Stochastic Frontier Analysis

Modern textbook presentations of production economics typically treat producers as successful optimizers. Conventional econometric practice has generally followed this paradigm, and least squares–based regression techniques have been used to estimate production, cost, profit, and other functions. In such a framework deviations from maximum output, from minimum cost and cost-minimizing input demands, and from maximum profit and profit-maximizing output supplies and input demands are attributed exclusively to random statistical noise. However, casual empiricism and the business press both make persuasive cases for the argument that, although producers may indeed attempt to optimize, they do not always succeed. This book develops econometric techniques for the estimation of production, cost and profit frontiers, and for the estimation of the technical and economic efficiency with which producers approach these frontiers. Because these frontiers envelop rather than intersect the data, and because the authors continue to maintain the traditional econometric belief in the presence of external forces contributing to random statistical noise, the work is titled *Stochastic Frontier Analysis*.

Subal C. Kumbhakar is Associate Professor of Economics at the University of Texas, Austin, where he has taught for the past 12 years. He received his Ph.D. from the University of Southern California. Professor Kumbhakar has written more than 60 articles in refereed journals and holds an honorary doctorate from Gothenburg University, Sweden. He is a member of the editorial board of *Technological Forecasting and Social Change: An International Journal* and *The Journal of Productivity Analysis*, and is Associate Editor of the *American Journal of Agricultural Economics*.

C. A. Knox Lovell holds the C. Herman and Mary Virginia Terry Chair in the Department of Economics in the Terry College of Business, University of Georgia, and is Professor in the School of Economics at the University of New South Wales, where he serves as Director of the Centre for Applied Economic Research. He previously taught at the University of North Carolina, Chapel Hill, for 25 years. Professor Lovell is coordinator of the Georgia Productivity Workshop, Editor-in-Chief of *The Journal of Productivity Analysis,* and Associate Editor of *Operations Research*. He has written more than 100 academic papers and authored or edited 6 books, including *Production Frontiers* (with Rolf Färe and Shawna Grosskopf, Cambridge University Press, 1994) and *The Measurement of Productive Efficiency* (with Harold Fried and Shelton Schmidt).

Stochastic Frontier Analysis

SUBAL C. KUMBHAKAR
University of Texas, Austin

C. A. KNOX LOVELL
University of Georgia
University of New South Wales

CAMBRIDGE
UNIVERSITY PRESS

Contents

Preface

Modern textbook presentations of production economics, from Samuelson's innovative *Foundations of Economic Analysis* to the present day, treat producers as successful optimizers. They produce maximum outputs allowable by the technology in place and the resources at their disposal. They minimize the cost of producing whatever outputs they choose to produce, given the technology in place and the input prices they face. Cost-minimizing input demands are derived from the resulting minimum cost function by means of Shephard's lemma. They also maximize profit, given the technology in place and the output and input prices they face. Profit-maximizing output supplies and input demands are derived from the resulting profit function by means of Hotelling's lemma.

Conventional econometric practice, beginning with the pioneering work of Cobb and Douglas, has generally followed this theoretical paradigm. Thus least squares–based regression techniques are used to estimate the parameters of production, cost, and profit functions. In such a framework departures from maximum output, from minimum cost and cost-minimizing input demands, and from maximum profit and profit-maximizing output supplies and input demands, are attributed exclusively to random statistical noise.

However casual empiricism and the business press both make persuasive cases for the argument that, although producers may indeed *attempt* to optimize, they do not always succeed. It is desirable, therefore, to develop a theory of producer behavior in which the motivations are unchanged, but in which success is not guaranteed. It is similarly desirable to modify conventional econometric techniques for the estimation of production, cost, and profit relationships that allow for failure in efforts to optimize. Finally it is desirable for the modified econometric approach to allow some producers to be relatively more successful than others, so as to provide the basis for a subsequent investigation into the determinants of variation in the efficiency with which producers pursue their objectives.

This book is concerned with the development of a modified econometric approach to the estimation of productive efficiency. We call it *Stochastic Frontier Analysis* because we are concerned with the estimation of frontiers, which envelop data, rather than with functions, which intersect data. We associate proximity to estimated frontiers with the degree of efficiency with which producers pursue their objectives. Their objectives can be purely technological or economic in nature, and so we are concerned with the estimation of production frontiers and also with the estimation of cost and profit frontiers. Finally, the frontiers we estimate are stochastic, because we continue to maintain the traditional econometric belief in the presence of external forces contributing to random statistical noise.

