1

Efficiency and Productivity

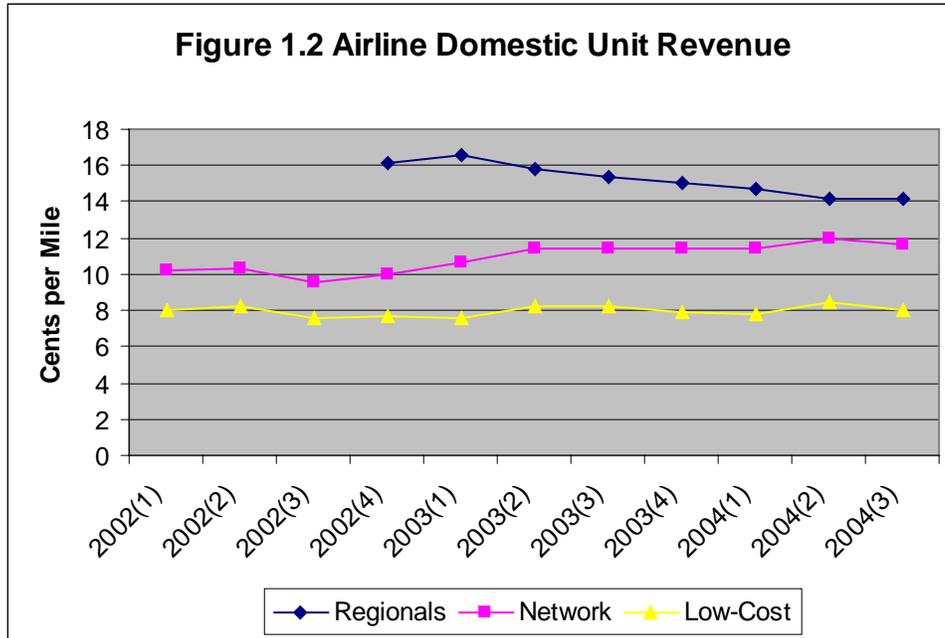*Harold O. Fried, C. A. Knox Lovell and Shelton S. Schmidt*

CONTENTS

1.1 Introduction

US airlines have encountered difficulties since September 11, 2001, particularly on domestic routes. Figure 1.1 plots quarterly operating profit margins (profit from domestic operations as a percent of operating revenue) for three segments of the industry, the small regional airlines, the medium-size low-cost airlines and the large network airlines. The regional airlines have performed relatively well, earning over a 10% profit margin, and the low-cost airlines have performed adequately, earning a considerably smaller but nonetheless positive profit margin. However the network airlines have performed poorly, earning a large negative profit margin. Some have sought bankruptcy protection.
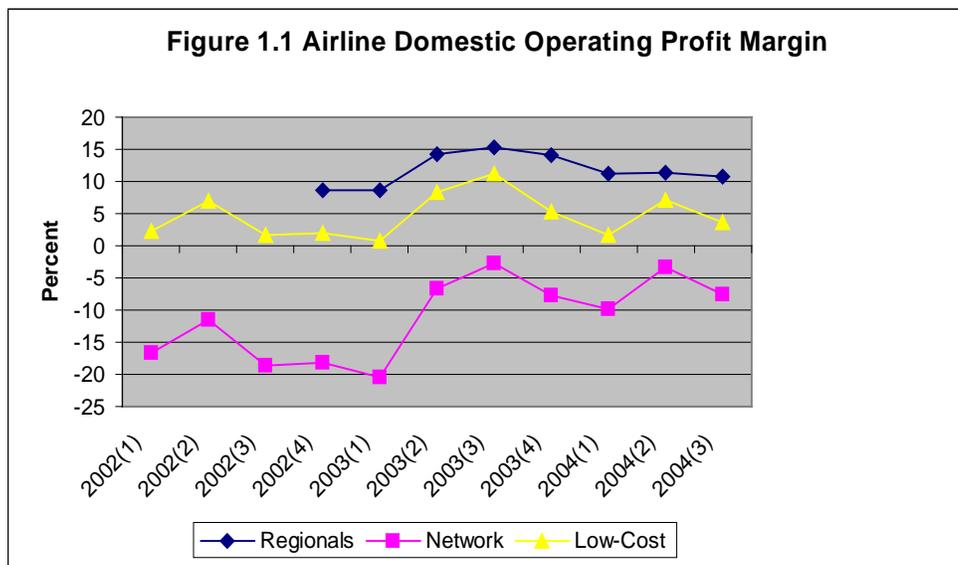
Source: http://www.bts.gov/press_releases/2004/bts035_04/html/bts035_04.html

