge of

$$\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right].$$

One can then extract from $V(\hat{\mathbf{l}})$ those components which correspond to the D_{ii}'s and use them as a basis for statistical inference. The disadvantage of this approach, however, is that there may be cases in which one may not be able to obtain D_{ii}'s from $\hat{\mathbf{B}}$, that is, $\hat{\mathbf{B}}$ may not be Cholesky factorizable. In that case, one will have to fall back on the solution given by Problem 1.[40]

The testing of the hypotheses of monotonicity and quasiconvexity is slightly more complicated. The difficulty lies in the fact that it is sometimes not possible to reject conclusively that there may exist a sufficiently large positive scalar constant λ such that $(\mathbf{B}+\lambda\boldsymbol{\alpha\alpha}')$ is positive semidefinite. The null hypothesis to be considered is: $\exists\lambda, \lambda > 0$, such that corresponding to that λ,

$$\alpha_i \geqq 0, \qquad D_{ii} \geqq 0, \qquad \forall i.$$

Obviously, for each λ, one can carry out a test of this type; however, each rejection is not conclusive, because there always remains the possibility (only a possibility) that for some larger λ, the hypotheses may not be rejected.

[40]Since not all real symmetric matrices are Cholesky factorizable, a legitimate question at this point is whether the power of our test of positive semidefiniteness of $\mathbf{B}$ is reduced by restricting our consideration to the class of real symmetric matrices that are Cholesky factorizable. The answer is no if the distribution function of the errors in the equation is continuous and smooth everywhere, as in the case of the normal distribution since the set of real symmetric matrices which are not Cholesky factorizable is of measure zero. Heuristically, the situation is comparable to that of maximizing a likelihood function with respect to a single parameter α and subject to the restriction that $\alpha \neq 0$. The resultant distribution of $\hat{\alpha}$ is essentially the same as that of an estimator derived without the restriction $\alpha \neq 0$. Constrained estimation, on the other hand, requires no such justification since all positive semidefinite real symmetric matrices are Cholesky factorizable.

Our proposed solution is to decompose the test procedure into two stages. First, we make use of Theorem 2.5 and Theorem 3.4 which together imply that a necessary and sufficient condition that there exists a vector $\mathbf{a}$ such that $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ whenever $\mathbf{a}'\mathbf{x} = 0$ is that the number of non-negative Cholesky values of $\mathbf{A}$, an $n \times n$ Cholesky factorizable real symmetric matrix, must be greater than or equal to $(n-1)$. Thus, if the Hessian matrix is Cholesky factorizable, this implies that there are at least $(n-1)$ non-negative Cholesky values (or alternatively at most one negative Cholesky value) everywhere. This provides a basis for testing the hypothesis of quasi-convexity at the point of approximation. However, because one has no knowledge *a priori* which one of the n Cholesky values will turn out to be negative, if any, this theorem does not lead to a corresponding procedure for constrained estimation. We therefore test the hypothesis that $\mathbf{B}$ has at least $(n-1)$ non-negative Cholesky values simultaneously with the hypothesis that $\alpha_i \geqq 0, \forall i$. Thus

$$
\begin{aligned}
H_0{:}\ & \alpha_i \geqq 0, \quad \forall i, \\
& D_{ii} \geqq 0, \quad \forall i, \quad \text{except possibly one.}
\end{aligned}
$$

There are three possible outcomes with respect to the D_{ii}'s. First, we do not reject that $D_{ii} \geqq 0, \forall i$. This implies that not only can we not reject the hypothesis of quasi-convexity, but also that of convexity as well. Second, we do not reject the hypothesis that $D_{ii} \geqq 0$ for only $(n-2)$ i's. This implies that the function cannot possibly be quasi-convex at the point of approximation, by virtue of Theorem 2.5. Third, we do not reject the hypothesis that $D_{ii} \geqq 0$ for only $(n-1)$ i's. This leaves open the possibility that one may find a λ and $\mathbf{D}$ such that $\mathbf{LDL}' = (\mathbf{B} + \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}')$ with $D_{ii} \geqq 0, \forall i$, that is, the function may still be quasi-convex.

Since the computer can only handle finite arithmetic, it is not possible to check through all possible λ's. The next best alternative is therefore to set $\lambda = \bar{\lambda}$, where $\bar{\lambda}$ is the largest positive scalar constant that the computer can recognize, and to test the hypothesis that the resultant D_{ii}'s are all non-negative. This constitutes the second stage of the test procedure.[41]

[41] After this paper was essentially completed, Jorgenson and Lau (1975b) have proposed a procedure, also based on the Cholesky factorization, that avoids this inconclusiveness.

5.2. Bonferroni t-Statistics

Suppose that one is interested in testing the null hypothesis

$$H_0: \alpha_i \geqq 0, \qquad i = 1,\ldots,n \quad (n \neq 1),$$

against the alternative hypothesis that at least one $\alpha_i < 0$, a natural way to proceed is to use n one-tailed t-statistics. However in order to control for the overall level of significance, the level of significance of each one-tailed test must be scaled down accordingly. If the desired overall level of significance is set at α, then the individual levels of significance α^i must satisfy the following Bonferroni inequality

$$1 - \alpha \geqq 1 - \alpha^1 - \alpha^2 - \cdots - \alpha^n.$$[42]

In the case that the α_i's are distributed independently of each other, we actually have

$$(1 - \alpha) = \prod_{i=1}^{n} (1 - \alpha^i).$$

Let $V(\hat{\alpha})$ be the estimated asymptotic variance–covariance matrix of the estimator of α, so that

$$T_i = \frac{\hat{\alpha}_i}{(V_{ii}(\hat{\alpha}))^{1/2}}, \qquad i = 1,\ldots,n,$$

is distributed asymptotically as Student's t with infinite degrees of freedom. Let $t_\infty^{\alpha/n}$ be the upper α/n percentile points of the t-distribution with infinite degrees of freedom. Then with probability greater than or equal to $(1 - \alpha)$, simultaneously,

$$0 \leqq \hat{\alpha}_i + t_\infty^{\alpha/n}(V_{ii}(\hat{\alpha}))^{1/2}, \qquad i = 1,\ldots,n.$$

We note that the distribution of Student's t with infinite degrees of freedom is precisely the uninormal distribution. Thus one can equivalently define the intervals above in terms of a uninormal distribution. For each component interval above, the level of significance is set at α/n. Equal significance levels can be abandoned, and unequal allocation substituted. Any combination of significance levels summing to α will produce the same bound α for the probability error rate. The reader is referred to Miller (1966) for further details.

Thus, one can apply the Bonferroni t statistics to construct simul-

[42]Superscripts are used to distinguish the levels of significance from the parameters.

taneous rejection regions for the monotonicity hypothesis, which for both the quadratic function (at $\mathbf{x} = [\mathbf{0}]$) and the transcendental logarithmic function (at $\mathbf{x} = [\mathbf{1}]$) amounts to $\alpha_i \geqq 0, \forall i$; and for the convexity and quasi-convexity hypotheses which for both the quadratic and transcendental logarithmic function amount to respectively $D_{ii} \geqq 0, \forall i$, and $D_{ii} \geqq 0, \forall i$, except possibly one, for the Cholesky factorization of appropriate matrices. For the generalized linear function, one can test for global monotonicity and convexity simultaneously,

$$\alpha_i \leqq 0, \qquad B_{ij} \leqq 0, \qquad i \neq j, \quad \forall i,j.$$

Finally, it should be noted that in the case of interval constraints, such as $\bar{\alpha}_i \geqq \alpha_i \geqq \underline{\alpha}_i, \forall i$, a similar Bonferroni t statistics procedure resulting in simultaneous intervals of the type

$$\bar{\alpha}_i + t_\infty^{\alpha/2n}(V_{ii}(\hat{\alpha}))^{1/2} \geqq \hat{\alpha}_i \geqq \underline{\alpha}_i - t_\infty^{\alpha/2n}(V_{ii}(\hat{\alpha}))^{1/2}, \quad \forall i,$$

will apply.

5.3. *Distribution of Extremes*

A second alternative, which is the original proposal made by Lau (1974), makes use of the theory of ordered statistics. Again consider the null hypothesis

$$H_0: \alpha_i \geqq 0, \qquad i = 1,\dots,n.$$

This null hypothesis may be transformed to the following one:

$$H_0: \min_i\{\alpha_1,\alpha_2,\dots,\alpha_n\} \geqq 0.$$

Now given a known joint distribution of α, one can presumably derive the distribution of the minimum element of α, making use of the techniques of the theory of ordered statistics.[43] Knowing the distribution of $\min_i\{\alpha_i\}$, one can then immediately construct a rejection region for the hypothesis that $\min_i\{\alpha_i\} \geqq 0$ for any given level of significance. The advantage of this test procedure is that it is exact, unlike the Bonferroni t-statistics procedure, which only gives a bound on the overall level of significance and it is in principle quite feasible. The disadvantage of this

[43]For an excellent introductory exposition to the theory of ordered statistics, see Kendall and Stuart (1969, Vol. 1, Ch. 14, pp. 325–346).

procedure, is, of course, that elaborate computations are required, even in the case that $V(\hat{\alpha})$ is diagonal.[44] In general, a multivariate normal integral must be evaluated by numerical methods as appropriate tables do not yet exist for $n \geqq 3$. To make matters worse, $V(\hat{\alpha})$ is in general unknown and only an estimate is available. Thus in computing the distribution of $\min_i\{\hat{\alpha}\}$ one will have to integrate over the distribution of $V(\hat{\alpha})$ as well. This all seems to be an extremely high price to pay just for obtaining an exact level of significance. However, this may be the only possible procedure if the Bonferroni t-statistics procedure does not give clearcut results; for example, if zero is on the boundary of the rejection region.

For the null hypothesis

$$H_0:\alpha_i \geqq 0, \quad \forall i, \quad \text{except possibly one.}$$

The appropriate transformation is the following:

$$H_0:\text{second } \underset{i}{\text{smallest}} \{\alpha_1,\alpha_2,\ldots,\alpha_n\} \geqq 0.$$

Again, this problem may in principle be solved by making use of the theory of ordered statistics. Actual computation is of course a different matter.

5.4. *Likelihood Ratio Tests*

It is well known that a null hypothesis of the type $\alpha_i = 0, i = 1,\ldots,n$, may be tested by the likelihood ratio procedure. Essentially, $-2 \ln \lambda$, where λ is the ratio of the maximized likelihood function under the null hypothesis to the unconstrained maximized likelihood function, is asymptotically distributed as a χ^2 variable with n degrees of freedom. However, the same is not true when the null hypothesis consists of inequalities rather than equality constraints. Here there is no reduction in the dimensionality of the feasible parameter space under the null hypothesis. Asymptotically, the value of the inequality constrained maximized likelihood function will converge to the value of the unconstrained maximum likelihood function and $-2 \ln \lambda$ is identically zero.

Despite this shortcoming, the use of the likelihood ratio procedure in

[44] Note that if $V(\hat{\alpha})$ is indeed diagonal, it implies that the components of $\hat{\alpha}$ are distributed independently of one another, and the Bonferroni t-statistics procedure is exact.

this context does have a certain amount of intuitive appeal. If the unconstrained estimates satisfy the constraints, the likelihood ratio will be identically one. Otherwise, it will be less than one. A rejection region may be constructed on the basis of λ. Consider for the sake of simplicity a one-dimensional example. We want to test the hypothesis of $\alpha \geqq 0$. Let $\hat{\alpha}$ be the unconstrained estimator of α. The likelihood ratio is identically one if $\hat{\alpha} \geqq 0$. The likelihood ratio is given by

$$\lambda = \frac{L(\alpha = 0)}{L(\hat{\alpha})}, \qquad \hat{\alpha} < 0,$$

which will be less than one and decreases monotonically as $\hat{\alpha}$ decreases. (Of course, in practice, we use the one-tailed t-test for this case.)

However, in order to make use of the likelihood ratio in this manner it is necessary to compute either its exact distribution or at least a finite sample approximation to the exact distribution. Unfortunately, the finite sample distribution will most likely depend on the values of both the dependent and independent variables, and will have to be computed on a case by case basis. Further research on the finite sample distribution theory of inequality-constrained estimators is needed before the likelihood ratio procedure can be fruitfully employed.

