5-1990

Production and Inefficiency

Arunava Bhattacharyya
Utah State University

PRODUCTION AND INEFFICIENCY

by

Arunava Bhattacharyya

A dissertation submitted in partial fulfillment of
the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Economics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1990

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my major professor, Dr. T. F. Glover, for the encouragement he provided all through my study. Without his guidance and support it would not have been possible for me to complete this dissertation. I have come across very few teachers comparable to him, and I will never be able to thank him enough for all I have learned from him.

I would like to thank Dr. Adele Cutler for her helpful suggestions at all stages of this dissertation. The support that she has given me in statistical estimation was invaluable.

I have received much help from Dr. Basudeb Biswas, academic as well as personal, for which I am very much greatful. I would like to thank Dr. Lewis and Dr. Keith for their participation as committee members and for their helpful comments on this dissertation.

I would like to thank Dr. Lyon for his help in writing the programme for reading the data file. Dr. Samar Datta of the Agro Economic Research Centre provided the data set for which I am very much greatful. I would like to thank my friend, Dr. Subal Kumbhakar for his helpful comments and suggestions in this dissertation. Jane Post and Bob Bayn of computer services have extended their help in using computer facilities at Utah State.

My special thanks go to my parents for their encouragement of my studies abroad. I would also like to thank my wife, Anjana, for her personal support and encouragement through this study.

Finally, I wish to express my sincere thanks and gratitude to all my colleagues and teachers at Utah State and in India for their encouragement for this research.

This study was partially funded by the Utah Agricultural Experiment Station and the Graduate School through President Fellowship, for which I am greatful. However, the content of the dissertation remains my responsibility.

Arunava Bhattacharyya

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Production and Inefficiency

by

Arunava Bhattacharyya, Doctor of Philosophy
Utah State University, 1990

Major Professor: Dr. T. F. Glover
Department: Economics

The overall purpose of this three-part dissertation is to specify and estimate various components of inefficiency in the production and profit-generating processes. Flexibility in inefficiency-measurement techniques is introduced using stochastic functional forms to overcome the restrictions of the simplifying assumptions used in previous studies. In addition, the profit function approach is used to measure firm specific inefficiency and to view profit inefficiency in the multiple output context. Empirical application of each approach is also attempted. Application of the measurement of the inefficiency component in the first two essays is made using data taken from Indian agriculture. The multiple output model of the third essay is applied to data of the U. S. unit bank taken from the Functional Cost Analysis programme of the Federal Reserve banking system.

In the first essay, a quasi-translog production function is introduced and allocative, technical, and scale inefficiencies are estimated for Indian agriculture with large and small farm divisions. Results obtained contradict earlier conclusions regarding the efficiency of Indian farms.

In the second essay, a Normalized Restricted Profit function is used to estimate allocative, scale, and profit inefficiency for the same set of farms. Empirical results confirm the conclusions of the first essay. Technical inefficiency cannot be isolated in this case, because the impact of technical inefficiency is confounded in the measure of profit inefficiency.

In the third essay, a translog profit function is used to estimate profit and allocative inefficiency in U. S. banking operations.

(162 pages)

CHAPTER I
INTRODUCTION AND BACKGROUND IN MEASURING INEFFICIENCY

Firm efficiency has long been a concern of economists. Better use of existing technology has been shown to provide substantial cost savings and improvements in profitability [e.g., Yotopolous and Lau, Lau and Yotopolous (1971, 1972), Caves and Barton, Kumbhakar, and Ali and Flinn among others]. However, much of the modeling of firm technology assumes that efficiency, in the sense of optimal employment of inputs and supply of output, is operative and that the stochastic term appended to empirical production structure is only statistical white noise. Considering the vast literature on firm efficiency, little has been done to isolate elements of inefficiency in the production process, cost minimization behavior, or maximizing profit.

In fact, the idea of separating inefficiency components in the error term of production structure (production function, cost function, or profit function) has been quite controversial, and no uniformly accepted procedure has emerged. At this point, researchers are left with three choices: 1) not to measure inefficiency; 2) to assume all deviations from optimal supply conditions can be explained by elements of risk; or 3) to measure the components of inefficiency and change firm organization to bring about optimality. In this study, the latter course has been taken.

The idea of production inefficiency has not received sufficient attention in the mainstream of modern neoclassical economics. The field developed as an auxilliary logic and placed more emphasis on the quest for empirical estimation than on the development of full-scale theoretical direction. However, a theoretical basis for empirical estimation of the components of inefficiency has crystalized over the past four decades through contradiction of a variety of models set up for the investigation of a wide range of efficiency-related topics.

The development of the stochastic frontier functional specification is primarily responsible for recent interest in measuring inefficiency. This development extended the early work of Farrell. Inefficiency up to this point has mainly been examined by separating out two components, technical and allocative inefficiency (to be defined later), with the former usually derived from production function or programming models and the latter obtained by comparing input marginal prod-

ucts with normalized input prices. Some developments of the empirical profit function now allow economists to test for differences in technical and allocative efficiency among groups of firms, but these approaches do not allow for measurement of firm-specific efficiency. Only recently has the concept of profit inefficiency been introduced in the literature. This idea extends the notion of the production frontier to the profit frontier and allows for extended measurement of inefficiency via the empirical profit function; however, very little research has been done to make this empirical extension.

The overall purpose of this three-part study is to specify and estimate various components of inefficiency in the production and profit-generating processes. A specific goal is to introduce flexibility in inefficiency-measurement techniques using stochastic functional forms to overcome restrictions imbedded in oversimplified assumptions in previous studies of firm efficiency. In addition, a second specific goal is to extend the profit function approach to measure firm-specific inefficiency and to view profit inefficiency in the multiple output context. This attempt progresses from the use of the production function to a system of relationships derived from and including the profit system, with an empirical application of each approach. Inefficiency components are measured in the first two essays using data taken from Indian agriculture. In the third essay, the stochastic multiple output profit system is used to examine inefficiency in a set of U.S. unit banks using data taken from the Functional Cost Analysis (FCA) programme of the Federal Reserve banking system.

In the first essay, a quasi-translog production frontier is developed to measure allocative, technical, and scale inefficiency. While not as versatile as a fully flexible functional form, the quasi-translog specification does allow for cross effects between two groups of factors of production. More importantly, this specification derives analytical solutions for the various inefficiency measures, their interactions, and their economic impacts. Use of fully flexible functional forms for similar purposes is greatly hindered for a number of reasons, as pointed out in the second essay.

The second essay, then, introduces a fully flexible functional form to examine inefficiency. A dual profit function specification is developed and used to overcome some inherent problems of using other value-dual relationships. Use of such a relationship is successful in measuring all three types of inefficiency. Finally, the third essay estimates profit inefficiency in a multiple output framework, along

with allocative and scale inefficiencies for each output and input (both termed as netputs). Use of the profit system is extended to the multiple output and panel data case using the banking firm data.

Definition of Inefficiency

Until recently, modern neoclassical theory has typically ignored the possibility that producers might operate inefficiently. All pioneering works of the neoclassical school have naively assumed that the individual producer allocates each economic resource in an efficient manner. It has also been assumed that individual producers are efficient relative to their production technology and that the structure of the input and output markets are subject to behavioral goals set by the private producer. The technological constraint in the producers' behavioral optimization is assumed to be binding, which eliminates the possibility of technical inefficiency. Similarly, satisfaction of the first-order conditions for behavioral optimization eliminates all behavioral inefficiency. Satisfaction of all regularity conditions for binding production technology eliminates the possibility of structural inefficiency through satiation of second-order conditions.[1] This led to the full efficient production model. Modification of the fully efficient model has not been explored until recently, although it was recognized that modification of naive assumptions through more realistic production structures would enrich the testable hypothesis.

Recent development of the duality approach to production theory was mainly to establish equivalance between the structure of production technology and the cost, revenue, and profit functions. As a result, duality theory does not contain the efficiency issue. In fact, the strict duality theorem (more precisely, the conditions required for establishing the dual relationships) effectively excludes all types of inefficiency. Efficient production with strictly positive input and output prices eliminates the possibilities of operation in the region of technology that fails to satisfy the regularity conditions of monotonicity and curvature of a production technology; thus, the very characteristic of the technology that gives rise to structural inefficiency is discarded on efficiency grounds. As a result, the conventional hypothesis about producers' behaviors based on duality also ignores the possibility of inefficiency; however, dual relationships may be modified to investigate the

[1] If the production point occurs at the interior of the production possibility set, the firm is said to be structurally efficient.

implication of inefficiency. Since any type of inefficiency is costly, one can examine the implication of extra cost or forgone profit (or unexploited revenue) by looking at the value duals.

Analysis of economic efficiency started with the work of Koopmans (1951, 1957), Debreu, and Farrell. While Koopmans and Debreu mainly confined their study to technical inefficiency, Farrell first provided the partial decomposition of inefficiency into technical and allocative components. Koopmans defined and characterized technical inefficiency, while Debreu first provided a measure of the degree of technical efficiency and its associated cost. Following the tradition of Koopmans and Debreu, Eichhorn (1978a, 1978b) developed a price-dependent notion of the technical and economic effectiveness of such inefficiencies in terms of a system of functional equations, but his measure is an index of overall private efficiency that cannot be decomposed into constituent parts.

Farrell not only proposed a practical decomposition of private efficiency into technical and allocative components, he also proposed indexes of technical, allocative, and overall inefficiency. Farrell's index of technical inefficiency is similar to Debreu's coefficient of resource utilization. This coefficient is computed as one minus the maximum equiproportionate reduction in all inputs consistent with the continued production of existing outputs, referred to in the literature as a radial measure of inefficiency. Farrell's decomposition proved very appropriate for the restrictive technology he considered but inapplicable to less restrictive technologies.

The restrictive nature of Farrell's approach prompted development of two separate and more general approaches. The first, suggested by Färe and Grosskopf (1983a), augments Farrell's approach in which overall efficiency is decomposed into technical, structural, and allocative components, and the measure of efficiency still remains radial. The second development is the non-radial measure of inefficiency, which preserves a two-way decomposition into allocative and technical inefficiencies. This measure of inefficiency allows disproportionate input reductions and/or disproportionate output increases and thereby assigns different shares to the technical and allocative components of overall private inefficiency.

Besides these two types of inefficiency, the size inefficiency of a firm relative to the industry's performance could be measured by comparing actual to optimal firm size, producing a scale inefficiency measure. An index of scale inefficiency based on this technically optimal firm size was proposed by Forsund and Hjal-

marsson (1974, 1979). Färe, Grosskopf, and Lovell later showed how to determine whether scale inefficiency is due to increasing or decreasing returns to scale.

There are five types of inefficiency measures to be derived for one feasible production plan, namely:

1. the classical Farrell input measure of technical inefficiency, FITE;
2. the weak input measure of technical inefficiency, WITE, as extended by Färe and Grosskopf (1983a, 1983b); and
3. the overall measure of inefficiency, OIE, as proposed by Farrell and Eichhorn (1978a).

Then, from these three primary measures can be derived two other measures of input inefficiency:

4. an input congestion measure of inefficiency, ICE, proposed by Färe and Svensson and Färe and Grosskopf (1983b), and
5. allocative input measure of inefficiency, AIE, as proposed by Farrell.

These five measures of inefficiency can be both radial and non-radial measures. Since radial input inefficiency measures are consistent with Farrell's definition, this study uses the radial measure of inefficiency. This is called radial because inefficiency relations are proportional measures along a radius vector. These radial input efficiency measures can be expressed either as input inefficiency measures or output inefficiency measures. In the first case, efficiency of an input vector is measured by comparing the observed input vector to smaller feasible input vectors in a feasible production plan. In the second case, efficiency in output is measured by comparing the observed output to higher output feasible from a given input vector. Here again, this study will utilize input inefficiency measures because Farrell's definitions are in terms of input vectors.

Before addressing the formal definition of individual inefficiency, a formal definition of input inefficiency is needed. Let X be a vector of inputs, $X \in R^n_+$, which can be used to produce a vector of outputs $Y : Y \in R^m_+$. The input correspondance for Y can be stated $X \in L(Y)$. The feasible production plan, therefore, is $\{(Y, X) : X \in L(Y)\}$. If an input vector $X' \leq X$ exists such that $\{(Y, X') : X' \in L(Y),\}$ then X is inefficient for Y.

Farrell's Input Measure of
Technical Inefficiency

Suppose a feasible production plan such as $\{(Y, X) : X \in L(Y)\}$ exists, as represented in figure 1, and is defined over a non-negative orthant in $\{X, Y\}$. Point K represents an input vector X used to produce an output vector Y, where X belongs to the interior of $L(Y)$. For any $Y \in R^m_+$, $L(Y)$ denotes the subset of all input vectors for $X \in R^n_+$ that yields at least Y. The production plan is so defined that X belongs to the interior of $L(Y)$. Define a non-positive orthant in R^2_- with origin at K and slide that down the radius vector OK such that a minimum nonempty intersection with $L(Y)$ is maintained. The transformed non-positive orthant is represented by $Q'\,Q\,Q''$, with a non-empty intersection with $L(Y)$ at C. Suppose the production unit faces an exogeneous input price vector $P \in R^n_{++}$. P with X generates a lower halfspace $H_-(X, P) := \{Z : PZ \leq PX; Z \leq X\}$ that is represented by MM'. The input cost is minimized when MM' intersects $L(Y)$ at point D.

Given this production plan, Farrell's measure of input technical inefficiency (FITE) is the ratio of the smallest feasible radial contraction of an observed input vector to the observed input vector itself, in terms of figure 1 given by (OB/OK). If both inputs x_1 and x_2 $(x_1, x_2 \in X)$ are scaled down proportionately by (OB/OK), the resulting input vector at B $(B \in Isoq L(Y))$[2]is techanically efficient in Farrell's sense because the same level of output can be produced with less of both factors, keeping their ratio constant. If there exists a minimum λ, say $\lambda^\circ < 1$, then the firm can produce the given output vector Y with a radially smaller input vector $\lambda^\circ X \in Isoq L(Y)$, and the observed input vector X is inefficient technically. The FITE of a firm is (OB/OK), and then corresponding technical inefficiency of the firm is $(1 - (OB/OK))$.

A cost interpretation of FITE can also be given because FITE also measures resultant cost savings from moving from K to B (in figure 1) for any price vector $P \in R^n_{++}$, since input cost at B is a fraction (OB/OK) of input cost at K. (OB/OK) measures the ratio of input cost at $\lambda^\circ X \in Isoq L(Y)$ to observed input cost. This ratio is independent of input prices, since FITE can be expressed in

[2] $Isoq$ represents isoquant. $A := B$ means A is defined by B. $A \in B$ means A ia an element in B. $R^k :=$ Euclidean space of dimension k. $R^k_+ := x : x \in R^k, x \geq 0$ and $R^k_{++} := x : x \in R^k, x > 0$.

value terms as

$$\frac{P \times (\lambda^{\circ} X)}{PX} = \lambda^{\circ}.$$

Weak Input Technical Efficiency

The idea of a weak input technical efficiency (WITE) measure was first introduced in the literature by Färe and Grosskopf (1983a, 1983b). If both elements of the observed input vector X are scaled down proportionately by the scalar (OQ/OK), the resulting input vector at point Q is technically efficient in the weak sense. At Q the output vector Y is not obtainable as $Q \notin L(Y)$. However, by disposing QC of x_2, Y is obtainable as $C \in L(Y)$. Thus, if Q is available and x_2 is strictly disposable, then the input vector Q is weakly input technically efficient (WITE), such that $Q \in L(Y)$ in *a weak sense*; therefore, $WITE = OQ/OK$. Thus, if a nonempty intersection exists between the transformed nonpositive orthant $Q'QQ''$ and $L(Y)$ along with a strong disposability of inputs, then WITE measures the ratio of the smallest radial contraction of X to X itself, such that the radial contraction belongs to the weakly efficient input subset.

A cost savings interpretation of WITE can also be given because cost of production falls due to a movement from K to C. Input cost at C is a fraction (OQ/OK) of input cost at K for any input price vector. In this case, input vector Q is purchased while C is actually employed, so $\overline{QC}$ amount of input is purchased but not employed. This is an added cost of weak technical inefficiency. Assume that a minimum $\lambda = \lambda^{*}$ exists and $\lambda^{*} < 1$. The firm can produce output vector Y with radially smaller input vector $\lambda^{*}X$, disposing of congesting inputs, and the observed input vector is technically inefficient. Since λ^{*} and $P \in R^{n}_{++}$ exist, WITE is the ratio of the input cost at Q to observed input cost PX, i.e.,

$$WITE = \frac{P \times (\lambda^{*} X)}{PX} = \lambda^{*}.$$

This ratio is also price independent.

Overall Input Efficiency

The overall input efficiency (OIE) measure, first introduced by Farrell and refined by Eichhorn (1978a), was devised to measure production unit success in minimizing the production cost of a certain output vector by adjusting its input vector in light of input prices. In the strict Farrell sense, OIE represents the

radial shrinkage (OE/OK). Although $E \notin L(Y)$ but $D \in L(Y)$ where vectors D and E represent the same level of input cost, production at D is feasible, given prices. Since this OIE involves a radial shrinkage of input vector and a change in input mix (due to price), the OIE measure involves both technical and non-technical components and is price dependent. The non-technical elements are the allocative component.

In terms of the figure 1 for $X \in L(Y)$, $P \in R^n_{++}$ the lower half space $H_-(X, P)$ can be scaled down along the radius vector OK under the condition that the transformed lower half space $H_-(X, P)$ has a non-empty intersection with $L(Y)$. Since at D, input cost of producing Y vector is minimized at input price $P \in R^n_{++}$, i.e., $D \in CM(Y, P)$, OIE also measures the cost savings due to movement from K to D. Thus, OIE measures the change in cost (which can be either positive or negative) incurred by adjusting inputs from $X \notin L(Y)$ to point $D \in CM(Y, P)$. Thus, an input vector X is OIE for prices P if $X \in CM(Y, P)$. In this case, $CM(Y, P)$ is the reference set used by the OIE measure. Total efficiency of the firm is (OE/OA), and the corresponding total inefficiency of the firm is $(1 - (OE/OK))$.

From these three measures of input inefficiency, two derived measures of ineffciency can be formulated — input congestion inefficiency and allocative input inefficiency.

Input Congestion Inefficiency

The idea of input congestion inefficiency was first introduced by Färe and Svensson. In a general sense, input congestion occurs whenever the increase in some input(s) leads to a fall in output or a decrease of some input(s) leads to a rise in output. According to Färe, Grosskopf and Lovell, a feasible production technology is input congested if for some non-negative output vector say Y' and input vector Z, such that $Z \in L(Y)$ there exists an input vector $X \geq Z$, such that $X \in L(\theta Y')$ for $0 \leq \theta \leq 1$ and $X \notin L(Y')$. The input vector $X \in L(\theta Y')$ is said to be congesting if by decreasing X to Z, output can be increased to Y'. In terms of figure 1, input congestion inefficiency is $ICE = (OQ/OB)$ or $ICE = (WITE/FITE)$. An input reduction and a consequent cost savings of (OQ/OB) could be realized if the technology was not input congested, after which $\overline{QC}$ units of congesting input could be disposed of to generates $C \in L(Y)$. Thus, ICE measures the proportional reduction in inputs from the isoquant of the parent

technology to the weak efficient subset of the corresponding input congestion-free technology.

Allocative Input Efficiency

This inefficiency component measures the input mix error relative to the input price faced by the production unit. An allocative input inefficiency measure, $A(Y, X, P)$, measures the minimal ray distance from the input congestion-free technology to the hyperplane yielding the minimum cost of producing output Y at input prices P. From the definition of overall inefficiency and weak input technical inefficiency, the allocative efficiency measure derived is $A(Y, X, P) = (OE/OQ) = (OIE/WITE)$. A reallocation of input from Q to D eliminates allocative or input price inefficiency, thereby reducing input cost to a fraction $AIE = (OE/OQ)$ of its value at Q. The formal definition of Farrell's measure of allocative efficiency is that an input vector $X \in L(Y)$ for $Y \geq 0$ is allocatively efficient for $P \in R^n_{++}$ if there exists a $\lambda \in [0, 1]$, such that $\lambda X \in CM(Y, P)$. Allocative efficiency of the firm is given by (OE/OQ), and the corresponding allocative inefficiency is $(1 - (OE/OQ))$.

Scale Inefficiency

Even if a producer is completely efficient in the technical and allocative sense, the resulting combination of input usage and output supply may be suboptimal, given the market price of the product. A combination of technical and allocative efficiency is necessary but not sufficient for profit maximization. A divergence may exist between the actual size of the production unit and the ideal size of that unit. Ideal means that input-output combination which corresponds to a zero profit long-run competitive equilibrium solution. Divergence from this optimal solution is referred to as scale inefficiency. A firm is said to be scale efficient if

$$p = CM_y(y, W, Z),$$

where p is the market price of output, y, W is the vector of hiring prices of endogeneous inputs, and Z is the vector of exogeneous fixed inputs. A firm is said to be scale inefficient if

$$p \neq CM_y(y, W, Z).$$

Therefore, a firm can be profit efficient if, and only if, the firm is technically, allocatively, and scale efficient, given that factors are freely disposable.

A Survey of Theories on Efficiency Measurement

This section describes the background literature on measurement of production inefficiency using the frontier production function.[3] A production function represents the maximum possible level (technologically) of output, given a set of inputs. However, estimation of a production function by any standard econometric technique would generate both positive and negative residuals since least squares estimates mean output, not maximum output, which contradicts the basic definition of a production function. A firm can produce below the technically possible maximum level but not above it. From a statistical point of view, it is not obvious why the maximum of the distribution of output is more interesting than the mean, but to answer some economic questions, particularly the measurement of inefficiency of a firm or plant, the maximum possible output is relevant. To say that a firm is 10% inefficient, i.e., that it is produces 90% of the output it could actually produce given input usage, we need to know what 100% output is. To answer this question, the modern literature of efficiency measurement was developed based on the idea of the frontier production function. The idea of frontier production function incorporates the idea of maximality that sets a limit to the range of possible observations. This limit is defined by an exogenously given technology.

Since the publication of Farrell's seminal paper on decomposition of productive inefficiency, development of the technique of inefficiency measurement has followed two major trends, the first being the measurement of non-statistical frontiers and the second being statistical frontier measurement. The second method depends on the statistical properties of the data, while the first method does not. The way in which econometricians look at the measurement of the production frontier has undergone substantial changes in last the two decades. The earlier work (1957–1977) assumed what is known as a "deterministic frontier" and led to the development of the "stochastic frontier" approach. The next subsection which discusses the deterministic frontier approach to the problem of inefficiency measurement, will be followed by a summary of the stochastic frontier approach. The final section discusses estimation methods and summarizes some fundamental literature on the estimation of inefficiency using the stochastic frontier approach.

[3] Non-frontier and non-parametric approaches will not be discussed here; however, a brief analysis of non-frontier methodology of efficiency estimation is included in appendix A.

Deterministic Frontier Approach

The idea of the deterministic frontier approach was developed in the writings of Farrell, Farrell and Fieldhouse, and Afriat. Empirical applications of this type of model were made by Aigner and Chu, Seitz, Richmond, Forsund and Jansen, and others. In each case, estimates are obtained by solving a constrained linear programming model. The full frontier model has the theoretical appeal of forcing the fitted function to correspond to the underlying theory in terms of the signs of the residuals. Moreover, they are maximum likelihood estimates for certain stochastic specifications. However, econometricians like Timmer, Greene (1980a), and Schmidt (1976) pointed out both conceptual and statistical problems with estimating of the deterministic frontier, which lead to the development of stochastic frontier production function.

The deterministic frontier approach assumes a function that gives maximum possible output as a function of certain inputs, maximal in the sense that output is bounded from above by a deterministic (non-stochastic) production function. For the ith firm this can be expressed as

$$y_i = f(X_i;\beta)\,, \tag{1}$$

where y_i is the maximum output obtainable from the non-stochastic input vector X_i, and β is an unknown parameter vector to be estimated. Aigner and Chu suggested application of mathematical programming methods to estimate β, a problem written as

$$\min_{\beta} \sum_{i=1}^{n} |y_i - f(X_i;\beta)| \tag{2}$$

$$\text{subject to} \qquad y_i \le f(X_i;\beta) \qquad \forall\ i.$$

This can be solved by the standard linear programming method, which suggests minimization of the sum of absolute residuals from the logarithm of the functional specification [4] while constraining all residuals to be negative. Aigner and Chu further suggested a quadratic minimization problem of the form,

$$\min_{\beta} \sum_{i=1}^{n} \left\{ y_i - f(X_i;\beta) \right\}^2 \tag{2a}$$

[4] Aigner and Chu used the Cobb-Douglas specification, $y = \alpha \prod_i x_i^{\alpha_i} u$, for their model.

$$\text{subject to} \qquad y_i \leq f(X_i;\beta).$$

The basic problem with such estimates is that their statistical properties are unknown and remain unknown unless further assumptions are made. This type of analysis, first done by Afriat, implicitly assumes a one-sided error term in the frontier, which leads to the introduction of a one-sided error term with a specific distributional assumption (exponential or gamma), after which the likelihood function is derived and maximum likelihood estimates are calculated. Schmidt (1976) showed that the Aigner and Chu linear programming technique is equivalant to the maximum likelihood estimation of the model

$$\ln y = \ln A + \alpha_1 \ln x_1 + \alpha_2 \ln x_2 - \epsilon, \tag{3}$$

where $\epsilon = -\ln u$ and has an exponential distribution,

$$f_\epsilon(\epsilon) = \lambda e^{-\lambda\epsilon}, \qquad \text{for} \qquad \epsilon \geq 0; \lambda > 0. \tag{4}$$

Schmidt further showed that the quadratic programming estimator of Aigner and Chu is maximum likelihood if ϵ has a half-normal distribution,

$$f_\epsilon(\epsilon) = \frac{2}{\theta\sqrt{\pi}} \exp\left\{\frac{-\epsilon^2}{2\theta^2}\right\}, \qquad \text{for} \quad \epsilon \geq 0,\ \theta > 0. \tag{5}$$

Forsund and Jansen first applied the duality relationship for estimating the frontier model when studied the production frontier by estimating a dual cost frontier. Starting with a homothetic production function

$$y^\beta e^{\gamma y} = A \prod_{j=1}^{M} x_j^{\alpha_j} u, \tag{6}$$

where y is output, x_j is inputs, $\sum_j^M \alpha_j = 1$, and u is assumed to have a distribution

$$f_u(u) = (1+\alpha)\, u^\alpha, \qquad \text{for} \quad 0 < u \leq 1,\ \alpha > -1. \tag{7}$$

The dual cost function, therefore, is

$$C = B y^\beta e^{\gamma y} \prod_{j}^{M} p_j u^{-1}, \tag{8}$$

where p_j is the unit price of x_j, and C is the total cost. Forsund and Jansen showed that maximum likelihood estimates are obtained using linear programming by minimizing the sum of the positive residuals of their dual cost frontier. Greene (1980a) showed that Forsund and Jansens' stochastic framework is equivalant to the Aigner and Chu specification if u is transformed to $\epsilon = \ln u^{-1}$, resulting in the following distribution of ϵ:

$$f_\epsilon = (1+\alpha)\, e^{-(1+\alpha)\epsilon}, \qquad \epsilon \geq 0. \tag{9}$$

Apart from conceptual questions of whether the frontier really ought to be deterministic, severe statistical problems occur in estimating of the deterministic frontier.

Although the one-sided disturbance was implicitly assumed in the Aigner and Chu process, no assumption was made about its properties. As a result, the parameters are not estimated in any statistical sense but are computed via the mathematical programming technique. In the second case (Afriat), a one-sided disturbance term was explicitly assumed, but one of the regularity conditions of MLE was violated (see Greene (1980a) for a detailed discussion); thus, properties of the estimated parameters are uncertain. Greene (1980a) pointed out some sufficiency conditions for the distribution of the one sided error term so that all regularity conditions of a MLE could be maintained; however, these conditions are very restrictive.

Another problem of the deterministic frontier approach is that it is extremely sensitive to outliers. To solve this problem, the probabilistic frontier model (Timmer, Dugger) was developed. In the probabilistic frontier approach, a deterministic frontier is estimated after discarding outliers from the sample. The selection of the outlier is completely arbitrary and lacks explicit economic and statistical justification. Moreover, since a probabilistic frontier is just a deterministic frontier, it remains vulnerable to the criticism that parameters are computed rather than estimated, rendering hypothesis testing impossible. Another problem involves reconciling the observations above the frontier with the concept of the frontier as a maximal possible output. Typically, this can be accomplished by measuring errors in the extreme observations, though it seems preferable to incorporate the possibility of measurement errors and other unobserved shocks in a more statistically meaningful way. This is accomplished through the development of stochastic frontier functional specification.

Stochastic Frontier Approach

To correct the problems associated with both the deterministic and probabilistic frontier approaches, Aigner, Lovell, and Schmidt and Meeusen and van den Broeck specified a stochastic frontier. In such a specification, the output of each firm is bounded above by a stochastic frontier whose placement is allowed to vary randomly across firms. From the economic standpoint, this approach permits firms to be technically inefficient relative to their own frontier rather than to some sample norm. Interfirm variation of the frontier presumably captures the effects of exogenous shocks, favourable and unfavourable, beyond the firms' control. Errors of observation and measurement of output, on the other hand, constitute another source of variation in the frontier. The statistical specification assumes the disturbance term is made up of two parts, namely a symmetric (normal) component that captures randomness outside the control of the firm, and a one-sided (non-positive) component that captures randomness under the control of the firm (inefficiency). Estimation of the frontier is then a statistical problem, and the usual types of statistical inference are possible.

In the stochastic frontier model, output is assumed to be bounded from above by a stochastic production function. For the ith firm, this can be written as

$$y_i = f(X_i;\beta) + \nu\,, \tag{10}$$

where ν represents a random error term. The output shortfall from the defined frontier is represented by τ $(\tau \leq 0)$. Equation (10) then can be expressed as

$$y_i = f(X_i;\beta) + \nu + \tau\,. \tag{10a}$$

This method simply distinguishes production inefficiency from other sources of disturbances beyond the firm's control. This is a great leap forward compared to the deterministic or probabilistic frontier approach. In principal, productive efficiency of the ith firm can be measured by the ratio

$$\frac{y_i}{\left\{f(X_i;\beta) + \nu\right\}}, \tag{11}$$

whereas in the deterministic case efficiency is measured by the ratio

$$\frac{y_i}{\left\{f(X_i;\beta)\right\}}. \tag{12}$$

Measure (11) clearly distinguishes productive inefficiency from other sources of disturbance beyond the firm's control, e.g., the farmer whose crop is destroyed by flood is unlucky in terms of measure (11) but inefficient by measure (12).

All stochastic frontier models discussed so far assume the error that represents statistical white noise, v, to be independent and identically distributed (iid) normally. At least three different types of distributions occur in the literature to specify the one-sided error term, τ, the most common of which is the half-normal. Stevenson generalized the half-normal by considering the normal truncated from below at zero. Greene (1980b) considered a two-parameter gamma distribution. If the two errors are assumed independent of each other and the inputs, then the likelihood function can be derived and maximum likelihood estimates can be estimated. A similar specification is possible using dual behavioral relationships like cost functions and profit functions.

Estimation of a Stochastic Frontier

Essentially two types of issues are involved in the estimation of stochastic frontiers, the first being mainly an economic issue and the second mainly statistical. Handling the first issue is relatively easy if the production model is a single equation model and if the analysis is confined to homothetic relationships. The economic issue involves specification of the model, while the statistical problems are generally related to specification of the distribution of the error terms. Of course, the severity of the second problem depends on the modeling of the economic problem.

Just as output must lie below the production frontier, so cost must lie above the cost frontier and profit below the profit frontier. Any of the statistical techniques used to estimate the production frontier can, therefore, be applied to estimate profit and cost frontiers with appropriate modifications. The choice of whether to estimate a production frontier or a cost or profit frontier depends on the exogeniety assumptions. It is more appropriate to estimate a production function if inputs are exogeneous; a cost function if output is exogeneous; or a profit function if prices are different for different observations. In every case, technical, allocative, and scale efficiency for each observation can be estimated provided the underlying technology is self dual.

It is also possible to estimate a system of equations consisting of the production, cost, or profit function plus auxiliary equations that represent the first-order

conditions for whatever problem the firm will attempt to solve. Such an approach increases the efficiency of the parameter estimates by exploiting the cross-equation restrictions that implicitly arise because the parameters of the behavioral function that appear in the first-order conditions (Greene, 1980b, Schmidt and Lovell (1979,1980). This simultaneous equation estimation also allows computation of allocative and scale inefficiency directly within one system (Kumbhakar, Biswas, and Bailey). Of course, very strong assumptions must be made regarding the nature of association among disturbance terms of the different equations, which imposes very restrictive conditions on the nature of association between different types of inefficiencies that may be counter intuitive. In the case of a non-homothetic functional specification, this problem becomes further complicated (as further discussed in the second essay).

Starting with a Cobb-Douglas production function, Schmidt and Lovell (1979, 1980) considered a system of equations

$$y_i = A + \sum_j \alpha_j x_{ij} + \nu - v\,, \tag{13a}$$

$$x_{il} - x_{ij} = P_{ij} - P_{il} + \ln \alpha_l - \ln \alpha_j + \epsilon_{ij}\,, \tag{13b}$$

and

$$\ln C_i = K + \frac{1}{r} y_i + \sum_{j=1}^{k} \frac{\alpha_j}{r} P_{ij} - \frac{1}{r}(\nu - v) + \{E - \ln r\}\,, \tag{13c}$$

where y is the log output, X is the vector of log inputs, P is the vector of log prices, i indexes firms, and j indexes inputs. v represents the technical inefficiency and ϵ represents the allocative inefficency of the individual firm. The ν is a random disturbance term, $r = \sum_j \alpha_j$ is the returns to scale, and E is a well-defined function of ϵ and parameters. Equation (13a) is the production function, and equation (13b) represents the set of first-order conditions. Using these two relations, the cost function (13c) is derived. The cost of technical inefficiency is given by $r^{-1}v$, while the cost of allocative inefficiency is represented by the term $\{E_i - \ln r\}$.

Estimation of this system of equations requires a set of strong assumptions regarding the disturbance (inefficiency) terms. In their first model, Schmidt and Lovell assumed away any correlation between technical and allocative inefficiency, i.e., between v and ϵ. In their second model, they took into account the correlation between v and ϵ, which not only made the estimation complex but also

produced results not much different from that of the first model. Since then all simultaneous equations models have assumed this sort of independence between different inefficiency parameters. It is true that different distributional assumptions and different data sets could generate results contrary to that of Schmidt and Lovell (1979, 1980); this type of inter-inefficiency correlation can be established only in the case of homothetic functional specification (Schmidt and Lovell, 1980). In the case of nonhomothetic dual functions, such statistical relationships cannot be established (Greene 1980b), a point that will be discussed in detail in the context of the second essay.

Measures of all required inefficiency parameters can be arrived at without estimating the full system of equations by simply estimating the technology. Estimating only the stochastic production function (or cost function or profit function) one can get efficient estimates of the relevent parameters. For example, in the above cost function case, estimating (13c) we can obtain estimates of the technology parameters (α_j) and then estimates of the inefficiency parametes ϵ_j can be obtained from (13b), without using (13a) and (13c), directly. The only reason for estimating the whole system of equations simultaneously is to improve the precision of the parameter estimates. The disadvantage is that consistency then hinges on correct specification of the entire system. The possible efficiency gains occur because of the cross-equation restrictions, not just because of cross-equation error correlations that are very important to flexible functional forms.

Almost all empirical applications of stochastic frontier functions have used the maximum likelihood method for estimation. Consistent estimates of all parameters of the frontier function can be obtained by modifying the ordinary least squares estimators, a process called the corrected least square method (Schmidt, 1985-6), however, distribution of the one-sided error term is essentially asymmetric. Maximum likelihood estimators that make use of this information should be more efficient asymptotically. In the frontier literature, different assumption have been made regarding distribution of the error terms. In a simultaneous equation estimation process, assumptions regarding error terms and specification of relationships between error terms play crucial role. In this case, the full information maximum likelihood method or application of the seemingly unrelated regression method does not resolve the problem, although both methods yield maximum likelihood estimates when iteration converges. This is due to the asymmetric distribution structure of the error term. The most convenient method, therefore, is

to define the likelihood function for the system of equations and numerically (or analytically) maximize that function using any suitable iterative algorithm (see Greene [1982] for details of this method and possible algorithms.). Obtaining the maximum likelihood estimate of the parameters, one can obtain the inefficiency estimates (technical, cost or profit) for each observation following Jondrow, Lovell, Materov, and Schmidt.

Since publication of the Aigner, Lovell, and Schmidt paper on specification of the stochastic frontier production function, interest in efficiency measurement and frontier functions has grown rapidly. Two issues of the *Journal of Econometrics* have been devoted solely to the topic of frontier functions, which clearly shows rapid diversification in this area. Earlier studies of frontier production functions either used single equation production function or dual cost function. The direct estimation of the production function (non-optimizing) gives consistent estimates of the production function parameters only when inputs are treated as exogeneous. The Zellner, Kmenta, and Drèze argument of expected profit maximization can be used to treat inputs as exogeneous only if technical inefficiency is completely unknown to the firm. Since technical inefficiency includes poor management, composition and vintage of capital stock, local labour quality, and more, the assumption of unknown technical inefficiency may not be fully justified. If technical inefficiency is known to the firm (as proposed by Aigner, Lovell, and Schmidt), estimates of the production function parameters obtained directly from the production function will be inconsistent. This inconsistency can be avoided by following a simultaneous equation estimation method.