Portions of this book owe a debt of gratitude to the innovators whose ideas we have borrowed and extended. Without generating a list, which could be very lengthy, we want to mention, in roughly chronological order, Harold Hotelling, Ronald Shephard, and Erwin Diewert for their work on duality theory; Tjalling Koopmans and Michael Farrell for their original work on efficiency measurement; Harvey Leibenstein for his insistence in the face of considerable academic skepticism that X-inefficiency really does Xist in the real world; and Sydney Afriat, Dennis Aigner, Peter Schmidt, and Bill Greene for their early and influential contributions to the stochastic approach to efficiency analysis. Although our book is based on the work of these innovators, it also contains much original material which hopefully extends their work in useful ways.

We wish to express our gratitude to Tim Coelli, whose comments were most helpful in revising Chapters 3 and 4; to Julie Millington, who created all the figures in the text, compiled the indexes, and simply took command during the final stages of the preparation of the manuscript; and to Scott Parris, our patient and understanding editor at Cambridge University Press.

1 Introduction

1.1 THE OBJECTIVES OF THE BOOK

A recent article in the *Economist* notes that most large European banks spend about two-thirds of their revenues on rent and employee expense. Credit Suisse and Deutsche Bank have expense ratios in excess of 70%. But Sweden's Svenska Handelsbanken has an expense ratio of barely 45%, despite relatively high wage rates. The article then proceeds to describe how, ever since a near-fatal financial crisis in the late 1960s, managements at Svenska Handelsbanken have striven to cut costs while, at the same time, increasing revenues. The impression one is left with is that, far from operating in a more favorable environment, Svenska Handelsbanken is more efficient than other large European banks. This and countless other examples raise the question of how conventional microeconomic theory and conventional econometric analysis deal with variation in productive efficiency.

Typical microeconomics texts develop models of production, cost, and profit in something like the following sequence. They begin with a production function, and producers are assumed to operate on their production functions, maximizing outputs obtainable from the inputs they use. First-order conditions for cost minimization are then introduced, and producers are assumed to satisfy these conditions, allocating inputs efficiently and ending up on their cost functions. Finally, first-order conditions for profit maximization are introduced, and producers are assumed to satisfy these conditions as well, allocating outputs and inputs efficiently and ending up on their profit functions.

For many years econometricians have implemented the textbook paradigm by estimating production, cost, and profit functions, on the assumption that producers actually operate on these functions, apart from randomly distributed statistical noise. Cobb and Douglas (1928), Arrow et al. (1961), Berndt and Christensen (1973), and their followers have estimated increasingly flexible production functions in an effort to learn something about the structure of production. Nerlove (1963) was perhaps the first to exploit duality theory to estimate a cost function for the same purpose. Christensen, Jorgenson, and Lau (1973) were perhaps the first to estimate a flexible profit function. It is notable that each of these studies, and the vast majority of subsequent studies, have used least squares techniques, or variants of least squares techniques, in which error terms were assumed to be symmetrically distributed with zero means. The only source of departure from the estimated function was assumed to be statistical noise.

However the anecdotal evidence cited previously, and much other empirical evidence as well, suggests that not all producers are always so successful in solving their optimization problems. Not all producers succeed in utilizing the minimum inputs required to produce the outputs they choose to produce, given the technology at their disposal. In our jargon, not all producers are technically efficient. Consequently not all producers succeed in minimizing the expenditure required to produce the outputs they choose to produce. In addition, even if they are technically efficient, not all producers succeed in allocating their inputs in a cost-effective manner, given the input prices they face, and this misallocation of inputs contributes further to their failure to minimize the expenditure required to produce the outputs they choose to produce. In our jargon, not all producers are cost efficient. Consequently not all producers succeed in maximizing the profit resulting from their production activities. In addition, even if they are cost efficient, not all producers succeed in allocating their outputs in a revenue-maximizing manner, given the output prices they face, and this misallocation of outputs contributes further to their failure to maximize profit. In our jargon, not all producers are profit efficient.