6. Conclusion

In this paper we have outlined an implementable procedure for testing the hypotheses of monotonicity, convexity and quasi-convexity of estimated functions, and to obtain parametric estimates of functions under the constraints of these hypotheses. Since these hypotheses are of fundamental importance in economic analysis, it is essential to test their validity under as unrestrictive a maintained hypothesis as possible. Functions which provide second-order numerical approximation to arbitrary functions are therefore well suited for this purpose. The availability of a simple procedure for constrained estimation to ensure monotonicity, convexity, or quasi-convexity also enhances immensely the usefulness of the new functional forms in practical economic applications. The proposed procedure may be applied to any functional form with the second-order numerical approximation property, including the class of general linear profit functions introduced by McFadden (Chapter II.2) and others yet to be invented.

The procedures considered here may be easily extended to solve other types of problems: linear inequality constraints, interval constraints, estimation of variance–covariance matrices (which must be positive semidefinite) and transition probability matrices (which must be non-negative and have column sums equal to ones), and estimation of saddle functions (convex-concave or concave-convex functions).

BIBLIOGRAPHY

Aczel, J. (1966), Lectures on Functional Equations and Their Applications, New York: Academic Press.

Afriat, S.N. (1972), "Efficiency estimation of production functions", International Economic Review, vol. 13, no. 3, pp. 568–598.

Afriat, S.N. (1968), "Test of laws of production from the operating data of firms and the construction of production functions", mimeo., Chapel Hill, NC: University of North Carolina.

Afriat, S.N. (1967), "The construction of utility functions from expenditure data", International Economic Review, vol. 8, no. 1, pp. 67–77.

Alexandrov, A.D. (1939), "Almost everywhere existence of second differentials of convex functions" (in Russian), Leningrad State University Annals, Mathematics Series 6, pp. 3–35.

Allen, R.G.D. (1938), Mathematical Analysis for economists, 2nd ed., London: Macmillan.

Amemiya, T. (1974), "The nonlinear two stage least squares estimator", Journal of Econometrics, vol. 2, no. 2, pp. 105–110.

Arrow, K.J. (1974), "The measurement of real value added", in: P. David and M. Reder, eds., Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz, New York: Academic Press.

Arrow, K.J. and A.C. Enthoven (1961), "Quasi-concave programming", Econometrica, vol. 29, no. 4, pp. 779–800.

Arrow, K.J. and F.H. Hahn (1971), General Competitive Analysis, San Francisco, CA: Holden-Day.

Arrow, K.J. and M. Hoffenberg (1959), A Time Series Analysis of Interindustry Demands, Amsterdam: North-Holland.

Arrow, K.J., B.H. Chenery, B.S. Minhas and R.M. Solow (1961), "Capital–labor substitution and economic efficiency", Review of Economics and Statistics, vol. 43, no. 3, pp. 225–250.

Atkinson, S.E. and R.F. Halvorsen (1976), "Interfuel substitution in electric power generation", Journal of Political Economy, vol. 84, no. 5, pp. 959–978.

Attiyeh, R. (1967), "Estimation of a fixed coefficient vintage model of production", Yale Economic Essays, New Haven, CT: Yale University.

Averch, H. and L. Johnson (1962), "Behavior of a firm under regulatory constraint", American Economic Review, vol. 52, no. 5, pp. 1052–1069.

Bailey, E.E. (1973), Economic Theory of Regulatory Constraint, Lexington, MA: Lexington Books.

Bailey, E.E. and R.D. Coleman (1971), "The effect of lagged regulation in an Averch–Johnson model", Bell Journal of Economics and Management Science, vol. 2, no. 1, pp. 278–292.

Bailey, E.E. and J. Malone (1970), "Resource allocation and the regulated firm", Bell Journal of Economics and Management Science, vol. 1, no. 1, pp. 129–142.

Barna, Tibor, ed. (1963), Structural Interdependence and Economic Development, New York: St. Martin's Press.

Barten, A.P. and E. Geyskens (1975), "The negativity condition in consumer demand", European Economic Review, vol. 6, no. 3, pp. 227–260.

Bartle, Robert G. (1964), The Elements of Real Analysis, New York: Wiley.

Barzel, Y. (1964), "The production function and technical change in the steam-power industry", Journal of Political Economy, vol. 72, no. 2, pp. 133–150.

Barzel, Y. (1963), "Productivity in the electric power industry, 1929–1955", Review of Economics and Statistics, vol. 45, no. 4, pp. 395–408.

Baumol, W. and A. Klevorick (1970), "Input choice and rate of return regulation: An overview of the discussion", Bell Journal of Economics and Management Science, vol. 1, pp. 162–190.

Beale, E. (1960), "Confidence regions in non-linear estimation", Journal of the Royal Statistical Society, vol. 22, p. 41.

Belinfante, A. (1969), "Technical change in the stream electric power generating industry", unpublished Ph.D. dissertation, Berkeley, CA: University of California.

Bellman, R.E. (1970), Introduction to Matrix Analysis, 2nd ed., New York: McGraw-Hill.

Ben-Israel, A. (1966), "A Newton–Raphson method for the solution of systems of equations", Journal of Mathematical Analysis and Applications, vol. 15, no. 2, pp. 243–252.

Berge, C. (1963), Topological Spaces, Edinburgh: Oliver & Boyd.

Berlinguette, V.R. and F.H. Leacy (1961), "The estimation of real domestic product by final expenditure categories and by industry of origin in Canada", in: Output, Input and Productivity Measurement, National Bureau of Economic Research, Studies in Income and Wealth, vol. 25, Princeton, NJ: Princeton University Press, pp. 303–243.

Berndt, E. and L. Christensen (1974), "Testing for the existence of a consistent aggregate index of labor inputs", American Economic Review, vol. 64, no. 3, pp. 391–404.

Berndt, E. and L. Christensen (1973a), "The translog function and the substitution of equipment, structures and labor in U.S. manufacturing, 1929–68", Journal of Econometrics, vol. 1, no. 1, pp. 81–113.

Berndt, E.R. and L.R. Christensen (1973b), "The internal structure of functional relationships: Separability, substitution and aggregation", Review of Economic Studies, vol. 40, no. 3, pp. 403–410.

Berndt, E.R. and N.E. Savin (1977), "Conflict among criteria for testing hypotheses in the multivariate linear regression model", Econometrica, vol. 45, no. 5, pp. 1263–1277.

Berndt, E.R. and D.O. Wood (1975), "Technology, prices and the derived demand for energy", The Review of Economics and Statistics, vol. 52, no. 3, pp. 259–268.

Bernstein, B. and R. Toupin (1962), "Some properties of the Hessian matrix of a strictly convex function", Journal für Mathematik, vol. 210, nos. 1/2, pp. 65–72.

Blackorby, C., D. Primont and R.R. Russell (1978), Duality, Separability and Functional Structure: Theory and Empirical Applications, New York: American Elsevier.

Blackorby, C., D. Primont and R. Russell (1977a), "On testing separability restrictions with flexible functional forms", Journal of Econometrics, vol. 5, no. 2, pp. 195–209.

Blackorby, C., D. Primont and R.R. Russell (1977b). "Dual price and quantity aggregation," Journal of Economic Theory, vol. 14, pp. 130–148.

Box, G.E.P. and D.R. Cox (1964), "An analysis of transformations", Journal of the Royal Statistical Society (Series B), vol. 26, pp. 211–243.

Bradu, D. and Y. Mundlak (1970), "Estimation in lognormal linear models", Journal of the American Statistical Association, vol. 65, no. 329, pp. 198–211.

Bridge, J.L. (1971), Applied Econometrics, Amsterdam: North-Holland.

Brooks, J. (1963), The Fate of the Edsel and Other Business Adventures, 1st ed., New York: Harper & Row.

Brown, M. and J. de Cani (1963), "Technical change and the distribution of income", International Economic Review, vol. 4, no. 3, pp. 289–309.

Brown R.S., D.W. Caves and L.R. Christensen (1975), "Modelling the structure of production with a joint cost function", SSRI W.P. 7521, Madison, WI: University of Wisconsin.

Bruno, M. (1973), "Protection and tariff change under general equilibrium", Journal of International Economics, vol. 3, no. 3, pp. 205–226.
Bruno, M. (1969), "Fundamental duality relations in the pure theory of capital growth", Review of Economic Studies, vol. 36, no. 1, pp. 39–53.
Bruno, M. (1968), "Estimation of factor contribution to growth under structural disequilibrium", International Economic Review, vol. 9, no. 1, pp. 49–62.
Burgess, D. (1976), "Tariffs and income distribution: Some empirical evidence for the United States", Journal of Political Economy, vol. 84, no 1, pp. 17–45.
Burgess, D. (1975), "Duality theory and pitfalls in the specification of technologies", Journal of Econometrics, vol. 3, no. 2, pp. 105–121.
Burmeister, E. and K. Kuga (1970), "The factor-price frontier, duality and joint production", Review of Economics Studies, vol. 37, no. 1, pp. 11–19.
Busemann, H. (1958), Convex Surfaces, New York: Interscience.
Busemann, H. and W. Feller (1935–36), "Bemerkungen zur Differentialgeometrie der konvexen Flächen" (3 parts), Mat. Tidsskr. (Series B), Part I (1935), nos. 1–2, pp. 24–36; Part II (1935), no. 4, pp. 87–114; Part III (1935), no. 3, pp. 41–70.
Cameron, Burgess (1952), "The production function in Leontief models", Review of Economic Studies, vol. 20, no. 1, pp. 62–69.
Carlson, S. (1939), A Study in the Pure Theory of Production, London: P.S. King & Son.
Carter, Anne P. (1970), Structural Change in the American Economy, Cambridge, MA: Harvard University Press.
Carter, Anne P. and Andrew Brody, eds. (1970a), Contributions to Input–Output Analysis, vol. 1, Amsterdam: North-Holland.
Carter, Anne P. and Andrew Brody, eds. (1970b), Applications of Input–Output Analysis, vol. 2, Amsterdam: North-Holland.
Cass, D. (1974), "Duality: A symmetric approach from the economist's vantage point", Journal of Economic Theory, vol. 7, no. 3, pp. 272–295.
Cavallo, D.F. (1976), "A note on consistent estimation of a production function with partially transmitted errors", Department of Economics mimeo., Cambridge, MA: Harvard University.
Chow, G.C. (1960), "Tests of equality between sets of coefficients in two linear regressions", Econometrica, vol. 28, no. 3, pp. 591–605.
Christensen, L.R., D.W. Jorgenson and L.J. Lau (1975), "Transcendental logarithmic utility functions", American Economic Review, vol. 65, no. 3, pp. 367–383.
Christensen, L.R., D.W. Jorgenson and L.J. Lau (1973), "Transcendental logarithmic production frontiers", Review of Economics and Statistics, vol. 55, no. 1, pp. 28–45.
Christensen, L.R., D.W. Jorgenson and L.J. Lau (1971), "Conjugate duality and the transcendental logarithmic functions", Econometrica, vol. 39, no. 4, pp. 255–256 (abstract).
Christensen, L.R. and D.W. Jorgenson (1970), "U.S. real product and real factor input, 1929–67", Review of Income and Wealth, vol. 16, no. 1, pp. 19–50.
Christensen, L.R. and D.W. Jorgenson (1969), "The measurement of U.S. real capital input, 1929–67", Review of Income and Wealth, vol. 15, no. 4, pp. 293–320.
Corden, W.M. (1969), "Effective protective rates in the general equilibrium model: A geometrical note", Oxford Economic Papers, vol. 21, no. 2, pp. 135–141.
Courant, R. (1937), Differential and Integral Calculus, vol. I, 2nd ed., New York: Interscience.
Courant, R. (1936), Differential and Integral Calculus, vol. II, 1st ed., New York: Nordeman.
Courant, R. and H. Hilbert (1953), Methods of Mathematical Physics, vols. I and II, New York: Interscience.
Courville, L. (1974), "Regulation and efficiency in the electric utility industry", Bell Journal of Economics and Management Science, vol. 5, no. 1, pp. 53–74.

Cowing, T.G. (1974), "Technical change and scale economies in an engineering production function: The case of steam electric power", Journal of Industrial Economics, vol. 23, no. 2, pp. 135–152.

Cowing, T.G. (1970), "Technical change in steam-electric generation: An engineering approach", unpublished Ph.D. dissertation, Berkeley, CA: University of California.

Danskin, J.M. (1967), The Theory of Max–Min, New York: Springer.

Dantzig, G.B. (1949), "Programming of interdependent activities, Part II: Mathematical model", Econometrica, vol. 17, no. 3–4, pp. 200–211.