Most dual specifications for inefficiency estimation that use the stochastic specification have used the cost function,[5] which represents the constant output cost minimization process. Although simultaneous equation approaches have been followed (Schmidt and Lovell, 1979, 1980, Greene 1980b, Bauer) and both technical and allocative inefficiencies were estimated, it was not until Kumbhakar's 1987 paper that an unconstrained profit maximizing system was estimated. Although technical and allocative inefficiency have received much attention in the literature, scale inefficiency did not received its due importance. A firm can be inefficient due to any of these three inefficiencies or any combination thereof. Kumbhakar, Biswas, and Bailey estimated all three inefficiencies simultaneously.

[5] Ali and Flinn have used a stochastic profit function to estimate profit inefficiency. This is, no doubt, an unconstraint profit maximization process; however, they have used a single equation estimation method.

One basic shortcoming of the frontier literature is that flexible functional specifications have not been used in spite of the fact that the Cobb-Douglas specification is very restrictive and fails to capture the possibility of efficiency gains due to factor substitutions. Greene (1980b) first introduced the flexible functional forms in estimation of stochastic functional form. In a single equation framework, application of the flexible functional form works perfectly well with the usual economic and statistical qualifications of their homothetic counterparts. Only when estimating a system of equations with flexible functional forms, the economic relationship between behavioral function and first-order conditions become impossible to establish. This difficulty—called Greene's Problem— will be discussed in detail in the second essay, which attempts to estimate inefficiency bypassing the Greene's problem.

The purpose of these essays is to introduce flexibility in the measurement of inefficiencies. In the first essay, the conventional Cobb-Douglas production function is modified to take into account interactions between fixed factor and variable factors of production. This quasi-translog production function will then be used to derive all three types of inefficiencies and their impacts on the level of profit. In the second essay, a fully flexible normalized restricted profit function will be introduced to estimate profit inefficency and its underlying inefficiencies. The third essay is basically an extension of the second essay where profit inefficiency is estimated using an unrestricted translog profit function in a multiple output, multiple input framework.

A Brief Note on Efficiency of Indian Agriculture

The first two essays utilize a data set from Indian agriculture, the economic characteristics and efficiency aspects of which are highlighted in this section. Since the publication of T. W. Schultz's *Transforming Traditional Agriculture*,[6] studies on economic efficiency of Indian agriculture have gained momentum. These studies can be broadly classified into two categories of theoretical and empirical description, depending on whether a primal or dual approach is used.

In the primal approach, the production function, mostly Cobb-Douglas (CD), is directly estimated by the ordinary least square (OLS) technique. After obtain-

[6] Hopper actually pioneered the process. Results of his small but pioneering study form the basic evidence advanced by Schultz in support of his hypothesis on allocation efficiency in Indian agriculture.

ing the parameter estimates, marginal products (MP) of each endogeneous input are generally calculated at the geometric mean of the data. The presence of allocative efficiency is tested by equating the value of marginal products of inputs with their respective input prices. In the dual approach, on the other hand, the profit function is estimated along with input share (in profit) equations derived by using Hotelling's lemma. Since the profit function accommodates allocative inefficiency, the hypothesis of profit maximization can be tested by imposing parametric restriction on the profit function. Most studies using this approach have used the CD specification of the profit function.

Some Stylized Facts
About Indian Agriculture

The following are some basic facts and economic characteristics of Indian farming covered in different studies on Indian farming over the last thirty years.

1. *There exists an inverse relationship between farm size and productivity of land.*

Various farm management survey reports have supported this fact, including studies by Bhagwati and Chakravarti, Mazumder, Sen, Rao, Bardhan, Rudra, Saini, and Bhattacharyya and Saini who have tried to explain this phenomenon both on analytical and empirical levels. As usual, economists are divided on explaining this result but they support existence of the phenomenon. Rudra pointed out that an inverse relationship between land-holding size and productivity is due to a spurious relationship arising out of the process of aggregation over villages involved in developing farm management data. However, studies by Saini, Sampath, Bhattacharyya and Saini, based on disaggregated micro-level data for many regions and over several years, have confirmed this inverse relationship.

This point requires some elaboration mainly because efficiency analysis in our first two essays has been done separately for large and small farms, and log likelihood ratio tests show that structural differences exists between the large and the small farms. Sen explained the inverse relationship between productivity and the size of land holdings from dualistic nature of Indian agriculture, assuming that small farms are predominantly subsistence operators and large farms are predominantly capitalist operators. Thus, Sen argued that capitalist farms stop hiring inputs at the point where the marginal product of a factor is equal to its market hiring price. In contrast, small farms continue to employ inputs to the point where the marginal products of some inputs (mainly human labour) are almost zero or at least below the prevailing market wage rate. This argument suggests

that while large farms follow the profit maximization principle, small farms do not produce exactly for profit maximization.

Sen, of course, provided theoretical justification for his postulate. He argued that in the peasants' equilibrium, the marginal product of labour is determined by the pressure on the family for consumption and the family's opportunity for production. Bhagwati and Chakravarti questioned Sen's explanation, pointing out that if the two agrarian systems coexist, the opportunity cost of the peasant family labour may be the wage that the market offers for employment by capitalist farmers. Thus, if a family is making a decision on overall income derived from the input of work hours by the family, then the opportunity cost of work on both types of farms should be equalized. Empirically, it has been established that small farms hire labour at the margin and to a great extent join the farm labour force (Rosenzweig), so the opportunity cost of labour on the small farm is likely to be very real. Moreover, later studies (Lau and Yotopolous (1971, 1972), Yotopolous and Lau, and Kalirajan (1981a, 1981b) have suggested that small farms are both economic and profit efficient.

Bhagwati and Chakravarti provided an alternative theory for lower productivity of larger farms in terms of fragmentation of land holdings. They pointed out that although large farms have larger operational land holdings, these larger holdings may have accumulated through purchases of separate plots. Such fragmentation of cultivated land adversely affects the productivity per unit of land. Hanumantha Rao emphasized the adverse distributional impacts on Indian agriculture due to technological changes introduced through high yielding variety (HYV). The high cost of HYV led to distressed sales of land by small farmers, which caused further concentration of land ownership without consolidation of plots; however, no conclusive statistical evidence could be found to establish that point. Other studies (Kalirajan, 1981b, Parthasarathy and Prasad, Rajagopalan) have established that small farms have not lagged behind in deriving benefits from the use of HYV.

Whatever be the analytics, it has became an accepted fact of Indian agriculture that size of holding and yield per land unit are inversely related. This has very important policy implications for land reforms and has indeed affected policy makers' decisions regarding land reforms.

2. *When family labour employed in agriculture is given an imputed value in terms of ruling wage rate, much of Indian agriculture seems unremunerative.*

Sen propagated this point, though empirical eveidence based on data collected in the 1970s shows that this may not be true (Sampath, Rosenzweig). Almost all studies on economic efficiency of Indian farms have used the market wage rate to evaluate family labour, and their results show that small farms are no less efficient than their larger counterparts.

3. *By and large, the profitability of Indian agriculture increases with the size of land holdings.*

Studies based on data from Punjab support this view, but generalization to all of India does not hold, at least based on data collected during the 1960s. Small farms are found to be no less profit efficient than large farms (Lau and Yotopolous (1971, 1972), Yotopolous and Lau, Kalirajan, 1981a, 1981b). In fact, studies suggest that small farms are more profit efficient than their larger counterparts. Studies based on data collected from southern India show that small farms are more profit efficient than their larger counterparts.

4. *Fragmentation decreases with an increase in the size of land holdings.*

This point, as emphasized by Bhradwaj, actually depends on the regions of India from which data were collected. Various parts of India have different land management systems and ways of controlling property rights to land.

5. *The intensity of cropping is inversely related to the size of holdings.*

In the Indian situation, in most cases farming is the only source of income for small farms. Farmers with large land holdings are also engaged in other types of economic activities; as a result, they prefer to lease out land during the dry season.

6. *There is generally an inverse relation between earners per hectare and the size of holdings.*

7. *Labour days spent per hectare are inversely related to the size of holdings.*

The above mentioned so called "stylized facts" of Indian agriculture are neither mutually exclusive nor collectively exhaustive. Most are related to one another, and existence of one leads to existence of the other. Reviewing these points summarize the fundamental characteristics of the Indian farming system before providing the major results of studies on efficiency aspects of Indian agriculture, as reviewed in the following section.

Review of Past Studies on Economic Efficiency of Indian Farms

Empirical studies on allocative efficiency in Indian agriculture can be divided into two categories based on whether a primal or a dual approach is used. The major findings of previous studies which followed the primal approach[7] are discussed in the next section, with results of the studies which followed the dual method are discussed in the following subsection.

The Primal Approach

Among the major studies, Hopper, Saini, and Sampath found Indian farms to be allocatively efficient by estimating the CD production function using OLS. The presence of allocative efficiency was tested by equating the value of the marginal product (at the geometric mean) of each input to its market price. An extension of this approach is considered in Ram, who analyzed allocative inefficiency in terms of education and extension services and found significant positive effects of these variables on allocative efficiency. Researchers like Dey and Rudra and Hati and Rudra, using the CD technology representation, concluded that Indian farmers are not profit maximizers. Under both constrained maximization and unconstrained maximization, they could not conclusively establish that Indian farmers were allocating resources optimally.

Problems with the use of a primal approach are basically twofold. First, OLS estimates of production function parameters are likely to be biased and inconsistent if OLS is used directly to estimate a production function, ignoring the fact that some inputs are endogeneous and, therefore, correlated with the error term in the production function (Nowshirvani). Thus, estimates of allocative inefficiency obtained from these parameters may be inappropriate. Second, use of a restrictive functional form like CD, which assumes unit factor substitution, biases estimates of the MPs and hence the allocative inefficiencies.

Studies using the CD functional specification have defined the average productivity of factors by ratio of geometric mean of output to geometric mean of input. There is a basic contradiction in the use of geometric mean. It divides the whole set of observations into two groups, namely, one group that overallocates the resource and another that underallocates.

[7] Some example of studies in this line are Hopper, Chennareddy, Saini, Dey and Rudra, and Sampath.

The Dual Approach

Studies that used the dual approach (mostly the profit function) can be divided into two groups, depending on whether the Cobb-Douglas (CD) specification or a flexible functional form specification was used. Lau and Yotopolous (1971, 1972), Wise and Yotopolous, and Yotopolous and Lau used the CD specification, while Junkunkar (1979, 1980), Kalirajan (1981a, 1981b), Sidhu, and Kumbhakar and Bhattacharyya applied flexible functional specifications (mostly translog). The advantage of using the dual approach is that price variables are exogenous to the system; consequently, no inconsistency occurs even if a single equation estimation method is used. However, efficiency of the estimators can be improved by using the input demand and output supply functions (or share equations in the case of the translog specification) derived from the profit function. This approach assumes that farms behave according to a decision rule, say profit maximization, subject to a set of exogenous variables such as prices and fixed factors of production.

Two sets of problems were addressed by these studies. Given different price regimes of variable factors of production and quantities of fixed factors of production, do the farms behave according to a decision rule, say profit maximization? If the profit maximization rule is generally applicable, then are there inter-farm group differences in the levels of economic efficiency?

The dual approach, first applied by Wise and Yotopolous compared the systematic with the random element of farm behavior to determine whether or not to examine economic efficiency of Indian farms. They found that Indian farms not only behave rationally but are also remarkably price efficient. Studies by Yotopolous and Lau, and Lau and Yotopolous (1971, 1972) tried to capture not only allocative but also technical efficiency of Indian farms. They concluded that, given the fixed factors of production and within the ranges of observed prices of output and variable inputs, small farms have higher actual profit and higher allocative and/or technical efficiency.

The methodology used by Lau and Yotopolous (1971, 1972) and Yotopolous, Lau, and Somel is no doubt commendable, but their empirical analysis was based on very few observations (34) drawn from aggregate size class averages, which introduced aggregation bias in the estimates. Moreover, they used the CD specification for the profit function which, as already mentioned, is too restrictive to measure allocative inefficiency.

Kalirajan (1981a, 1981b) and Kumbhakar and Bhattacharyya using the translog specification of the profit function, estimated the full system of equations to examine the economic efficiency of Indian farms. In both studies, Kalirajan concluded that farmers growing HYVs are efficient economically regardless of farm size. His results support the findings of Ghosh who found no significant differences between small and large farms provided both are given the same improved technology. Junnunkar (1989), on the other hand, questioned the applicability of the profit function in the context of Indian agriculture and argued that the hypothesis that Indian farmers are profit maximizers is incorrect. Using CD profit function, "... the results we get with wrong signs on (most) variable input prices imply that farmers do not behave in consistent maximizing fashion" (Junnunkar 1980, p. 58). He further pointed out that a simple static model of behavior under competitive conditions does not apply to Indian farmers.

Kumbhakar and Bhattacharyya addressed Junnunkar's problem by developing a shadow profit function that takes account for different market distortions introduced by institutional and socio-cultural factors. The model implicitly internalized these distortions through shadow prices. Using the shadow profit function, they tested the validity of neoclassical profit maximization behavior and also related allocative inefficiency to some explanatory variables. The major results of this study indicated a positive correlation between the level of education and allocative efficiency of a farm. Both small and large farms were inefficient in allocating outputs in response to market prices. Finally, farms are found to be maximizing their shadow profit but not their actual profit. This study extended the results of Ram and Huffman,who considered the allocative implication of education regarding the use of modern inputs like chemical fertilizer. Kumbhakar and Bhattacharyyas' study concludes that education also enables farmers to allocate their conventional inputs, like human and bullock labour, more efficiently.

The present study looks at inefficiency of Indian farms using stochastic flexible functional forms under the neoclassical hypothesis of profit maximization. This approach features three advantages over earlier studies. First, the use of stochastic functional specifications allows more precise measurement of farm specific inefficiencies. Introduction of flexible functional forms gives a better measure of inefficiencies since no *a priori* structure is assumed. Finally, all three types of inefficiencies can be estimated and their impact on level of profit can be examined.

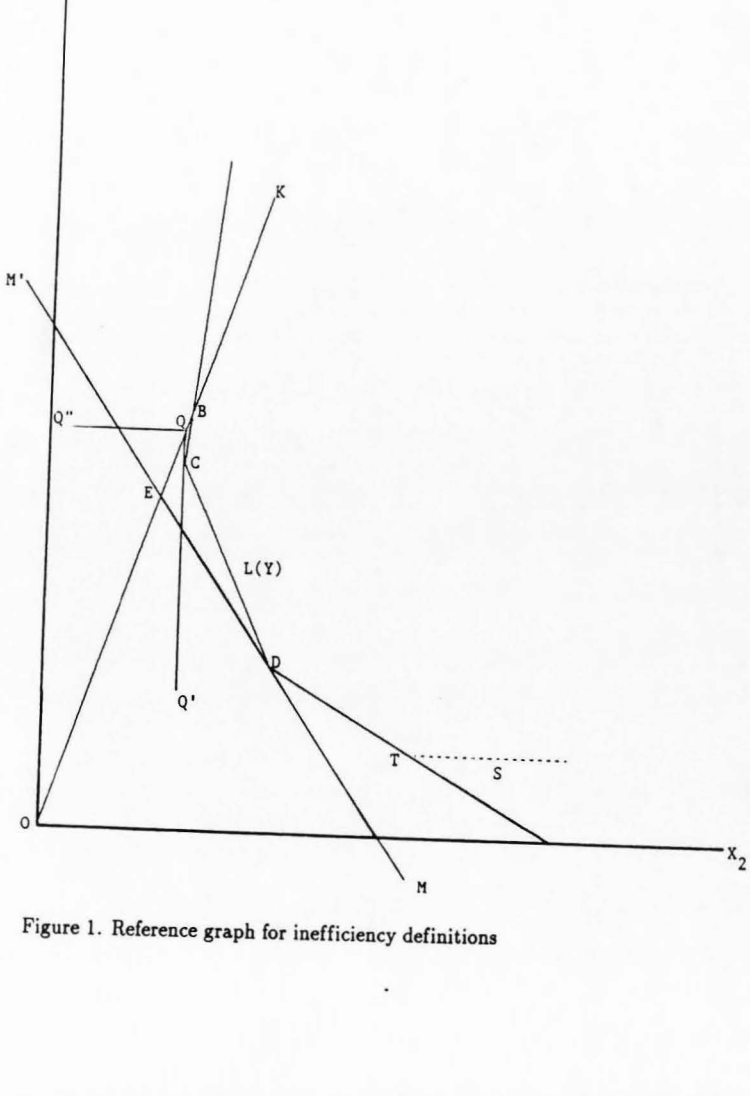

Figure 1. Reference graph for inefficiency definitions

References

Afriat, S. N. "Efficiency Estimation of Production Functions." *Int. Econ. Rev.* 13(1972):568-98.

Aigner, D. J., and S. F. Chu. "On Estimating the Industry Production Function." *Amer. Econ. Rev.* 58(1968):826-39.

Aigner, D. J., C. A. K. Lovell, and P. Schmidt. "Formulation and Estimation of Stochastic Frontier Production Function Models." *J. Econometrics* 6(1977):21-37.

Ali, M., and J. C. Flinn. "Profit Efficiency Among Basmati Rice Producers in Pakistan Punjab." *Amer. J. Agr. Econ.* 71(1989):302-10.

Bardhan, P. K. "Size, Productivity and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture." *J. Polit. Econ.* 81(1973):1370-86.

Bauer, P. W. *An Analysis of Multiproduct Technology and Efficiency Using the Joint Cost Function and Panel Data.* Ph. D. Dissertation, University of North Carolina at Chapel Hill, 1985.

Bhagwati, J. N., and S. K. Chakravarti. "Contribution to Indian Economic Analysis: A Survey." *Amer. Econ. Rev.* 59(1969):1-72.

Bhattacharyya, N., and G. R. Saini. "Farm Size and Productivity: A Fresh Look." *Econ. and Polit. Weekly* 7(1972):63-72.

Bhradwaj, K. *Production Condition of Indian Agriculture.* Cambridge: Cambridge University Press, 1974.

Caves, R., and D. Barton. *Efficiency in U.S. Manufacturing Industries*, Mass: MIT Press, 1990.

Chennareddy, V. "Production Efficiency in Traditional Indian Agriculture." *J. Farm Econ.* 49(1967):816-20.

Debreu, G. "The Coefficient of Resource Utilization." *Econmertica* 19(1951):173-92.

Dey, A. K., and A. Rudra. "A Test of Hypothesis of Rational Allocation of Resources Under Cobb-Douglas Technology." *Econ. and Polit. Weekly* 8(1973):609-12.

Dugger, R. *An Application of Bounded Nonparametric Estimating Functions to the Analysis of Bank Cost and Production Functions.* Ph.D. Dissertation, University of North Carolina at Chapel Hill, 1974.

Eichhorn, W. "What is an Economic Index? An Attempt of an Answer." *Theory and Applications of Economic Indices*, eds. W. Eichhorn, R. Henn, O. Optiz, and R. W. Shephard, Würzburg: Physica-Verlag, 1978a.

——. *Functional Equation in Economics.* Reading, Mass.: Addison-Wesley, 1978b.

Färe, R., and S. Grosskopf. "Measuring Output Efficiency." *European J. Operational Research* 13(1983a):173-79.

——. "Measuring Congestion in Production." *Zeitschrift für Nationalökonomie*

43(1983b):257-71.

Färe, R., S. Grosskopf, and C. A. K. Lovell. *Measurement of Efficiency of Production.* Boston: Kluwer-Nijhoff Publishing, 1985.

Färe, R., and L. Svensson "Congestion of Production Factors." *Econometrica* 48(1980) :1745-53.

Farrell. J. M. "The Measurement of Productive Efficiency." *J. Royal Statist. Soc.* Series A(general) 21(1957):253-81.

Farrell, J. M., and M. Fieldhouse. "Estimating Efficient Production Under Increasing Returns to Scale." *J. Royal Statist. Soc.* Series A(general) 125(1962):252-67.

Forsund, F. R., and L. Hjalmarsson. "On the Measurement of Productive Efficiency." *Swedish J. Econ.* 76(1974):141-54.

——. "Generalized Farrell Measure of Efficiency: An Application to Milk Processing in Swedish Dairy Plants." *Econ. J.* 89(1979):274-315.

Forsund, F. R., and E. S. Jansen. "On Estimating Average and Best Practice Homothetic Production Function Via Cost Function." *Int. Econ. Rev.* 18(1977):463-72.

Ghosh, A. K. "Farm-Size and Land Productivity in Indian Agriculture: A Reappraisal." *J. Dev. Studies* 16(1979):27-49.

Greene, W. H. "Maximum Likelihood Estimation of Econometric Frontier Functions." *J. Econometrics* 13(1980a):27-56.

——. "On the Estimation of a Flexible Frontier Production Model." *J. Econometrics* 13(1980b)101-15.

——. "Maximum Likelihood Estimation of Stochastic Frontier Production Model." *J. Econometrics* 18(1982):285-89.

Hanumantha, Rao. *Technological Change and the Distribution of Gains in Indian Agriculture.* New Delhi: Macmillan, 1975.

Hati, A. K., and A. Rudra. "Calculation of Efficiency Indices for Farmers: A Numerical Exercise." *Econ. and Polit. Weekly* 8(1973):A17-A26.

Hopper, W. D. "Allocative Efficiency in a Traditional Indian Agriculture." *J. Farm Econ.* 47(1965):611-24.

Huffman, W. E. "Allocative Efficiency: The Role of Human Capital." *Quart. J. Econ.* 47(1977):59-79.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *J. Econometrics* 23(1982):269-74.

Jununkar, P. N. "The Test of Profit Maximization Hypothesis: A Study of Indian Agriculture." *J. Dev. Studies* 16(1979):186-203.

——. "Do Indian Farmer Maximize Profit?" *J. Dev. Studies* 17(1980):48-61.

——. "The Response of Peasant Farmers to Price Incentives: The Use and Misuse of

Profit Function." *J. Dev. Studies* 25(1989):169-72.

Kalirajan, K. "Testing Hypothesis of Equal Relative Economic Efficiency Using Restricted Aitken's Least Square Estimation." *J. Dev. Studies.* 17(1981a):307-16.

——. "The Economic Efficiency of Farmers Growing High-Yielding, Irrigated Rice in India." *Amer. J. Agr. Econ.* 63(1981b):566-70.

Koopmans, T. C. An Analysis of Production as an Efficient Combination of Activities." *Activity Analysis of Production and Allocation,* ed. T. C. Koopmans. Cowels Commission for Research in Economics. Monograph No. 13, New York: John Wiley and Sons, 1951.

——. *Three Essays on the State of Economic Science.* New York: McGraw Hill Book Company, 1957.

Kumbhakar, S. C. "The Specification of Technical and Allocative Inefficiency in Stochastic Production and Profit Frontiers." *J. Econometrics* 34(1987):335-48.

Kumbhakar, S. C., and A. Bhattacharyya. "Education Farm-Size and Allocative Efficiency in Indian Agriculture: A Restricted Profit Function Approach." Working Paper # 89-04, Utah State University, Utah, 1989.

Kumbhakar, S. C., B. Biswas, and D. V. Bailey. "A Study of Economic Efficiency of Utah Dairy Farmers: A System Approach." *Rev. Econ. Statis.* 71(1989):595-604.

Lau, L. J., and Pan A. Yotopolous,"A Test of Relative Economic Efficiency and Application to Indian Agriculture."*Amer. Econ. Rev.* 61(1971):94-109.

——. "Profit, Supply and Factor Demand Functions." *Amer. J. Agr. Econ.* 54(1972):11-18.

Mazumder, D. "Size of Firm and Productivity: A Problem of Indian Peasant Agriculture." *Economica* 32(1965):161-73.

Meeusen, W., and J. van den Broeck. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Errors." *Int. Econ. Rev.* 18(1977):435-44.

Nowshirvani, V. "Allocation Efficiency in a Traditional Indian Agriculture: Comment." *J. Farm Econ.* 49(1967):218-21.

Parthasarathy, G., and D. S. Prasad. "Response to the Inpact of the New Rice Technology by Farm Size and Tenure–Andhra Pradesh, India." *Interpretive Analysis of Selected Papers from Changes in Rice Farming in Selected Areas of Asia,* ed. International Rice Research Institute, pp. 111-27, Los Banos, Philippines: IRRI, 1978.

Rajagopalan, V. "North-Arcot, Tamil Nadu." *Changes in Rice Farming in Selected Areas of Asia,* ed. International Rice Research Institute, pp. 71-91, Los Banos, Philippines: IRRI, 1975.

Ram, R. "Role of Education in Production: A Slightly New Approach." *Quart. J. Econ.* 95(1980):365-73.

Rao, A. P. "Size of Holding and Productivity." *Econ. Polit. Weekly* Nov(1964):1989-91.

Richmond, J. "Estimating the Efficiency of Production." *Int. Econ. Rev.* 15(1974):515-21.

Rosenzweig, M. R. "Agricultural Development, Education and Innovation." *The Theory and Experience of Economic Development: Essays in Honour of Sir W. Arther Lewis*, eds. M. R. Rosenzweig, C. F. Diaz-Alejandro, Gustav Ranis and M. Gsovitz. pp. 107-34, London: George Allen & Unwin, 1982.

Rudra, A. "Allocative Efficiency of Indian Farmers-Some Methodological Doubts." *Econ. and Polit. Weekly* 8(1973):107-12.

Saini, G. R. "Holding Size, Productivity, and Some Related Aspects of Indian Agriculture" *Econ. and Polit. Weekly.* 6(1971):79-84.

Sampath, R. K. *Economic Efficiency of Indian Agriculture: Theory and Measurement.* New Delhi :Macmillan, 1979.

Schmidt, P. "On Statistical Estimation of Parametric Frontier Production Functions." *Rev. Econ. Statist.* 58(1976):238-39.

——. "Frontier Production Functions." *Econometric Rev.* 4(1985-86):289-328.

Schmidt, P., and C. A. K. Lovell. "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontier." *J. Econometrics* 9(1979):343-66.

——. "Estimating Stochastic Production and Cost Frontiers When Technical and Allocative Efficiencies are Correlated.' *J. Econometrics* 13(1980):83-100.

Schultz, T. W. *Transforming Traditional Agriculture.* New Haven: Yale University Press, 1964.

Seitz, W. D. "Productive Efficiency in the Steam-Electricity Generating Industry." *J. Polit. Econ.* 79(1971):876-86.

Sen, A. K. "Peasants and Dualism With or Without Surplus Labour." *J. Polit. Econ.* 74(1966):425-50.

Sidhu, S. S. "Relative Economic Efficiency in Wheat Production in the Indian Punjub." *Amer. Econ. Rev.* 63(1974):742-21.

Stevenson, R. E. "Likelihood Functions for Generalized Stochastic Frontier Estimation." *J. Econometrics* 13(1980):57-66.

Timmer. C. P. "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency." *J. Polit. Econ.* 79(1971):776-94.

Wise, J., and Pan A. Yotopolous. "A Test of Hypothesis of Economic Rationality in a Less Developed Economy." *Amer. J. Agr. Econ.* 50(1968):395-97.

Yotopolous, Pan A., and L. Lau, "A Test of Relative Economic Efficiency and Application to Indian Agriculture: Some Further Results." *Amer. Econ. Rev.* 63(1973):214-23.

Yotopolous, Pan A., L. Lau, and K. Somel. "Labor Intensity and Relative Efficiency in

Indian Agriculture." *Food Res. Inst. Studies in Agri. Econ., Trade Dev.* 9(1970)43-55.

Zellner, A., J. Kmenta, and J. Drèze. "Specification and Estimation of Cobb-Douglas Functions." *Econometrica* 34(1966):784-95.

CHAPTER II
SPECIFICATION AND ESTIMATION OF PRODUCTION INEFFICIENCY USING A QUASI-TRANSLOG STOCHASTIC PRODUCTION FUNCTION

Earlier empirical applications of the frontier production function model have typically used the restrictive Cobb-Douglas functional specification of the production technology; however, it is likely that the restrictive nature of technology affects measures of productive inefficiency. In this essay, the main objective is to introduce limited cross effects (or limited flexibility) in a functional (technology) specification and to specify all three types of efficiencies as previously defined, after which the model will be applied to a data set on Indian agriculture. In order to estimate a simultaneous equation model, a general optimization model must first be specified for analysis.[1] The next section develops the general optimization model, and section three introduces inefficiency in the model where the required specification for estimation is provided. A development of estimation procedure is followed by section five, which describes the data set. This is followed by the discussion of results obtained from empirical application of the model. Finally, a summary of the major findings, limitations, and possibilities of further extension of this study are given.

General Model

In any production system, the fixed inputs set the limit (in the short run) on output and possibilities of factor substitution. The interaction between fixed inputs, $Z\,(=\, z_1,\ldots,z_J)$, and variable inputs, $X\,(=\, x_1,\ldots,x_n)$, is, therefore, important for any production analysis. To capture this interaction and to make the production function specification more flexible than the conventional Cobb-Douglas (CD), a semi-translog production function is introduced as

$$(1) \qquad y(X,Z) = a\prod_i x_i^{\alpha_i}\prod_j z_j^{\beta_j}\exp\Big(\sum_i\sum_j \gamma_{ij}\ln x_i \ln z_j\Big),$$

where y is the output, $i = 1,\ldots,n$, and $j = 1,\ldots,J$. Given this technology, we assume that the firm is a profit maximizer. This condition can be obtained in two ways. The first method is termed the unconstrained (direct) profit maximization technique where

$$\Pi(W,P,Z) = P'Y - W'X$$

[1] Single equation estimation does not require any optimizing condtion.

is maximized. Where Π is nominal profit, P is a $(m \times 1)$ output price vector, Y is a $(m \times 1)$ output vector, W is a $(n \times 1)$ vector of hiring prices of variable inputs, and X is a $(n \times 1)$ vector of variable inputs. Z is a $(J \times 1)$ fixed input vector. In the alternative two-step method, the first step derives the optimum input demands using cost minimization, which are constant output input demands. To obtain the profit maximizing solution, solve the profit maximizing output supply. The second step obtains the output supply conditions using the first-order conditions of cost minimization. Information generated from these two steps is equivalant to that of the unconstrained maximization process. This study follows the second approach because the derivation of the cost of allocative inefficiency is easier and enables the use of the Schmidt and Lovell (1979, 1980) specification to measure the cost of allocative inefficiency. The optimizing problem, therefore, is

$$\begin{aligned} (2) \qquad & \min_{x} \quad C(W, y, Z) = W'X \\ & \text{subject to} \qquad y = f(X, Z), \end{aligned}$$

where $f(X, Z)$ is the production function, i.e., equation (1).

From the set of first-order conditions, the equilibrium condition

$$MRTS_{ks} = \left(\frac{w_k}{w_s}\right) \qquad \forall\, k,\ s = 1, \ldots, n;\ \ k \neq s,$$

is derived as

$$(3) \qquad \frac{w_k}{w_s} = \frac{x_s}{x_k} \frac{(\alpha_k + \sum_j \gamma_{kj} \ln z_j)}{(\alpha_s + \sum_j \gamma_{sj} \ln z_j)},$$

where w_k and w_s are kth and sth elements of the W vector. Taking logarithms of both sides of (3) obtains

$$(4) \qquad \ln x_k = \ln x_s - \ln w_k + \ln w_s + \ln \left\{ \frac{\alpha_k + \sum_j \gamma_{kj} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j} \right\}.$$

From (4), k logarithms of input demands equations can be solved in terms of input, x_s. Substituting k equations (4) in the constraint equation and solving for $\ln x_s$ produces

$$\begin{aligned} (5) \qquad \ln x_s = {} & \theta \ln y - \theta \ln \left\{ a \prod_i \left\{ \frac{\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j} \right\}^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \right\} \\ & + \theta \sum_i \left\{ \alpha_i + \sum_j \gamma_{ij} \ln z_j \right\} \ln \left(\frac{w_i}{w_s} \right) - \theta \sum_j \beta_j \ln z_j, \end{aligned}$$

where

$$\theta = \frac{1}{\sum_i \alpha_i + \sum_i \sum_j \gamma_{ij} \ln z_j}.$$

Taking the antilog, the solution input demand equation for input s is given as

$$(5a) \qquad x_s^* = y^\theta \Big\{ a \prod_i \Big\{ \frac{\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sk} \ln z_j} \Big\}^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \Big\}^\theta$$

$$\prod_i \Big(\frac{w_i}{w_s} \Big)^{\theta(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \prod_j z_j^{-\theta\beta_j}.$$

Substituting (5) in (4), $\ln x_k$ is given as

$$\ln x_k = \theta \ln y + \ln \Big\{ \frac{w_s}{w_k} \prod \Big(\frac{w_i}{w_s} \Big)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)\theta} \Big\} - \sum_j \theta\beta_j \ln z_j$$

$$(6) \qquad + \ln \Big\{ \frac{\alpha_k + \sum_j \gamma_{kj} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j} \Big\{ a \prod_i \Big(\frac{\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j} \Big)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \Big\}^{-\theta} \Big\}.$$

The antilog of the above equation is

$$x_k^* = y^\theta \Big\{ \frac{\alpha_k + \sum_j \gamma_{kj} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j} \Big\{ a \prod_i \Big(\frac{\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j} \Big)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \Big\}^{-\theta} \Big\} \times$$

$$(6a) \qquad \frac{w_s}{w_k} \prod_i \Big(\frac{w_i}{w_s} \Big)^{\theta(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \prod_j z_j^{-\theta\beta_j}.$$

The set of equations (5a) and (6a) are the constant output input demand equations. Substituting this set of equations in the objective function, i.e., of problem (2), the cost minimizing cost function becomes

$$C^*(W, y, Z) = W' X^*,$$

where X^* is the solution input vector. The logarithm of the solution cost function can be expressed as

$$\ln C^*(W, y, Z) = \theta \ln y - \theta \ln \Big\{ a \prod_i \Big(\frac{\alpha_i + \sum_j \gamma_{ik} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j} \Big)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \Big\}$$

$$(7) \qquad + \theta \sum_i \Big(\alpha_i + \sum_j \gamma_{ij} \ln z_j \Big) \ln \Big(\frac{w_i}{w_s} \Big) + \ln w_s - \theta \sum_j \beta_j \ln z_j$$

$$+ \ln\left\{\frac{\sum_i(\alpha_i + \sum_j \gamma_{ij} \ln z_j)}{\alpha_s + \sum_j \gamma_{sj} \ln z_j}\right\}.$$

Taking the antilog of the above equation, the solution cost function is

$$C^*(W, y, Z) = y^\theta\left\{a \prod_i \left\{\frac{\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j}\right\}^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)}\right\}^{-\theta} w_s \prod_j z_j^{-\theta\beta_j} \times$$

$$\prod_i \left(\frac{w_i}{w_s}\right)^{\theta(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \left\{1 + \frac{\sum_k(\alpha_k + \sum_j \gamma_{kj} \ln z_j)}{\alpha_s + \sum_j \gamma_{1j} \ln z_j}\right\}. \tag{7a}$$

This cost function is a constant output cost-minimizing solution. At the optimum, the firm supplies the optimum (profit-maximizing) output when

$$\frac{\partial C^*(W, y, Z)}{\partial y} \equiv MC = p \tag{8}$$

Marginal cost (MC) is given by

$$\frac{\partial C^*(W, y, Z)}{\partial y} = \theta y^{\theta-1}\left\{a \prod_i \left(\frac{\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j}\right)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)}\right\}^{-\theta} w_s \times$$

$$\prod_i \left(\frac{w_i}{w_s}\right)^{\theta(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \prod_j z_j^{-\theta\beta_j} \left\{\frac{\sum_i(\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j}\right\}. \tag{9}$$

Equating the logarithm of (9) to $\ln p$ and solving for $\ln y$ provides the logarithm of the output supply as

$$\ln y = \frac{1}{\theta - 1} \ln p + \frac{1}{\theta - 1}\left\{\ln w_s + \theta \sum_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j) \ln\left(\frac{w_i}{w_1}\right)\right\}$$

$$- \frac{1}{\theta - 1}\left\{\ln\theta + \ln\left\{a \prod_i \left(\frac{\alpha_i + \sum_j \gamma_{ij} \ln z_j}{\alpha_s + \sum_j \gamma_{sj} \ln z_j}\right)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)}\right\}^{-\theta}\right\} \tag{10}$$

$$- \frac{1}{\theta - 1}\left\{\theta \sum_j \beta_j \ln z_j + \ln\left\{\frac{\sum_i(\alpha_i + \sum_j \gamma_{ij} \ln z_j)}{\alpha_s + \sum_j \gamma_{sj} \ln z_j}\right\}\right\}.$$

This system of equations [(5a), (6a), and (10a)] or [(5), (6), and (10)] gives full profit maximizing solutions. Parameters of this set of equations can be estimated

by full information maximum likelihood (FIML) or the seemingly unrelated regression method (SUR), both of which yield maximum likelihood estimates when iteration converges.