In light of the evident failure of (at least some) producers to optimize, it is desirable to recast the analysis of production, cost, and profit away from the traditional functions toward frontiers. Thus a *production frontier* characterizes the minimum input bundles required to produce various outputs, or the maximum output producible with various input bundles, and a given technology. Producers operating on their production frontier are labeled technically efficient, and producers operating beneath their production frontier are labeled technically inefficient. A dual *cost frontier* characterizes the minimum expenditure required to produce a given bundle of outputs, given the prices of the inputs used in its production and given the technology in place. Producers operating on their cost frontier are labeled cost efficient, and producers operating above their cost frontier are labeled cost inefficient. Similarly a dual *revenue frontier* characterizes the maximum revenue obtainable from a given bundle of inputs, given the prices of the outputs produced and given the technology in place. Producers operating on their revenue frontier are labeled revenue efficient, and producers operating beneath their revenue frontier are labeled revenue inefficient. Finally a dual *profit frontier* characterizes the maximum profit obtainable from production activity, given the prices of the inputs used and the prices of the outputs produced and given the technology in place. Producers operating on their profit frontier are labeled profit efficient, and producers operating beneath their profit frontier are labeled profit inefficient. In each of these four cases interest naturally centers on the magnitude of each type of inefficiency and on the determinants of each type of inefficiency.

The econometric implication of this proposed reformulation from functions to frontiers is that symmetrically distributed error terms with zero means are no longer appropriate when analyzing producer behavior. The possibility remains that a producer will end up above the deterministic kernel of an estimated production, revenue, or profit frontier (or beneath an estimated cost frontier) due to an unusually favorable operating environment. But it is considerably more likely that a producer will end up beneath an estimated production, revenue, or profit frontier (or above an estimated cost frontier), because two factors work in this direction. First, if environmental effects are random as is typically assumed, then an unfavorable operating environment is just as likely to occur as is a favorable operating environment, and this causes a producer to end up beneath an estimated production, revenue, or profit frontier (or above an estimated cost frontier). Second, failure to optimize in each of the senses discussed previously also causes a producer to end up beneath an estimated production, revenue, or profit frontier (or above an estimated cost frontier).

Consequently error terms associated with frontiers are "composed" error terms, composed of a traditional symmetric random-noise component and a new one-sided inefficiency component. These composed error terms cannot be symmetric and they cannot have zero means. They must be skewed (negatively in the case of production, revenue, and profit frontiers and positively in the case of cost frontiers), and they must have nonzero means (negative in the case of production, revenue, and profit frontiers and positive in the case of cost frontiers).

In this reformulation production, cost, revenue, and profit frontiers are stochastic, due to random variation in the operating environment, and deviations from these stochastic frontiers are one-sided, due to various types of inefficiency. The retention of symmetric error components designed to capture the effects of random variation in the operating environment is in keeping with the older least squares–based approach to the estimation of production, cost, revenue, and profit functions. The introduction of one-sided error components designed to capture the effects of inefficiency is new, and constitutes the econometric contribution to the estimation of production, cost, revenue, and profit frontiers. Consequently we refer to this body of work as *Stochastic Frontier Analysis*, the title of our book.

1.2 A BRIEF HISTORY OF THOUGHT

In this section we recall some of the more influential antecedents, both theoretical and empirical, of stochastic frontier analysis (which we abbreviate SFA). We continue by recalling some of the origins of SFA, the events that led to the original developments in the field. We conclude with a brief summary of what we believe to have been some of the most significant developments in SFA since its inception in 1977. Many of these developments are discussed in detail in the remainder of the book.

1.2.1 Intellectual Antecedents of Stochastic Frontier Analysis

Many years ago Hicks (1935; 8) observed that "people in monopolistic positions . . . are likely to exploit their advantage much more by not bothering to get very near the position of maximum profit, than by straining themselves to get very close to it. The best of all monopoly profits is a quiet life." Hicks's suggestion that the absence of competitive pressure might allow producers the freedom to not fully optimize conventional textbook objectives, and, by implication, that the presence of competitive pressure might force producers to do so, has been adopted by many writers. Thus Alchian and Kessel (1962; 166) asserted that "[t]he cardinal sin of a monopolist . . . is to be too profitable." In a similar vein Williamson (1964) argued that, given the freedom to do so, managers would seek to maximize a utility function with staff and "emoluments" as arguments in addition to profit.

