This is to certify that the

dissertation entitled

EFFICIENCY IN FRONTIER PRODUCTION FUNCTIONS

presented by

RICHARD ALMY BARCLAY

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Ag Economics

Stephen B. Hoch
Major professor

Date 2 July 1986

# EFFICIENCY AND FRONTIER PRODUCTION FUNCTIONS

By

Richard Almy Barclay

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1986

ABSTRACT

EFFICIENCY AND FRONTIER PRODUCTION FUNCTIONS

By

Richard Almy Barclay

This dissertation raises serious questions regarding the validity of frontier production functions which are considered better than traditional production functions because they distinguish between "technical efficiency" (TE) and "price efficiency" (PE).

In empirical work there are always variations in the data. Frontier production functions represent a new theory explaining the cause of these variations predicated on the existence of an interior to an isoquant at the same level of production. This dissertation first explores unit isoquants and solid production possibilities sets. It is shown that specification and/or aggregation error of the sub-production function, due to a lack of attention to the fixed inputs, account for the variations attributed to TE and PE. It is also shown that the discrepancies between observations within a unit isoquant attributed to differences in TE are due to the averaging process itself, so that the unit isoquant represents the boundary between Stages I and II of production.

Frontier production functions and distance functions are frequently regarded as the same. It is demonstrated

that the principles of duality implicit in distance functions conflict with the distinguishing characteristics of frontier functions. Specifically, duality proves that TE and PE are always identically equal by definition.

The two appendices present background theory and agruments that substantiate the thesis that "TE" and "PE" represent specification and/or aggregation error. A brief review of thermodynamics shows that thermal efficiency and economic efficiency are the same, and that "technical inefficiency" as used in the frontier production function literature is physically impossible.

TABLE OF CONTENTS

Page

CHAPTER

Page

CHAPTER

Page

CHAPTER

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER ONE

# INTRODUCTION

It is becoming increasingly fashionable to estimate production functions in a 'new and better way.' Increasingly researchers estimate a "frontier" production function rather than a traditional production function. The frontier production function is supposed to be a new way to explain differences in the observed production behavior of different firms [King, 1980]. The frontier function is supposed to be 'better' because it represents an estimate of the best performance observable.

> A frontier production function may be thought of as a "best practice" production function (Forsund and Jansen) or a function that expresses the maximum product obtainable from various combinations of factors given the existing state of technical knowledge. [King, 1980, page 1]

## 1.1. VARIATION IN OBSERVATIONS IN APPLIED WORK

Undoubtedly the frontier production function approach has some intuitive appeal, or it would not be gaining such wide acceptance and popularity. The intuitive appeal of the frontier production function approach is probably due to the fact that in applied work it is not uncommon to observe two firms Q and P, which appear to be using the same technology with different amounts of the same inputs to produce the same amount of the same output.

One should remember that all applied problems assume a consistent theoretical basis for all the observations included in the data. This theoretical basis is understood to be an abstraction from reality which helps one distinguish what is important from what is unimportant. In firm level production analysis this frequently means that all observed firms are assumed to use the same technology, use the same inputs, pay the same prices for their inputs, and receive the same price for their output. These same assumptions are generally made for frontier production functions, too. It is clear that the assumptions do not strictly hold in reality; that firms do not use the same technology, do not use the same inputs, do not pay the same prices for their inputs, and do not get the same price for their output. Traditionally, reseachers have assumed that the real observations represent a distribution around the theoretical points of "same" technology, "same" inputs, "same" input prices, and "same" output price. This distribution is caused by "noise," or random uncontrollable factors affecting the real observations. Consequently, one uses the data to estimate parameters of the system at the mean of the data, i.e., at the theoretical points of "sameness." That is, with careful attention to one's assumptions, one in effect theoretically insures the variation does not exist. Dealing with the variation that does exist in the data is an empirical problem, rather than a theoretical one.

Traditionally, one deals with the variation by carefully specifying the sub-production function in terms of the levels of the fixed inputs, so that all observations are on the "same" sub-production function, and by aggregating inputs that are very close to the "same" for all observations. How close is "very" close is a matter of judgement. When the specifying and aggregating is not "very close," one commits either a specification or an aggregation error. That is, if one includes data from two firms in estimating a function which are on two different sub-production functions, one has a specification problem. Or, if one includes two firms using different inputs which are treated as the "same," one has an aggregation problem. One expects that the actual variation left after specifying the sub-production function and aggregating the inputs will be variation that deviates from the mean of the data points, the point of theoretical "sameness," and not from one extreme. Thus, such observed differences between firms are not due to "technical" efficiencies or inefficiencies but, rather, to either implicit specification and/or aggregation errors.

Frontier production functions represent not only a empirical departure from the traditional approach but also a theoretical justification for this departure. Frontier production functions constitute a new theory of production, not simply a new method for explaining variation in data. No longer are observations simply deviations from their means, they are deviations from their extremes.

> Why not use a statistically estimated <u>average</u> unit isoquant, rather than a frontier isoquant? The answer is that the frontier function, which determines "best" practice in the industry and which all firms are attempting to emulate, may <u>not</u> be a nuetral transformation of the average function. The frontier production function may have entirely different factor elasticities from the average function. [Timmer, 1971, page 779]

Because the frontier function is the "best practice" function, it reputedly allows the researcher to discriminate between firms on the basis of "technical" efficiency (TE) and "price," or allocative efficiency (PE) to explain the difference between Q and P. That is, some of the variation in the data, due to its lack of "sameness," is explained as being due to differences in "technical" and/or "price" efficiency between observations.

## 1.2. WHY FRONTIER PRODUCTION FUNCTIONS ARE "BEST"

The theory of frontier production functions is still evolving. This means that in understanding frontier production functions, one must accept the presence of two obstacles. These two obstacles result in problems of inconsistency, ambiguity, and vagueness, and are the result of the fact that frontier production functions prove, in every case, to be misleading interpretations of reality.

First, there is more than one definition of a frontier production function and hence a lack of consensus among authors as to what represents a frontier production function. There might be disagreement or inconsistencies between authors regarding some aspects of what constitutes frontier production function theory.

Second, for a particular representation of a frontier production function, not all the underlying preconditions (assumptions) are necessarily clear. Nor are all the implications or ramifications of a particular characteristic stated, or even clearly implied.

Despite these two obstacles all frontier production functions share two concepts: (1) TE $\neq$ PE due to an intrinsic separation between physical decisions and value decisions, and therefore (2), there is a technical efficiency and a technical inefficiency that are exclusively different.

The first characteristic shared by all representations of frontier production functions is that TE is separate from PE; that technical efficiency and allocative efficiency are mutually exclusive phenomenon. Recent equivocation, that allocative efficiency means only getting inputs in the proper ratios but not necessarily in the proper quantites [Kopp, 1981a, Schmidt and Lin, 1983], has not altered the basic premise that one can observe technical efficiency (or inefficiency) in isolation from observing price efficiency.

> Contrary to price efficiency which is a purely behavioral concept, technical efficiency is purely an engineering concept. It entirely abstracts from the effect of prices. [Lau and Yotopoulous, 1971, page 95]

If technical efficiency "entirely abstracts from prices," in so doing, it ignores the fact that physical commodities without value (price) are considered to be "free" goods, and therefore are of no economic consequence. Quite simply, some frontier production theory suggests that one may get something for nothing (more output) by simply increasing technical efficiency. Since technical efficiency is separate from price efficiency the increase in efficiency should be costless.

> These results also show, however, that on average there is 22.82 (=1.0 - 0.7718) percent and 23.27 (= 1.0 - 0.7673) percent technical inefficiency in the cases of crop and mixed farm samples, respectively. This means that actual (observed) output is about 23 percent less than maximal output which can potentially be achieved from the existing level of inputs. In other words, through the efficient use of existing inputs the farm output can be increased by almost 23 percent without any additional cost to the farmers. [Bagi and Huang, 1983, page 255]

This quote not only exemplifies that TE and PE are regarded as mutually exclusive, but also exemplifies the second concept that all frontier production function interpretations share.

This second concept is that technical efficiency means one might get more than one quantity of output given the same technology (the same set of identical fixed and variable inputs) and that the failure to achieve the same level of output cannot be corrected using economic theory as a guide.

> The economic decision-making process can fail in two different ways. The whole core of economic theory is concerned with the first of these - the marginal revenue products of some or all factors might be unequal to their marginal costs. If this is true the allocative decision is said to be inefficient. The second source of failure is the technical production function - a failure to produce the greatest possible output from a given set of inputs means the technical decision is inefficient. [Timmer, 1971, page 776]

The suggestion that one might get two or more possible outcomes when using homogeneous inputs in identically specified production processes has lead some to suggest explicitly that the production set is a solid rather than a surface; that an isoquant is a plane rather than a line [Jamison and Lau, 1982].

As can be seen from the quotes above, both these concepts are so interdependent that there is virtually an if-and-only-if connection between them. This is because both of them rely entirely on the existence and interpretation of <u>interiors to isoquants</u>. Being on the isoquant means being technically efficient, while being within the interior of the isoquant means being technically inefficient. Being at the point where the budget constraint is tangent to the isoquant, or being on a ray from the origin through this point, means being price efficient. Being away from the tangency point, or the ray through it, means being allocatively inefficient.

## 1.3. WHY IS THIS IMPORTANT?

The frontier production function is an attempt to explain the cause of variation in data in empirical work. This is a practical problem rather than a theoretical one. Unfortunately, frontier production functions do not merely represent a method for dealing with an empirical problem. They also represent a theory of production which is physically and logically questionable. This confusion of theoretical issues with empirical issues has resulted in an ambiguous and misleading frontier production function literature.

Eliminating this confusion is important because the distinction between technical and price efficiency reduces the credibility of economic theory in explaining reality. The distinction places economists at odds with engineers and all physical scientists. Frontier production function theory encourages researchers to believe that efficiency can be solely concerned with physical production relationships (TE), or solely concerned with value relationships (PE). This suggests that economic efficency is not merely minimizing opportunity cost, but that there is a purely physical aspect that is unassociated with value.

The current production literature increasingly contains examples of research that have measured these 'two types' of efficiency [Bagi and Huang, 1983, Bravo-Ureta, 1983, Charnes, Cooper, and Rhodes, 1978, Charnes, Cooper, and Rhodes, 1981, Forsund, Lovell, and Schmidt, 1979, Forsund and Hjalmarsson, 1974, Forsund and Hjalmarsson, 1979, Hall and Le Veen, 1978, Lesser and Greene, 1980, Schmidt and Lovell, 1977, Schmidt and Lovell, 1978]. Since there is a fundamental difficulty in the logic of the proposition that there are two, or more, types of efficiency, these studies are fundamentally flawed and reach erroneous conclusions. Researchers have measured what is in reality variation in the data due either to (1) combining observations from two or more different sub-production functions (the specification problem) or (2) aggregating heterogeneous inputs (the aggregation problem). These differences are then attributed to differences in "technical" efficiency between firms. There is a danger that the conclusions of these studies will be used in formulating public policy or in entreprenuerial decision making. More importantly, the credibility and legitimate development of the science of economics is jeoprodized by institutionalizing illogical and misleading theory.

Frontier production functions are both a theory and an explanation of variation in data. If the theory is invalid, frontier production functions, <u>properly interpreted</u>, can still be an empirically valid way of detecting

and explaining variation in the output of firms. The methods currently used in frontier production function analysis can be used diagnostically to detect aggregation and/or specification errors.

## 1.4. OUTLINE OF CHAPTERS

### 1.4.1. CHAPTER TWO: UNIT ISOQUANTS AND PRODUCTION SOLIDS

Two cases of frontier production function theory and two mutually exclusive types of efficiency are discussed in this chapter. First, the unit isoquant is defined. This concept is used in the original and fundamental definition of a frontier production function. It compares firms on the basis of input used per unit of output obtained. The next section considers two cases of a solid production set. A solid production set implies that two identical sets of the same inputs can produce different output using the same technology.

These cases preserve the essential frontier production function characteristic that there is an interior to an isoquant at the same level of production. The last section explains why such an interior can appear to be observed in reality only if a specification or aggregation errors exist. Specification error means one is making a comparison across sub-production functions or input requirement sets, rather that within them. Aggregation error means one has aggregated heterogeneous inputs and identified them as homogeneous.

Apparent differences in TE are attributed to firms having different amounts of fixed inputs, or aggregating heterogeneous inputs. Frontier production function analysis often ignores the presence of fixed inputs and the economies of investment/disinvestment necessary to change the amount of fixed input used in production (see Edwards, 1958).

### 1.4.2. CHAPTER THREE: FRONTIER FUNCTIONS AND DISTANCE FUNCTIONS

Frontier production functions are often assumed to conform to the same underlying assumptions as distance functions; concavity and monotonicity. Therefore, it might be inferred that frontier production functions are a special case of distance functions. First, it will be shown why frontier production functions and distance functions may mistakenly be interpreted as being the same. To many, the frontier production function <u>appears</u> to be a by-product of the attempt to provide a rigorous mathematical proof to duality theory. The third and fourth sections prove by contradiction that frontier production functions and distance functions are incompatible.

The proofs in the third and fourth sections rely heavily on an interpretation of duality theory presented by McFadden (1978). McFadden's duality theory excludes Stage III of production because he postulates that marginal physical products are always non-negative. His duality theory implies that TE and PE are identical; i.e., that

also

recip

mappi

Stag

valu

fre

aff

carr

usin

disp

cor

mea

con

ext

inp

the

1.(

bec

Th

is

ti

fr

ti

that all points on a production function that are TE are also PE by definition. This is demonstrated using polar reciprocal sets. Polar reciprocal sets make a one to one mapping from physical to value relationships (excluding Stage III); the physical aspect of production (TE) and the value aspect of production (PE) are inseparable. Therefore, the third section demonstrates that there can be no difference between TE and PE.

The fourth section demonstrates that an isoquant cannot have an interior at the same level of production using duality theory and its inherent principle of free disposal. Free disposal in production space means the corresponding cost is also freely disposed. Free disposal means that <u>both</u> input and output are freely disposed and completely removed from accounting, i.e., it is as if the extra input and output never existed; consequently "extra" input/output cannot affect technical efficiency because they can not be included in production or cost.

### 1.4.3. CHAPTER FOUR: A NEW CASE

The last case presented is a new case that has not been considered in the frontier production literature. This case presents a valid interpretation of the unit isoquant where there is neither specification nor aggregation error. This case is fundementally different from a frontier production function because there is no distinction between TE and PE, and the unit isoquant may have

both <u>interior</u> and <u>exterior</u> points. It defines "technical efficiency" as obtaining maximum average physical output. Therefore, the point of "technical efficiency" can be viewed as roughly equivalent to the boundary between Stage I and II in production and is <u>truly</u> efficient only if the firm is in long run equilibrium under perfect competition.

### 1.4.4. CHAPTER FIVE: THE FRONTIER PRODUCTION FUNCTION LITERATURE

Chapter Five deals with the frontier function literature itself. Careful reading shows that TE is due to specification error and/or aggregation error. Only the salient theoretical literature is dealt with since much of the literature is reptitious. The "other literature" is also briefly described. This literature falls into one of two categories, (1) it develops a method for measurement of TE and PE, or (2) it develops related concepts that encounter the same basic difficulties.

## 1.5. OUTLINE OF APPENDICES

The Appendices present two mutually exclusive arguments, each of which raises serious doubts about frontier production functions and all of their ramifications; (1) that a firm can be "technically inefficient" given a set of inputs (system), an output (useful work and any change in the state of the system and its environment), and a sub-production function (processes), and (2), that one may separate physical efficiency (TE) from value efficiency

(PE). Both appendices demonstrate that there can be no relevant logical interior to an isoquant.

### 1.5.1. APPENDIX ONE: EFFICIENCY AND THE LAWS OF THERMODYNAMICS

There is no single term for "technical efficiency" in engineering. Rather there are a number of different types of efficiency which describe aspects of what some economists call technical efficiency. Thermal efficiency is an example of one such technical efficiency. The focus of this appendix is on thermodynamics, the laws of which determine thermal efficiency. Thermal efficiency is defined as the ratio of useful output to costly input [Dixon, 1975], so that in thermodynamics the physical aspect of production is not separated from the value aspects of production.

The first part of the appendix defines the important terms and concepts from thermodynamics.

The second section presents the first and second laws of thermodynamics. This establishes the one to one relationship between total input and total output, i.e., establishes that one cannot get more of the same output with less of the same input using the same processes.

Finally, thermal efficiency, as a case of technical efficiency found in thermodynamics, is described. In the appendix it is shown that, (1) thermal efficiency is an evaluation of output which means TE and PE are inseparable, and (2), thermal efficiency is not the TE of frontier production literature. It is also shown that differences in thermal efficiency can be found only in situations where one is comparing two different sub-production functions (cycles) or two different sets of inputs (systems). Differences in technical efficiency using the same bundle of inputs, and the same production processes, are not possible in thermodynamics.

## 1.5.2. APPENDIX TWO: EFFICIENCY IN ECONOMIC THEORY

Microeconomic theory, simply but clearly stated, has no logical place for the frontier production function concept. Using set theory notation, the first section of the appendix carefully defines terms, including input and output, production functions and sub-production functions, input requirement sets and isoquants, and distance functions.

The second section of this appendix states the fundamental assumptions of production sets. The implications of these assumptions are critical to understanding, (1) why frontier production functions are invalid, and (2) the reality they misrepresent.

The assumption of concavity implies the importance of fixed inputs. In order to evaluate efficiency, something must always be fixed. Therefore, the consequences of fixed inputs are critical to evaluating efficiency.

Each of the relevant consequences of fixed inputs is taken up in turn in this appendix. Without fixed inputs, the law of diminishing returns implies constant returns to scale. Even within this context one may conclude that total input or output must be fixed or constrained in order to discuss efficiency. The law of diminishing returns states that fixed inputs affect the productivity of the variable inputs. The law of diminishing returns is important if one is going to freely dispose of the input/output of one firm in order to make it "comparable" to another firm; precisely what is implicitly done in the frontier production function theory. The last, often over looked, consequence of a fixed input is the existence of Stage III of production. While it is true that additional input always results in additional output (monotonicity), it does not always result in additional useful output, or product. The idea that one can get more or the same useful output with fewer of the same inputs, is only true for a move from Stage III towards Stage II. Stage III also means that marginal physical products can be negative. Thus, producing in Stage III is not economically rational, not because it is not technically efficient nor because it allows one to distinguish between TE and PE, but because it does not pay to be there, which is simultaneously a physical and an economic phenomenon. Therefore, being in Stage III is inefficient.

The assumption of monotonicity means there can be free disposal in production. Free disposal has little if any counterpart in the real world. Theoretically, it exists for perfect complements generally assumed in linear prog

as a

wide

excl

with

Thi

tar

jus

Si

ec

on

ju

li

th

ni

programming. It is used in the conventional theory simply as a means to an end [McFadden, 1978]. In order to provide a rigorous mathematical proof for duality theory, excluding Stage III of production, one ordinarily deals with convex sets rather than convex functions.

> The theory establishing the dual relation between cost functions and production functions was introduced into economics by Shepard (1953), who drew heavily on properties of convex sets discovered by Fenchel (1953). [McFadden, 1978, page 5]

This means one hs an 'interior' to the production (distance) function, since it is a convex set, rather than just a convex function.

> Even though the identical word <u>convex</u> appears in both the term <u>convex</u> <u>set</u> and the term <u>convex</u> <u>function</u>, it has a widely different connotation in each context. In describing a <u>set</u>, the word convex is concerned with whether the set has any holes in it, whereas, in describing a <u>function</u>, the word has to do with how a curve or surface bends. [Chiang, 1974, page 643]

Since an interior is clearly a violation of nature and economic theory, free disposal is used to reconcile being on and off the isoquant at the same time. This is how one justifies an input requirement set, which has an 'interior' of higher order isoquants, as a description of the sub-production function or distance function.

A misunderstanding of free disposal leads to a definition of TE that is in conflict with the laws of diminishing returns. Free disposal does not mean that a production function is a solid rather that a surface, i.e., it does not mean that what one usually conceives of as an isoquant has an interior at the same level of production. The 'interior' of an input requirement set is a collection of true isoquants which is consistent with the fact that given identical sets of inputs one always gets identical sets of outputs. Frontier production theory misrepresents free disposal by applying the concept to only one side of the input:output equation. It suggests one may retain all the input bundles within an input requirement set and equate them to the output represented by the "bounding" lowest isoquant of the set by ignoring ("freely disposing") the additional output of each of the higher isoquants within the set. "Technical inefficiency" implies one does not ignore the additional inputs themselves that are consumed in producing the additional output that is freely disposed, or ignored. That is, the extra inputs are included in the accounting of what is required to achieve the given lower level of output.

It can be shown that a distance function (a generalized production function excluding Stage III that displays free disposal) is from the same set as its cost function. Therefore, there is a one to one mapping from distance functions to cost functions. Factor-price requirement sets in cost space are analogous to input requirement sets in production space, and are a direct consequence of duality theory. Together they are polar reciprocal sets. Polar reciprocal sets mean that free disposal in production space is always accompanied by free disposal in cost space.

The third section briefly describes profit maximizing behavior which constitutes a valid definition of efficiency. The ratio that determines the efficient point of production in economic theory is the same ratio that determines efficient production in thermodynamics; i.e., equating the ratios of the marginal physical products with their respective prices.

The last section explains how and why frontier production functions violate the tenets of microeconomic theory.

## 1.6. SUMMARY

First and last this thesis is a discussion of efficiency. It is predicated on the fundamental fact of physical reality that there is a 1:1 physical relationship between input and output (product, waste, and pollutants) in production; a one to one accounting between everything that goes in to everything that comes out. Given two sets of identical inputs, using the same sub-production function, one cannot get two different amounts of identical output. Therefore, isoquants cannot have interiors at the same level of production. One cannot observe technical inefficiency without simultaneously observing price inefficiency. One cannot distinguish between them or

discuss them separately. One cannot measure technical inefficiency except with respect to price(s) or value. Therefore, there is only one type of economic efficiency, involving the physical and value attributes.

Frontier functions are illegitimate since, (1) they postulate that one may get something for nothing, i.e., more output with the same inputs and nothing else of value in addition, and (2), they attempt to separate technical efficiency from price efficiency.

## CHAPTER TWO

## UNIT ISOQUANTS AND PRODUCTION SOLIDS

In this chapter, frontier production theory, including its distinction between technical and price efficiency will be summarized briefly. The first section, Section 2.1, will develop the concepts of technical efficiency (TE) and price efficiency (PE) by describing the unit isoquant. The frontier production function originated as a unit isoquant in Farrell (1957), and is still often so described. Much of the first section follows the presentation of frontier production functions made by Bressler (1966).

The second section, Section 2.2, will describe two cases of frontier production functions that are not explicitly sets of unit isoquants, but retain the essential characteristic of the unit isoquant: there is an interior to the production surface (frontier) that is made up of "technically inefficient" points of production. Without an interior so characterised, there can be no frontier production functions. The laws of diminishing returns and utility indicate that the interiors described by frontier production functions cannot exist (see Appendices One and Two).

func

posi

diff

in

sep

or

In

pro

2.1

cle

193

197

in

pr

re

ti

ar

th

ti

re

un

th

Section 2.3 explains why the frontier production function interpretation of these interiors arose from the misinterpretion of observed phenomena. Apparent observed differences in firm performance attributed to differences in technical efficiency, are in fact due to comparing two separate sub-production functions (specification error), or aggregating heterogeneous inputs (aggregation error). In neither of these cases do identical sets of inputs produce different levels of identical output.

## 2.1. UNIT ISOQUANTS

### 2.1.1 WHAT IS "TECHNICAL EFFICIENCY?"

In much of the frontier literature, technical efficiency (TE) means being on a "unit isoquant [Farrell, 1957, Bressler, 1966, King, 1980, Nerlove, 1965, Timmer, 1971]."

Consider an input requirement set with two variable inputs $v_1$ and $v_2$, and one fixed input $z_3$, that together produce some output Y. Since there is only one input requirement set, one is implicitly considering identical sub-production functions. The inputs and outputs are assumed to be homogeneous. These conditions assume that technology is fixed and identical for all observations. The technology of this production process is represented by the unit isoquant $y_0$ in Figure 2.1. The unit isoquant is found by dividing the input quantities by the output quantities they produce, i.e. the unit isoquant

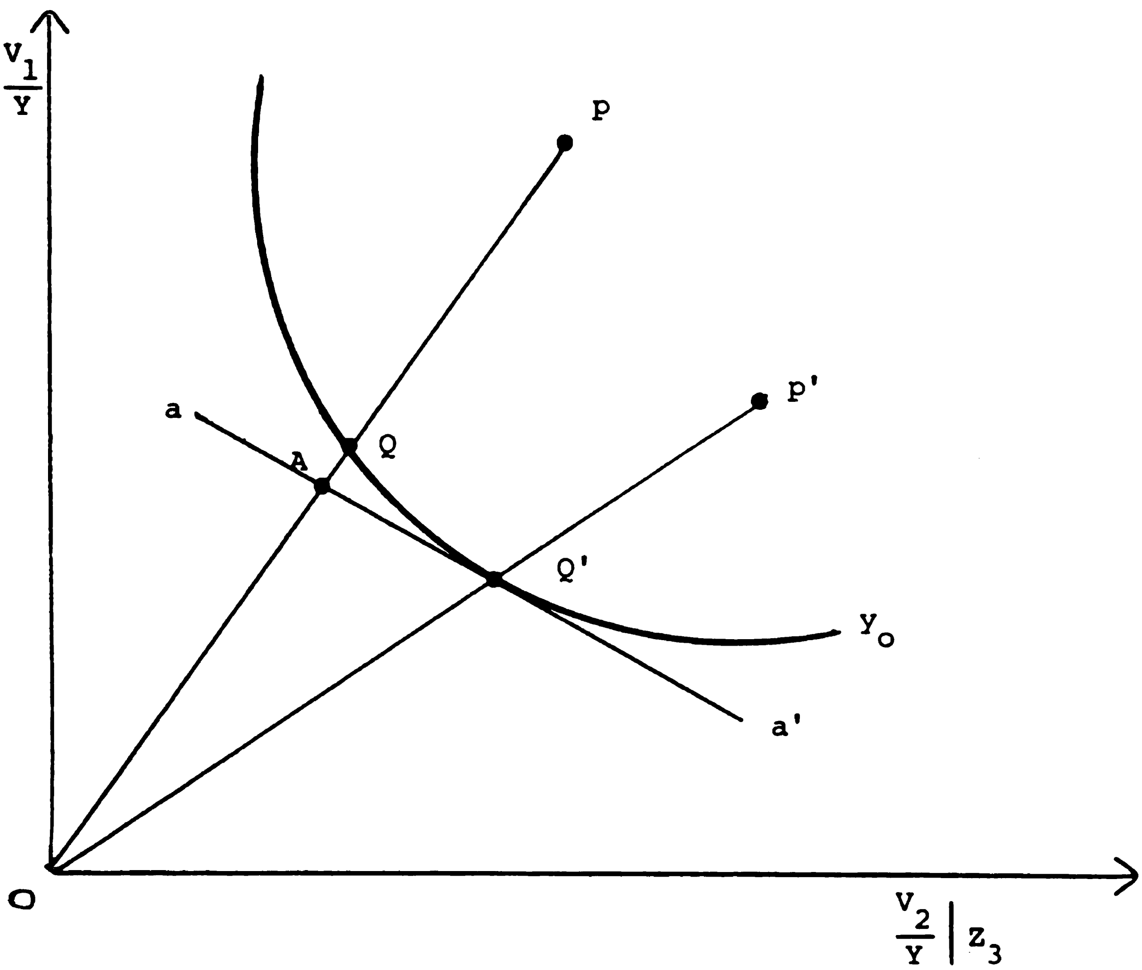



FIGURE 2.1

A UNIT ISOQUANT WITH BUDGET A CONSTRAINT maps the average variable input required per unit of output (not the marginal input bundle that produces one unit of output at the margin). Such a mapping can produce observations located at widely varying positions in the quadrant. The points lying closest to the axes are connected to produce the unit isoquant, or the "technically efficient" isoquant. That is, all points on this unit isoquant are regarded as "technically efficient," while all points within it are regarded as "technically inefficient." Consider an observation on firm P. The firm uses the inputs $v_1$ and $v_2$ in the same relative proportions as the firm Q, but uses more of both $v_1$ and $v_2$ on average in producing a unit of output as Q; P gets less output per unit of variable input on average than Q.

> The distance OP relative to OQ measures the extent to which the same amount of output could be produced with fewer inputs used in the same proportion.... [Nerlove, 1965, page 88]

The ratio OQ/OP is the measure of technical inefficiency. This means that P can produce the same quantity of output, with less of the same homogeneous inputs, using the same processes or sub-production function, merely by becoming technically efficient. TE means getting different output given the same inputs and the same production processes. The firms on the unit isoquant are technically (physically) efficient since they produce on average the most output per unit of input, or conversely, they use on average the least input per unit of output. The interior points are technically inefficient because they use greater physical quantities of both $v_1$ and $v_2$ on the average to produce a unit of output.

### 2.1.2. WHAT IS "PRICE EFFICIENCY?"

Technical efficiency does not imply price efficiency in frontier production theory since they are mutually exclusive phenomena. Price efficiency (PE) implies inputs being used in their least cost combination ratios. It reflects the proportions of $v_1$ to $v_2$ but not necessarily the quantities. In Figure 2.1 Q and Q' are both technically efficient, but only Q' is price efficient since it is tangent with aa', the budget constraint. The ratio that reflects the degree of price efficiency for both P and Q is OA/OQ.