David, Paul A. (1966), "Measuring real net ourput: A proposed index", Review of Economics and Statistics, vol. 48, no. 4, pp. 419–425.

David, Paul A. (1962), "The deflation of value added", Review of Economics and Statistics, vol. 44, no. 2, pp. 148–155.

David, P. and Th. Van de Klundert (1965), "Biased efficiency growth and capital–labor substitution in the U.S., 1899–1960", American Economic Review, vol. 55, no. 3, pp. 356–394.

Davis, E.G. (1973), "A dynamic model of the regulated firm with a price adjustment mechanism", Bell Journal of Economics and Management Science, vol. 4, no. 1, pp. 270–282.

Debreu, G. (1959), Theory of Value, New York: Wiley.

Debreu, G. (1952), "A social equilibrium existence theorem", Proceedings of the National Academy of Sciences of the U.S.A., vol. 38, no. 10, pp. 886–893.

Denny, M. (1974a), "The relationship between functional forms for the production system", Canadian Journal of Economics, vol. 7, no. 1, pp. 21–31.

Denny, M. (1974b), "On the sensitivity of productivity measures in Canadian manufacturing", mimeo., Toronto: University of Toronto.

Denny, M. (1972), "Trade and the production sector: An exploration of models of multi-product technologies", unpublished Ph.D. dissertation, Berkeley, CA: University of California.

Denny, M. (1971), "Multiple outputs and inputs in aggregate production models: Imports and the Canadian economy", mimeo., Toronto: University of Toronto.

Denny, M. (1970), "Notes on the specification of technologies", Institute for Policy Analysis Working Paper 7008, Toronto: University of Toronto.

Denny, M. and M. Fuss (1977), "The use of approximation analysis to test for separability and the existence of consistent aggregates", American Economic Review, vol. 67, no. 3, pp. 404–418.

Denny, M. and J.D. May (1978), "A representation of Canadian manufacturing technology", Applied Economics, forthcoming.

Dent, W.T. (1973), "A note on least squares fitting of functions constrained to be either non-negative, non-decreasing or convex", Management Science, vol. 20, pp. 130–132.

Dhrymes, P.J. (1967), "On a class of utility and production functions yielding everywhere differentiable demand functions", Review of Economic Studies, vol. 34(4), no. 100, pp. 399–408.

Dhrymes, P.J. (1963), "A comparison of productivity behavior in manufacturing and service industries", Review of Economics and Statistics, vol. 45, no. 1, pp. 64–69.

Dhrymes, P.J. and M. Kurz (1964), "Technology and scale in electricity generation", Econometrica, vol. 32, no. 3, pp. 287–315.

Diamond, P.A. (1965), "Disembodied technical change in a two sector model", Review of Economic Studies, vol. 32(2), no. 90, pp. 161–168.

Diewert, W.E. (1976), "Exact and superlative index numbers", Journal of Econometrics, vol. 4, no. 2, pp. 115–146.

Diewert, W.E. (1974a), "Applications of duality theory", in: M.D. Intriligator and D.A. Kendrick, eds., Frontiers of Quantitative Economics, vol. II, Amsterdam: North-Holland, pp. 106–171.

Diewert, W.E. (1974b), "Functional forms for revenue and factor requirement functions", International Economic Review, vol. 15, no. 1, pp. 119–130.

Diewert, W.E. (1974c), "A note on aggregation and elasticities of substitution", Canadian Journal of Economics, vol. 7, no. 1, pp. 12–20.

Diewert, W.E. (1974d), "Afriat and revealed preference theory", Review of Economic Studies, vol. 40(3), no. 123, pp. 419–425.

Diewert, W.E. (1973a), "Functional forms for profit and transformation functions", Journal of Economic Theory, vol. 6, no. 3, pp. 284–316.

Diewert, W.E. (1973b), "Separability and a generalization of the Cobb–Douglas cost, production, and indirect utility functions", Institute for Mathematical Studies in the Social Sciences Technical Report no. 86, Stanford, CA: Stanford University.

Diewert, W.E. (1971), "An application of the Shephard duality theorem, a generalized Leontief production function", Journal of Political Economy, vol. 79, no. 3, pp. 481–507.

Diewert, W.E. (1969a), "Canadian labor markets: A neo-classical econometric approach", Project for the Evaluation and Optimization of Economic Growth, Institute of International Studies Technical Report no. 20, Berkeley, CA: University of California.

Diewert, W.E. (1969b), "Functional form in the theory of production and consumer demand", unpublished Ph.D. thesis, Berkeley, CA: University of California.

Domar, E. (1961), "On the measurement of technological change", Economic Journal, vol. 71, no. 284, pp. 709–729.

Dorfman, R., P.A. Samuelson and R.M. Solow (1958), Linear Programming and Economic Analysis, New York: McGraw-Hill.

Douglas, P.C. and C.W. Cobb, (1928), "A theory of production", American Economic Review, vol. 18, no. 1, pp. 139–165.

Dunford, N. and J. Schwartz (1958), Linear Operators, vol. I, New York: Interscience.

Eisenpress, H. (1968), "Non-linear equations and systems, estimation and prediction (IBM S/360)", 360D 13.6.005.

Eisenpress, H. and J. Greenstadt (1966), "The estimation of nonlinear econometric systems", Econometrica, vol. 34, no. 4, pp. 851–861.

Electrical World Steam Station Cost Survey: Electrical World, biannual (odd numbered years), usually in October.

Epstein, L. (1975), "A disaggregated analysis of consumer choice under uncertainty", Econometrica, vol. 43, nos. 5–6, pp. 877–892.

Epstein, L. (1973a), "Consumer behaviour under uncertainty", Research Projects Group mimeo., Ottawa: Department of Manpower and Immigration.

Epstein, L. (1973b), "Duality and optimal consumption under uncertainty", Research Projects Group mimeo., Ottawa: Department of Manpower and Immigration.

Fan, K. (1959), Convex Sets and Their Applications, Argonne, IL: Argonne National Laboratory.

Farrell, M.J. (1957), "The measurement of productive efficiency", Journal of Royal Statistical Society (Series A), vol. 120, no. 3, pp. 253–290.

Federal Power Commission (1964), National Power Survey.

Federal Power Commission (a), Statistics of Electric Utilities in the United States, Classes A and B Privately Owned Companies, annual.

Federal Power Commission (b), Steam-Electric Plant Construction Cost and Annual Production Expenses, annual.

Fenchel, W. (1953), Convex Cones, Sets, and Functions, Lecture Notes, Princeton, NJ: Princeton University.

Fenchel, W. (1949), "On conjugate convex functions", Canadian Journal of Mathematics, vol. 1, no. 1, pp. 73–77.

Ferguson, C. (1963), "Cross-section production functions and the elasticity of substitution in American manufacturing industry", Review of Economics and Statistics, vol. 45, no. 3, pp. 305–313.

Finsler, P. (1936–37), "Über das Verkommen definiter und semidefiniter Formen in Scharen quadratischer Formen," Commentarii Mathematicii Helveticii, vol. 9, pp. 188–192.

Forsyth, A.R. (1906), Theory of Differential Equations, vol. I–VI, Cambridge: The University Press.

Frenger, P. (1975), "Factor substitution and inconsistent aggregation", unpublished Ph.D. dissertation and Center for Research in Management Science Working Paper IP-238, Berkeley, CA: University of California.

Frisch, R. (1965), The Theory of Production, Chicago, IL: Rand-McNally.

Frisch, R. (1931), "Einige Punkte einer Preistheorie mit Boden und Arbeit als Produktionsfaktoren", Zeitschrift für Nationalökonomie, vol. 3, no. 1, pp. 62–104.

Fuss, M.A. (1977a), "The demand for energy in Canadian manufacturing: An example of the estimation of production functions with many inputs", Journal of Econometrics, vol. 5, no. 1, pp. 89–116.

Fuss, M.A. (1977b), "The structure of technology over time: A model for testing the 'putty-clay' hypothesis", Econometrica, vol. 45, no. 8, pp. 1797–1821.

Fuss, M.A. (1970), "The time structure of technology: An empirical analysis of the 'putty-clay' hypothesis", unpublished Ph.D. dissertation, Berkeley, CA: University of California.

Galatin, M. (1968), Economies of Scale and Technological Change in Thermal Power Generation, in: Contributions to Economic Analysis, vol. 53, Amsterdam: North-Holland.

Gale, D. (1960), The Theory of Linear Economic Models, New York: McGraw-Hill.

Gale, D., V. Klee and T. Rockafellar (1968), "Convex functions on convex polytopes", Proceedings of the American Mathematical Society, vol. 19, no. 4, pp. 867–873.

Gale, D. and H. Nikaido (1965), "The Jacobian matrix and global univalence of mappings", Mathematische Annalen, vol. 159, no. 2, pp. 81–93.

Gantmacher, F.R. (1959), The Theory of Matrices, vols. I and II, New York: Chelsea.

Geary, R.C. (1944), "The concept of net volume of output, with special reference to Irish data", Journal of the Royal Statistical Society, N.S., vol. 107, pp. 251–259.

Georgescu-Roegen, N. (1935), "Fixed coefficients of production and the marginal productivity theory", Review of Economic Studies, vol. 3, no. 1, pp. 40–49.

Goldman, S.M. and H. Uzawa (1964), "A note on separability in demand analysis", Econometrica, vol. 32, no. 3, pp. 387–399.

Gorman, W.M. (1970), Personal communication.

Gorman, W.M. (1968a), "Measuring the quantities of fixed factors", in: J.N. Wolfe, ed., Value, Capital and Growth: Papers in Honour of Sir John Hicks, Chicago, IL: Aldine.

Gorman, W.M. (1968b), "Conditions for additive separability", Econometrica, vol. 36, nos. 3–4, pp. 605–609.

Gorman, W.M. (1965), "Production functions in which the elasticities of substitution stand in fixed proportion to each other", Review of Economic Studies, vol. 32(3), no. 91, pp. 217–224.

Gorman, W.M. (1953), "Community preference fields", Econometrica, vol. 21, pp. 63–80.

Griliches, Z. and V. Ringstad (1971), Economies of Scale and the Form of the Production Function, Amsterdam: North-Holland.

Grunbaum, B. (1967), Convex Polytopes, New York: Interscience.

Hadley, G. (1964), Nonlinear and Dynamic Programming, Reading, MA: Addison-Wesley.

Haldi, John (1959), "A test on two hypotheses concerning interpretation of the input–output coefficients", Weltwirtschaftliches Arkiv, vol. 83, no. 1, pp. 1–14.

Hall, R. E. (1973), "The specification of technology with several kinds of output", Journal of Political Economy, vol. 81, no. 4, pp. 878–892.