An argument related to Williamson's, arising from the property rights literature, asserts that public production is inherently less efficient than private production. This argument, due originally to Alchian (1965), asserts that concentration and transferability of private ownership shares creates an incentive for private owners to monitor managerial performance, and that this incentive is diminished for public owners, who are dispersed and whose ownership is not transferable. Consequently public managers have greater freedom to pursue their own objectives at the expense of conventional objectives. Thus Niskanen (1971) argued that public managers are budget maximizers, de Alessi (1974) argued that public managers exhibit a bias toward capital-intensive budgets, and Lindsay (1976) argued that public managers exhibit a bias toward "visible" inputs.

Ownership forms are more variegated than just private or public. Hansmann (1988) identifies investor-owned firms, customer-owned firms, worker-owned firms, and firms without owners (nonprofit enterprises). Each deals differently with problems associated with hierarchy, coordination, incomplete contracts, and monitoring and agency costs. This leads to the expectation that different ownership forms will generate differences in performance. Much of the theoretical literature on which this expectation is based is surveyed by Holmstrom and Tirole (1989).

At a somewhat more micro level, Simon (1955, 1957) analyzed the performance of producers in the presence of bounded rationality and satisficing behavior. Later Leibenstein (1966, 1975, 1976, 1978, 1987, and elsewhere) argued that production is bound to be inefficient as a result of motivation, information, monitoring, and agency problems within the firm. This rather amorphous type of inefficiency, inelegantly dubbed "X-inefficiency," has been criticized by Stigler (1976), de Alessi (1983), and others, on the grounds that it reflects an incompletely specified model rather than a failure to optimize. Unfortunately the difficult problem of model specification – including a complete list of inputs and outputs, and perhaps conditioning variables as well, a list of constraints, technological, and other (e.g., regulatory), and a proper specification of the objective function – has faced us forever, and will continue to do so.

The extent to which the literature just cited actually influenced the development of SFA is not obvious. Suffice it to say that most of us were aware of this literature, but that it did not exert the impact that hindsight suggests that it should have. Most of us were more directly influenced by another literature. Nonetheless, in retrospect this literature does suggest that the development of SFA was a useful idea if it could be used to shed empirical light on the theoretical issues raised.

The literature that did directly influence the development of SFA was the theoretical literature on productive efficiency, which began in the 1950s with the work of Koopmans (1951), Debreu (1951), and Shephard (1953). Koopmans provided a definition of technical efficiency: A producer is technically efficient if, and only if, it is impossible to produce more of any output without producing less of some other output or using more of some input. Debreu and Shephard introduced distance functions as a way of modeling multiple-output technology, but more importantly from our perspective as a way of measuring the radial distance of a producer from a frontier, in either an output-expanding direction (Debreu) or an input-conserving direction (Shephard). The association of distance functions with technical efficiency measures was pivotal in the development of the efficiency measurement literature.

Farrell (1957) was the first to measure productive efficiency empirically. Drawing inspiration from Koopmans and Debreu (but apparently not from Shephard), Farrell showed how to define cost efficiency, and how to decompose cost efficiency into its technical and allocative components. He also provided an empirical application to U.S. agriculture, although he did not use econometric methods. His use of linear programming techniques inspired the unfortunately neglected work of Boles (1966), Bressler (1966), Seitz (1966), and Sitorus (1966) in agricultural economics and eventually influenced the development of data envelopment analysis (DEA) by Charnes, Cooper, and Rhodes (1978). DEA is by now a well-established nonparametric (but in practice largely nonstochastic) efficiency measurement technique widely employed in management science.

Of greater significance in the present context is the influence Farrell's work exerted on Aigner and Chu (1968), Seitz (1971), Timmer (1971), Afriat (1972), and Richmond (1974), for it was the work of these writers that led directly to the development of SFA. Although the contributions of these authors differ in a number of important respects, it is probably fair to say that each "estimated" a deterministic production frontier, either by means of linear programming techniques or by modifications to least squares techniques requiring all residuals to be nonpositive. Afriat (p. 581) went so far as to note that "a production function $f(x)$, together with a probability distribution $\rho_\theta(e)$ of efficiency, is constructed so that the derived efficiencies $e_i = y_i/f(x_i)$ have maximum likelihood." Afriat suggested a beta distribution for $\rho_\theta(e)$ and a gamma distribution for $\rho_\theta[-\ln(e)]$ in log-linear models, an idea that Richmond (1974) explored further. Later Schmidt (1976) showed that the programming estimators of Aigner and Chu were consistent with maximum likelihood "estimation" with one-sided errors distributed as either exponential or half normal. Thus began the association of technical inefficiency with specific one-sided error distributions. However it is worth reiterating that the only source of error in these models was inefficiency; they were purely deterministic frontier models lacking a symmetric random-noise error component. However Aigner and Chu recommended, and Timmer experimented with, variants of chance-constrained programming in an ex post attempt to test the sensitivity of their "estimates" to outlying observations.