When we ask why airline performance has varied so much, we naturally think of revenues and costs. Figure 1.2 plots quarterly operating revenue per available seat mile (a measure of average revenue), and Figure 1.3 plots quarterly operating expense per available seat mile (a measure of average cost). On the revenue side, the regional airlines earned the highest operating revenue per available seat mile, trailed in order by the network airlines and the low-cost airlines. On the cost side, the low-cost airlines incurred the lowest operating cost per available seat mile, appropriately enough, trailed by the network airlines and the regional airlines.

Figure 1.2 Airline Domestic Unit Revenue

Cents per Mile

[Chart showing unit revenue (Cents per Mile) from 2002(1) to 2004(3) for three airline segments: Regionals (blue, ~16 declining to ~14), Network (magenta, ~10 rising to ~12), and Low-Cost (yellow, ~8 roughly flat).]

Source: as Figure 1.1

It appears that the regional airlines have been the most profitable segment of the domestic airline industry despite having had the highest unit costs. The low-cost airlines have been marginally profitable because their low unit revenues have been offset by even lower unit costs. Finally, the network airlines have lost money primarily because of their high unit costs.

On the cost side, three hypotheses spring quickly to mind, each being inspired by conventional economic theory. First, the pattern of unit operating costs may reflect a pattern of scale economies that generates a U-shaped minimum average cost function favoring the medium-size low-cost airlines. Second, it may reflect higher input prices paid by the regional and network airlines. This hypothesis rings true for the older network airlines, which at the time were burdened by high labor costs attributable in large part to onerous pension obligations. Third, it may reflect different technologies embedded in a "low-cost business model" employed by the low-cost airlines and an inefficient "hub-and-spoke" system employed by the network airlines. Support for this hypothesis comes from the network airlines themselves, which predict efficiency gains and cost savings as they gradually abandon the system they adopted three decades ago.

Figure 1.3 Airline Domestic Unit Cost
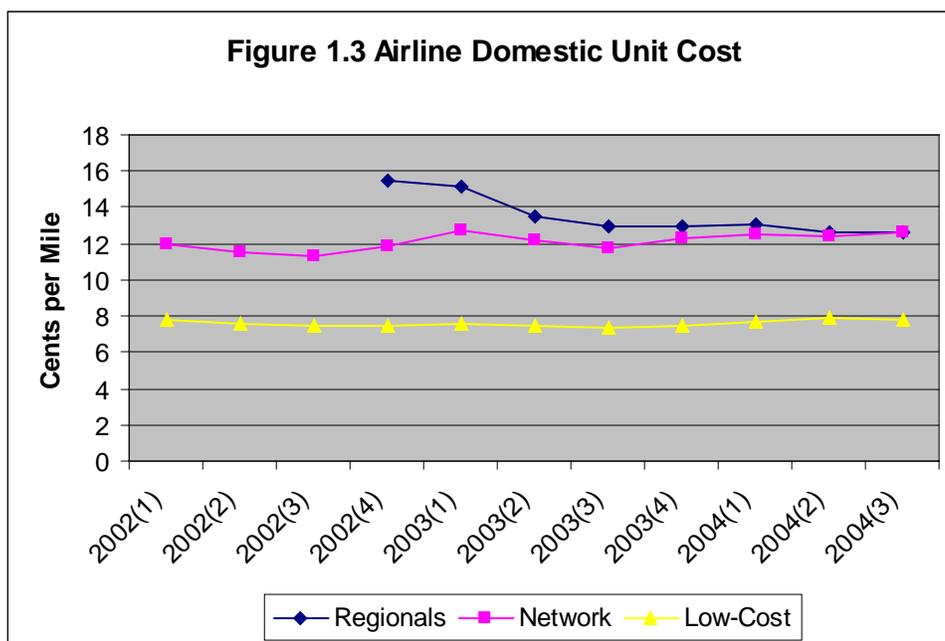
Source: as Figure 1.1

On the revenue side, differential pricing power is a possible explanation, although it is not clear why the small regional airlines would have such an advantage. A more likely explanation is variation in rates of capacity utilization as measured by load factors (the percent of available seats actually sold), which might have favored the regional airlines and penalized the low-cost airlines.