Hall, R. E. (1968), "Technical change and capital from the point of view of the dual",

Review of Economic Studies, vol. 35(1), no. 101, pp. 35–46.
Hall, R. and D. Jorgenson (1967), "Tax policy and investment behavior", American Economic Review, vol. 57, no. 3, pp. 391–414.
Halperin, M. (1963), "Confidence interval estimation in non-linear regression", Journal of the Royal Statistical Society (Series B), vol. 25, no. 2, pp. 330–333.
Halter, A.N., H.O. Carter, and J.G. Hocking (1957), "A note on transcendental production functions", Journal of Farm Economics, vol. 39, no. 4, pp. 966–974.
Hanoch, G. (1977), "Risk aversion and consumer preferences", Econometrica, vol. 45, no. 2, pp. 413–426.
Hanoch, G. (1975a), "Production and demand models with direct or indirect implicit additivity", Econometrica, vol. 43, no. 3, pp. 395–420.
Hanoch, G. (1975b), "The elasticity of scale and the shape of average costs", American Economic Review, vol. 65, no. 3, pp. 492–497.
Hanoch, G. (1971), "CRESH production functions", Econometrica vol. 39, no. 5, pp. 695–712.
Hanoch, G. (1970), "Homotheticity in joint production", Journal of Economic Theory, vol. 2, no. 4, pp. 423–426.
Hanoch, G. and M. Rothschild (1972), "Testing the assumptions of production theory: A nonparametric approach", The Journal of Political Economy, vol. 80, no. 2, pp. 256–275.
Harberger, A. (1959), "The corporation income tax: An empirical appraisal", in: Committee on Ways and Means, Tax Revision Compendium, vol. 1, Washington, DC: U.S. Congress House.
Heady, E.O. and J.L. Dillon (1961), Agricultural Production Functions, Ames, IA: Iowa State University Press.
Hicks, J.R. (1963), The Theory of Wages, 2nd ed., London: Macmillan.
Hicks, J.R. (1956), A Revision of Demand Theory, 1st ed., Oxford: Clarendon Press.
Hicks, J.R. (1946), Value and Capital, 2nd ed., Oxford: Clarendon Press.
Hicks, J.R. (1932), The Theory of Wages, 1st ed., London: Macmillan.
Hildreth, C. (1954), "Point estimates of ordinates of concave functions", Journal of the American Statistical Association, vol. 49, pp. 598–619.
Hoch, I. (1962), "Estimation of production function parameters combining time and cross-section data", Econometrica, vol. 30, no. 1, pp. 34–53.
Hoch, I. (1958), "Simultaneous equations bias in the context of the Cobb–Douglas production function", Econometrica, vol. 28, no. 4, pp. 566–578.
Holthausen, D.M. (1974), "The firm under regulation: An analysis of uncertainty, incentives and regulatory lag", unpublished Ph.D. dissertation, Evanston, IL: Northwestern University.
Hotelling, H. (1935), "Demand functions with limited budgets", Econometrica, vol. 3, no. 1, pp. 66–78.
Hotelling, H. (1932), "Edgeworth's taxation paradox and the nature of demand and supply functions", Journal of Political Economy, vol. 40, no. 5, pp. 577–616.
Householder, A.S. (1964), The Theory of Matrices in Numerical Analysis, New York: Blaisdell.
Houthakker, H.S. (1965), "A note on self-dual preferences", Econometrica, vol. 33, no. 4, pp. 797–801.
Houthakker, H.S. (1960), "Additive preferences", Econometrica, vol. 28, no. 2, pp. 244–257.
Houthakker, H.S. (1950), "Revealed preference and the utility function", Economica, N.S., vol. 17, no. 66, pp. 159–174.
Hudson, D.J. (1969), "Least-squares fitting of a polynomial constrained to be either non-negative, non-decreasing or convex", Journal of the Royal Statistical Society (Series B), vol. 31, no. 1, pp. 113–118.
Hudson, D.J. and D.W. Jorgenson (1974), "U.S. energy policy and economic growth: 1975–2000", Bell Journal of Economics, vol. 5, no. 2, pp. 461–514.

International Monetary Fund: International Financial Statistics, supplement to 1966/67 issue, Washington, DC: IMF.

Intriligator, M.D. (1971), Mathematical Optimization and Economic Theory, Englewood Cliffs, NJ: Prentice-Hall.

Jacobsen, S.E. (1972), "On Shephard's duality theorem", Journal of Economic Theory, vol. 4, no. 3, pp. 458–464.

Jacobsen, S.E. (1970), "Production correspondences", Econometrica, vol. 38, no. 5, pp. 754–770.

Jensen, J.L.W.V. (1906), "Sur les functions convexes et les inegalities entre les valeurs moyennes", Acta Mathematica, vol. 30, pp. 175–193.

Jorgenson, D.W. (1967), "The theory of investment behavior", in: National Bureau of Economic Research, Determinants of Investment Behavior, New York: Columbia University Press.

Jorgenson, D.W. (1966a), "Rational distributed lag functions", Econometrica, vol. 34, no. 1, pp. 135–149.

Jorgenson, D.W. (1966b), "The embodiment hypothesis", Journal of Political Economy, vol. 74, no. 1, pp. 1–17.

Jorgenson, D.W. (1963), "Capital theory and investment behavior", American Economic Review, vol. 53, no. 2, pp. 247–259.

Jorgenson, D.W. and L.J. Lau (forthcoming), Duality and Technology, Amsterdam: North-Holland.

Jorgenson, D.W. and L.J. Lau (1975a), "The structure of consumer preferences", Annals of Economic and Social Measurement, vol. 4, no. 1, pp. 49–102.

Jorgenson, D.W. and L.J. Lau (1975b), "The integrability of consumer demand functions", Institute of Economic Research Discussion Paper no. 425, mimeo., Cambridge, MA: Harvard University.

Jorgenson, D.W. and L.J. Lau (1974a), "The duality of technology and economic behaviour", Review of Economic Studies, vol. 41(2), no. 126, pp. 181–200.

Jorgenson, D.W. and L.J. Lau (1974b), "Duality and differentiability in production", Journal of Economic Theory, vol. 9, no. 1, pp. 23–42.

Jorgenson, D.W. and Z. Griliches (1967), "The explanation of productivity change", Review of Economic Studies, vol. 34(3), no. 99, pp. 249–283.

Joskow, P.L. (1974), "Inflation and environmental concern: Structural change in the process of public utility price regulation", Journal of Law and Economics, vol. 17, no. 2, pp. 291–327.

Joskow, P.L. (1973), "Pricing decisions of regulated firms: A behavioral approach", Bell Journal of Economics and Management Science, vol. 4, no. 1, pp. 118–140.

Joskow, P.L. (1972), "The determination of the allowed rate of return in a formal regulatory hearing", Bell Journal of Economics and Management Science, vol. 3, no. 2, pp. 632–644.

Judge, G.C. and T. Takayama (1966), "Inequality restrictions in regression analysis", The Journal of the American Statistical Association, vol. 66, no. 313, pp. 166–181.

Kahn, A.E. (1970), The Economics of Regulation: Principles and Institutions, vol. I, New York: Wiley.

Kakwani, N.C. (1967), "The unbiasedness of Zellner's seemingly unrelated regression equations estimators", Journal of the American Statistical Association, vol. 62, no. 317, pp. 141–142.

Karlin, S. (1959), Mathematical methods and theory in games, programming and economics, vol. I: Matrix games, programming and mathematical economics, Reading, MA: Addison-Wesley.

Katzner, D.W. (1970), Static Demand Theory, New York: Macmillan.

Kelley, J. and I. Namiuka (1963), Linear topological spaces, Princeton, NJ: Van Nostrand.

Kendall, M.G. and A. Stuart (1969), The Advanced Theory of Statistics, vol. 1: Distribution theory, 3rd ed., New York: Hafner.

Kendall, M.G. and A. Stuart (1967), The Advanced Theory of Statistics, vol. 2: Inference and relationship, 2nd ed., New York: Hafner.

Kendrick, J. and R. Sato (1963), "Factor prices, productivity, and economic growth", American Economic Review, vol. 53, no. 5, pp. 974–1003.

Khang, C. (1971), "An isovalue locus involving intermediate goods and its applications to the pure theory of international trade", Journal of International Economics, vol. 1, no. 3, pp. 315–325.

Klee, V. (1969), "Separation and support properties of convex sets – a survey", Mathematical Note no. 599, Mathematics Research Laboratory, Boeing Laboratories.

Klee, V., ed. (1963), Convexity, American Mathematical Society Proceedings of the Symposium on Pure Mathematics, vol. 7.

Klein, Lawrence R. (1953), A Textbook of Econometrics, Evanston, IL: Row, Peterson, and Co.

Klein, Lawrence R. (1952–1953), "On the interpretation of Professor Leontief's system", Review of Economic Studies, vol. 20(2), no. 52, pp. 131–136.

Klevorick, A.K. (1973), "The behavior of a firm subject to stochastic regulatory review", Bell Journal of Economics and Management Science, vol. 4, no. 1, pp. 57–88.

Kmenta, J. (1967), "On estimation of the CES production function", International Economic Review, vol. 8, no. 2, pp. 180–189.

Komiya, R. (1962), "Technical progress and the production function in the United States steam power industry", Review of Economics and Statistics, vol. 44, no. 2, pp. 156–166.

Konyus, A.A. and S.S. Byushgens (1926), "*K* probleme pokupatelnoi cili deneg", Voprosi Konyunkturi, vol. 2, no. 1, pp. 151–172.

Koopmans, T.C. (1957), "Allocation of resources and the price system", in: T.C. Koopmans, Three Essays on the State of Economic Science, New York: McGraw-Hill.

Koopmans, T.C. (1951), "Analysis of production as an efficient combination of activities", in: T.C. Koopmans, ed., Activity Analysis of Production and Allocation, New York: Wiley.

Koopmans, T.C., P.A. Diamond and R.E. Williamson (1964), "Stationary utility and time perspective", Econometrica, vol. 32, no. 1, pp. 82–100.

Kurz, M. and A. Manne (1963), "Engineering estimates of capital–labor substitution in metal machining", American Economic Review, vol. 53, no. 4, pp. 662–681.

Lanczos, C. (1966), The Variational Principles of Mechanics, 3rd ed., Toronto: University of Toronto Press.

Lau, L.J. (1977), "Empirical approaches to systems of consumer demand functions through duality", in: M.D. Intriligator, ed., Frontiers of Quantitative Economics, vol. III, Amsterdam: North-Holland.

Lau, L.J. (1976a), "A characterization of the normalized restricted profit function", Journal of Economic Theory, vol. 12, no. 1, pp. 131–163.

Lau, L.J. (1976b), "A note on elasticity of substitution functions", Review of Economic Studies, vol. 43, no. 134, pp. 353–358.

Lau, L.J. (1975), "Statistical determination of functional structure: An expository note", Department of Economics Working Paper, mimeo., Stanford, CA: Stanford University.

Lau, L.J. (1974), "Comments on applications of duality theory", in: M.D. Intriligator and D.A. Kandrick, eds., Frontiers of Quantitative Economics, vol. II, Amsterdam: North-Holland, pp. 176–199.

Lau, L.J. (1972), "Profit functions of technologies with multiple inputs and outputs", Review of Economics and Statistics, vol. 54, no. 3, pp. 281–289.

Lau, L.J. (1969a), "Duality and the structure of utility functions", Journal of Economic Theory, vol. 1, no. 4, pp. 374–396.

Lau, L.J. (1969b), "Duality and utility structure", unpublished Ph.D. dissertation, Berkeley, CA: University of California.

Lau, L.J. (1969c), "Some applications of profit functions", Center of Research in Economic Growth Memoranda nos. 86A and 86B, mimeo., Stanford, CA: Stanford University.

Lau, L.J. and P.A. Yotopoulos (1971), "A test for relative efficiency and an application to Indian agriculture", American Economic Review, vol. 61, no. 1, pp. 94–109.

Lehmann, E.L. (1959), Testing Statistical Hypotheses, New York: Wiley.

Leontief, W.W. (1951), The Structure of the American Economy, 1919–1939, 2nd ed., New York: Oxford University Press.

Leontief, W.W. (1947), "A note on the interrelations of subsets of independent variables of a continuous function with continuous first derivatives", Bulletin of the American Mathematical Society, vol. 53, no. 4, pp. 343–350.

Leontief, W.W. (1936), "Composite commodities and the problem of index numbers", Econometrica, vol. 4, no. 1, pp. 39–59.

Liew, C.K. (1976), "Inequality constrained least squares estimation", Journal of American Statistical Association, vol. 71, pp. 746–751.

Ling, S. (1964), Economies of Scale in the Steam-Electric Power Generating Industry, Amsterdam: North-Holland.

Litzenberger, R. and C. Rao (1971), "Estimates of the marginal rate of time preference and average risk aversion of investors in electric utility shares: 1960–1966", Bell Journal of Economics and Management Science, vol. 2, no. 1, pp. 265–277.

Lucas, R. (1963), "Substitution between labor and capital in U.S. manufacturing; 1929–1958", unpublished Ph.D. dissertation, Chicago, IL: University of Chicago.

McFadden, D. (1967), Book review of Hildebrand–Liu, Journal of American Statistical Association, vol. 62, no. 317, pp. 395–399.

McFadden, D. (1966), "Cost, revenue and profit functions: A cursory review", Institute for Business and Economic Research Working Paper no. 86, Berkeley, CA: University of California.

McFadden, D. (1964), "Estimation techniques for the elasticity of substitution and other production parameters", Paper read at the Meeting of the Econometric Society, 1964 (Chapter IV.1).

McFadden, D. (1963), "Further results on C.E.S. production functions", Review of Economic Studies, vol. 30(2), no. 83, pp. 73–83.

McFadden, D. (1962), "Factor substitutability in the economic analysis of production", unpublished Ph.D. dissertation, Minneapolis, MN: University of Minnesota.

McKenzie, L. (1960), "Matrices with dominant diagonals and economic theory", in: K.J. Arrow et al., eds., Proceedings of a Symposium on Mathematical Methods in the Social Sciences, Stanford, CA: Stanford University Press.

McKenzie, L. (1957), "Demand theory without a utility index", Review of Economic Studies, vol. 24(3), no. 65, pp. 185–189.