Aigner, Amemiya, and Poirier (1976) proposed a model in which errors were allowed to be both positive and negative, but in which positive and negative errors could be assigned different weights. Ordinary least squares emerges as a special case of equal weights, and a deterministic frontier model emerges as another special case (weights of zero and one in the case of production, revenue, and profit frontier models). They considered estimation for the case in which the weights are known, and for the considerably more difficult case in which the weights are unknown and are to be estimated along with the other parameters in the model. They did not actually estimate the model, and to our knowledge no one else has estimated the model. Nonetheless, it is a short step from the Aigner, Amemiya, and Poirier model with larger weights attached to negative errors to a composed error stochastic production frontier model. The step took a year.

1.2.2 The Origins of Stochastic Frontier Analysis

SFA originated with two papers, published nearly simultaneously by two teams on two continents. Meeusen and van den Broeck (MB) (1977) appeared in June, and Aigner, Lovell, and Schmidt (ALS) (1977) appeared a month later. The ALS paper was in fact a merged version of a pair of remarkably similar papers, one by Aigner and the other by Lovell and Schmidt. The ALS and MB papers are themselves very similar. Both papers were three years in the making, and both appeared shortly before a third SFA paper by Battese and Corra (1977), the senior author of which had been a referee of the ALS paper.

These three original SFA models shared the composed error structure mentioned previously, and each was developed in a production frontier context. The model can be expressed as $y = f(x; \beta) \cdot \exp\{v - u\}$, where y is scalar output, x is a vector of inputs, and β is a vector of technology parameters. The first error component $v \sim N(0, \sigma_v^2)$ is intended to capture the effects of statistical noise, and the second error component $u \geq 0$ is intended to capture the effects of technical inefficiency. Thus producers operate on or beneath their stochastic production frontier $[f(x; \beta) \cdot \exp\{v\}]$ according as $u = 0$ or $u > 0$. MB assigned an exponential distribution to u, Battese and Corra assigned a half normal distribution to u, and ALS considered both distributions for u. Parameters to be estimated include β, σ_v^2, and a variance parameter σ_u^2 associated with u. Either distributional assumption on u implies that the composed error $(v - u)$ is negatively skewed, and statistical efficiency requires that the model be estimated by maximum likelihood. After estimation, an estimate of mean technical inefficiency in the sample was provided by $E(-u) = E(v - u) = -(2/\pi)^{1/2}\sigma_u$ in the normal–half normal case and by $E(-u) = E(v - u) = -\sigma_u$ in the normal–exponential case.

1.2.3 Developments in Stochastic Frontier Analysis since 1977

In an early survey of various approaches to frontier analysis and efficiency measurement, Førsund, Lovell, and Schmidt (1980; 14) wrote that "the main weakness of the stochastic frontier model [is that] it is not possible to decompose individual residuals into their two components, and so it is not possible to estimate technical inefficiency by observation. The best that one can do is to obtain an estimate of mean inefficiency over the sample." Smart audiences in Washington and Moscow in the winter of 1980–1981 quickly detected the error in that statement. The result was the Jondrow et al. (1982) (JLMS) paper, in which either the mean or the mode of the conditional distribution $[u_i|v_i - u_i]$ was proposed to provide estimates of the technical inefficiency of each producer in the sample. The possibility of obtaining producer-specific estimates of efficiency has greatly enhanced the appeal of SFA.

The half normal and exponential distributions assigned to the one-sided inefficiency error component are single-parameter distributions, and researchers soon developed more flexible two-parameter distributions for the inefficiency error component. Drawing inspiration from Afriat and Richmond, Greene (1980a, b) proposed a Gamma distribution, and Stevenson (1980) proposed Gamma and truncated normal distributions. Other, even more flexible, distributions followed; Lee (1983) even proposed the four-parameter Pearson family of distributions. Nonetheless the two original single-parameter distributions remain the distributions of choice in the vast majority of empirical work.