> The distance [OA] <u>relative</u> to OQ measures the fraction of costs for which the output could be produced if the <u>relative</u> use of inputs were altered. [Nerlove, 1965, page 88, underlining added]

### 2.1.3. WHAT IS "ECONOMIC EFFICIENCY?"

Economic efficiency (EE) in the frontier production function literature is defined as the product of technical efficiency and price efficiency:

$$(2.1) \quad EE = TE * PE = OQ/OP * OA/OQ$$
$$= OA/OP$$

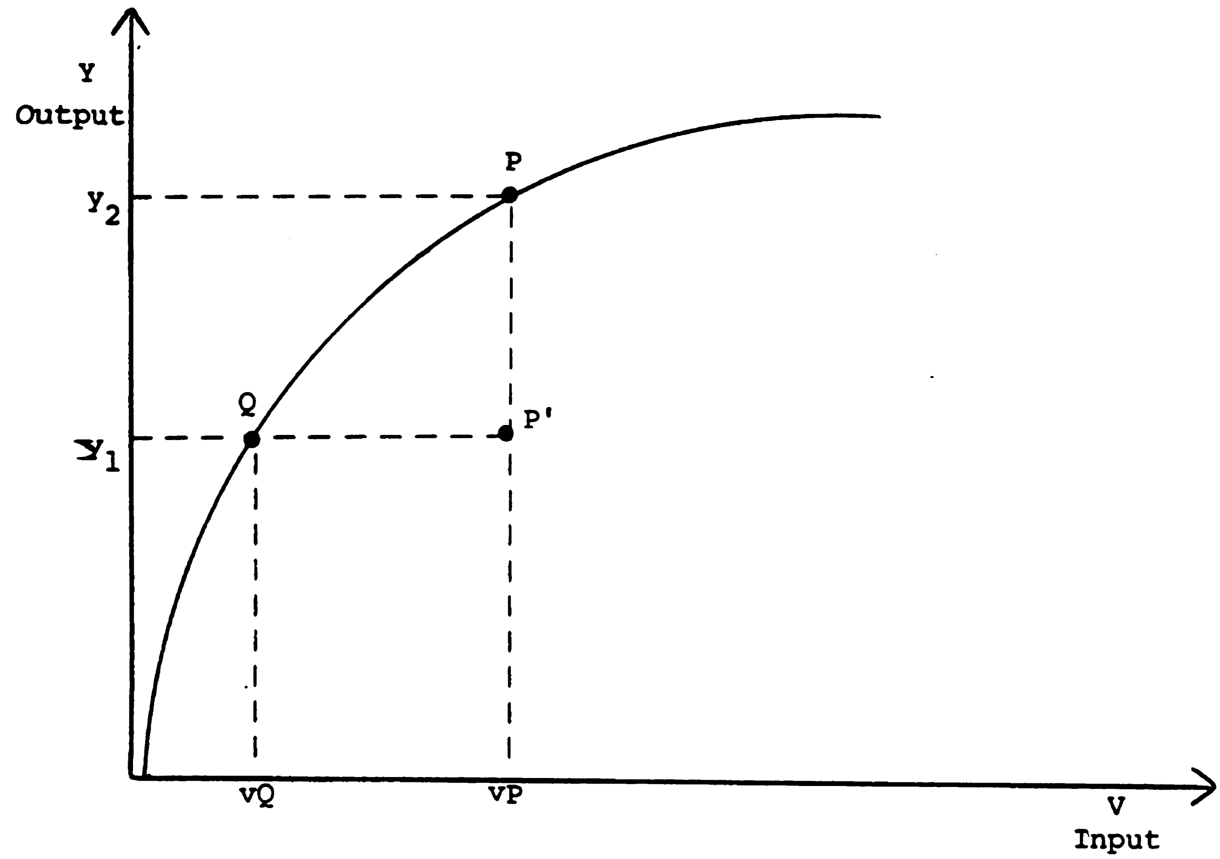



FIGURE 2.2

TECHNICALLY EFFICIENT AND INEFFICIENT POINTS OF PRODUCTION WITHIN A PRODUCTION POSSIBILITY SET

While this formulation makes a distinction between technical and price efficiency, economic efficiency remains connected with prices or values.

> Note that price and economic efficiency measures are in fact cost comparisons: The "price line" aa' represents total expenditure per unit of output, with slope representing the inverse ratio of the given factor prices; lines parallel to aa' through Q and P represent the higher unit expenditures. Thus, the economic or overall efficiency ratio is equivalent to the ratio of the average cost of producing at Q' to the average cost of producing at P. [Bressler, 1966, page 130]

In order to apply this theory to "reality," one must collect observations on firms producing the same amount of the same output (or assume constant returns to scale), using the same technology (same sub-production function), and different amounts of the same variable inputs. If one could find such observations, the next step would be to plot this unit requirement in the positive quadrant and connect the observations closest to the axes. These connected observations define the unit isoquant from which all efficiency comparisons are made. A set of such unit isoquants, for different levels of production, define what is called the <u>frontier production function</u>.

### 2-1.4. SUMMARY OF UNIT ISOQUANTS:

The essential and critical aspect of the unit isoquant as described above is the existence and interpretation of the interior points. Consequently, frontier production functions are valid only if one, or both, of two things are true about reality.

First, given two sets of variable inputs that differ only in quantity, e.g., one set is "a" times larger, and using the same fixed inputs, one may produce the same quantity of identical output. That is, different quantities of given inputs produce the same output using the same technology. It is producing a different quantity of output from identical inputs and technology that explains the existence of the interior to a unit isoquant for a given level of output. Without an interior to the unit isoquant for firms producing the same output there is no need for the concept of a frontier production function. Consequently, the existence and the nature of this interior is critical to the interpretation of the frontier production function.

Secondly, one must be able to separate TE from PE. In the frontier production formulation, TE is purely a physical concept, PE is purely a value concept, and TE and PE are observable mutually exclusively of each other. Therefore, measurements of physical quantities have no intrinsic relationship with their corresponding prices or values. In Figure 2.1 for firm P, TE is measured at the point P, while PE is measured at point Q.

Though TE and PE are regarded as mutually exclusive, TE and total cost may not be. That is, the measurement of PE is the same for firms P' and Q'. Both firms P' and Q'

are price efficient since the measurement of PE = OQ'/OQ' are the same for both. Since P' uses more of both $v_1$ and $v_2$ in total and on average to produce one unit of output, the total cost for P' must be higher than for Q'.

Current frontier production function literature often deals with a frontier production function that is not explicitly defined as a collection of unit isoquants since there is no averaging process. This type of frontier production function clearly represents a production set that is solid, rather than the surface that one ordinarily associates with a production function in traditional production theory.

## 2.2. PRODUCTION SOLIDS

It should be evident from the description of the unit isoquant given above that the production set represented by a frontier production function is characterized as a surface (the traditional production function) and its interior which is a set of "technically inefficient" points of production. The two cases of production sets that are solids, rather than shells, presented in this section retain this critical aspect of unit isoquants, while ignoring the average aspect of unit isoquants. Therefore these two cases highlight the critical aspect of a frontier production function that its interior is a set of technically inefficient points. In so doing, it demonstrates the deficiency of frontier production functions by showing that this interior is either physically impos-

ble, or a misinterpretation of what is observed. Both ses fundamentally revolve around specifying the appro- iate sub-production functions, and the economic ad- stment of investment and disinvestment implicit in ving from one sub-production function to another. Both ses arise from either committing a specification error d/or an aggregation error in identifying the sub-produc- on function described by the set of observed production ints. These are empirical issues rather than theoreti- l ones.

Consider the production possibility set represented Figure 2.2. In Figure 2.2, points Q and P are

> ... technically efficient in the sense that they are in the production possibility set, and there is no way to obtain more output than depicted by these points without using more of the input. Point [P'] is technically inefficient, in that more output could be obtained with no more input. [Jamison and Lau, 1982, page 54].

"frontier" of this frontier production function is esented by the surface upon which points Q and P lie. t P' represents a point within the solid interior to frontier. P' is technically inefficient since one d get the same amount of output, $Y_1$, "with no more t;" i.e., by using less input and producing at point Additionally, point P' is technically inefficient, "in more output," $Y_2$, "could be obtained with no more same] input," by producing at point P. In both cases it is implicit that neither does the total input bundle nor does the technology change, only the degree to which the technology is used technically efficiently. That is, the input $v_P$ can produce either $y_1$ or $y_2$ using the same technology merely as a matter of technical efficiency.

### 2.2.1. CASE 1:

Case 1 is the obvious case of a solid production possibility set for the variable inputs given the same technology and the same fixed inputs. That is, the technology used and the quantities of the same fixed inputs are identically equal at points Q, P, and P'. An obvious difficulty with this is that both points P and P' use the same input, $v_P$, but get different outputs. Since the variable input is the same, the fixed input is the same, and the technology is the same, there is nothing to account for the marginal physical product being positive at P and zero between P and P'. Jamison and Lau (1982) explain the difference as follows:

> Technical inefficiency results from combining available inputs poorly; for example, by plowing the insecticide into the ground or spraying fertilizer on the plant leaves. [Jamison and Lau, 1982, page 54]

There are two problems with this explanation. The first is, one can increase TE without additional cost, since one simply uses less of the same variable input to produce the same output, i.e. moving from P' to Q [Bagi and Huang, 1983]. In frontier production theory the only physical (technical) difference between Q and P' is the quantity of variable input used. In fact, improving TE should result in a savings since one is using less costly variable input. Therefore, one can increase profit at no cost; get something for nothing. Therefore, either there is no opportunity cost to instituting the changes that will bring about the improvement in TE, or the net opportunity cost is always positive by definition and no marginal analysis is needed to consider changes to improve TE. In the examples quoted above: there is no cost in taking care not to plow any insecticide into the ground nor in taking care not to spray any fertilizer on the leaves.

The second is that in traditional theory, inputs are defined with respect to time, location, and quality (see Appendix Two), so that one cannot combine inputs "poorly" or "better" since any particular combination of inputs is time, location, and quality specific. Changing any of these aspects of the combination means changing the inputs by definition. How inputs are combined in practice is an applied problem and not a theoretical one. The frontier production literature confuses theory and application. Insecticide plowed into the ground is not the same homogeneous input as insecticide not plowed into the ground because they are not in the same location. Similarly, fertilizer sprayed on the plant leaves is not the same as fertilizer sprayed on the ground. The problem of treating heterogeneous inputs as "homogeneous" groups for the purposes of analysis is a specification or aggregation problem. It is a problem in application that is assumed away in theory by the homogeneity conditions.

The <u>theory</u> of frontier production functions maintains that the same homogeneous inputs (systems) can be used with the same sub-production function (processes) to get different output (actual work and/or final states). This is clearly in violation of the laws of thermodynamics (see Appendix One) and diminishing returns (see Appendix Two). Therefore, Case 1 is physically impossible; i.e., interiors to unit isoquants for the same levels of total output.

### 2.2.2. CASE 2:

The second case is the case of a traditional production possibility set, where all the points are also within the feasible production set. Therefore, if a point P' uses more of the same homogeneous variable inputs, and the same technology as point Q, and only gets the same amount of identical output, then P' and Q must be on different sub-production functions; i.e., be using different amounts or kinds of fixed inputs. That is, point P' represents the 'interior' to the production set, but only because it is on a different sub-production function than Q and P. The interior in Figure 2.2. is in reality a collection of sub-production functions, for each of which the level or nature of the fixed input is different. Because P' uses less of the fixed input than Q, output of the variable input is less by the law of diminishing marginal returns. Consequently, for Figure 2.2 there are two problems.

The first is an indexing problem, due to the fact that the different levels of the fixed input are not clearly specified for Q and P, and P'. The input bundle $v_P$ at P is different than at P'. Therefore, P' is not "technically inefficient." At P' one cannot get more output with the same ("no more") input, i.e., move to P, because P' does not have enough fixed input. Nor can P' get the same output by using less ("no more") variable input, i.e., move to Q, for the same reason.

The second problem with Figure 2.2 is related to the first. Figure 2.2 represents a case of not properly treating the economics of adjusting the use of the "fixed" inputs. The problem of making an adjustment with respect to fixed inputs, the analysis of investment and disinvestment, has been treated by Edwards (1958). Clearly there is a cost in moving from P' to either Q or P since P' must invest in more of the fixed input to make the change. The economics of this adjustment is totally ignored in the frontier production function literature which treats such adjustments as costless technical changes.

## 2.3. THE EXPLANATION FOR "OBSERVING" CASES 1 AND 2

In reality, when one makes observations on firms Q, P, and P', one is observing differences between the firms that are due to either <u>specification error</u> or <u>aggregation error</u>.

The specification problem, involving the amounts of which variables are included and which ones are not, is almost indistinguishable from the aggregation problem. When inputs are defined as homogeneous with respect to time, form, and space, production systems and processes become interdependent. It is also due to output being both useful output (work) and the changes in the final states from the initial states, one a function of the processes (work) and the other a function of the systems' properties (see Appendix One).

### 2.3.1. THE SPECIFICATION PROBLEM:

Ordinarily, a comparison of firms assumes all firms are using the "same" sub-production function. Specification means identifying the sub-production function clearly and accurately. Specification error means one has inaccurately identified two different sub-production functions as being the same. That is, one commits a specification error in classifying the the firms as representatives of the same sub-production function. A comparison between firms is valid only if their sub-production function(s) are properly identified, so that differences between firms can be properly explained. If differences are due to the firms having different sub-production functions, then those differences must be properly identified. The specification problem is an empirical problem not a theoretical one. In applied work specification problems arise in several ways.

One way one creates a specification problem is by aggregating observations across sub-production functions in specifying the sub-production function to be analyzed. This means that inputs that are <u>fixed</u> at the observation level, e.g. tillable acres for crop production, are <u>assumed</u> <u>variable</u> within the sub-production function, i.e., across all observations.

Another way is to ignore the <u>real</u> impact of some unspecified input(s). One generally avoids this problem precisely by assuming that the uncontrollable random variables (the unspecified variables) are equal to some "average" value across all observations when in fact they are not, so they can be treated like fixed inputs, i.e. having the "same" impact on all observations.

A third way of mis-specifying a sub-production function is to ignore an input entirely, or to aggregate two inputs inconsistently (this latter situation is also considered an aggregation problem, which is considered in the following section). If one omits a relevant variable input from the specification of the sub-production function, a variable input that is present in the production processes in different amounts across observations, and this omitted variable input is correlated with one or more included variable inputs, then the marginal physical products for identical quantities of the included variable inputs will appear to be different due to their including the additional marginal product of the unspecified variable input. This means that the TE observed is due only to the differences in the marginal physical products for the unspecified variable input. Note that the differences in the quantities of the random uncontrollable variables also has this effect. If the unspecified input is a fixed input, and it is present in different amounts across observations, then one is measuring and comparing marginal physical products for the variable inputs across different sub-production functions. The "technical inefficiency" is not inefficiency, but a difference in the productivity of the variable inputs due to the point of diminishing marginal returns for the variable inputs starting at different points for the different levels of the fixed input.

### 2.3.2. THE AGGREGATION PROBLEM:

The aggregation problem is also a real world problem rather than a theoretical one. In order for analysis to be legitimate, one need only insure that this error is within some recognized tolerable bound prior to conducting the analysis, i.e., the inputs aggregated are "very close."

Aggregation error is suspect when one finds differences in marginal physical products for "homogeneous" inputs. One reason may be specification error as described above. The other reason the same amounts of two "like" inputs may have different marginal physical products, when measured with respect to the identical amounts of fixed, or unspecified, input, is they are not in fact homogeneous; they are not "very close" to the "same" input. Aggregation error means inaccurately identifying heterogeneous inputs as homogeneous inputs. Economists recognize that labor is not homogeneous across laborers, but in applied work this labor is aggregated and specified as one input. This results in error being introduced into the analysis due to the differences in quality (form), time, or space of the inputs aggregated.

In either case, specification or aggregation, the assumptions one makes about the error term (the unspecified variables), the inputs which are fixed (which sub-production functions are included), and/or how inputs are combined into "same" inputs implicitly means that the differences in marginal physical products that do in fact exist are "unimportant" to the analysis.

## 2.4. SUMMARY

This chapter has outlined the theory of frontier production functions as unit isoquants and production solids. Frontier production function theory suggests production within the interior of an isoquant at the same level of production is possible so that an isoquant is a plane rather than a line. That is, one can produce different amounts of the same output using identical amounts of the same inputs and the same technology. Interiors are critical since they allow one to separate TE and PE as mutually exclusive phenomena which implies that TE $\not\equiv$ PE. It also means that technical efficiency is purely physical, which is contrary to the laws of nature, e.g., the laws of thermodynamics. This means that the physical quantites of inputs used in production, and the value of those inputs in production do not have an intrinsic one-to-one correspondence, since they cannot have the same marginal value products if they are not producing the same output.

If one uses more inputs in one situation to produce the same amount of work as in another situation, those additional inputs have an opportunity cost. Yet the total value of the bundles must remain the same if they produce equal output.

If one is measuring a "frontier production function," specification error or aggregation error accounts for the discrepancies in observations attributed to "technical inefficiency." That is, "technical inefficiency" is a measurement of specification error or aggregation error. If the two sets of "identical" inputs produce different amounts of work, and if one tries to compare them by insisting that they are from the same input requirement set, using the same processes or sub-production function, then one has committed either an specification error, they do not belong to the same input requirement set, or an

ggregation (measurement) error, they are not the same
nputs used in the same proportions. It is a logical
osurdity to conduct an analysis where one creates or
ntroduces discrepancies between observations by mis-
pecifying the function or by aggregating heterogeneous
nputs, measures the discrepancy, calls this discrepancy
ignificant, and attributes it to some inherent differ-
nce (TE) between the observations. Yet, this is exactly
hat is done in order to obtain measurements of "technical
nefficiency" between firms.

## CHAPTER THREE

## FRONTIER PRODUCTION FUNCTIONS AND DISTANCE FUNCTIONS

In this chapter two hypotheses will be tested. The first hypothesis is that frontier production functions and distance functions are the same. The second hypothesis is that technical efficiency (TE) and price, or allocative, efficiency (PE) are not the same. These two hypotheses are both tested in this chapter because each has a bearing on the other. If it is true that frontier production functions are distance functions, then the property of frontier production functions that TE and PE are not the same will be true by transitivity. Conversely, if TE and PE are not the same, as frontier production function theory maintains, then this condition will be consistent with duality theory as demonstrated by distance functions.

The reason it is important to test the first hypothesis, that frontier production functions and distance functions are the same, is that distance functions appear to provide a legitimate theoretical basis for the existence of frontier production functions within the current set theoretic approach to microeconomic theory [Malinvaud, 1972, McFadden, 1978, Quirk and Saposnick, 1968, Varian, 1978]. It might appear that a frontier production function is simply a special case of a distance function;

that frontier production functions and distance functions arise from the same theoretical basis and display the same properties. If frontier production functions are distance functions, then they are legitimate functions and all their conditions, notably that TE and PE are not the same, are valid.

The reason it is important to test the second hypothesis, that TE and PE are not the same, is two fold. First, the apparent difference between TE and PE is a consequence of frontier production functions. If TE and PE are not mutually exclusive, the two terms as they are now used in the frontier production function theory are confusing, inconsistent, and misleading because they suggest that there is more than one type of efficiency, when in fact there is only one type of efficiency [Knight, 1933].

In order to test the first hypothesis, it will be shown that frontier production functions and distance functions are not the same because they have different theoretical origins.

In order to test the second hypothesis it will be demonstrated that the theory of frontier production functions is incompatible with duality theory and free disposal (or monotonicity). The duality principles implicit within a distance function prove that TE $\equiv$ PE, while free disposal eliminates the relevance of any 'interior' points to the lowest bounding isoquant of an input requirement set.

This chapter will rely heavily on duality theory as
esented by McFadden (1978). The purpose of McFadden's
978) discussion of microeconomic theory is to provide a
ormal mathematical duality between [distance] and cost
nctions...." A modified presentation of this theory is
tlined in Appendix Two. An important consequence of
ality theory is that a distance function and a cost
nction have a unique one to one correspondence with each
her which can be conceptually and graphically represen-
d by what are termed "polar reciprocal sets [McFadden,
78]."

## 3.1. TEST OF THE FIRST HYPOTHESIS

### 1.1. WHY FRONTIER PRODUCTION FUNCTIONS APPEAR TO BE DISTANCE FUNCTIONS:

A superfical interpretation of a distance function
ggests a similiarity with a frontier production
nction. Mathematically, a distance function is defined
McFadden (1978) as:

(3.1) $F(y, v) = \text{Max}[\ a > 0 \mid 1/a * v \in V(y)]$

difference between this definition and the one found
equation (A2.8) in Appendix Two is that (3.1) neglects
distinguish between variable and fixed inputs. In
gure 3.1A, TE = OQ/OP. In Figure 3.1B, a = A/B, so that

(3.2) $1/a = B/A = ov'/ov$

both cases $y_o$ looks like the "frontier" of a frontier
oduction function so that TE = 1 when P = Q while "a" =

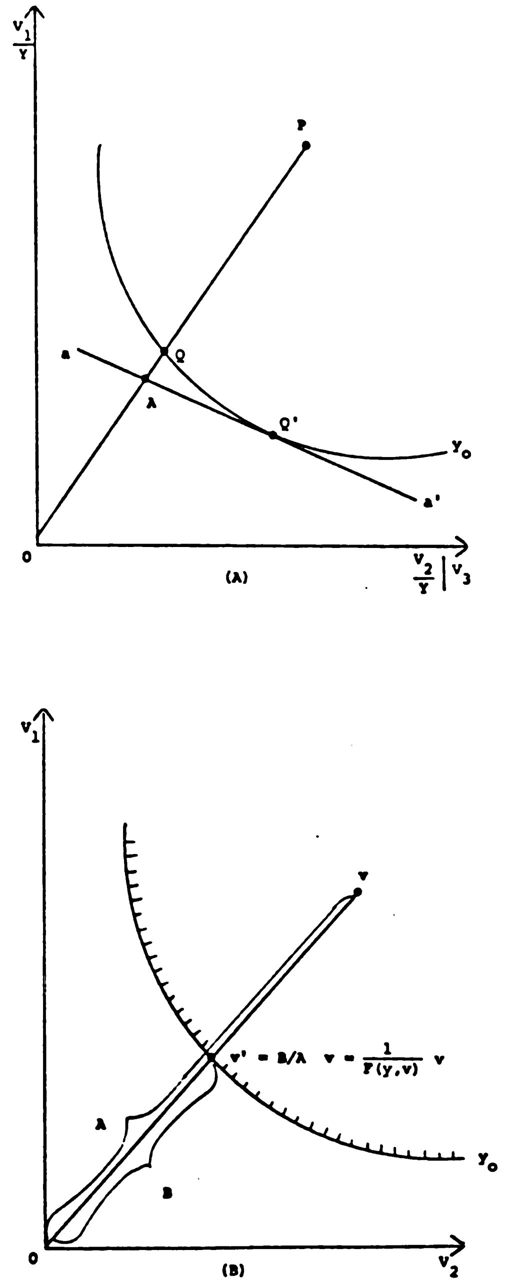



FIGURE 3.1

A UNIT ISOQUANT DERIVED FROM A FRONTIER PRODUCTION FUNCTION (A) AND AN INPUT REQUIREMENT SET DERIVED FROM A DISTANCE FUNCTION (B)

1 when v = v'. In addition, consider the following quotes:

> While reformulation of duality in terms of distance functions is potentially useful in application, its primary appeal comes from the fact that it allows us to establish a full, formal mathematical duality between [distance] and cost functions, .... [McFadden, 1978, page 24, underlining added]

> It is sometimes useful to extend the definition of the distance function to all non-negative input bundles v by applying the formula [3.1] provided v/[a] is in V(y) for some positive scalar [a], and setting F(y, v) = 0 otherwise.... In applications, it is sometimes useful to employ this extended definition of the distance function. [McFadden, 1978, page 28]

> When y contains more than one element, efficient production of y can be described in terms of the distance function
>
> [F](y, v) = max[[a]>0 | 1/[a]*v ∈ V(y)]
>
> for (v, y) Y and v strictly positive; the frontier satisfies [F](y, v) = 1. [Fuss, McFadden, Mundlak, 1978, page 227]

From this one could easily, but mistakenly, infer a basis for frontier production functions. Unfortunately, some authors make the connection directly:

> In addition, admissable frontier functions must be continuous, quasi-concave, and exhibit strong free disposability of inputs. [Kopp and Diewert, 1982, page 322]

Notice that the conditions for the "frontier functions," continuity, quasi-concavity, and free disposal, are exactly those conditions that apply to distance functions (see Sections A2.1.2.4. and A2.2 in Appendix Two).

### 3.1.2. FRONTIER PRODUCTION FUNCTIONS ARE NOT DISTANCE FUNCTIONS:

Frontier production functions are not distance functions because the two types of functions have different origins; they are derived from different theoretical premises. A frontier production function is derived from a production solid. The frontier production function describes the surface of a production solid that includes both "technically efficient" and "technically inefficient" points. "Technically efficient" points are those points that are on the production surface (or function) and are supposed to represent those points of production for which it is true that the firm represented by that point "yields the greatest output for any set of inputs, given its particular location and environment [French, 1977, page 94]." What is frequently not made explicit, but can be easily inferred from this definition of "technical efficiency," is that a firm can be "technically inefficient" which is represented by a point within the interior of the surface of the production solid at the same level of output. That is, "technical inefficiency" means that a firm 'yields <u>less</u> that the greatest output for any set of inputs, given its particular location and environment.' (It will be shown in testing the second hypothesis that this distinction between "technically efficient" and "technically inefficient" points of production is inconsistent with reality, and therefore confusing and misleading.)

A distance function does not originate as a description of a surface to a production solid. A distance function is a true production function or surface. McFadden's (1978) fundamental purpose in defining the distance function is to obtain the 'interiors' to isoquants necessary to satisfy the convexity conditions required to make his "formal" mathematical proof to duality theory. But, there are no true or observable interior points of production that represent a solid interior to a distance function. The 'interior' to a distance function is represented in Figure 3.1B by point v in $V(y_0)$. This is not the same 'interior' as the interior to a frontier production function, point P in Figure 3.1A. Properly understood, the 'interior' of a distance function is not that of a solid (points at the same level of production), but is composed of higher order isoquants (see Appendix Two). That is, point $v \in V(y_0)$ but $f(v) \neq y_0$. Rather $f(v) = y_1$ where $y_1 > y_0$. In addition, this 'interior' is 'swept out' for <u>all practical purposes</u> by invoking free disposal. Free disposal is used to make $f(v') = f(v)$. Therefore, any 'interior' points from production solids that might exist within a distance function are for the sake of mathematical convenience physically and economically superfluous in any description of reality. Free disposal eliminates _all_ the differences between v and v' whatever their origin.

Frontier production functions are different than distance functions and the first hypothesis is rejected. Frontier production functions describe the surface of a production solid. The existence of observable interior points to an isoquant at the _same_ level of production is a fundamental condition for frontier production functions. Only the _observable_ interior points allow one to identify "technical inefficiency;" that is, being in the interior of a surface isoquant of a frontier production function. This means that the isoquants for the complete production set are planes rather than lines. This is not true for a distance function which obtains its 'interior' points by mapping higher order isoquants into a lower one. The 'interior' to the lowest bounding isoquant of an input requirement set is not made up of additional points of equal production, but of points of higher or greater production. That is, an input requirement set is created conceptually by mapping the production surface of three dimensional space into a plane. It looks like a conventional isoquant map. It is true that every point within a given input requirement set, designated by the lowest bounding isoquant, produces "at least as much" output as the points on the lowest bounding isoquant, since all the points within the lowest bounding isoquant are points on higher isoquants which therefore produce more ("as least as much") output.

## 3.2 TEST OF THE SECOND HYPOTHESIS

The second hypothesis to be tested is that TE is not the same as PE. Because a frontier production function describes the surface to a production solid, frontier production function theory distinguishes between two separate and mutually exclusive types of efficiency, TE and PE. In frontier production function theory, TE is getting the most output from a given set of inputs, given a process, being on an isoquant rather than in the interior of an isoquant. PE is using inputs in their least cost combination ratio. The condition that TE $\neq$ PE has two implications. The first is that one can have two firms that are both price efficient when only one of them is technically efficient while the second is that one can have two firms that are both technically efficient when only one of them is price efficient.

The second hypothesis will be tested in two steps. The first step will test whether or not the first implication, that one can have two firms that are both price efficient when only one of them is technically efficient, conforms to the principles of duality theory and free disposal. The second step will test whether or not the second implication, that one can have two firms that are both technically efficient when only one of them is price efficient, conforms to the principles of duality theory.

It will be shown that both implications violate duality. Conversely, it will be shown that TE $\equiv$ PE. Additionally, it will be shown that the term "technical efficiency (TE)," as it is used in the frontier production function theory, is ambiguous, confusing, and misleading.

### 3.2.1. THE CONDITIONS FOR THE TESTS:

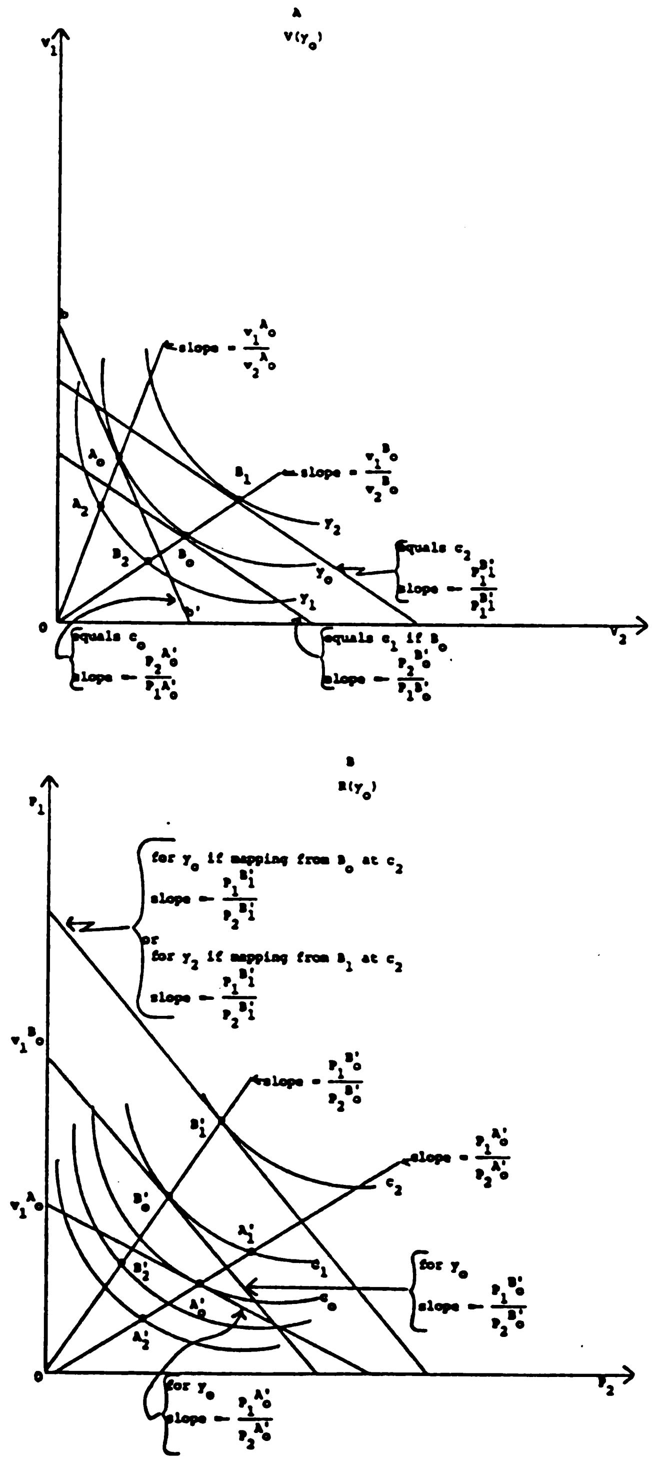
Whether measuring distance functions or frontier production functions, one assumes that any two firms, "A" and "B", are using the <u>same</u> process and the <u>same</u> cost function, and that the principles of duality apply to the distance (production) or frontier production function, and the cost function, which is to say, that for both firms their production or cost functions are identical. Therefore, any technically efficient point on a frontier production function is implicitly associated with corresponding prices in cost space.