Inefficiency Specification

Following Aigner, Lovell, and Schmidt and Meeusen and van den Broeck, the composite error structure is introduced into the production function (1). The production function can be related to the stochastic frontier production function by specifying the intercept term a as

$$a = \alpha_0 \exp(\tau), \tag{11}$$

where $\tau \leq 0$. The stochastic variant of the production function (1) becomes

$$y = \alpha_0 \prod_i x_i^{\alpha_i} \prod_j z_j^{\beta_j} \exp\Big(\sum_i \sum_j \gamma_{ij} \ln x_i \ln z_j + \nu + \tau\Big). \tag{1a}$$

The error component ν represents the symmetric statistical white noise that is beyond the firm's control. The τ is the technical efficiency parameter that varies across firms but is not observable. If $\tau = 0$, then the firm is on the frontier, producing the maximum possible level of output given the technology. The resultant specification of the production function is stochastic because $y \leq f(x_i, z_j; \alpha_0, \alpha_i, \beta_j, \gamma_{ij}) + \nu$. The economic logic is that the production process is subject to two economically distinguishable random disturbances with different characteristics, both economic and statistical. The non-positive disturbance τ reflects the fact that each firm's output must lie on or below its frontier $y \leq f(x_i, z_j; \alpha_0, \alpha_i, \beta_j, \gamma_{ij}) + \nu$. Any such deviation results from factors under the firm's control, such as technical and economic inefficiency caused by poor management, the composition and vintage of capital stock, or local labour quality, but the frontier itself can vary randomly across firms or over time for the same firm. Given this interpretation, the function is stochastic, with random disturbance term, ν, resulting from favourable and unfavourable external events such as luck, climate, and machine performance. Errors of observation and measurement on y constitute another source of variation in the random variable ν. Technical inefficiency can be due either to excessive input use (for a given level of output following Farrell) or to less output produced relative to the optimum for a given input combination (as defined in stochastic frontier literature). Either

way, technical inefficiency is costly. Since cost is not minimized, profit is not maximized.

Allocative inefficiency results from employing inputs in incorrect proportions. Following Schmidt and Lovell (1979, 1980), a firm is allocative inefficient if the marginal rate of technical substitutions between factors differs from the ratio of their hiring prices, i.e., $\text{MRTS}_{qs} \neq (w_q / w_s)$. This can be specified as

$$\frac{f_q}{f_s} = \left\{\frac{w_q}{w_s}\right\} \exp(u_q) \quad \forall\, q,\ s = 1, \ldots, n;\ q \neq s; \tag{12}$$

or

$$\exp(u_q) = \left\{\frac{f_q}{f_s}\frac{w_s}{w_q}\right\}; \qquad \text{for} \quad u_q \gtrless 0. \tag{12a}$$

Given the stochastic production function (1a), the marginal productivity of the ith input is given by

$$f_i = \frac{\partial y}{\partial x_i} = MP_i = \frac{y}{x_i}\left(\alpha_i + \sum_j \gamma_{ij} \ln z_j\right). \tag{13}$$

Therefore, allocative inefficiency of the qth input relative to the sth input can be expressed as

$$\exp(u_q) = \left\{\frac{x_s}{x_q}\frac{w_s}{w_q}\right\}\left\{\frac{(\alpha_q + \sum_j \gamma_{qj} \ln z_j)}{(\alpha_s + \sum_j \gamma_{sj} \ln z_j)}\right\}, \quad \text{for} \quad u_q \gtrless 0. \tag{14}$$

Observed expenditure $W'X$, therefore, coincides with the minimum cost $C^*(W, y^*, Z)$, and/or observed profit Π coincides with the maximum profit Π^* if, and only if, the firm is both technically and allocatively efficient. If a difference exists between observed and optimum (cost and/or profit), this difference may be due to technical inefficiency alone, allocative inefficiency alone, or some combination of allocative and technical inefficiencies. Observed input demands will be equal to the optimum input demand if the firm is both technically and allocatively efficient. A combination of technical and allocative inefficiencies causes the firm to deviate from its optimum. Therefore, any inefficiency study should simultaneously take into account both technical and allocative inefficiency.

Technical and allocative efficiency are necessary but not sufficient for profit maximization. A firm could still be scale inefficient if MC $\neq$ p. This can be

specified as

$$\frac{\partial C^*(W, y, Z)}{\partial y} \neq p. \tag{15}$$

Since there is no inefficiency at the optimum, the cost function can be represented by equation (9). MC, therefore, is

$$\frac{\partial C^*}{\partial y} = \theta\, y^{\theta-1} \prod_i w_i^{\theta(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \prod_j z_j^{(-\theta\beta_j)} \Big\{ \sum_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j) \Big\} \times$$

$$\Big\{ \Big\{ \alpha_0 \prod_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \Big\}^{-\theta} \Big\}. \tag{16}$$

The firm is scale inefficient if

$$\exp(\xi) = \left\{ \frac{1}{p} \frac{\partial C^*}{\partial y} \right\} \quad \text{where } \xi \lesseqgtr 0. \tag{17}$$

The parameter ξ measures the deviation of observed output supply from the profit-maximizing output supply. If the scale efficiency parameter, ξ, is zero, then the firm is supplying the profit-maximizing level of output, given p. ξ is positive (negative) if the firm is supplying more (less) than the optimum level. Equation (17) can be expressed as

$$p \exp(\xi) = \theta\, y^{\theta-1} \Big\{ \sum_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j) \prod_i w_i^{\theta(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \prod_j z_j^{(-\theta\beta_j)} \Big\}$$

$$\Big\{ \Big\{ \alpha_0 \prod_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \Big\}^{-\theta} \Big\}. \tag{18}$$

Taking logarithms of equations (1a), (14), and (18), the following system of equations is obtained:

$$\ln y = \ln \alpha_0 + \sum_i \alpha_i \ln x_i + \sum_j \beta_j \ln z_j + \sum_i \sum_j \gamma_{ij} \ln x_i \ln z_j + \nu + \tau; \tag{1b}$$

$$\ln x_s - \ln x_q = \ln w_q - \ln w_s + \ln\Big\{ \alpha_s + \sum_j \gamma_{sj} \ln z_j \Big\} - \ln\Big\{ \alpha_q + \sum_j \gamma_{qj} \ln z_j \Big\} + u_q; \tag{14a}$$

$$(18a) \qquad \ln y = (\theta - 1)^{-1} \ln p - (\theta - 1)^{-1} \ln \theta + \ln K - \theta \sum_j \beta_j \ln z_j + \theta \sum_i \Big(\alpha_i + \sum_j \gamma_{ij} \ln z_j \Big) \ln w_i - \xi ,$$

where

$$\theta = (r + k); \qquad r = \sum_i \alpha_i ; \qquad k = \sum_i \sum_j \gamma_{ij} \ln z_j ; \qquad \text{and}$$

$$K = \sum_i \Big\{ \alpha_i + \sum_j \gamma_{ij} \ln z_j \Big\} \Big\{ a(\alpha_i + \sum_j \gamma_{ij} \ln z_j)^{\left(\alpha_i + \sum_j \gamma_{ij} \ln z_j \right)} \Big\}^{-\theta} .$$

The system of equations (16a), (14a), and (18a) involve $(n + 1)$ equations to solve for $(n + 1)$ unknowns ($\ln Y$ and $\ln x_i$; $\forall\, i = 1, \ldots, n$). A solution to these $(n + 1)$ equations gives n unconditional input demand equations and output supply equations. The conditional input demand equations can be obtained from the first n equations, and the output supply function can be obtained from the last equation. To be economically efficient, all three efficiency parameters (τ, ξ and u_q) must be equal to zero, both individually and simultaneously (Forsund, Lovell, and Schmidt).

Estimation Method

Direct estimation of the production function gives consistent estimates only when inputs can be treated as exogenous. The Zellner, Kmenta, and Drèze argument of expected profit maximization treats inputs as exogenous only if technical inefficiency is completely unknown to the firm. Since technical inefficiency includes the possibility of poor management, the composition of vintage of capital stock, and local labour quality, the assumption of unknown technical inefficiency may not be fully justified. If technical inefficiency is known to the firm, the estimates of the production function parameters obtained directly from the production function will be inconsistent. This inconsistency can be avoided if a simultaneous equation approach is used that allows estimation of all three inefficiency parameters.

A simultaneous equation method such as full information maximum likelihood (FIML) gives inconsistent estimates of the parameters. If τ is known to the firm, then the use of FIML gives upward bias to the intercept. On the other hand, if τ is unknown, then it is not possible to disentangle τ and ν. The use

of the maximum likelihood estimation (MLE) method is warranted here because an assumed distribution for the disturbance terms can be specified; therefore, a simultaneous equation MLE method has been used for estimation of various inefficiencies.

The system of equations can be written as

$$(1c) \qquad \tau + \nu = \ln y - \ln \alpha_0 - \sum_i \alpha_i \ln x_i - \sum_j \beta \ln z_j - \sum_i \sum_j \gamma_{ij} \ln x_i \ln z_j ;$$

$$(14b) \qquad u_q = \ln x_s - \ln x_q + \ln w_s - \ln w_q + \ln\Big(\alpha_q + \sum_j \gamma_{qj} \ln z_j\Big) - \ln\Big(\alpha_s + \sum_j \gamma_{sj} \ln z_j\Big) ;$$

$$(18b) \qquad \xi = (\theta - 1)\ln y - \ln p + \ln \theta + \ln K + \theta \sum_i \Big\{\alpha_i + \sum_j \gamma_{ij} \ln z_j\Big\} \ln w_i - \theta \sum_j \beta_j \ln z_j .$$

This system of equations are more useful than equations (1b), (14a), and (18a). To estimate the system of equations (1c), (14b), and (18b) the probability density function (pdf) of the error vector must be derived, with specified distributional assumptions for each error term. The error vector from the system of equations (1c), (14b), and (18b) is

$$(19) \qquad \begin{pmatrix} \tau + \nu \\ \tilde{u} \\ \xi \end{pmatrix},$$

where $\tilde{u} = \tilde{u}_q = [u_2, \ldots, u_n]'$. Assumptions regarding the distribution of error terms are:

1. $\tau \sim iid\ N(0, \sigma_\tau^2)$ truncated at zero from above;
2. $\nu \sim iid\ N(0, \sigma^2)$;
3. $\tilde{u} \sim MVN(\mu, \Sigma)$, i.e., iid normal over firms;
4. $\xi \sim iid\ N(0, \sigma_\xi^2)$;
5. τ, ξ, $\tilde{u}$ and ν are independent of each other and also independent of w_j and exogenous fixed inputs Z.

Given the above set of assumptions, the pdf for the error vector (19) can be derived. Let $b_1 = \tau + \nu$, $b_2 = \tilde{u}$, and $b_3 = \xi$. By variable transformation,

$$f_1(b_1, b_2, b_3, \tau) = f(\nu, \tau, \tilde{u}, \xi)\,|J|, \tag{20}$$

where J is the Jacobian of the transformation from $(\tau, \nu, \tilde{u}, \xi)$ to (b_1, b_2, b_3, τ), i.e.,

$$J = \frac{\partial(\nu, \tau, \tilde{u}, \xi)}{\partial(b_1, b_2, b_3, \tau)}.$$

Since $|J| = 1$, here, and τ, ξ, ν, and $\tilde{u}$ are independent of each other, we can write

$$f_1(b_1, b_2, b_3, \tau) = f(\tau)\, f(\nu)\, f(\tilde{u})\, f(\xi).$$

This can be written as

$$f(b_1, b_2, b_3, \tau) = \frac{2\exp(-d/2)}{(2\pi)^{\frac{1}{2}(n+2)}\,\sigma_\xi\,\sigma_\nu\,\sigma_\tau\,|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\,\frac{(\tau-\mu_\tau)^2}{\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma_\xi^2}\xi^2\right\} \exp\left\{-\frac{1}{2}(b_2-\mu)'\,\Sigma^{-1}(b_2-\mu)\right\}; \tag{21}$$

where $\sigma^2 = \left\{\frac{\sigma_\tau^2\sigma_\nu^2}{\sigma_\tau^2+\sigma_\nu^2}\right\}$; $\mu_\tau = \left\{\frac{(b_1\sigma_\tau^2)}{(\sigma_\tau^2+\sigma_\nu^2)}\right\}$ and $d = \left\{\frac{b_1^2}{(\sigma_\tau^2+\sigma_\nu^2)}\right\}$.

Equation (21) provides the joint pdf of b_1, b_2, b_3, and τ. To obtain the joint pdf of b_1, b_2, and b_3, τ can be integrated out from (21) as

$$\begin{aligned} f(b_1, b_2, b_3) &= \int_{-\infty}^{0} f(b_1, b_2, b_3, \tau)\, d\tau, \\ &= \frac{2\exp(-d/2)}{(2\pi)^{\frac{n+1}{2}}\,\sigma_\nu\,\sigma_\xi\,\sigma_\tau\,|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(b_2-\mu)'\,\Sigma^{-1}(b_2-\mu)\right\} \\ &\qquad \exp\left\{-\frac{1}{2}\,\frac{\xi^2}{\sigma_\xi^2}\right\} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \exp\left\{-\frac{1}{2\sigma^2}(\tau-\mu_\tau)^2\right\} d\tau; \end{aligned}$$

$$= \frac{2\exp(-d/2)\,\sigma\,\Phi(-\mu_\tau/\sigma)}{(2\pi)^{\frac{1}{2}(n+1)}\,\sigma_\xi\,\sigma_\nu\,\sigma_\tau|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\,\frac{\xi^2}{\sigma_\xi^2}\right\} \exp\left\{-\frac{1}{2}(b_2-\mu)'\,\Sigma^{-1}(b_2-\mu)\right\}, \tag{22}$$

where $\Phi(\cdot)$ is the cumulative density function (cdf) of a standard normal distribution since

$$\sigma \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{-\frac{1}{2\sigma}(\tau-\mu_\tau)^2\Big\}\, d\tau = \sigma\,\Phi\Big(-\frac{\mu_\tau}{\sigma}\Big). \tag{23}$$

By variable transformation, move from the pdf of the error vector to the pdf of the endogenous variable vector

$$f(\ln Y, \ln X_i) = f(b_1, b_2, b_3)\,|J|, \tag{24}$$

where $|J|$ is the determinent of the Jacobian of the transformation from ν, τ, $\tilde{u}$, and ξ to endogeneous variables $\ln Y$ and $\ln x_i$. The log likelihood function for T firms can be expressed as

$$L = \sum_{t=1}^{T} \ln f(b_1, b_2, b_3) + T \ln |J|. \tag{25}$$

In this model,

$$\begin{aligned} |J| &= \frac{\partial(b_1, b_2, b_3)}{\partial(\ln Y, \ln x_i)} \\ &= \frac{\partial(\tau+\nu, u_2, \ldots, u_n, \xi)}{\partial(\ln Y, \ln x_1, \ldots, \ln x_n)} \qquad (26) \\ &= \begin{vmatrix} 1 & -\alpha_1 & -\alpha_2 & -\alpha_3 & \ldots & -\alpha_n \\ 0 & 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ (\theta-1) & 0 & 0 & 0 & \ldots & 0 \end{vmatrix} = (\theta-1)\sum_i \alpha_i . \end{aligned}$$

Therefore,

$$f(\ln Y, \ln x_i) = f(b_1, b_2, b_3)\,(\theta-1)\sum_i \alpha_i . \tag{27}$$

The likelihood function for a single observation, therefore, can be expressed by taking the logarithm of equation (25) and introducing the Jacobian of transformation:

$$L = \ln 2 - (d/2) + \ln\sigma + \ln\Phi(-\mu_\tau/\sigma) - \frac{1}{2}(n+1)\ln(2\pi) - \ln\sigma_\xi$$

$$(28)\qquad -\ln\sigma_\nu - \ln\sigma_\tau - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(b_2-\mu)'\Sigma^{-1}(b_2-\mu) - \frac{1}{2}\left\{\frac{\xi^2}{\sigma_\xi^2}\right\} + \ln\left\{(\theta-1)\sum_i \alpha_i\right\}.$$

The burden of estimation can be reduced by concentrating the likelihood function. Here, the likelihood function can be concentrated with respect to μ and Σ, i.e., with respect to the mean and the variance. At the maximum of L, the estimate of μ is

$$(29)\qquad \mu_q = \frac{1}{T}\sum_{t=1}^{T} u_{qt} = \overline{\ln x_s} - \overline{\ln x_q} - \overline{\ln w_q} + \overline{\ln w_s} + \ln\left\{\frac{\alpha_q + \sum_j \gamma_{qj}\ln z_j}{\alpha_s + \sum_j \gamma_{sj}\ln z_j}\right\},$$

where bars over variable indicate the arithmetic means. Similarly, the elements of Σ, i.e., σ_{qk} can be estimated from

$$\sigma_{qk} = \frac{1}{T}\sum_{t=1}^{T}\left\{\ln x_s - \ln x_q - \ln w_q + \ln w_s - \overline{\ln x_s} + \overline{\ln x_q} + \overline{\ln w_q} - \overline{\ln w_s}\right\}$$

$$(30)\qquad \left\{\ln x_s + \ln w_s - \ln x_k - \ln w_k - \overline{\ln x_s} - \overline{\ln w_s} + \overline{\ln w_k} + \overline{\ln x_k}\right\},$$

which is independent of the parameters and, therefore, Σ becomes a constant. The term $\frac{1}{2}(b_2-\mu)'\Sigma^{-1}(b_2-\mu_2)$ is a scalar, which is equal to its trace, and the trace of a product is invariant to certain orderly permutation of the order of multiplication, so

$$(31)\qquad \begin{aligned} W &= tr\left\{(b_2-\mu)'\Sigma^{-1}(b_2-\mu)\right\} \\ &= tr\left\{\Sigma^{-1}(b_2-\mu)'(b_2-\mu)\right\} \\ &= tr\Sigma^{-1}(b_2-\mu)(b_2-\mu)' \\ &= tr\Sigma^{-1}\Sigma \\ &= trI \\ &= n-1. \end{aligned}$$

The concentrated likelihood function for a single observation, therefore, can be expressed as

$$(32)\qquad L_c = \kappa - \left(\frac{d}{2}\right) + \ln\sigma + \ln\Phi\left(\frac{-\mu_\tau}{\sigma}\right) - \ln\sigma_\xi - \ln\sigma_\tau - \ln\sigma_\nu$$

$$-\frac{1}{2\sigma_\xi^2}+\ln\left\{(\theta-1)\sum_i \alpha_i\right\},$$

where

$$\kappa=\ln 2-\frac{n+1}{2}\ln 2\pi-n+1+\ln|\Sigma|$$

is a constant and can be eliminated during estimation. The MLE of α_0, α_i, β_j, γ_{ij}, σ_τ, σ_ν, and σ_ξ can be obtained by maximizing equation (32) after adding over T firms. The MLE of μ and the elements of Σ can be obtained from the estimates of equations (29) and (30). This maximization of the log likelihood function yields consistent and asymptotic efficient estimates of the parameters.

Estimation of Efficiency Parameters

Given the estimates of the parameters, τ, ξ, and $\tilde{u}$ can be solved for each observation and each endogenous variable.

Technical Inefficiency

To derive the technical efficiency parameter, τ, for each firm, τ must be isolated from ν, where $\tau+\nu=b_1$ is the residual of the production function. Following Jondrow, Lovell, Materov, and Schmidt this can be done by estimating the conditional pdf of τ given b_1. It can be shown that τ given b_1 is normally distributed with mean μ_τ and variance σ_τ truncated at zero, i.e.,

$$(33)\qquad h(\tau\mid b_1)=\frac{1}{(2\pi)^{1/2}\sigma\,\Phi(-\mu_\tau/\sigma)}\exp\left\{-\frac{1}{2\sigma^2}(\tau-\mu_\tau)^2\right\}$$

$$=\frac{\phi((\tau-\mu_\tau)/\sigma)}{\Phi(-\mu_\tau/\sigma)}.$$

Technical inefficiency can be estimated at a point estimator of τ with either mean or mode as the point estimator. The mean of the conditional distribution of τ is given by

$$(34)\qquad \mathrm{E}(\tau\mid b_1)=\int_{-\infty}^{0}\tau\,h(\tau\mid b_1)\,d\tau.$$

The mean and mode of τ are, respectively,

$$(35)\qquad \tau_{mean}=\mu_\tau-\sigma\frac{\phi(\mu_\tau/\sigma)}{\Phi(-\mu_\tau/\sigma)};$$

$$\tau_{mode} = \begin{cases} \mu_\tau, & \text{if } (\mu_\tau/\sigma) \leq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

Estimates derived by using either mode or mean are consistent and asymptotically efficient (Jondrow, Lovell, Materov, and Schmidt).

Given the estimated values of μ_τ and σ, τ can be estimated. Therefore, given X and Z, the percentage loss of output due to technical inefficiency, τ, can be obtained from

$$YL_\tau = \frac{(Y - Y^*)}{Y^*}. \tag{37}$$

The parametric expression for the above equation is

$$YL_\tau = 1 - \exp(\tau), \tag{37a}$$

where Y^* is the frontier output obtained by setting $\tau = 0$, and Y is the observed output.

Allocative Inefficiency

The relative allocative inefficiency of each firm for each input (relative to input s) can be estimated from the equation (14b):

$$u_q = \ln x_s - \ln x_q + \ln w_s - \ln w_q + \ln\left\{\frac{\alpha_q + \sum_j \gamma_{qj} \ln z_j}{\alpha_s + \sum_j \gamma_{1j} \ln z_j}\right\}. \tag{14b}$$

The u_q can be both positive and/or negative, but in either case it increases the cost of production due to allocative ineffciency. Following Schmidt and Lovell (1979), the cost impact of allocative inefficiency can be measured from the equation (see appendix B.1 for derivation of the cost function given below)

$$\begin{aligned} \ln C = {} & \ln\left\{\sum_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j)\left\{\alpha_0 \prod_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)}\right\}^{-\theta}\right\} \\ & \theta \ln y + \theta \sum_q (\alpha_q + \sum_q \gamma_{qj} \ln z_j) u_q - \theta(\nu + \tau) \\ & - \theta \sum_j \beta_j \ln z_j + \theta \sum_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j) \ln w_i \\ & + \ln\left\{\frac{(\alpha_k + \sum_j \gamma_{kj} \ln z_j) + \sum_q (\alpha_q + \sum_j \gamma_{qj} \ln z_j)}{\sum_i (\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \exp(-u_q)\right\}. \end{aligned} \tag{38}$$

After some algebraic manipulation, this can be expressed as

$$(39)\quad \ln C(W,y,Z) = \ln K + \theta \ln y + \theta \sum_i \Big(\alpha_i + \sum_j \gamma_{ij} \ln z_j\Big) \ln w_i - \theta(\nu + \tau) - \theta \sum_j \beta_j \ln z_j + \Big\{\rho - \ln(r+k)\Big\},$$

where

$$\rho = \ln\Big\{\alpha_1 + \sum_j \gamma_{1j} \ln z_j + \sum_q \Big\{\alpha_q + \sum_j \gamma_{qj} \ln z_j\Big\} \exp(-u_q)\Big\}$$

$$\theta \sum_q (\alpha_q + \sum_j \gamma_{qj} \ln z_j)\, u_q\,;$$

$$\theta \equiv \frac{1}{r+k} = \frac{1}{\sum_i \{\alpha_i + \sum_j \gamma_{ij} \ln z_j\}}\,;\ \text{and}$$

$$\ln K = \ln\theta - \theta \sum_i \Big\{\alpha_i + \sum_j \gamma_{ij} \ln z_j\Big\} \ln\Big\{\alpha_i + \sum_j \gamma_{ij} \ln z_j\Big\}.$$

The cost of allocative infficiency following Schmidt and Lovell (1979, 1980) is therefore represented by the term

$$(40)\quad CA = \Big\{\rho - \ln r + k\Big\}$$

of (39). Clearly ρ is minimized when $u_2 = u_3 = \ldots = u_n = 0$ and equals $\ln(r+k)$. In that case, the cost function does not include any allocative inefficiency parameter. Otherwise, the values (non-negative) of $\{\rho - \ln(r+k)\}$ are the addition to the logarithm value of cost ($\ln C(W,y,Z)$) due to allocative inefficiency. It is interesting to note that the actual cost exceeds the frontier cost for two reasons: (i) technical inefficiency, reflected by the term $\theta(\tau)$ and (ii) allocative inefficiency, represented by the term $\{\rho - \ln(r+k)\}$ in (39).

Scale Inefficiency

Scale inefficiency can be measured from equation (18b), i.e.,

$$(39)\quad \hat{\xi} = (\theta - 1)\ln y - \ln p + \ln\theta + \theta \sum_i \Big\{\alpha_i + \sum_j \gamma_{ij} \ln z_j\Big\} \ln w_i - \theta \sum_j \beta_j \ln z_j + \ln K\,.$$

This measure of scale inefficiency is different from the KBB measure of scale inefficiency of Kumbhakar, Biswas, and Bailey. The KBB measure can, however, be obtained from our measure of scale inefficiency (see appendix B.2 for derivation of the KBB measure for this model).

The loss of output (over or under supply) due to scale inefficiency alone is

$$YL_{\xi} = 1 - \exp\left\{\frac{\xi}{(\theta - 1)}\right\}. \tag{41}$$

Impact on Profit

The percentage loss of profit due to inefficiency is given by the relation

$$\Pi L = \frac{\Pi^{*}(p, W, Z, \nu) - \Pi(p, W, Z, \tau, \xi, \nu, \tilde{u})}{\Pi^{*}(p, W, Z, \nu)}, \tag{43}$$

where $\Pi(p, W, Z, \xi, \tilde{u}, \tau) = py - W'X = py - C(W, Y)$, and the frontier profit or efficient profit frontier is

$$\Pi^{*}(p, W, Z) = pY_{|\xi=\tilde{u}=\tau=0} - \sum_{i} W' X^{*}_{|\tau=\xi=\tilde{u}=0}.$$

The percentage loss in profit due to technical inefficiency alone can be expressed as

$$\Pi L_{\tau|\xi=\tilde{u}=0} = 1 - \frac{\Pi_{\tau}(p, W, Z, \tau, \nu)}{\Pi^{*}(p, w, z, \nu)}. \tag{44}$$

After some algebraic manipulation, the parametric expression for the loss of profit due to technical inefficiency alone becomes

$$\Pi L_{\tau|\xi=\tilde{u}=0} = 1 - \exp\left\{\frac{\theta\tau}{(\theta - 1)}\right\}. \tag{45}$$

Similarly, the loss in profit due to scale inefficiency alone is

$$\Pi L_{\xi} = 1 - \exp\left\{\frac{\xi}{(\theta - 1)}\right\}\left\{1 - \frac{\exp(\xi)}{\theta}\right\}. \tag{46}$$

Data Description

The data set used in this study was collected from three agro-climatic regions[2] of the Indian state of West Bengal. A cross section sample of 300 household farms was interviewed over the period 1980–1983. Since these regions were surveyed over a four-year period, the bias of the data set caused by time specific problems (such as flood or drought) was avoided.[3] Over the last two decades, the success of technological change, introduction of high-yielding seeds, security of land tenure, availability of production loans, and relaxation of government controls over purchase of output and sale of vital factors like fertilizers, irrigation water, and pesticides have made the agricultural production system very much market oriented. In such an environment, profit-maximizing behavior on the part of producers would be a reasonable maintained hypothesis.

Three regions covered in this survey cover the entire rice production belt of West Bengal. Topographical conditions do not differ very much between regions, and all three regions are situated on the Gangatic Plain. As a result, the irrigation water supply system is well developed and cheap or even free.[4] The transportation service is also quite well developed, and access to market for both inputs and outputs is fairly easy. After bank nationalization and establishment of a rural cooperative banking system, the expansion of the rural banking system has to a great extent been able to reduce and in some places abolish monopoly control by the village money lenders over the rural production system. The new land tenancy system of the present state government has ensured land tenancy and share tenancy rights, which allowed tenants to invest in land development and make permanent investments on their plots (like installing a pump set or digging irrigation wells). The expansion of extension services has enabled poor farmers to access modern technology and related information. Moreover, the expansion of the free schooling system all over the state has made a positive contribution to the state's agricultural system.

The agricultural production activities of West Bengal are spread over at least

[2] Region I covers two districts, Nadia and Murshidabad. Region II covers the districts of Burdwan and Birbhoom, and region III covers the districts of Howrah, Hooghly, 24 Parganas, and Midnapore. District is equivalant to U. S. county.

[3] Although the data was collected over a period of three years, it can be used for cross-section analysis. Greene (1989) pointed out that if a fairly large number of observations are collected over a short period of time, then cross-section use of that data set does not affect the quality of the results.

[4] Thanks to government irrigation and flood control programmes during the first three five-years plans.

three crop seasons during a one-year production cycle. The monsoon paddy is the single most important crop in all three regions. This crop is produced by almost all peasants in these three regions and almost 70% of the total arable land is used in this particular cultivation (Dasgupta).[5] The other two crop seasons are winter and summer, with wheat being the major summer crop and pulses and oilseeds during the winter.

The survey provides a wide range of information about both fixed and variable inputs used in each production. Among inputs, fertilizer (F), manure (M), human labour (H), and bullock labour (B) are used as endogenous variable inputs, while land (L), capital (K), off-farm income (O), and years of education (E) are used as fixed inputs. The same capital input was assumed for each crop. Fertilizer is measured in terms of kilograms, while both types of labour are measured in terms of labour hours. The monsoon[6] crop is the only output considered in this model. Total amount of paddy (Y) produced, rather than the marketable surplus, is considered the only output, which is measured in terms of quintal (one hundred kilograms). Manure is also measured in terms of hundred kilograms. Average characteristics of the farms covered by the survey are represented in the table 1. For analysis, the data set has been divided into two non-intersecting groups, large and small farms.

Since the size of operational land holdings is very small, the role of machine capital is very limited. Both bullock and human labour play very important roles. Human labour includes both family and hired labour; though it is not possible to separate these two components in the data set, effects of these two types of labour on output are quite different. Family labour plays a dual role, both as manual labour and also as managerial decision-making unit. Thus, the returns to these two types of labour are different. In this study, the market wage rate has been used to represent the hiring prices of all labour. Sampath has argued that in the context of Indian agriculture, the valuation of family labours at the ruling market wage rate is quite justified for all size classes of farms; however, Sen pointed out that if family labour in agriculture is given an imputed value in terms of ruling wage rate, much of Indian agriculture seems unremunerative. All major empirical works on Indian agriculture have used the prevailing market wage to value domestic labour, and their results show that Indian farmers are quite rational in allocating their labour factor based on market prices.[7] Since

[5] The quality of the crop may differ across regions and between individual plots of a single region; however, this quality difference is not introduced in the model.

[6] Monsoon cropping spans over the period June to September.

[7] About 65–70% of Indian farmers participate in the labour market (Rosenzweig).

usual labour contracts are for working days, the hours of human labour have been converted assuming eight hours per working day.

The use of animal power for cultivation is still dominant in India; therefore, bullock labour is very important in the family farming system. The market for animal labour is highly localized, and most farms own at least a pair of bullock. During the peak season, such as the monsoon season, a market for animal labour exists. To avoid double counting, the total number of use hours of animal labour has been treated as an explanatory variable, while the value of farm-owned animals (for farming) has not been included in fixed capital. Although the usual practice is to hire bullock labour on a per-day basis, bullock labour is not expressed in terms of working days but in terms of total working hours. Assuming eight hours per working day, the hours of bullock labour have been converted into a measure of input use.

Another group of variable factors is fertilizer and manure. The idea of using chemical fertilizer is relatively new in India. With the introduction of the high yielding variety (HYV) production system and multiple season cropping, the use of chemical fertilizers has gained momentum. Fertilizer represents the total dose of chemical fertilizer used in that cropping season. The market for chemical fertilizers has a dual nature. Since the supply of chemical fertilizers is partly regulated, a parallel market for this input exists, and the farmer has free access to that market; therefore, the actual price of chemical fertilizers rather than the controlled price is used in this analysis. Different farms use different types of chemical fertilizer (or combinations of different fertilizers) and at different doses. This data set does not provide detailed information regarding various types of fertilizers used on a farm. The total amount of fertilizer used by each farm for each crop can only be obtained from the data set and is measured in kilograms.

The market for manure is highly localized and more contract based than the usual market. Due to alternative uses of manure, the market is quite competitive, particularly during the monsoon season. The data set reflects quite a variation in the price of manure even within one region, mainly because of different types of forward contracts between buyers and sellers of manure and transportation costs involved.

Besides these four variable endogenous inputs, four fixed inputs considered in this model are land, physical capital, years of education, and off-farm income. Land is actually the allocable fixed input, but in our model, land has been treated as a fixed input. A farmer can allocate land among different crops in the same season or among seasons, given the total amount of land, due to three reasons.

First, farmers may keep a portion of land fallow as part of their optimizing decision. Second, the total operational land holding of a household farm can be fragmented with qualitatively different plots. Qualitatively, some plots may be unsuitable for cultivation of a specific crop, e.g., for a summer crop that needs generous irrigation, a plot away from a water source may not be suitable, or shallow land may not be suitable for monsoon paddy. Third, there is a well-developed market for leasing land with tenancy and share cultivation systems being a large part of Indian agriculture. A farmer can lease all of his land or part of it and earn rent. In this model, land has been treated as a fixed factor, using the area under cultivation rather than the total acreage owned by a household farm.[8]

Capital is defined as a combination of values of all durable farm assets, namely the farmhouse, irrigation equipment, and farm equipment. The total value of capital has been deflated by the user cost of capital, defined as the sum of the yearly depreciation rate and the market rate of interest less the yearly inflation rate, assuming a straight line depreciation. The interest rate used is the average market interest rate over the period of survey. Measurement of capital is the *Achilles heel* of neo-classical economics. This study follows convention to derive a measure of capital in value terms.

Off-farm income is measured in terms of hundred rupees. In this study, off-farm income is included in the production function as a control variable and enters the function as a fixed factor. Importance of off-farm income on the level of productivity was first highlighted by Schultz. Whether participation in off-farm employment will lead to enhancement of productivity of the household farm is an open question. Gronau's full income household production model also emphasized the role of off-farm income on household productivity. He pointed out that, given the opportunity set, labour force participation is associated with a decline both in leisure and work at home through time allocation. Even without a full income model in this analysis, by introducing off-farm income as a regressor, to a certain extent, the impact of off-farm income on the inefficiency of Indian household farms can be verified.

In agriculture, the degree of stochastic element is high, and the need for on-the-spot decisions is very important. If the farmer is engaged in off-farm employment, he cannot make such quick decisions, which means productivity and output may fall. On the other hand, off-farm employment allows the farmer to invest more

[8] Under the present land management system in West Bengal, share-croppers and tenancy rights are protected by law. Therefore, user right rather than ownership of land is important from the viewpoint of estimation.

cash in his farming (say by buying a pumpset or doing some permanent development of the farm) that might increase productivity. Therefore, exact signs of the impact of off-farm income on productivity are difficult to determine *a priori* and empirical analysis is needed to assess the direction of impact. Measuring the importance of off-farm income is crucial in characterizing the Indian agricultural system. The underlying reasons are, of course, different for large and small farmers. Small farmers participate in the local labour market both as agricultural and non-agricultural labourers. Rosenzweig has shown that about 65-70 % of Indian farmers participate in the labour market. Large farm families, on the other hand, are usually involved in some local trade or services. In fact, local agri-businesses are usually controlled by large farm families, which allows them better access to market information compared to the small farms.

Since the seminal papers of Ram and Huffman on the impact of education on production decisions and allocation of scarce inputs, treatment of education as a regressor input in production relationship (especially of agricultural production) has become commonplace. In a dynamic environment with imperfect information, education has a significant impact on allocative efficiency because the ability to acquire and process information is enhanced with increased education. Indian agriculture, as mentioned above, has undergone rapid modernization and commercialization; therefore, education would likely influence allocative and technical efficiency on individual farms. In our model, using education as a fixed factor, the years of education of adult male members of the farm household are used as the measure for this input.[9] In the Indian situation, educated members of large farms are usually employed in off-farm positions in the services or trade sectors away from their village, so they take with them the education input that could enhance efficiency on the farm. On the other hand, they contribute a part of their income to the family which, in turn, might influence productivity of the family farm. In the case of small farms, this process is little different. Small farmers usually stay in the village, and their off-farm jobs usually are in local services or in large farms as labourers. As a result, the human capital and off-farm income stays in the household.

The data set shows that the hiring prices of factors and output prices vary not only between years covered by the survey but also across regions. Variation in output and input prices can be attributed in part to differences in transport cost,

[9] Unfortunately, the data set only provides such information for male members of the household; however, female members also work in the field and play an important role in farm management.

storage cost, types of forward contracts, and types of land tenancy contracts. Quality differences in output and inputs may be another factor responsible for these price variations.[10] Though floor wage and support prices were in effect during the years of study, the data set reflects wide variation on either side of official rates. The price received by an individual farmer for his produce also depends on how long he can hold the sale. Large farms are expected to have enough financial strength to hold their sale longer than their smaller counterparts, or they can also sell their output in portions over time. This is probably why the data set reflects the fact that large farms receive better prices for both inputs and outputs, except for manure. The pattern of agriculture in India to a great extent depends on the mercy of the monsoon, to which farmers match their farming schedules. This is another reason for wide fluctuations in rupee prices of some inputs like fertilizer, bullock labour and human labour.

Empirical Results

In view of the diverse farm size and wide dispersion of profit and output, these observations have been divided into two nonintersecting subsets of large and small farms. Table 1 showed some average characteristics of farms used for the study. Land holding size has been used to characterize farms into these two subgroups. Following Bardhan, this study chose seven acres as the dividing line between large and small farms. Three separate models have been estimated to check whether the production structure of small farms (Model B) and that of large farms (Model C) differ from the production structure reflected by the pooled data (Model A). If so, then an analysis of farm specific inefficiency based on pooled data (Model A) is inappropriate. The likelihood ratio test indicates evidence to reject the hypothesis of single production function for all size classes. The test statistic in the case of large farms, $-2(L_U - L_R) = 4417.7$, is greater than the χ^2 with 37 degrees of freedom for any reasonable significance level. Similarly, for small farms the test statistic, $-2(L_U - L_R) = 5431.35$, is greater than the χ^2 with 37 degrees of freedom for any reasonable level of significance. Thus, the use of a single production function (Model A) for our purpose does not appear to be appropriate; however, for comparison, parameter estimates of Model A are reported in table 2.