It is a simple matter to change the sign of the inefficiency error component u and convert the stochastic production frontier model to a stochastic cost frontier model $E = c(y, w; \beta) \cdot \exp\{v + u\}$, where E is expenditure, $[c(y, w; \beta) \cdot \exp\{v\}]$ is a stochastic cost frontier, and u is intended to capture the cost of technical and allocative inefficiency. The JLMS technique may be used to provide an estimate of overall cost inefficiency, but the difficult remaining problem is to decompose the estimate of u into estimates of the separate costs of technical and allocative inefficiency. Schmidt and Lovell (1979) accomplished the decomposition for the Cobb–Douglas case. In a wonderful example of why researchers attend international conferences, Kopp and Diewert (1982) obtained the decomposition for the more general translog case, although econometric difficulties with their decomposition remain to this day.

Cross-sectional data provide a shapshot of producers and their efficiency. Panel data provide more reliable evidence on their performance, because they enable us to track the performance of each producer through a sequence of time periods. Long ago Hoch (1955, 1962) and Mundlak (1961) utilized panel data to purge agricultural production function parameter estimates of bias attributable to variation in what Hoch called technical efficiency and what Mundlak called management bias. Eventually Pitt and Lee (1981) extended cross-sectional maximum likelihood estimation techniques to panel data, and Schmidt and Sickles (1984) extended the pioneering work of Hoch and Mundlak by applying fixed-effects and random-effects methods to the efficiency measurement problem, where the effects are one-sided. The objective of these latter studies was not so much to eliminate bias from parameter estimates as to obtain producer-specific estimates of technical efficiency, or of the management effect. A significant advantage of (sufficiently long) panels is that they permit consistent estimation of the efficiency of individual producers, whereas the JLMS technique does not generate consistent estimators in a cross-sectional context.

Early panel data models were based on the assumption of time-invariant efficiency. The longer the panel, the less tenable this assumption becomes. Eventually this assumption was relaxed, in a series of papers by Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990), and Battese and Coelli (1992).

If efficiency varies, across producers or through time, it is natural to seek determinants of efficiency variation. Early studies adopted a two-stage approach, in which efficiencies are estimated in the first stage, and estimated efficiencies are regressed against a vector of explanatory variables in a second stage. More recent studies, including those of Kumbhakar, Ghosh, and McGuckin (1991), Reifschneider and Stevenson (1991), Huang and Liu (1994), and Battese and

Coelli (1995), have adopted a single-stage approach in which explanatory variables are incorporated directly into the inefficiency error component. In this approach either the mean or the variance of the inefficiency error component is hypothesized to be a function of the explanatory variables.

Abramovitz (1956) referred to productivity change, the residual between an index of the rates of growth of outputs and an index of the rates of growth of inputs, as a measure of our ignorance. Early studies of productivity change, such as Solow (1957), associated productivity change with technical change. As we became less ignorant, productivity change was decomposed into the magnitude and biases of technical change, and the effect of scale economies. However if productive efficiency changes through time, then it must also contribute to productivity change. Eventually Bauer (1990a) and others incorporated efficiency change into models of productivity change. Griliches (1996) provides an illuminating survey of research into "the residual," although the research surveyed makes only passing reference to the role of efficiency change.

1.3 THE ORGANIZATION OF THE BOOK

Chapter 2 is devoted to the analytical foundations of producer theory and efficiency measurement. In Section 2.2 we characterize production technology with production frontiers in the single-output case and with distance functions in the multiple-output case. We also characterize technology with dual cost, revenue, and profit frontiers, which provide increasingly exacting standards against which to measure producer performance. In Section 2.3 we define producer performance in terms of technical efficiency, and we measure technical efficiency with distance functions. In Section 2.4 we define producer performance in terms of economic (cost, revenue, and profit) efficiency, and we measure economic efficiency relative to cost, revenue, and profit frontiers. We also show how to decompose each type of economic efficiency into technical and allocative components.

Chapter 3 is concerned with the estimation of technical efficiency. In Section 3.2 we develop and show how to estimate cross-sectional production frontier models, both deterministic and stochastic, although most of our effort is directed toward stochastic production frontiers. In Section 3.3 we develop and show how to estimate panel data production frontier models, in which technical efficiency is initially time invariant and then is allowed to vary through time. In Section 3.4 we discuss the problem of heteroskedasticity in stochastic production frontier models.

Chapter 4 is concerned with the estimation and decomposition of cost efficiency. In Section 4.2 we develop and show how to estimate cross-sectional stochastic cost frontier models, in both single-equation and simultaneous-equation settings. In Section 4.3 we develop and show how to estimate panel data stochastic cost frontier models, again in single-equation and simultaneous-equation settings. In Section 4.4 we discuss a pair of novel approaches to the estimation of cost efficiency.