Each of these hypotheses is suggested by economic theory, and may or may not be refuted by the evidence. We now put forth an additional pair of refutable hypotheses that, although not suggested by conventional economic theory, should not be dismissed *a priori*.

One hypothesis concerns the cost side, and posits that part of the observed pattern of unit operating cost may be a consequence of cost inefficiency at the regional and network airlines. Cost inefficiency can be "technical," arising from excessive resource use given the amount of traffic, or "allocative," arising from resources being employed in the wrong mix, given their prices. Perhaps the low-cost airlines had relatively low unit costs because they utilized part-time labor and because they leased, rather than purchased, aircraft. Either strategy would reduce idleness and down time. More generally, perhaps the low-cost airlines had relatively low unit costs because their resources, human and physical, were well managed. This would place them on the minimum average cost function, whereas cost inefficiency at the regional and network airlines would place them above the minimum average cost function.

The second hypothesis concerns the revenue side, and posits that part of the observed pattern of unit operating revenue may be a consequence of revenue inefficiency at the network and low-cost airlines. Revenue inefficiency can be "technical," arising from a failure to provide maximum service from the available resources, or "allocative," arising from the provision of services in the wrong mix, given their prices. Perhaps the regional airlines were nimble enough to adjust their route structures to respond quickly to fluctuations in passenger demand. Perhaps the regional airlines have faster gate turnaround times than the network airlines, whose hub-and-spoke technology leaves aircraft and crew idle and sacrifices revenue. This would place the regional airlines on the maximum average revenue function, whereas revenue inefficiency at the network and low-cost airlines would place them beneath the maximum average revenue function.

The point of the foregoing discussion is not to engage in a deep exploration into airline economics, about which we are blissfully ignorant. We are merely frequent fliers who happen to be curious economists wondering what might explain the observed variation in the recent domestic performance of US airlines. The point is to suggest that variation in productive efficiency, in both the management of resources and the management of services, may be a potentially significant source of variation in financial performance. Inefficient behavior is assumed away in conventional economic theory, in which first-order and second-order optimizing conditions are satisfied. But it exists in the real world, as a perusal of almost any trade publication will verify, and as the hordes of consultants armed with their buzzwords will testify.

Productive inefficiency exists, and it deserves to be included in our analytical toolkit because it can generate refutable hypotheses concerning the sources of variation in business performance. This book is devoted to the study of inefficiency in production and its impact on economic and financial performance. The study ranges from the underlying theory to the analytical foundations, and then to the quantitative techniques and the empirical evidence.

Chapter 1 sets the stage. Section 1.2 provides background material, and focuses on hypotheses that have been proposed in the literature that would explain variation in producer performance. This Section also provides a glimpse at the empirical literature, and demonstrates that the search for variation in producer performance has been conducted in a wide variety of settings. Section 1.3 lays the theoretical foundation for the measurement of productive efficiency. It provides definitions of alternative notions of productive efficiency, and it provides corresponding measures of efficiency. Section 1.4 offers a brief introduction to alternative techniques that have been developed to quantify inefficiency empirically. Section 1.5 introduces various econometric approaches to efficiency estimation, while Section 1.6 introduces variants of the mathematical programming approach to efficiency estimation. Section 1.7 introduces the Malmquist productivity index, and shows how to decompose it into various sources of productivity change, including variation in productive efficiency.

Section 1.8 describes three ways of approximating a Malmquist productivity index: the use of superlative index numbers, the use of econometric techniques and the use of mathematical programming techniques. Section 1.9 offers some concluding observations.

Chapter 2 extends Section 1.5 by providing a detailed survey of the econometric approach to efficiency estimation. Chapter 3 extends Section 1.6 by providing a detailed survey of the mathematical programming approach to efficiency estimation. Chapter 4 recasts the parametric and statistical approach of Chapter 2, and the nonparametric and deterministic approach of Chapter 3, into a nonparametric and statistical approach. Chapter 5 extends Sections 1.7 and 1.8 by discussing alternative approaches to the measurement of productivity change, with special emphasis on efficiency change as a source of productivity change.