Table 1 shows some average characteristics of sample farms. The large farms on average have a land holding twice the size of small farms. Per acre productivity of large farms is higher compared to that of small farms. Large farms receive

[10] In this model we have not introduced quality differences. Information about this aspect can not be obtained from the data set.

a higher price for their produce and pay lower prices (except manure) for their inputs. There is not much difference in per acre input utilization, but high value of standard deviations indicates wide fluctuation across farms, both for small and large farms.

The system of estimated equations can be expressed in terms of all exogenous and endogenous variables as

$$\ln y = \ln \alpha_0 + \sum_i \alpha_i \ln x_i + \sum_j \beta_j \ln z_j + \sum_i \sum_j \gamma_{ij} \ln x_i \ln z_j + \nu + \tau ;$$

$$\ln x_s - \ln x_q = \ln w_q - \ln w_s + \ln \left\{ \alpha_s + \sum_j \gamma_{sj} \ln z_j \right\} - \ln \left\{ \alpha_q + \sum_{qj} \gamma_{qj} \ln z_j \right\} + u_q ;$$

$$\ln y = (\theta - 1)^{-1} \ln p - (\theta - 1)^{-1} \ln \theta + \ln K - \theta \sum_j \beta_j \ln z_j + \theta \sum_i \left\{ \alpha_i + \sum_j \gamma_{ij} \ln z_j \right\} \ln w_i - \xi ,$$

where, in this model

$$x_i = F, M, H, B; \quad z_j = K, L, E, O; \quad w_q = w_B, w_H, w_F; \text{ and } w_s = w_M.$$

$$\theta = (r + k); \quad r = \sum_i \alpha_i; \quad k = \sum_i \sum_j \gamma_{ij} \ln z_j, \text{ and}$$

$$K = \sum_i \left\{ \alpha_i + \sum_j \gamma_{ij} \ln z_j \right\} \left\{ \alpha_0 \left(\alpha_i + \sum_j \gamma_{ij} \ln z_j \right)^{(\alpha_i + \sum_j \gamma_{ij} \ln z_j)} \right\}^{-\theta}.$$

x_i represents variable inputs, fertilizer (F), manure (M), human labour (H), and bullock labour (B). z_j represents capital (K), labour (L), education (E), and off-farm income (O). w_q represents prices of variable inputs, fertilizer (w_F), manure (w_M), human labour (w_H), and bullock labour (w_B).

Parameter estimates of the likelihood functions are reported in table 2. The most relevent parameters are of appropriate sign and appear to be statistically

significant.[11] In the econometric analysis, the derivative properties of economic behaviors are reflected by the estimated coefficients; therefore, some economic conclusions can be drawn from the estimated values and signs of coefficients of the estimated model. Note the impacts of education and off-farm income on the level of production. For large farms, the off-farm income elasticity is negative (-0.0296) but insignificant at any reasonable level of significance. On the other hand, for small farms the same elasticity is positive (0.2226) and significant[12] at the 10% level of significance. One explanation for these findings may be the fact that large farms quite often emphasize the return from their off-farm sources of income more than their farm income, and the expansion of off-farm economic activities is probably more profitable than the expansion of their farming activities. On the other hand, small farms invest their off-farm income in the farm to enhance their farm income.[13] The negative effect of off-farm income on the productivity of the large farms indicates that the larger the off-farm income, the less time the operator spends on the farm. Consequently, the production decision will be based on less economic information, which tends to reduce efficiency.

The impact of education on the level of production is positive and significant for both groups, but the impact appears much stronger for small farms compared to large. This is consistent with earlier findings (Ram, Kumbhakar and Bhattacharyya) in the literature. The impact of education on productivity of large farms is very much linked to off-farm income. In most cases, educated members of large farm households obtain employment in the nonagricultural service sector, which accounts for the statistical insignificance of the off-farm income coefficient for large farms. From a policy point of view, this result supports of the benefits of the recent government policy that provides very low cost education for the rural population.

Another important coefficient to note is that of fertilizer, α_F, which is positive for both groups. The marginal productivity of fertilizer seems to be stronger for small farms compared to that of large farms; on the other hand, marginal productivity of both types of labours appear comparatively higher for the large

[11] In this type of complicated multiple input model the possiblility of multicollinearity is very high, which can affect the signs of the parameters. It is very difficult to interpret the signs of coefficients involved in the presence of multicollinearity.

[12] Elasticities are measured at the approximation point of the regressors.

[13] Due to present land management legislation that protects the rights of share croppers, the potentiality of earning monopoly rent from agriculture is not very promising. Instead, other agri-economic activities are now a potential source for earning monopoly rent.

farms.

Regarding the estimates of technical inefficiency, τ, allocative inefficiency, u_j, and scale inefficiency, ξ, these inefficiencies are estimated using equations (35), (14b), and (39). Though τ and ξ can be interpreted as deviation of actual output produced and supplied from their respective optimum levels, it is impossible to give a unique interpretation of the allocative inefficiency parameters, u_j. Following the Schmidt and Lovell (1979, 1980) definition of allocative inefficiency, the measure depends on the choice of numerare, in this case manure. The estimated values of u_j would change if some other variable were used.

From the policy point of view, the impact of inefficiencies on the level of profit, cost, and output is important, as reported for both groups of farms in table 3. The percentage of potential output loss due to technical inefficiency, YL_τ, is estimated using equation (37a). The percentage loss of output due to scale inefficiency alone, YL_ξ, is estimated by equation (41). The percentage increase in cost due to allocative inefficiency alone, CA, is estimated using equation (40). Similarly, the percentage loss of profit due to technical inefficiency alone, ΠL_τ, is estimated from equation (44a). The percentage loss of profit due to scale inefficiency alone, ΠL_ξ, is estimated from equation (46).

The estimates of output loss due to technical inefficiency, YL_τ, indicates that large farms are technically more inefficient than their smaller counterparts. Apparently, small farms could have increased their output by at least 16% without increasing inputs use had they been operating on their frontier. On the other hand, large farms could have similarly increased their output by 21%. Thus, our estimates indicate that the group of large farms is $(.20803-.15337)\times 100 = 5.47\%$ technically less efficient on average relative to their smaller counterparts. From these estimates, small farms are more scale inefficient compared to large farms. The output loss due to scale inefficiency, YL_ξ, for small farms shows that on average they supply 8.7% over their optimum level, while their larger counterparts oversupply by only 6.32%. This indicates that small farms are 2.38% more scale inefficient compared to large farms. The potential profit loss due to scale inefficiency is also higher for small farms compared to large. The small farms on average could have increased their profit over their actual profit by 17.43% by reducing their output supply by 8.67%. Large farms, on the other hand, sacrificed 8.97% on average of their potential profit due to 6.32% over production.[14]

[14] One explanation for this scale inefficiency of both small and large farms may be that the output variable represents the total production, not marketable surplus. Farm famillies usually keep a part of their produce for domestic consumption.

These estimates indicate that small farms incur greater cost as a result of allocative inefficiency compared to large farms. The cost of production of small farms on average increases by 9.14% due to allocative inefficiency, while the increase for large farms is 6.14%. Although small farms are technically 5.47% more inefficient than their larger counterparts, the loss of profit due to technical inefficiency alone is less for small farms. Small farms on average lose 15% of their potential profit due to technical inefficiency alone, while large farms lose 30.02%. This indicates that technical inefficiency is more problematic on large farms.

Conclusions

The concluding section will summarize the major results of this essay, and highlight limitations and possible extensions.

Major Results

To bring some degree of flexibility in measuring of production inefficiencies, a quasi-translog stochastic production function has been introduced. This production function takes into account the interaction between fixed and variable factors. This specification also allows for specifications and analytical derivations of all three inefficiencies. As a result, a system approach is followed to estimate all three inefficiencies simultaneously using the maximum likelihood estimation method. Newton's algorithm was used to solve the optimization problem. The empirical application was done using a data set from Indian agriculture. Estimated maximum likelihood parameters were used to obtain firm specific inefficiency estimates. For comparison, the data set was partitioned into two groups, and the log likelihood ratio test justified such truncation.

At least two distinguishing features of this model and its empirical application can be identified. First, the model takes account of intergroup factor substitution (due to its quasi-translog specification), and all three types of inefficiency are estimated using the framework for each farm. Second, all three types of farm-specific inefficiencies for Indian farms are estimated using stochastic production function and micro level data. The economic impacts of such inefficiencies are also highlighted, with policy implications following directly from the inefficiency estimates.

The empirical results generate at least four fundamental results. First, the small farms appear less technically inefficient than their larger counterparts. The percentage of profit loss due to technical inefficiency is larger for large farms compared to the smaller. Second, large farms are less scale inefficient compared to

small farms, partly due to the fact that the data set does not provide information on marketable surplus which is more important for measuring scale inefficiency in the Indian context (especially for small farms). Third, the cost of allocative inefficiency is higher for smaller farms compared to the large farms, a conclusion which contradicts the results of earlier studies. Four, loss of profit due to scale inefficiency is higher for small farms than for large farms.

Estimates suggest that education has a very high positive and statistically significant impact on productivity. Apparently education does influence inefficiency by increasing managerial ability, and hence productivity of capital and labour.[15] Off-farm income, on the other hand, has negative but insignificant impact on the productivity of large farms. The same variable has positive and significant impact on the productivity of small farms. This suggests that the higher the off-farm income, the less time the farm-operator spends managing farm operations.

Results of this study differ from earlier results of efficiency studies on Indian agriculture. As summarized before, most studies indicated Indian agriculture is profit, allocative, and technically efficient, whereas this study contradicts earlier results (e.g. Hopper, Kalirajan), mainly due to three factors. First, this study applieded a stochastic frontier approach using the micro unit instead of an aggregate approach. Second, this study used a more flexible functional specification compared to earlier studies.[16] Finally, most earlier studies used data either from Punjab or southern India. Studies of efficiency using micro level data from the eastern region of India have not been done up to this point. Agricultural systems among different parts of India vary so much that generalization of conclusions drawn from a study in one part of India is not necessarily appropriate for another region.

The set of assumptions made regarding the distribution of the stochastic error terms is quite strong. Specification of the model and, hence, the results of this estimation process depend on the correctness of such strong assumptions, therefore, the assumptions must be checked empirically. A normal probability plot for each error term has been checked, and the empirical distribution of each set of estimated residuals is close to the normal distribution. Some empirical distributions are short-tailed, but none is long-tailed. From this check, it seems the set of assumptions does not abstract the model from reality or prove too strong for

[15] Of course a positive causal relationship between the level of education and the level of inefficiency cannot be drawn from these results.

[16] Kalirajan and Jununkar used a flexible functional specification for their analysis.

the data set.

Limitations

The basic limitation of this approach is its inability to measure allocative inefficiency associated with the employment of individual inputs. The definition of Forsund, Lovell, and Schmidt expresses allocative inefficiency in relative terms (ratio or marginal productivity of two inputs); thus, change in the denominator input will change measures of the allocative inefficiency parameters, u_q. Input specific allocative inefficiency, however, can be computed (percentage of over or under utilization). The next two studies attempt to estimate input specfic allocative inefficiency.

The production function used in this model is not a fully flexible production function and, therefore, lacks the versatility of a fully flexible production functions. Another shortcoming of this type of model arises out of the error structure specification. The possibility of interaction between three types of inefficiencies are assumed away. Schmidt and Lovells' (1980) justification for adopting such a method avoids many complications in estimation at the cost of economic empiricism.

Possible Extensions

The next essay will introduce flexible functional forms for inefficiency estimation. Problems associated with the use of flexible functional forms for such modelling are also discussed in the next essay. Interaction between different inefficiencies is another interesting extension that can be checked with a different data set. The risk element involved in production has not been considered in the inefficiency literature. Technical inefficiency, as considered in this approach, is known to the firm and it can be treated as a function of some economic and noneconomic factors. For instance technical inefficiency can be composed of a deterministic component and a random component, a possible extension that would allow a farm to determine factors responsible for technical inefficiency.

Inspite of some of limitations, this methodology is quite complete and helpful for policy prescriptions for each individual farm covered by the sample.

Table II.1 Average Farm Characteristics*

Category	Small Farm B	Large Farm C
Number of Observations	245	55
Land Holdings	4.16	9.59
	(1.14)	(2.07)
Value of Capital	30.71	87.74
	(34.69)	(77.71)
Off Farm Income	46.05	52.08
	(77.03)	(10.09)
Years of Education	8.18	10.20
	(4.31)	(1.02)
Output	28.40	71.28
	(14.37)	(28.82)
Fertilizer	19.63	49.89
	(20.74)	(35.31)
Manure	38.73	83.79
	(21.64)	(52.39)
Bullock Labour	34.67	77.69
	(17.74)	(30.65)
Human Labour	138.71	313.46
	(118.34)	(48.14)
Price of Output	142.02	145.44
	(17.32)	(20.71)
Price of Fertilizer	5.50	5.23
	(1.09)	(0.97)
Price of Manure	6.88	7.27
	(2.91)	(3.36)
Hiring Price of Human Labour	10.09	9.05
	(2.19)	(1.88)
Hiring Price of Bullock Labour	8.98	8.67
	(1.42)	(1.77)

* Standard deviation in parentheses.

Table II.2 Maximum Likelihood Parameter Estimates*

Parameter	Pooled Model A	Small Farm B	Large Farm C
α_0	-3.6901	-10.4407	-3.0606
	(.4699)	(.8364)	(1.3168)
α_F	.3895	.9268	.0215
	(.0573)	(.3821)	(.0494)
α_M	.3771	.6583	.3149
	(.0749)	(.3001)	(.0783)
α_H	.8107	.0892	.2579
	(.1150)	(.2193)	(.0533)
α_B	.9878	.4044	.6759
	(.1382)	(.0533)	(.5173)
β_K	.1019	.7046	.3466
	(.0909)	(.1649)	(.2213)
β_L	.9696	.2314	.7808
	(.2815)	(.6332)	(.4104)
β_E	-.9842	.8989	.0989
	(.1950)	(.3709)	(.0468)
β_O	.3933	.3372	-.0177
	(.0830)	(.1195)	(.9675)
γ_{KF}	.5071	-.0071	.0298
	(.0072)	(.0409)	(.0220)
γ_{KM}	.2450	-.0647	-.0039
	(.0102)	(.0407)	(.0033)
γ_{KH}	-.1079	-.0060	-.0189
	(.0154)	(.0272)	(.0293)

* Standard errors in parentheses.

continued

Parameter	Pooled Model A	Small Farm B	Large Farm C
γ_{KB}	−.6884	−.1806	−.0355
	(.0092)	(.0548)	(.0205)
γ_{LF}	.5484	.1253	.1247
	(.0151)	(.2301)	(.0762)
γ_{EF}	.9004	−.1592	.0304
	(.0138)	(.1043)	(.0076)
γ_{EM}	−.4420	−.0241	−.0814
	(.0176)	(.0875)	(.0301)
γ_{OH}	1.1038	.0007	−.0019
	(.0147)	(.1981)	(.0023)
γ_{LM}	.6583	.2594	−.0298
	(.0306)	(.1878)	(.0165)
γ_{LH}	−.0809	.0921	−.0564
	(.0018)	(.1091)	(.0414)
γ_{LB}	3.1749	−1.0312	−.1524
	(.0255)	(.2758)	(.0733)
γ_{EH}	.1171	−.0598	−.0052
	(.0166)	(.0631)	(.0113)
γ_{EB}	−.0359	−.1111	−.0279
	(.0073)	(.1321)	(.0254)
γ_{OF}	.0014	−.0393	−.0030
	(.0025)	(.0386)	(.0021)
γ_{OM}	.8786	−.0774	−.0045
	(.0078)	(.0335)	(.0022)
γ_{OB}	.0108	.0021	−.0044
	(.0028)	(.0428)	(.0142)

continued

Parameter	Pooled Model A	Small Farm B	Large Farm C
σ_τ	.4543	.6059	.8854
	(.2273)	(.5228)	(.1749)
σ_ν	.1992	.6899	1.4021
	(.2356)	(.1112)	(.3670)
σ_ξ	.9817	.2657	2.1704
	(.1217)	(.0184)	(1.1225)
Incidental Parameters			
μ_F	2.6805	1.7961	2.1330
μ_H	−.5914	−3.1531	−3.3865
μ_B	.0751	1.6435	2.3372
σ_{FF}	1.7471	1.8032	1.4924
σ_{BB}	.7766	.7584	.8150
σ_{HH}	2.5089	2.4937	2.4959
σ_{FB}	.7355	.7346	.7253
σ_{FH}	1.5805	1.6328	1.3286
σ_{BH}	1.0799	1.0534	1.1398

Table II.3 Measures of Inefficiency by Farm Size Class

Measures	Small Farm B	Large Farm C
YL_τ	.1534	.2080
σ_τ	.0186	.0316
t_{τ_1}	2.1434	
YL_ξ	.0867	.0632
σ_ξ	.1330	.2491
t_ξ^1	.0832	
CA	.0914	.0614
σ_{CA}	.0766	.0169
t_{CA}^1	3.4172	
ΠL_τ	.1503	.3002
σ_τ	.0546	.0202
t_τ^1	3.4172	
ΠL_ξ	.1732	.0897
σ_ξ	.3361	.0454
t_ξ^1	.2495	

1 t statistic for testing H_0 : $\text{Mean}_B = \text{Mean}_C$.

σ = Standard deviation.

YL_τ = loss of output due to technical inefficiency alone.

YL_ξ = loss of output due to scale inefficiency alone.

CA = cost of allocative inefficiency alone.

ΠL_τ = loss of potential profit due to technical inefficiency.

PiL_ξ = loss of potential profit due to scale inefficiency.

References

Aigner, D. J., C. A. K. Lovell, and P. Schmidt. "Formulation and Estimation of Stochastic Frontier Production Function Models." *J. Econometrics* 6(1977):21-37.

Bardhan, P. K. "Size, Productivity and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture." *J. Polit. Econ.* 81(1973):1370-1386.

Dasgupta, B. "Monitoring and Evaluation of the Agrarian Reform Program of West Bengal." Comprehensive Area Development Program, Calcutta, December, 1987.

Farrell, J. M. "The Measurement of Productive Efficiency." *J. Royal Statist. Soc.* Series A(general) 21(1957):253-81.

Forsund, F. R., C. A. K. Lovell, and P. Schmidt. "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurements." *J. Econometrics* 13(1980):5-25.

Greene, W. H. *Econometric Analysis*. New York: Macmillan Publishing Company, 1989.

Gronau, R. "Leisure, Home Production, and Work – The Theory of the Allocation of Time Revisited." *J. Polit. Econ.* 85(1977):1099-125.

Hopper, W. D. "Allocative Efficiency in a Traditional Indian Agriculture." *J. Farm Econ.* 47(1965):611-24.

Huffman, W. E. "Allocative Efficiency: The Role of Human Capital." *Quart. J. Econ.* 47(1977):59-79.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt. "On The Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *J. Econometrics* 23(1982):269-74.

Jununkar, P. N. "The Response of Peasant Farmers to Price Incentives: The Use and Misuse of Profit Function." *J. Dev. Studies* 25(1989):169-72.

Kalirajan, K. "The Economic Efficiency of Farmers Growing High Yielding Irrigated Rice in India." *Amer. J. Agr. Econ.* 63(1981):566-70.

Kumbhakar, S. C., and A. Bhattacharyya. "Education Farm-Size and Allocative Efficiency in Indian Agriculture: A Restricted Profit Function Approach." Working Paper # 89-04. Utah State University, Utah, 1989.

Kumbhakar, S. C., B. Biswas, and D .V. Bailey. "A Study of Economic Efficiency of Utah Dairy Farmers: A System Approach." *Rev. Econ. Statis.* 71(1989):595-604.

Meeusen, W., and van den Broeck. "Efficiency Estimation From Cobb-Douglas Production Functions With Composed Errors." *Int. Econ. Rev.* 18(1977):435-44.

Ram, R. "Role of Education in Production: A Slightly New Approach." *Quart. J. Econ.* 95(1980):365-73.

Rosenzweig, M. R. "Agricultural Development, Education and Innovation," *The Theory*

and Experience of Economic Development: Essays in Honour of Sir W. Arther Lewis, eds. M. R. Rosenzweig, C. F. Diaz-Alejandro, Gustav Ranis and M. Gsovitz, pp. 107-34. London: George Allen & Unwin, 1982.

Sampath, R. K. *Economic Efficiency of Indian Agriculture–Theory and Measurement*. New Delhi: Macmillan, 1979.

Schmidt, P., and C. A. K. Lovell. "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontier." *J. Econometrics* 9(1979):343-366.

——. "Estimating Stochastic Production and Cost Frontiers when Technical and Allocative Efficiencies are Correlated." *J. Econometrics* 13(1980):83-100.

Schultz, T. W. *Transforming Traditional Agriculture*. New Haven: Yale University Press, 1964.

Sen, A. K. "Peasants and Dualism With and Without Surplus Labour." *J. Polit. Econ.* 74(1966):425-50.

Zellner, A., Kmenta, J., and J. Drèze. "Specification and Estimation of Cobb-Douglas Functions." *Econometrica* 34(1966):784-95.

CHAPTER III
SPECIFICATION AND ESTIMATION OF INEFFICIENCIES USING FLEXIBLE FUNCTIONAL FORM

Since the seminal paper of Greene (1980), the use of flexible functional forms for inefficiency estimation has gained momentum. Although the introduction of these forms has allowed the researcher to overcome restrictions imposed by homothetic functional specifications, a new set of problems has been introduced inherent to the nonhomothetic functional specification. Decomposition of inefficiency into its components, e.g., decomposition of cost inefficiency into its three constituent components—technical, allocative, and scale inefficiency—becomes impossible. Three major attempts have been made so far to make this decomposition possible, but none has amicably resolved the problem of decomposition. This essay will look at the decomposition problem from another angle. The amount of potential profit loss (or additional cost incurred) due to inefficiency is the main purpose of searching for the decomposition of inefficiency components. Without going into the impasse of decomposition, this study proposes a measure of potential profit loss (profit inefficiency) due to all three types of inefficiencies and specifies measures of the underlying inefficiency components.

The body of the essay is developed in seven sections. The next section develops the background literature for the study. The third section presents the model using a dual profit function. The estimation procedure follows in the fourth section, applying a data set on Indian agriculture. The fifth section provides a brief description of the data. Analysis of the empirical results is completed in the sixth section. The final section mainly summarizes the major conclusions and indicates further extensions of the study.

Background

This section explains the theoretical background for the model, and the theoretical background for the essay is developed in five sub-sections.

Why Duality?

The economic theory of production, as presented in such classic treatises as Hicks' *Value and Capital* and Samuelson's *Foundation of Economic Analysis* is based on the profit maximization principle subject to a production function. The objective of this theory is to characterize demand and supply functions using only

the restrictions on producers' behavior that arise from optimization. The principle analytical tool for this purpose is the implicit function theorem.

Unfortunately, the characterization of demand and supply equations as implicit functions of relative prices is inconvenient for econometric applications. In specifying an econometric model of producers' behavior, the demands and supplies must be expressed as explicit functions. These functions can be parameterized by treating measures of substitution, technical change, and economies of scale as unknown parameters estimated on the basis of empirical data.

The traditional approach to modeling producers' behavior begins with the assumption that production function is additive and homogenous. Under these restrictions, the demand and supply functions can be derived explicitly from the production function and the necessary conditions for producer equilibrium. However, this approach has the disadvantage of imposing constraints on patterns of production and thereby frustrates the objective of determining these patterns empirically.

The traditional approach, as originally developed by Cobb and Douglas, has been employed in empirical research since 1928. The limitation of this approach was made strikingly apparent by Arrow, Chenery, Minhas, and Solow, who pointed out that the Cobb-Douglas (CD) production function imposes *a priori* restrictions on the patterns of substitution among inputs. In particular, elasticities of substitution among all inputs must be equal to unity. Attempts to cure these shortcomings in the primal approach basically follow two different approaches. The first used a modification of the functional specification within the realm of additivity and homogeneity. The constant elasticity of substitution (CES) and variable elasticity of substitution (VES) functional forms are a result of that process; however, the CES and VES production functions retain the assumption of additivity and homogeneity, and the CES particularly imposes very stringent limitations on patterns of substitution. McFadden (1963) and Uzawa (1962) have shown that elasticities of substitution among all inputs must be the same in the case of CES. The other frontier of development was the development of flexible functional forms (FFF) like the Translog, Quadratic, and Generalized Symmetric McFadden. These specifications do not impose any *a priori* restrictions on the patterns of production; however, they fail to satisfy the global curvature properties required for global optimization.[1]

The neoclassical definition of a production technology is based on the pro-

[1] Each flexible functional form, however, satisfies curvature properties, at least, over a region. See appendix C.1 for definition of FFF.

duction function, which defines the maximum output from a specified set of inputs. Moreover, the single output production function, say $f(x)$, where x is an nonempty input vector, is defined as a positive, continuous, twice differentiable function with certain properties of monotonicity and concavity. In the economically relevant region of the production surface defined by this function, the production function is strictly concave, with positive marginal products and nonincreasing marginal rate of technical substitution. From these properties and the assumption of profit maximization or cost minimization, the firms' economic behavior can be deduced in terms of output supply and input demand functions. The properties of the production function, in combination with the behavioral assumption of profit maximization or cost minimization, determine the properties of these function.

Usually three types of difficulties are encountered in the neoclassical or primal approach to the analysis of firm behavior. First, the closed form analytic solutions of the first-order conditions of profit maximization or cost minimization for the firms' supply and input demand functions are generally not obtainable. That is, given a system of first-order condition equations, a closed form solution for the demand and supply functions can be obtained if, and only if, the system satisfies some properties of the implicit function theorem (Rudin). Second, even if the demand and supply functions can be solved, obtaining comparative static results is difficult because the properties of demand and supply functions must be derived from the properties of the production function. Third, the production function representation of the multiple-output technology is inconvenient for analytical and empirical purposes.

During the last three decades, production theory moved from the neoclassical approach, based on the production function and differential calculus, to a modern approach based on the analysis of technology sets using the mathematical theory of convex sets and the assumptions of maximizing or minimizing behavior. This approach permits duality relations between the production technology and cost, revenue, and profit functions to be derived rigorously and elegantly from the properties of the envelope theorem. The firms' production technology is defined in terms of a production possibility set, which in turn defines the production output set (bounded above by the production possibility frontier) and the input requirement set (bounded below by the isoquant). The modern approach overcomes previous difficulties by establishing a correspondence between the properties of the production possibilities set and the firm's cost, revenue, and profit functions. Due to duality theorem properties of the production possibilities set and the firm's

cost, revenue, and profit functions, and also due to the derivative properties of the dual functions, closed form expressions for input demand and output supply functions can be obtained for a variety of functional forms. Moreover, by virtue of the convexity and derivative properties of the dual functions, the comparative static properties of demand and supply functions are readily established. The multiple output versions of cost, revenue, and profit functions are straightforward generalizations of their single product counterparts.

The derivative properties of the neoclassical production function can be translated into restrictions on the input requirement set. Nonnegative marginal productivity is interpreted in terms of monotonicity or free disposibility. The convexity of isoquants due to nonincreasing marginal rate of technical substitution translates into the restriction that the input requirement set be convex from below.

The development of flexible functional forms and the subsequent development of the dual formulation of production theory have made it possible to overcome the limitations of the traditional approach to econometric modeling. The dual formulation was introduced by Hotelling and later revived and extended by Samuelson (1953, 1960) and Shephard (1953, 1970).[2] The key features of the dual formulation are, first, to characterize the production function by means of a dual representation such as a price or cost function, and, second, to generate explicit demand and supply functions as derivatives of the price or cost function (Diewert, 1982).

In the first essay, the starting point was a predefined, unobserved technological relationship which is typically called the production function. The optimizing conditions, hence all behavioral relationships, are derived on the basis of that unseen technology. In this second essay, dual relationships are used to derive the same set of information. The gain from such exercises is basically of two types. First, from the theoretical point of view using the dual relation production relationships (technology) can be inferred based exclusively on the observable economic information. Second, from the estimation point of view, use of the dual allows efficient estimation of parameters by avoiding the simultaneity bias caused in the estimation of a single equation production function[3] since the explanatory

[2] Hotelling and Samuelson (1953, 1960) developed the dual transformation of production theory on the basis of the Legendre transformation. This approach was later used by Jorgenson and Lau (1974a, 1974b) and Lau (1976, 1978). On the other hand, Shephard (1953, 1970) utilized distance functions to characterize the duality between cost and production functions. This approach is employed by Diewert (1974, 1982), Hanoch, McFadden (1978), and Uzawa (1964).

[3] Single equation production function estimation leads to inefficient estimates of the parameters involved because explanatory variables are correlated to the

variables are exogenous to the system. The implication of using dual functions (cost and/or profit function) to describe the technology is that the specification of a well-behaved cost function or profit function is equivalent to the specification of a well-behaved production function. All economically relevant information about the technology can be gleaned from the corresponding profit or cost function (Diewert, 1982).

Primal Dual Correspondence

For a given technology and given endowments of fixed factors of production, a profit frontier expresses the maximum profit of a firm as a function of prices of output, prices of variable factors, and, in the case of restricted profit function, also a function of the quantities of fixed factors of production. The underlying assumptions are (i) firms are profit maximizers, (ii) firms are price takers in both factors and product markets, and (iii) the production function is concave in variable inputs. Every concave production function has a dual that is a convex profit function (McFadden, 1978).[4] The local duality results of one-to-one correspondence between the production function and the profit function (both satisfying respective regularity conditions) can be established by the following identity between the Hessian of a profit function and that of a production function (Lau, 1976).[5]

$$(1) \qquad \begin{pmatrix} \Pi_{qq} & \Pi_{qz} \\ \Pi_{zq} & \Pi_{zz} \end{pmatrix} \equiv \begin{pmatrix} -f_{xx}^{-1} & f_{xx}^{-1} f_{xz} \\ (f_{xx} f_{xz})' & f_{zz} \end{pmatrix}$$

where Π is the profit function, and f is the production function. q, x, and z are price, input, and fixed factor vectors, respectively. Without loss of general-

dependent variable. However, the simultaniety bias could be avoided by following the simultaneous equation estimation procedure described in essay 1.

[4] McFadden (1978) extended the concept of cost function to the revenue function and the profit function and proved for the first time the McFadden Duality Theorem–the profit function analog of the Shephard (1953)–Uzawa (1964) Duality Theorem on cost and production function.

[5] Consider a firm, that maximizes restricted profit by choosing variable inputs

$$\max_{x} \; f(X, Z) - W'X = \Pi(W, Z),$$

where Z is the fixed input vector, and X is the variable input vector. $f(\cdot)$ is the production function. The first-order condition yields $f_{x_i}(\cdot) - w = 0$, i.e., $F(X; Z, W) = 0$. By the implicit function theorem $\frac{\partial x_i}{\partial w_i} = -\frac{F_{w_i}}{F_{x_i}} = -\frac{-1}{f_{x_i x_i}}$. Similarly, $\frac{\partial x_i}{\partial z_j} = -f_{x_i x_i}^{-1} f_{x_i z_j}$.

ity, either a profit function or a production function can be used to analyze the production structure of a profit maximizing price taking firm without explicit specification of the technology. This also allows movement back and forth from the primal to the dual and vice versa.

Duality provides a great deal of flexibility in empirical studies. Starting with a profit function, the duality properties assures that the resulting system of supply and demand functions is also obtainable from the maximization of a concave production function subject to given fixed inputs under competitive market conditions. The output supply and the input demands so obtained are only functions of the normalized input prices and the quantities of the fixed inputs, variables that are determined independently of the firm's behavior. These variables are exogenous to the system and, therefore, estimation of the system of equations allows avoidance of the simultaneous equation bias in the system.

Why the Profit Function?

Under certain regularity conditions, a production function, a cost function, and a profit function represent the same technology. A profit function that is much more flexible does not presume any *a priori* information compared to the other two functional specifications. The profit function, if not otherwise specified, leads to unconstrained maximization, which also implies that cost minimization (profit maximizing minimum cost) and output maximization (profit maximizing maximum output) have been achieved. In the case of cost function, cost minimizing process yields optimizing conditions for a given level of output that can differ from the profit-maximizing level of output. Similarly, using the production function, the condition of equality between the marginal cost and the output price, i.e., $MC = P$, is necessary for system-wide optimization, so additional conditions are needed when the production function and/or cost function are used. No auxiliary condition, however, is needed to guarantee system-wide optimization when the profit function is used.

Empirically, a production function approach may be inappropriate to estimate the economic efficiency of an individual firm. Theoretically, all price-taking profit-maximizing firms may be expected to produce identical output with identical inputs, provided they have the same production technology and are facing the same set of prices. In practice, firms produce the same set of outputs with varying factor intensities because they face different factor prices and may have different factor endowments. As a result, they have different production functions and different optimal operating points (Ali and Flinn). An estimate of firm-specific

inefficiency should incorporate, as arguments in the analysis, firm-specific prices and the levels of fixed factors. Two firms of equal technical efficiency that have successfully maximized their respective profit would still have different levels of profit as long as they faced different prices (Lau and Yotopolous, 1971, 1972). Inefficiency specifications using the primal approach fails to capture this firm-specific information. The primal approach considers the case where two firms can have varying degrees of technical, allocative, and scale inefficiency but face identical prices. This problem could be avoided by using the dual profit function to represent the production technology. Moreover, the use of the profit function also allows firms to pay and receive different prices for homogenous variable factors and output. Thus, the use of a profit function allows for inter-firm differences in equating the marginal value product (VMP) of variable inputs with their firm specific prices.

Problems of Using Flexible Functional Forms

A recurring theme in econometrics is the conflict between structure and flexibility. The more structure imposed on a model, the better the estimates, provided the structure imposed is correct. When estimating frontiers, the usual choice problem of selecting proper functional form is particularly crucial because the model specification becomes very complicated and sometimes requires modification of conventional definitions of inefficiencies.

Most earlier studies of inefficiency that used stochastic functional specifications with composite error term employed the Cobb-Douglas (CD) specification. One part of the composite error term represents statistical white noise and is generally assumed to follow a normal distribution. The other part represents inefficiency and is assumed to follow a particular one-sided distribution. The major advantage of using the CD functional form specification is its ability to generate closed-form formulae for resultant input demands and for the output supply functions. They can further be used to specify allocative and scale inefficiencies of the individual firm and to quantify the cost associated with these inefficiencies. The disadvantage is the restrictive theoretical specifications (discussed previously in essay 1). Since the estimates of efficiency would be affected by the degree of restrictiveness of the functional form of the technology, it is desirable to use flexible functional forms that do not impose any *a priori* restrictions on the degree of factor substitution. The methodology suggested by Aigner, Lovell, and Schmidt for specifying stochastic functional forms is quite general and can accommodate

any functional form as long as the purpose is limited to the estimation of technical inefficiency and the associated cost. Forsund, Lovell, and Schmidt have shown that the stochastic frontier specification can be extended to dual relationships such as the stochastic cost frontier (Schmidt and Lovell 1979, 1980) and stochastic profit frontier (Ali and Flinn).

Just and Pope pointed out that a useful stochastic functional specification should have sufficient flexibility for the effects of inputs on the deterministic component of the functional form to be different than the effects on the stochastic component.[6] Greene (1980) proposed the use of a stochastic frontier functional specification in the estimation of flexible function forms. Schmidt and Lovell (1979, 1980) and Greene (1980) also emphasized the importance of dual relations. Schmidt and Lovell employed the CD cost function, while Greene (1980) proposed the use of a translog cost function and highlighted the fundamental problem associated with its use. This problem is sometimes referred to in inefficiency literature as "Greene's Problem." Using the conventional definition of inefficiency, a cost function system can be specified that allows for cost inefficiency as

$$\ln C = \ln C(Y, W) + \ln \tau + \ln U + V, \tag{2a}$$

and

$$S = S^{*}(Y, W) + \varepsilon, \tag{2b}$$

where C is the observed cost, $C^{*}(Y, W)$ is the deterministic minimum cost frontier, Y is the vector of outputs, W is the vector of input prices, $\ln \tau$ is the nonnegative term reflecting increased cost due to technical inefficiency, $\ln U$ is the nonnegative term representing increased cost due to allocative inefficiency, V represents statistical white noise, S is the observed share vector, $S^{*}(Y, W)$ is the efficient share vector, and ε is the disturbance vector of share equations that represents the mixture of allocative inefficiency and white noise.[7]

This type of dual representation of inefficiency following Farrell's definition has been used in many applications using the translog cost function. The following characteristics are typical of this type of modeling. The allocative inefficiency

[6] Just and Pope stated that in the familiar CD (production function) case with log-linear disturbances, a risk-increasing effect is incorrectly imposed on output. In the model used here, risk has not been incorporated.

[7] Since any inefficiency leads to increase in cost, both technical and allocative inefficiencies are represented in the cost function by one-sided disturbance terms.

disturbance term in the cost equation, $\ln U$, is related to the allocative inefficiency disturbance term of the input share equations, ε. In the input share equations, allocative inefficiency and white noise may increase or decrease a given input's cost share (i.e., ε can take any sign). In the cost equation, on the other hand, the disturbance terms that represent technical and allocative inefficiency only increases the observed cost (i.e., $\ln \tau$ and $\ln U$ are by definition nonnegative). The white noise term, V, can affect the observed cost either way. Lastly, technical inefficiency does not appear in the input share equations when considered from the cost perspective since output is exogenously determined in the cost minimization framework.