Duality between product space and cost space means that if one holds output constant and changes prices, e.g., by a scalar, then one maps from the same isoquant in product space to different isocosts in cost space. Similiarly, if one holds total cost constant and varies input quantities, e.g., by a scalar, then one moves from the same isocost line in cost space to different isoquants in product space. Finally, if one holds both output and total cost constant then duality means one isoquant maps to one isocost line (this is a special case). If one holds neither output nor total cost constant, then duality does not apply since points in product space can map to more than one point in cost space, and vice versa.

Polar reciprocal sets represent the set of points for which duality exists between points in input (product) space (excluding Stage III) and cost space. In order to define the full set of mappings one must assume that both price and input ratios vary in both spaces simultaneously. This means that a duality mapping is always between a point that is technically efficient in input space and a point that is price efficient in cost space. All points in cost space represent least cost bundles by definition of a cost function.

The theory of frontier production functions assumes that firms are purchasing their inputs in perfectly competitive markets so that input prices are fixed and the same for all firms. This means that only one price ratio defines PE, the ratio of the given fixed prices.

Consider the polar reciprocal sets of a distance function and what frontier production functions might suggest is true that would assure that frontier production functions conform to the same principles of duality that distance functions do (see Figure 3.2). Figure 3.2A, V(y), is the input-conventional input requirement set from a distance function and 3.2B, R(y) is the input conventional factor-price requirement set from a cost function. Together they represent the distance function and the total cost function respectively for both firms "A" and




FIGURES 3.2

AN INPUT REQUIREMENT SET V(y) FROM A DISTANCE FUNCTION (A)
AND A FACTOR-PRICE REQUIREMENT SET R(y)
FROM A COST FUNCTION (B)

"B." Let points $A_0$ and $B_0$ represent two firms "A" and "B" respectively. Either all the firms on the isoquant $y_0$ produce at the same total cost (a special case), or one firm is the least cost producer. Assume that for output $Y_0$ firm "A" (point $A_0$) is the least cost producer. The prices for which $A_0$ has a mapping in cost space are $P_1^{A'0}$ and $P_2^{A'0}$; i.e., $A_0$ maps to $A'_0$ in Figure 3.2B at output $Y_0$ and at a total cost of $c_0$. Similiarly, $B_0$ maps to $B'_0$ at a total cost $c_1 > c_0$. In Figure 3.2A, the ray $OA_0$ represents a path along which the ratio of the prices for the inputs $(P_2^{A'0}/P_1^{A'0})$ remains constant. At point $A_0$ in Figure 3.2A the line bb' represents the budget constraint with slope $- P_2^{A'0}/P_1^{A'0}$. Point $A_0$ maps to point $A'_0$ in Figure 3.2B when the input prices are $P_1^{A'0}$ and $P_2^{A'0}$ and total cost equals $c_0$. Point $B_0$ in Figure 3.2A maps to point $B'_0$ in Figure 3.2B when the input prices are $P_1^{B'0}$ and $P_2^{B'0}$, and total cost $c_1$, and to point $B'_1$ when input prices are $P_1^{B'1} = kP_1^{B'0}$ and $P_2^{B'1} = kP_2^{B'2}$, and total cost equals $c_2 = kc_1$ (where k equals some coefficient of proportionality greater than 1.0). Therefore, the ray $OB'_1$ in Figure 3.2B represents the locus of points in cost space that map from $B_0$ in product space as prices and total cost increase in proportion. Notice that at any point on the ray $OB'_1$, the ratio of input quantities $(v_1^{B_0}/v_2^{B_0})$, remains constant since all points on the ray map from $B_0$. This is analogous to all the points on the ray $OA_0$ having a constant price ratio. Point $A_0$ maps to $A'_0$, and point $B_0$ maps to $B'_0$, and vice versa, by duality.

The price level is the set of prices that will keep the total cost relationship between firms constant when output is held constant but the ratio of input quantities used varies. Notice again that if the general price level faced by "A" and "B" were to increase by $|1-k|$, $A_0$ would map to $A'_1$ and $B_0$ would map to $B'_1$. That is, relative prices and relative total cost would remain the same in comparing "A" and "B," but the ceteris paribus conditions would not be violated only if the prices for both "A" and "B" increased by the same proportion $|1-k|$; if the prices paid by "A" and "B" remain of the same magnitude. The magnitude, or price level, paid by the firms is important since "A" and "B" would both continue to produce $y_0$ at the higher price level only if the price for y were also increased by the proportion $|1-k|$. Otherwise, "A" and "B" would produce $y_1$, $y_1 < y_0$, at $A_2$ and $B_2$ respectively since the income from the sale of the output would only allow them to purchase a lesser amount of both $v_1$ and $v_2$ at the higher input price level.

In demonstrating why duality theory refutes frontier production functions, close attention to three aspects of duality theory is especially important: (1) Both the input (product) space and the cost space for the firms considered must conform to the assumptions of being input-conventional. This means the input requirement sets and the factor-price requirement sets for the firms considered must display not only regularity conditions, but also display (i) monotonicity, or free disposal, and (ii) strict convexity from below [McFadden, 1978].

> 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 the 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. [McFadden, 1978, 22]
>
> ... 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. [McFadden, 1978, page 26]

(2) The dual prices represented in the cost space of the duality mapping are the prices which would make the purchased input bundle the least cost bundle. The prices associated with a particular input bundle may or may not be market prices. Market and dual prices are not necessarily equal, and may necessarily be different. One must keep very close track of whether or not prices are market prices, dual prices, or both. Between firms for which duality is assumed to hold, only the dual prices are relevant for the purposes of making comparisons between the firms whether or not they are market prices. (3) Finally, and most importantly, the <u>level</u> of prices must

remain constant across firms being compared. This means that the difference between the dual prices of two firms may be different with respect to their ratios, but one firm cannot have dual prices of a greater absolute magnitude than another. To do so would violate the usual ceteris paribus conditions. For example, firm "A" may have dual prices $(P_1{}^A0,\ P_2{}^A0)$ and firm "B" dual prices $(P_1{}^B0,\ P_2{}^B0)$ where $P_1{}^A0/P_2{}^A0 \neq P_1{}^B0/P_2{}^B0$. But firm "A" cannot have dual prices $(P_1{}^A0,\ P_2{}^A0)$ and firm "B" dual prices $(kP_1{}^B0,\ kP_2{}^B0)$ where $|1-k|$, and $k \neq 0$ represents some percentage difference in the prevailing price level faced by "A".

### 3.2.2. WHY FRONTIER PRODUCTION FUNCTIONS ARE NOT COMPATIBLE WITH DUALITY THEORY - FIRST PART:

This section will test the implication that two firms can be price efficient when only one of them is technically efficient. This means that both firms are on the price efficient ray, e.g., line $OA_0$ in Figure 3.2, but one of them is <u>not</u> on an isoquant (it is a true interior point), while the other one is on an isoquant. That is, one firm is technically efficient, and the other one is "technically inefficient," meaning it could produce more output with the same resources than it is <u>in fact</u> producing without changing the amounts of the inputs being consumed in production.

In this section a proof will be offered to demonstrate that frontier production functions are incompatible with duality theory and free disposal given the characteristic of frontier production functions that there is an interior to an isoquant at the same level of production. Apart from the fact that it has already been shown that production sets cannot be solid due to the laws of thermodynamics (see Chapter Two and Appendix One), it will be shown that solid production sets are contrary to duality and free disposal.

3.2.2.1. THE INTUITIVE ARGUMENT:

Consider Figure 3.3 which represents the presumed input requirement set and factor-price requirement set for a frontier production function. If one can produce the same output as one produces at $B_0$ using less of both inputs and the same process, i.e., produce at $A_0$ with absolutely no other changes, then why cannot one produce the same output as one produces at $A_0$ using less of both inputs? Differences in TE, as the term is used in the frontier production function theory, must involve something else in addition to merely a change in the level of $v_1$ and $v_2$. Since that something else is never defined or explicitly included, its cost is implicitly ignored, and therefore comparisons of "technical efficiency" are incomplete and invalid.

Consider the polar reciprocal sets in Figure 3.3 for two firms "A" and "B" which have the same technology and

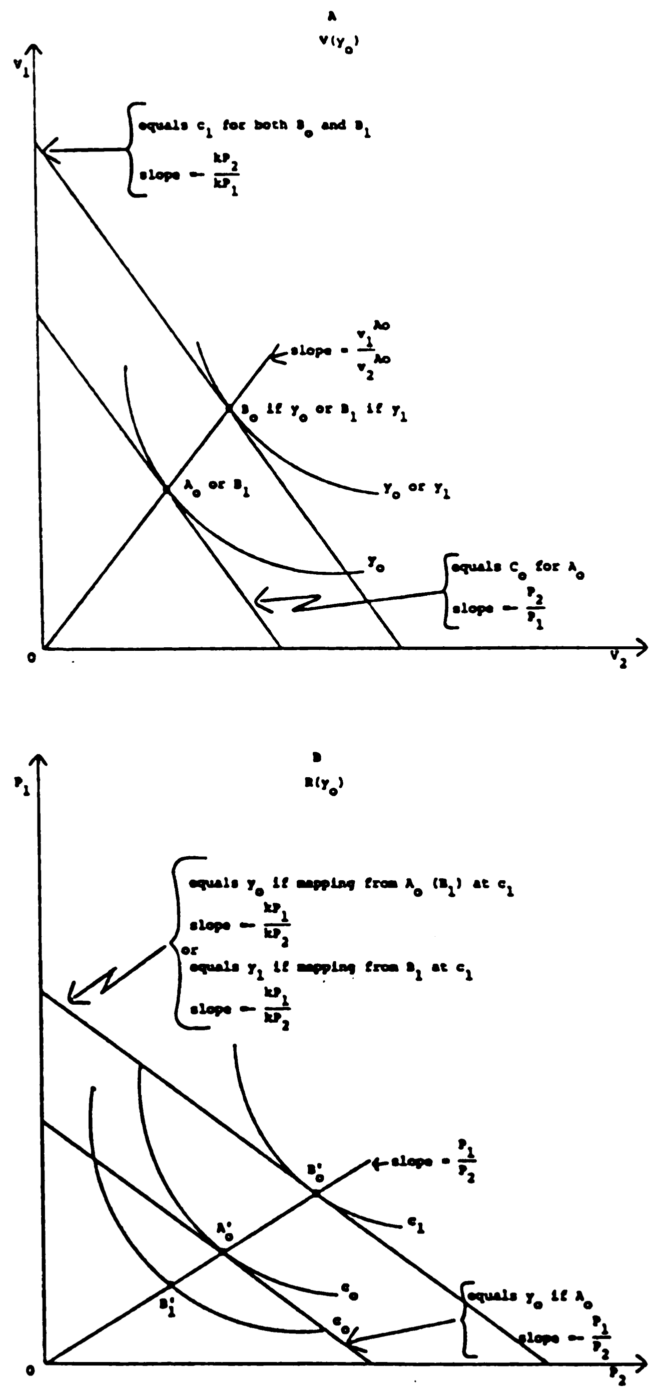



FIGURE 3.3

AN INPUT REQUIREMENT SET V(y) FROM A DISTANCE FUNCTION (A)
AND A FACTOR-PRICE REQUIREMENT SET R(y)
FROM A COST FUNCTION (B)

the same cost function conforming to the assumptions and principles of duality.

Assume for an initial situation that firm "A" is represented by point $A_0$ and firm "B" is represented by point $B_0$ in Figure 3.3A, that both firms produce exactly the same product $y_0$, and that both firms purchase their inputs in a perfectly competitive market at fixed prices ($P_1$, $P_2$). According to the theory of frontier production functions, this means that firm "A" is both technically efficient and price efficient, while firm "B" is price efficient but not technically efficient, as $B_0$ is a "true" interior point to the isoquant $y_0$. That is, firm "B" uses more of both inputs, $v_1$ and $v_2$, to produce the same amount of product as firm "A."

Therefore in Figure 3.3, using frontier production function theory it appears that $A_0$ and $B_0$ map to points $A'_0$ and $B'_0$ respectively. Indeed, if one calculates the total cost for firms "A" and "B" at the given market prices ($P_1$, $P_2$), then $C(B_0) = c_1$ is greater than $C(A_0) = c_0$, and those values would be located in cost space at $B'_0$ and $A'_0$ respectively.

3.2.2.2. PROOF USING DUALITY:

Notice that at $B'_0$, the dual prices for firm "B" are at a higher price level than the market prices and dual prices for "A," ($P_1$, $P_2$), but in the same ratio, i.e, they are higher by a constant, $|1-k|$, where $k > 1.0$. Now,

using duality, find the points in product space that map from $A'_0$ and $B'_0$; the points in product space for which $A'_0$ and $B'_0$ represent the least cost bundle as required by the definition of a cost function. Notice $B'_0$ will map to $B_1$ since $B'_0$'s dual prices are $P_1 = kP_1$ and $P_2 = kP_2$: that "B" has dual prices that are $|1-k|$ times greater than "A." But at $B_1$, $f(B_1) = y_1 = y_0$ only if $B_1$ is at $A_0$, since there cannot be two different least cost bundles at the same level of output. Otherwise, production at $B_1$ must equal $y_1$, where $y_1 > y_0$; that is, there are two least cost bundles, but at two different levels of output. Notice production at $B_1$ is not equal to production at $B_0$ so that $B_1$ is not lacated at $B_0$.

In neither case does $B'_0$ map to $B_0$. Therefore, our initial assumption is inconsistent with duality since firm "B" cannot be at both points $B_0$ and $A_0$, or points $B_0$ and $B_1$, at the same time. Thus, there cannot be any frontier production function "technically inefficient" points that conform to duality. Two firms cannot be price efficient when only one is technically efficient.

### 3.2.3.3. PROOF USING FREE DISPOSAL:

One might reason that since point $B_0$ is a "true" interior point, it does not map to a point $B'_0$ in cost space, but to $B'_1$. For instance, McFadden (1978) suggests that the 'interiors' to the isoquant of a distance functions map to exteriors of factor-price requirement sets, and vice versa. In so doing one keeps production at a constant level in product space and total cost at a constant level in cost space therefore maintaining duality. In order to map $B_0$ to $B'_1$, one must have dual prices for the greater quantities of the inputs used at $B_0$ that are lower than the market prices ($P_1$, $P_2$). lower by a factor $|1-k|$, where $k < 1.0$, which is a violation of the ceteris paribus conditions since this implies a change in the price level paid by $B_0$. Therefore, a correction or adjustment must be made to keep all comparisons between "A" and "B" on an equal footing. This means that given fixed market prices ($P_1$, $P_2$), one must freely dispose of the difference in total cost between "B" and "A," $c_1 - c_0$, in order to have points in cost space for which total costs are equal (as they are for $B'_1$ and $A'_0$).

When one uses free disposal in product space one freely disposes of both input <u>and</u> output (see Appendix Two). In cost space, total cost corresponds to the output of product space, while prices correspond to the inputs of product space. Therefore, when one freely disposes of the difference in total cost between "B" and "A," this free disposal is the same as disposing of the difference between the dual prices for $B_0$ and the market (also dual) prices for $A_0$. It is the same as moving from $B'_1$ to $A'_0$ in cost space by free disposal. By the same argument, free disposing of the total cost difference between "B" and "A" is the same as disposing of the difference between the dual prices for $B'_0$ and the dual prices for $A'_0$; that is, it is the same as moving from $B'_0$ to $A'_0$ in cost space by free disposal. But by duality, if one freely disposes in cost space, one must freely dispose in product space and vice versa. Therefore, moving from $B'_1$ to $A'_0$, or from $B'_0$ to $A'_0$, in cost space by free disposal means one is necessarily moving from $B_1$ to $A_0$ (where $B_1 \neq A_0$), or $B_0$ to $A_0$, in product space respectively at the same time, which eliminates any difference in "technical efficiency," as the term is used in the frontier production function theory, between firms "A" and "B."

The possibility that one might define or locate a point $B'_1$ in cost space, where cost is equal to $c_0$, raises two other issues. First, this may be the reason that some frontier production function authors maintain that improvements in "technical efficiency" are costless [Bagi and Huang, 1983], while others maintain that there is a cost to "technical inefficiency [Kopp, 1981a, and Schmidt and Lin, 1983]." Second, and more importantly, the existence of a point $B'_1$ at a total cost $c_0$ would seem to violate the assumption of strict convexity from below that is implicit in a cost function. Therefore, if cost functions cannot have <u>exteriors</u> at a constant total cost, in order to maintain strict convexity from below, distance functions cannot have <u>interiors</u> at a constant production. Thus, the implication that two firms can be price efficient when only one firm is technically efficient is refuted. Any firm that is price efficient is always technically efficient.

### 3.2.3. WHY FRONTIER PRODUCTION FUNCTIONS ARE NOT COMPATIBLE WITH DUALITY THEORY - SECOND PART:

Section 3.2.2. refuted the implication of frontier production functions that TE and PE are mutually exclusive and that two firms can be price efficient when only one of them is technically efficient. This section will consider the other implication that two firms can be technically efficient when only one of them is price efficient since the fact that PE implies TE does not necessarily mean that TE implies PE.

Recall the frontier production function definition of TE, that one produces the greatest amount of output given a bundle of inputs and a specific process. This definition leaves open the possibility that one can produce something less than the greatest amount of output and still consume all the inputs. This would mean that production functions have true interiors at the same level of production. Section 3.2.2. eliminated using this possibility to distinguish between PE and TE. This is not surprising since the theory of thermodynamics clearly demonstrates the same thing, where thermal efficiency is considered to be an example of technical efficiency (see Appendix One). In thermodynamics, technical efficiency, or thermal efficiency, is determined only within the context of comparing value of input (heat) to value of output (work), a point made in economics by Knight (1933) and Boulding (1981).

Therefore, the term "technical efficiency (TE)," as it is used in the frontier production function theory is ambiguous, confusing, and misleading. The term efficiency implies a comparison. In order to make a comparison, one must first have a basis for making the comparison; one must in effect have a numeraire. The frontier production function theory definition of TE, that one produces the greatest amount of output given a bundle of inputs and a specific process, is ambiguous, confusing, and misleading since for any point on a production function there is only <u>one</u> level of output possible. That is, any point on a production function not only represents the greatest amount of output given that bundle of inputs, it also represents the smallest amount of output given that same bundle of inputs, if <u>all</u> the inputs are consumed (which is true by definition). Therefore, to define TE as being on any point on an isoquant is ambiguous, confusing, and misleading since by itself there is no basis for comparison; there is no counter point that represents any other degree of "technical efficiency" given the same set of inputs and the same process. Production means creating utility by changing the time, form, or location of inputs. Therefore, efficieny involves implicit comparisons of value [Knight, 1933; Boulding, 1981]. Value must be measurable for there to be a basis for comparison between two points of production: one needs a numeraire.

Physical measures alone divorced from their specific value (i.e., from a numeraire) are an inadequate for comparing efficiency. Marginal physical products of inputs, which are determinate once a production function is defined, have value, as does the output they represent. They are technical coefficients of production. But as physical quantities their specific values are not indicated and, as such, they are an inadequate basis for making efficieny comparisons.

Consider the case where one is faced with a set of measurements of marginal physical products for various sets of the same types of inputs used in the same production process. If TE was determined by the set of coefficients of the largest magnitudes (hence the largest value), then TE would imply selecting those input bundles where average physical products are their largest, which represents the unit elasticity point on the production function, or the boundary between Stages I and II (for a further discussion of this case see Chapter Four). If price efficient points are always technically efficient, as Section 3.2.2. showed, then the set of techniclaly efficient points must include points that are at points of unit elasticity on the production function. Even if one assumes that all the unit elasticity points will lie on one isoquant, which is not generally true, one might still ask if any one of the unit elasticity points is more "technically efficient" than the others. This is the same as asking if any one point on an isoquant can be more "technically efficient" than the others.

Consider two sets of inputs, $v_1{}^0$ and $v_2{}^0$ that produce a given level of output, $y_0$. The marginal physical products are determinate with $MPP(v_1{}^0) = X^0{}_1$ and $MPP(v_2{}^0) = X^0{}_2$. If one changes the amounts of the inputs used, e.g. to $v_1{}^1$ and $v_2{}^1$, then the marginal physical products will no longer be $X^0{}_1$ and $X^0{}_2$, but $MPP(v_1{}^1) = X^1{}_1$ and $MPP(v_2{}^1) = X^1{}_2$, where $X^0{}_1 \neq X^1{}_1$ and $X^0{}_2 \neq X^1{}_2$. If $X^0{}_1 > X^1{}_1$ and $X^0{}_2 > X^1{}_2$, then $f(v_1{}^0, v_2{}^0) = y_1$ where $y_1 > y_0$. Therefore, in general, if $X^0{}_1 > X^1{}_1$, then $X^0{}_2 < X^1{}_2$, for $f(v_1{}^0, v_2{}^0) = f(v_1{}^1, v_2{}^1) = y_0$. In this case, notice that the inherent values of $v_1{}^1 \neq v_1{}^0$, and $v_2{}^1 \neq v_2{}^0$ so that one might ask if one set of inputs $(v_1{}^0, v_2{}^0)$ or $(v_1{}^1, v_2{}^1)$ is "technically more efficient" in producing $y_0$. To answer this question, one needs to have a basis of comparison, a basis for valuation; i.e., one needs a numeraire or a set of relative prices.

In economics, the utility function provides a basis for measuring value, by allowing one to derive the demand for all commodities based on the value (utility) they each provide. When one selects a numeraire, one commodity, the utility function serves as a basis for evaluating the relative value of all other commodities (including money) by comparing them to the numeraire; specifically by establishing how much of a specific quantity of a specific commodity will be equal in value (utility) to a specific quantity of the numeraire. A change in the underlying utility function will change the relative measures of value.

In economics, relative prices represent specific measures of value because they are determined by normalizing commodities on a numeraire. Clearly, associated with any specific production function are an infinite number of sets of relative prices, each set determining a unique total cost function. The relationship between the relative prices for any two points on the production function remains the same, in direction if not in magnitude, regardless of which set of relative prices is selected to measure the value of the physical quantities if they are based on the same preselected utility function. In general, changes in relative prices from one set of relative prices to another is accomplished either by vector multiplication (changing the utility function) of all prices and/or a change in the choice of numeraire which determines the measurable degree of difference between relative prices. A change in the utility function causing a change in relative prices is accomplished by vector multiplication of prices, where the elements of the vector are not equal. Any comparisons using the resulting two sets of prices is fundamentally a comparison of the change in utility. A change in numeraire, holding utility constant, will change relative prices, but not the direction of change between those relative prices, only the magnitude; i.e., the ranking of preference for the individual commodities will be unchanged.

Therefore, while the prices for the physical quantities represented by a point on a production function can change, any set of prices is as valid a representation of the instrinsic value of those physical quantities as any other set of prices, if they are derived from the same utility function; i.e., one holds the utility function constant. If one holds the utility function constant, then any change in prices will be the result of either scalar multiplication, multiplying all prices by a constant, which leaves relative prices unchanged, or a change in the numeraire, which leaves the fundamental values of the commodities unchanged. This is why economists prefer to make comparisons based on sets of "real" prices, sets of prices using the same numeraire that are not different by a scalar, and why in a duality mapping a valid comparison of points must hold relative prices at the same level of magnitude, or price level, as was pointed out in Section 3.2.1.

In this thesis, all the prices are designated as algebraic unknowns or variables, which means they can be any set of <u>real</u> prices from the infinite number of sets of prices with no change in the arguments presented, assuming there is only one underlying utility function upon which all comparisons are based. This is a necessary ceteris paribus condition for comparisons to be on an equal footing. If two points on a production function are designated as <u>equally</u> technically efficient at the same level of output, as the term is used in the frontier production function theory, and the two points represent two different levels of utility, then since production is by definition a change in the utility of inputs through a change in time, form, or space, the two points are in fact on two different value productivity functions since they represent two different levels of utility (or output). Specifically, this means that any two points on an isoquant that are designated as equally technically efficient must reflect the same level of utility on the same utility function; i.e., two or more points on an isoquant can be equally technically efficient only if everything is held to be the same. That is, once one has normalized all prices by selecting a utility function and a numeraire, there is only one set of relative prices associated with the value of, or physical quantity of, any given bundle of inputs and output.

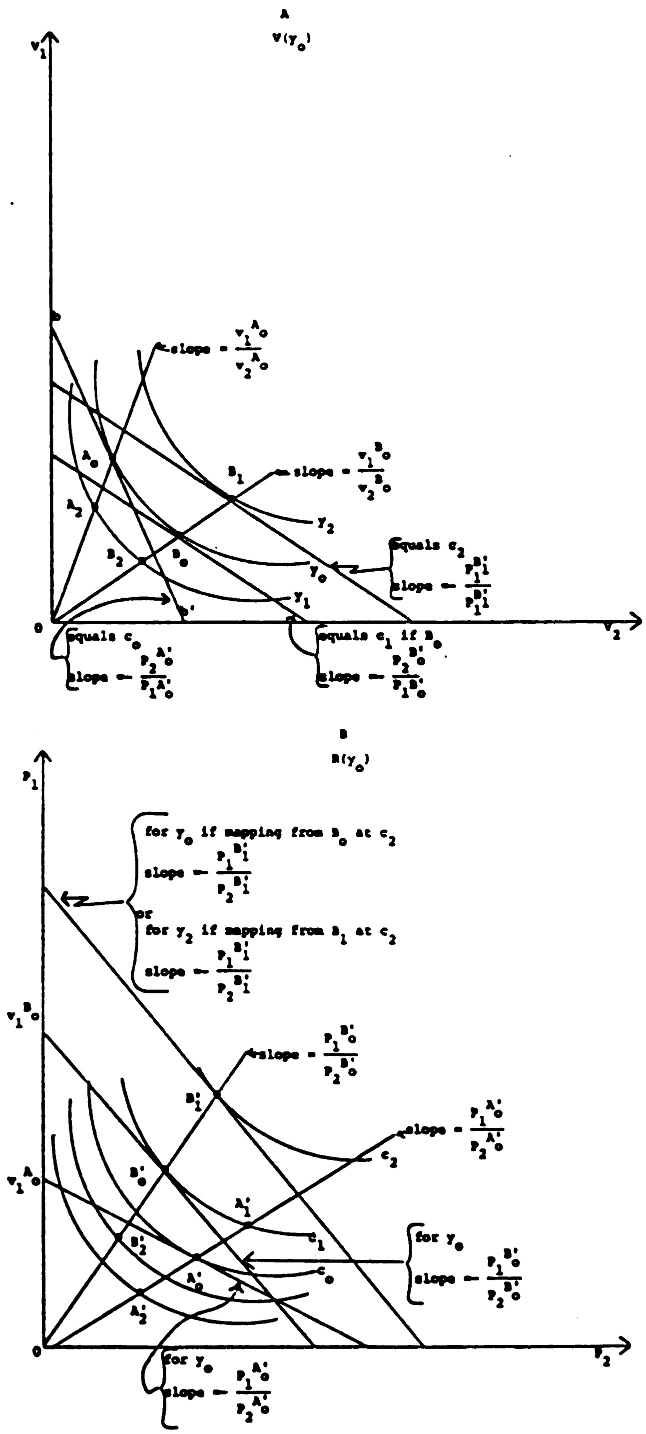
Therefore, "TE," as the term is used in frontier production function theory, is either associated with specific relative prices by virtue of being associated with a given utility function and a numeraire, or it is meaningless since there is no basis for the comparison of efficiency of "TE" points. Specifically, if "TE" is not meaningless, then any point that is techniclly efficient is also price efficient. For the purposes of testing the implication that two points can be technically efficient when only one of them is price efficient, it will be assumed that the term TE is meaningfully distinct from PE.

### 3.2.3.1. THE INTUITIVE ARGUMENT:

Consider the implication of frontier production functions that TE and PE are mutually exclusive and that two firms can be equally technically efficient at the same level of output (utility) when only one of them is price efficient. If two firms are equally technically efficient, it means that the two firms are on the same isoquant and can produce no more than that output with the (value of, or quantity of) the resources at their disposal when input (and output) prices are fixed. If two firms "A" and "B" are producing the same output $Y_0$, and they are purchasing their inputs in perfectly competitive markets, so that only one of the two firms, assume it is "A", is price efficient, then the firm that is not price efficient, "B," is not producing as much output (utility) as it could given the (value of, or quantity of) the resources at its disposal. Therefore, "B" is not technically efficient. If "B" is not price efficient, then its cost is not as low as it could be if it were a least cost producer; its total cost is higher than it would be if it were a least cost producer. Therefore, in the perfectly competitive markets, "B" can costlessly rearrange its input bundle so that the inputs are used in the same ratio as they would be were "B" a least cost producer, i.e., in the same ratio as the input bundle used by "A," and at its same (original) cost produce more output that "A."