Chapter 5 is concerned with the estimation and decomposition of profit efficiency. In Section 5.2 we develop and show how to estimate single-output stochastic profit frontier models, using both primal and dual approaches. In Section 5.3 we develop and show how to estimate multiple-output stochastic profit frontier models, again using both primal and dual approaches. We pay little attention to the distinction between cross-sectional and panel data models, since the various estimation techniques developed in Chapters 3 and 4 apply equally well to the estimation of stochastic profit frontiers.

In Chapters 3–5 inefficiency is modeled by introducing additional error components and assigning distributions to them. Inefficiencies are then estimated as functions of the parameters of these distributions. In Chapter 6 we take a very different approach, in which both technical and allocative inefficiencies are modeled parametrically, on the assumption that producers optimize with respect to shadow prices, which are parametrically related to actual prices. In Section 6.2 we develop and show how to estimate and decompose both cost and profit efficiency in a cross-sectional setting. In Section 6.3 we do the same thing in a panel data setting. This parametric approach has some advantages, and some disadvantages, relative to the error component approach developed in Chapters 3–5.

Estimation of efficiency is the first of two tasks. The second task is to explain variation in estimated efficiency. In Chapter 7 we discuss alternative approaches to the explanation of variation in efficiency. In Section 7.2 we discuss some early approaches to explanation, and we find these approaches wanting. In Section 7.3 we discuss a variety of recent approaches to explanation, which we find superior to the early approaches. Essentially these recent approaches achieve explanation by making the one-sided inefficiency error component a function of the explanatory variables.

If efficiency varies, either across producers or through time, its variation constitutes a source of producer performance variation. One measure of performance is productivity change, and so if efficiency changes through time, it makes a contribution to productivity change. Chapter 8 concludes the book by incorporating efficiency change into models of productivity change, which heretofore have tended to neglect the contribution of efficiency change. In Section 8.2 we develop a primal approach, based on a stochastic production frontier, to the estimation and decomposition of productivity change. In Sections 8.3 and 8.4 we develop a pair of dual approaches, based on stochastic cost and profit frontiers, to the estimation and decomposition of productivity change.

At least four topics are missing from the book. First, we do not discuss the estimation and decomposition of revenue efficiency relative to stochastic revenue frontiers. This is because all of the techniques developed in Chapter 4 for the estimation and decomposition of cost efficiency relative to stochastic cost frontiers can readily be applied to the revenue efficiency problem. Variables and regularity conditions change, as discussed in Chapter 2, but nothing else of import changes.

Second, we do not explore the efficiency with which producers pursue nonconventional objectives. One prominent example is provided by Shephard's (1974) indirect production frontier, relative to which it is possible to estimate and decompose both cost-indirect output-oriented technical efficiency and revenue-indirect input-oriented technical efficiency. The former allows the measurement of the performance of producers seeking to maximize output (or revenue) subject to a conventional technology constraint and a budget constraint. The latter allows the measurement of the performance of producers seeking to minimize input use (or cost) subject to a conventional technology constraint and a revenue target. The two indirect models of producer behavior are analyzed in Färe, Grosskopf, and Lovell (1988, 1992) and Färe and Grosskopf (1994), and the estimation techniques developed in Chapters 3 and 4 can be adapted to the estimation of cost-indirect and revenue-indirect efficiency. Another example is provided by the literature on labor-managed firms pioneered by Ward's (1958) "Illyrian" firm and Domar's (1966) Soviet collective farm. The duality properties of a stylized labor-managed firm model have been worked out by Neary (1988) and Kahana (1989), and the econometric techniques discussed in Chapters 3–6 can be modified to estimate primal and dual efficiencies in this framework.

The third and fourth omissions are perhaps more serious. We do not discuss the Bayesian approach to stochastic frontier analysis, and we do not discuss semiparametric approaches to stochastic frontier analysis. Our reason for omitting these two topics is that the two literatures are small and not yet influential. However we refer interested readers to van den Broeck et al. (1994) and Osiewalski and Steel (1998) for good treatments of the Bayesian approach to SFA, and to Park and Simar (1994) and Park, Sickles, and Simar (1998) for good treatments of the semiparametric approach to SFA.