In this type of model specification, the basic problem is how to relate the two-sided disturbance term in the input share equation, ε, with the nonnegative allocative inefficiency disturbance term, $\ln U$, in the cost function, as first noted by Greene (1980) and termed "Greene's Problem." As long as the constant output cost minimization framework is used, the additive disturbance term, ε, may be interpreted as a true representation of the allocative inefficiency because output is exogenous. If, on the other hand, the dual representation of a profit maximization framework is used where the level of output is endogenously determined, the interpretation of the allocative inefficiency term becomes complicated. Schmidt and Lovell (1979, 1980) used the CD cost function where the underlying production structure was homothetic, and, as a result, the input ratios and factor cost shares were independent of the exogenously given output. Thus, the additive disturbance term of the share equations could be interpreted as truly representative of allocative inefficiency. On the other hand, if the production structure is not homothetic, the factor shares and input ratios are not independent of the level of output. Hence, in that case the usual Farrell type of interpretation of allocative inefficiency is not clear. Even if some interpretation of allocative inefficiency in the non-Farrell sense can be given, the essential statistical problem of relating two-sided disturbances of the share equation and the one-sided disturbance of the frontier function still remains.[8]

[8] An analytical solution to this problem is not possible if a non-homothetic profit function is considered; but, in the case of the non-homothetic cost function where output is exogenous to the system and inputs are endogenous, a solution can be derived. The output level in the cost function no longer remains a random variable and cannot possibly be affected by any error. If output is endogenous, it would be affected by technical inefficiency and thus affect the disturbance term of the share equations.

To this point, three possible modeling approaches have been used to remedy Greene's problem. Schmidt and Lovell (1979) and Kumbhakar (1989) tried to find an analytical relationship between the allocative inefficiency disturbances, $\ln U$ and ε. Schmidt (1984) modeled the relationship using an approximating function, imposing all the structure one knows *a priori*. Greene (1980) modeled the system by ignoring the relationship among disturbances in the cost and input share equations. Schmidt and Lovell (1980) also pointed out that given certain dynamic considerations, a model that assumes independence between disturbances may well be appropriate.

The first approach is generally preferred since the exact analytic solution of the relationship can be derived. However, an analytic relationship can be established only when fairly restrictive functional forms are assumed, such as the CD. When a closed-form representation of both the primal and dual functional relations exists, an analytic representation of the relationship among the disturbances is possible. Schmidt and Lovell (1979, 1980) were the first to develop this system approach.

Schmidt and Lovell (1979, 1980) started from a CD production function,

$$y = A \prod X^{\alpha} \exp(\epsilon), \tag{3}$$

where y is the scalar measure of output (due to the cost minimization process), X is the input vector, α is coefficient of X, and $\epsilon (= v+u)$ is the composite error term, where v is the one-sided error term that represents technical inefficiency and u is the symmetric white-noise term. The system of cost and factor demand equations are

$$\ln C = K + \frac{1}{r}\ln y + \frac{1}{r}\sum_{m=1}^{M} \alpha_j \ln w_j + \left\{E - \ln r\right\} - \frac{1}{r}(v+u); \tag{4a}$$

$$\ln x_1 = \ln k_1 + \frac{1}{r}\ln y + \ln\left\{\sum_{m=1}^{M} \frac{p_m^{(\alpha_m/r)}}{p_1}\right\} + \sum_{m=2}^{M} \frac{\alpha_m}{r}\epsilon_m - \frac{1}{r}(v+u); \tag{4b}$$

$$\ln x_j = \ln k_j + \frac{1}{r}\ln y + \ln\left\{\sum_{m=1}^{M} \frac{p_m^{(\alpha_m/r)}}{p_j}\right\} - \epsilon_j + \sum_{m=2}^{M} \frac{\alpha_m}{r}\epsilon_m - \frac{1}{r}(v+u); \tag{4c}$$

$$j = 2, \ldots, M,$$

where w_j is the hiring price of the j^{th} factors,

(5) $$r = \sum_{m=1}^{M} \alpha_m ,$$

is the returns to scale parameter, and

$$k_j = \alpha_j \Big\{ a \prod_{m=1}^{M} \alpha_m^{(\alpha_m)} \Big\}^{-(1/r)} \quad \text{and} \quad K = \ln \Big\{ \sum_{m=1}^{M} k_m \Big\} .$$

Through the cost minimization process Schmidt and Lovell formulated the expression

(6) $$E = \sum_{m=2}^{M} \frac{\alpha_m}{r} + \ln \Big\{ \alpha_1 + \sum_{m=2}^{M} \alpha_m \exp(-\epsilon_m) \Big\},$$

where M is the number of inputs. The term E involves all production function coefficients and the disturbances of the factor demand equations. The errors, both noise and inefficiency, from each equation are analytically present in every equation of the system. Greene's problem is solved because disturbances in the factor demand equations are functionally mapped into the allocative inefficiency term, $\ln U$, in the cost equation. Technical inefficiency, $((1/r)\, v)$, is simply a function of the returns to scale and the one-sided disturbance term, v, in the production function.

Kumbhakar (1989) avoided this problem by estimating the cost minimizing input demand equations derived from a symmetric generalized McFadden cost function that used input-specific, technical inefficiency measures. Since technical inefficiency was the point of interest, the cost equation was not needed to identify all the parameters in this derivation. The cost equation was, therefore, dropped from the system for estimation. The one-sided error term that represented the level of technical inefficiency was added with the cost-minimizing, input demand equation along with a regular error term. With only the input demand equations being estimated, no problem occurs in relating inefficiency in the input share equations to the cost equation. The symmetric generalized McFadden cost function and input demand equations are given as

$$C^*(\cdot) = g(p)y - \sum_i b_i p_i + \sum_i b_{ii}\, p_i y + \sum_i \sum_j d_{ij}\, p_i\, q_j\, y + \sum_j a_j \Big\{ \sum_i \alpha_{ij}\, p_i \Big\}\, q_j$$

$$(7a) \qquad + b_{yy}\Big\{\sum_i \beta_i p_i\Big\}y^2 + \sum_k \sum_j \delta_{kj}\Big\{\sum_i \gamma_{ijk}\, p_i\Big\} q_k\, q_j\, y\,;$$

$$\frac{x_i}{y} = \frac{\sum_j s_{ij}\, w_i}{\Big(\sum_r \theta_r\, w_r\Big)^2} - \frac{\theta_i}{2}\frac{\sum_r \sum_j s_{ij}\, w_j\, w_r}{\Big(\sum_r \theta_r\, w_r\Big)^2} + b_{ii} + \frac{b_i}{y} + \sum_k d_{ik}\frac{q_k}{y} + \sum_k \alpha_{ik}\frac{q_k}{y}$$

$$(7b) \qquad + \beta_i\, y + \sum_l \sum_k \gamma_{lk}\, q_k\, q_l - \tau_i + \nu_i\,;$$

$$j, r = 1, \ldots, n\,; \quad \text{and} \quad k, l, = 1, \ldots, m\,;$$

where y is the scalar output, x is a vector of n variable inputs with corresponding input price vector w, q is a vector of m fixed inputs, ν_i represents white noise that is allowed to follow a first-order autoreregressive process, and τ is the one-sided error term. The function $g(\cdot)$ is defined as $g(p) = (p'Sp/2\theta'p)$ with S being an $n \times n$ symmetric negative semidefinite matrix such that $S'p^* = 0$ and $\theta = (\theta_1, \ldots, \theta_n)'$, a vector of nonnegative constants not all 0. The parameters in θ, along with α_{ij}, β_j, and γ_{ijk}, are assumed to be exogenously given since these cannot be identified. The other parameters, i.e., a_j, b_i, b_{ii}, d_{ij}, b_{yy}, s_{kj}, and δ_{kj}, are estimated to specify the technical inefficiency parameter, τ. Since in the above set of equations (7a and 7b) the input demand function, (7b), exhausts all required parameters and incorporates the technical inefficiency parametr, τ, the cost function is dropped from the estimated system. The efficiency of the ith input is defined as

$$(8) \qquad TE_i = \frac{x_i^*}{x_i} = 1 + \frac{\tau_i}{(yx_i)}\,,$$

where $x_i^* = x_i + \tau_i + \nu_i$ is the minimum quantity of input required to produce a given level of output keeping all other inputs unchanged at the optimum level. Thus, a measure of overall efficiency is defined as

$$(9) \qquad TE = 1 - \frac{w'\tau}{C(y, w, q)}\,.$$

This approach deviates from the conventional Farrell measure of inefficiency and allows some inputs to be used more efficiently than others. The problem with such a measure lies in its treatment of allocative inefficiency. Although not addressed in this model explicitly, allocative inefficiency was present in the model in the sense that the ratio of cost-efficient input demands, derived from the cost

function, is not necessarily equal to the ratio of observed input usage. Moreover, Greene's problem remains unsolved.

For a system of equations as defined by (2a) and (2b) above, Schmidt (1984) proposed an approximation solution. Following Schmidt and Lovell (1979), Schmidt proposed an error specification where $|\varepsilon_i|$ and the $\ln U$ terms are positively correlated. Since the exact analytical form of this relationship is unknown in the case of non-homothetic functional forms, Schmidt proposed modeling the relationship between allocative inefficiency in the cost and input share equations as

$$\ln U = \varepsilon' E \varepsilon , \tag{10}$$

where E is a $M \times M$ positive semidefinite matrix. This specification ensures (or imposes) that $\ln U$ and $|\varepsilon|$ are positively correlated, and when $\varepsilon_i = 0$, the $\ln u_i = 0$. Schmidt suggested that

$$E = D^{1/(M-1)} \Sigma^{+} , \tag{11}$$

where the scalar D is the product of the non-zero eigenvalues of Σ, Σ^{+} is the generalized inverse of Σ, and Σ is the covariance of the input share equation disturbances. As a rationale to this type of specification, Schmidt pointed out that for any fixed E, doubling each element of Σ would double $E(\ln U)$. With the above specification of A, scaler multiplication of Σ raises the $E(\ln U)$ by the same multiple. This is obvious since $\ln U$ is quadratic in ε; and $E(\ln U) = trA\Sigma$. Given this assumption, the variance of disturbances of the input share equations is positively correlated with $\ln U$.

Since the resulting likelihood function is very difficult to estimate under such assumptions, Melfi simplified Schmidt's specification by imposing further structure on the system. Melfi modeled $E = I_M$, where I is an identity matrix of M dimensions and assumed no cross-equation correlation among the input share equations making $\ln U$ the sum of the squared errors of all the share equations. The basic drawback of this specification is that estimates of allocative inefficiency are forced towards zero. The input share residuals are less than one in absolute value, so the sum of the squares will almost necessarily be small. Bauer proposed modeling E as a positive semidefinite diagonal matrix whose elements are treated as parameters to be estimated, setting $E = \delta I$ where δ is a scalar to be estimated

so that the sum of the squared errors from the input share equations can be scaled up (or down) by δ.

Despite these modifications, some problems remain. First, even for a small number of outputs and inputs, a large number of parameters must be estimated. Parameters such as the off-diagonal elements of E and of Σ would be very difficult to estimate in practice without imposing additional structure. Second, solving Greene's problem by modeling this type of relationship among allocative efficiency disturbances does not necessarily lead to better estimates of the cost frontier. Ignoring these relationships may yield better estimates than imperfect modeling (Schmidt and Lovell, 1980).

Greene (1980) tried to solve this problem by ignoring the statistical relationship among allocative inefficiency disturbances across the equations. Disturbances on the input share equations were assumed to follow a multivariate normal distribution with mean zero. Greene recognized the statistical interdependence among allocative inefficiency terms but treated the share equation disturbance term as independent of the one-sided disturbance term of the cost equation. Schmidt and Lovell (1980), considering the appropriateness of this approach, pointed out that, given certain dynamic considerations, a model that assumes independence may well be appropriate. They have amply demonstrated that dispensing with the assumption has almost no effect on their estimates of the frontier, and avoids complication in modeling. This approach is not statistically fully efficient for modeling allocative inefficiency disturbances, but it does not necessarily yield results that are worse than an approach which models the relationship incorrectly.

In this essay Green's methodology (1980) has been followed by ignoring the statistical relationship between the disturbance terms of optimizing conditions and the behavioral equation. Specific analytical relationships are needed only when the impact of allocative inefficiency on the level of cost (or the impact of allocative and scale inefficiencies on the level of profit) is under investigation. However, without decomposing the combined effect of all inefficiencies (represented by the one-sided disturbance term), degrees of allocative and scale inefficiencies (only allocative in the case of cost function and both in the case of profit function) can be inferred from the optimizing equations derived from their derivative properties.[9] Information about the level of technical inefficiency can also be derived if the behavioral functional forms used are self-dual.

[9] Of course, technical inefficiency affects both scale and allocative inefficiencies.

Specification of the Functional Form

The problem of flexible functional forms (FFFs) arises not only because of their inability to yield closed form solutions for the input demands and output supply functions but mainly from the nature of error specifications of the model. The usual multivariate normal (MVN) error specification for a system of equations may fail to adequately capture certain obvious features of the relationships between equations.

Availability of closed form solutions for the duals is not sufficient to show a specific functional relationship between the error terms of the system of equations. Greene (1980) proposed a translog specification to capture the factor interdependence. The basic problem of specifying the functional relationship between error terms remains unsolved (discussed in the last section); however, to achieve the greater level of generality afforded by the flexible functional forms, some modification of the familiar frontier estimations is required.

The translog functional form has been the most frequently used among all flexible functional forms, but the nonavailability of analytical solutions for the input demands and the output supply make it impossible to quantify economic impacts associated with technical, allocative, and scale inefficiencies. Moreover, derivation of the optimizing conditions (share equations) implicitly assumes that the optimum input demands and output supply have been achieved. The disturbance terms associated with the optimizing conditions represent the deviation of the observed share from the theoretical optimum. This can be interpreted as a measure of inefficiency (Greene, 1980), but information about over- or underemployment of inputs and over- or under-supply of outputs could not be obtained from this approach. From the policy perspective, the impact of various inefficiencies on the levels of profit and/or cost cannot be derived, so how much of whic input is over- or underemployed cannot be inferred nor how much of which output is over- or undersupplied and the costs associated with such deviation from optimum.

In this essay, estimation of inefficiency parameters is achieved by using a normalized restricted quadratic profit function (NRQPF) specification. The normalized restricted profit function (NRPF) is defined as the maximized value of normalized profits, given the value of normalized prices of variable commodities and quantities of fixed commodities. Theoretical properties of the NRPF were proved by Lau (1976). Econometric application of the NRPF was first made

by Lau and Yotopolous (1971, 1972) and Yotopolous and Lau in their extensive studies on agricultural production. The NRQPF which is no less flexible than a translog functional form, imposes no constraints on the substitution among inputs and does not assume separability. It can generate closed-form solutions for the input demand and output supply, and it is self-dual. The resulting input demand and output supply functions are linear in parameters. This allows examination of the deviation of optimum from the observed amount of input demands and output supplies. However, like the translog, the NRQPF fails to satisfy the curvature properties globally. The convexity property of the normalized restricted profit function can be verified by examining whether each principal minors of the Hessian matrix is positive semidefinite. The neoclassical restriction of linear homogeniety in prices is represented through normalizing the nominal profit function.

Model

In the econometric literature, the profit function, $\Pi(P, W, Z)$, where P, W, and Z are, respectively, the output price vector, the variable input price vector, and the fixed inputs vector, typically represents a frontier. It characterizes optimizing behavior on the part of an efficient producer, and, therefore, places a limit on the possible values of their dependent variables. A NRQPF can be expressed as[10]

$$\Pi(P,W,Z) = \alpha_0 + \alpha' P + \beta' W + \gamma' Z + \frac{1}{2} P' \delta P + \frac{1}{2} W' \theta W + \frac{1}{2} Z' \eta Z + P' \varphi W + P' \phi Z + W' \psi Z , \tag{12}$$

where $P \in R_+^m$ and $W \in R_+^{n-1}$ are normalized prices of output, and variable inputs. $Z \in R_+^k$ represents the quasi-fixed factors. The nth input price is used as a normalizing factor.

The profit maximization behavior of the firm is exhibited by the profit function its properties of linear homogeneity, symmetry, and convexity. By normalizing the profit function by one of the prices, linear homogeneity in prices is maintained,[11]

[10] Maximizing the NRQPF, $\Pi(\cdot)$, also implies maximization of the nominal profit function, say Π_n, to yield identical values of the optimal input demands and optimum output supplies (Lau, 1976).

[11] It makes no difference, economically or statistically, which price is used to normalize the equation (Schmidt and Lovell, 1979).

which implies the behavioral assumption of price-taking and profit maximization (Lau, 1976). The twice continuous differentiability of the profit function implies symmetry restrictions on the output supply and variable factor demand functions. This guarantees regularity in the process of maximization. The convexity property of the NRQPF depends on the positive semi-definiteness of each principal minor of the Hessian matrix.

A set of dual transformation relations that connects the production function and the profit function can be obtained by applying Hotelling's lemma, viz., $\frac{\partial \Pi(\cdot)}{\partial w_i} = -x_i^*$, and $\frac{\partial \Pi(\cdot)}{\partial p_j} = y_j^* \quad \forall i = 1,\ldots,n-1$, and $j = 1,\ldots,m$, where $*$ represents the solution level. The input demand and output supply equations are, respectively,

$$Y^*(P,W,Z) = \alpha + \delta' P + \varphi' W + \phi' Z\,, \tag{13}$$

and

$$-X^*(P,W,Z) = \beta + \theta' W + \varphi' P + \psi' Z\,, \tag{14}$$

where $X^* \in R_+^{n-1}$ and $Y_+^* \in R_+^m$ are solution output supply vector and solution input demand vector, respectively. The solution input demand for the normalizing factor, x_n^*, can be obtained by substituting the X^* and Y^* in the normalized profit equation, $x_n^* = P'Y^* - W'X^*$. The system of factor demand and product supply functions, (13) and (14), respectively, are linear in normalized prices and fixed inputs.

The stochastic specification in equation (12) can be expressed as

$$\Pi^s(P,W,Z) = \Pi(P,W,Z) + \epsilon\,, \tag{12a}$$

where ϵ is a random disturbance term, and $\Pi^s(\cdot)$ is the stochastic profit frontier. Profit inefficiency is defined as the ability of the firm to attain the profit frontier, given prices and fixed factors, and ϵ represents the deviation from the frontier. Following Aigner, Lovell, and Schmidt and Meeusen and van den Broeck, ϵ could be formulated as the sum of a symmetric component that represents the effects of factors outside the control of the firm (statistical white noise) and a one-sided component that represents profit inefficiency. Thus, the stochastic profit function can be expressed as

$$\Pi^s(P,W,Z) = \Pi(P,W,Z) + \upsilon + \nu\,, \tag{12b}$$

where $v + \nu = \epsilon$. The classical disturbance term is represented by ν, and v represents the one-sided, nonpositive disturbance term ($v \leq 0$)[12] representing profit inefficiency. If $v < 0$, the firm is operating below the profit frontier and is profit inefficient. If, on the other hand, $v = 0$, the firm is operating on the profit frontier. The parameter v, therefore, captures how close an individual producer's production plan (profit level) comes to the maximum level of the profit frontier given prices and fixed factors. Thus, a negative value of v can be interpreted as the percentage reduction in potential profit due to inefficiency. Inefficiency is defined as any downward deviation from the optimum point on the frontier.

Equation (12b) can be estimated by the maximum likelihood method, given pre-specified distributional assumptions regarding the disturbance terms involved, which results in both consistent and efficient estimates. The profit inefficiency of the individual firm can be estimated following the method developed by Jondrow, Lovell, Materov, and Schmidt. However, the full system of equations, i.e., the stochastic profit function and the auxiliary conditions obtained from it (Kumbhakar, Biswas, and Bailey), can be estimated using the envelope theorem. Estimating the full system of equations simultaneously, instead of a single equation, not only improves the precision of the parameter estimates but also yields information regarding the sources of profit inefficiency of the individual firm. However, consistency of the estimated parameters hinges on the specification of the full system. The possible efficiency gain in estimation occurs not only because of cross-equation error correlations but also due to cross-equation restrictions imposed by the model specification. Thus, consistency depends to a great extent on the correctness of the distributional assumptions and on the specification of the deterministic part of the system of equations (Schmidt, 1985-86).

Equations (13) and (14) can be expressed, appending disturbance terms, as

$$Y^*(P,W,Z) = \alpha + \delta' P + \varphi' W + \phi' Z + \Omega ; \tag{13a}$$

and

$$-X^*(P,W,Z) = \beta + \theta' W + \varphi' P + \psi' Z + U . \tag{14a}$$

[12] Note that the one-sided disturbance term v is an integral part of the model specification. The profit function is stochastic in this sense. Aigner, Lovell, and Schmidt decomposed the intercept term for stochastic specificaton : $\alpha_0 \exp(v)$. Here, we specify the one-sided disturbance as an additive term, rather than multiplicative (Just and Pope).

The disturbance term vectors Ω and U can take any value. The terms Ω and U represent deviation of the observed output supplies and input demands from the optimum output supplies and input demands, respectively. Terms Ω and U may appropriately interpreted as measures of scale and allocative inefficiency, respectively (Greene, 1980), and as long as Ω and U represent deviations from the optimum, they can be treated as measures of inefficiency. This definition of allocative inefficiency is somewhat different from the conventional definition provided by Forsund, Lovell, and Schmidt,[13] which obtains the allocative inefficiency measure for every input from the optimizing conditions. Forsund, Lovell, and Schmidt defined a firm to be allocatively inefficient if $f_i/f_j \neq w_i/w_j$, $\forall\, i,j = 1,2,\ldots,n \;\; i \neq j$, where f_i is the marginal product of input x_i, and w_i is the hiring price of input x_i. The dual transformation of the same definition in the case of the profit function is $\Pi_i/\Pi_j \neq x_i/x_j$, where Π_i is the partial derivative of profit function with respect to the ith input price, w_i. Schmidt and Lovell (1979) defined allocative inefficiency as $\ln(x_j/x_1) \neq \ln(x_j^*/x_1)$, i.e., with x_1 is used as the numeriare. Both types of allocative inefficiencies can be derived from this model.

If a firm is supplying the appropriate output, i.e., where $MC = P$ is achieved for the firm, then $\Omega = 0$. Deviation in either direction would increase the cost of production and reduce the level of profit. From the $MC = P$ condition, the output supply equation can be derived by substituting the solution input demands, X^*, in the direct cost function, $C = W'X^*$. If $MC \neq P$, then scale inefficiency exists. In the dual transformation, this relation is represented by (13a). If $\Omega \neq 0$, the firm is not supplying the right output. Similarly, the firm is hiring the appropriate combination of inputs if the $\text{MRTS}_{it} = \frac{w_i}{w_t}, \forall\; i,\, t;\;\; i \neq t$. (MRTS represents the marginal rate of technical substitution between inputs.) Again, a profit-maximizing firm would hire factors at the point when the hiring price of a factor equals the value of the marginal product, i.e., $VMP = W$. If $VMP \neq P$, then allocative inefficiency exists. This relationship is reflected by (14a). Deviation on either side increases cost and reduces level of profit. The vector U, therefore, represents the pure allocative inefficiency provided the profit function is homothetic. The influence of the output price vector, P, on Π^* is manifested only through an output adjustment. Hence, if the cost minimization input ratio is independent of output, the associated profit-maximizing input ratios

[13] Forsund, Lovell, and Schmidt's definition is in terms of a pair of inputs, i.e., allocative inefficiency of one input with respect to another input.

must also be independent of output price, P. If $\varphi' l = 0$ is forced (where l is an unit vector), then Ω and U represent solely the scale and allocative inefficiency parameters, respectively.

In the model, the definition of allocative inefficiency must to be modified. Since output price enters the input demand equation, it appears that technical inefficiency also affects X^*. In the cost minimization case (constant output), the output appears in the input demand equation; but output, being exogenous, does not affect the level of allocative inefficiency. By similar reasoning, although P appears in the input demand equations when using the profit function, it does not affect allocative inefficiency. Thus, u_i is interpreted as the degree of over- or underemployment of a factor that causes the firm to operate below the profit frontier. Similarly, allocative inefficiency in this model does not affect the level of scale inefficiency since W input prices are exogenous to the system. In this case, technical inefficiency cannot be derived but rather confounded in the profit inefficiency term, v.

Even if homotheticity is assumed, profit inefficiency as given by v is affected by both Ω and U, i.e., $v = v(\Omega, U)$, but it is not clear how Ω and U are related to v because Ω and U enter the profit function in a very complicated manner (the Greene problem). Regardless of the sign of Ω and U, their contribution to the profit function is always negative. The exact statistical relationship between v, Ω, and U is unclear, particularly in view of the contribution of Ω and U to v. Moreover, another important component of inefficiency, i.e., technical inefficiency, cannot be identified separately in this type of model because it is confounded in the profit inefficiency term, v. When monotheticity is assumed in a model, one way of avoiding this complication is to assume them to be independent and estimate the equation system. In this case, only the values of the allocative and scale inefficiency parameters can be obtained from the estimates of U and Ω from (3a) and (4a). This is no doubt a simplifying assumption. Schmidt and Lovell (1980) have amply demonstrated that dispensing with this assumption has almost no effect on their estimates of the frontier and further that its elimination saves algebric complications. Given certain dynamic considerations, a model that assumes independence may well be appropriate (see the subsection Problems of Using FFF for detailed discussion). In this model, Ω and U are interpreted as modified measures of scale and allocative inefficiencies subject to the assumption of mutual independence among v, Ω, and U.

Ali and Flinn recently estimated the profit inefficiency of rice producers in Pakistan using a single equation translog profit function. This essay extends their analysis to formulate a simultaneous equation model for estimating profit inefficiency and to shed light on the underlying allocative and scale inefficiencies.

To estimate the profit, scale, and allocative inefficiencies, the system of equations defined by (12b), (13a), and (14a) must be estimated. The system contains $(n+1)$ independent equations ($[n-1]$ input demend equations, output supply equation, and the profit equation) to solve for $(n+1)$ endogenous variables ($[n-1]$ inputs, X, output, Y, and profit, Π). Parameters involved in the above set of equations can be estimated and used to estimate the inefficiency parameters, i.e., v, Ω, and U.

Estimation Procedure

In principle, the parameters in equation (10a) are estimated using ordinary least square (OLS) or instrumental variables (IV) methods. Of course, this neglects the nature of the disturbance terms involved and does not separate v from ν. The maximum likelihood estimation method will capture the specification of the disturbance terms; however, as the number of inputs increases, the number of parameters becomes quite large. If the model contains more than two inputs, the question of multicollinearity will likely become severe. Thus, the single equation estimation will be imprecise.

Berndt and Christensen have shown how this problem can be avoided in a standard setting by using auxiliary optimizing conditions, i.e., factor demand and output supply equations along with the behavioral function. The resulting system could be estimated using methods like full information maximum likelihood, (FIML), or seemingly unrelated regression, (SUR), which yield maximum likelihood estimates when the iteration process converges. These simultaneous equation methods take care of the cross equation dependence but fail to separate the effects of v and ν. Therefore, the maximum likelihood method will best estimate the set of equations (12b), (13a), and (14a) with specific distributional assumptions regarding the disturbance terms involved. Unless one has panel data, specific distributional assumptions about the one-sided component of the error term must be made in order to obtain individual firm inefficiency. When panel data is available, specific distributional assumption for the one-sided error term can be avoided. Of course, in this case a structure of efficiency variation

over time must be imposed. However, the estimation process becomes manageable if specific distributional assumption is made for the one-sided error term, and application of the maximum likelihood method enables separation of the effects of v from that of ν.

The system of simultaneous equations incorporating profit inefficiency, allocative, and scale inefficiency can be expressed as

$$(12c)\quad v+\nu = \Pi^s - (\alpha_0 + \alpha' P + \beta' W + \gamma' Z + \frac{1}{2}P'\delta P + \frac{1}{2}W'\theta W + \frac{1}{2}Z'\eta Z + P'\varphi W + P'\phi Z + W'\psi Z);$$

$$(13b)\quad -U = X^* + \beta + \theta' W + \varphi' P + \psi' Z;$$

and

$$(14b)\quad \Omega = Y^* - (\alpha + \delta' P + \varphi' W + \phi' Z).$$

To estimate the above system of equations, the joint probability density function (pdf) of the error vectors must be derived with specified distributional forms. At this point, the dimension of our model can be reduced from multiple output framework to single output, by assuming that Y^* is a (1×1) vector, which also implies the Ω vector to be a (1×1) vector. The error vector from the system of equations (12c), (13b), and (14b)

$$(15)\quad \begin{pmatrix} v+\nu \\ \omega \\ u_1 \\ \vdots \\ u_{n-1} \end{pmatrix}$$

forms the probability density function (pdf) of v, ν, ω, and u_i, where, $i = 1, \ldots, n-1$. Assumptions regarding the distribution of the individual error terms of the error vector (15) are

1. $v \sim iid\ N(0, \sigma_v^2)$ truncated at zero from above;
2. $\nu \sim iid\ N(0, \sigma_\nu^2)$;
3. $\omega \sim iid\ N(0, \sigma_\omega^2)$;
4. $U \sim MVN(\mu, \Sigma)$; and

5. v, ω, ν, and U are independent of each other and also independent of prices and fixed factors.

Let $b_1 = v + \nu$, $b_2 = \omega$, and $b_3 = U$. The joint pdf of b_1, b_2, and b_3 is

$$f(b_1, b_2, b_3, v) = f(v, \nu, \omega, U)\,|J|, \tag{16}$$

where J is the Jacobian of transformation. The Jacobian for variable transformation in this case is given by

$$J = \frac{\partial(v,\ \nu,\ \omega,\ U)}{\partial(b_1,\ b_2,\ b_3,\ v)}.$$

In this case, $|J| = 1$, and since, by assumptions, v, ν, ω, and U are independent, equation (16) can be expressed as

$$f(b_1, b_2, b_3, v) = f(v)\, f(\nu)\, f(\omega)\, f(U), \tag{17}$$

where $f(v)$, $f(\nu)$, $f(\omega)$, and $f(U)$ represent the pdf of v, ν, ω, and $f(U)$, respectively. After some algebraic manipulations, the joint pdf can be expressed as

$$f(b_1, b_2, b_3, v) = \frac{2\exp(-d/2)}{(2\pi)^{\frac{n+2}{2}}\,\sigma_v\,\sigma_\omega\,\sigma_\nu\,|\Sigma|^{\frac{1}{2}}} \exp\Big\{-\frac{1}{2}\Big(\frac{v-\mu_v}{\sigma}\Big)^2\Big\}$$

$$\exp\Big\{-\frac{1}{2}(b_2-\mu)'\,\Sigma_U^{-1}(b_2-\mu)\Big\}\exp\Big\{-\frac{1}{2\sigma_\omega^2}\,\omega^2\Big\}, \tag{17a}$$

where $d = \left\{\frac{b_1}{\sigma_v^2+\sigma_\nu^2}\right\}$; $\mu_v = \left\{\frac{b_1\sigma^2}{\sigma_\nu^2}\right\} = \left\{\frac{b_1\sigma_v^2}{\sigma_v^2+\sigma_\nu^2}\right\}$; $\sigma^2 = \left\{\frac{\sigma_v^2\sigma_\nu^2}{\sigma_v^2+\sigma_\nu^2}\right\}$.

To get the joint pdf of b_1, b_2, and b_3, integrate out the v from (17a), i.e.,

$$f(b_1, b_2, b_3) = \int_{-\infty}^{0} f(b_1, b_2, b_3, v)\,dv, \tag{18}$$

$$= \frac{2\exp(-d/2)}{(2\pi)^{\frac{n+2}{2}}\sigma_v\,\sigma_\nu\,\sigma_\omega\,|\Sigma|^{\frac{1}{2}}} \exp\Big\{-\frac{1}{2}(b_2-\mu)'\Sigma_U^{-1}(b_2-\mu)\Big\}$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} \exp\Big\{-\frac{1}{2\sigma^2}(v-\mu_v)^2\Big\}dv.$$

Since

$$(19) \qquad \sigma \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(v-\mu_v)^2\right\} dv = \sigma\,\Phi\left(-\frac{\mu_v}{\sigma}\right),$$

where $\Phi(\cdot)$ is the cumulative density function (cdf) of a standard normal, the joint pdf of b_1, b_2, and b_3 can be expressed as

$$(20) \qquad f(b_1,b_2,b_3) = \frac{2\exp(-d/2)\,\sigma\,\Phi(-\mu_v/\sigma)}{(2\pi)^{\frac{1}{2}(n+1)}\,\sigma_v\,\sigma_\nu\,\sigma_\omega\,|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\frac{\omega^2}{\sigma_\omega^2}\right\} \exp\left\{-\frac{1}{2}(b_2-\mu)'\,\Sigma_U^{-1}\,(b_2-\mu)\right\}.$$

From the joint pdf of the error vector (15), the pdf of the endogeneous variable vector, $(\Pi^s\, X\, Y)'$ can be derived. By variable transformation,

$$(21) \qquad f(\Pi^s, Y, X) = f(b_1,b_2,b_3)\,|J|\,.$$

The log likelihood function for the T firms is, therefore,

$$(22) \qquad L = \sum_{t=1}^{T} \ln f(\Pi_s, Y, X) + T \ln |J|\,,$$

where J is again the Jacobian of transformation. Because $|J| = 1$, the log likelihood function for a single firm can be expressed as

$$(23) \qquad L = \ln 2 - \frac{d}{2} + \ln\sigma + \ln\Phi\left(-\frac{\mu_v}{\sigma}\right) - \frac{m+2}{2}\ln(2\pi) - \ln\sigma_\omega - \ln\sigma_\nu - \ln\sigma_v - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(b_2-\mu)'\,\Sigma_U^{-1}\,(b_2-\mu) - \left\{\frac{1}{2\sigma_\omega^2}\omega^2\right\},$$

where

$$\mu = \frac{1}{T}\sum_{t=1}^{T}(X^* + \beta + \theta' W + \varphi' P + \psi' Z)\,.$$

The maximum likelihood estimates of the model's parameters can be estimated by maximizing (23) (or minimizing the negative of [23]); however, the estimation process can be simplified by concentrating the likelihood function. Since the

additive constant terms are disregarded in maximizing the likelihood function, L, the terms $\ln 2$ and $\frac{n+2}{2}\ln(2\pi)$ are eliminated. The term $(b_2-\mu)'\Sigma_U^{-1}(b_2-\mu)$, on the right-hand side of (23) is a scalar, which equals its trace, and the trace of a product is invariant to certain orderly permutations of the order of multiplication, so that

$$
\begin{aligned}
C &= tr(b_2-\mu)'\,\Sigma_U^{-1}(b_2-\mu) \\
&= tr\left(\Sigma_U^{-1}(b_2-\mu)(b_2-\mu)'\right) \\
&= tr\,\Sigma_U^{-1}\Sigma_U \\
&= trI \\
&= (n-1).
\end{aligned} \tag{24}
$$

The concentrated likelihood function, therefore, becomes

$$
\begin{aligned}
L_c = \Lambda - \frac{d}{2} + \ln\sigma + \ln\Phi\left(-\frac{\mu_v}{\sigma}\right) - \ln\sigma_v - \ln\sigma_\nu - \frac{1}{2}\ln|\Sigma| \\
- \ln\sigma_\omega - \ln\frac{1}{2\sigma_\omega^2}\omega^2 .
\end{aligned} \tag{25}
$$

The constant term $\Lambda = \ln 2 + n - 1 - \frac{n+2}{2}\ln 2\pi$ can be dropped for estimation purposes. The MLE of all relevant parameters, i.e., α, β, γ, δ, θ, η, φ, ϕ, ψ, σ_v, σ_ω, σ_ν, and the elements of Σ_U can be estimated by maximizing (25), and the resulting estimates are asymptotically efficient and consistent.

Estimation of Inefficiency Parameters

Profit inefficiency is defined as the combined inefficiency. Given the MLE of (25), v, ω, and U must be estimated.

Profit Inefficiency

To derive v, ν is to be isolated from v, where $b_1 = \nu + v$ is the residual of the profit function. This can be done by estimating the conditional pdf of v given b_1. The conditional pdf of v given $b_1 \sim N(\mu_v, \sigma_*^2)$ truncated at zero (see appendix 3.a for proof). Following Jondrow, Lovell, Materov, and Schmidt, profit inefficiency can be estimated at a point estimate of v. Either mean or mode can be used as the point estimator of v. The mean of the conditional distribution of

v is

$$E(v|b_1) = \int_{-\infty}^{0} v f(v|b_1)\, dv, \tag{26}$$

where the mean and mode of v are respectively given by

$$v_{mean} = \mu_v - \sigma\left\{\frac{\phi(\mu_v/\sigma)}{\Phi(-\mu_v/\sigma)}\right\}; \tag{27}$$

and

$$v_{mode} = \begin{cases} \mu_v, & \text{if } (\mu_v/\sigma) \leq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

(See appendix C.1 and C.2 for proof.)

Allocative Inefficiency

The level of allocative inefficiency for each firm for each input can be obtained from the estimate of the equation,

$$-\hat{U} = X^* + (\hat{\beta} + \hat{\theta}'W + \hat{\varphi}'P + \hat{\psi}'Z). \tag{29}$$

The allocative inefficiency of the tth firm, for the ith input is therefore, $\hat{u}_{ti}$. The elements of $\hat{U}$ i.e., $\hat{u}_j$ can be both positive or negative; in either case the cost of production increases.