3.2.3.3. PROOF BY DUALITY:

Consider Figure 3.4, which is essentially like Figure 3.2. Frontier production function theory suggests that firms "A" and "B" are initially at points $A_0$ and $B_0$ respectively, while prices are fixed at $(P_1^{A'}0,\ P_2^{A'}0)$ for both firms. That is, both "A" and "B" are equally technically efficient for output $y_0$, but only "A" is price efficient. Recall that TE means producing the greatest amount of output (utility) possible from a given set of inputs in a given process. It was pointed out above that this definition is meaningless unless one recognizes that with every set of physical quantities there is associated a specific set of relative prices (subject to change only by vector multiplication or a change in numeraire). For the frontier production function definition of TE to be rhetorically and internally consistent, it must not change whether prices are considered or not; the definition of TE must be invariant with respect to prices, or a lack of prices, if there is to be a meaningful separation of TE from PE. But this is not true. Including the PE conditions, concurrent with the TE conditions, for the puposes of making comparisons between firms means relative dual prices are established. Frontier production functions cannot be used to consider TE without concurrently considering prices as a measurement of the value of the technically efficient bundle.



FIGURES 3.4

AN INPUT REQUIREMENT SET V(y) FROM A DISTANCE FUNCTION (A)
AND A FACTOR-PRICE REQUIREMENT SET R(y)
FROM A COST FUNCTION (B)

But <u>fixing</u> prices (which fixes the efficient price ratio) by assuming that both "A" and "B" buy their inputs in the same competitive market, and that the prices they face are the prices that make $A_0$ price efficient, $(P_1{}^{A'}0, P_2{}^{A'}0)$, means PE is determined, but that $B_0$ does not map to $B'_0$. For $B_0$ to map to $B'_0$, firm "B" would have to have dual prices equal to $(P_1{}^{B'}0, P_2{}^{B'}0)$, at a total cost of $c_1$ when it is paying market prices equal to $(P_1{}^{A'}0, P_2{}^{A'}0)$.

Therefore, calculate $C(B_0)$, the cost at "B," at prices $(P_1{}^{A'}0, P_2{}^{A'}0)$ using the cost function shared by "A" and "B," and find the location for the resulting value in the factor-price requirement set.

Notice that $C(B_0) \neq c_1$ at prices $(P_1{}^{A'}0, P_2{}^{A'}0)$. If $C(B_0) = c_1$, then $B'_0$ represents a least cost producer at both market prices $(P_1{}^{A'}0, P_2{}^{A'}0)$ and at market prices $(P_1{}^{B'}0, P_2{}^{B'}0)$ at a total cost $c_1$, using inputs $v_1{}^{B}0$ and $v_2{}^{B}0$ to produce $y_0$. But this would violate the condition that if one holds output constant and changes prices, then one moves from the same isoquant in product space to different isocosts in cost space since $P_1{}^{A'}0 \neq P_1{}^{B'}0$ and $P_2{}^{A'}0 \neq P_1{}^{B'}0$.

Notice, too, that $C(B_0) \nless c_1$ at prices $(P_1{}^{A'}0, P_2{}^{A'}0)$. If $C(B_0) < c_1$, then $P_1{}^{B'}0$ and $P_2{}^{B'}0$ are not the least cost prices at $B_0$, rather $(P_1{}^{A'}0, P_2{}^{A'}0)$ are.

The location of $C(B_0)$ must be on a ray from the origin in the factor-price requirement set so that the ratio of the inputs in the input bundle purchased by firm "B," $(v_1{}^{B}0/v_2{}^{B}0)$, is not changed. Therefore, $C(B_0)$ is located not at $B'_0$, but at $B'_1$ and at a total cost $c_2 > c_1$. Given $B_0$ at fixed prices ($P_1^{A'}0$, $P_2^{A'}0$) using the cost function it shares with "A," firm "B" is located at $B'_1$ in the <u>shared</u> factor-price requirement set. That is, if firms "A" and "B" share the same production and cost function, then at fixed market prices the locations for those firms in cost space are at $A'_0$ and $B'_1$ respectively.

The prices ($P_1^{B'}1$, $P_2^{B'}1$) represent not the market prices which were used to calculate the value $C(B_0)$, but the derived prices that make the input bundle represented by $B'_1$ a least cost bundle, and the price ratio that would make the firm "B" PE, since all points on the cost function are least cost bundles by definition of a cost function. The bundle purchased by "B" is in fact not price efficient, since the ratio of the dual prices, ($P_1^{B'}1/P_2^{B'}1$), is not equal to the ratio of the dual prices, ($P_1^{A'}0/P_2^{A'}0$) for the price efficient producer, "A." Notice that at $B'_1$ the dual prices ($P_1^{B'}1$, $P_2^{B'}1$) are absolutely greater than the prices at $B'_0$, ($P_1^{B'}0$, $P_2^{B'}0$). That is, $P_1^{B'}1 = kP_1^{B'}0$ and $P_2^{B'}1 = kP_2^{B'}0$, where $k > 1.0$. At the fixed prices $P_1^{A'}0$ and $P_2^{A'}0$, "B" enters the duality mapping at $B'_1$. This implies that "B" is not only generating dual prices that are inefficient, $P_1^{B'}1/P_2^{B'}1 \neq P_1^{A'}0/P_2^{A'}0$, but also generating dual prices that are higher than the <u>level</u> of prices paid by "B." That is, the magnitude of the dual prices paid by "B" are greater than the magnitude of the dual prices paid by "A," by $|1-k|$.

Because of this change in the magnitude of the dual prices generated by "B" at $B'_1$, one must make an adjustment to eliminate this implied difference in the magnitude of the dual prices generated by "A" and "B" in order to maintain ceteris paribus conditions in the comparison. Now, having located $B'_1$ in cost space, under the assumption that both firms face the same fixed market prices, where does $B'_1$ map back to in its corresponding shared technical (product) space by duality? That is, what point in the technical (product) space would map to $B'_1$ in cost space where the dual prices $P_1^{B'1}$ and $P_2^{B'1}$ would be at the same price level as the dual prices generated by "A." The point is either $B_1$ or $B_2$, not to $B_0$. Therefore, the bundle purchased by firm "B" no longer represents the same technically efficient level of output (utility) "A" does. By duality, points in product (cost) space have a unique one-to-one mapping with points in cost (product) space.

If one wishes to maintain "B"'s greater expenditure capacity, "B"'s larger budget, or total cost, $c_2$, then "B" must produce a greater output $y_2 > y_0$, to provide "B" with the greater compensating income to cover its greater cost $c_2$. This means "B" would be represented by $B_1$ in the duality mapping.

If one wishes to maintain the same budget constraints for both "A" and "B" at $c_0$ and $c_1$, thereby maintaining the original total cost relationship between the two firms, then "B" must purchase fewer inputs and produce a lower output $y_1$, where $y_1 < y_0$ at $c_1$, and therefore be represented by $B_2$ in the duality mapping.

In either case, within the duality mapping "B" cannot be technically efficient at a production level $y_0$ while paying market prices $(P_1{}^{A'}0, P_2{}^{A'}0)$. Therefore, there appears to be a contradiction between the initial situation, two firms "A" (represented by $A_0$) and "B" (represented by $B_0$) both technically efficient at the same level of ouput (on the same isoquant), and the final situation, two firms "A" (represented by $A_0$) and "B" (represented by $B_1$, or $B_2$), on two different isoquants. The firm "B" cannot be at $B_0$ and $B_1$, or $B_2$, (on two isoquants) at the same time. The conclusion is that being off the isocost line, not being price efficient, means by duality that the firm is also off the isoquant $y_0$ and not technically efficient at the same level of output $y_0$. Therefore, the only firm that is technically efficient on the $y_0$ isoquant is "A," because it is the only firm that is price efficient. Firm "B" could produce more output with the budget (resources) that allows it to produce $y_0$ (when its not price efficient), or it could produce less output if it can spend no more than $c_1$. Therefore, TE $\equiv$ PE always, by duality. This should not be a surprise to persons familiar with Knight's 1933 and Boulding's 1981 arguments.

Buying $B_0$ inputs at market prices $(P_1{}^{A'}0, P_2{}^{A'}0)$ is identical to buying $B_1$, or $B_2$, inputs at dual prices

$(P_1^{B'}1,\ P_2^{B'}1)$. The duality exists only between $B'_1$ and $B_1$, or $B_2$, for firm "B" at the fixed market prices $(P_1^{A'}0,\ P_2^{A'}0)$. That is, there is no duality mapping on the cost function for $B_0$ at prices $(P_1^{A'}0,\ P_2^{A'}0)$ and an output level $y_0$, since $B_0$ is not a least cost bundle for $y_0$ and prices $(P_1^{A'}0,\ P_2^{A'}0)$. Therefore, in measuring efficiency for firms for which the duality principle is assumed, the relevant points for firm "B" are $B_1$, or $B_2$, and $B'_1$, which means that at the assumed market prices "B" is neither technically efficient nor price efficient at output level $y_0$; only "A" is either, and it is both, technically efficient and price efficient at output level $y_0$.

### 3.4. SUMMARY

Distance functions are based on production functions which are input-conventional, meaning they conform to regularity conditions, free disposal (monotonicity), and strict convexity from below. Using duality, one can find corresponding cost functions which must also display free disposal and strict convexity from below (quasi-concavity). Distance functions are valid representations of reality (excluding Stage III of production). It might easily be mistakenly inferred that frontier production functions are also valid representations of reality because they look the same as distance functions and bear a superfical appearance to distance functions when graphs of the two are compared (see Figure 3.1).

Se

distan

the S

origi

that

isog

duct

sol

rep

The

in

fr

P

Section 3.1 tested, and rejected, the hypothesis that distance functions and frontier production functions are the same, since they arise from different theoretical origins. Distance functions arise from a production set that has no true interior points of production to its isoquants at the same level of production. Frontier production functions arise from a production set that is a solid set with interior points to its isoquants that represent the same level of output as the isoquant. Therefore, distance functions can have no "technically inefficient" points of production, as the term is used in frontier production function theory, within its corresponding production set.

Frontier production functions have the fundamental property that TE and PE are mutually exclusive conditions of reality; that there are in reality these two different types of efficiency, rather than only one type of efficiency, simply efficiency. If these two kinds of efficiency validly reflects reality, they should conform to the principles of free disposal and duality between product and cost space. The hypothesis that TE and PE are not the same was tested in Section 3.2 in two parts by testing separately the two implications of the hypothesis. The first implication is that two firms can be price efficient when only one of them is technically efficient. It was shown in Section 3.2.2. that this first implication is contradicted by free disposal and duality and therefore that a firm that is price efficient is necessarily technically efficient at the same time. The second implication is that two firms can be equally technically efficient at the same level of output when only one of them is price efficient. Section 3.2.3. demonstrated that this implication is contradicted by duality, and therefore that any firm that is technically efficient is necessarily price efficient at the same time. Therefore, TE and PE are always identically equal and the two adjectives "technical" and "price" add nothing to the discussion.

The consequences of testing these two hypotheses are twofold. Frontier production functions are not the same as distance functions since they arise from a different theoretical origin than distance functions, that of a solid production set including both "technically efficient" and "technically inefficient" points of production. Therefore, in production theory, only distance functions should be used as the basis for theoretical explanations of reality (excluding Stage III of production), or for any empirical analysis that excludes Stage III of production, since distance functions alone reflect the physical reality that production sets are hyperplanes (surfaces) in input space rather than solids (see Appendices One and Two).

The consequences of testing the second hypothesis, that TE and PE are not the same, are that the terms "TE" and "PE" as they are used in the frontier production function theory are ambiguous, misleading, and confusing, since they

There is o

is by def

the util

ly meas

determi

produc

parin

one

des

cia

m

since they imply that there is two kinds of efficiency. There is only one -- simply efficiency. Since production is by definition a change in the time, form, or space of the utility of the inputs, a production function implicitly measures utility or value. Therefore, in order to determine the degree of efficiency for any point on a production function, one must first have a basis of comparing the intrinsic value represented by a point; i.e., one must first choose a numeraire. Once a numeraire is designated, all points on a production function are associated with a set of relative prices that serve as a measure of their value or utility. A change in those relative prices can be accomplished only by vector multiplication, which means that one has changed production (utility) functions, or by a change in the numeraire, which will not change the direction of differences between the original relative prices, only the magnitudes of those differences. Therefore, "TE" as it is used in the frontier production function theory must be implicitly associated with specific relative values, or prices, or it implies a comparison of points without a basis for comparison. In this case, the adjective "technical" loses its meaning as the distinction between "TE" and "PE" is without foundation. If technically efficient points in frontier production function theory are associated with specific relative values, or prices, which would be implicitly logical since they stands as counter points to points that are price efficient, points which are explicitly as

are tec

vice ve

techni

versa

not p

vers

adj

le

d

v

citly associated with values, or prices, then points that are technically efficient must also be price efficient and vice versa. TE always means PE by duality; a firm that is technically efficient is always price efficient and vice versa. The converse is also always true; a firm that is not price efficient is not technically efficient and vice versa. Consequently, in discussions of efficiency, the adjectives "technical" and "price" are ambiguous, misleading, and confusing, and might best be dropped from the discussions. Efficient production is always determined by values as Knight (1933) pointed out over fifty years ago.

f

T

bot

repre

ferent

subjec

$z_3$. T

functior

duction,

This mea

and $v_2$ P

the bour

# CHAPTER FOUR

## A NEW CASE

There is one other case worth considering. It arises as a result of calculating a unit isoquant. This chapter will show how a point on the production surface becomes an 'interior' point to the unit isoquant for another point of lower production on the production surface due to the averaging process itself. This is a new case since it has not previously been dealt with in the frontier production function literature. However, unlike a frontier production function, the definition of technical efficiency is fundamentally different, there is no distinction between TE and PE, and the unit isoquant so described may have both 'interior' and 'exterior' points.

Consider points Q and P in Figure 4.1. Both points represent production of the same output, y, using different quantities of the same variable inputs, $v_1$ and $v_2$, subject to the same quantity of an identical fixed input, $z_3$. Therefore both points are on the same sub-production function (surface). P represents a higher level of production, but is a member of Q's input requirement set. This means that both P and Q can be projected into the $v_1$ and $v_2$ plane. In the $(v_1,v_2)$ plane the isoquant for Q is the boundary to Q's input requirement set, within which P

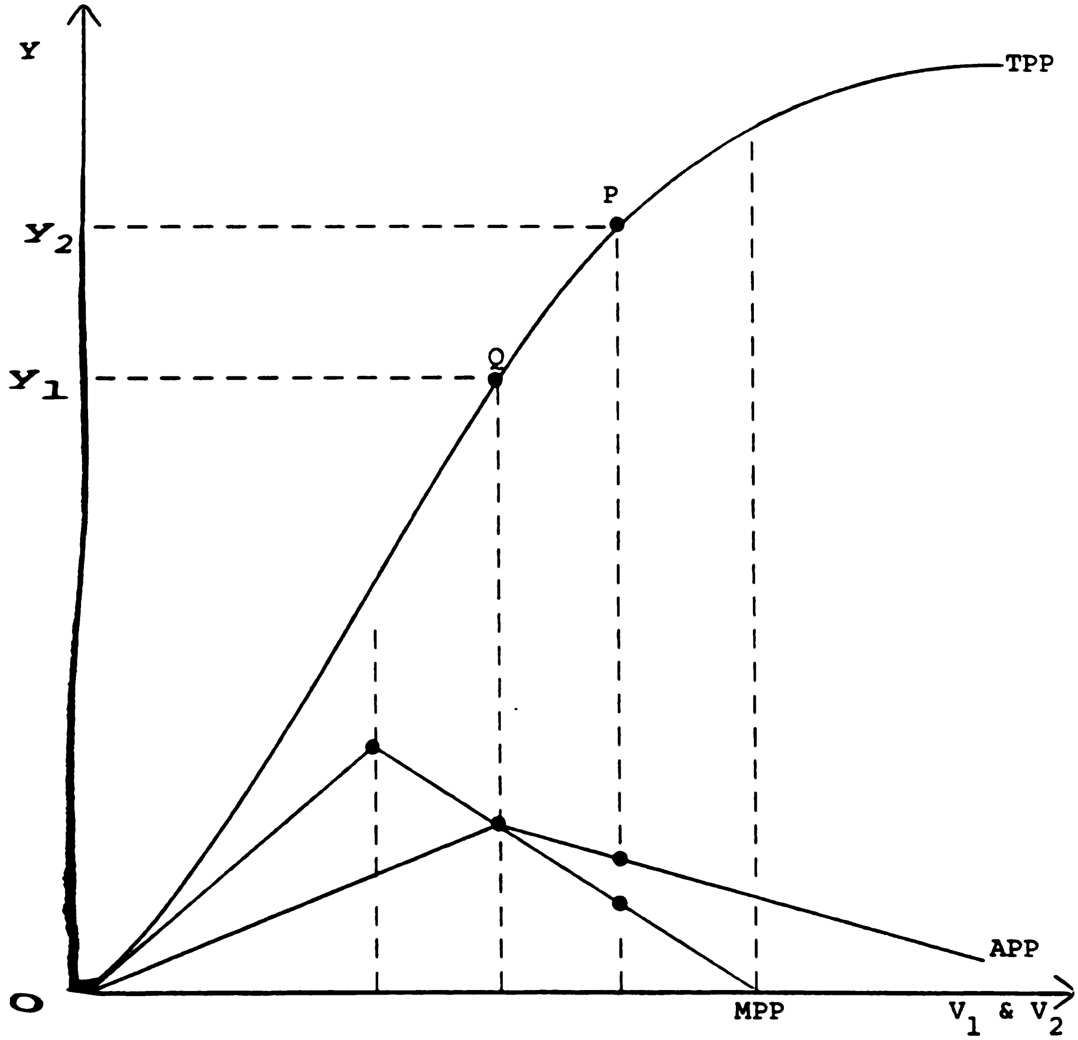



FIGURE 4.1

A PRODUCTION FUNCTION (SURFACE) SHOWING TOTAL PHYSICAL PRODUCT (TPP), AVERAGE PHYSICAL PRODUCT (APP), AND MARGINAL PHYSICAL PRODUCT (MPP) OF $v_1$ AND $v_2$ IN LEAST COST COMBINATION GIVEN $z_3$ EQUAL TO A CONSTANT will lie (see Figure 3.1A in Chapter Three). Notice that Q represents the point of production for which APP (average physical product) is at a maximum, and therefore represents the boundary between Stages I and II of production.

Now construct the unit isoquant through point Q for the set of production points from which these two observations are drawn (see Figure 4.2). Point Q maps to Q' and point P maps to P'. Point P' is an 'interior' point to the unit isoquant through Q'. Assuming that both Q and P use inputs in their least cost combination, both points Q' and P' are "price efficient," but point P' appears to be "technically inefficient" as compared to Q'. Notice that at Q one obtains maximum output per unit input, i.e., maximum APP. At P one obtains less output per unit input, i.e, lower APP. Therefore, Q is on the unit isoquant where one is using minimum input per unit output, and P is in the 'interior' where one uses more input per unit output. By definition of the unit isoquant, P' represents less output per unit input than point Q' because point P represents a lower APP than point Q. This means that the discrepancies between Q' and P' in Figure 4.2 come about as a direct consequence of the averaging process itself. The difference between Q' and P' are not due to any intrinsic "technical" or physical inefficiency originating at a true interior point to Q as would be construed from frontier production function theory. In constructing the unit isoquant one in effect shifts one's attention from

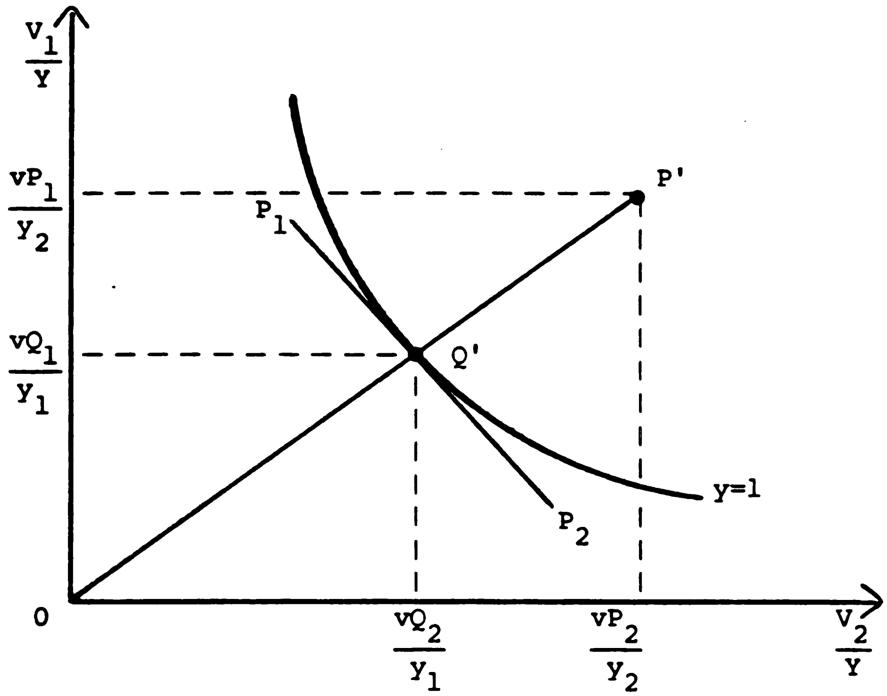



FIGURE 4.2

A UNIT ISOQUANT

points on the TPP (total physical product) curve to corresponding points on the APP curve (see Figure 4.1). That is, clearly there is less output per unit input, i.e., more input per unit output, at P (P') than at Q (Q') due to the law of diminishing returns. At P (P') the MPP (marginal physical product) is lower than it is at Q (Q').

### 4.1. IS THIS A NEW DEFINITION OF TE ?

This new definition of TE does not depend on interior points to a production function. It does not mean one can get more output with the same inputs at P', as would be concluded from frontier production function theory. Therefore, to conclude that Q (Q') is "technically efficient" would be misleading. The point P' (P) is only less "technically efficient" than Q (Q') in the sense that one can achieve a higher APP using less input than is used at point P (P') by moving back to point Q (Q'). If one were to define this difference in average physical output potential between points P (P') and Q (Q') as a difference in "technical efficiency" between the two points, one would be creating a definition for TE that is fundamentally different than the "TE" that is defined in the frontier production function literature. Recall that in the frontier production literature a point is TE if, and only if, one can not get more output (TPP) from the same inputs. Clearly, at point P (P') one cannot get more output from the same inputs; one can only get a higher APP from the same variable inputs by leaving some variable inputs unused. This difference in potential APP is clearly understood in traditional microeconomic theory. Equally clear from traditional microeconomic theory is the inappropriateness of recommending that one always maximize APP; that one maximize TE where TE means maximum average physical product. In traditional microeconomic theory the relevant issue is whether or not a higher APP is worth producing. This is a case of properly treating economic adjustment with respect to variable inputs.

## 4.2. TE NOT SEPARATE FROM PE

This case still does not separate TE and PE. A point P (P') would be off the unit isoquant for Q (Q') due to: (1) either P (P') not properly equating its MRP (marginal revenue product) to its MFC (marginal factor cost), or (2) to P (P') paying lower prices for inputs $v_1$ and $v_2$ (having a higher budget constraint). P' cannot be allocatively efficient, a least cost producer, if Q' is, and vice versa, when both firms have the same budget constraint and pay the same prices and opportunity costs for all their inputs, both variable and fixed. This is clear due to the same arguments that were advanced in Chapter Three regarding distance functions and frontier production functions.

If P' (P) and Q (Q') are considering different within firm opportunity costs for the identical quantity of fixed input $z_3$, then both might be least cost producers and be producing different levels of output. Recall from traditional microeconomic theory that inputs become fixed in production when their within firm opportunity cost (shadow price) is between their out of firm opportunity costs (acquisition and salvage price) [Johnson and Quance, 1972]. The different firms represented by P' (P) and Q (Q') might maximize the within firm opportunity cost for the fixed input $z_3$, the economic returns to $z_3$, at different values between the acquisition price and salvage price that both firms face. If P' (P) allocates less of its budget to compensate for the use of $z_3$, pays less rent to $z_3$ than Q (Q') does, then P' (P) can "purchase" more of the two variable inputs $v_1$ and $v_2$ than Q (Q') can. This problem can be avoided if both firms endogenize the quantity allocation of $z_3$ using the method outlined by Edwards (1958); i.e., both firms will be evaluating $z_3$ at the same within firm opportunity cost.

### 4.3. 'INTERIOR' AND 'EXTERIOR' POINTS

Unless the unit isoquant drawn through Q (Q') is the boundary between Stage I and Stage II due to the firm being in long run equilibrium within a perfectly competitive industry, there are both 'interior' and <u>'exterior'</u> points. Notice that if point P in Figure 4.1 were to represent the least cost producer, and one were therefore to draw a unit isoquant through P (P'), then Q (Q') would become an <u>'exterior'</u> point to the unit isoquant through P

(P'). That is, one could still increase "TE", increase physical output per unit input, by operating at point Q (Q'). In this case $z_3$ for Q (Q') would be earning economic rent; $z_3$ for Q (Q') would have a higher within firm opportunity cost (shadow price) than for P (P').

## 4.4. SUMMARY

This chapter has considered a valid case where one can construct unit isoquants that have 'interior' points. It is clear that the interpretation of these unit isoquants is not the same as the interpretation of a frontier production function. The 'interior' points to the unit isoquant represent higher levels of total production than points on the unit isoquant, rather than equal levels of production. An 'interior' point becomes an 'interior' point due to the fact that at higher levels of production the law of diminishing physical returns results in lower total output per unit input which means one uses more input per unit output. In addition, unless the unit isoquant represents the boundary between Stages I and II of production, where average physical production is at a maximum, a production set for firms that are in long run equilibrium in a perfectly competitive industry, the unit isoquant will have both 'interior' and 'exterior' points. The term "technical efficiency" is really a misnomer in this case, since it has nothing to do with getting maximum output from a given set of inputs. In this case, technical efficiency implies only that one is producing where average physical product is at a maximum. Clearly, one chooses to be technically efficient or maximize APP only if it is allocatively efficient.

# CHAPTER FIVE

## THE FRONTIER PRODUCTION FUNCTION LITERATURE

While the popularity of frontier production functions is relatively recent, the arguments for their existence have been around for some time. The erroneous distinction between TE and PE has flourished largely unchallenged. The distinction between TE and PE seems to have been simply postulated and accepted without any serious attention to its logic or its conformity with the laws of nature. How did such misguided logic establish itself in economic theory? What are the original sources of error? If the logic of frontier production functions is inconsistent with legitimate economic theory, the frontier production function literature should have discredited itself.

This chapter will examine some of the more salient literature relevant to frontier production functions. An exhaustive review of all the literature is unnecessary since it is highly repetitious. The chapter will attempt to explain briefly how the errors established themselves and evolved despite the internal inconsistences revealed in the foregoing chapters. It will be shown that these inconsistences substantiate the arguments offered in Chapter Two; that frontier production functions arise due to specification error and/or aggregation error.

## 5.1. LINEAR PROGRAMMING AND FREE DISPOSAL

### 5.1.1. KOOPMANS:

In tracing the development of technical and price efficiency, free disposal, and frontier production functions, there does not appear to be any literature prior to the development of linear programming, or activity analyisis.

One of the earliest references to technical efficiency may be found in Koopmans (1951). The article is a discussion of how one finds the efficient set amongst the feasible set. Koopmans creates ambiguity by saying that he will be discussing the "efficient point set in the commodity set." It is clear that he means isoquant in using the term feasible set rather than the efficient set (a maximum profit point, or an expansion path), since he is referring to a production function. His postulate that the isoquant is "the efficient set" is never substantiated; indeed what an "efficient set" means is not discussed. If an isoquant is itself efficient, then the tangency of the isoquant and the budget constraint is a different type of efficiency and one can infer that there is technical and price efficiency. If the isoquant is a technically efficient set, then one can infer that a technically inefficient set must also exist, i.e., the interior to the isoquant. None of this is explored in the article. Koopmans' use of "efficient" is an unfortunate word choice that is never explained. The same is true of his use of the term free disposal.

The concept of free disposal is not developed, but alluded to within the context of developing the conditions and characteristics of activity analysis. It is evident from the introductory remarks that free disposal depends on the perfect complementarity assumed for each activity being analyzed.

> ... and situations where some factors can only be combined, within the technological principle involved, in fixed ratios to each other (limitational factors). The second type of situation can only be reconciled with the notion of a production function defined in the whole factor space by allowing the production manager to discard parts of the factor quantities specified as <u>being available.</u> The corresponding production functions have kinks at the points where the ratios of available factor quantities coincide with the technical ratios specific to the process in question. [Koopmans, 1951, page 33-34; underlining added]

In linear programming this type of free disposal is represented by slack and surplus activities, which serve as a place "freely" to "dispose," or "discard," "available" but <u>unused</u> amounts of resources in order to establish the equality constraints of the algorithm. Seldom are the quantities of inputs in the slack and surplus activities interpreted as having been consumed in the production process: they are not inputs to production. Instead surplus (slack) activities can be thought of as warehouses in which inputs can be stored at zero cost, i.e. disposed of freely -- at no cost. Having inputs available does not mean they are consumed. In addition, it is worth noting that in a linear program any input <u>always</u> has a shadow price associated with it. Consequently, in linear programs the physical quantities of inputs are not separated from their values.

Free disposal probably originated in the slack activities of linear programming. This in itself casts some suspicion on its theoretical rigor since linear programming is not a general theory of production but is, instead, a operational method for optimizing returns under a set of extremely restrictive and often unrealistic assumptions which simplify calculations.