McKinnon, R. (1963a), "The C.E.S. production function applied to two-digit manufacturing and three mining industries for the United States", mimeo., Stanford, CA: Stanford University.

McKinnon, R. (1963b), "Some difficulties in the use of value added production functions", mimeo., Stanford, CA: Stanford University.

Malinvaud, E. (1970), Statistical Methods of Econometrics, 2nd ed., Amsterdam: North-Holland.

Malinvaud, E. (1966), Statistical Methods of Econometrics, Chicago, IL: Rand-McNally.

Mangasarian, O.L. (1969), Nonlinear Programming, New York: McGraw-Hill.

Marschak, J. and W. Andrews (1944), "Random simultaneous equations and the theory of production", Econometrica, vol. 12, nos. 3–4, pp. 143–205.

Metcalf, L. and V. Reinemer (1967), Overcharge, New York: David McKay.

Miller, M.H. and F. Modigliani (1966), "Some estimates of the cost of capital to the electric utility industry, 1954–1957", American Economic Review, vol. 56, no. 3, pp. 333–391.
Miller, Jr., R.G. (1966), Simultaneous Statistical Inference, New York: McGraw-Hill.
Minasian, J. (1961), "Elasticities of substitution and constant output demand curves for labor", Journal of Political Economy, vol. 69, no. 3, pp. 261–270.
Minhas, B. (1963), An International Comparison of Factor Costs and Factor Use, Amsterdam: North-Holland.
Monk, K. (1976), Personal correspondence.
Moreau, J.J. (1965), "Sous-differentiabilité", in: W. Fenchel, ed., Proceedings of the Colloquium of Convexity, Copenhagen: Matematisk Institut, pp. 185–201.
Mukerji, V. (1963), "Generalized SMAC function with constant ratios of elasticities of substitution", Review of Economic Studies, vol. 30(3), no. 84, pp. 233–236.
Mundlak, Y. (1973a), "Functional forms of production functions: A survey", mimeo., Washington, DC: Development Research Center of the World Bank.
Mundlak, Y. (1973b), "Estimation of production functions: A survey", mimeo., Washington, DC: Development Research Center of the World Bank.
Mundlak, Y. (1968a), "Cross-section analysis", in: D.L. Siels, ed., International Encyclopedia of the Social Sciences, New York: Macmillan.
Mundlak, Y. (1968b), "Elasticities of substitution and the theory of derived demand", Review of Economic Studies, vol. 35(2), no. 102, pp. 225–236.
Mundlak, Y. (1964), "Transcendental multiproduct production functions", International Economic Review, vol. 5, no. 3, pp. 273–284.
Mundlak, Y. (1963), "Estimation of production and behavioral functions from a combination of cross-section and time series data", in: C. Christ et al., eds., Measurement in Economics, Stanford, CA: Stanford University Press.
Mundlak, Y. (1961), "Empirical production function free of management bias", Journal of Farm Economics, vol. 43, no. 1, pp. 44–56.
Mundlak, Y. and I. Hoch (1965), "Consequences of alternative specifications in estimation of the Cobb–Douglas functions", Econometrica, vol. 33, no. 4, pp. 814–828.
Mundlak, Y. and A. Razin (1971), "On multi-stage, multi-product production functions", American Journal of Agricultural Economics, vol. 53, no. 2, pp. 491–499.
Mundlak, Y. and A. Razin (1969), "On aggregation index numbers and the measurement of technical change", Review of Economics and Statistics, vol. 51, no. 2, pp. 166–175.
Mundlak, Y. and Z. Vulcani (1973), "The correspondence efficiency frontier as a generalization of the cost function", International Economic Review, vol. 14, no. 1, pp. 223–233.
Murray, W., ed. (1972), Numerical Methods for Unconstrained Optimization, New York: Academic Press.
Myers, S.C. (1973), "A simple model of firm behavior under regulation and uncertainty", Bell Journal of Economics and Management Science, vol. 4, no. 1, pp. 304–315.
Nerlove, M. (1967), "Recent empirical studies of the CES and related production functions", in: M. Brown, ed., The Theory and Empirical Analysis of Production, New York: Columbia University Press.
Nerlove, M. (1963), "Returns to scale in electricity supply", in: C.F. Christ et al., eds., Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld, Stanford, CA: Stanford University Press.
Newman, P. (1969), "Some properties of concave functions", Journal of Economic Theory, vol. 1, no. 3, pp. 291–314.
Newman, P. (1965), The Theory of Exchange, Englewood Cliffs, NJ: Prentice-Hall.
Nikaido, H. (1968), Convex Structures and Economic Theory, New York: Academic Press.
Nikaido, H. (1954), "On von Neumann's minimax theorem", Pacific Journal of Mathematics, vol. 4, no. 1, pp. 65–72.
Norway, Statistisk Sentralbyraa (1965), Nasjonalregnskap med fjorten og fem produksjonssektorer, 1949–61, Central Bureau of Statistics: "National accounts classified by

fourteen and five industrial sectors, 1949–61". Vol. 1: Norges Offisielle Statistikk A 116 (1965). Vol. 2: Norges Offisielle Statistikk A 159 (1966).
Norway, Statistisk Sentralbyraa (1954–61), Kredittmakedstatistikk.
Ohta, M. (1975), Production technologies of the U.S. boiler turbo-generator industries and hedonic price indexes for their products", Journal of Political Economy, vol. 83, no. 1, pp. 1–26.
Oi, W.Y. (1961), "The desirability of price instability under perfect competition", Econometrica, vol. 29, no. 1, pp. 58–64.
Parks, R.W. (1971), "Price responsiveness of factor utilization in Swedish manufacturing, 1870–1950", The Review of Economics and Statistics, vol. 53, no. 2, pp. 129–139.
Parthasarathy, T. (1962), Selection Theorems and Their Applications, Lecture Notes in Mathematics no. 263, Berlin: Springer-Verlag.
Petersen, H.C. (1975), "An empirical test of regulatory effects", Bell Journal of Economics, vol. 6, no. 1, pp. 111–126.
Phillips, C.F. (1969), The Economics of Regulation, Homewood, IL: R.D. Irwin.
Pollak, R.A. (1972), "Generalized separability", Econometrica, vol. 40, no. 3, pp. 431–453.
Pollak, R.A. (1969), "Conditional demand functions and consumption theory", Quarterly Journal of Economics, vol. 83, no. 1, pp. 60–78.
Popoviciu, T. (1945), Les Fonctions Convexes, Paris: Hermann.
Powell, A.A. and F.H.G. Gruen (1968), "The constant elasticity of transformation frontier and linear supply system", International Economic Review, vol. 9, no. 3, pp. 315–328.
Raiffa, H. and R. Schlaifer (1968), Applied Statistical Decision Theory, Cambridge, MA: M.I.T. Press.
Rao, C.R. and S.K. Mitra (1971), Generalized Inverse of Matrices and Its Applications, New York: Wiley.
Reidemeister, K. (1921), "Uber die singulären Randpunkte eines konvexen Körpers", Mathematische Annalen, vol. 83, pp. 116–118.
Revankar, N. (1971), "A class of variable elasticity of substitution production functions", Econometrica, vol. 39, no. 1, pp. 61–72.
Roberts, A.W. and D.E. Varberg (1973), Convex Functions, New York: Academic Press.
Rockafellar, R.T. (1972), "State constraints in convex control problems of Bolza", SIAM Journal of Control, vol. 10, no. 4, pp. 691–715.
Rockafellar, R.T. (1970), Convex Analysis, Princeton, NJ: Princeton University Press.
Rockafellar, R.T. (1969), "Convex functions and duality in optimization problems and dynamics", in: H.W. Kuhn and G.P. Szego, eds., Mathematical Systems Theory and Economics, Berlin: Springer-Verlag.
Rockafellar, R.T. (1966), "Characterization of the subdifferentials of convex functions", Pacific Journal of Mathematics, vol. 17, no. 3, pp. 497–510.
Rosenlicht, M. (1968), Introduction to Analysis, Glenview, IL: Scott & Foresman.
Rothenberg, T.J. (1973), Efficient Estimation with A Priori Information, New Haven, CT: Yale University Press.
Rothenberg, T.J. (1966), "Econometric estimation with a priori information", unpublished Ph.D. dissertation, Cambridge, MA: M.I.T. Press.
Rothenberg, T.J. and C.T. Leenders (1964), "Efficient estimation of simultaneous equation systems", Econometrica, vol. 32, no. 1, pp. 57–76.
Roy, R. (1942), De L'Utilite: Contribution a la Theorie des Choix, Paris: Hermann.
Russell, Robert (1975), "Functional separability and partial elasticities of substitution", Review of Economic Studies, vol. 42(1), no. 129, pp. 79–86.
Salvas-Bronsard, L.D. Leblanc, Y. Zarra and C. Bronsard (1973), "The Rotterdam model and the negativity of the Slutsky matrix", Department de Sciences Economiques Cahier no. 7312, mimeo., Montréal: Université de Montréal.
Samuelson, P.A. (1969a), "Pure theory of public expenditure and taxation", in: J. Margolis and H. Guitton, eds., Public Economics, New York: St. Martin, pp. 98–123.

Samuelson, P.A. (1969b), "Corrected formulation of direct and indirect additivity", Econometrica, vol. 37, no. 2, pp. 355–359.

Samuelson, P.A. (1966), "The fundamental singularity theorem of non-joint production", International Economic Review, vol. 7, no. 1, pp. 34–41.

Samuelson, P.A. (1965a), "Using full duality to show that simultaneously additive direct and indirect utilities implies unitary price elasticity of demand", Econometrica, vol. 33, no. 4, pp. 781–796.

Samuelson, P.A. (1965b), "The Le Chatelier principle in linear programming", Santa Monica, CA: Rand. Reprinted in: J.E. Stiglitz, ed., The Collected Scientific Papers of Paul A. Samuelson, Cambridge, MA: M.I.T. Press.

Samuelson, P.A. (1962), "Parable and realism in capital theory: The surrogate production function", Review of Economic Studies, vol. 29(3), no. 80, pp. 193–206.

Samuelson, P.A. (1960), "Structure of a minimum equilibrium system", in: R.W. Pfouts, ed., Essays in Economics and Econometrics, Chapel Hill, NC: University of North Carolina Press, pp. 1–33.

Samuelson, P.A. (1958), "Frank Knight's theorem in linear programming", Zeitschrift für Nationalökonomie, vol. 18, no. 3, pp. 310–317.

Samuelson, P.A. (1953–54), "Prices of factors and goods in general equilibrium", Review of Economic Studies, vol. 21(1), no. 54, pp. 1–20.

Samuelson, P.A. (1947), Foundations of Economic Analysis, Cambridge, MA: Harvard University Press.

Samuelson, P.A. (1938), "A note on the pure theory of consumer's behaviour", Economica, N.S., vol. 5, no. 17, pp. 61–71.

Sargan, J.D. (1971), "Production functions", in: Layard et al., Qualified Manpower and Economic Performance, London: Allan Lane.

Sato, K. (1967), "A two-level CES production function", Review of Economic Studies, vol. 34(2), no. 98, pp. 201–218.

Sato, R. (1970), "The estimation of biased technical progress and the production function", International Economic Review, vol. 11, no. 2, pp. 179–208.

Scheffe, H. (1959), The Analysis of Variance, New York: Wiley.

Seitz, W.D. (1971), "Productive efficiency in the steam-electric generating industry", Journal of Political Economy, vol. 79, no. 4, pp. 878–886.

Seitz, W.D. (1968), "The measurement of productive efficiency", unpublished Ph.D. dissertation, Berkeley, CA: University of California.

Sevaldson, Per (1970), "The stability of input–output coefficients, in: Anne P. Carter and Andrew Brody, eds., Applications to Input–Output Analysis, vol. 2, Amsterdam: North-Holland.

Sevaldson, Per (1963), "Changes in input–output coefficients", in: T. Barna, Structural Interdependence and Economic Development, London: Macmillan.

Sevaldson, Per (1960), Norway, Statistisk Sentralbyraa: Kryssløpsanalyse av produksjon og innsats i norske naeringer 1954, Central Bureau of Statistics: "Input–output analysis of production and factor inputs in Norwegian industries, 1954". Samfunnsøkonomiske Studier no. 9, Oslo. Publication prepared by Per Sevaldson.

Shapiro, P. (1966), "Aggregate time series production functions and technical change: An econometric study", unpublished Ph.D. dissertation, Berkeley, CA: University of California.