Scale Inefficiency

Scale inefficiency can be obtained from the output supply equation as

$$\hat{\omega} = Y^* - (\hat{\alpha} + \hat{\delta}'P + \hat{\varphi}'W + \hat{\psi}'Z), \tag{30}$$

where $\hat{\omega}$ can take any value.

Impact of Inefficiency

The impact of various inefficiencies on the level of profit, i.e., the potential profit loss due to inefficiency, is of prime importance from the policy point of view. Profit inefficiency, v, gives the measure of potential profit loss due to the

combined effect of all inefficiencies. Given Z and prices, the percentage loss of profit due to profit inefficiency of the tth firm becomes

$$\Pi L_{v_t} = \frac{\Pi_t - \Pi_t^*}{\Pi_t^*} = \frac{v_t}{\Pi_t^*}. \tag{31}$$

This loss of profit is due to allocative, scale, and confounded technical inefficiency.

The degree of allocative inefficiency (defined as over or underemployment of inputs) and scale inefficiency (defined as over- or underproduction) may be obtained by using the input demand and output supply equations. The percentage deviation from the optimum allocation point for the tth firm for jth input can be obtained from

$$UL_{jt} = \frac{x_{jt} - x_{jt}^*}{x_{jt}^*} = \frac{\hat{u}_{jt}}{x_{jt}^*}. \tag{32}$$

Similarly, the percentage of over or under-production can be obtained from

$$\Omega L_t = \frac{y_t - y_t^*}{y_t^*} = \frac{\hat{\omega}_t}{y_t^*}, \tag{33}$$

where $\hat{u}_{jt}$ and $\hat{\omega}_t$ can be obtained from (29) and (30), respectively.

Data Description

The data set used here has been collected from three agro-climatic regions[14] of the Indian state of West Bengal. A cross-section random sample of 250 household farms were interviewed over the period of 1980–1983. Since these regions were surveyed over a four-year period, the bias of the data set caused by time specific problems (such as flood, drought etc.) could be avoided.[15] These three regions cover the entire rice production belt West Bengal. Over the last two decades, the success of technological change, introduction of high-yielding seeds, security of

[14] Region I covers two districts, Nadia and Murshidabad. Region II covers the districts of Burdwan and Birbhoom, and region III covers the districts of Howrah, Hooghly, 24 Parganas, and Midnapore. A district is equivalant to a U. S. county jurisdiction.

[15] Although, the data were collected over a period of four years, it can be used for this cross section analysis. Greene (1989) pointed out that if a fairly large number of observations is collected over a short period of time, then cross section use of that data set does not affect the quality of results.

land tenure, availability of production loans, and relaxation of government controls over purchase of output and sale of some vital factors, (e.g., fertilizers, irrigation water, and pesticides) have made the agricultural production system very much market oriented. Under such an environment, profit-maximizing behavior on the part of producers is a reasonable maintained hypothesis.

The agricultural production activities of West Bengal are spread over at least three crop seasons during a one-year production cycle. The monsoon paddy is the single most important crop in all three regions. This crop is produced by almost all peasants in these three regions, with almost 70% of the total arable land used in this particular cultivation (Dasgupta).[16] The other two crop seasons are winter production of pulses and oilseeds and summer crops of wheat.

The survey provides a wide range of information about both fixed and variable inputs used in one production activities. Among inputs, fertilizer (F), human labour (H), and bullock labour (B) are used as endogenous variable inputs, while land and capital are two fixed inputs. The same capital base is assumed for each crop. Fertilizer is measured in terms of kilograms, while both types of labour are measured in terms of labour hours. The monsoon[17] crop is the only output considered, measured in terms of kilograms. In this model, fertilizer is defined as a combination of manure and chemical fertilizer expressed per unit of fertilizer.

Only during the monsoon season is scarcity of resource considered a consequence (Hopper). At no other period in their farms' operations are the allocation choices more crucial for an Indian farmer.[18] All factors, therefore, are truly competitive during this season. Since the size of the operational land holdings is very small, the role of machine capital is very limited. Both bullock and human labour play very important roles. Human labour includes both family and hired labour, and it is impossible to separate these two components in the data set, although effects of these two types of labour on production are quite different. Family labour plays a dual role, both as manual labour and as a managerial decision-making unit. Thus, the returns to these two types of labour are different. One

[16] The quality of the crop may differ accross regions and between individual plots of a single region; however, this quality difference is not introduced in this model. Prices of factors of production and output to certain extent capture the quality difference in this model.

[17] Monsoon cropping spans the period June to September.

[18] Due to this reason, almost all empirical studies on Indian agriculture have typically used data for the monsoon crop.

way of capturing the impact of the managerial skill of family labour may be to introduce the level of education of the household as an explanatory variable; in this essay, however, education is not considered. Rather wages received by each labourer are used to represent the hiring prices of all labour. It has been argued that (Saini) in the context of Indian agriculture, valuation of family labour at the ruling market wage rate is quite justified for all size classes of farms. Sen pointed out that if family labour employed in agriculture is given an imputed value in terms of ruling wage rate, much of Indian agriculture seems unremunerative. However, all major empirical work on Indian agriculture has used the prevailing market wage to value domestic labour, and results show that Indian farmers are quite rational in allocating their factors based on market prices. Moreover, about 65–70% of Indian farmers participate in the labour market (Rosenzweig), so it seems rational to use market wage to value all types of human labour. Human labour is measured in terms of working days (eight hours per day).

The market for animal labour is highly localized, and most farmers own at least a pair of bullock. During the peak season, such as the monsoon, a market exists for animal labour. To avoid double counting the total number of hours use of animal labour has been treated as an explanatory variable, while the value of farm-owned animals (for farming) has not been included in fixed capital. Although the usual practice is to hire bullock labour on a per day basis, in this data set bullock labour is not expressed in terms of working hours. For estimation, total working hours are converted to working days (eight hours per working day).

Another group of variable factors is fertilizer, which includes both chemical fertilizer and manure. The market for manure is highly localized but quite competitive, particularly during the monsoon season. On the other hand, the market for chemical fertilizers has a dual nature. Since the supply of chemical fertilizer is partly regulated, a parallel market exists for this input, and the farmer has free access to that market. Therefore, the actual price of chemical fertilizer rather than the controlled price is used in this analysis. Very likely, different farms use different types of chemical fertilizer (or combination of different fertilizers) at different doses. This data set does not provide detailed information regarding various types of fertilizers used by a farm. The total amount of fertilizer used by each farm for each crop can only be obtained from the data set.

Besides these four variable endogenous inputs, two fixed inputs considered are land and physical capital. Land is actually the allocable fixed input, but in our

model land has been treated as fixed input.[19] In this static model, however, it does not matter if the land input is treated as a fixed factor. The amount of land (acres) under cultivation, rather than the amount of land owned a by farmer, has been considered in this model.

The other fixed factor, capital, is defined as combined of value of all durable farm assets, namely farmhouse, irrigation equipment, and farm equipment. The total value of capital has been deflated by the user cost of capital, which is defined as the sum of the yearly depreciation rate and the market rate of interest less the yearly inflation rate, assuming a straight line depreciation. The interest rate is the average market interest rate over the period of survey.

The data set shows that hiring prices of factors vary not only between years covered by the survey but also across regions. Variation in output and input prices can be attributed in part to differences in transport cost, storage cost, types of forward contracts, and types of land tenancy contracts. Quality differences in outputs and inputs may be another factor in price variations. Though floor wage and support prices were in effect during the years covered by the data set, there is wide variartion on either side of the official rate. The price received by an individual farmer for his produce also depends on how long he can hold the sale. The pattern of agriculture in India to a great extent depends on the mercy of the monsoon, and farmers match their farming schedule the monsoon pattern. This is another reason for wide fluctuation in rupee prices of some inputs like fertilizer, bullock labour and human labour.

Empirical Results

Because of the diverse farm size and high dispersion of profit (and also following convention), these observations have been divided into two nonintersecting subsets

[19] A farmer can allocate land between products given the total amount of land, mainly due to three reasons. First, some farmers keep a part of their land fallow as part of their optimizing decision. Second, the total operational land holding of a household farm can be fragmented and the plots qualitatively different rather than being homogenous. Qualitatively, some plots may be unsuitable for cultivation of a specific crop, e.g., a summer crop that needs generous irrigation. Similarly, shallow land might not be used for monsoon paddy. Third, a well developed market exists for leasing land under the land tenancy and share cultivation system. From the point of view of the farmer, there is a process for allocation of variable factors but also for the allocation of quasi-fixed factors e.g., land.

of large and small farms.[20] Table 1 gives some descriptive statistics regarding both types of farms under consideration. Land holding size has been used to characterize farms into these subgroups. Various authors used different farm sizes as the point of truncation between large and small farms, following Bardhan, this study choose seven acres as the point of truncation.[21] Farms with more than seven acres of land under cultivation were classified as large farms, and farms holding less than seven acres were classified as small. Three separate models were estimated to check whether the production structure of small farms (Model B) and that of large farms (Model C) differ from the production structure reflected by the pooled data (Model A). If so, then an analysis of farm-specific inefficiency based on pooled data is not appropriate. A likelihood ratio test rejects the hypothesis of a single profit function for all size classes. The test statistic in the case of small farms, $-2(L_U - L_R) = 212.12$ is greater than χ^2 with 27 degrees of freedom for any reasonable significance level. Similarly, for large farms the test statistic, $-2(L_U - L_R) = 388.37$, is greater than χ^2 with 27 degrees of freedom for any reasonable level of significance. Thus, the use of a single profit function (Model A) for our purpose is not appropriate; however, parameter estimates of Model A are reported in table 2 for comparison.

Parameters of the likelihood function are reported in table 2. Most relevant parameters are significantly different from zero.[22] In the econometric analysis, the derivative properties of economic behaviors are reflected by the estimated coefficients. Therefore, some economic conclusions can be drawn from values and signs of the coefficients of the estimated model. For the large farms, the intercept term, α_0, is negative, indicating a high fixed-cost component. All price coefficients (input and output) are of correct sign,[23] as are coefficients of the fixed factors. Although the contribution of capital to profit is positive for both farm sizes, the impact is only significant for large farms. For small farms, the

[20] In Indian agriculture, especially in the context of West Bengal, this distinction is very important from a policy point of view.

[21] Some authors have used a larger farm size as a point of truncation between large and small farms, e.g., Lau and Yotopolous (1971) who used ten acres as the dividing line. In the eastern part of India, where the fragmentation of land is very high, average size of land holdings is lower than northern and southern India.

[22] For small farms, coefficients α_0, α_P, β_H, β_B, and γ_L are significant at the 5% level. For large farms, coefficients α_0, α_P, β_B, and γ_K are significant at the 5%.

[23] In this type of model the possibility of multicollinearity is very high, which may affect the signs of estimated coefficients.

impact of land is significant, but the influence of capital is not. The impact of the exogenous output price change on the level of profit is higher for large farms than for small farms. On the other hand, the impact of input prices on level of profit is stronger and significant for small farms relative to large farms; thus, the profit of small farms is more affected by input price changes than are profits of large farms. However, the profit structure of the large farms depends more on variation of output prices relative to that of input prices. This is typical of any underdeveloped country's agriculture system where the resource constraint, especially financial resources, is the most important binding factor. Small farm operators try to maximize their profit by minimizing their factor cost, whereas large farms try to maximize profit by maximizing revenue.

Earlier studies of the economic efficiency of Indian farms have suggested that Indian farms are both allocative and profit efficient and further that small farms are allocatively more efficient than their larger counterparts. Most of these studies, however, tested the hypothesis of efficiency associated with input use and/or the hypothesis of the rationale of exact profit maximization by Indian farmers. Earlier studies, all parametric, have used nonfrontier functional specification to investigate their hypotheses. The idea of a nonfrontier function assumes that all farms share a common family of production, cost, and profit frontier, with all variation in farm performance attributed to variation in firm efficiencies relative to the common family of frontiers. Although this scenario conforms to its theoretical justification, it proves difficult to justify empirically. To this point, stochastic functional specification has not been applied to investigate profit, allocative, and scale inefficiency questions of Indian farms, a gap this study will bridge.

The upper part of table 3 shows the percentage of potential profit loss due to inefficiency. To conserve space, only the mean inefficiency rather than individual farm inefficiency of relevant classes is reported. The second portion shows allocative and scale inefficiencies associated with the different size classes. Using equation (31), the potential profit losses (in percentage terms) are estimated for all three models. Small farms which are more profit efficient than their larger counterparts, on average lose only 6% potential profit due to inefficiency, whereas the large farms on average lose just over 15%. These figures clearly support earlier results which suggested that smaller farms are more efficient in the context of Indian farming. On the other hand, this result is contrary to earlier results that emphasized profit efficiency of Indian farms.

The causes of these profit inefficiencies and differences can be obtained from the second part of table 2. Using equation (32), the percentage of allocative inefficiency (mean) is estimated for all three variable inputs. Small farms are more efficient than larger farms in allocating inputs like fertilizer. Small farms over utilize fertilizer 7.4% relative to the requirement of profit maximization, while large farms use 16% more than the optimum requirement. On the other hand, larger farms allocate their bullock labour input more efficiently than small farms. Large farms overemploy this input by only 8.5%, while small farms underutilize the same input by more than 28% compared to the optimum. Again, while small farms overutilize human labour by 31%, large farms on average underemploy this factor by almost 40%. An inference from this information may be that small farmers substitute human labour for fertilizer and bullock labour, whereas large farms substitute fertilizer for human labour. This is also reflected in the cross-price effects.

Assuming farmers are profit maximizers, an important issue is whether farms are supplying the optimum level of output for the prices involved. Estimates of the output supply equation and its percentage deviation from the actual output suggest that both types of farms overproduce relative to the optimum level of production. Large farms on average produce 12% more than the optimum amount, while small farms are supplying on average 34% over the optimum. This suggests that in general farming is characterized by overproduction, especially the small farms, relative to the prices sustaining such supply. The tendency of small farms is to substitute human labour for bullock labour; instead of hiring other market factors, they utilize family labour. On the other hand, fertilizer is substituted for labour on large farms because in the Indian situation large land holders usually are involved in other economic activities. So the opportunity cost of family labour is quite high for the large farms.

Conclusions

The concluding section will summarize the major results of this essay and highlight limitations and possible extensions. The first section covers major conclusions, while the second section points out some limitations of this study. In the final section, some possible future extensions of this approach will be proposed.

Major Results

A stochastic NRQPF has been estimated along with its auxiliary optimizing conditions to determine profit inefficiency of Indian farmers. In the process, allocative inefficiency for each input and scale inefficiency are estimated using the nonlinear simultaneous equation maximum likelihood method. Newton's algorithm was used to solve the system of equations. Estimated maximum likelihood parameters were applied to obtain inefficiency estimates. The allocative inefficiency definition in this essay is different from the definitions used by Schmidt and Lovell (1979, 1980) and Forsund, Lovell, and Schmidt whose estimates of allocative inefficiency can be estimated from these estimates.

This method has at least two advantages. First, use of a NRQPF allows estimation of inefficiency without any *a priori* restrictive assumptions on the structure of factor substitutions. Second, without going into the impasse of Greene's problem, over- and/or underallocation of inputs and output underlying the profit inefficiency of an individual farm can be estimated. The total profit loss due to inefficiencies (allocative, technical, and scale) can also be obtained from the profit function.

The application of this model to Indian rice cultivation provides some interesting results. Log likelihood ratio tests allow decomposition of the data set into two subgroups, small and large farms. Three models have been estimated to compare efficiency performance between two groups of farms. The empirical results generate five fundamental results. First, small farms are more profit efficient than large farms. Second, large farms are less scale inefficient with respect to allocating their output compared to small farms. Third, small farms are less inefficient than large farms in allocating inputs like fertilizer. Four, large farms are less efficient than the small farms in allocating bullock labour input. Finally, regarding use of human labour, both small and large farms are inefficient; small farms overutilize human labour while large farms underutilize it.

The set of assumptions regarding distribution of the stochastic error terms are quite strong. Specification of the model and results of the estimation process depend on the correctness of such strong assumptions; which must be checked empirically. A normal probability plot for each error term has been chacked, and the empirical distribution of each set of estimated residuals is quite close to normal distribution. The set of assumptions made appears consistent with the model for the data set used.

Limitations

The basic limitation of this study is Greene's problem. The percentage loss of potential profit and underlying over- or underallocation of inputs and output can be determined; however, the loss of profit due to individual inefficiency elements cannot be determined. More precisely, the profit inefficiency parameter, v, cannot be decomposed into its constituent elements. The profit inefficiency experienced by each farm results from the combined effect of technical (τ), allocative (U), and scale (ω), inefficiencies, from the policy point of view a very important point. However, policy prescription regarding allocation of inputs and output can be made from the study's estimate. Similarly, the cost of allocative inefficiency, as defined by Schmidt and Lovell (1980, 1979), cannot be determined in this case due to nonhomothetic functional specification.

This essay contains no measure of the technical inefficiency of farms, but technical inefficiency is very much a part of total profit inefficiency. Technical efficiency as defined by Farrell must be independent of prices, so, technical inefficiency cannot be determined directly from the profit function as derived. Profit inefficiency may, of course, be defined as a sum of three independent inefficiencies (technical, allocative, and scale), but not without complications in the estimation process and model specification. The following section proposes two ways of incorporating technical inefficiency into this model.

Possible Extensions

Four possible extensions of this model might be considered. Technical inefficiency can be introduced in this model in two different ways. First, because a NRQPF is self-dual, the analytic form of the underlying production function can be derived and the technical inefficiency in the production structure specified.[24] A second way specifies the term v as $v = \tau + \rho + \kappa$, where κ is the potentional profit loss due to scale inefficiency, ρ is loss of profit due to allocative inefficiency, and τ is the loss of potential profit due to technical inefficiency (i.e., in our model $\epsilon = \tau + \rho + \kappa + \nu$). In the first case, a simultaneous equation approach would be a good method to solve the parameters of interest. For the second case, a single equation approach seems appropriate when using a nonhomothetic dual functioal specification.

[24] This may be done in two ways: obtain the X^* from the profit function using Hotelling's lemma and substitute them into the direct profit function and solve for Y, or use the integrability theorem to solve for the production function.

This model does not derive the analytical equation to determine the impact of various inefficiencies on level of profit and the impact of allocative inefficiency on cost, though some estimates of the cost of allocative inefficiency and loss of profit associated with various inefficiencies are possible.

Table III.1 Average Farm Characteristics

Category	Pooled Data A	Small Farm B	Large Farm C
Number of Observations	250	202	48
Land Holding	5.11	4.04	9.63
Value of Capital	36.73	25.53	83.89
Output	36.25	28.12	70.45
Fartilizer	93.60	63.65	185.45
Bullock Labour	42.56	33.72	79.76
Human Labour	187.67	159.01	305.79
Output Price	140.95	140.23	143.96
Fartilizer Price	5.38	5.43	5.16
Hiring Price of Bullock Labour	8.84	8.94	8.42
Hiring Price of Human Labour	9.97	10.20	9.00
Years of Education	8.52	8.04	10.54
Off-farm Income	50.24	49.72	52.49
Profit	25.27	18.67	53.02

land in acres;
capital and off-farm income in hundred rupees;
output in quintals;
fertilizer in kilograms;
labours in working day (8 hours per day).

Table III.2 Maximum Likelihood Parameter Estimates *

Parameter	Pooled Model A	Small Farm B	Large Farm C
α_0	.4745	.8169	-3.3487
	(.4696)	(.3703)	(1.3295)
α_P	1.7835	1.3685	2.0373
	(.1294)	(.3854)	(.5776)
β_H	-1.3491	-1.9617	-1.1008
	(.2751)	(.4518)	(.8849)
β_B	-2.7498	-1.1482	-1.2569
	(.5273)	(.2364)	(.9388)
γ_L	1.1014	.6278	6.7913
	(.3032)	(.3545)	(7.3594)
γ_K	.4179	.4115	.1057
	(.3635)	(.3945)	(.0322)
δ_{PP}	.5946	.1876	.7848
	(.1226)	(.1089)	(.2720)
θ_{HH}	.3491	-1.0534	-.5249
	(.1736)	(.3257)	(.2009)
θ_{BB}	1.6024	-.2496	-.2633
	(.3839)	(.1522)	(.6683)
η_{LL}	.0384	.0865	-3.6619
	(.0633)	(.1127)	(4.0287)
η_{KK}	-.1048	-.3729	.0398
	(.1347)	(.1521)	(.0100)
φ_{PH}	-.3843	-.2687	-.0118
	(.1331)	(.1002)	(.1509)
φ_{PB}	-.2983	-.3315	-.3797
	(.0944)	(.0653)	(.2404)

* Standard errors in parentheses

continued

Parameter	Pooled Model A	Small Farm B	Large Farm C
φ_{HB}	.2342	.0447	.1675
	(.1638)	(.1234)	(.3417)
ϕ_{PL}	.2568	.1546	.2827
	(.0552)	(.0704)	(.6558)
ϕ_{PK}	.1716	.1149	.0485
	(.0685)	(.0785)	(.0324)
ψ_{HL}	-.5467	-.3920	-.5020
	(.0808)	(.1226)	(.3911)
ψ_{HK}	-.1191	-.0224	-.0162
	(.0978)	(.1184)	(.0191)
ψ_{BL}	-.3590	-.2533	-.7249
	(.0385)	(.0738)	(.5139)
ψ_{BK}	.0581	.0109	-.0079
	(.0551)	(.0702)	(.0282)
η_{KL}	.0947	.1067	.0557
	(.0572)	(.1088)	(.1273)
σ_τ	.6482	.2869	1.3566
	(.0984)	(.0984)	(.4338)
σ_ν	1.1599	1.1268	2.2506
	(.0617)	(.0614)	(.2736)
σ_ξ	1.5779	1.0098	1.5033
	(.0740)	(.0539)	(.1433)
σ_{HH}	1.7672	1.1849	1.0297
	(2.3307)	(.1011)	(.3139)
σ_{BB}	2.6585	1.0659	.9264
	(18.5492)	(5.3931)	(.2825)
σ_{BH}	1.9231	.5131	.4458
	(1.5126)	(1.6857)	(.2846)

Table III.3 Impact of Inefficiency by Farm Size Class

Measures	Pooled Model A	Small Farm B	Large Farm
$\Pi L_\tau{}^*$	.2302	.0588	.1484
σ_τ	.2248	.4309	.1468
$t_\tau{}^{**}$		2.4108	
UL_H^1	-.0133	.3113	-.3987
σ_H	.7310	.6284	.4698
$t_H{}^{**}$		-10.7832	
UL_B^1	.0261	-.2858	.0845
σ_B	.4051	4.9205	.3072
$t_B{}^{**}$		.8445	
UL_F^1	.1084	.0735	.1549
σ_F	12.7447	6.3055	4.7091
$t_F{}^{**}$		1.7310	
ΩL_Y^2	.3457	.3453	.1234
σ_Y	.7592	.4645	.4225
$t_Y{}^{**}$		3.1557	

* percentage loss of profit due to combined inefficiency

** t statistics for testing $H_0 : Mean_B = Mean_C$.

1 percentage of over or under utilization of variable inputs

2 percentage of over or under supply of output

References

Aigner, D. J., C. A. K. Lovell, and P. Schmidt. "Formulation and Estimation of Stochastic Frontier Production Function Models." *J. Econometrics* 6(1977):21-37.

Ali, M., and J. C. Flinn. "Profit Efficiency Among Basmati Rice Producers in Pakistan Punjab." *Amer. J. Agr. Econ.* 71(1989):302-10.

Arrow, K. J., H. B. Chenery, B. S. Minhas, and R. M. Solow. " Capital-Labour Substitution and Economic Efficiency." *Rev. Econ. Statist.* 63(1961):225-47.

Bardhan, P. K. "Size, Productivity and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture." *J. Polit. Econ.* 81(1973):1370-386.

Bauer, P. W. *An Analysis of Multiproduct Technology and Efficiency Using the Joint Cost Function and Panel Data.* Ph. D. Dissertation, University of North Carolina at Chapel Hill, 1985.

Berndt, E. R. and L R. Christensen. "The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing 1929-1968." *J. Econometrics* 1(1973):81-114.

Cobb, C. W., and P. H. Douglas. "A Theory of Production." *Amer. Econ. Rev.*, 8(1928):139-65.

Dasgupta, B. "Monitoring and Evaluation of the Agrarian Reform Programme of West Bengal." Comprehensive Area Development Programme, Calcutta, December, 1987.

Diewert, W. E. "Application of Duality Theory." *Frontiers in Quantitative Economics.* vol 2, pp. 106-71, ed. M. D.Intrigellitor and D. Krendric. Amsterdam: North-Holland, 1974.

——. "Duality Approach to Microeconomic Theory." *Handbook of Mathematical Economics*, Vol. 2, pp. 535-91, ed. K. J. Arrow and M. D. Intriligator. Amsterdam: North-Holland, 1982.

Farrell, M. J. "The Measurement of Productive Efficiency." *J. Royal Statist. Soc.* Series A (general) 21(1957):253-81.

Forsund, F. R., C. A. K. Lovell, and P. Schmidt. "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurements." *J. Econometrics* 13 (1980):5-25.

Greene, W. H. "On the Estimation of a Flexible Frontier Production Model." *J. Econometrics* 13(1980):101-15.

——. *Econometric Analysis*. New York: Macmillan Publishing Company, 1989.

Hanoch, G. "Symmetric Duality and Polar Production Functions." *Production Economics: A Dual Approach to Theory and Application.* vol. 1, pp. 111-32, ed. M. Fuss and D. McFadden. Amsterdam: North-Holland, 1978.

Hicks, J. R. *Value and Capital*. 2nd ed. Oxford: Oxford University Press, 1946.

Hopper, W. D. "Allocative Efficiency in a Traditional Indian Agriculture." *J. Farm Econ.* 47(1965):611-24.

Hotelling, H. S. "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functios." *J. Polit. Econ.* 40(1932):577-616.

Jondrow, J. C., C. A. K. Lovell, I. S. Materov, and P. Schmidt. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *J. Econometrics* 23(1982):269-74.

Jorgenson, D. W., and L. Lau. "Duality and Differentiability in Production. *J. Econ. Theory* 9(1974a):23-42.

——. "The Duality of Technology and Economic Behavior." *Rev. of Econ. Studies* 41(1974b):181-200.

Just, R. E. and R. D. Pope. "Stochastic Specification of Production Functions and Economic Implications." *J. Econometrics* 7(1978):67-86.

Kumbhakar, S. C. "Estimation of Technical Efficiency Using Flexible Functional Form and Panel Data." *J Bus. and Econ. Statis.* 7(1989):253-58.

Kumbhakar, S. C., B. Biswas, and D. V. Bailey. "A Study of Economic Efficiency of Utah Dairy Farmers: A System Approach." *Rev. Econ. Statis.* 71(1989):595-604.

Lau, L. J. "Characteristics of the Normalized Restricted Profit Function." *J. Econ. Theory* 12(1976):131-63.

——. "Application of Profit Functions." *Production Economics: A Dual Approach to Theory and Applications* vol. 1, pp. 133-216, ed. M. Fuss and D. McFadden, Amsterdam: North-Holland, 1978.

Lau, L. J., and Pan A. Yotopolous, "A Test of Relative Economic Efficiency and Application to Indian Agriculture."*Amer. Econ. Rev.* 61(1971):94-109.

——. "Profit, Supply, and Factor Demand Functions." *Amer. J. Agr. Econ.* 54(1972):11-18.

McFadden, D. "Further Results on CES Production Functions." *Rev. of Econ. Studies* 30(1963):73-83.

——. "Cost, Revenue, and Profit Functions." *Production Economics: A Dual Approach to Theory and applications.* vol. 1, pp. 1-110, ed. M. Fuss and D. McFadden, Amsterdam: North-Holland, 1978.

Meeusen, W., and J. van den Broeck. "Efficiency Estimation From Cobb-Douglas Production Functions with Composed Error." *Int. Econ. Rev.* 18(1977):435-44.

Melfi. C. A. *Estimation and Decomposition of Productive Efficiency in A Panel Data Model: An Application to Electric Utilities.* Ph. D. Dissertation, University of North Carolina at Chapel Hill, 1984.

Rosenzweig, M. R. "Agricultural Development, Education and Innovation." *The Theory and Experience of Economic Development: Essays in Honour of Sir W. Arther Lewis.*

pp. 107-34, ed. M. R. Rosenzweig, C. F. Diaz-Alejandro, Gustav Ranis, and M. Gsovitz. London: George Allen & Unwin, 1982.

Rudin, W. *Principles of Mathematical Analysis.* 2nd ed. New York: McGraw-Hill, 1964.

Saini, G. R. "Farm Size, Productivity and Returns to Scale." *Econ. and Polit. Weekly.* 6(1971):79-81.

Samuelson, P. A. "Prices of Factors and Goods in a General Equilibrium." *Rev of Econ. Studies* 21(1953):1-20.

——. "Structure of a Minimum Equilibrium System." *Essays in Economics and Econometerics.* pp. 1-33, ed. R. W. Pfouts, Chapel Hill: University of North Carolina Press, 1960.

——. *Foundation of Economic Analysis.* 2nd ed. Cambridge: Harvard University Press, 1983.

Schmidt, P. "An Error Structure for Systems of Translog Cost and Share Equations." Economic Workshop Paper 8309, Michigan State University, 1984.

——. "Frontier Production Functions." *Econometric Rev.* 4(1985-86):289-328.

Schmidt, P., and C. A. K. Lovell. "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontier." *J. Econometrics* 9(1979):343-66.

——."Estimating Stochastic Production and Cost Frontiers When Technical and Allocative Efficiencies Are Correlated." *J. Econometrics* 13(1980):83-100.

Sen., A. K. "Peasants and Dualism With or Without Surplus Labor." *J. Polit. Econ.* 74(1966):425-50.

Shephard, R. W. *Cost and Production Functions.* Princeton: Princeton University Press, 1953.

——. *Theory of Cost and Production Functions.* Princeton: Princeton University Press, 1970.

Uzawa, H. "Production Function With Constant Elasticity of Substitution." *Rev. Econ. Studies* 29(1962):291-99.

——. "Duality Principles in the Theory of Cost and Production." *Int. Econ. Rev.* 5(1964):216-20.

Yotopolous, Pan A., and L. J. Lau, "A Test of Relative Economic Efficiency and Application to Indian Agriculture: Some Further Results." *Amer. Econ. Rev.* 63(1973):214-23.

CHAPTER IV
SPECIFICATION AND ESTIMATION OF PROFIT INEFFICIENCY OF MULTI-PRODUCT BANKING FIRMS USING A FLEXIBLE FUNCTIONAL FORM

In the literature, there is a growing acceptance that the money supply by financial firms should be viewed as a behavioral decision process which these firms make in attempting to perform financial intermediation in an optimal manner. Previous to this development, traditional portfolio theory methodology was most often used to analyze the behavior of financial firms. This theory explains the pattern of holding different assets by the firm, but firm size as well as the structure of liabilities are assumed to be exogenously determined, while the allocation of funds is viewed as a simple allocation among various assets. This approach neglects banking cost aspects and thus cannot be used to explain supply characteristics of outputs or services offered by the financial firm. The portfolio approach cannot be used to address concerns about efficiency in the production sense and firm behavior in the money supply mechanism. In this essay, a microeconomic theory of the firm characterization of the banking firm is developed in contrast to the portfolio approach and then evaluated for efficiency, with the previously defined components of inefficiency measurement, of a set of banks in the supply of monetary services.

This approach is not without precedent. Several attempts have been made to explain the behavior of financial firms using concepts of the theory of the firm. Construction of a neoclassical money supply function from the application of this theory appears to be fruitful both in explaining the decisions of firms and tracing policy implementations and their impacts on the economy. Monetary and other regulatory policies are, in principle, transmitted to the economy via financial firms that provide monetary services through intermediation between borrowers and lenders. In this context, microeconomic analysis of optimal behavior of banking firms appears essential in providing a clear understanding of why such institutions exist and how the money supply mechanism operates through these firms. A microeconomic foundation for monetary theory and macroeconomics is provided by such an approach, as well as a neoclassical money supply function rather than the conventional multiplier-type supply function. In addition, an assessment of inefficiency (or level of efficiency) in the supply of monetary services by banking

firms is made using the microeconomic approach coupled with a framework to measure elements of inefficiency.

This production approach views banking firms as producing demand and time deposits that results from loans using capital, labour, and materials as inputs. A controversy runs in the banking literature about what constitutes inputs and outputs and how to measure them. According to Sealey and Lindley, economic output is best measured by the dollar value of earning assets, that is, dollar amount of outstanding loans and total dollar amounts of securities. Deposits in this approach are regarded as inputs. Implementation of this view is described in the work of Murray and White, and Clark. Another view argues that banks provide deposit services and, therefore, the later should be regarded as output (Bell and Murphy, Benston). Using a profit function, Hancock (1985a, 1985b) showed that demand and time deposits are outputs produced by banking firms using labour, materials, capital, and most importantly cash reserves. Through its cash reserve liquidity management, banking firms generate deposits that produce monetary assets; thus, through intermediation, banking firms do generate output. This method was practiced by Gilligan, Smirlock, and Marshall, who specified a multiproduct banking technology defined by total deposits and total loans. This essay follows Hancock's methodology. The level of outstanding accounts provides the measure of output, and total costs include all operating costs incurred in the production of these accounts. Using these measures for outputs and inputs, the technology of the financial firm can be characterized and issues of efficiency addressed as a primary concern. Again, the concepts of inefficiency as previously defined are used in the variable profit function of the banking firm to assess banking inefficiency, if any, in the money supply mechanism. Of particular concern is the efficiency of the use of labour and capital in banking in the United States, an issue then will be addressed using a sample of banks from the Functional Cost Analysis Data of the Federal Reserve Bank system.

Previous Studies of Banking Firm Production

A rather extensive literature has developed from cost studies of banking firms. These studies were conducted primarily to provide regulatory agencies with an empirical basis to evaluate the influence of bank mergers on competition and the cost structure of the financial firm. Early studies were completed by Alhadeff, Horvitz, Schweiger and McGee, and Gramley, which related bank costs and out-

puts, though measures of outputs varied. In fact, a considerable debate on output produced by financial firms, started by these studies now continues in the literature.

Later, Greenbaum, Powers, and Schweitzer measured output as total bank revenue, a departure from the original debate over loans, investment, and deposits as output measures. Another stage of bank cost studies was ushered in by Benston, and Bell and Murphy, who used yet another measure of output the number of accounts.

Longbrake, Longbrake and Haslem, and Mullineaux followed with studies on the influence of branch banking on the cost structure of the banking firm. Their results suggested that the cost structure differs among branch banks, unit banks, and affiliates of holding companies. However, no concensus in output measurement developed by the addition of these studies.

All bank operations and technology were characterized in a later study by Benston, Hanweck, and Humphrey. They also used a flexible functional form representation for the bank technology, avoiding the restrictions imposed by the assumed technology of previous studies. The output measure used in this study was the number of accounts, represented by a multilateral divisia index of the aggregate number of accounts for demand and time deposits and different loan categories. Their results suggest that economies of scale are exhausted at a very low level of deposits ($25 million) and that the cost structure is different for unit banks relative to branch banks. Studies of economies of scope followed with work by Murray and White; Gilligan, Smirlock, and Marshall; Benston, Berger, Hanweck, and Humphrey; Clark; and Hunter and Timme. More flexible functional form specifications were used in these studies (Translog and Box-Cox generalized forms) to estimate economies of scope as well as scale economies. Murray and White investigated the British Columbia Credit Union in Canada and found rather significant economies of scale and scope in the subsets of outputs that they introduced. The other studies investigated U. S. banking (both unit and branch operations) using various definitions of output, to find that slight interproduct cost complementarities exist in unit banks and that economies are exhausted at a relatively low level of deposits. Tests for the existence of natural monopoly suggested little evidence for such market power.

Several major issues emerged from the studies mentioned above, the most predominant being the lack of general consensus regarding the appropriate definition

of output. The definition of bank output has long been a controversial subject, in particular because of its potential importance in the estimation of economies of scale and bank efficiency. Also, with the possible exception of the Hunter and Timme study, the influence of technical change and other forces that impact efficiency and scale economies were never addressed. Also, little was done to introduce non price-taking behavior in the modeling of the financial firm, an issue not addressed in this study.

With the monetary asset aggregation studies by Barnett(1980, 1981, 1986, 1987), the criticisms of earlier studies by Santomero, and the monetary aggregate testing work by Hancock (1985a, 1985b, 1987), the definition of bank output narrowed conceptually to monetary assets that are deposit account types of assets. That is, using the microeconomic framework, outputs of the banking firm are real balances of produced monetary assets that are supplied using capital, labour, and materials as inputs. Barnett (1980, 1981) first introduced the use of neoclassical aggregation and index number theory into monetary economics following the work on superlative indexes by Diewert (1976). In his 1986 study, Barnett fully developed the neoclassical multiproduct theory of financial firms and suggested the framework for characterizing the technology of these firms. Hancock at the same time estimated banking technology using similar specifications and tested the monetary aggregation theory and the appropriateness of alternative definitions of banking ouput in the multiple output model.

To this point, with the exception of a paper by Ferrier and Lovell, little has been done to develop measures of inefficiency in the banking industry with the approach proposed in these essays. The Ferrier-Lovell paper does develop a cost frontier and mathematical programming technique to calculate a production frontier but not the inefficiency measurement using the variable profit function system, as is later developed here.

Model

Banks are assumed to be competitive profit maximizers, an approved loan produces a vector of monetary assets, real balances, m_t, and yields a returns, w_t. Banking firm use real units of excess reserve (cash), say c_t, labour, q_t, along with some fixed factors, say z_t, to produce loans. Real balances of m_t and c_t are defined to equal nominal balances divided by the true cost of living index, p_t.