### 5.1.2. BOLES:

Boles' (1966, undated) contribution is worth noting since he develops an algorithm closely related to a linear programming algorithm. Boles restricts his discussion to the mechanical aspects of constructing an algorithm that will compute technical and price efficiency indices. While he provides no new insight into the theory, the algorithm he develops could have a very practical application in applied work, in determining the degree of specification or aggregation error one has in a sample and the need to carry out additional detailed investment/disinvestment analysis of the causes of apparent interior points -- that is, by recognizing that Boles measures of technical and price efficiency are essentially measures of specification or aggregation error, his algorithm serves a potentially legitimate and useful diagnostic value.

## 5.2. FRONTIER PRODUCTION FUNCTION THEORY

### 5.2.1. FARRELL:

Farrell (1957) is credited with formally introducing the theory of frontier production functions. It is clear that Farrell developed his ideas from association with linear programming.

> ... the treatment of the efficient production function is largely inspired by activity analysis.... [Farrell, 1957, page 11]

Except for extending his theory to include situations where constant returns to scale need not be assumed [Farrell and Fieldhouse, 1962], and some elaboration and interpretation [Bressler, 1966], Farrell's theory has been largely accepted without critcal challenge. Nerlove (1965) and Yotopoulous (1974) point out some critical ambiguities, but still accept the fundamental premise.

> It is the virtue of the present method that it separates price from technical efficiency. [Farrell, 1957, page 264]

Farrell himself states that the concept of technical efficiency is a consequence of aggregating inputs, management, and measurement error. Recognizing this destroys the distinction between TE and PE.

Price efficiency deals with choosing the optimum set of inputs along an isoquant. Price efficiency (allocative efficiency) is the point where an isoquant is tangent to a budget constraint. This determines the quantities of variable inputs V(J) purchasable with a given budget that will maximize production. Similarly, technical efficiency maximizes physical output from a given bundle of inputs. If technical efficiency and price efficiency are different, then it implies that V(J) can produce different levels of output, each level having a different value. There are two possible explanations.

Technical efficiency implies that given amounts of the V(J) can produce different amounts of an output. Clearly, two different outputs can not be on the same isoquant since one output is smaller than the other. If the sub-production function is the same for both sets of inputs, then the two isoquants are in the same input requirement set, and the "extra" input and output must be freely disposed in order to move from the higher to the lower isoquant. But because of polar reciprocal sets, the TE and PE of the two points on the two isoquants are the same and there can be no difference between TE and PE. This flatly contradicts the statement that Farrell's method "separates price from technical efficiency."

The second case suggests that the quantity of output forthcoming from V(J) is not uniquely determined by the production processes. Nowhere is there an explanation of how identical inputs can produce different output; that is, there is never an explanation of what accounts for the differences in production. Therefore, price efficiency does not insure maximum value, only a proportionate relationship; between the v's in V(J). If the bundle is used technically efficiently one gets one output but if its used technically inefficiently then one gets a different output. Therefore, the value of V(J) is not uniquely determinable.

If one looks closely, what Farrell treats as technical inefficiency results from comparisons across sub-production functions (specification error) or from aggregating heterogeneous inputs (aggregation error). Indeed, one can infer that <u>mis-aggregation of inputs creates an apparent</u> "<u>technical inefficiency</u>" from his suggestion that dis-aggregating inputs improves the technical efficiency of the firm.

> It will be seen that, in accordance with the theory developed in Section 3, the introduction of a new factor of production in the analysis cannot lower, and in general, raises the technical efficiency of any particular [observation]. Thus, the more factors that are considered, the more are apparent differences in efficiency explained as being due to differing inputs of the factors. [Farrell, 1957, page 269]

Since inputs are defined to be quality (form) specific, one cannot have differences of quality between units of the same input, without committing an aggregation error.

Management is better regarded as choosing production processes, input levels, and hence, output levels than as a factor of production. Consequently, it does not affect the "technical efficiency" of a given set of inputs. Farrell is ambiguous when he says,

> ... the technical efficiency of a firm or plant indicates the undisputed gain that can be achieved by simply 'gingering up' the management. [Farrell, 1957, page 260]

Management chooses the input requirement set. Choosing one set over another set is not maximizing output from a given set of inputs, or 'gingering up' technical efficiency, but an economic adjustment among the alternative sets [Edwards, 1958]. For different input sets, the physical output will be different, but not for identical sets. This is a physical (technical) phenomenon, but implicit is a comparison of two input requirement sets with different fixed inputs. If a one chooses one set rather that another, one is implicitly changing sub-production functions in the process.

The two problems mentioned above can arise due to measurement error. In discussing the practical problems of measuring inputs, i.e., measurement error, Farrell aknowledges,

> ... that a firm's technical efficiency will reflect the quality of its inputs .... [Farrell, 1957, page 260]

Farrell goes on to explain that discrepancies in the measurement of inputs also affects price efficiency, "but in rather a complex way, so that problems are best discussed ad hoc." As an example he states,

> If labour input were measured in man-hours, as is conceptually correct, this too would affect price efficiency, but if in numbers of men employed, it would affect the firm's technical efficiency. [Farrell, 1957, page 261]

This suggests that Farrell regards men as fixed inputs but man hours as variable inputs. Farrell seems to be suggesting that price efficiency is associated with variable inputs and technical efficiency with fixed inputs and that variable inputs cannot be used technically inefficiently while "fixed" inputs cannot be used in uneconomic proportions. When Farrell attempts fully to illustrate the difference between price and technical efficency he attributes the differences to specification error (comparisons across sub-production functions), or error in aggregating inputs.

Farrell's paper contains repeated instances of ambiguity about what one is to assume initially about the production function that gives rise to the "efficient" production function. He is repeatedly ambiguous about how price and technical efficiency can be separated and isolated without interactions. Finally, he repeatedly neglects the implications noting that what he calls "technical efficiency" depends on input mis-specification and/or aggregation errors.

**5.2.2. BRESSLER:**

Bressler's (1966) summary of the frontier production function theory has been presented in Chapter Two. In addition, Bressler offers three explanations for inefficiency in a firm: (1) operating with excess capacity, (2) inefficient use of technology, and (3) using an outmoded or inefficient technology [Bressler, 1966].

Operating with "excess capacity" clearly focuses attention on the role of fixed inputs. If having excess capacity means that to be profitable one must expand use of variable inputs, then the firm is simply inefficient. A second interpretation is more conventional: excess capacity suggests the need to make length of run adjustments (changing sub-production functions), which is treated below.

Since the production possibilities set defines the technology, Bressler appears to be using a different definition of technology in cases two and three. Technology is not chosen, but given by a state of knowledge. Bressler's technology appears to be the same as the input requirement set, or technique, of Appendix Two. Bressler's cases two and three, efficient use of technology and choice of technology, imply efficiency in general rather than just "technical efficiency." In light of the theory of production functions presented in Appendix Two, one can conclude that the choice of technology and the efficient use of technology are not merely physical (technical) decisions but are inherently economic decisions based on evaluating the costs and returns involved. Choosing a technology raises the issue of technical change and makes the selection of fixed and variable inputs carrying old and new technologies endogeneous to the system (see Edwards, 1958). Changing technology implies changing sub-production function. Using a technology efficiently means first, finding the tangencies between the budget constraints and isoquants and then the high profit level of production. "Using an outmoded or inefficient technology" suggests that one has not properly considered the economies of investing in or using inputs carrying new technologies and disinvesting in or ceasing to use inputs carrying old technologies.

Bressler's comparisons of efficiency are directly attributable to making comparisons across sub-production functions. As mentioned in Chapter Two, Bressler points out that economic efficiency has a direct relationship to average costs. Specifically, if one maps the inverse of a production function's longer-run average total cost curve, one finds a curve representing maximum efficiency for the production function at each level of production, e.g. point A maps to point D in Figure 5.1. Each level of production represents an isoquant. If one observes a point within this "efficient envelope," point E, to what may it be attributed? There are only two possible explanations.

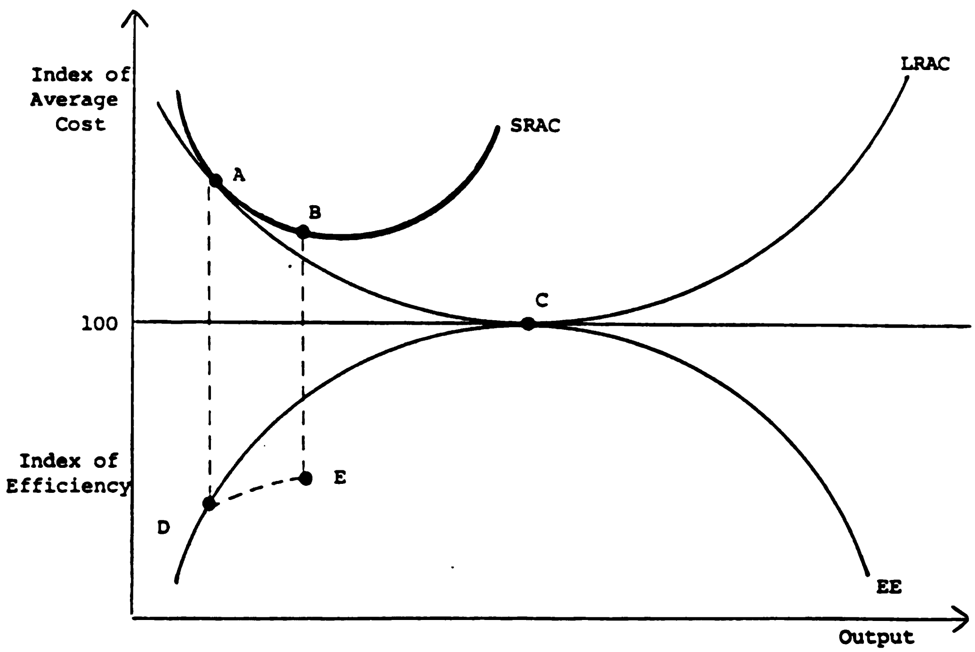



FIGURE 5.1

A SHORTER-RUN AVERAGE COST CURVE (SRAC), A LONGER-RUN AVERAGE COST CURVE (LRAC), AND AN "EFFICIENT' ENVELOPE" (EE)

First, if the efficiency envelope represents maximum efficiency at each level of output (for each isoquant), then the efficient envelope is an expansion path. One would be inefficient if one used inputs in proportions different from the proportions determined by the tangency of the isoquant and the budget constraint. One would be inefficient, but still be on the isoquant. Point E must be on the same isoquant as D, but not on the budget constraint. D and E produce the same amount of output. Thus, E is inefficient from an allocative point of view and does not illustrate technical inefficiency. Technical efficiency cannot be determined from this presentation. Technical inefficiency, as Bressler uses the term, is ambiguous.

The second possiblity is that the interior of the efficient envelope maps back to the interior of the longer-run average cost curve, e.g., point E to point B. But what is in the interior of a longer-run average cost curve? The longer-run average cost curve is an envelope of <u>shorter-run average cost curves</u>. Consequently, being off an efficient envelope means being on a different short-run average cost curve which, by duality, means making a comparison across different sub-production functions without considering the economies of shifting between them; i.e., of making a comparison between firms which are using different levels of fixed inputs without due attention to investment and disinvestment theory.

Consequently, being at point B (off the efficient envelope) is due to the law of diminishing returns when certain inputs are fixed at different levels for A and B. This is a curious treatment of efficiency because the values of the difference in the amount of fixed inputs are not taken into account.

If only the variable inputs are taken into consideration, a firm operating on the SRAC at B in Figure 5.1 appears inefficient because it is not operating on its longer-run average cost curve at A. This is inappropriate. The efficient firm would still be experiencing decreasing costs on its sub-production function at A, and would continue to expand output thereby reducing average total cost at least to B. Furthermore, A is only pertinent in the longer-run, and with respect to the longer-run, A is not efficient, C is. Therefore, A is not an efficient point on either the SRAC or the LRAC. It is simply inappropriate to compare A and B without considering the economies of investing and disinvesting in the inputs which are treated as fixed.

In his conclusion, Bressler aknowledges the aggregation problem, but without aknowledging any of the critical problems it raises as to the legitimacy of the frontier production function.

> First, all these methods are subject to essentially similar problems of aggregation; for example, if we use some aggregate measure of capital inputs in any of these approaches, we are ignoring the fact that capital is a non-

> homogeneous input and $1.00 of capital applied marginally in a small or a large business may represent quite different real inputs. Stated another way, it makes a lot of difference if our marginal capital input is in the form of power shovels or hand shovels. [Bressler, 1966, page 136]

In effect Bressler recognizes that aggregation hides specification error and that when one compares firms, one frequently compares two different sub-production functions. Further, he implicitly recognizes that differences between firms can be attributed to the influence of differences in fixed input on output.

In summary, Bressler, too, fails to make valid distinctions between price efficieny and technical efficiency. The difference which appears to him to exist between the two kinds of efficiency are due to different sub-production functions (specification error) or aggregation error.

### 5.2.3. NERLOVE:

As was mentioned earlier, Nerlove (1965) discusses some of the ambiguities in Farrell (1957). Despite recognizing the short comings of Farrell's approach, Nerlove attempts to retain the technical/price efficiency dichotomy and "attempt[s] to generalize Farrell's work."

Nerlove summarizes the conventional assumptions about production and concludes that there will be no difference in output for profit maximizing firms using the same inputs in the same environment (subject to the same uncontrollable random inputs of the same magnitudes).

> Those differences which do exist must therefore be due to differences among firms with respect to their technical knowledge and fixed factors, their ability to maximize profits, and their economic environments. [Nerlove, 1965, page 87]

Different technical knowledge implies different production possibilities sets and/or technical change. Different technical knowledge and different fixed factors suggests different sub-production functions which alone are not grounds for evaluating differences in efficiency between firms using a physical criteria alone. The ability to maximize profits involves, among other things, finding the tangency between the budget constraint and an isoquant which is not determined solely by technology. Different environments result in different sub-production functions, as defined in Appendix Two. None of these situations allows for "technical inefficiency."

Nerlove quickly pinpoints the weakness of Farrell's argument.

> His measure may be divided into two components. The first, technical efficiency, relates to an improper choice of production function among all those actually in use by firms in the industry. The second, price efficiency, refers to the proper (or improper) choice of input combinations. [Nerlove, 1965, page 88]

The choice of "production function," which corresponds to the choice of sub-production function of Appendix Two, means choosing the amount of the fixed inputs. Nerlove does not define what "improper" means. Had he done so he would have had to consider the economies of investing in variable inputs and of disinvesting in what might be fixed inputs. This may be similiar to Bressler's (1966) reference to excess capacity. But his comment makes it clear that differences in "technical" efficiency are due to being on different sub-production functions, i.e., different levels of fixed input. He does not consider the question of the opportunity costs associated with choosing one sub-production function over another, i.e., whether the total net value of one bundle of resources, both variable and fixed, is greater than or lesser than another bundle of resources.

Nerlove elaborates on this same essential issue in his discussion of Farrell's "quasi-factors" [Farrell, 1957]. Quasi-factors are those inputs which are defined in Appendix Two as uncontrolled random inputs, the collection of which Nerlove calls "the environment." Nerlove implies that firms with different amounts of quasi-factors will produce different levels of output, which is to be expected inasmuch as the quasi-factors act as fixed inputs and influence the onset of diminishing returns. Regarding firms purely with different amounts of quasi-factors as the "same" constitutes a specification error -- e.g., different sub-production functions are regarded as the same.

> The last point is clearly crucial to any definition of relative economic efficiency; in general, some attempt must be made to standardize environment in the construction of the measure, or else the measure will reflect not merely differences in efficiency but also the degree to which the environment of a particular firm is favorable or unfavorable. [Nerlove, 1965, page 90]

He continues by pointing out that distinguishing between price efficiency and technical efficiency is tantamount to mixing short-run considerations (what he calls price efficiency), and long-run considerations (what he calls technical efficiency).

If price efficiency applies only in the short-run, then it deals with the variable inputs. To make efficiency comparisons across firms one must hold the levels of fixed input identically constant among firms. Consequently, differences in efficiency will be due only to differences in the success of each firm in accurately finding the tangency between its budget contraint and an isoquant and its high profit points.

If technical efficiency applies only in the long-run, then clearly it deals with economic adjustment or changing the amount of fixed input -- with investment and disinvestment. This means that changing the amount of fixed input will change technical efficiency and technical efficiency must result from inappropriate comparisons across different sub-production functions. This creates two problems: (1) what are the criteria for making changes in fixed inputs in order to change what is called technical efficiency; and (2), a firm cannot be technically efficient and price efficient simultaneously except at point C in Figure 5.1. A firm cannot be both shorter-run and longer-run efficient except at point C since in general there will be the inherent contradiction noted above that B is efficient but A is not.

Unfortunately, after clearly identifying the short comings of the frontier production function approach, Nerlove equivocates by attempting to "generalize Farrell's" method. In so doing he implicitly assumes that a firm with less fixed input than another firm is "technically" less efficient. In assuming that firms are <u>free</u> to choose whatever level of fixed input they want in order to minimize long-run average costs Nerlove implicitly assumes there is no cost in changing sub-production functions. If this were true, all firms would be operating at the minimum of their long-run average total cost curve and there would be no such thing as inefficiency, "technical" or otherwise.

In Appendix Two it is noted that Edwards (1958) offers a way out of this dilemma by showing that the fixity or variability of inputs can be endogenized to the system for the purpose of finding the most efficient point, without changing any of the basic conclusions of the theory presented.

**5.2.4. YOTOPOULOUS:**

Yotopoulous (1974) comes closest to dispelling the technical efficiency fallacy. He repeatedly identifies the real sources of "technical inefficiency."

> The difference in output between the "average" firm and the extreme positive outlier is used to measure the technical inefficiency of the average firm. Another interpretation, of course, could have the "average" firm representing the norm and positive outlier representing an unusual endowment of some fixed factor of production, such as entrepreneurship, or luck. It may represent the classical source of error in measurement or of noise in the universe, and as such it can imply nothing systematic about efficency. [Yotopoulous, 1974, page 264]

This clearly suggests that given a comparison between two firms, the "technically efficient" firm is so due to some additional amount of some fixed factor though entrepreneurship is a poor candidate for reasons discussed earlier in considering Farrell's contribution. When firms have the same amount of fixed input, and face the same prices, then differences in efficiency are due either to error in maximizing profits which is simply inefficiency; or to differences in the uncontrollable random fixed inputs, which cannot be appropriately called "technical efficiency."

> ... the remaining differences in observable input mixes can be attributed to two factors. First, they can be traced to differences in nonmeasured fixed inputs of production. These can be

> readily captured through the analysis of variance as used to measure management bias. They constitute the component of technical efficiency. Second, the results can be attributed to residual differences that are due to imperfect equalization of marginal products to opportunity costs. These constitute the component of price efficiency. [Yotopoulous, 1974, page 269]

Like Nerlove (1965), Yotopoulous failed to exploit his insights; instead he attributed "technical" efficency to the "environment" and to uncontrolled random inputs. This makes differences in "technical" efficiency among firms due to chance or to being on different sub-production functions. Yotopoulous constructs, hypothetically, the set of situations where one might observe differences in technical efficiency between two firms. He specifically attributes the differences in "technical" efficiency to being on different sub-production functions: The differences in firms which he treats as differences in "technical" efficiency are due to differences in the amount of fixed input being used.

> In Panel II comparison of technical efficiency becomes possible since the isoquants belongs to production functions that differ only by the constant. This term represents differences in endowments of fixed factors as well as the impact of nonmeasurable inputs, such as entrepreneurship. Technical efficiency is the shorthand notation for such differences. [Yotopoulous, 1974, page 266]

In his discussion of these various cases, only the "efficiency" of the variable inputs is actually compared. He does not introduce the economies of investing or disinvesting to change sub-production functions. In effect, technical efficiency becomes a comparison of the productivities of the variable inputs, ignoring his own critical point that these productivities will differ with differences in the quantites of fixed inputs. This is the same thing as ignoring the contribution to production and the costs of the fixed input(s).

Despite repeated statements demonstrating that technical efficiency cannot exist if one maintains the usual set of assumptions about production, as stated in Appendix Two, Yotopoulous retains a belief that something called "technical efficiency," and its logical corollary "technical inefficiency," exist. Firms that have the same production function, but different amounts of fixed input, are said to have "neutral differences in technical efficiency." That is, firms on different sub-production functions display no differences in technical efficiency, which contradicts everything he has previously developed. He states:

> Technical inefficiency, on the other hand, is related to the fixed resources of the firm. It is an engineering datum and as such, at least in the short run, it is exogeneous and part of the environment that is taken as given. [Yotopoulous, 1974, page 271]

What this suggests is that the amount of fixed input used is either not a choice -- an example of an uncontrollable input -- or that the choice is made without any reference to costs.

Like Nerlove (1965), Yotopoulous fails to exploit the opportunities revealed by his own observations and insights. Consequently, he presents a "computational form that combines the three elements, technical, price, and economic efficiency." Prehaps because his "form" excludes consideration of fixed inputs, and fixed costs, he does not detect the inconsistency of his "form" with his own arguments.

## 5.3. FORMAL MICROECONOMIC THEORY

Unfortunately, some of the leading microeconomic texts of the last twenty years have institutionalized the error of distinguishing between "technical" and "price" efficiency. As in the frontier production function literature itself, the distinction is postulated with little, if any, attention to the logic of such a distinction, or its contradictions with the usual assumptions made regarding production processes.

### 5.3.1. HENDERSON AND QUANDT:

In the "Basic Concepts" of production theory, Henderson and Quandt (1971) state that the "production function states the quantity of his [the manager's or entrepreneur's] output as a function of the quantities of his variable inputs." One or more additional inputs are considered fixed within the production function. Consequently, for two firms to have the same production function, they would have to be constrained by the same amounts of the same fixed inputs. This corresponds to the firms having the same sub-production function as defined in Appendix Two.

In defining technology, and explaining how it is different than a production function, Henderson and Quandt contradict the laws of thermodynamics. They begin by stating that,

> The entrepreneur's technology is all the technical information about the combination of inputs necessary for the production of his output. it includes all physical possibilities. [Henderson and Quandt, 1971, page 54]

It would seem they have defined the production possibilities set. Appendix One and Two make it clear that the laws of thermodynamics substantiate that the production function is a surface and not a solid; no isoquant has an interior due to identical quantities of variable inputs producing more than one quantity of output given the same fixed inputs. However, Henderson and Quandt contradict this in their next sentences.

> The technology can state that a single combination of [V1] and [V2] can be utilized in a number of different ways and therefore can yield a number of different output levels. The production function differs from the technology in that it presupposes technical efficiency and states the <u>maximum</u> output obtainable from every possible input combination. The best utiliza-

> tion of any particular input combination is a technical, not an economic, problem. The selection of the best input combination for the production of a particular output level depends upon input and output prices and is the subject of economic analysis. [Henderson and Quandt, 1971 page 54]

This suggests that within the production possibilities set, two identical sets of inputs can produce different sets of output. The crucial phrase is "utilized in a number of different ways." The inputs can be used in different ways only if they are used in different relationships to each other with respect to their time, form, or location. In any of these cases the inputs are not identical. As is pointed out in Appendix Two, the time, form, and space (location) of an input is held constant in its definition; two apparently identical inputs that differ in either time, form, or space are in fact different inputs. Consequently, "a single combination" of two identical variable inputs cannot be "utilized in a number of different ways." The "way" the inputs are "utilized" is captured within each input vector in the input requirement set. The choice among the different ways of using inputs is as economic as any other choice: it depends on whether or not "it pays" to make the change in time, form, or location, and is therefore economic and not merely "technical." If one observes different quantities of output as a result of using two identical bundles of variable inputs, then the difference must be due to differences in the amount of the fixed input[s] used. Then, by Henderson and Quandt's own definition of a production function, the different outputs are produced by different production functions, not by a difference in technical efficiency.

**5.3.2. McFADDEN:**

McFadden's (1978) development of microeconomic theory is ironic because he maintains a distinction between technical efficiency and price efficiency, and at the same time develops the duality theory of polar reciprocal sets that eliminates any possible distinction between the two. In fact it is done within the same context, distance functions (see Appendix Two).

His treatment of distance functions creates ambiguity and misunderstanding as was suggested in Chapter Three. The similiarity of the unit isoquant of a frontier production function, and a distance function is more than coincidence (see Figure 3.1). It was suggested that a frontier production function is mistakenly identified as a distance function, where "a" is a measure of "technical efficiency." This suggests that the scaling process implicit in distance functions changes the quantities of v, but not the original marginal physical products associated with them, which is not generally observed if there are fixed inputs.

> A positive input bundle (v, z) is <u>efficient</u> for an output bundle y and distance function F if F(y,v, z) = 1

> and any distinct postive input bundle (v', z) with (v', z) <= (v, z) has F(y, v', z) < 1. Alternately, define an input bundle (v, z) to be <u>efficient</u> for an input requirement set V(y) if any distinct input bundle (v', z) with (v', z) <= (v, z) has (v', z) $\notin$ V(y). [McFadden, 1978, page 30; z's added]

What gets lost in this scaling is the corresponding expected changes in marginal physical products. However, this frontier function, or "efficient set" is slightly different than the frontier in Chapter Two. In including a reference to a graph of an isoquant (see McFadden, 1978, page 17), McFadden indicates that "inefficient" points are points <u>on</u> the isoquant. Since the distance function moves all "interior" points to the isoquant by free disposal, the inefficient points must be points that are not on the expansion path for a given level of output.

McFadden developes distance functions chiefly as one way to prove duality between production functions (excluding Stage III) and cost functions. The one to one correspondence between prices and inputs, between physical quantites and their values, is summarized in the definition of the polar reciprocal sets that were discussed in Chapter Three. Polar reciprocal sets prove there is no possible distinction between technical and price efficiency.

## 5.4. OTHER LITERATURE

The foregoing has dwelt on the literature important in establishing the erroneous theory of frontier production functions. There is a growing volume of other literature that may be grouped into three categories: (1) methods for estimating frontier production functions, (2) applications of frontier production functions, and (3) miscellaneous or related concepts. This other literature will be dealt with only briefly since it does not focus on the development of the theory of frontier production functions per se, but rather accepts the premise that frontier production functions exist.

The first group concerns itself with how one might estimate frontier production functions, or measure technical and price efficiency between firms. This literature would be of more value if it focused instead on issues of specification error or aggregation error which together constitute the discrepanies between firms that are mistakenly attributed to differences in technical efficiency. The contributions of Boles (1966, undated) within the context of linear programming, or activity analysis, has been alluded to above. Timmer (1971) attempted to measure "technical efficiency" using a specific functional form and a mathematical programming algorithm. It marks one of the first attempts to measure "technical efficiency" parametrically. The more recent literature attempts to develop a method of estimating frontier production functions parametrically using an econometric approach [Forsund, Lovell, and Schmidt, 1979, Forsund and Hjmalmarsson, 1974, Forsund and Hjmalmarsson, 1979, Schmidt and Lovell, 1977, Schmidt and Lovell, 1978]. The focus of this literature is on the specification of the error term, which has either a one sided distribution (all the errors are of one sign), or is divided into two components, one representing the usual error term while the other is one sided representing differences due to "technical inefficiency."

Despite the attention to the development of the theory of, and methods for estimating, frontier production functions, there have been few attempts to apply frontier production functions [Bravo-Ureta, 1983, Hall and LeVeen, 1978, Lesser and Greene, 1980, Seitz, 1966]. Most applications have used a non-parametric method, while applications using a parametric method have served chiefly as examples of a new or improved method.

The last group contains literature that does not always refer explicitly to frontier production functions but clearly offer theories of management or decision-making that are closely akin to the concept of a frontier production function [Charnes, Cooper, and Rhodes, 1978, Charnes, Cooper, and Rhodes, 1981]. The best known, and prehaps best example of this literature, is Leibenstein's theory of "X-Efficiency" [ Leibenstein, 1966]. The theory of X-Efficiency suggests that different firms using the same inputs to produce the same output will have different degrees of efficiency depending on how the production process is organized in practice. While the concept may be useful to a decision-maker in a very applied sense, in the strictest sense, it violates the definition of homogeneous inputs given in Appendix Two as the inputs are not time, form, and location specific and ignores the costs and returns involved in investing and disinvesting.

## 5.5. SUMMARY

Perhaps the oddest aspect of the frontier production literature is that it appears to be a theory developed as a consequence of a technique (linear programming), rather than the other way around. In the earliest literature, dealing with activity analysis, what is called technical efficiency is assumed or postulated without a clear description of what technical efficiency means, or how it differs from the traditional concept of economic efficiency. Since there is no explicit discussion that would suggest that an isoquant can be represented as a plane, one would expect more suspicion of an "unit isoquant" which creates a plane by dealing in <u>average</u> production, rather than the more traditional focus of <u>total</u> production.

Implicit aggregation and specification error is evident from a careful reading of Farrell [Farrell, 1957, Farrell and Fieldhouse, 1962], and several successors, including Bressler (1966), Nerlove (1965), and Yotopoulos (1974). Despite this, each author eventually (uncritically) accepts the existence of a frontier production function without attempting to specify it in a proper manner. Thus, the concepts of frontier production functions have been uncritically incorporated into statements of microeconomic theory [Henderson and Quandt, 1971, Malinvaud, 1972, McFadden, 1978, Quirk and Saposnick, 1968, Varian, 1978]. Ironically, McFadden's effort to prove duality mathematically provides the theoretical refutation of the existence of frontier production functions by describing the characteristics of distance functions and polar reciprocal sets [McFadden, 1978]. None-the-less, he simultaneously maintains the logically inconsistent notion of "technical efficiency." This inconsistency appears to be the result of his discounting the importance of those aspects of traditional theory that are commonly considered to be 'economically irrelevant;' i.e., Stages III, because normally all the points in Stages III are inefficient [McFadden, 1978]. Unfortunately, this oversight by McFadden (1978) creates ambiguity in those aspects of the theory that he preserves.

Finally, there is the other literature described in 5.4 above, which adds little to the debate of whether or not frontier production functions exist, but if properly directed, might provide methods for dealing with the problems of specification and aggregation error and lead to incorporation of investment/disinvestment theory.