Shephard, R.W. (1970), Theory of Cost and Production Functions, Princeton, NJ: Princeton University Press.

Shephard, R.W. (1953), Cost and Production Functions, Princeton, NJ: Princeton University Press.

Sibley, D.S. and E.E. Bailey (1974), "The dynamics of price adjustment in regulated industries", mimeo., Bell Laboratories.

Sidhu, S.S. (1974), "Relative efficiency in wheat production in the Indian punjab", American Economic Review, vol. 64, no. 4, pp. 742–757.

Sims, C.A. (1969), "Theoretical basis for a double deflated index of real value added", Review of Economics and Statistics, vol. 51, no. 4, pp. 470–471.

Solow, R.M. (1967), "Some recent developments in the theory of production", in: M. Brown, ed., The Theory and Empirical Analysis of Production, New York: Columbia University Press.

Solow, R.M. (1964), "Capital, labor, and income in manufacturing", in: National Bureau of Economic Research, The Behaviours of Income Shares, in: Studies in Income and Wealth, vol. 27, Washington, DC: NBER.

Solow, R.M. (1957), "Technical change and the aggregate production function", Review of Economics and Statistics, vol. 39, no. 3, pp. 312–320.

Spann, R. M. (1974), "Rate of return regulation and efficiency in production: An empirical test of the Averch–Johnson thesis", Bell Journal of Economics and Management Science, vol. 5, no. 1, pp. 38–52.

Star, S. (1974), "Accounting for the growth of output", American Economic Review, vol. 64, no. 1, pp. 123–135.

Statistics Canada (1974a), National Income and Expenditure Accounts, 1926–73, Ottawa.

Statistics Canada (1974b), Flows and Stocks of Fixed Non-Residential Capital, Ottawa.

Statistics Canada (1972), Fixed Capital Flows and Stocks, Manufacturing, Ottawa.

Statistics Canada (1963), Indexes of Real Domestic Product by Industry of Origin, 1935–61, Ottawa, pp. 61–505.

Stigler, G. (1939), "Production and distribution in the short run", Journal of Political Economy, vol. 47, pp. 305–357. Reprinted in: A.E.A., Readings in the Theory of Income Distribution, Homewood, IL: American Economic Association.

Stiglitz, J.E. (1969), "Behavior toward risk with many commodities", Econometrica, vol. 37, pp. 660–667.

Syrquin, M. (1970), "A note on inferior inputs", The Review of Economic Studies, vol. 37(4), no. 112, pp. 591–597.

Takayama, A. (1969), "Behavior of the firm under regulatory constraint", American Economic Review, vol. 59, no. 3, pp. 255–260.

Tilanus, C.B. (1966), Input–Output Experiments, The Netherlands 1948–1960, Rotterdam: Rotterdam University Press.

United Nations (1968), Department of Economics and Social Affairs, Statistical Office of the U.N.: A System of National Accounts, Studies in Method, Series F, no. 2, rev. 3, prepared by R. Stone et al., New York: U.N.

Uzawa, H. (1964), "Duality principles in the theory of cost and production", International Economic Review, vol. 5, no. 2, pp. 216–220.

Uzawa, H. (1962), "Production functions with constant elasticity of substitution", Review of Economic Studies, vol. 29(4), no. 81, pp. 291–299.

Valentine, F.A. (1964), Convex Sets, New York: McGraw-Hill.

Valentine, F.A. (1937), "The problem of Lagrange with differential inequalities as added side conditions", in: Contributions to the Calculus of Variations, 1933–1937, Chicago, IL: Chicago University Press, pp. 407–447.

Wall Street Journal (1962), "The Electric Conspiracy Case". Reprinted in: M.L. Joseph, N.C. Seeber and G.L. Bach, eds. (1966), Economic Analysis and Policy, 2nd ed., New York: Prentice-Hall.

Walters, A.A. (1963), "Production and cost functions: An econometric survey", Econometrica, vol. 31, nos. 1–2, pp. 1–66.

Weddepohl, H.N. (1970), Axiomatic Choice Models and Duality, Rotterdam: Rotterdam University Press.

Wellisz, S.H. (1963), "Regulation of natural gas pipeline companies: An economic analysis", Journal of Political Economy, vol. 71, no. 1, pp. 30–43.

Whitman, E.B. and E.C. North (1953), "Trending public utility construction cost indices", Public Utilities Fortnightly, vol. 52, no. 5, pp. 285–290.

Whitman, Requardt and Associates, The Handy–Whitman Index of Public Utility Construction Costs, semi-annual.

Wiley, D.E., W.H. Schmidt and W.J. Bramble (1973), "Studies of a class of covariance structure models", Journal of the American Statistical Association, vol. 68, no. 342, pp. 317–323.

Wilkinson, J.H. (1965), The Algebraic Eigenvalue Problem, Oxford: Clarendon Press.

Wilks, S.S. (1963), Mathematical Statistics, New York: Wiley.

Wilks, S.S. (1938), "The large-sample distribution of the likelihood ratio for testing composite hypotheses", Annals of Mathematical Statistics, vol. 9, pp. 60–62.

Winter, S. (forthcoming), "On the Existence of a Profit-Maximizing Production Plan", Econometrica.

Wold, H. (1944), "A synthesis of pure demand analysis, Part 3", Skandinavisk Akturarietidskrift, vol. 27, nos. 1–2, pp. 69–120.

Wold, H. in association with L. Jureen (1953), Demand Analysis, New York: Wiley.

Woodland, A.D. (1977), "Joint outputs, intermediate inputs, and international trade theory", International Economic Review, vol. 18, pp. 517–534.

Woodland, A.D. (1975), "Substitution of structures, equipment and labor in Canadian production", International Economic Review, vol. 16, no. 1, pp. 171–187.

Yotopoulos, P.A. and L.J. Lau (1973), "A test for relative efficiency: Some further results", American Economic Review, vol. 63, no. 2, pp. 214–223.

Yotopoulos, P.A., L.J. Lau and W. Lin (1976), "Microeconomic output supply and factor demand functions in the agriculture of the province of Taiwan", American Journal of Agricultural Economics, vol. 57, no. 2, pp. 333–340.

Zangwill, W.I. (1969), Nonlinear Programming: A Unified Approach, Englewood Cliffs, NJ: Prentice-Hall.

Zellner, A. (1963), "Estimators for seemingly unrelated regressions: Some exact finite sample results", Journal of the American Statistical Association, vol. 58, no. 304, pp. 977–992.

Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias", Journal of the American Statistical Association, vol. 57, no. 298, pp. 348–368.

Zerban, A.H. and E.P. Nye (1956), Power Plants, 2nd ed., Scranton, PA: International Textbook.

AUTHOR INDEX

Aczel, J. I.163
Afriat, S.N. I.412
Alexandrov, A.D. I.390
Allen, R.G.D. I.287; II.41, II.78
Andrews, W. I.261
Arrow, K.J. I.38, I.238, I.288, I.416; II.3, II.18, II.19, II.41, II.53, II.74, II.81, II.262
Atkinson, S.E. II.227
Averch, H. I.319; II.xx

Bailey, E.E. II.217, II.236
Barna, T. II.261
Barten, A.P. I.413
Bartle, R.G. I.383, I.406
Barzel, Y. II.150, II.163, II.187
Baumol, W. II.217
Belinfante, A. I.viii; II.188, II.191, II.200, II.210, II.212, II.261
Bellman, R.E. I.415
Ben-Israel, A. I.445
Berge, C. II.21, II.39, II.50
Berlinguette, V.R. II.19
Berndt, E. I.245, I.246, I.268; II.41, II.42, II.56, II.60, II.206, II.233
Blackorby, C. II.19, II.57
Bridge, J.L. I.267
Brooks, J. II.162
Brown, M. II.81, II.250
Bruno, M. I.ix, I.37, I.242, I.245; II.10, II.20, II.36, II.53, II.54
Burgess, D. I.248, I.266, I.268; II.250, II.262
Busemann, H. I.383, I.390

Cameron, B. II.262
Cass, D. I.138
Carter, A. I.242; II.270
Cavallo, D.F. I.261
Chenery, B.H. I.38, I.238; II.74, II.81
Chow, G.C. II.276
Christensen, L.R. I.viii, I.131, I.170, I.196, I.237, I.245, I.246, I.268, I.281, I.410; II.227, II.251, II.266
Cobb, C.W. I.230, I.238
Corden, W.M. II.19, II.41
Courant, R. I.118, I.142
Courville, L. II.215, II.237
Cowing, T.C. I.viii, I.ix; II.xx, II.191

Danskin, J.M. II.23
David, P. II.4, II.8, II.10, II.19, II.143
Davis, E.G. II.217
Debreu, G. I.5; II.21
Denny, M. I.viii, I.100, I.185, I.245, I.246, I.268; II.3, II.4, II.11, II.20, II.57, II.59, II.63, II.250, II.254
Dent, W.T. I.412
Dhrymes, P.J. I.243, I.288, I.297; II.81, II.88, II.96, II.103, II.188, II.230
Diamond, P.A. II.129, II.181
Diewert, W.E. I.viii, I.ix, I.5, I.38, I.112, I.113, I.114, I.115, I.122, I.124, I.129, I.131, I.138, I.170, I.183, I.226, I.237, I.238, I.270, I.282, I.289, I.292, I.293, I.410, I.412, I.415; II.3, II.9, II.10, II.20, II.28, II.29, II.30, II.32, II.35, II.36, II.40, II.42, II.53, II.54, II.57, II.92, II.251, II.271, II.272
Dillon, J.L. I.410
Domar, E. II.4, II.13
Douglas, P.C. I.230, I.238

Eisenpress, H. I.444
Enthoven, A.C. I.416
Epstein, L. II.18, II.25, II.26, II.27, II.28, II.31, II.33

Fan, K. I.383
Fenchel, W. I.5, I.12, I.95, I.383, I.385, I.387, I.389, I.390

Ferguson, C. II.81
Finsler, P. I.415
Frenger, P. I.viii, I.268; ii.288
Fuss, M.A. I.viii, I.214, I.246, I.248, I.268, I.340, I.341, I.342; II.63, II.65, II.187, II.189, II.190, II.193, II.209, II.254

Galatin, M. II.150, II.163, II.188, II.191, II.211
Gale, D. I.70, I.391
Gantmacher, F.R. I.431
Geary, R.C. II.19
Geyskens, E. I.413
Goldman, S.M. I.128, I.245, I.302; II.19
Gorman, W.M. I.viii, I.128, I.187, I.243, I.288, I.294; II.17, II.20, II.23
Griliches, Z. I.239, I.267; II.4, II.11
Grunbaum, B. I.383, I.385, I.386, I.387

Hadley, G. I.118
Haldi, J. II.271
Hall, R.E. I.viii, I.ix, I.170, I.185, I.248; II.3, II.153, II.251
Halter, A.N. I.242
Hanoch, G. I.5, I.24, I.38, I.58, I.127, I.129, I.130, I.230, I.239, I.241, I.242, I.243, I.244, I.288, I.294, I.295, I.296, I.299, I.301, I.305, I.307, I.412; II.18, II.33, II.40, II.193
Harberger, A. II.81
Heady, E.O. I.410
Hicks, J.R. I.viii, I.149, I.287; II.17, II.23, II.78, II.181
Hilbert, H. I.142
Hildreth, C. I.412
Hoch, I. I.249, I.261
Hocking, J.G. I.242
Holthausen, D.M. II.217
Hotelling, H. I.viii, I.5, I.14, I.145; II.30, II.223
Householder, A.S. I.422, I.429
Houthakker, H.S. I.112, I.116, I.128, I.195, I.297, I.309; II.24
Hudson, D.J. I.268, I.412

Intriligator, M.D. I.377

Johnson, L. I.319
Jorgenson, D.W. I.viii, I.5, I.131, I.134, I.138, I.170, I.196, I.212, I.237, I.238, I.246, I.268, I.281, I.410, I.413; II.152, II.186, II.191, II.200, II.227, II.266, II.275
Joskow, P.L. II.217, II.235

Kakwani, N.C. II.276
Karlin, S. I.95, I.383, I.389, I.406
Katzner, D.W. I.415
Kendall, M.G. I.441, I.450
Kendrick, J. II.81
Khang, C. II.20, II.36
Klee, V. I.70, I.383, I.386, I.391
Klein, L.R. I.258, I.285, I.410; II.271
Klevorick, A.K. II.217
Komiya, R. II.212
Konyus, A.A. II.29, II.30
Koopmans, T.C. I.214
Kurz, M. I.243, I.288, I.297; II.81, II.88, II.96, II.103, II.188, II.230