The technology of the bank is represented by a transformation function; F,

$$F(m_t, q_t, l_t, c_t; z_t) = 0\,, \tag{1}$$

which satisfies all required regularity conditions. This function, in sense, outlines how banks manage reserve and liquidity, which plays a central role in the money supply mechanism. The production set is defined by

$$S = (m_t, q_t, l_t, c_t; z_t) \geq 0\,. \tag{2}$$

If the set $(m_t, c_t) = 0$, then no financial intermediation takes place, no value added is created, and no loans are made. In this case, banks work as vaults, all of $\sum_i p_{ti} m_{ti}$ are reserves, and excess reserves are $p_t c_t = \sum_i m_i(1 - k_i)p_i$, where k_i is the reserve requirement, p is the true cost of living index, and t indexes time.

Banks can borrow from the discount window. The discount rate, d_t, may differ from the market rate, r_t, due to regulations. If $d_t < r_t$ banks will borrow excess reserves from the Federal Reserve and there are no free excess reserves. If, $d_t > r_t$, then there is no borrowing from the Federal Reserve, and free reserves equal excess reserves. The variable profit is defined as the difference between variable revenue and variable cost. The variable revenue at the end of period t is

$$\left\{ \sum_i (1 - k_{it}) m_{it} p_t - c_t p_t - q' Q_t \right\} r_t + c_t p_t (r_t - r_t^\circ)\,, \tag{3}$$

where $r_t^\circ = min(r_t, d_t)$, and q_t is the user cost price of variable factors, Q_t. Variable cost at time t is then defined as

$$\sum_i m_{it} p_{it} e_{it} - q_t' Q_t - w_t' l_t\,, \tag{4}$$

where w_t is the vector of wages, l_t is labour, and e_t is a vector of yields paid by the firm on produced monetary assets, m_t. $w_t' l_t$ is paid at the end of the period and is not substracted out of loan quantities produced at the beginning of the period. All Q_t are purchased at the beginning of the period for use during the period, and banks pay for them at the beginning of the period. Interest on produced monetary assets is paid at the end of the period, and interest received from loans is paid at the end of the period.

The variable profit, Π_T, at the end of the period t is obtained by substracting (4) from (3). Profit, Π_t, is divided by $(1 + r_t)$ in order to discount the variable profit to the beginning of period t. Then the present value of period t variable profit is

$$(5) \qquad \Pi(m_t, Q_t, l_t, c_t, p_t, q_t, r_t, e_t, w_t, k_t; z_t) = \frac{(m_t' g_t - q_t' Q_t - w_t' l_t)}{(1 + r_t)} - h_t c_t ,$$

where g_t is the nominal user-cost price of monetary assets defined as

$$(6) \qquad g = \frac{\{p(1-k)r - e\}}{(1+r)} .$$

h_t is the nominal user-cost price of reserve cash, defined as

$$(7) \qquad h = \{pr^{\circ}/(1+r)\} .$$

The real user-cost prices for m_t and c_t, therefore, are (g_t/p) and (h_t/p).

By applying Hotelling's lemma on the value dual profit function, netputs as $x_{it}(p, Z) = \frac{\partial \Pi_t(p,Z)}{\partial p_{it}}$ are derived. The derived demand for high-powered (base) money is then given in real terms as

$$m_t^h = c_t + \sum_i h_{it} m_{it} .$$

Banks' nominal demand for high-powered money is $p_t m_t^h$.

Profit in this analysis as represented by a normalized restricted translog profit function (NRTPF), can be expressed as

$$(8) \qquad \ln \Pi(P, Z) = \alpha_0 + \alpha' \ln P + \beta' \ln Z + \frac{1}{2} \ln P' \delta \ln P + \frac{1}{2} \ln Z' \theta \ln Z + \ln P' \eta \ln Z,$$

where $P \in R_{n-1}^{+}$ is the $(n-1) \times 1$ vector of prices of $(n-1)$[1] netputs, and $Z \in R_m^{+}$ is a $m \times 1$ vector of m quasi-fixed inputs. δ, θ, and η matrices are symmetric. A set of dual transformation relations that connects the production function and the profit function can be obtained by applying Hotelling's lemma,

[1] Price of the nth netput is used as normalizing factor. It does not matter which particular price is used for normalization.

viz., $\frac{\partial \ln \Pi(\cdot)}{\partial \ln P} = S^*$, where $S^* \in R^+_{n-1}$, is a $(n-1) \times 1$ vector of profit shares,[2] where * represents the solution level of profit shares. The elements of the S^* vector represents shares of each netput in total profit. Differentiating equation (8) with respect to $\ln P$ we get following system of equations:

$$\frac{\partial \ln \Pi(\cdot)}{\ln P} \equiv S^* = \alpha + \delta \ln P + \eta \ln Z . \tag{9}$$

The above system involves $(n-1)$ profit share equations. Equations (8) and (9) involve n simultaneous equations. Since shares add up to unity, one share equation is dropped to avoid singularity in estimation. The resulting system of $(n-1)$ equations can be solved for unknown parameters by any standard simultaneous equations estimation technique after appending disturbance terms to each equation (e.g., full information maximum likelihood [FIML], seemingly unrelated regression [SUR], or three stage least square [3SLS]).

Because interest lies in estimating firm and netput specific inefficiencies, a stochastic specification of equation (8) is used following Aigner, Lovell, and Schmidt; and Meeusen and van den Broeck. The stochastic specification of equation (8) can be expressed as

$$\begin{aligned} \ln \Pi^s(P,Z) = \alpha_0 + \alpha' \ln P + \beta' \ln Z + \frac{1}{2} \ln P' \delta \ln P \\ + \frac{1}{2} \ln Z' \theta \ln Z + \ln P' \eta \ln Z + \epsilon, \end{aligned} \tag{8a}$$

where ϵ is a random disturbance term composed of two different error terms. The economic significance and statistical properties of these two terms are completely different. ϵ could be formulated as the sum of a symmetric component representing the effects of factors outside the control of the firm (statistical white noise) and a one-sided component representing profit inefficiency known to firm. The stochastic profit function, therefore, can be expressed as

$$\begin{aligned} \ln \Pi(P,Z) = \alpha_0 + \alpha' \ln P + \beta' \ln Z + \frac{1}{2} \ln P' \delta \ln P \\ + \frac{1}{2} \ln Z' \theta \ln Z + \ln P' \eta \ln Z + \omega + \nu, \end{aligned} \tag{8b}$$

where $\epsilon = \omega + \nu$. The classical disturbance term, as represented by ν and ω, signifies the one-sided nonpositive disturbance term ($\omega \leq 0$). If $\omega < 0$, then

[2] This definition of share requires assumptions that factors are paid their marginal products, and production is characterized by constant returns to scale.

the production plan of the firm is profit inefficient (i.e., the firm is earning less than the maximum level of profit that it could earn if all inputs and outputs are optimally allocated). The inequality, $\omega < 0$, therefore, indicates that the firm is operating below the profit frontier. On the other hand, if $\omega = 0$, the firm is operating on the profit frontier. The parameter, ω, measures how close an individual producer's production plan (profit level) is to the maximum level of profit frontier, given prices and fixed factors. Thus, $\omega < 0$ can be interpreted as the percentage reduction in potential profit due to inefficiency. Profit inefficiency is, therefore, defined as any downward deviation from the optimum point on the frontier.

Equation (1b) is estimated by maximum likelihood methods, given prespecified assumptions regarding the distribution of disturbance terms involved. The resulting estimates are efficient and consistent. The profit inefficiency of the individual firm can be estimated following the method developed by Jondrow, Lovell, Materov, and Schmidt. The full system of equations, i.e., the stochastic profit function and the set of share equations obtained using the envelope theorem, can also be estimated. Estimating the full system of equations simultaneously, instead of a single equation, not only improves the precision of the parameter estimates but also yields information regarding the sources of profit inefficiency of the individual firm. However, consistency of the estimated parameters depends on the specification of the full system. The possible efficiency gain occurs not only due to cross-equation error correlation but also due to cross-equation restrictions imposed by model specification. Consistency depends to a great extent on the correctness of the distributional assumptions and on the specification of the deterministic part of the system of equations (Schmidt, 1985-86).

After appending the disturbance terms, the set of share equations can be expressed as

$$S^* = \alpha + \delta \ln P + \eta \ln Z + U, \tag{9b}$$

where U is a $(n-1) \times 1$ vector, elements of which are fractions and can take any sign. The term, U, represents the deviation of the observed profit share of netputs from their respective theoretical optimum levels. If firms fail to allocate their netputs optimally, then the share of that netput will deviate from its theoretical optimum level; therefore, it is appropriate to consider U as the effect of allocative inefficiency (Greene). The important point to note is that the netputs

include both inputs and outputs. The first-order derivative of the profit function, with respect to input prices, yields profit shares of inputs in total profit. On the other hand, the first-order derivative of the profit function, with respect to output prices, gives the profit share of the specific input in total profit (given the underlying assumption of perfect competition and constant returns to scale). Elements of U obtained from the first case represent the misallocation of outputs from the theoretical optimum, and this misallocation reduces the level of profit. The elements of U in the second case represent misallocation of inputs from their theoretical optimum and, therefore, reduce actual profit from its maximal level. This definition of allocative efficiency is somewhat different from that provided by Schmidt and Lovell (1979, 1980); and Forsund, Lovell, and Schmidt. However, their estimates of allocative inefficiencies can be derived from this definition.[3]

As mentioned in the previous essay, the exact statistical relationship between U and the profit inefficiency term ω in equation (8b) cannot be established because they are assumed to be statistically independent. This is no doubt a very simplifying assumption. Schmidt and Lovell (1980), who dispensed with this assumption, found almost no effect on their estimate of the frontier. Schmidt and Lovells' work involved a homothetic production relation, where a statistical relationship between U and ω could be established. In the case of a nonhomothetic production specification, this relationship is impossible to derive (Greene).

This essay extends the analysis of Ali and Flinn and that of last essay by formulating a simultaneous equation model to estimate profit inefficiency in a multiple output-multiple input setting. To estimate profit inefficiency and allocative inefficiency of associated netputs, the system of equations defined by (8b) and (9b) must be estimated. The system contains n independent equations ($[n-1]$ share equations and the profit function) to solve for n endogenous variables ($[n-1]$ shares, S, and Π^s). The parameters involved in the above set of equations can be estimated and used to get the estimates of the inefficiency parameters, i.e., ω and U.

Estimation Procedure

In principle, parameters in the equation (8b) can be estimated using ordinary least squares (OLS) or instrumental variables (IV) techniques by using ϵ instead

[3] See the essay 2 for detailed discussion of this theoretical issue.

of $\omega + \nu$. The resulting estimates of the parameters are consistent and efficient since the regressors in the case of a normalized restricted profit function (NRPF) are independent of ϵ. This process does not estimate the profit inefficiency independent of statistical white noise. To capture specification of the profit function as shown in the previous section, we need to apply the maximum likelihood estimation (MLE) method. A single equation estimation method can estimate all relevant parameters to compute all relevant inefficiency parameters, ω and U. However, as the number of inputs and outputs increases, the number of parameters becomes quite large. In a multiple input and multiple output case, the question of multicollinearity is likely to become severe. Thus, the single equation estimation is likely to be imprecise.

Simultaneous equation estimation methods like full information maximum likelihood (FIML) and seemingly unrelated regression (SUR), which yield maximum likelihood eastimates when the iteration process converges, can solve the above system of equations, but the basic problem of separating ω from ν remains. The maximum likelihood methods may be used to estimate the system of equations (8b) and (9b) and to separate ω from ν, with assumption of a specific distibution for ω. In the case of panel data, one can avoid specific distributional assumption for the one-sided error term, ω. Of course, a structure that specifies variation of efficiency over time must be imposed, though it is very difficult to justify the imposition of any arbitrary time structure (dynamic structure) on technology. In the literature, use of specific distributional assumption for this purpose is commonplace, and Monte Carlo studies have been done to justify their use. Here, it is assumed that profit inefficiency of each firm is time variant.

The system of equations (8b) and (9b) with firm, (f), and time, (t), subscripts can be expressed as

$$\begin{aligned}
\ln \Pi^{s}_{ft} &= \ln \alpha_0 + \sum_i \alpha_i \ln P_{ift} + \sum_j \beta_j \ln Z_{jft} + \frac{1}{2} \sum_i \sum_r \delta_{ir} \ln P_{ift} \ln P_{rft} \\
&\quad + \frac{1}{2} \sum_j \sum_q \theta_{jq} \ln Z_{jft} \ln Z_{qft} + \sum_i \sum_j \eta_{ij} \ln P_{ift} \ln Z_{jft} + \omega_{ft} + \nu_{ft}, \\
&= \ln \alpha_0 + \alpha' \ln P_{ft} + \beta' \ln Z_{ft} + \frac{1}{2} \ln P'_{ft} \delta \ln P_{ft} + \frac{1}{2} \ln Z'_{ft} \theta \ln Z_{ft} \\
&\quad + \ln P'_{ft} \eta \ln Z'_{ft} + \omega_{ft} + \eta_{ft}. \qquad (8c)
\end{aligned}$$

$$S_{ift} = \alpha_i + \sum_r \delta_{ir} \ln P_{rft} + \sum_j \eta_{ij} \ln Z_{jft} + u_{ift}$$

$$= \alpha + \delta' \ln P_{ft} + \eta' \ln Z_{ft} + U_{ft}. \tag{9c}$$

$$i = 1, \ldots, (n-1); \quad j = 1, \ldots, m; \quad f = 1, \ldots F; \text{ and } t = 1, \ldots T.$$

To estimate the above system of equations, the joint probability density function (pdf) of the error vectors ω_{ft}, ν_{ft}, and U_{ft} must be derived with specified distributional forms. Since the sum of shares adds to unity, $I'U = 0$, where I is an unit vector of dimension conformable to that of U, and to avoid singularity in estimation, one share equation is dropped (Berndt and Christensen). The error vector from the system of equations (8c) and (9c), therefore, is

$$\begin{pmatrix} \omega_{ft} + \nu_{ft} \\ u_{1ft} \\ \vdots \\ u_{(n-2)ft} \end{pmatrix}. \tag{10}$$

From (10), the pdf of ν_{ft}, ω_{ft}, and $u_{1ft}, \ldots, u_{(n-2)ft}$ are formed. The usual assumptions for distribution of the aforementioned disturbance terms in the literature (Battese and Coelli, Kumbhakar (1987, 1988), Pitt and Lee, Schmidt (1988), and Schmidt and Sickles) are:

1. ω_{ft} are time variant and distributed as $iid\ N(0, \sigma_\omega^2)$ truncated as zero from above,
2. $\tilde{u}_{ft} \sim iid\, N(0, \Sigma)$ for all f and t, where $\tilde{u}_{ft} = (u_{1ft}, \ldots, u_{(n-2)ft})'$,
3. $\nu_{ft} \sim iid\, N(0, \sigma_\nu^2)$ for all f and t, and
4. ω_{ft}, $\tilde{u}_{ft}$, and ν_{ft} are independent of each other for all f and t, and also independent of P_{ft} and Z_{ft},

where f indexes firm ($f = 1, \ldots, F$) and t indexes time ($t = 1, \ldots, T$).

Letting $b_1 = (\omega_{ft} + \nu_{ft})$ and $b_2 = \tilde{u}_{ft}$, the joint pdf of b_1 and b_2 is

$$f(b_1, b_2, \omega_{ft}) = f(\omega_{ft}, \nu_{ft}, \tilde{u}_{ft})\,|J|, \tag{11}$$

where J is the Jacobian of transformation. The Jacobian for the variable transformation in this case is

$$J = \frac{\partial(\nu_{ft}, \omega_{ft}, \tilde{u}_{ft})}{\partial(b_1, b_2, \omega_{ft})}.$$

In this case $|J| = 1$ and by assumption (4), the equation (11) can be expressed as

$$f(b_1, b_2, \omega_{ft}) = f(\omega_{ft})\, f(\nu_{ft})\, f(\tilde{u}_{ft}), \tag{12}$$

where $f(\omega_{ft})$, $f(\nu_{ft})$, and $f(\tilde{u}_{ft})$ represent the pdf of ω_{ft}, ν_{ft}, and $\tilde{u}_{ft}$ respectively. After some manipulations, the joint pdf can be expressed as

$$f(b_1, b_2, \omega_{ft}) = \frac{2\exp(-d/2)}{(2\pi)^{(n/2)}\,\sigma_\nu\,\sigma_\omega\,|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left(\frac{\omega_{ft}-\mu_\omega}{\sigma}\right)^2\right\} \exp\left\{-\frac{1}{2}(b_2'\,\Sigma^{-1}\,b_2)\right\}, \tag{13}$$

where

$$d = \left\{\frac{b_1}{\sigma_\nu^2+\sigma_\omega^2}\right\};\ \mu_\omega = \left\{\frac{\sigma^2 b_1}{\sigma_\omega^2}\right\} = \left\{\frac{\sigma_\omega^2 b_1}{\sigma_\omega^2+\sigma_\nu^2}\right\};\ \text{and } \sigma^2 = \left\{\frac{\sigma_\omega^2\,\sigma_\nu^2}{\sigma_\omega^2+\sigma_\nu^2}\right\}.$$

To get the joint pdf of b_1 and b_2, the ω_{ft} must be integrated out from equation (13). After integrating out ω_{ft} and some algebraic manipulation, the joint pdf of b_1 and b_2 can be expressed as

$$f(b_1, b_2) = \frac{2\exp(-d/2)\,\sigma\Phi(-\mu_\omega/\sigma)}{(2\pi)^{\frac{(n-1)}{2}}\,\sigma_\omega\,\sigma_\nu\,|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(b_2'\,\Sigma^{-1}\,b_2)\right\}. \tag{14}$$

From the joint pdf of the error vectors (12), the endogeneous variable vector, $(\Pi^s, S)'$ can be derived. By variable transformation,

$$f(\Pi^s, S) = f(b_1, b_2)|J|. \tag{15}$$

The log likelihood function for the F firms observed over T years is

$$L = \sum_{t=1}^{T}\sum_{f=1}^{F} \ln f(\Pi^s, S) + TF \ln |J|, \tag{16}$$

where J is the Jacobian of transformation. Since, in this case $J = 1$, the log likelihood function can be expressed as

$$L = TF\ln 2 - \frac{TFd}{2} + TF\ln\sigma + \sum_t\sum_f \ln\Phi\left(-\frac{\mu_\omega}{\sigma}\right) - \frac{TF(n-1)}{2}\ln 2\pi \tag{17}$$

$$- TF \ln \sigma_\nu - TF \ln \sigma_\omega - \frac{TF}{2} \ln |\Sigma| - \frac{1}{2} \sum_t \sum_f (b_2' \Sigma^{-1} b_2).$$

To simplify the likelihood function for estimation, it is convenient to rewrite the frontier profit function and share equations for the fth firm at tth time as

$$\Pi^*_{ft} = \Delta \tilde{X}_{ft} + \omega_{ft} + \nu_{ft}; \tag{8d}$$

$$\tilde{S}_{ft} = \Lambda \tilde{X}_{ft} + U_{ft}, \tag{9d}$$

where $\tilde{X}$ is the full vector of K regressors ($[n-1]$ prices and m fixed factors). Δ is a $(1 \times K)$ vector of all parameters. In the share equation, $\tilde{S}_{ft}$ is a $((n-2) \times 1)$ vector of observations on the factor shares. Λ is the $((n-2) \times K)$ parameter matrix that contains all parameters of Δ. The Λ matrix contains a substantial number of zeros since none of the quadratic terms of (8d) appear in the share equations, and every non-zero element in Λ is equal to some element in Δ.

The log likelihood function (17) could be estimated to obtain parameters for the efficiency estimation, but the burden of estimation could be reduced considerably by concentrating the likelihood function with respect to Σ and the constant terms. $((n-2) \times (n-2))$ sample estimates of Σ can be obtained from

$$G_{ft} = \frac{1}{2} \sum_{t=1}^{T} \sum_{f=1}^{F} \left(S_{ft} - \Lambda X_{ft}\right) \left(S_{ft} - \Lambda X_{ft}\right)'. \tag{18}$$

Dropping firm and time indexes, the log likelihood function for one observation without irrelevant constants is expressed as

$$L = -\frac{d}{2} + \ln \sigma + \ln \Phi\left(-\frac{\mu_\omega}{\sigma}\right) - \ln \sigma_\omega - \ln \sigma_\nu - \frac{1}{2}\left\{\ln |\Sigma| + tr(\Sigma^{-1} G)\right\}. \tag{19}$$

At the maximum of L,

$$\frac{\partial L}{\partial \Sigma} = \Sigma^{-1}(\Sigma - G)\Sigma^{-1} = 0.$$

This implies that given Λ, G is the MLE of Σ. Substituting this result in the likelihood function, concentrated log likelihood function can be obtained. The log likelihood function for the fth firm observed at the tth year is expressed (again

dropping time and firm indexes) as

$$L_C = -\frac{d}{2} + \ln\sigma + \ln\Phi\Big(-\frac{\mu_\omega}{\sigma}\Big) - \ln\sigma_\nu - \ln\sigma_\omega - \frac{1}{2}\ln|G|, \tag{20}$$

which can be maximized with respect to the parameters involved. This log likelihood equation gives a set of non-linear optimizing equations when differentiated with respect to each unknown parameter. These sets of equations are solved simultaneously to obtain estimates of the unknown parameters.

Estimation of Efficiency Parameters

Given the MLE of (20), ω_{ft} and $\breve{u}_{ft}$ can be estimated. Profit inefficiency is defined as the combination of all three types of inefficiency (technical, allocative, and scale).

Profit Inefficiency

To derive ω_{ft}, ν_{ft} is isolated from ω_{ft} by estimating the conditional pdf of ω_{ft} given b_1. $f(\omega_{ft}|b_1)$ is a $N(\mu_{\omega_{ft}}, \sigma^2)$ variable truncated at zero. Following Jondrow, Lovell, Materov, and Schmidt , the profit inefficiency estimated at a point estimator of ω_{ft}. The point estimate of ω_{ft} can be obtained either by the mean or the mode of ω_{ft}. Thus either

$$E(\omega_{ft}|b_1) = \mu_{\omega_{ft}} - \sigma\Big\{\frac{\phi(\mu_{\omega_{ft}}/\sigma)}{\Phi(-\mu_{\omega_{ft}}/\sigma)}\Big\} \tag{21}$$

or

$$M(\omega_{ft}|b_1) = \begin{cases} \mu_{\omega_{ft}}, & \text{if } b_1 \le 0; \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

can be used as a point estimator of ω_{ft}.

Allocative Inefficiency

The level of allocative inefficiency for the firm f at time t can be expressed as

$$\hat{U}_{ft} = S^* - (\hat{\alpha} + \hat{\delta}\ln P_{ft} + \hat{\eta}\ln Z_{ft}), \tag{23}$$

where $\hat{U}$ is a $((n-1)\times 1)$ vector, elements of which can be obtained by using the estimated values of the parameters $\hat{\alpha}$, $\hat{\delta}$, and $\hat{\eta}$. Estimated values of $\hat{u}_{ift}$ can take any sign.

Impact of Inefficiency

The impact of various inefficiencies on the level of profit, i.e., the potential profit loss due to inefficiency, is of prime importance from the policy point of view. Profit inefficiency, ω_{ft}, gives the percentage of potential profit loss due to combined inefficiency. Given Z and prices, the percentage loss of potential profit due to inefficiency of the fth firm becomes

$$\Pi L_{\omega_{ft}} = \frac{\Pi_{ft} - \Pi^*_{ft}}{\Pi^*_{ft}} = \frac{\hat{\omega}_{ft}}{\Pi^*_{ft}}, \tag{24}$$

where $\hat{\omega}_{ft}$ can be obtained from (21). This loss of profit is due to misallocation of inputs, outputs, and technical inefficiency. The degree of misallocation of the ith netput of the fth firm can be obtained from

$$UL_{ift} = \frac{S_{ift} - S^*_{ift}}{S^*_{ift}} = \frac{\hat{u}_{ift}}{S^*_{ift}}, \tag{25}$$

where $\hat{u}_{ift}$ and S^*_{ift} can be obtained from equations (23) and (9c) directly. Elements of UL obtained by using the output share equations represent the misallocation of outputs, while UL related to input share equations represents misallocation of inputs.

Data Description

The primary source of the data used in this study is the Federal Reserve's Functional Cost Analysis (FCA) program. Through this voluntary program, participating member banks provide the Federal Reserve with detailed information on balance sheet and income statement composition. Information on operating expenses and income is broken down and allocated by various bank functions such as loans, deposits, safe deposit boxes and trust functions.

The sample data are for a five-year (1979-83) time series of 41 banks, drawn from member banks of Federal Reserve District 7. The sample, therefore, concentrates on this individual district, and the banks are primarily unit banking firms operating within that district. One bank has deposits at a level less than $25 million in 1983. Four banks have deposits ranging from $25 million to $50 million. Twenty-eight have total deposits ranging from $50 million to $200 million during the period covered. Eight of the banks have total deposits exceeding $200 million.

The median of total deposits is \$122 million. The minimum deposit level is \$21.1 million, and the maximum level is \$241.7 million.

In this study, financial firms are considered to produce monetary assets (deposit account type) that play the essential role in providing payment or transaction services to the economy of District 7. Demand and time deposits are defined as outputs following the specification of Hancock (1985a, 1987). Inputs are identified as cash balances (excess reserves), labour, materials, and capital.

The quantities of the two outputs are real balances of demand and time deposits from the balance sheet; that is, their quantities are dollar values from the balance sheet deflated by the price index. It is assumed that there exists exact economic quantity aggregates for demand and time deposits, which can be treated like elementary goods. The cash balance quantity measure is real excess reserves above required reserves. Real balances of monetary assets from the balance sheet are stock variables. Alternatively, monetary assets are commonly considered durable goods that yield continuous monetary service flows. To convert the stocks to flows, user cost prices for the monetary assets (which are output) (6) are used. Similarly, the real excess reserves (cash) are converted using the user-cost price of excess reserves (7). The corresponding real user-costs are (g/p) and (h/p).

With k symbolizing the required reserve ratio on a specific deposit, say the ith, then $(1 - k_i)$ is available for each dollar deposited and this amount can be invested in loans and other investments. Thus, $r_i k_i$ in the user-cost formula for g is considered the implicit tax paid by the banking firm due to required reserve. The rate, r_i, for the ith loan investment is operationally defined as the aggregate loan rate. The own yield paid on the ith monetary good, i, by the banking firm, w_i, includes interest paid, service charges earned, and the Federal Deposit Insurance Corporation (FDIC) premium; therefore, w_i equal interests paid on m_i, minus the service charge rate earned, plus the FDIC insurence premimum rate paid. The interest rate paid on demand deposits includes only those for interest bearing deposits. Therefore, the interest rate on the demand deposit monetary asset is the explicit rate.[4]

The paid interest rate is the annual gross interest paid on the ith monetary asset divided by the total amount of the ith deposit. The service charge is service chage income divided by the total amount of the ith deposit, and the FDIC

[4] Klein derived the implicit rate approach to yields on demand deposits.

premium rate paid is the total FDIC premium paid divided by total deposits. In the FCA data, the paid FDIC insurance premium is reported for total deposits, not allocated to each deposit category.

Required reserve ratios are derived using member bank reserve requirements reported in the *Federal Reserve Bulletin*. Reserve requirements for the 1979–1980 period are calculated according to the schedule of requirements before implementation of the Monetary Control Act. The reserves for the 1981–1983 period are computed according to the schedule, Depository Institution Requirements after Implementation of the Monetary Control Act. Information on the maturity of time deposits is not available from the FCA data. Therefore, an average of 3% is used to compute reserve requirements following procedures in the *Federal Reserve Bulletin* for the five-year period.

The model uses two outputs, demand and time deposits from the monetary grouping of quantities, and then assumes excess reserves as an aggregate is an input like labour, materials, and capital. The quantities of the two outputs are real balances of demand deposits and time deposits on the financial firm balance sheet deflated by the price index. It is, therefore, assumed that exact economic quantity aggregates for demand and time deposits exist. It is assumed that banking firm manages liquidity and in so doing increases or decreases assets that are a part of the money supply mechanism. Implicitly assumed in these definitions of output is that loans become deposits within the same bank, a simplifying abstraction in the definitions given here. Loans generated at one bank may not directly generate deposits at the same bank. Inefficiencies in generating output may simply be due to this kind of abstraction and not entirely representative of the liquidity or liability management choice. The nominal balance of excess reserves is derived as

$$pc = C - \sum_{i}(k_i u_i), \qquad i = D, T.$$

Here, C is total cash and dues on the balance sheet of FCA, and u is the ith produced monetary asset, with k being the required reserve ratio for u.

An aggregate loan rate, r, is used to develop the indexes; so some assumptions and derivations are needed to make that aggregation of the various types of loans. Five types of fund-using functions of loan and investments are reported in the FCA data: investments, real estate mortgage loans, installment loans, credit card loans, and commercial and industrial loans. In this sense, r should be the marginal cost

of borrowing an additional dollar for one period, which is the marginal return on an additional dollar of bank loans, but this is extremely difficult to measure. Thus, the dual price index of the aggregate loan is derived on the gross annual average rates of returns earned on loans and used as a proxy. In this study, the aggregate loan quantity does not enter the profit function directly as described by Hancock (1985a). However, this quantity index is constructed to obtain the corresponding price index of the loan quantity aggregate (r) used to construct the user-cost prices of monetary goods in the study.

In the construction of the aggregate loan quantity, the Divisia multilateral index (or Tornqvist-Theil-Translog index as it is sometimes called) previously developed by Diewert (1976) is used. This type of index is within the superlative class of index numbers as defined by Diewert (1976). It is a transitive index while a chained index (or bilateral index) for time series is not. This transitive index is suitable for time series and cross-section observation such as contained in the FCA data.

Data on the quantities and prices of non-monetary goods such as labour, materials, and capital inputs are not directly available from the FCA income statement. These quantities are constructed using various available information from the FCA income statements, the *Survey of Current Business*, and the *Federal Reserve Bulletin*. Quantities are derived by using the Divisia index concept, as are prices of input quantities (non-monetary).

In the FCA data, all operating expenses for labour, materials, and capital inputs are classified and allocated by the various bank functions such as loans and deposits. Labour is classified as two groups of employment, viz., managerial (or officers) and other employees. These two forms of employment were aggregated using the Divisia index approach for each bank and for each year. The corresponding price index was computed using the Fisher factor reversal test procedure. This index is based on the base year 1979. For materials, all expenditures are reported in the FCA files, and again a Divisia index is developed to provide an aggregate quantity. Producer prices of the items used in materials are given in *Producer Prices and Price Indexes* and *Survey of Current Business*. The price index is developed by aggregating producer price indexes for the items in the materials category. The material quantity is then derived from Fisher's factor reversal test, i.e., the material quantity is the ratio of total material expenditures to the Divisia price index, with 1979 again being used as the base year.

Real physical capital expenditures have been approximated by total expenses allocated for furniture, equipment, and bank premises. In developing a new variable capital item and a new fixed capital item, expenditures on computers and other equipment have been incorporated as reported in the FCA files. These latter data are rather tenuous for the 1979–1981 period. Physical fixed capital is approximated by book value. The prices of capital have accordingly been approximated by dividing the physical capital expenditures by the book value less depreciation of the respective items in each category. The Fisher ideal price index is used for the true cost-of-living index, defined as the geometric mean of Laspeyres and Paasche indexes. The Consumer Price Index of the Bureau of Labor Statistics is used as the Laspeyres index, while the Implicit Price Deflator of the Commerce Department is used as the Paasche index.

In summary, the developed output aggregates are demand deposits and time deposits. The inputs aggregates include labour, cash (excess reserve as a monetary aggregate), variable capital (materials, some computers and other equipment expenses, and depreciation from the original physical capital aggregate), and capital (bank premises and furniture). The variable inputs are labour, cash, and the variable capital in the form of bank furniture, assumed to be the fixed inputs.

Empirical Application

This multiple input-multiple output model is applied to a data set of U. S. banks to determine the profit inefficiency of these institutions. Six netputs are empirically defined and classified into three groups: variable inputs, variable outputs, and a quasi-fixed input. Variable inputs include cash (C), variable capital (K), and labour (L), with corresponding prices are P_C, P_K, and P_L. Two outputs considered in this model are demand deposits (D) and time deposits (T), with prices outputs are P_D and P_T. The fixed input considered in this model is capital (Z).

The normalized restricted stochastic profit function (NRSPF) in terms of these netputs for the fth firm at time t can be expressed as (dropping firm and time indexes and adding the classical disturbance term and the one-sided error term)

$$\ln \tilde{\Pi}^s = \alpha_0 + \alpha_D \ln \tilde{P}_D + \alpha_T \ln \tilde{P}_T + \alpha_L \ln \tilde{P}_L + \alpha_K \ln \tilde{P}_K + \beta_Z \ln Z$$

$$(26) \qquad + \frac{1}{2}\Big\{\delta_{DD}(\ln \tilde{P}_D)^2 + \delta_{TT}(\ln \tilde{P}_T)^2 + \delta_{LL}(\ln \tilde{P}_L)^2 + \delta_{KK}(\ln \tilde{P}_K)^2\Big\}$$

$$+\frac{1}{2}\theta_{ZZ}(\ln \tilde{Z})^2 + \delta_{DT}\ln \tilde{P}_D \ln \tilde{P}_T + \delta_{DL}\ln \tilde{P}_D \ln \tilde{P}_L + \delta_{DK}\ln \tilde{P}_D \ln \tilde{P}_K$$

$$+\eta_{DZ}\ln \tilde{P}_D \ln Z + \delta_{TL}\ln \tilde{P}_T \ln \tilde{P}_L + \delta_{TK}\ln \tilde{P}_T \ln \tilde{P}_K + \eta_{TZ}\ln \tilde{P}_T \ln Z$$

$$+\delta_{LK}\ln \tilde{P}_L \ln \tilde{P}_K + \eta_{LZ}\ln \tilde{P}_L \ln Z + \eta_{KZ}\ln \tilde{P}_K \ln Z + \omega + \nu,$$

where the following definitions regarding normalization are used:

$$\tilde{P}_D \equiv \left(\frac{P_D}{P_C}\right); \tilde{P}_T \equiv \left(\frac{P_T}{P_C}\right); \tilde{P}_L \equiv \left(\frac{P_L}{P_C}\right);$$

$$\tilde{P}_K \equiv \left(\frac{P_K}{P_C}\right); \tilde{\Pi}^s \equiv \left(\frac{\Pi^s}{P_C}\right) \text{ and } \tilde{P} = \left(\tilde{P}_D\ \tilde{P}_T\ \tilde{P}_L\ \tilde{P}_K\right)'.$$

The constant returns to scale and symmetry restrictions are imposed directly. From the above NRSPF, the following profit share equations (with appended error terms) can be obtained by Hotelling's lemma:

$$S_D = \alpha_D + \delta_{DD}\ln \tilde{P}_D + \delta_{DT}\ln \tilde{P}_T + \delta_{DL}\ln \tilde{P}_L + \delta_{DK}\ln \tilde{P}_K + \delta_{DZ}\ln Z + u_D, \tag{26a}$$

$$S_T = \alpha_T + \delta_{TT}\ln \tilde{P}_T + \delta_{DT}\ln \tilde{P}_D + \delta_{TL}\ln \tilde{P}_L + \delta_{TK}\ln \tilde{P}_K + \eta_{TZ}\ln Z + u_T, \tag{26b}$$

$$S_L = \alpha_L + \delta_{LL}\ln \tilde{P}_L + \delta_{DL}\ln \tilde{P}_L + \delta_{TL}\ln P_T + \delta_{LK}\ln \tilde{P}_K + \eta_{LZ}\ln Z + u_L, \tag{26c}$$

and

$$S_K = \alpha_K + \delta_{KK}\ln \tilde{P}_K + \delta_{DK}\ln \tilde{P}_D + \delta_{TK}\ln \tilde{P}_T + \delta_{KL}\tilde{P}_L + \eta_{KZ}\ln Z + u_K. \tag{26d}$$

Equations (26) and (26a–26d) constitute the full system of equations. For estimation, one equation is dropped from the above set of share equations to avoid singularity (Barten, Berndt and Christensen, and Greene) since symmetry and constant returns to scale are imposed. In this system, the profit function cannot be dropped since an estimate of ω_f is needed. The system is a four-equation multivariate system.

The set of equations (excluding the equation dropped [26d]) for the fth firm at time t in matrix notation is given as

$$\Pi_{ft} = \Delta' \tilde{X}_{ft} + \omega + \nu, \tag{27}$$

$$\tilde{S}_{ft} = \Lambda \tilde{X}_{ft} + U_{ft}, \tag{28}$$

where $\tilde{X}_{ft}$ is the full vector of 21 regressors (21×1), and Δ is the full vector of 21 unknown parameters (21×1). $\tilde{S}_{ft}$ is a (3×1) vector of observed shares, while Λ is the (3×21) matrix of unknown parameters that contains a substantial number of zeros, because none of the quadratic terms in (27) appear in the share equations. There will also be a number of equality restrictions on Λ due to the symmetry restriction of the second-order terms; but every non-zero element in Λ is equal to some element in Δ. Another element of log likelihood function that must be specified is G, defined as

$$G = \frac{1}{TF} \sum_{t=1}^{T} \sum_{f=1}^{F} \left(\tilde{S}_{ft} - \Lambda \tilde{X}_{ft}\right)\left(\tilde{S}_{ft} - \Lambda \tilde{X}_{ft}\right)', \tag{29}$$

where G is a (3×3) sample estimate of Σ based on random sample, U_{ft}. Therefore,

$$|G| = \begin{vmatrix} \sigma_{DD} & \sigma_{DT} & \sigma_{DL} \\ \sigma_{TD} & \sigma_{TT} & \sigma_{TL} \\ \sigma_{LD} & \sigma_{LT} & \sigma_{LL} \end{vmatrix},$$

where

$$\sigma_{SM} = \frac{1}{F} \sum_{f=1}^{F} \{u_M u_S\}.$$

Results

The above model was estimated using maximum likelihood and estimates of the parameters and asymptotic standard errors are given in table 1. The translog variable profit function, in this case, is a quadratic approximation to an arbitrary, multiproduct variable profit funtion around a point of expansion (1,1,1,1,1). The profit function should possess the properties of monotonicity and convexity. This can be evaluated from the coefficient estimates given in table 1.