# CHAPTER SIX

# CONCLUSION

Frontier production functions are supposed to be different from, and better than, traditional production functions. They are supposed to represent the "best performance" obtainable by a firm given some set of endowments and some given technology. They do this by distinguishing between the "technical" and "price" efficiency of the 'best performer' and the other firms included in the comparison. This dissertation maintains that rather than measuring "best performance," frontier production functions measure specification and/or aggregation error. This is because the theoretical basis upon which frontier production functions identify "best performance" is invalid.

## 6.1. VARIATIONS IN DATA IN APPLIED WORK

In any applied work there will be variation in the data one collects. The real world does not conform to theoretical conditions of homogeneity, perfectly competitive markets, etc. This lack of conformity does not mean that microeconomic theory is failing to explain reality.

Microeconomic theory, like all theory, is a guide to analysis after one has decided what is important and what is unimportant to the issue being investigated. One must take a preliminary step of using one's judgement as to what constitutes "sameness" and what constitutes "close enough" when specifying a sub-production function and aggregating data across observations for the purposes of the investigation. Once the degree of "sameness" and "closeness" has been determined, the remaining variation in the data is implicitly "unimportant" and irrelevant to the inferences to be drawn from the analysis. The statistical properties assumed to be exhibited by the data as specified and aggregated serves as a means of filtering out the remaining variation. The remaining "noise" is assume to be captured in an "error term" and estimation is done at the mean, or average value, of the data.

The variation that remains means that if one plots the data there is a scattering of points, a distribution of the observations. The question examined in this disseration is essentially which observations in this distribution should serve as the bench mark from which comparisons of the observations might be made. As such, it examines the nature of the remaining variation; it explains what causes the variation one observes between observations after one has specified a common functional representation for the technical relationship among the specified and aggregated variables of production. If one chooses the traditional bench mark of the data's mean, then one is implicitly assuming that the remaining variation is due to random unexplainable and inconsequential phenomena, that the variation that remains is simply due to the "unimportance" of measuring and explaining the remaining imperfections of the real world.

If one chooses the theory of frontier production functions, then one is assuming a bench mark of one extreme of the observations, which is an implicit assumption that the remaining variation is systematic and due to consequential and measureable causes, "technical efficiency" and "price efficiency." Therefore, the "less efficient" observations all lie within the interiors of the isoquants of the extreme observations.

## 6.2. ISOQUANTS DO NOT HAVE INTERIORS

Chapter Two, Three, Four, and Five discuss the frontier production function and traditional theory explanation of these interiors to isoquants.

Chapter Two is a summary of the characteristics of frontier production functions that make them different from traditional production functions. Frontier production functions originated from the concept of the "unit isoquant." Certainly plotting real production data, with its inherent "remaining" variation, on a per unit of input to per unit of output basis will reveal a scattering of the observations, as was suggested above. Two cases for inferring that this scatter of observations represent solid production sets was discussed. It was shown that both cases of these frontier production functions were a violation of traditional microeconomic theory due to specification and/or aggregation error in identifying the frontier production function.

Chapter Three explored the critical case of whether or not frontier production functions are accurate representations of distance functions. Superficially, they appear to be the same. They have the same basic assumptions of concavity and monotonicity, and they both have 'interiors.' The issue is whether or not the unit isoquant is the same as an input requirement set. Using the principles of the duality theory that distance functions serve to prove [McFadden, 1978] and the free disposal that is simultaneously assumed for both production and cost space, it was proved by contradiction that frontier production functions are a violation of the theory of distance functions. This was accomplished by a careful accounting of what is in the 'interior' of a distance function (higher isoquants) compared to the interior points conceived to be within the surface of a frontier production function.

Chapter Four presented a new case of an unit isoquant. It explained what the unit isoquant demonstrates in reality. It returned to the original formulation of the frontier production function and correctly inteprets the fact that observations on production on a per unit of input to per unit of output basis provides an unit isoquant with an 'interior.' It pointed out that "technical efficiency" would identify the point of maximum average physical output, the boundary between Stages I and II on a production function with both stages in the production of firms in long-run perfectly competitive equilibrium. As such, "technical efficiency" is identical to traditional economic efficiency.

Chapter Five reviewed the salient frontier production function theory revealing that the proponents of this new theory repeatedly and consistently reveal the apparent interior points of frontier production functions to be due to specification and/or aggregation error yet do not explore the theoretical consequences of such revelations.

## 6.3. THE INTERIORS IN FRONTIER PRODUCTION FUNCTION THEORY AND "TECHNICAL EFFICIENCY"

Frontier production functions represent not just a change in the "bench mark" observations in applied work from the mean points of the data to the extreme points of the data. Frontier production function theory is questionable as the apparent 'interior' points result from specification and/or aggregation errors.

The traditional view of the world is that identical circumstances result in identical outcomes (see Appendices One and Two). It is this stability of the real world that makes events predictable. By contrast, the fundamental postulate of frontier production function theory is that identical circumstances may result in different outcomes and that the differences are not attributable to chance variations in uncontrolled variables. This is what explains an interior to frontier production functions that is non-existent in physical reality.

Frontier production function theory suggests that two production events can differ in their degree of "technical efficiency" and that a "technically less efficient" producer can become more "techncially efficient." The new theory does not explain how this can be accomplished. If one is to improve one's "technical efficiency," then one presumably must "improve" or change one's initial circumstance in some way. But frontier production functions assume there is no difference in the initial circumstances of the "technically efficient" and "technically inefficient" producers; they have the same inputs, the same output, and the same technology. If the change involves changing the time, form, or location of some aspect of production (other than the differences in time, form, or location that are assumed to be "unimportant"), then a specification and/or aggregation error has been committed in identifying the differences in "technical efficiency." If one is to change one's "technique" then one must change the sub-production function one is using in production, and "techncial efficiency" becomes an incomplete comparison across sub-productions.

The error involved in conceiving that frontier production functions have interiors which can be technically corrected is that of ignoring the contribution and opportunity cost of changing sub-production functions by investing or disinvesting in "fixed" inputs.

### 6.4. INTERIORS TO FRONTIER PRODUCTION FUNCTIONS WOULD MEAN "TE" $\neq$ "PE"

If all points of equi-production are on an isoquant, then a producer cannot be "price efficient" without simultaneously being "technically efficient." Being "price efficient" means being on the expansion path of production. Moving along the expansion path of a full production function means changing sub-produciton functions by investing or disinvesting in fixed inputs. This means that the fixed inputs are temporarily variable inputs. In order to completely account for "price efficiency" one must account for the prices (opportunity costs) of the fixed inputs (see Edwards, 1958).

Can a firm be "technically efficient," but not "price efficient" by being on the same isoquant as the "price efficient" firm, but off the expansion path? If the prices paid for all inputs are different for the two firms, then in a world in which duality is assumed to exist, both firms will be "price efficient" if they are both "technically efficient," as was demonstrated in Chapter Three. If prices are the same for both firms then the firm that is not "price efficient" is not "technically efficient," since it can achieve more output with its resources by selling and/or buying its inputs without changing its total investment or expenditures. That is, the "price inefficient" firm will pay a greater cost for its bundle of inputs, so that its bundle of inputs has a greater value. That is, the firm implicitly has a greater budget constraint. By trading some of its inputs for more of the other inputs in the market place it can "costlessly" rearrange its bundle to be a "price efficient" bundle still of greater value than the bundle of the originally "price efficient" firm. Thus, the adjusting firm will be implicitly able to produce a greater output, since it will have more of all inputs due to its larger constraint.

The opportunity cost principle means that physical quantities of commodities are inseparable from their value (prices). The measure of thermal efficiency (an example of "technical" efficiency in engineering) is defined as useful output to costly input. In this treatment both inputs and outputs are measured in a common physical denominator. When this ratio is unity, efficieny is 100 percent. Similarly, economic efficiency is marginal revenue product to marginal factor cost. When this ratio equals unity for all inputs, <u>including fixed inputs</u>, efficiency is 100 percent. In this definition both commdities are measured in a common denominator of prices. When one finds the profit maximizing point by equating ratios of marginal physical products to the corresponding price ratios, one is implicitly comparing useful output to costly input where a common denominator measurment that is either physical or value is lacking.

## 6.5. THERE IS ONLY ONE TYPE OF EFFICIENCY

Production efficiency is measurable only within some clearly defined set of circumstances or constraints. If the preconditions for comparing the efficiency of two or more firms are identified, the decisions made as to what constitues "sameness" and "close enough," then efficiency can be measured. This efficiency can not be separated into "technical efficiency" and "price efficiency" since the two are identically equal. If they were not equal, as suggested by frontier production function theory, then some aspect of the established preconditions has been violated, either in theory, or in practice. Efficiency is efficiency which is a maximizing of profit, when inputs of given value are used to produce the greatest value.

## 6.6. FINAL CONSIDERATIONS

There are three issues that might deserve further attention.

(1) For expediency, the economics of making input fixity and variability endogeous to the input requirement set has been avoided in this work. That one can endogenize decisions to invest and disinvest fixed inputs has been established by Edwards (1958). That so doing does not alter the conclusions about efficiency has been postulated, rather than proved. The fixed inputs were not made endogeneous lest they become confused with variable inputs. Even when they are endogeneous, they behave as

fixed inputs, not as variable inputs.

(2) The methods for separately measuring "technical efficiency" and "price efficiency" on frontier production functions clearly do not do so. What is actually measured is specification error and/or aggregation error when production function analysts measure technical inefficieny. Both of these errors create significant problems in applied work since both bias estimates of a system's parameters. Proper identification and measurement of these errors might allow one to aggregate or disaggregate, specify and respecify, production relationships in applied work in order to minimize, or at least explicitly account for, the degree of specification or aggregation error present in the analysis.

(3) Finally, more attention should be paid to whether or not the proof of duality theory requires free disposal, input requirement sets, and distance functions when nonstochastic interior points are known to be absent for a production function. These three concepts assist proving duality theory by excluding Stage III of the production function and by eliminating possible 'interior' points. Aside from this they add little if any insight into production theory. They are often misunderstood, and consequently, researchers attempt to introduce the concepts into the analysis of applied problems resulting in questionable inferences, and misleading prescriptions.

Free disposal, which has no counterpart in reality, is especially misleading. Without free disposal there can be no input requirement sets and distance functions. Free disposal is accomplished mechanically by introducing a scalar into the analysis. The scalar, is the factor by which all additional inputs have to be reduced to sink to the bounding lowest isoquant, within a given input requirement set, the 'scale' by which all the output and input levels must be scaled back to shrink production to the level of the lowest isoquant. It is a measure of the "distance" between the two isoquants. Thus, by using free disposal, within the context of a distance function, one can make any two sub-production functions identical simply by reducing one to the other by scalar multiplication. The only means for determining the scalar is as a function of what needs to be disposed in order to eliminate any difference between two sub-production functions, or two isoquants. The scalar has no counterpart in physical reality, and suggests consequences that obscure the true differences between two sub-production functions or isoquants, and what accounts for those differences.

# APPENDICES

## APPENDIX ONE

## EFFICIENCY AND THE LAWS OF THERMODYNAMICS

Economic theory is largely a theory about the behavior of people. It focuses on how people make choices, and offers guidance in selecting the best among competing choices. Each choice represents an opportunity; an opportunity to experience the utility embodied in that choice. To say that something has utility means that it has value for the consumer. That is, if a commodity can provide utility it has value. In economics the value of an opportunity (in some sense a measure of utility) is captured by the concept of opportunity cost. In physics the concept of utility is captured in the concept of work. Because work can provide utility it has value.

In production economics one speaks of transforming "inputs" into "outputs." Production is a _cycle_, wherein energy is transformed. It is assumed that one is able to derive utility from the output. Since the transformation is usually physical in nature, and often involves changes in the _state_ of the production _system_, technical efficiency is relevant. Thermal efficiency, hereafter treated as synonomous with technical efficiency, is one of several types of technical efficiency. By examining what thermal efficiency is, one can appreciate what technical efficiency means in general.

This chapter will first define the terms and relationships of thermodynamics that are pertinent to economic production theory as it relates to efficiency. The second section presents the fundamental definitions. The third section will focus particularly on the first and second laws of thermodynamics and demonstrate that the "technical efficiency" of frontier production functions cannot exist. The fourth section will define and explain technical efficiency, or thermal efficiency, as it is used in thermodynamics. It will be shown that technical efficiency in thermodynamics is the same as economic efficiency. The last section will explain specifically how and why TE and PE are inconsistent with thermodynamic theory.

## A1.1. DEFINITIONS FROM THERMODYNAMICS

The laws of thermodynamics are observations on the physical relationships in the transformation of energy from heat to work or vice versa. Energy is the ability to do work. There are two types of energy; energy that is stored and energy that is in transition. Work and heat are energy in transition.

The laws of thermodynamics explain the work (heat) that can be obtained from resources, and some of the restraints for doing it. The first law is the well known principle of the conservation of energy. The second law deals with the the amount of thermal energy that will become "useful" work by means of a given cycle. It leads to a definition of thermal efficiency which is a type of technical efficiency.

**A1.1.1. SYSTEMS:**

In thermodynamics, one investigates the nature and behavior of physical systems. Systems interact with each other, or their environment, producing work or heat in the process; energy is transformed. In order to study the interaction between two systems, they must be insulated from the environment so that no thermal energy escapes into the environment and is thereby unaccounted for in the interaction between the two systems. This is called an adiabatic system. In order to simplify the investigation of the principles of thermodynamics, it will be assumed that one has a system and its environment (which is in effect another system) and that together they represent an adiabatic system.

A system is either a particular collection of matter, a closed system, or a particular region of space, an open system. Briefly, in a closed system no matter can cross the system boundary, while in open system matter can cross the boundary. Interactions will occur when the state of the system is out of equilibrium with the state of the environment, and the system is not insulated from the environment; when there is no barrier to the exchange of energy between the system and the environment. When there is an exchange of energy between a system and its environment, the energy is transformed such that the system may change state, and/or some of the energy is transformed into heat and/or work. This appendix will be confined to exploring the characteristics of closed systems for the sake of brevity. All of the principles governing the characteristics and behavior of closed systems can be easily assigned to open systems.

A1.1.1.1. STATES:

The state of a system is evaluated by its equation-of state. In a very simple case, for an ideal gas, this relationship might take the form,

(A1.1) $pv = RT$

where

p = PRESSURE

v = VOLUME

R = CONSTANT; USUALLY A CONSTANT ASSOCIATED WITH A PARTICULAR GAS

T = ABSOLUTE TEMPERATURE

When a system exchanges energy with its environment, it often results in a change in the values of the variables in the equation of state. For example, for a change in state of a system (A1.1), P, v, or T would change in value.

## A1.1.1.2. PROCESS:

The path which describes the exchange of energy between a system and its environment is called a process.

> A process is any transformation of a system from one equilibrium state to another. A complete description of a process typically involves specification of the initial and final equilibrium states, the path (if identifiable), and the interactions which take place across the boundaries of the system during the process. Path in thermodynamics refers to the specification of a series of states through which the sytem passes. [Wark, 1983, page 10]

In thermodynamics, a process transforms a system from one state to another state. A process is a path function since it encompasses the system's changing states from its inital to its final state. The values of the initial and final states of the system in thermodynamics may, or may not, be equal. That is, the system may return to its initial state. If the initial and final state of the system are equal then the inital state does not equal the final state for the system's environment. That is, either the state of the system changes, or the state of the environment changes, or both. The change occurs because energy has been exchanged between the two. It is the work and the final states of both the system and its environment that correspond to output in economics.

A1.1.1.3. PROPERTY:

A property of a system is a characteristic, or parameter, of the system. It is a point function, since it is the value of a property at a point. Examples of properties are pressure, volume, temperature, energy, mass, and entropy, but not heat nor work (the difference between temperature and heat is explained below). In Figure A1.1, a system may change state, from S1 to S2, by either Process A or Process B. In either case, the properties of S1 and S2 respectively are the same; they are not uniquely determined by A nor B. The properties p,v, and T, become the means whereby the change in the state of the system can be measured.

When the properties of the system are different than those of its environment, i.e. it is out of equilibrium with its environment, the system can

> ...interact with the environment and produce work until the system reaches a state where such potential differences do not exist. For any system, this state is called the dead state because the system can do nothing more. [Dixon, 1975, page 231]

Whenever a system is out of equilibrium with its environment, the two states differ, there is the potential for an exchange or transfer of energy between the two which will result in work becoming heat or vice versa, until the system reaches its dead state.

## A1.1.1.4. HOMOGENEITY:

Properties of a system are classified as either <u>extensive</u> or <u>intensive</u>. The distinction is important since it has a bearing on the definition of <u>homogeneity</u> in economics. The distinction is as follows:

> Imagine a whole system divided into a number of parts. If the value of a property for the whole system is equal to the sum of its values for the various parts of the system, then it is called <u>extensive</u>. ...intensive properties have meaning at a point or locally. That is, we can talk about the local pressure or temperature but not about the local mass or volume because the latter have no meaning. [Dixon, 1975, page 59-60]

Examples of extensive properties are mass (M) and volume (V), while examples of intensive properties are pressure (p) and temperature (T). Extensive properties may be effectively converted into intensive properties by dividing them by mass, which then equals an <u>average specific property</u>, and finding the limiting value of this quotient at a point, which is called a <u>local specific property</u>. For example, the local specific volume of a system is:

$$(A1.2) \quad v = \lim_{\Delta V \to 0} (\Delta V / \Delta M)$$

where

v = LOCAL SPECIFIC VOLUME OF THE SYSTEM

V = VOLUME OF THE SYSTEM

M = MASS OF THE SYSTEM

Extensive properties are such that they cannot vary within the system; they are the same everywhere. If the same is true for the local specific properties, the system is said to have uniform properties throughout. A system whose properties are uniform throughout is homogeneous. It will be assumed throughout that all systems are homogeneous. This is the same as assuming that all inputs, both variable and fixed, into a production system are homogeneous.

A1.1.1.5. REVERSIBILITY:

Thus far there has been no discussion of the direction of change when a system changes state. If one starts with state S1 in Figure A1.1, and arrives at state S2 by means of process A, and then 'backs up' from S2 to S1 by means of reversing process A, then the process is called reversible.

> A process executed by a system is called reversible if the system and its environment can be restored to their initial states and leave no other effects anywhere. Another term for reversible might be completely restorable. The definition requires that work and heat exchanged between a system and its environment in a reversible process can be restored to each in exactly the same form so that both [the system and its environment] are returned completely to their their initial states. [Dixon, 1975, page 174]

A process is reversible if its effects on a system and the system's environment can be completely returned without

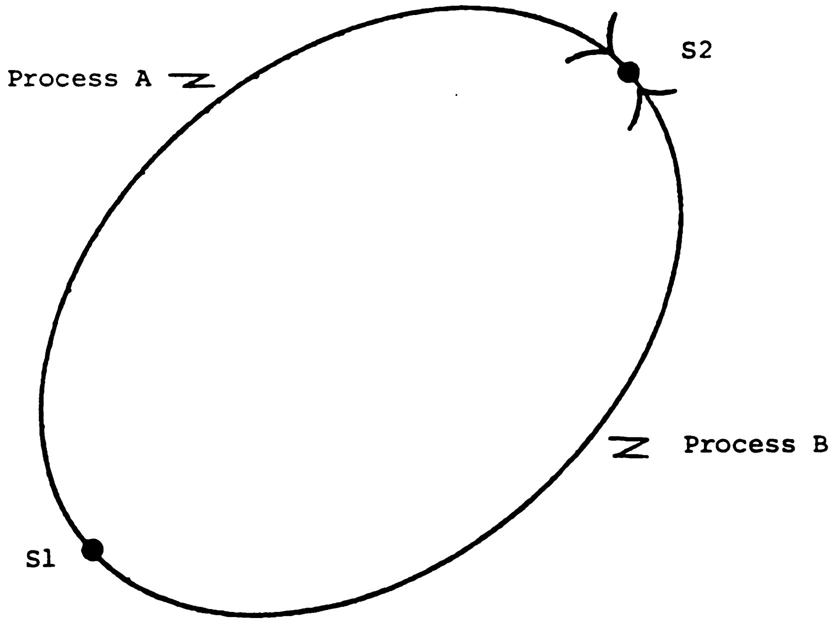



**FIGURE A1.1**

**THE CHANGE IN THE STATE OF A SYSTEM (S1 TO S2) BY MEANS OF EITHER PROCESS A OR B**

additional inputs. In reality, no real process is reversible, as a consequence of the second law. As an example, consider a block sliding down an incline. As it does, it produces heat, i.e., the 'environment becomes hotter.' If one were then to slide the block back up the incline, the environment does not become cooler; that is, the process is not reversible.

A1.1.1.6. CYCLE:

A cycle is a sequence of processes operating on a system such that the final state is identical to the initial state. Figure A1.2 is a cycle, since it takes the system from S1 to S2 where S1 = S2. Recall that although there is no change in the initial and final states of the system, there is necessarily a change in the state of its environment, since real cycles are not reversible. As will be seen later, the development of the second law requires that one can "imagine" a reversible cycle. In particular, the definitions of technical efficiency require using the hypothetical reversiblity of a cycle in order to measure thermal, or technical, efficiency.

**A1.1.2. WORK AND HEAT:**

A1.1.2.1. WORK:

As was noted above, work and heat are not properties of a system. Work is defined at the boundary of the system. Work is a form of energy, commonly measured as

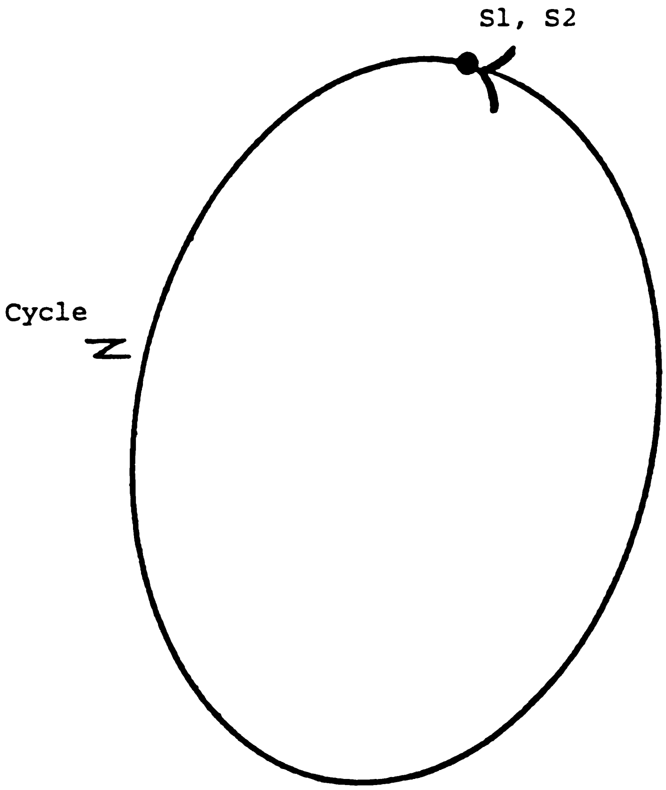



**FIGURE A1.2**

**THE CHANGE IN THE STATE OF A SYSTEM (S1 TO S2) BY MEANS OF A CYCLE**

force times distance:

$$(A1.3)\quad W = \int F\, dX$$

where

W = WORK IN FOOT POUNDS

F = FORCE IN POUNDS

X = DISTANCE IN FEET

In order that the relationship between a system, its environment, and work are explicitly clear, the following definition of work will be used.

> Work is done by a system (on another) when the sole effect external to the system could be the rise of a weight. The amount of work done is the product of the weight (force) times the distance lifted. By convention, work done <u>by</u> a system (which could lift weights in the environment) is taken as positive for that system; work done <u>on</u> a system (the environment lifts weights within the system boundaries) is taken as negative. [Dixon, 1975, page 106]

Clearly, there are a number of different types of systems that can do work. The type of work as measured by (A1.3) is linear mechanical work. Table A1.1 is a partial list of other types of systems and the types of work they do and how that work is measured. Notice that none of the WORK equations are expressed as inequalities; they are all equalities. That is, a given system cannot be "technically inefficient," as the terms are used in the frontier production literature, in performing work given the same processes. The amount of work produced by the system is

TABLE A1.1

SUMMARY OF VARIOUS WORK EQUATIONS

| SYSTEM | FORCE | DISPLACEMENT | WORK |
|---|---|---|---|
| LINEAR MECHANICAL | FORCE(F) | DISTANCE(dX) | $dW = F\ dX$ |
| ROTATIONAL MECHANICAL | TORQUE($\mathcal{T}$) | ANGLE($d\alpha$) | $dW = \mathcal{T}\ d\alpha$ |
| ELECTRICAL CHARGE | VOLTAGE(e) | COULOMBS(dQ) | $dW = -e\ dQ$ |
| ELECTRICAL FIELD | VOLT/METER(E) | POWER(dP) | $dW = -VE\ dP$ |
| MAGNETIC FIELD | AMPERE/METER(H) | WEBER/METER$^2$(dM) | $dW = -VH\ dM$ |

SOURCE: DIXON

exactly determinant for a given cycle, and does not vary. One system can be technically inefficient in performing work as compared to a <u>different</u> system given the same cycle, or one cycle can be technically inefficient in performing work as compared to a different cycle given the same system. Given two <u>identical</u> systems, and identical cycles, the same force and the same displacement (the same inputsand the same technology), one will not get different amounts of work from each; the energy of one system will not be less efficient than the other.

## A1.1.2.2. HEAT:

<u>Heat</u> is another form of energy. It may provide utility and be obtained by transforming work energy. Like work, heat is measured at the boundary of the system and its environment and therefore is not a property of the

system or its environment. It is thermal energy in transition due to a temperature difference. Typically, heat is defined operationally in Btu's; the energy required to raise the temperature of one pound of water at atmospheric pressure from 59.5 to 60.5 degrees Fahrenheit. It should be noted that the temperature of a system may change without heat being transferred either to or from the system, e.g., by a change in pressure. One should also be careful to distinguish between a difference in temperature between a system and its environment, and a change in temperature within a system. These two points emphasize that heat is an interaction between a system and its environment and not a property of either.

### A1.1.3. SUMMARY OF DEFINITIONS:

Energy in most forms is a property of a system, while work and heat are not. Energy is exchanged between a system and its environment in the form of work or heat. Thus, work and heat are defined at the boundary of the system. A process is the means by which the transfer of energy is effected between a system and its environment. Work may be transformed into heat, and vice versa, by a process. Except in the case of a cycle, the results of a process acting on a system are changes in the state of the system. Since real processes are not reversible there is either a change in the state of the system or in the state of its environment, or both, in all cases of a real process acting on system. In production economics, the inital states of the system and the environment corresponds to the inputs, the production function corresponds to cycle, and work or heat and the change in the state of either the system and/or its environment corresponds to output.

## A1.2. THE LAWS OF THERMODYNAMICS

### A1.2.1. THE FIRST LAW:

The first law defines the well known principle that in the absence of nuclear changes or approaching the velocity of light, energy is neither created nor destroyed. Energy can be transferred or exchanged between a system and its environment in the form of work or heat by means of a cycle. Within the cycle, work may be transformed into heat or vice versa. But, neither within the cycle, nor as a result of the cycle, is there a change in the sum of the energy of the system and its environment. Therefore, the first law can be stated as:

$$\text{(A1.4)} \quad \oint (dQ - dW) = 0$$

where

$\oint$ = INTEGRATION OVER A CYCLE

Q = HEAT IN JOULES

W = WORK IN JOULES

This suggests that input = output. Given two identical cycles acting on two identical systems, there can be no differences between them: there is no difference in technical efficiency between them. Given the same inputs, and the same production function one always gets the same output.

If work is energy that provides utility, i.e., work has value, then technical (physical) efficiency must implicitly be a comparison of values. Therefore, there is no basis for a difference between technical (physical) efficiency and price (value) efficiency.

### A1.2.2. THE SECOND LAW:

The second law accounts for the energy that is "lost" when energy is exchanged between a system and its environment. Due to the first law, the energy is not truly lost; rather it is degraded in quality so that it is no longer available to become work, and therefore loses its value. This is called degradation of energy. By observation of the real world:

$$\text{(A1.5)} \quad \oint d'Q/T < 0$$

where

Q = HEAT FLOW IN THE SYSTEM OR ENVIRONMENT

T = ABSOLUTE TEMPERATURE AT WHICH THE HEAT FLOWS

This is the Clausius Inequality. If a cycle were reversible, then in conformance with the above,

$$\text{(A1.6)} \quad \oint d'Q/T = 0$$

where

Q = HEAT FLOW IN THE SYSTEM

T = ABSOLUTE TEMPERATURE AT WHICH THE HEAT FLOWS

The Clausius Inequality is statement of the fact that in reality no process is reversible so that,

(A1.7) $\oint d'Q/T <= 0$

where

Q = HEAT IN THE SYSTEM

T = ABSOLUTE TEMPERATURE AT WHICH THE HEAT FLOWS

and

THE EQUALITY HOLDS FOR HYPOTHETICALLY REVERSIBLE CYCLES

THE INEQUALITY HOLDS FOR REAL CYCLES

Equation (A1.7) is a statement of the second law. It reveals the existence of another property of systems known as <u>entropy</u> (s). Since entropy is a property it is not a function of the process, but a function of the end states of the system. It is defined as:

(A1.8) $ds = d'Q/T$

where

s = ENTROPY IN THE SYSTEM OR ENVIRONMENT

Q = HEAT FLOW IN THE SYSTEM OR ENVIRONMENT

T = ABSOLUTE TEMPERATURE AT WHICH THE HEAT FLOWS

when the process is reversible. While s is always a property, in general $d'Q/T$ is <u>not</u> a property, because of (A1.5), the Clausius Inequality: it is a property only in the limiting cases of a reversible process. One must <u>imagine</u> a reversible process in order to measure the change of entropy in the system and/or environment. Since entropy is a property, it may increase or decrease within a system, or within its environment. But since in reality any process goes only in one direction, due to the Clausius Inequality, the change in the total entropy of a system and its environment must be positive. Therefore, in general, for real systems:

(A1.9) $ds > d'Q/T$

or

(A1.10) $ds > 0$

in an adiabatic real process since $dQ = 0$, due to the definition of adiabatic, and $T > 0$.