Lau, L.J. I.viii, I.5, I.38, I.39, I.100, I.112, I.116, I.120, I.124, I.128, I.129, I.131, I.134, I.138, I.153, I.160, I.163, I.166, I.169, I.170, I.184, I.187, I.192, I.198, I.237, I.238, I.241, I.245, I.246, I.267, I.268, I.281, I.410, I.413, I.418, I.419, I.450; II.20, II.25, II.28, II.56, II.223, II.226, II.227, II.249, II.272
Leontief, W.W. II.9, II.41, II.269
Liew, C.K. I.ix
Ling, S. II.197
Litzenberger, R. II.242
Lucas, R. II.81

Malinvaud, E. II.202, II.204, II.276
Mangasarian, O.L. I.383
Manne, A.S. II.81
Marschak, J. I.261
May, D. I.245; II.59
McFadden, D.L. I.viii, I.5, I.112, I.115, I.122, I.126, I.129, I.133, I.170, I.192, I.197, I.214, I.238, I.240, I.241, I.245, I.248, I.267, I.287, I.288, I.292, I.299, I.308, I.411, I.452; II.3, II.4, II.20, II.28, II.33, II.40, II.66, II.80, II.131, II.151, II.156, II.163, II.173, II.223, II.249, II.273
McKenzie, L. I.5, I.14, I.47; II.5

McKinnon, R. II.81
Metcalf, L. II.162
Myers, S.C. II.217
Miller, R.G. I.449
Minasian, J. II.81
Minhas, B. I.38, I.238; II.74, II.81
Mork, K. I.94
Mukerji, V. I.243, I.288, I.297
Mundlak, Y. I.170, I.240, I.241, I.242, I.246, I.247, I.249, I.250, I.261, I.267, I.287, I.288, I.290, I.291
Murray, W. I.444

Nerlove, M. I.viii, I.5, I.192, I.197, I.267; II.xx, II.74, II.127, II.151, II.177, II.188
Newman, P. II.25

Ohta, M. I.ix; II.xx

Parks, R.W. I.267; II.42
Peterson, H.C. II.215, II.219, II.229, II.237
Phillips, C.F. II.218
Pinto, C. I.268
Pollak, R.A. I.116, I.128; II.18, II.25
Popoviciu, T. I.390

Raiffa, H. II.201
Rao, C.R. I.445
Razin, A. I.242, I.246
Reidemeister, K. I.390
Revankar, N. I.242
Ringstad, V. I.239, I.267; II.4, II.11
Roberts, A.W. I.135
Rockafellar, R.T. I.12, I.13, I.15, I.55, I.70, I.100, I.109, I.137, I.140, I.141, I.142, I.383, I.385, I.386, I.387, I.389, I.390, I.391, I.395, I.397; II.21, II.39, II.47, II.48, II.50
Rosenlicht, M. I.383
Rothenberg, T.J. I.441, I.443, I.444
Rothschild, M. I.412
Roy, R. I.viii, I.5; II.29

Salvas-Bronsard, L.D.L. I.421
Samuelson, P.A. I.viii, I.5, I.15, I.46, I.47, I.112, I.116, I.128, I.184, I.309; II.3, II.5, II.18, II.20, II.40
Sato, K. I.242, I.244
Sato, R. II.81, II.125
Sargan, J.D. I.239
Scheffe, H. II.204
Seitz, W.D. II.150, II.162, II.185
Sevaldson, P. II.270, II.271
Shapiro, P. II.156
Shephard, R.W. I.viii, I.5, I.19, I.21, I.24, I.37, I.112, I.113, I.114, I.115, I.133, I.153, I.170, I.290; II.25, II.40, II.58, II.65, II.223, II.272
Sibley, D.S. II.217
Sims, C.A. II.4, II.10, II.19, II.41, II.55
Solow, R.M. I.38, I.238; II.81, II.162, II.250
Spann, R.M. II.215, II.229, II.237
Star, S. II.60
Stigler, G. I.313
Stiglitz, J.E. II.18, II.33
Syrquin, M. II.188

Takayama, A. I.412
Tilanus, C.B. II.271, II.284

Uzawa, H. I.viii, I.5, I.19, I.21, I.112, I.114, I.128, I.133, I.197, I.240, I.242, I.245, I.288, I.299, I.302; II.19, II.76, II.77

Valentine, F.A. I.383, I.412
Varberg, D.E. I.135
Vulcani, Z. I.250

Walters, A.A. I.267
Weddepohl, H.N. II.25
Wellisz, S.H. II.217
Wiley, D.E. I.421
Wilkinson, J.H. I.422
Wilks, S.S. II.231, II.261
Winter, S. I.71, I.385, I.394
Wold, H. II.17, II.23, II.30
Wood, D.O. I.246, I.268; II.41, II.42, II.56, II.60
Woodland, A.D. I.266; II.20

Yotopoulos, P.A. I.207, I.267

Zangwill, W.I. II.28
Zellner, A. II.61, II.204, II.205, II.231, II.261, II.276

SUBJECT INDEX

ACMS (CES) I.38, I.238; II.74
Activity analysis I.333
Addilog normalized profit function I.195
Adding-up condition I.232; II.255
Additive separability I.158
Affine hull I.384
Aggregate commodities II.17
Aggregated utility function II.22
Aggregation I.284, I.222
  consistent II.289
  of outputs II.251
Almost homogeneous functions I.163
Approximation (of production relations) I.232, I.236, II.28, II.56, II.253
Asymptotically irreversible I.271
Asymptotic cone I.63, I.384
Average cost I.313
  ex ante and ex post I.313
Averch–Johnson Model II.217–22

Barrier hyperplane I.384
Bernstein polynomial I.236
Bias of technical change II.128, II.182–85
Bordered matrix I.368
Boundedly polyhedral I.391
Bounding hyperplane I.16
Bounds (implied by technical progress) II.139

Capacity (of plant) II.99
Capital augmentation II.140
Capital share II.128
Capital stock in manufacturing II.60
CDE cost function I.241, I.291, I.295
CES I.38, I.238, II.74
  combined with Cobb–Douglas I.193
  composition I.57
  cost function I.38, I.299
  multiple-output frontier I.307, I.308
  non-homothetic I.298
  production function I.38, I.192, I.238, I.299; II.74
  structure II.82–88
Characteristic value I.366
Cholesky factorization I.413, I.421–37
Cholesky values I.429
Classical cost function II.76
Classical production functions II.75, II.127
Classical programming problem I.375
  second-order conditions I.380
Clay-clay model II.190
Closed function I.67, I.385
Closed half-space I.384
Cobb–Douglas I.6, I.38
  composition I.56
  cost function I.38, I.300
  production function I.38, I.190, I.238, I.253, I.300
  restricted profit function I.93, I.190
  variants I.242, I.301, I.309
Comparative statics I.46, I.147, I.231
Complements I.48
Composite good II.9
Composition I.49, I.95
  composition operator I.247
  convolution of input requirement sets I.51
  examples I.100
  intersection of input requirement sets I.50
  of cost functions I.49
  of restricted profit functions I.95
  scaling I.50
  sums of input requirement sets I.51
  union of input requirement sets I.50
Concave function I.12, I.127
  quasi-concave I.12
  strictly quasi-concave I.30
  strictly differentiably quasi-concave

Concave function—*contd*
I.30
strongly concave I.135
Cone I.384
asymptotic I.63, I.384
normal I.62, I.385, I.389
pointed I.384, I.387, I.388
polar I.385
recession I.63, I.384
Conical function I.271
Conjugate dual I.390
Conjugate functions I.139
Constant marginal share (CMS) production function I.242
Constrained estimation I.412, I.413, I.437–43
generalized linear function I.439
maximum likelihood I.441–43
method of squaring I.412
non-linear least squares I.444–45
quadratic function I.438
translog function I.438
Continuity I.12
of correspondences I.63, II.21
of cost function I.12
of functions I.12, II.20
of restricted profit function I.68
Conventional production possibility set I.10
Convex function I.67, I.385
constrained estimation (quasi-convex) I.440
examples I.274
Hessian matrices I.414–20
properties of approximating function I.418–20
quasi-convex I.12, I.414–20
strictly differentiably quasi-convex I.89, I.403
strictly quasi-convex I.89, I.403
Convex set I.8, I.384
convex from below I.8
convex hull I.384
properties I.385–386
strictly convex from below I.16
Correspondence I.63, I.391, II.21
bounded (semibounded) range I.391
continuous I.63, I.392
convex-valued I.391
lower hemicontinuous I.63, I.391
lower semicontinuous II.21
maximand I.393, I.399
semibounded-valued I.397
strongly continuous I.63, I.391
upper hemicontinuous I.63, I.391
upper semicontinuous II.21
Cost function I.10
asymptotes I.46
CDE I.241, I.291, I.295
definition I.10
duality I.19
derivative property I.14
ex ante and ex post II.189, II.190
generalized Leontief I.340, II.271–73
geometry I.39
input-conventional I.20
joint II.253
"kinks" and "flats" I.44
nested I.340
non-jointness II.256
non-homothetic II.64
polar I.112
properties I.13, I.228, II.75
separable II.253
Shephard's lemma I.14
transcendental logarithmic II.56, II.243, II.253–55
value-added II.37–40
Cost mapping I.21
Cost-polar transformation I.121
Cournot aggregation condition I.232
CRES (CRESH) production function I.240, I.243, I.291, I.294, I.295
Criteria for functional forms I.224

Demand functions I.14
inverse II.29
short-run II.29
Depreciation II.154
Derivative property I.389, I.272
gradient I.74, I.389
Hessian I.74, I.389
of cost function I.14
of restricted profit function I.74, I.272
subdifferential I.74, I.394
Diminishing returns I.48

Direction of recession I.137
Directional shadow elasticity of substitution II.291–96
inconsistent II.296–300
Disposal I.66
of inputs I.66
of outputs I.66
Distance function I.24, I.113, I.226
duality I.28
input-conventional I.26
properties I.26
Distribution I.221
Divisia index II.12, II.162
Dominant own-price effect I.279
Double deflation II.7
Duality I.19, I.390, I.138
conjugate duality I.390
extensions for cost functions I.35
extensions for restricted profit functions I.85, I.90, I.151–69
for cost functions I.21, I.115
for distance functions I.28
for restricted profit functions I.82
for valued-added functions II.37
homogeneity and separability I.151–169
Legendre transformations I.143 fn
multiple output I.173–182
polar reciprocal duality I.401
Shephard–Uzawa theorem I.21, I.115

Efficiency I.30, I.70, I.222, I.313
Elasticity of demand I.290, II.209, II.264
Elasticity of substitution I.42, I.197, I.240, I.344, II.73, II.78, II.128
Allen (AES) I.239, I.290, II.80
between outputs I.292
direct (DES) I.129, I.240, II.80
directional II.291
inconsistent II.287
OOES I.240
properties II.82 fn
relation to production function II.131
relation to separability II.42
shadow (SES) I.130, I.240, II.80, II.273
TOES I.240, I.288, I.291
TTES I.240, I.288
Elasticity of transformation I.289, I.293
TOET I.289, I.293

Electricity system design I.335
Electric power industry II.150, II.157–62
Empirical production analysis I.267
Engle aggregation condition I.232
Energy (intermediate input) II.41
Error structure I.259
Estimation methods (for production parameters) I.261
Euler's law I.11, I.152
Eventually decreasing returns I.48
Eventually increasing costs I.48
Ex ante envelope profit function I.329
Ex ante technology I.313, I.324, I.345
envelope I.325
in steam-electric generation II.172, II.193
of a regulated electric utility II.225–27
parameters II.190
Exponential production function I.195
Exposed functions I.402
Exposed points I.68, I.396
Exposed sets I.396
Ex post cost function II.190
Ex post technology I.313, I.324, I.345
in steam-electric generation II.164, II.191
variable I.325
Ex post total profit function I.328

Factor augmentation II.128
Factor augmenting technical change II.135
conditions for II.135–37
finite parameter II.143
implied bounds II.139
Factor price elasticities (ex ante) II.209
Factor price frontier I.37, I.116, II.288
Factor price requirement set I.37
Factor requirement function I.129
Factor substitution II.269
Final demand II.250
Firm effect I.253
Fixed proportions II.8, II.69
Flat (affine subspace) I.384
Flexibility (of technology) I.313, I.221
Forward price I.325
Free disposal I.8, I.66
Free disposal hull I.8