Monotonicity requires variable profit to increase with output prices and decrease with input prices, and these properties hold if the function has correct gradients with respect to output and input prices. A check for monotonicity is made by viewing the average value shares of each of the outputs and variable

inputs defined in the banking profit model. For the translog variable profit function, the share for each output and input in profit is approximated by the α_i for each share, as given in the previous equations (26a–26d) at the point of approximation. These estimated parameters are positive for both demand and time deposits and are both negative for labour and variable capital inputs, suggesting the gradients are correct.

However, a check of the convexity conditions suggests that the Hessian matrix is not positive semidefinite because all principal minors are not positive given these estimates. All Eigean values are not positive as checked at the point of approximation, indicating that curvature conditions have not met.[5] It is well known that estimated translog functions frequently fail to satisfy the convexity or concavity conditions implied by economic theory, and this study is no exception. A number of researchers have imposed the curvature property globally or locally on the translog profit function prior to estimation (e.g., Lau, and Diewert and Wales [1985, 1987]); however, the constrained estimation with the imposition of global curvature complicates the estimation. Lau also suggested that not all families of approximation functions can be made globally convex without severely restricting the parameters. In the case of translog, imposition of global convexity would also impose restriction of unitary elasticity of substitution between all pairs of commodities. Since there was no reason to expect this substitution result, no such restriction was imposed. Nonetheless, with the incorrect curvature conditions, some values of the substitution (transformation) elasticities and either the Hicks-Allen or Marshallian supply and input demand elasticities take on unexpected signs.

In this sample of banking firms, the estimated parameters suggest that labour and capital are complementary. This result is not different from some results of studies on the technology of banks using either the cost function or the profit function approach. Other studies have not used the variable capital definition used here. These previous studies used fixed factor capital and a variable input that are generally classified as materials and also found that labour and materials are generally complementary inputs. Evidence of complementarity in the study may also be related to some of the related estimates of inefficiency discussed below. The own-price elasticity of labour is of correct sign, indicating a downward sloping

[5] The possibility of multicollinearity, which is very high in such models, can affect the signs of the parameter estimates.

labour demand, but the own-price elasticity for variable capital is of unexpected positive sign. Similarly, the supply elasticity for demand deposits is of incorrect sign, while the same elasticity for time deposits is correct in sign, though very large in value, a result contrary to findings of most previous studies.

The negative intercept parameter shown in table 1 indicates a high level of fixed factors in the bank sample. The estimated parameter associated with the effect of fixed capital on variable profit is also negative. Fixed capital is primarily comprised of the value of bank premises and furniture. The current study attempted separate premises capital from equipment services, mainly computer capital (which was combined with materials to develop the variable capital input) to define variable and fixed capital. This breakout was dependent on information given in the FCA data; however, incomplete reporting of these items means accurate delineation of variable versus fixed input is a problem that influences estimates.

The inefficiency component of the disturbance term is given in table 1, while the estimates of inefficiency derived therefrom are given in table 2. It appear that on average, this set of unit banks is losing approximately 45% of their profit potential for the period of years sampled due to combined effect of allocative, technical, and scale inefficiency; however, technical inefficiency cannot be isolated in the model here specified.

The sample period involves years of change in the regulation of banking in the United States. These inefficiency estimates may reflect the regulatory structure and other operating practices of banks, but, no direct analysis of the impact of regulation on bank efficiency was done here. What has been estimated is the relative magnitudes of inefficiency that occur in the unit banking system for a particular geographic location.

The estimates of inefficiency suggest that the sample banks on average are not achieving the optimum output shares, given the definitions of outputs in this study. The demand deposit shares are found to be, on average, 8.24% less than optimum, while the time deposit shares are 3.5% less than optimum. The inefficiency measure is only in terms of output shares since estimates are derived from profit function, and the same comparison is made for input shares. Observed netput (outputs and inputs) cannot be compared with optimum netput using this specification. The difference in the estimated and optimum shares reflects inefficiencies in liquidity and liability management processes. Actual management

of reserves to generate loans and, therefore, deposits is not perfect, as reflected by the estimates, but a 3 to 8% range does not appear to be an unreasonable calculation error in such management. This difference could also reflect the loan-to-deposit abstraction that is assumed to define demand and time deposits as outputs. Number of accounts may be a better measure of bank output when measuring inefficiency.

The profit share of the labour input is found to be, on average, 38% higher than the optimum share. Given hiring prices of labour and the observed profit of the banks, this could indicate that banks overemploy labour relative to optimal employment. The literature has long debated this question, and many cost studies of banking suggest evidence of overemployment relative to efficient employment. However, inefficiency could also be due to higher wages than optimally required or to a combination of higher wages and overemployment since we are dealing with the labour profit share. This inefficiency may be related to the labour–variable capital complementarity discussed previously.

The profit share of variable capital is 13% less than the optimal share for the sample. The literature suggests that labour is overemployed and capital underutilized in the banking industry, and the findings supports that notion but in relative share terms. Estimates of the profit model do indicate complementarity between labour and capital and not substitutability, which also suggests labour–capital employment problems might exist in American banking. Even with the problems encountered in this study, banks could substantially improve their profit potential by better allocation of inputs and management of reserves and liabilities.

Suggestions for Further Study

Incomplete data on the banking system influenced the findings of this study. Additional and more complete data must be developed to provide a more complete representation of banking technology. Direct estimation of the impacts of regulatory policy and its changes over the years was not accomplished. To measure inefficiency this same approach should be used with specific regulatory information and epochs to develop estimates of the impacts of specific regulatory policies on banking organization and delivery of monetary services. Monetary asset aggregation and the supply of monetary services was not dealt with in this study, but information does exist on this subject. The concept of inefficiency measurement could be combined to study the supply of monetary services to the economy.

Using the translog approximation to the profit function and imposing the inefficiency specification on that system to obtain estimates of inefficiency, output supply, and derived input demand conditions met with rather mixed results. Particularly, the desired curvature conditions were not entirely met. As mentioned above, the translog variable profit function is a Taylor series expansion and provides only a local approximation to an arbitrary function. The flexible functional forms that use a Taylor series expansion do not have global properties and often violate the curvature conditions implied by economic theory. A particularly constructive extension would be to apply a flexible functional form that posesses global properties to produce more empirically consistent elasticity estimates. Two such flexible functional forms are the fourier flexible functional form (Gallant) and the minflex Laurent flexible functional form (Barnett[1983]). A generalized version of the latter, termed the symmetric generalized Barnett flexible functional form, could prove particularly useful. However, much more detailed in specification of the measure of inefficiency must be developed in order to use any of these specifications, but effort may prove fruitful.

This study of banking did not deal with issues of price-taking behavior relative to price-setting behavior in imperfect competition (Gilbert). Additionally, the issues of scale economies and economies of scope were not taken up, which might best be handled with a cost function approach. The issues of dynamics in banking technology likewise were not addressed. The literature contains some information, but much more is needed to develop bank policies for and to understand banking services in the macroeconomy.

The definition of bank output remains controversial. Even though some recent work has attempted to sort out production issues in banking within a microfoundations framework of money supply, considerable refinement in these definitions is needed. Deposit creation and liability management should be further addressed. Only then can estimated measures of inefficiency be traced directly to management decisions in bank operations. Development of tests to identify inefficiency from risk elements is critically needed. Finally, the further reconciliation of these measures of inefficiency behavior and technology with other concepts of inefficiency and risk elements may cause firms to operate on different frontiers.

Table IV.1 Maximum Likelihood Parameter Estimates*

Parameter	Parameters		
α_0	-0.3998	δ_{DL}	-0.2024
	(0.1406)		(0.8443)
α_D	0.8896	δ_{DK}	-0.3091
	(0.3206)		1.2128
α_T	0.3436	η_{DZ}	-1.6885
	(0.1171)		0.0336
α_L	-1.2183	δ_{TL}	1.6002
	(0.4400)		(0.5301)
α_K	-1.3221	δ_{TK}	-1.3222
	(0.5029)		(0.8146)
β_Z	-0.4590	η_{TZ}	0.3268
	(0.1569)		(0.0599)
δ_{DD}	1.1685	δ_{LK}	-0.3109
	(0.4963)		(0.8181)
δ_{TT}	0.3431	η_{LZ}	-1.5733
	(0.1171)		(0.3842)
δ_{LL}	-0.5572	η_{KZ}	0.2781
	(2.0055)		(0.0550)
δ_{KK}	-0.5263	σ_ω	0.7655
	(1.9043)		(0.0254)
δ_{ZZ}	-0.2061	σ_ν	0.3422
	(0.0660)		(0.0114)

* Standard errors in parentheses

Table IV.2 Inefficiency (Mean) Estimates

Measures		Measures	
ΠL^1_ω	44.87		
$\sigma^2_{\Pi L}$	(0.5478)		
UL^3_D	-8.24	UL^3_L	38.81
σ^2_D	(0.2777)	σ^2_L	(0.2219)
UL^3_T	-3.50	UL^3_K	-13.12
σ^2_T	(0.5325)	σ^2_K	(0.1232)

1 percentage loss of profit due to combined inefficiency

2 Standard deviation

3 percentage of allocative inefficiency of netputs

References

Aigner, D. J., C. A. K. Lovell, and P. Schmidt. "Formulation and Estimation of Stochastic Production Function Models." *J. Econometrics* 6(1977):21-37.

Alhadeff, David A. *Monopoly and Competition in Banking.* Berkeley: University of California Press, 1954.

Ali, M., and J. C. Flinn. "Profit Efficiency Among Basmati Rice Producers in Pakistan Punjab." *Amer. J. Agr. Econ.* 71(1989):302-10.

Barnett, W. A. "Economic Monetary Aggregates: An Application of Aggregation and Index Number Theory." *J. Econometrics* 14(1980):11-48.

——. "The New Monetary Aggregates." *J. Money, Credit, and Banking* 13(1981):485-89.

——. "New Indices of Money Supply and the Flexible Laurent Demand System." *J. Bus. Econ. Statis.* 1(1983):7-23.

——. "The Microeconomic Theory of Monetary Aggregation." Paper, Second International Symposium in Economic Theory and Econometrics, 1986.

——. "The Microeconomic Theory of Monetary Aggregation." *New Approaches to Monetary Economics.* Chapter 6, pp. 115-68, ed. W. A. Barnett and K. J. Singleton. Cambridge: Cambridge University Press, 1987.

Barten, A. P. "Maximum Likelihood Estimation of a Complete System of Demand Equations." *European Econ. Rev.* 1(1969):7-73.

Battese, G. E., and T. J. Coelli. "Prediction of Firm Level Technical Efficiencies With a Generalized Frontier Production Function and Panel Data." *J. Econometrics* 38(1988):387-99.

Bell, F., and N. Murphy. "Costs in Commercial Banking: A Quantitative Analysis of Bank Behavior and Its Relation to Bank Regulation." Research Report No. 41, Federal Reserve Bank of Boston, 1968.

Benston, G. J. "Branch Banking and Economies of Scale." *J. of Finance* 20(1965):312-41.

Benston, G. J., A. N. Berger, G. A. Hanweck, and D. B. Humphrey. "Economies of Scale and Scope in Banking." Research Paper in Banking and Funancial Economics, Board of Governors of the Federal Reserve System, 1983.

Benston, G. J., G. A. Hanweck, and D. B. Humphrey. "Scale Economies in Banking: A Restructuring and Reassessment." *J. Money, Credit and Banking* 14(1982):435-56.

Berndt, E. R., and L. R. Christensen. "The Translog Function and the Substatution of Equipment, Structures, and Labor in U. S. Manufacturing 1929-1968." *J. Econometrics* 1(1973):81-114.

Clark, J. A. "Estimation of Economies of Scale in Banking Using a Generalized Func-

tional Form."*J. of Money, Credit, and Banking* 16(1984):53-75.

Diewert, E. W. "Exact and Superlative Index Numbers." *J. of Econometrics.* 4(1976) :115-46.

Diewert, E. W., T. J. Wales. "Flexible Functional Forms and Global Curvature Conditions." Discussion Paper No. 85-19, University of British Columbia, Vancouver, Canada, 1985.

——. "Flexible Functional Forms and Curvature Conditions." *Econometrica* 55(1987) :43-68.

Ferrier, G. D., and C. A. K. Lovell. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence." Paper, Department of Economics, Southern Methodist University, Dallas, Texas, 1990.

Forsund, F. R., C. A. K. Lovell, and P. Schmidt. "A Survey of Frontier Production Functions and Their Relationship to Efficiency Measurement." *J. Econometrics* 13(1980):5-25.

Gallant, A. R. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form. *J. Econometrics* 15(1981):211-45.

Gilbert, A. R. "Bank Market Structure and Competition: A Survey." *J. Money, Credit, and Banking* 14(1984):617-45.

Gilligan, T., M. Smirlock, and W. Marshall. "Scale and Scope Economies in the Multi-Product Banking Firm." *J. Monetary Econ.* 13(1984):393-405.

Gramley, L. E. "A Study of Scale Economies in Banking." Report, Federal Reserve Bank of Kansas City, Missouri, 1962.

Greenbaum, S. I. "A Study of Bank Cost." *National Banking Review* 37(1967):415-34.

Greene, W. H. "On the Estimation of a Flexible Frontier Production Model." *J. Econometrics* 13(1980):101-15.

Hancock, D. C. "The Financial Firm: Production with Monetary and Non-Monetary Goods." *J. Polit. Econ.* 93(1985a):859-80.

——. "Bank Profitability, Interest Rates, and Monetary Policy." *J. Money, Credit, and Banking* 17(1985b):189-202.

——. "Aggregation of Monetary Goods: A Production Model." *New Approaches to Monetary Economics,* Chapter 9 pp. 200-18, ed. W. A. Barnett and K. J. Singleton. Cambridge: Cambridge University Press, 1987.

Horvitz, P. M. "Economies of Scale in Banking." *Private Financial Institutions.* ed. P. Horvitz Englewood Cliffs: Prentice-Hall, 1963.

Hunter, W. C., and S. G. Timme. "Technical Change, Organization Form, and the Structure of Bank Production." *J. Money, Credit and Banking* 18(1986):152-66.

Jondrow, J. C., C. A. K. Lovell, I., S. Materov, and P. Schmidt. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *J.*

Econometrics 23(1982):269-74.

Klein, B. "Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money." *Amer Econ. Rev.* 64(1974):931-99.

Kumbhakar, S. C. "The Specification of Technical and Allocative Inefficiency in Stochastic Production and Profit Frontiers." *J. Econometrics* 34(1987):335-48.

——. "On Estimation of Technical and Allocative Inefficiency in Stochastic Frontier Functions: The Case of U. S. Class 1 Railroads." *Int. Econ. Rev.* 29(1988):727-43.

Lau, L. J. "Testing and Imposing Monotonicity, Convexity, and Quasi-Convexity Constraints." *Production Economics: A Dual Approach to Theory and Applications* vol I, pp. 409-54, ed. M. Fuss and D. McFadden. Amesterdam: North-Holland Publishers, 1978.

Longbrake, W. A. "The Differential Effects of Single-Plant, Multi-Plant and Multi-Firm Organizational Forms on Cost Efficiency in Commercial Banks." Working Paper 74-7, Federal Deposit Insurance Corporation.

Longbrake, W. A., and J. A. Haslem. "Productive Efficiency in Commercial Banking: The Effects of Size and Legal Form of Organization on the Cost of Producing Demand Deposit Services." *J. Money, Credit and Banking* 7(1975):717-30.

Meeusen, J. van den Broeck. "Efficiency Estimation From Cobb-Douglas Production Functions With Composed Error." *Int. Econ. Rev.* 18(1977):435-44.

Mullineaux, D. J. "Economies of Scale and Organizational Efficiency in Banking: A Profit-Function Approach." *J. Finance* 33(1978):259-80.

Murray, J. D., and R. W. White. "Economies of Scale and Economies of Scope in Multiproduct Financial Institutions: A Study of British Columbia Credit Unions." *J. Finance.* 38(1983):887-902.

Pitt, M., and L. F. Lee. "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry." *J. Dev. Studies* 9(1981):43-64.

Powers, J. "Branch Versus Unit Banking: Bank Output and Cost Economies." *Southern Econ. J.* 36(1969):153-64.

Santomero, A M. "Modeling the Banking Firm: A Survey." *J. Money, Credit and Banking* 14(1984):576-602.

Schmidt, P. "Frontier Production Function." *Econometric Rev.* 4(1985-86):289-328.

——. "Estimation of a Fixed-Effect Cobb-Douglas System Using Panel Data." *J. Econometrics* 37(1988):361-80.

Schmidt, P., C. A. K. Lovell. "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers." *J. Econometrics* 9(1979):343-66.

——. "Estimating Stochastic Production and Cost Frontiers When Technical and Allocative Inefficiency are Correlated." *J. Econometrics* 13(1980):83-100.

Schmidt, P., R. C. Sickles. "Production Frontiers and Panel Data." *J. Bus. and Econ.*

Statis. 4(1984):367-74.

Schweiger, I. and J. S. McGee. "Chicago Banking." *J. Business* 34(1961):203-366.

Schweitzer, S. A. "Economies of Scale and Holding Company Affiliation in Banking." *Southern Econ. J.* 39(1972):258-66.

Sealey, C. W., and J. T. Lindley. "Input, Output, and a Theory of Production and Cost of Depository Financial Institutions." *J. Finance* 32(1977):1251-1266.

APPENDICES

Appendix A. Non-Frontier (Parametric) Efficiency Models

In frontier efficiency analysis, the one-sided error term is forced to represent inefficiency; however, it is possible to investigate efficiency without using one-sided error terms, and both primal and dual relations can be used to investigate inefficiency. Both technical and price inefficiencies (combination of allocative and scale inefficiencies) can be estimated for n types of firms. The production relation can be defined as

$$y_i = \alpha_i f(x_i), \qquad \text{for } i = 1, \ldots, n.$$

The term $\alpha_i > 0$ is the index of technical inefficiency, x_i is the input vector of the ith firm, and y_i is the output of ith firm. n types of firms are equally technically efficient if

$$\alpha_1 = \alpha_2 = \cdots = \alpha_i = \cdots = \alpha_n .$$

From the first order conditions for profit maximization,

$$\frac{\partial \alpha_i f(x_i)}{\partial x_{ij}} = \beta_{ij} \frac{w_{ij}}{p_i} \qquad i = 1, \ldots, n; \text{ and } j = 1, \ldots, m .$$

The term, $\beta_{ij} > 0$, is the index of price efficiency. n types of firms are equally price efficient if

$$\beta_{1j} = \beta_{2j} = \ldots = \beta_{nj}, \qquad \forall\, j = 1, \ldots, m.$$

Similarly, n types of firms are equally economic efficient if

$$\alpha_1 = \alpha_2 = \ldots = \alpha_n ;$$

and

$$\beta_{1j} = \beta_{2j} = \ldots = \beta_{nj}, \quad \forall\, j = 1, \ldots, m.$$

From the above information, n types of firms can be proven equally profit efficient if their respective profit functions coincide. For the homothetic functional specification the profit function can be derived from production function. If the technology is not homogenous, application of such model becomes very restrictive. Works of Lau and Yotopolous (1971, 1972), Yotopolous and Lau, and Trosper on measurement of inefficiency are based on this framework. Modifications of this

model by Toda and Atkinson and Halvorsen made use of nonhomothetic functional specification possible.

Although inefficiency of n types of firms can be studied, but it cannot be extended to investigate inefficiency on a firm-by-firm basis. Theoretically, any functional form (as per Toda's approach) can be used, but empirical application is quite difficult because most flexible functional forms are not self-dual. However, compared to estimation of stochastic function the econometric technique required in this type of modeling is fairly simple.

Appendix B

Appendix B.1 Derivation of the Cost of Allocative Inefficiency

From the production function (1b) and the first-order condition with allocative inefficiency, (14) we can derive the input demand equations

$$(B2.1) \qquad \ln x_q = \ln x_s + \ln\left\{\frac{w_q}{w_s}\frac{\alpha_s + \sum_j \gamma_{sj}\ln z_j}{\alpha_q + \sum_j \gamma_{qj}\ln z_j}\right\} + u_q, \quad \forall\ q.$$

Substituting these input demand equations in the constraint equation of our cost minimization problem (the log of the production function (1b)) and solving for $\ln x_q$, we get

$$\begin{aligned}\ln x_q^* = \ln\Big\{&(\alpha_q + \sum_j \gamma_{qj}\ln z_j)\Big\{\alpha_0 \prod_i (\alpha_i + \sum_j \gamma_{ij}\ln z_j)^{(\alpha_i + \sum_j \gamma_{ij}\ln z_j)}\Big\}^{-\theta}\Big\} \\ (B2.2) \qquad &- \ln w_q - \theta\sum_j \beta_j \ln z_j + \theta\sum_i (\alpha_i + \sum_j \gamma_{ij}\ln z_j)\ln w_i \\ &+ \theta\ln y - \theta(\nu+\tau) - u_q - \theta\sum_q \Big(\alpha_q + \sum_j \gamma_{qj}\ln z_j\Big)u_s.\end{aligned}$$

The x_q^* are solution input demand equations (constant output). Substituting x_i^* in the objective function gives the solution cost function, which is not the optimum solution cost. This is minimum cost in the presence of allocative and technical inefficiencies. Substituting x_q^* in $C = W.X$ and taking the log gives

$$\begin{aligned}\ln C^*(\cdot) = \ln\Big\{&\alpha_0 \prod_i \Big(\alpha_i + \sum_j \gamma_{ij}\ln z_j\Big)^{(\alpha_i + \sum_j \gamma_{ij}\ln z_j)}\Big\}^{-\theta} \\ &\theta\ln y + \ln w_i^{\theta\sum_i (\alpha_i + \sum_j \gamma_{ij}\ln z_j} \\ (B2.3) \qquad &- \theta\Big\{\sum_j \beta_j \ln z_j + \sum_q (\alpha_q + \sum_j \gamma_{qj}\ln z_q)u_q + (\nu+\tau)\Big\} \\ &+ \ln\Big\{(\alpha_s + \sum_j \gamma_{sj}\ln z_j) + \sum_s \Big(\alpha_s + \sum_j \gamma_{sj}\ln z_j\Big)\exp(-u_s)\Big\}\end{aligned}$$

where $\ln C(\cdot) = \ln C(y, W, \tau, u_j, Z)$.

After some rearrangment and algebraic manipulation, the log of the cost function with allocative and technical inefficiency terms is

$$(B2.4) \qquad \ln C^*(\cdot) = \theta \ln y + \ln K + \theta \sum_i \left(\alpha_i + \sum_j \gamma_{ij} \ln z_j\right) \ln w_i$$
$$- \theta(\nu + \tau) - \theta \sum_j \beta_j \ln z_j + \left\{\rho - \ln(r + k)\right\},$$

where the term on the right-hand side $\{\rho - \ln(r + k)\}$ represents the cost of allocative inefficiency.

Appendix B.2 KBB Measure
of Scale Inefficiency

The measure of scale inefficiency used in this study is that of Forsund, Lovell, and Schmidt. The definition used is

$$\frac{\partial C^*(W,y,Z)}{\partial y} = p\,\exp(\xi)\,,$$

where $C^*(\cdot)$ is the cost net of allocative and technical inefficiencies. On the other hand, the KBB measure uses the definition

$$\frac{\partial C(W,y,Z)}{\partial y} = p\,\exp(\xi_{KBB})\,,$$

The ξ_{KBB} is not net of other inefficiencies. Forsund, Lovell, and Schmidts' definition is based on efficient marginal cost, while the KBB definition is based on actual cost (or observed cost). In the second case, the optimum level of output is conditional on technical and allocative inefficiency. Two measures are not same and may lead to different results. Conceptually, both measures are appealing, but Forsund, Lovell, and Schmidts' measure is consistent with the error structure imposed for the estimation process. For the quasi-translog model, the KBB measure of scale inefficiency can be derived from the output supply equation using the KBB definition. The resulting output supply equation is

$$\ln y = \ln x_s - \ln \alpha_s + \ln p + \ln w_s - \ln\left\{1 + \exp(-u_j)\right\} - \xi_{KBB}\,.$$

The resulting measure of scale inefficiency, therefore, is

$$\hat{\xi}_{KBB} = \ln x_s - \ln y + \ln w_s - \ln \hat{\alpha}_s - \ln p + \ln\left\{1 + \exp(-\hat{u}_j)\right\}$$

The optimum level of output corresponding to Forsund, Lovell, and Schmidts' measure is greater than that of the KBB measure. The magnitude of difference will increase with the size of technical and allocative inefficiencies that are embedded in the KBB measure.

Appendix C

Appendix C.1

The art of functional form specification has dramatically advanced in recent years, primarily because of the pioneering work of Diewert (1971, 1973). The main criterion for specifying a functional form is that it be capable of satisfying the appropriate regularity conditions, at least over a region, if not globally. Generally, two types of criteria are available in the literature to specify a flexible functional form. The first one is called the weak approximation criterion, first proposed by Diewert, the second one is Taylors' series approximation, usually called the strong approximation criterion.

Diewert pointed out that "the functional form contains precisely the number of parameters needed to provide a *second order approximation* to an arbitrary twice differentiable...function satisfying the appropriate regularity conditions...."(1973; p. 285). Diewert called such a function a flexible functional form if the parameters of the functional form can be chosen to make values of its first- and second-order derivatives equal to the first and second derivatives of the function being approximated at any point (of approximation) in the domain. That is, a function $\tilde{F}$ is a second-order approximation to the true function F at the point, say X^*, if the following set of conditions are satisfied:

$$\tilde{F}(X^*) = F(X^*),$$

$$\frac{\partial \tilde{F}(X^*)}{\partial x_i} = \frac{\partial F(X^*)}{\partial x_i}$$

$$\frac{\partial^2 \tilde{F}(X^*)}{\partial x_i^2} = \frac{\partial^2 F(X^*)}{\partial x_i^2}.$$

This is the weak approximation criterion for specifying a flexible function form.

Using the strong criterion for specifying the flexible functional form, a Taylor's series approximation, the true function is defined as

$$F(x_1, \ldots, x_n) = \tilde{F}(f_1(x_1), \ldots, f_n(x_n)),$$

where $f_i(.)$ are arbitrary, twice differentiable transformations of the original variables. Taking a twice differentiable monotonic transformation, say θ, of both

sides gives

$$\theta^{\circ} F(X) = \theta^{\circ} \tilde{F}(f_1(x_1), \ldots, f_n(x_n)).$$

By Taylor's series expansion of $\theta^{\circ}\tilde{F}$ about $(f_1(\bar{x}_1), \ldots, f_n(\bar{x}_n))$,

$$\begin{aligned}
\theta^{\circ} F(X) &= \theta^{\circ} \tilde{F}(f_1(\bar{x}_1, \ldots, f_n(\bar{x}_n)) \\
&\quad \sum_{i=1}^{n} \Big(\frac{\partial}{\partial x_i} \Big\{ \theta^{\circ} \tilde{F}(f_1(\bar{x}_1), \ldots, f_n(\bar{x}_n) \Big\} \Big) \Big(f_i(x_i) - f_i(\bar{x}_i) \Big) \\
&\quad + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \Big(\frac{\partial^2}{\partial x_i \partial x_j} \Big\{ \theta^{\circ} \tilde{F}(f_1(\bar{x}_1), \ldots, f_n(\bar{x}_n)) \Big\} \Big\{ f_i(x_i) - f_i(\bar{x}_i) \Big\} \times \\
&\quad \Big\{ f_j(x_j) - f_j(\bar{x}_j) \Big\} + \ldots \\
&= \bar{\alpha}_0 + \sum_{i=1}^{n} \bar{\alpha}_i \Big\{ f_i(x_i) - f_i(\bar{x}_i) \Big\} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{\beta}_{ij} \Big\{ f_i(x_i) - f_i(\bar{x}_i) \Big\} \times \\
&\quad \Big\{ f_j(x_j) - f_j(\bar{x}_j) \Big\} + \ldots \\
&\quad \bar{\beta}_{ij} = \bar{\beta}_{ji}, \qquad \forall i, j = 1, \ldots, n
\end{aligned}$$

where $\bar{\alpha}_0$, $\bar{\alpha}_i$, and $\bar{\beta}_{ij}$, are values of the first and second partial derivatives of $\theta^{\circ}\tilde{F}$ at the point $\bar{X}$. Rearranging and eleminating third and higher-order terms givesthe second-order Taylor's series approximation to $\theta^{\circ}\tilde{F}$ at $\bar{X} = 1$,

$$\begin{aligned}
\overline{\theta^{\circ} F(X)} &= \Big(\bar{\alpha}_0 - \sum_{i=1}^{n} \bar{\alpha}_i f_i(\bar{x}_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{\beta}_{ij} f_i(\bar{x}_i) f_j(\bar{x}_j) \Big) \\
&\quad + \sum_{i=1}^{n} \Big(\bar{\alpha}_i - \sum_{j=1}^{n} \beta_{ij} f_j(\bar{x}_j) \Big) f_i(x_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{\beta}_{ij} f_j(x_i) f_j(x_j)
\end{aligned}$$

or

$$\begin{aligned}
\overline{\theta^{\circ} \tilde{F}(X)} &= \alpha_0 + \sum_{i=1}^{n} \alpha_i f_i(x_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} f_i(x_i) f_j(x_j), \\
\beta_{ij} &= \beta_{ji} \qquad \forall i, j = 1, \ldots, n,
\end{aligned}$$

where

$$\alpha_0 = \bar{\alpha}_0 - \sum_{i=1}^{n} \bar{\alpha}_i f_i(\bar{x}_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{\beta}_{ij} f_i(\bar{x}_i) f_j(\bar{x}_j),$$

$$\alpha_i = \bar{\alpha}_i - \sum_{j=1}^{n} \bar{\beta}_{ij} f_j(\bar{x}_j) \quad \forall i, j = 1, \ldots, n,$$

and

$$\beta_{ij} = \bar{\beta}_{ij}, \quad \forall i, j, = 1, \ldots, n.$$

Thus, any functional form that can be written like the right-hand side of the above equation has a Taylor's series interpretation; that is, it can be interpreted as a second-order Taylor's series approximation in transformed variables to a monotonic transformation of the true underlying function.

Appendix C.2

If v is a non-positive value of a $N(b, \sigma_v^2)$ variable truncated at zero, then the conditional probability density function of v, given say b_1, $h(v|b_1)$, is $N(\mu_v, \sigma^{*2})$ truncated at zero.

Proof

$$h(v|b_1) = \frac{f(b_1, v)}{f(b_1)}$$

By variable transformation,

$$f(b_1, v) = f(\nu) f(v) |J|$$

where $|J|$ is the Jacobian of transformation, and $|J| = 1$.

$$f(b_1, v) = \frac{\exp(-d/2)}{\pi \sigma_\nu \sigma_v} \exp\Big\{ -\frac{1}{2} \Big(\frac{v - \mu_v}{\sigma} \Big)^2 \Big\}$$

To derive the $f(b_1)$, v must be integrated out of $f(b_1, v)$ as

$$\begin{aligned} f(b_1) &= \int_{-\infty}^{0} f(b_1, v) dv \\ &= \frac{2\exp(-d/2)}{\sqrt{2\pi} \sigma_\nu \sigma_v} \sigma \Phi\Big(-\frac{\mu_v}{\sigma} \Big), \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative normal density. Therefore,

$$\begin{aligned} h(v|b_1) &= \frac{\sigma \frac{1}{\sqrt{2\pi}\sigma} \exp\{ -\frac{1}{2} (\frac{v - \mu_v}{\sigma})^2 \}}{\sigma \Phi(-\frac{\mu_v}{\sigma})} \\ &= \frac{\phi((v - \mu_v)/\sigma)}{\Phi(-\mu_v/\sigma)}, \end{aligned}$$

which is an $N(\mu_v^*, \sigma^{*2})$ variable truncated at zero.

Appendix C.3

The mean of the truncated normal, when truncated at zero, is

$$E(v) = \mu_v - \sigma \frac{\phi(\mu_v/\sigma)}{\Phi(-\mu_v/\sigma)}.$$

Proof

Let $X \sim N(\mu, \sigma^2)$ truncated at a. Then the probability density of the truncated normal is

$$f^*(x) = \frac{f(x)}{\Phi((a-x)/\sigma)},$$

where $f(x)$ is the pdf of the untruncated random variable, and $\Phi(\cdot)$ is the cpdf of the standard normal at the point of truncation. The moment generating function of the truncated normal random variable, say x^*, is

$$M_{x^*}(t) = E(e^{tx}) = \int_{-\infty}^{a} e^{tx} f^*(x) dx$$

$$\int_{-\infty}^{a} e^{tx} \frac{f(x)}{\Phi((a-\mu)/\sigma)} dx$$

$$\exp(\mu t + \frac{1}{2} t^2 \sigma^2) \frac{\Phi\left(\frac{a-(\mu+t\sigma^2)}{\sigma}\right)}{\Phi\left(\frac{a-\mu}{\sigma}\right)}$$

at $t = 0$, $M_x^*(0) = 1$. Taking the log of both sides results in

$$\ln M_x^*(t) = \mu t + \frac{1}{2} t^2 \sigma^2 + \ln \Phi\left(\frac{a-(\mu+t\sigma^2)}{\sigma}\right) - \ln \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Then

$$M'_{x^*}(t) = \left\{\mu + t\sigma^2 + \frac{1}{\Phi\left(\frac{a-(\mu+\sigma^2 t)}{\sigma}\right)} \phi\left(\frac{a-(\mu+t\sigma^2)}{\sigma}\right)(-\sigma)\right\} M_{x^*}(t)$$

at $t = 0$,

$$M'_{x^*}(0) = \left\{\mu + \frac{\phi((a-\mu)/\sigma)}{\Phi((a-\mu)/\sigma)}(-\sigma)\right\}$$

$$= \mu - \sigma \frac{\phi((a-\mu)/\sigma)}{\Phi((a-\mu)/\sigma)}$$

is the mean of the truncated normal when the random variable x is truncated at a. In this case, τ is the random variable truncated at 0; therefore,

$$\hat{\tau} = \mu_\tau - \sigma\left\{\frac{\phi(\mu_\tau/\sigma)}{\Phi(-\mu_\tau/\sigma)}\right\}.$$

References

Atkinson, S. E., and R. Halvorsen. "Parametric Efficiency Tests, Economies of Scale, and Input Demand in U.S. Electric Power Generation." *Int. Econ. Rev.* 25(1984):647-62.

Diewert, W. E. "An Application of the Shephard Duality Theorem: Generalized Leontief Production Function." *J. Polit. Econ.* 79(1971):481-507.

——. "Functional Forms for Profit and Transformation Functions." *J. Econ. Theory* 6(1973):284-316.

Forsund, F. R., C. A. K. Lovell, and P. Schmidt. "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurements." *J. Econometrics* 13(1980):5-25.

Lau, L. J., and Pan. A. Yotopolous. "A Test of Relative Economic Efficiency and Application to Indian Agriculture." *Amer. Econ. Rev.* 61(1971):94-109.

——. "Profit, Supply, and Factor Demand Functions." *Amer. J. Agr. Econ.* 54(1972):11-18.

Toda, Y. "Estimation of Cost Function When the Cost is Not Minimum: The Case of Soviet Manufacturing Industries, 1958-71." *Rev. Econ. Statist.* 58(1976):259-68.

Trosper, R. L. "American Indian Relative Ranching Efficiency." *Amer. Econ. Rev.* 68(1978):503-16.

Yotopolous, Pan. A., and L. Lau. "A Test of Relative Economic Efficiency and Application to Indian Agriculture: Some Further Results." *Amer. Econ. Rev.* 63(1973):214-23.

VITA

Arunava Bhattacharyya

Candidate for the Degree of
Doctor of Philosophy

Dissertation: Production and Inefficiency

Major Field: Economics

Biographical Information :

Personal Data:

Born in Calcutta, India, January, 1953, son of Mr. Anilava Bhattacharyya and Mrs. Rani Bhattacharyya; married Mrs. Anjana Bhattacharyya, December, 1981.

Education :

Graduated from Maha Raja Manindra Chandra College, Calcutta University, as Bachelor of Arts, with major in Economics, in April 1974. Completed Masters of Arts in Economics from Calcullta University in 1977. Completed Master of Science in Economics from Utah State University in 1989. Received President's Fellowship from the Utah State University in 1989. Completed requirements for Doctor of Philosophy degree in Economics at Utah State University in 1990.

Professional Experience :

Lecturer in Economics, Bagnan College, West Bengal, India, from February 1979. Associate Fellow to the Centre for Urban Economic Studies, Department of Economics, University of Calcutta, Calcutta, India, from January 1981.