**A1.2.3. SUMMARY OF FIRST AND SECOND LAWS:**

The first law means that the amount of total energy in a system and its environment does not change. The second law means that when a real process acts on a system the total entropy in the system and its environment increases; that there is some amount of energy that becomes "bound up" as the increase in entropy, and is unavailable to be transformed into work. If one can <u>imagine</u> a reversible cycle <u>and</u> one can <u>imagine</u> a zero heat flow, then one can imagine a situation where there is no change in entropy for a closed system. In such a situation, all the energy exchanged between the system and its environment would be in the form of work. If there is <u>either</u> heat flow or a reversible cycle, then one can measure thermal (technical) efficiency.

Technical efficiency can be defined when one has a reversible cycle using heat as an input and work as an output. The measure of work output to heat input is the measure of technical, or thermal, efficiency.

## A1.3. TECHNICAL EFFICIENCY

In thermodynamics the definition of efficiency is a relationship between values; it is a ratio of value.

> Efficiency here means the <u>useful</u> output divided by the the <u>costly</u> input, both expressed in energy units. [Dixon, 1975, page 15]

In this definition one can substitute "work" for "useful output." <u>The definition alone eliminates any possible distinction between "technical efficiency" and "price efficiency" as they are used in the frontier production function literature</u>. Since technical efficiency in thermodynamics deals with work energy it implicitly deals with changes in value since work by definition has value; i.e. provides utility. "Costly" is used in the opportunity cost sense, since the input has an opportunity cost. Consequently, in thermodynamics there is no difference between technical efficiency and price efficiency; they are identical.

Recognizing that the second law is a reality, that some energy is "lost" when a system interacts with its environment, technical efficiency becomes a function of the work, which is a function of the cycle in the exchange of energy. In order to be perfectly clear that thermal efficiency is a function of the cycle, one must avoid misunderstanding another concept in thermodynamics, that of _potential work_, sometimes called optimum work, maximum work, or reversible work. Potential work is a function of the properties of the system and therefore is not a function of a process, nor heat nor work.

Potential work is a measure of energy _availability_ within a specific system. That is, it is a function of the magnitudes of the properties which determine the states of the system, measured between two different states.

> The maximum possible work output that can be produced by a system from a given state to its dead state is what is called, appropriately enough, the _work_ _potential_. The term _availability_ is also used.
>
> It should be noted by students that, for a given environment, work potential is a property of systems. ...it should be clear at this point that the _maximum_ work that can be produced by a system is not a function of the process. The actual work, of course, will be a strong function of the process but the maximum is the maximum regardless of how it is ob-

> tained. Hence work potential is a property. [Dixon, 1975, page 232]

Two things should be evident from this definition of work potential. The first is, that two systems with the same properties have the same work potential. Secondly, given that the second law and entropy account for a "loss" of some of the energy in a system when it changes state, the measure of potential work is less than the total energy in the system.

The definition of potential work suggests that one might measure efficiency by taking the ratio of work potential to the amount of work one actually observes given the operation of one process. This might appear to be a means of evaluating the "technical efficiency" of a system. Unfortunately, in the case of a cycle the measure of work potential equals zero, since in a cycle the system returns to its initial state. This eliminates any differences between the initial and final properties of the system with which to calculate a measure of work potential. Therefore, efficiency is a measure of all the processes which use heat input and produce work output during the cycle.

During a cycle some processes will have positive net work in and some will have negative net work in. If the cycle produces positive net work, it will use positive net heat in a greater amount due to the second law, even if the cycle is reversible. Therefore, the measure of efficiency will be less than 100%. Nevertheless, in the case of a reversible cycle one will have a measure of the maximum efficiency possible. The technical efficiency of a non-reversible (real) cycle will be less, but always the same given the same inputs and the same processes, ceteris paribus.

## A1.4. WHAT IS WRONG WITH FRONTIER PRODUCTION FUNCTIONS FROM THE PERSPECTIVE OF THERMODYNAMICS:

Within the context of thermodynamics thermal efficiency, or technical efficiency, has been defined. Does a comparison of firm P to firm Q, in Figure 2.1, reflect this type of technical efficiency? If it does then one of two situations must exist.

### A1.4.1. THE FIRST SITUATION:

In Figure 2.1, P and Q are clearly supposed to be within the same input requirement set, or sub-production function, so they are using the same cycle. Firms P and Q also use the same homogeneous inputs in the same proportions since they are both on a ray from the origin. Therefore, one can conclude that P and Q have the same properties in their initial states with the only difference being that P is some scalar a times greater in quantity than Q. This corresponds to the situation of getting the same quantity of useful output with different amounts of the same inputs in the same proportions. Since they are not on the same unit isoquant in their end states, assumed to be a dead state, then their end states differ by some factor $\underline{a'} \neq \underline{a}$. That is, if the amount of useful work is the same, then the amounts of waste in the two systems must be different. Specifically, firm P's end state will have a higher value of entropy. This means that the properties in the end states are not the same, so that the output from the two systems are not the same. This situation cannot be due to the processes since the states are by definition not a function of any of the processes. Therefore, the difference in the end states must be due to $\underline{a}$; the properties in P's end state that are different than the corresponding ones in Q must be a function of $\underline{a}$. Thus, $\underline{a}$, the factor of proportionality, would have to be an argument in the equation of state, which it is not in thermodynamics.

**A1.4.2. THE SECOND SITUATION:**

If the initial states of the two systems are the same and the amounts of useful output are different, then one has the situation of two sets of identical inputs producing different output. The difference between P and Q is the amount of actual work done by them. This is not a function of the properties of either their initial states or their end states, as was discussed above, but of the processes each uses. Specifically, if there is a difference in the amounts of actual work, then in keeping with the laws of thermodynamics it must be due to each system using different processes. Using different processes means opoerating with different fixed inputs, i.e., following different cycles or being on different sub-production functions. Comparing the two firms P and Q using a unit isoquant under these circumstances constitutes an error of specifying the same fixed inputs or process when in fact they are different.

## A1.5. SUMMARY

Given identical systems and identical processes, the measures of efficiency will always be identical. One cycle may be less efficient than another cycle given the same system since it is a different path from state to state, thereby producing different levels of net heat in and different levels of net work out as the system's properties assume different values at each point along the path. Similiarly, two systems might differ in technical efficiency using identical processes since the states for each system will differ initially and consequently at each point (state) along the path from initial to final state. Therefore, if one observes two different production situations that differ in efficiency, either the systems are different (non-homogeneous inputs) or the processes differ (different sub-production functions).

Two things should be clear from considering thermal efficiency in thermodynamics. The first is that there is no difference between technical efficiency and price efficiency since efficiency measures the ratio of value of output to the value of input. The second is that "technical efficiency" as used in the frontier production function literature cannot exist. If one uses identical inputs and the identical process, then one must get the identical output. Therefore, if in Figure 2.1, the process used by points Q and P are identical, point P can be less efficient than point Q only if P is producing a <u>different</u> level of output. Alternatively, if P and Q are producing identical output, then they it must be the result of Q and P using <u>different</u> processes. In either case, the laws of thermodynamics makes it clear that the frontier production function distinction between "TE" and "PE" are not valid without violating the basic assumptions of the theory. Therefore, frontier production functions are the result of either specification error, and/or aggregation error.

# APPENDIX TWO

## EFFICIENCY IN ECONOMIC THEORY

To understand much of the frontier production function literature one must be familiar with the "set theory" approach to microeconomic theory. Frontier production functions are sometimes conceived in that literature as "distance functions" displaying "strong disposability," [Kopp, 1981b, Kopp and Diewert, 1982] .

The first section of this chapter will define of terms. The next section will present the usual assumptions made about production sets and the implications of the more important two, concavity, which indirectly suggests the importance of fixed inputs, and monotonicity, which implies free disposal. Fixed inputs are frequently ignored in the frontier production function literature which may explain why aggregation and/or specification error have been identified as "technical efficiency" in frontier production functions. In Chapter Three free disposal and duality theory were used to show that frontier production functions cannot exist if they are distance functions that display strong (free) disposability. The third section correctly defines efficiency in relation to profit maximizing behavior, to clarify the inseparability of the physical and value aspects of production.

Finally, as a result of the first three sections, the explicit inconsistencies of frontier production function theory will be explained.

## A2.1. PRODUCTION THEORY

A production relationship maps inputs to outputs, where input is the range and output the domain. One should bear in mind at the outset that production is by definition the creation of value -- production processes are implicitly normative. Production means creating utility by changing the time, space, or form of commodities. Prehaps for this reason alone, one cannot consider physical (TE) and value (PE) efficiency separately.

### A2.1.1. WHAT ARE INPUTS AND OUTPUTS:

Inputs are the commodities and services with which one starts while outputs are the commodities and services with which one ends a production process even if there is no change in time, form (quality), or space. Thus, "left over" inputs are part of the output. Left over inputs would result from a system that changes state but does not reach its dead state. In this simple analysis, inputs and outputs are assumed to be individually homogeneous (have the same properties), and completely divisible. Additionally, homogeneous means that each input or output identified is defined to be alike with respect to time, place, and form. Therefore, the individual units of a quantity of identical input or output are indistinguishable from each other under any and all circumstances, and in particular, in the way they behave in production. Notice that this means that the processes becomes a function of the system and vice versa. One cannot differentiate between units of the same inputs, except with respect to order in which they are added to production. This means each pound in ten pounds of the input v(j) is exactly like any other one pound of that input v(j) in all its descriptive characteristics. Divisiblity means that the functional nature of the input (output) is independent of the units in which it is measured. Five pounds of the input v(j) added in one pound units will have the same affect on the output as five pounds of v(j) added in ten one half pound units.

The law of diminishing utility indicates that the utility of any particular input (output), will change at the margin as it becomes increasingly scarce or plentiful. Indeed, a good that is not scarce does not have exchange value. This issue of changing marginal value will be side stepped for the purposes of this disseratation by assuming atomistic competition so that inputs and outputs can be treated as having constant prices. In order that there be no confusion, the part of output that has net positive value will be called product, that part that has zero net value will be called waste, and that part that has negative net value will be called pollutant. Either utils or dollars are treated as being adequate common denominators of value.

**A2.1.2. PRODUCTION SETS:**

Output is produced by transforming inputs. From the laws of thermodynamics it is clear that there is a single valued relationship between the quantities of inputs transformed and the quantity of product generated, given specific processes. That is, using given processes and specific quantities of homogeneous input, only one quantity of product will result, ceteris paribus. Because of this relationship the terms output and product can be used interchangeably in most situations. For a particular producing unit, or firm, there is some finite set of production possibilities, that is described by the collection of all possibile input bundle combinations and the quantities of product that result from their transformation. A vector of input and output quantities can be variously called a production plan, activity vector, or netput.

> Suppose the firm has n possible goods to serve as inputs and/or outputs. We can represent a specific production plan by a vector y in $R^n$ [the positive quadrant of euclidean hyperspace] where y(i) is negative if the $i^{th}$ good serves as a net input and positive if the $i^{th}$ good serves as a net output. Such a vector is called a netput vector. The set of all feasible production plans - netput vectors - is called the firm's production possibilities set and will be denoted by Y, a subset of $R^n$. [Varian, 1978, page 3]

A2.1.2.1 PRODUCTION POSSIBILITIES SETS:

A production possibilities set is the collection of all feasible input/output vectors for some output vector Y. As such, it represents the technology for Y. The production possibilities set is:

$$\text{(A2.1)} \quad Y = (y(1),\ y(2),\ y(3),\ \ldots,\ y(M),\ V(1),\ V(2),\ V(3),\ldots,\ X(N),\ Z(1),\ Z(2),\ Z(3),\ \ldots,\ Z(P),\ U(1),\ U(2),\ U(3),\ldots,\ U(Q))$$

where

y(I) = PRODUCTS FOR I=1 TO M

V(J) = v(J,I) FOR INPUTS J=1 TO N AND PRODUCTS I=1 TO M

Z(K) = z(K,I) FOR INPUTS K=1 TO P AND PRODUCTS I=1 TO M

U(L) = u(L,I) FOR INPUTS L=1 TO Q AND PRODUCTS I=1 TO M

It is worth reiterating that this set is a feasible set, or set of those production plans that are physically possible, e.g., the set of netput vectors that conform to the laws of thermodynamics.

A2.1.2.2. TECHNICAL CHANGE:

Usually, technical change means that the original production possibilities set has been expanded by adding new input:output vectors, or by adding a dimension to the existing vectors. This is equivalent to adding a previously unknown input, or unknown way of combining known inputs (processes), to create new vectors, or a new dimension, to the production possibilities set. Note that this new input, or new way of combining old inputs, can be used in various amounts, which is why the change adds more than one vector to the set.

If this new input results in the possibility of a given amount of product y being produced at a lower cost, then one can conclude that the new production possibilities set is more efficient. To ignore the new activity bundles (to operate only with the opportunities of the old production set) would indeed be inefficient. Efficiency is achieved by making an economic adjustment resulting from comparing the cost of using the new input bundles as opposed to any of the original bundles. Strictly speaking, this means not only their cost in operation, but also the cost of switching from one set to the other by investing or disinvesting. If adopting or using the new input results in a lower net value from production, the new technology is less efficient than the old technology. Determining efficiency is a question of evaluating cost; the ratio of the value of useful work to the cost of the input.

Exactly what causes technical change and how technical change is accomplished are two complex issues that are beyond the scope of this dissertation.

### A2.1.2.3. PRODUCTION FUNCTIONS:

A production possibilities set may contain one or more <u>production functions</u> or processes. The production function specifies a functional relationship between inputs and an output such that for any given vector of inputs there is one, and only one, vector of output. That is, one cannot have two netputs in the same production function where:

$$(A2.2)\quad y^*(I) <> y(I)$$

and

$$(A2.3)\quad (V^*(J),\ Z^*(K),\ U^*(L)) = (V(J),\ Z(K),\ U(L))$$

In addition there is a distinction made between the groups of inputs in that the V(J)'s are *variable inputs*,the Z(K)'s are *fixed inputs*, and the U(L)'s are *random variable inputs*. These distinctions are critical.

A2.1.2.3.1. SUB-PRODUCTION FUNCTIONS:

A *sub-production function* is a "restricted" subset of the production possibilities set. It is restricted in the sense that there is only one product, y(I), and some subset of the inputs remain fixed or constant over the entire range of production possibilities. The sub-production function defines the technical relationship between inputs and output. It is one *technique* within the technology for Y. It defines the physical transformation of inputs into output. It is the totality of the processes that act on the inputs, or system, resulting in useful work, or output.

$$\text{(A2.4)} \quad f(y) = [f(v(1),\ v(2),\ v(3),\ldots,\ v(N),\ z(1),\ z(2),\ z(3),\ \ldots,\ z(P),\ u(1),\ u(2),\ u(3),\ \ldots,\ u(Q)) \mid (y,\ v,\ z,\ u) \in Y]$$

where

y = VECTOR FOR A PRODUCT I

v(J) = VARIABLE INPUTS FOR J=1 TO N FOR PRODUCT I

z(K) = FIXED INPUTS FOR K=1 TO P FOR PRODUCT I

u(L) = RANDOM VARIABLE INPUTS FOR L=1 TO Q FOR PRODUCT I

Variable inputs, v(J), are inputs over which the manager has control, and for which the manager may vary the quantities of the input in the sub-production processes within one production period. Variable inputs will be varied in the amounts used within the production processes as a result of assessing their costs, relative to the value of production within the firm, and relative to the value of production outside the firm. Their acquisition prices and their salvage values are always equal. Their within firm opportunity cost is the same as their out of firm opportunity cost. Within firm opportunity cost changes acquisition price and their salvage values and all three remain equal. Therefore, when product or variable input prices change, the amount used will necessarily be adjusted upward or downward in order to maintain efficiency.

Fixed inputs, z(K), are those inputs over which the manager may have control, but which are not varied in the amounts used within the production processes. They are defined for commodities and inputs for which acquisition costs exceed salvage values. The amount of a fixed resource acquired by the firm initially is to be determined after assessing its value in production. The fixed resource will be acquired up to the point that its marginal acquisition cost equals its value in production both expressed as stocks or services. Once acquired by the firm, its quantity is fixed in production so long as its within firm opportunity cost, or shadow price, is bounded by its acquisition and salvage value. That is, having been acquired, its acquisition price is greater than its salvage price. Therefore, a change in the fixed input's acqusition cost or salvage value, the latter reflecting out of firm opportunity costs, will not necessarily lead to an adjustment in the amount used, investment or disinvestment, by the firm. That is, the input is fixed in production.

Some fixed inputs are specialized in the sense that they have no within firm opportunity cost, i.e., they cannot be used to produce more than one product. The number of products they can produce is one, the Ith. If one considers unspecialized inputs capable of contributing to the production of more than one product, then such an input may be fixed to the firm, but not within the sub-production function for one product, since its amount might be varied between two, or more, sub-production functions for the multiple products it can produce in the firm. It should be noted that the fixity of inputs can be made endogeneous [Edwards, 1958] without changing the conclusions to be drawn regarding either the contribution of the fixed input to the production processes or to the definition of efficiency.

Thus, the optimal amount of an input whose acquisition price equals its salvage price always changes with its one price. However, the optimal amounts of inputs whose acqusition costs exceed their salvage values do not always react to changes either in their acquisition cost or salvage value. Indeed, if the within firm opportunity cost is between the input aquisition and salvage prices expressed as flow prices then the optimal amount to use will not vary and the input is fixed.

Random variable inputs, u(L), are inputs over which the manager has no control, and whose quantities (or quantities) vary randomly among firms and for individual firms from some normal, or average specification. This average quantity (or quantity) is usually the first moment of the distribution from which the quantities of u(L) are drawn. This average quantity is fixed. Examples would be inches of rain fall, number of hours of sun light per day, quantity or land, labor of capital; or air temperature. Since the random variable inputs are usually measured as deviations from their averages in any given production period the expected value so measured is equal to zero. Since the random variables are outside the control of the manager, and because their expected value is zero, the average value from which they deviate is fixed. This assures "average conformity" with the laws of thermodynamics. Unless otherwise noted, fixed inputs and the average value of random variable inputs are treated in the same manner.

The sub-production function does three things to a subset of the production possibilities set; it (1) holds the quantites and qualities of a subset of the inputs, the z(K)'s, at a constant level, absolutely or on the averages, for all the input vectors in the set, and (2) fixes the distribution from which the random variable inputs, the u(L)'s, are drawn, and (3) fixes the functional relationship or processes, the f(....), between the inputs and the output. Notice that this means that the sub-production function changes if the quantities or qualities of the fixed inputs change, or the average quantities or qualities of the random variable inputs change. Note that the levels of the fixed inputs are held constant; this means as one increases variable inputs, not necessarily in proportion, the constraining influence of a fixed input or an average random variable input may change.

A2.1.2.3.2. INPUT REQUIREMENT SETS:

A collection of all input vectors capable of producing <u>at least</u> some given level of y is called the <u>input requirement set</u>, V(y) [Varian, 1978].

(A2.5) $V(y) = [(v, z) \mid (y, v, z) \in Y \text{ and } y <= y']$

That is, if (v, z) is in V(y) and (v', z) >= (v, z) then (v', z) is in V(y), but (v, z) V(y'). Notice that this means that a sub-production plan (v', z) that is in V(y) will produce y', where y' > y, and still be in y's input requirement set. Notice too, that the z's are fixed in identical quantities for all the sub-production plans in the set V(y); if the quantity of one or more z's changes, one implicitly changes input requirement sets and sub-production functions.

A2.1.2.3.3. ISOQUANTS:

The collection of sub-production plans that produce exactly y are called <u>isoquants</u>, Q(y).

(A2.6) $Q(y) = (v, z) \mid (v, z) \quad V(y) \text{ but } (v, z) \notin V(y')$

where

(A2.7) $y' > y$

All of the input vectors (v', z) > (v, z) such that (v', z) $\in$ V(y) are not members of Q(y). This definition of isoquants should make it clear that input requirement sets are conceived to be, basically, a collection of isoquants. That is, that the input requirement set V(y) is the isoquant Q(y), and all the isoquants Q(y'), where y' > y. That means that Q(y) acts like a "frontier" in V(y), within which all the higher valued isoquants lie. Notice that this means a point can be an 'interior' point to an isoquant and still be in the same input requirement set as the isoquant. These 'interior' points are not on the isoquant, because they produce $y' > y$, but that they are members of the same input requirement set, $V(y)$. It is especially important to appreciate that these 'interior' points are fundamentally different than those of a unit isoquant or frontier production function (see Chapter Two).

A2.1.2.4. DISTANCE FUNCTIONS:

McFadden uses the concept of free disposal (discussed in a following section) to expand, modify, or generalize, the concept of a production function to what he terms a <u>distance function</u>.

> 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 [production] 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. [McFadden, 1978, page 24]

Formally, the definition of a distance function is:

(A2.8) $F(y,v, z) = \text{Max } (a > 0 \mid (1/a * (v, z) \in V(y))$

This is essentially an application of the implicit-function theorem [Chiang, 1974]. One should note that implicit-functions of mathematics do not correspond with real world production processes. Implicit-functions are, in this respect, similiar to reversible processes in thermodynamics; neither can exist in reality due to the second law.

Figure A2.1 represents a distance function.

> As illustrated in [Figure A2.1], the value of F(y,v, z) is given by the ratio of the length of the vector (v, z) to the length of a vector (v*, z) defined by the intersection of the "y-isoquant" and the ray through (v, z). [McFadden, 1978, page 25; z's added]

Note what this does: "a," in (A2.8), is an adjustment or scaling factor that reduces all input bundles on the isoquants greater than y* (in the interior of y*) to values equal to the bundles <u>on</u> the isoquant y*. Clearly, "a" needs to be a vector, with 1's corresponding to the z's so the z's are unchanged in value. Only that portion of the input bundle (v, z) equal to (v*, z) remains after scaling, where (v, z) > (v*, z); i.e., (v, z)/a = (v*, z). Note: this does not mean that (v, z) is getting less output from the same inputs as (v*, z), since both the inputs and the corresponding output are adjusted for (v, z). In effect, the distance function finds a scaling factor that takes any input bundle within the 'interior' of the lowest isoquant of an input requirement set and moves it back to that lowest isoquant; it transforms the

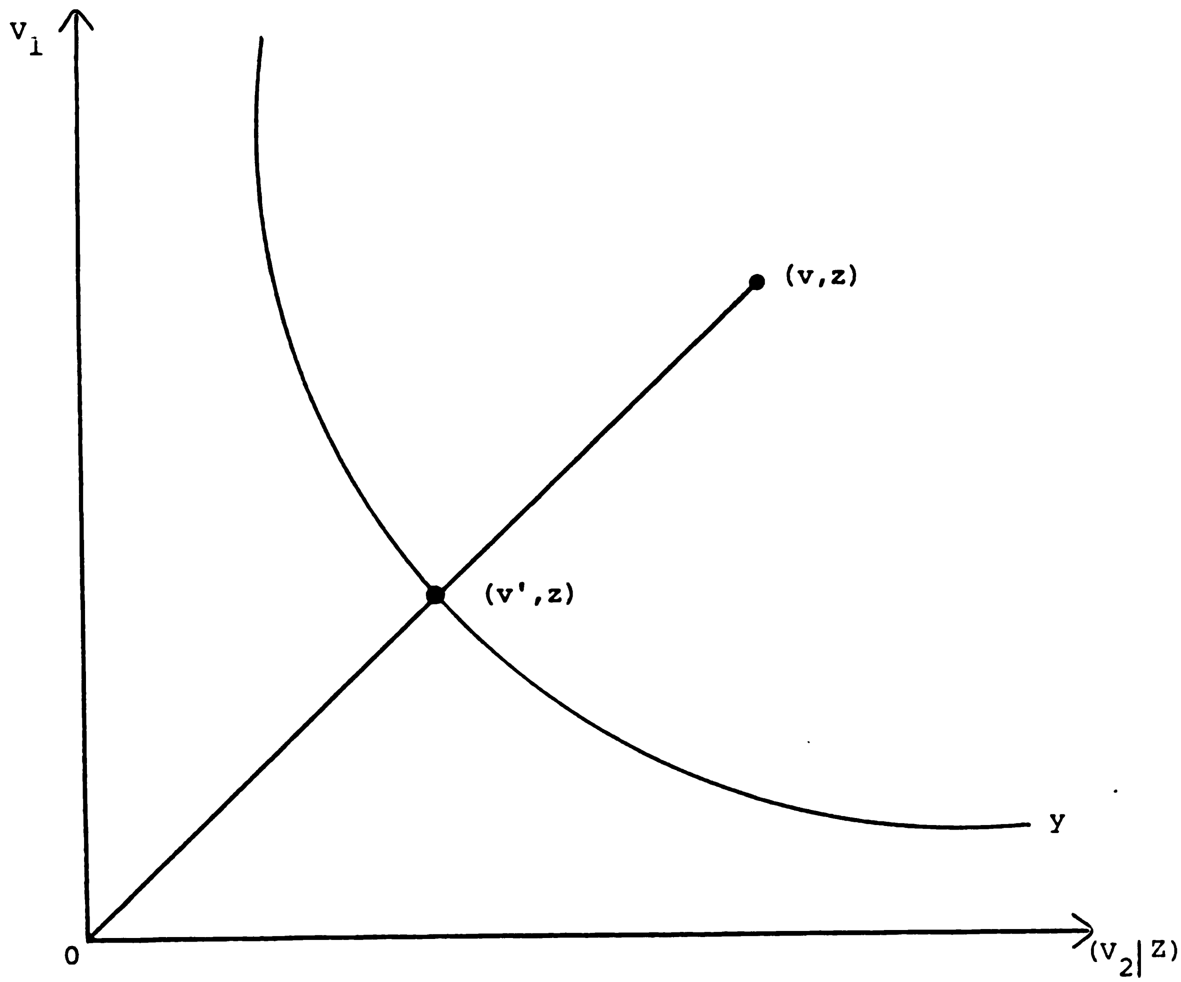



FIGURE A2.1

A DISTANCE FUNCTION

'interior' so that the 'interior' lies on its boundary. The physical quantities of "excess" input and the corresponding "excess" output, are entirely and completely removed and eliminated from consideration. It eliminates the 'interior' for all practical purposes.

There are three facts which must be noted. (1) Distance functions are made possible, conceptually, with free disposal. Consequently, the concept of free disposal is very important and therefore will be dealt with in a following section. (2) The distance function eliminates any 'interior' points within the lowest isoquant of an input requirement set while retaining the input requirement set's mathematical properties of convexity. The convexity conditions are necessary for McFadden to develop a concise mathematical proof to duality theory. These conditions would also be met in a monotonic production function, or in an analysis confined to Stage II (the rational area) of a traditional production function, as contrasted to a production solid. Thus free disposal appears necessary to deal with the 'interior' points of production solids. (3) McFadden's (1978) original formulation of distance functions did not include fixed inputs, the duality theory they are used to prove is valid only for production sets that exclude Stage III of production.

## A2.2. ASSUMPTIONS ABOUT PRODUCTION SETS

Theories of production often postulate three fundamental assumptions. (1) The first is that production is regular or input regular. The set of all possible production plans for some output or group of outputs is non-empty (you cannot get positive output from zero inputs), and closed (there is some bounded and describable set of feasible production plans). This is the standard economic assumption that there is "no free lunch." (2) production sets are concave functions, which means one can obtain input requirement sets that are convex sets. Input requirement sets are convex sets because the lowest isoquant within the set is a "boundary" to all the higher isoquants. Isoquants are not convex sets; they are convex functions (See Chiang, 1974). The distinction is critical since it is within the context of convex sets rather than convex functions that one may provide a proof of duality [McFadden, 1978]. In practice, this assumption is always further restricted so that input requirement sets are strictly convex sets and isoquants are strict convex functions. This assures that the isoquants are "well behaved" (no flat spots). This means, in effect, that perfect substitutes and perfect complements are ruled out or are combined into single inputs. Additionally, Malinvaud (1972) points out that convexity implies that the bundles of resources used to produce any given level of output display conditions of additivity and divisibility. This means changes in input quantites can be considered in infinitesimal amounts. This in turn suggests that sub-production plans would necessarily display constant returns to scale except for the presence of some fixed resource [Malinvaud, 1972]. This is why the consequences of fixed inputs in production is very important and will be dealt with in the following section. (3) Production displays monotonicity among the variable inputs; that is, additional variable inputs will yield additional output(s) within the constraint of the fixed input(s). This assumption assures that isoquants will not converge. This means points of equi-production are on the same isoquant. This assumption also implies free disposal among the variable inputs.

While many set theoretically inclined microeconomic theorists note that only the assumption that production is regular is essential for most of the important results of microeconomic theory, concavity (assumption two) and monotonicity (assumption three) are critical to the formal proofs of duality theory, which receives major emphasis in the more abstract forms of production theory currently taught in general economics departments (for example see Varian 1978).

Duality simply means that every point in production space (excluding Stage III) is associated, by a one to one mapping, with a corresponding point in cost space (polar reciprocal sets). Since value and technical relationships are inseparable, this one to one correspondence is intuititively obvious. This means the apparent difference between physical quantity and price is nothing more than a mathematical transformation, a difference in the units of measure. Input requirement sets that conform to these three assumptions are called input-conventional input requirement sets.

**A2.2.1. CONSEQUENCES OF FIXED INPUTS:**

Fixed inputs play a very important role since they act as constraints on the amount of total product one can achieve from adding more variable inputs. They determine the sub-production function one is using. Thus, the fixed inputs determine which technical relationships will exist among the factors of production and output, i.e., which sub-production function is relevant and, hence, the marginal physical products (MPP) for the various variable inputs.

One cannot ignore fixed inputs, especially in empirical work, because without some input fixed in the production processes, constant returns to scale would apply and one could continuously increase (variable) inputs in fixed proportions and obtain proportionate increases in physical output. In particular, the law of diminishing returns (or law of variable proportions) would not apply and marginal physical products (if not marginal value products) would always be positive and constant. The second partial derivatives of the production function are negative in all of

Stage II and part of Stage I, which means that marginal physical products are decreasing, given that all the other inputs become fixed. The absence of the law of diminishing returns, and marginal physical products that can be negative, result in an incomplete production function, a production function without Stage I and III, so that production can never reach a maximum. The three topics of constant returns to scale, the law of diminishing returns (or law of variable proportions), and the stages of production, will be reviewed in turn, demonstrating how they are important to a definition of efficiency.