Functions I.12
almost homogeneous I.163
closed I.385
conical I.271
concave I.12
continuous I.12
convex I.12, I.67, I.274, I.385
differentiability I.389
dom *f* I.271
epi *f* I.271
lower semicontinuous I.31, I.73
proper I.137
quasi-concave I.12, II.25
quasi-convex I.12, I.403, I.413
set-valued I.391
strictly differentiably quasi-convex I.403
strictly quasi-convex I.403
strongly continuous I.35
upper semicontinuous I.31
Future price I.325

Gauge function I.76, I.102, I.123, I.400
Generalized Cobb–Douglas production function I.238
Generalized concave production function I.238
Generalized Leontief production function I.238, II.271
Generalized linear production function I.410
General linear profit form I.272
Generally non-decreasing costs I.48
Generally non-increasing returns I.48
Gradient I.30, I.75, I.389
Gross output II.6, II.54, II.59
Gross substitutes property I.276, I.362, II.5
Growth equations II.129

Handy–Whitman Index II.161, II.212
Harrod neutral technical change I.202
Hemicontinuous I.63, I.391
Hessian I.30, I.75, I.149, I.150, I.389
Hicks aggregation II.23, II.69
Hicksian laws of production I.144–45
Hicks neutral technical change I.202, II.261

Homogeneous I.11, I.121, I.127, I.151–82
almost I.163
frontier I.305
of degree *k* I.152
test of homogeneity II.58
Homothetic I.58, I.120, I.128, I.153, I.222, II.77, II.66, II.67
frontier I.305
lack of in electric-power generation II.163, II.189
multiple output case I.173, I.305
partially I.167
test of homotheticity II.58, II.61
weakly II.65
Homothetic-polar I.121, I.128
Homothetic separability I.159
Hotelling's lemma I.145, I.230, II.223–25
Hyperplane I.16, I.383
barrier I.384
bounding I.16, I.384
separation property I.384, I.386
supporting I.16, I.384

Identification (technical change) II.133, II.152, II.155–57
Implicit gauge mapping I.82
Implicit input requirement set I.20
Implicit production possibility set I.20
Implicit profit mapping I.82
Implicit technology mapping I.82
Increasing function I.31, I.85
Indirect Harrod neutral technical change I.204
Indirect Hicks neutral technical change I.204
Indirect utility function II.24
Indirect production function I.116, II.26
Input-bundle I.6
Input-classical I.47
Input-conventional I.10, I.20
Input-homothetic I.58
Input (intermediate) II.11
Input–output coefficients II.272
Input–output ratio II.60
Input–output separable I.58
Input–regular I.7
Input requirement sets I.7, I.225
Intertemporal variability I.321

Inverse gauge mapping I.82
Inverse Legendre transformation I.143

Joint production I.121

Lagged adjustment I.212
Lagrangian multipliers I.375
  associated with rate-of-return constraint II.220–22
  critical point I.376
Legendre transformation I.142
Leontief hypothesis II.282
Lim inf I.67, I.385
Linear homogeneous I.11, I.271, I.385, II.27
Linear-in-parameters technology I.230, I.237, I.346
Lipschitzian I.389
Load curve I.335
Load factor I.335
Locally strongly concave I.135
Local solution I.375
Lower hemicontinuous correspondence I.63, I.391
Lower semicontinuous function I.73

Maintained hypotheses for estimation I.253–59
Mappings I.82
  duality mappings I.104–07
Marginal rate of transformation II.265
Matrices I.365
  bounded matrix I.368
  characteristic value I.366
  Cholesky factorization of I.422–25, I.430–32
  determinant I.366
  Frobenious root I.282
  positive definite subject to constraint I.368
  positive semidefinite I.277, I.365, I.427
  principal minor I.365
  real symmetric I.433–35
  relative definiteness I.278
Maximand correspondence I.393, I.399
Monopolistic profit function I.209, II.218 fn
Monotonicity I.31, I.135
Multiple output production functions I.171, I.291, I.292, I.302, II.251
Multi-product final demand technology II.250

Nominal value-added II.19
Non-decreasing I.31, I.85, I.114
Non-degenerate programming problem I.376
Non-increasing I.31, I.85
Neoclassical I.35
Nested cost functions I.340
Net output I.61
Netput I.61
Net supply correspondence I.71
Neutral scale effect II.90
Non-constant returns II.145, II.176
Non-joint production I.182, I.362, II.251, II.252
Non-systematic error I.253
Normal cone I.62, I.345
Normal inputs I.48
Normalized price I.134
Normalized profit function I.134
  addilog I.195
  example I.190
  functional forms I.184
  multiple output I.171
  properties I.139
  quadratic II.226
  reciprocal quadratic I.196
  structure I.151
  transcendental logarithmic I.196
Normal to hyperplane I.384

Objectives of production analysis I.220
Output bundle I.6
Own-price effect I.148

Paasche price aggregates II.289
Parsimonious flexible forms I.233
Plant design I.320, I.324
Pointed cone I.384, I.387–88
Polar cone I.385
Polar production function I.112, I.118, I.119
Polar reciprocal functions I.401

Polar reciprocal sets I.37, I.401
Polytope I.391
Positive definite subject to constraint I.368
Positively linear homogeneous I.11, I.67, I.114, I.385
Positive semidefinite matrix I.277, I.365
Price aggregate (inconsistent) II.296
Price expectations II.200, II.213
Price possibility set I.92, I.103
Prices
  forward I.325
  future I.325
  spot I.325
Principal minor I.365
Producible output set I.7
Production function I.6, I.25, I.134, I.226
  almost homogeneous I.163
  CES (AMCS) I.38, I.238, I.299
  classical II.127
  Cobb–Douglas I.38, I.238
  constant marginal share (CMS) I.242
  CRES I.243, I.291, I.294, I.295
  CRESH I.240, I.296
  direct addilog I.298
  exponential I.195
  generalized Cobb–Douglas I.238
  generalized concave I.238
  generalized Leontief I.131, I.238
  generalized linear I.131
  indirect I.116, I.297, II.26
  Leontief I.299
  linear I.299
  multiple output I.170, II.251
  polar I.112, I.118, I.119
  properties I.135, I.136
  quadratic I.131, I.194, I.238
  reciprocal indirect I.295
  strictly neoclassical II.127
  variable ES (VES) I.242
Production plan I.61
  ex post I.324
Production possibility set I.6, I.61, I.113, I.225
  asymptotically irreversible I.271
  conventional I.10
  irreversible I.62
  properties I.61–66, I.102
  regular I.7, I.122, I.126
  semi-bounded I.62
  strongly continuous I.63
Production technology I.5
Profit function I.61, I.122, I.229 (see also normalized profit, restricted profit)
  composition rules for I.95
  design-linear profit structure I.346
  examples I.100
  for a regulated firm II.222–28, II.245–46
  future I.357
  general linear I.272
  Hotelling's lemma I.230, II.223–25
  normalized I.134
  polar I.112
  polar variables I.129
  restricted profit function I.61
  saddle-functions I.101
  unit-output-price II.220, II.222
  variable I.129, II.20, II.30 fn
Profit polar frontier I.304
Profit polar production set I.126
Proper function I.137
Public goods II.18, II.32
Putty-clay technology I.315
  estimation in steam-electric generation II.208
  hypothesis I.362
  test of in steam electric generation II.204–08
Putty-putty technology I.315, II.188

Quadratic production function I.194, I.238
Quadratic profit function II.226
Quasi-concave I.12, II.33, II.76
Quasi-convex I.12, II.24
  estimation subject to I.437–43
  hypothesis testing of I.447–52

Rate of return regulation II.217–22
  test of in steam electric generation II.231–34
Rate of technical change II.128, II.168, II.177, II.180
Real value-added function II.35, II.69
  non-homothetic II.64
Recession cone I.63, I.384

Reciprocal quadratic normalized profit function I.196
Regressive inputs I.48
Regular growth paths II.131
Relative definiteness I.278
Relative efficiency I.206
Relative interior I.76, I.384
Restricted cost function II.192
Restricted profit function I.61
composition rules I.95
continuity I.68, I.70
definition I.66
examples I.93, II.3
joint continuity I.72
kinks and flats I.90, I.91
nested I.101
properties I.67, I.68, I.103
Returns to scale I.48, I.221, II.156
in steam-electric generation II.176, II.177, II.195
Revenue function I.302
polar I.303

Saddle function I.103, II.6
Scale bias II.90
Second-order approximation property I.233
of translog cost functions II.56, II.253–58
Semibounded I.62, I.384, I.387–89
Separability I.60, I.151–82, I.221, I.244 fn, II.10, II.41
across states I.356, I.362
additive I.158
approximate II.255
input–output I.60, I.128, I.175, I.302, I.305
in value-added functions II.9, II.19, II.55
of empirical production relations I.246, II.62, II.63, II.65, II.251
strong I.244
weak I.245
Separating hyperplane I.384
Separation property (of sets) I.384, I.386
Sets I.383–86
boundedly polyhedral I.391
closed I.383
convex hull I.384
free disposal hull I.8
interior I.383
polytope I.391
relative interior I.76, I.384
semi-bounded I.62, I.384, I.387–89
separation property I.384, I.386
strictly convex (rotund) I.16
unit profit I.122
Shadow elasticity of substitution II.273
estimates in Norway II.279
inconsistent II.287
Share of capital II.128
Shephard's lemma I.14, I.75, I.115
Shephard–Uzawa duality theorem I.21, I.115
Short-run demand functions II.29
Short-run profits I.134
maximization I.129
Specification (of ACMS model) II.80
Spot price I.325
Steam-electric generating industry II.73, II.95, II.157
Strictly neoclassical production function II.127
Strongly concave I.135
Strongly continuous I.35
Strongly lower semicontinuous function I.31
Strongly lower (upper) hemicontinuous correspondence I.35, I.391
Strongly non-degenerate programming problem I.376
Strongly regular I.353
Strongly upper semicontinuous function I.31
Strong nested profit form I.353
Strong separability I.244
Strong separation property of sets I.384
Structure (of technology) II.187
Subdifferential I.19, I.75, I.394
Substitutability-scale effect II.90
Substitutes I.48
Substitution (of inputs) I.221, I.42, II.131
Support function I.385
Supporting hyperplane (support) I.16, I.384
Symmetry of cross-price effects I.148, II.254

Taylor series I.233
Technical change I.201, I.221
  bias II.12, II.128, II.157, II.182–85
  capital augmentation II.140
  embodied and disembodied II.152–57
  factor augmenting I.204
  factor saving I.204
  Harrod neutral I.202
  Hicks neutral I.202
  indirect Harrod neutral I.202
  indirect Hicks neutral I.202
  in steam-electric generation II.158, II.159, II.168, II.177, II.180, II.196
Technical non-retrogression II.127
Technical progression II.127
Technological structure I.345, II.187
  design linear I.346, I.353
Technology I.5
  ex ante and ex post I.313, I.324, I.345, II.189
  linear-in-parameters I.346
  multi-product II.251
  non-joint II.251
  putty-clay I.315, II.190, II.208
Technology mapping I.21
Total gauge function I.103
Total profit function I.103
Total price possibility set I.103
Total productivity II.11
  rate of change II.12
Transformation function I.227
  polar I.303
Transcendental logarithmic forms I.196, I.238, II.56, II.253
Transcendental production function I.242
Translated technology I.76, I.79, I.155
Two-level structure I.345

Uncertainty (and profits) I.214, I.321
Uniformly increasing (decreasing) I.31, I.85
Unit cost frontier I.115, I.124
Unit cost set I.115
Unit value-added cost function II.37
Upper hemicontinuous correspondence I.63, I.391
Upper semicontinuous function I.31
Utility function II.24
  conditional indirect II.25
  indirect II.24
  macro II.23

Value-added II.92, II.54
  and separability II.8, II.61, II.92
  double deflation II.7, II.55, II.68
  function II.4–6, II.19, II.35–40
  imperfect competition II.15
  saddle-point property I.103, II.6
  value-added cost function II.38
Variable elasticity-of-substitution production function I.242, II.88
Variable profit function I.129, II.20, II.30 fn

Weak homotheticity II.65, II.66
Weakly non-degenerate programming problem I.376
Weak separability I.245, II.61, II.255
  of the translog cost function II.57, II.62, II.63
Wold's identity II.30
Wong–Viner I.316
Wronskian I.233