A2.2.1.1. CONSTANT RETURNS TO SCALE:

Constant returns to scale means that there will be proportionate increases in output resulting from proportionate increases in all the inputs. Therefore, in theory, constant returns to scale will exist when there are no fixed inputs. The relationship between height, width, volume and mass may also make strict constant returns to scale physically impossible in reality. Though this relationship does not fix any inputs, but it does prevent expansion of all inputs in constant proportions.

Under constant returns to scale (no fixed inputs), and with prices constant, production will never reach a maximum, so all production levels are equally efficient. One can discuss differences in efficiency only within one of two sets of circumstances; (1) where output is predetermined, or fixed, or (2) where the total amount of resources available, the budget constraint, is predetermined, or fixed. In the first case, the level and combinations of variable inputs that achieve minimum cost for producing the given level of output are efficient. In the second case, the set of given resources determines the input requirement set, and the maximum output possible.

With constant returns to scale there is only one sub-production function, since all inputs are variable. Generally, one sub-production function may be less efficient than another, i.e., unable to produce as many units of value per unit of value consumed in the production processes as another sub-production function. When one considers changing sub-production functions, one is asking whether or not is would pay to vary some hitherto fixed input. The economics of investment and disinvestment was developed by Edwards (1958). In order to understand why one sub-production function might be more efficient than another, one must understand the law of diminishing returns.

## A2.2.1.2. THE LAW OF DIMINISHING RETURNS:

The law of diminishing returns (or law of variable proportions) states that as one adds to production successively more equal units of a given variable input while holding one or more other inputs fixed, the marginal physical product of the variable input first increases at an increasing rate, then increases at a decreasing rate, then decreases absolutely. The law of diminishing returns presumes the existence of fixed inputs. In the case above, under constant returns to scale, efficiency was determinable only after something other than inputs became fixed, because otherwise marginal physical products remained constant and equal to average physical product. The laws of diminishing returns, and/or diminishing utility, influence efficiency since the former affect the marginal physical products and the latter the value of the product.

Because of the laws of thermodynamics (see Appendix One) if one has identical processes using identical variable inputs and the identical quantity of identical fixed inputs, the marginal physical products will be identical for every identical marginal change in the variable inputs. Consequently, if one observes two production processes using identical quantities of identical variable inputs to achieve different quantities of identical output(s), then the marginal physical products for some input(s) must be different. This can only be true if the amount of fixed input(s) is different for the two sets of processes. That is, the sub-production, or processes, with the lower level of fixed input(s) constrains the marginal physical product for some variable input(s) and thus constrains output.

A2.2.1.3. STAGE III OF PRODUCTION:

One last consequence to the inclusion of fixed inputs in sub-production function is Stage III, a region where product declines after reaching a maximum. Stage III is particularly important since many contemporary theorist maintain:

> A production plan y in Y is called <u>efficient</u> if there is no y' in Y such that y' >= y; that is, a production plan is efficient if there is no way to produce more [product] with the same inputs or to produce the same output with less inputs. [Varian, 1978, page 4]

One can produce more useful product with less input if one is in Stage III, i.e., by moving from point D to point B in Figure A2.2, but staying on the same isoquant. The monotonicity assumption means that more variable input always produces more output, but in Stage III the additional output is waste or pollutant. For example, water is a necessary input to crop production but too much water results in reduced product. What happens in Stage III is that the additional variable input(s) begins to block the ability of some other input to contribute to the production process thereby reducing useful output. When too much water is added to crops, the roots are unable to get as much oxygen as previously and this reduces useful output. This is anexample of the law of diminishing returns since one input is blocking the other(s) from contributing, e.g., due to the lack of space for both to

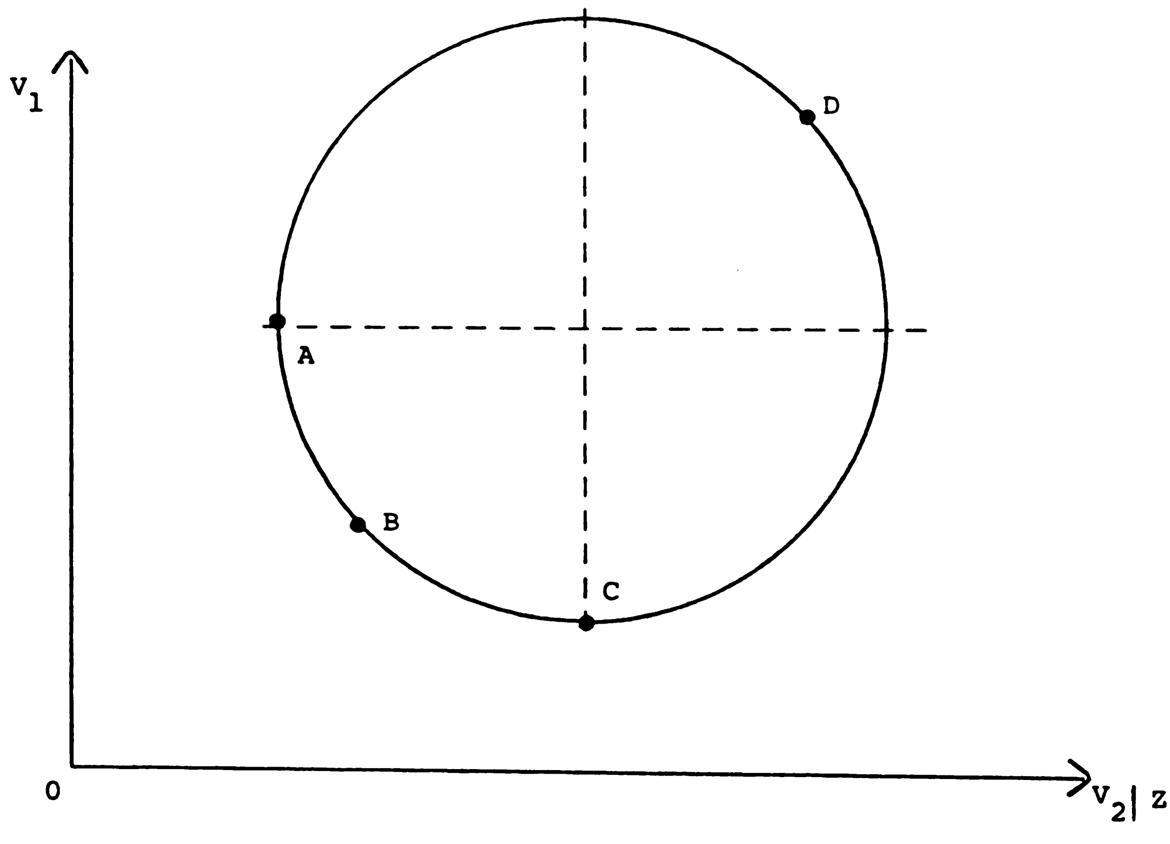



FIGURE A2.2

ISOQUANT SHOWING STAGE III

operate effectively. One should note that this is not a change in sub-production function since the levels of the fixed inputs remains constant. Ignoring Stage III means the production possibilities set is incomplete; that it does not contain all the feasible production plans.

In production theory, Stage III is often precluded or avoided since it is outside the range of "rational" production. In Figure A2.2, ABCDA is the full isoquant for some level of production. Usually, only ABC is considered rational since the portion ADC is in Stage III. If one where to observe a point D, one could conclude that one could achieve the same level of output using less of the same variable inputs, and conclude that point D was a point of "technical inefficiency;" that point D was in the interior of the isoquant ABC. Clearly, D is not in the interior of the isoquant but is on the isoquant and is inefficient for exactly the reason that Stage III is outside the range of rational production; net value of the output is not at a maximum.

**A2.2.2. CONSEQUENCES OF MONOTONICITY:**

A2.2.2.1. FREE DISPOSAL:

Free disposal arises as a consequence of the assumption of monotonicity [Varian, 1978]. Monotonicity implies that if a set of resources (A) is greater than another similiar set of resources (B) then the former set (A) can produce at least as much output(s) as the latter set(B).

> The idea is clear: if we can produce y with a certain input bundle v, we should be able to produce y if we have more of everything. This is sometimes referred to as the hypothesis of "free disposal." For if we can always costlessly dispose of anything we don't want, our technology must certainly satisfy the monotonicity assumption. [Varian, 1978, page 6]

The "anything we don't want" is both input and output. Due to the laws of thermodynamics and the one to one relationship between input and output, it is especially important to remember that despite the fact that input requirement sets are defined in terms of inequalities, in reality one is dealing with sets of equalities. Free disposal is the means of reconciling the appearance of producing exactly y with the input bundle (v', z) included in V(y), when (v',z) in fact produces exactly y', y < y'. The input requirement set for a particular isoquant includes the input bundles for that isoquant and all the input bundles for all the isoquants at higher levels of production. The input requirement sets for higher levels of production are proper subsets of the input requirement sets of lower levels of production [McFadden, 1978]. It means that all of the isoquants "within" a given isoquant y are within the input requirement set for that given isoquant y. Free disposal does not require that the production function is a solid instead of the surface one ordinarily associates with a production function. The concept of free disposal does not mean that an isoquant has interior points that are at the same level of output. Suggesting this is a misunderstanding of what constitutes free disposal, i.e. a violation of monotonicity.

The notion of free disposal is often misunderstood which is to be expected as it has no counterpart in the real world. The nearest exception to this is approached as two inputs approach perfect complemenatary. McFadden (1978) suggests that free disposal is essentially a gimmick to provide the conditions for the derivative conditions necessary to provide a rigorous mathematical treatment of the theory.

> However, the importance of [free disposal] in traditional production analysis lies in [its] analytical convenience rather than in [its] economic realism; [it] provide[s] the groundwork for application of calculus tools to the firm's cost minimization problem. [McFadden, 1978, pages 8 & 9]

Free disposal means one may "throw away' commodities without using up inputs in the disposal process [Pachico, 1980]." This does not mean using up <u>additional</u> inputs, instead it means that free disposal removes some amount of input, and its counterpart in output, from production. The freely disposed input/output is in no sense part of the production processes. When the additional output y' - y is freely disposed, the additional resources (v', z) - (v, z) are also disposed. One must remember that inputs cause output by being <u>consumed</u> in the processes. Freely disposed inputs are not consumed, which is what Varian (1978) means by "costlessly disposed." Because an input is never consumed it is never paid for nor does it have an opportunity cost. Proper accounting in production will only record those inputs consumed in production.

Free disposal means that one can move from a higher isoquant to a lower isoquant without cost, by the free disposal of the additional output. It also means that 'interior' points of production functions can be swept out. The inputs that created the output thrown are also treated as costless in polar reciprocal sets. In the definition or a distance function, the scaling factor, a, is the mechanical means of performing "free disposal." This scaling factor is needed in Figure A2.1 to transform (v, z) so that (v, z) will produce exactly y*, and not y, where y > y*, in keeping with the laws of thermodynamics. Those excess amounts of the inputs in the input bundle (v, z) are freely disposed, otherwise they would create output in excess of y* if used in production. The output is reduced precisely because the quantities of inputs used are reduced.

### A2.2.3. DUALITY THEORY AND POLAR RECIPROCAL SETS:

The duality theory proved by McFadden (1978) excludes Stage III of production since he assumes that all marginal physical products must be non-negative. Duality means that all the points in a input-conventional input requirement set have a unique correspondence to points in the associated cost space. Therefore, in duality theory the input-conventional input requirement sets of distance functions have an analogous counterpart in cost functions, called <u>factor-price requirement sets</u>, R(y) [McFadden, 1978]. Factor-price requirement sets are defined as:

(A2.9) R(y) = [r >= 0 | r * (v, z) >= F(y, v, z)
for all positive (v, z)]

where

r = VECTOR OF PRICES FOR ALL INPUTS

This means that the prices in the factor-price requirement set satisfy the condition that when multiplied by the input bundles in the corresponding input requirement set, the inner product is at least as large as the value of the relevant distance function F(y, v, z), which was defined earlier. This property means that that not only do input-conventional input requirement sets and factor-price requirement sets have a unique one-to-one mapping from one set into the other, but <u>both display the property of free disposal</u>. The logical consequence of this is that for any point within the "interior" of an isoquant, there is a mapping of this point into the "interior" of a isocost; i.e., if one is on a higher isoquant, one is producing at a greater cost. Inputs have an opportunity cost. If by invoking free disposal one ignores the additional output created by additional inputs in production space, then <u>one must necessarily ignore the additional cost for those</u>

inputs in cost space. If one can move from a higher isoquant to a lower isoquant by free disposal, then one is necessarily moving from a higher isocost to a lower isocost in cost space, by the same free disposal.

### A2.3. MAXIMIZING BEHAVIOR:

Maximizing behavior defines efficient behavior. It involves both technical (physical) and price (value) information. The most efficient production is seldom the maximum average production. Efficiency deals with the question of relative costs; inputs are used efficiently when they are used in least cost combination. This is obvious from the decision rule equating marginal cost with marginal revenue. This is identical to equating the ratios of marginal physical products to the ratios of the respective prices for all the inputs, where the marginal physical products represent the technical aspect of production and the prices represent the opportunity cost aspect. Some of the prices are internal opportunity costs for unspecialized inputs or "shadow prices" for specialized fixed inputs [Edwards, 1958]. Notice that this is identical to the definition of efficiency used in thermodynamics, i.e, the ratio of useful output to costly input [Dixon, 1975].

$$\text{(A2.10)}\quad MPP(v_1)/P(v_1) = MPP(v_2)/P(v_2) = \ldots\ldots = MPP(v_n)/P(v_n)$$

Take the case of two inputs with different marginal physical products. One input is necessarily being used inefficiently when compared to the other only if the two inputs have the same price or opportunity cost. One cannot allocate resources using marginal physical products alone or prices alone. If technology could be separated from value, it would seem reasonable to expect that there would be a rule for maximizing efficiency by equating marginal physical products without reference to their values and vice versa. This would suggest that one could allocate resources solely on technical or price criteria. What those criteria might be is unclear.

### A2.4. WHAT IS WRONG WITH FRONTIER PRODUCTION FUNCTIONS FROM THE PERSPECTIVE OF ECONOMIC THEORY

First, the analysis of frontier production functions deal with averages. Recall Figure 2.1. The unit isoquant represents average input per unit output. Bressler (1966) notes that the price line represents "average cost." Both technical and price efficiency are found by using average input and average cost rather than marginal input and marginal cost. This surely conflicts with conventional production theory wherein one equates marginal cost to marginal revenue to find the most efficient point of production. Indeed, only in the context of perfect competition is it true that average revenue equals marginal revenue. Only in the case of constant returns to scale and fixed input prices is it true that average cost equals marginal cost over the whole range of output. Consequently, the only case imaginable where it might be legitimate to define efficiency in terms of averages would be for constant returns to scale under perfect competition, a very restrictive and unrealistic case.

Even in the special case of constant returns to scale in perfectly competitive equilibrium, a firm P would not be within the interior of the unit isoquant. Firm P uses more of the variable inputs given the same amount of fixed input than a firm on the unit isoquant. The assumption of monotonicity means there must be more output. Constant returns to scale means that for additional input there is proportionate increases in output. So, in terms of the output per unit input, P must lie on the unit isoquant, or violate constant returns to scale, or use a different amount of the fixed input. Recall that with any fixed input, constant returns to scale only exists only within an infinitely small neighborhood around the point on the production function where average product equals marginal product for all inputs.

Assume for a moment that one observes a firm Q and a firm P both using the same identical sub-production function. If P is using more variable inputs then it is in fact producing more output. If P is getting less output per unit input than Q, it is because P is on a higher isoquant where the marginal physical products for the variable input are smaller due to the law of diminishing returns. In this case, the higher output of P is obscured by the averaging process in finding the unit isoquant. It is also true that P may be inefficient. P might be a dairy farmer who continues to feed his cows beyond the point where the return for the additional milk produced is greater than or equal to the cost of the additional feed. The inefficiency is due to value as well as physical considerations.

If P is producing more output than Q, can one invoke free disposal to make a comparison of the two firms, and the inputs they use to achieve the same output, and conclude that P is less efficient? No, because if one ignores the additional output then by duality one must ignore the cost of the additional output. By duality, free disposal in production space must be associated with free disposal in cost space. Therefore, if output is freely disposed in production space, the cost for the inputs that produced that output must be freely disposed in cost space, which makes those inputs free goods, and economically irrelevant. If one ignores the cost of the additional output, then by duality one must ignore the additional inputs. That is, if one includes inputs in the production accounting, then in order to make the production "ledger" balance one must also take account of the output produced by those inputs. Conversely, if the output is ignored by free disposal, then the corresponding input must also be ignored, or freely disposed. The inputs are irrelevant technically and economically because one has freely disposed of their output. In order to maintain the 1:1 ratio of energy input to energy output demanded by the law of thermodynamics, if one erases an amount from one side of the equation (output), one must erase an equal amount from the other side of the equation (input). If one were to maintain that in freely disposing of the output one were converting it from work output to waste output, then one is implicitly changing sub-production functions, or processes, ceteris paribus.

Finally, the quotation:

> But [P] uses more of both inputs than [Q] to produce the same level of output. [Timmer, 1971, page 777]

exemplifies an error repeatedly asserted in the frontier production function literature; namely that Q produces the same amount of output as P with fewer inputs. That P might actually be producing more output, which is freely disposed, has been discussed above. Suppose that P and Q are in fact producing the same amount of the same output. Figure 2.1 suggests that Q uses fewer inputs. This indicates an inherent indexing problem since in this case the Z3 for Q is not the same as the Z3 for P. That is, Q uses fewer <u>variable inputs</u>, but can only get more output at the margin with them if Q has <u>more fixed</u> input, due to the laws of thermodynamics and the law of diminishing returns to scale. One can compare the "size" of the two input bundles by <u>evaluating</u> them with respect to their opportunity costs. This would indicate that economic efficiency is identical to "price" efficiency, and that the frontier production function distinction between "price" efficiency and "technical" efficiency is meaningless.

## A2.5. SUMMARY

In economics, as in thermodynamics, it is impossible to separate the physical aspect of production from the price aspect of production. Efficiency is defined with respect to the relationship between the two; i.e., efficiency means equating marginal value products with their respective opportunity costs, or useful output to costly input.

In production theory the role that fixed inputs play in determining the technical relationships between all the inputs in the production process is critical. Indeed, the level of fixed input determines the sub-production function. Because some inputs are fixed, the law of diminishing returns operates, leading to variable returns for different levels of input, both variable input and fixed input. Constant returns to scale is a special case where no input is fixed, and where, consequently, there is nothing endogeneous to the production system that affects efficiency unless prices become functions of quantities. Stage III may result from fixing input and means that MPP >=< 0. This implies that efficiency is associated with the location on the isoquant at which one is producing. It also raises the question of whether or not isoquants can have "interiors."

Free disposal deals with the issue of the 'interior' to an isoquant. Free disposal is a necessary in order to understand that input requirement sets are a collection of upper level isoquants for "one" level of production: Clearly a paradox. Free disposal is often not clearly understood, since it has no observable counterpart in the real world. It serves only to reconcile the paradox that an input requirement set implies being on and off the isoquant at the same time. This reconciliation allows one to define a distance function and give a rigorous mathematical proof to duality (excluding Stage III). Duality is demonstrated by polar reciprocal sets. Duality means that the physical aspects of production, marginal physical products, are inseparable from the value aspects of production, prices, through their one to one mapping from input space (excluding Stage III) to cost space.

Since the physical aspect of production is inseparable from the price aspect of production, the notion of "technical inefficiency," as used in frontier production functions, has no logical basis. Since an isoquant does not have a <u>true</u> interior at the same level of ouput there is no theoretical basis for the definition of "TE" as it is used in the frontier production function literature. The definition of a frontier production function is a violation of the tenets of microeconmic theory.

# BIBLIOGRAPHY

# BIBLIOGRAPHY


Aigner, D.J. and S.F. Chu,"On Estimating The Industry Production Function,"American Economic Review, 58 (1968): 826-839

Bagi, Faqir Singh and Cliff J. Huang, "Estimating Production Technical Efficiency for Individual Farms in Tennessee," Canadian Journal of Agricultural Economics, 31(1983): 249-256

Boles, James N.,"Efficiency Squared - Efficient Computation of Efficiency Indexes," Proceedings of the Western Farm Economics Association Meeting 1966, Western Farm Economics Association, pp.137-142

Boles, James N.,The Measurement of Productive Efficiency: The Farrell Approach, University of California, Berkeley

Boulding, Kenneth E., Evolutionary Economics, Sage Publications, Beverly Hills, 1981

Bressler, R. G.,"The Measurement of Productive Efficiency, "Proceedings of the Western Farm Economics Association Meeting 1966, Western Farm Economics Association, pp.129-136

Bravo-Ureta, Boris E.,"Non-Parametric Measures Of Technical Efficiency In Milk Production," Journal of the Northeastern Agricultural Economics Council, 11 (1983): 69-76

Charnes, A., W.W. Cooper, and E. Rhodes,"Measuring the Efficiency of Decision Making Units," European Journal Of Operational Research, 2 (1978): 429-444

Charnes A., W.W. Cooper, and E. Rhodes,"Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis To Program Follow Through," Management Science, 27 (1981): 668-697

Chiang, Alpha C., Fundamental Methods of Mathematical Economics, Second Edition, McGraw-Hill Book Company, New York, 1974

Commoner, Barry, The Poverty of Power, Bantam Books, New York, 1976

Deaton, Angus, and John Muellbauer, Economics and Consumer Behavior, Cambridge University Press, U.K., 1980

Debreu, G., "The Coefficient Of Resource Utilization," Econometrica, 19 (1951): 273-292

Diewert, W.E.,"Capital and the Theory of Productivity Measurement," American Economic Review, 70 (1980): 260-267

Dixon, John R., Thermodynamics I: an introduction to energy, Prentice-Hall, Englewood Cliffs, N.J., 1975

Edwards, Clark, Resource Fixity, Credit Availability And Agricultural Organization, unpublished Ph.D. Thesis, Agricultural Economics, Michigan State University, 1958

Farrell, M.J.,"The Measurment of Productive Efficiency," Journal of the Royal Statistical Society, Series A, 120 (1957): 253-281

Farrell, M.J. and M. Fieldhouse,"Estimating Efficient Production Functions Under Increasing Returns to Scale," Journal of the Royal Statistical Society, Series A, 125 (1962): 252-267

Forsund, Finn R., C.A. Knox Lovell, and Peter Schmidt,"A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement," to appear in Journal of Econometrics, Mimeo, Draft, September, 1979

Forsund, F.R. and L. Hjalmarsson,"On The Measurment Of Productive Efficiency," Swedish Journal Of Economics, 72 (1974): 141-154

Forsund, Finn R. and Lennart Hjalmarsson,"Generalised Farrell Measures on Efficiency: An Application to Milk Processing In Swedish Dairy Plants," The Economic Journal, 89 (1979): 294-315

French, Ben C.,"The Analysis of Productive Efficiency in Agricultural Marketing: Models, Methods, and Progress," in A Survey of Agricultural Economics Literature: Volume 1, Traditional Fields of Agricultural Economics, 1940's to 1970's, Lee R. Martin, ed., University of Minnesota Press, Minneapolis, 1977

Fuss, Melvyn, Daniel McFadden, and Yair Mundlak, "A Survey of Functional Forms In The Economic Analysis Of Production," Chapter II.1, in Production Economics: A Dual Approach To Theory And Applications, Volume 1, Melvyn Fuss and Daniel McFadden, eds., North-Holland, New York, 1978

Gold, Bela, "Changing Perspectives on Size, Scale, and Returns: An Interpretive Survey," Journal of Economic Literature, 19 (1981): 5-33

Griliches, Zvi,"Measuring Inputs In Agriculture: A Critical Survey," Journal of Farm Economics, 42 (1960): 1411-1427

Hall, Bruce F. and E. Phillip LeVeen, "Farm Size and Economic Efficiency: The Case of California," American Journal of Agricultural Economics, 60 (1978): 589-600

Henderson, James M. and Richard E. Quandt, Microeconomic Theory: A Mathematical Approach, Second Edition, McGraw-Hill Book Company, New York, 1971

Hoch, Irving, "Estimation of Production Function Parameters Combining Time-Series And Cross-Section Data," Econometrica, 30 (1962): 34-53

Holland, David, Production Efficiency And Economics Of Size In Agriculture, Scientific Paper #5266, College of Agriculture, Agricultural Research, Washington State University, Pullman, Washington,

Jamison, Dean T. and Lawrence J. Lau, Farmer Education and Farm Efficiency, John Hopkins University Press, Baltimore, 1982

Johnson, Glenn L. and C. Leory Quance, editors, The Overproduction Trap in U.S. Agriculture, The John Hopkins University Press, 1972

Jorgenson, Dale W. and Lawrence J. Lau, "The Duality of Technology and Economic Behaviour," The Review of Economic Studies, 41 (1971): 181-200

King, Richard A., "The Frontier Production Function: A Tool For Improved Decision Making," Journal of the Northeastern Agricultural Economics Council, 9 (1980): 1-10

Knight, Frank H., The Economic Organization, University of Chicago, 1933

Koopmans, T.C., "Analysis of Production as an Efficient Combination of Activities," Chapter III in Activity Analysis of Production and Allocation, Cowles Commission Monograph 13, T.C. Koopmans, ed., John Wiley and Sons, New York, 1951, pg. 33-97

Kopp, Raymond J., "The Measurement Of Productive Efficiency: A Reconsideration," Quarterly Journal Of Economics, 96 (1981a): 477-503

Kopp, Raymond J., "Measuring the Technical Efficiency of Production: A Comment," Journal of Economic Theory, 25 (1981b): 450-452

Kopp, Raymond J. and W. Erwin Diewert, "The Decomposition Of Frontier Cost Function Deviations Into Measures Of Technical And Allocative Efficiency," Journal of Economics, 19 (1982): 319-331

Kopp, Raymond J. and V. Kerry Smith, "Frontier Production Function Estimates for Steam Electric Generation: A Comparative Analysis," Southern Economic Journal, 46 (1980): 1049-1059

Lau, Lawrence J. and Pan A. Yotopoulus, "A Test for Relative Efficiency and Application To Indian Agriculture," American Economic Review, 61 (1971): 94-109

Leibenstein, Harvey, "Allocative Efficiency vs. 'X-Efficiency'," American Economics Review, 56 (1966): 392-415

Lesser, W.H. and W.H. Greene, "Economies of Size and Operating Efficiency of Livestock Markets: A Frontier Function Approach," Journal of the Northeastern Agricultural Economics Council, 9 (1980): 37-40

Madden, J. Patrick, Economies of Size in Farming, Agricultural Economics Report #107, ERS, USDA, February 1967

Malinvaud, E.,Lectures On Microeconomic Theory, North-Holland Publishing Company, Amsterdam, 1972

McFadden, Daniel, "Cost, Revenue, and Profit Functions," Chapter I.1 in Production Economics: A Dual Approach To Theory And Applications, Volume 1, Melvyn Fuss and Daniel McFadden, eds., North-Holland, New York, 1978

Miller, Bill R., Measuring Economic Efficiency: The Farrell Approach, Staff Paper of Agricultural Economics Department, Auburn University

Miller, Thomas A., Gordon E. Rodewald, and Robert G. McElroy, Economies of Size in Field Crop Farming, USDA, ESS, Agricultural Economics Report #472, July 1981

Muller, Jurgen, "On Sources of Measured Technical Efficiency: The Impact of Information," American Journal Of Agricultural Economics, 56 (1974): 730-738

Nerlove, Marc, Estimation and Identification of Cobb-Douglas Production Functions, Rand McNally and Company, Chicago, 1965

Pachico, Douglas, Applying Efficiency Analysis To Small Farms In Low Income Countries: Some Theoretical And Empirical Considerations, Cornell International Agriculture Mimeograph 84, Department of Agricultural Economics, Cornell University, November 1980

Quirk, James and Rubin Saposnick, Introduction To General Equilibrium Theory And Welfare Economics, McGraw-Hill Book Company, New York, 1968

Shapiro, Kenneth H. and Jurgen Muller, Sources of Technical Efficiency: The Roles of Modernization and Information, Discussion Paper 36, Center For Research On Economic Development, University of Michigan, Ann Arbor, April 1974

Shapiro, Kenneth H. and Jurgen Muller, "Sources of Technical Efficiency: The Roles of Modernization and Information," Economic Development and Cultural Change, 25 (1977): 293-316

Schmidt, Peter and Tsai-Fen Lin, Simple Tests Of Alternative Specifications In Stochastic Frontier Models, Mimeo, 1983

Schmidt, Peter, and C.A. Knox Lovell, Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers, MSU Econometrics Workshop Paper #7702, December 1977

Schmidt, Peter, and C.A. Knox Lovell, "Estimating Stochastic Production and Cost Frontiers When Technical and Allocative Inefficiency Are Correlated," to appear in Journal of Econometrics, Draft November 1978

Schultze, Theodore W., Transforming Traditional Agriculture, Yale University Press, New Haven, 1964

Seitz, Wesley D., "Efficiency Measures For Steam-Electric Generating Plants," Proceedings of the Western Farm Economics Association Meeting 1966, Western Economics Farm Association, pp.143-151

Stigler, George J., "The Xistence of X-Efficiency," American Economics Review, 66 (1976): 213-216

Timmer, C.P., "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency," Journal of Political Economy, 79 (1971): 776-794

USDA, Measurement Of U.S. Agricultural Productivity: A Review Of Current Statistics And Proposals For Change, USDA, ESCS, Technical Bulletin #1614, February 1980

Varian, Hal R., Microeconomic Analysis, W.W. Norton & Company, New York, 1978

Wark, Kennth, Thermodynamics, Fourth Edition, McGraw-Hill Book Company, 1983

Yotopoulos, P.A., "Rationality, Efficiency, And Organizational Behavior Through The Production Function, Darkly," Food Research Institute Studies, 13 (1974): 263-274

Yotopoulos, Pan A., and Lawrence J. Lau, "A Test for Relative Economic Efficiency: Some Further Results," American Economic Review, 63 (1973): 214-223