1.2 Background

When discussing the economic performance of producers, it is common to describe them as being more or less "efficient," or more or less "productive." In this Section we discuss the relationship between these two concepts. We consider some hypotheses concerning the *determinants* of producer performance, and we consider some hypotheses concerning the financial *consequences* of producer performance.

By the *productivity* of a producer we mean the ratio of its output to its input. This ratio is easy to calculate if the producer uses a single input to produce a single output. In the more likely event that the producer uses several inputs to produce several outputs, the outputs in the numerator must be aggregated in some economically sensible fashion, as must the inputs in the denominator, so that productivity remains the ratio of two scalars. Productivity growth then becomes the difference between output growth and input growth, and the aggregation requirement applies here as well.

Variation in productivity, either across producers or through time, is thus a residual, which Abramovitz (1956) famously characterized as "a measure of our ignorance." Beginning perhaps with Solow (1957), much effort has been devoted to dispelling our ignorance by "whittling away at the residual" (Stone (1980)). Much of the whittling has involved minimizing measurement error in the construction of output and input quantity indexes. The conversion of raw data into variables consistent with economic theory is a complex undertaking. Griliches (1996) surveys the economic history of the residual, and state-of-the-art procedures for whittling away at it are outlined in OECD (2001). When the whittling is finished, we have a residual suitable for analysis.

In principle the residual can be attributed to differences in production technology, differences in the scale of operation, differences in operating efficiency, and differences in the operating environment in which production occurs. The US Department of Labor's Bureau of Labor Statistics (2005) and the OECD (2001) attribute variation in productivity through time to these same sources. Proper attribution is important for the adoption of private managerial practices and the design of public policies intended to improve productivity performance. We are naturally interested in isolating the first three components, which are under the control of management, from the fourth, which is not. Among the three endogenous components our interest centers on the efficiency component, and on measuring both its cross-sectional contribution to variation in productivity and its inter-temporal contribution to productivity change.

By the *efficiency* of a producer we have in mind a comparison between observed and optimal values of its output and input. The exercise can involve comparing observed output to maximum potential output obtainable from the input, or comparing observed input to minimum potential input required to produce the output, or some combination of the two. In these two comparisons the optimum is defined in terms of production possibilities, and efficiency is technical. It is also possible to define the optimum in terms of the behavioral goal of the producer. In this event efficiency is measured by comparing observed and optimum cost, revenue, profit, or whatever goal the producer is assumed to pursue, subject, of course, to any appropriate constraints on quantities and prices. In these comparisons the optimum is expressed in value terms, and efficiency is economic.

Even at this early stage three problems arise, and much of this Section is devoted to exploring ways each has been addressed. First, which outputs and inputs are to be included in the comparison? Second, how are multiple outputs and multiple inputs to be weighted in the comparison? And third, how is the technical or economic potential of the producer to be determined?

Many years ago Knight (1933) addressed the first question by noting that if all outputs and all inputs are included, then since neither matter nor energy can be created or destroyed, all producers would achieve the same unitary productivity evaluation. In this circumstance Knight proposed to redefine productivity as the ratio of useful output to input. Extending Knight's redefinition to the ratio of useful output to useful input, and representing usefulness with weights incorporating market prices, generates a modern economic productivity index. As a practical matter, however, the first problem is not how to proceed when all outputs and all inputs are included, but rather how to proceed when not enough outputs and inputs are included.

As Stigler (1976) has observed, measured inefficiency may be a reflection of the analyst's failure to incorporate all relevant variables and, complicating the first problem, to specify the right economic objectives and the right constraints.

Stigler was criticizing the work of Leibenstein (1966, 1976), who focused on inadequate motivation, information asymmetries, incomplete contracts, agency problems and the attendant monitoring difficulties within the firm, and who lumped all these features together and called the mix "X-inefficiency." When the agents' actions are not aligned with the principal's objective, potential output is sacrificed. Thus what appears as inefficiency to Leibenstein is evidence of an incomplete model to Stigler, who called it waste and concluded that "…waste is not a useful economic concept. Waste is error within the framework of modern economic analysis…" (p. 216) The practical significance of this exchange is that if Stigler's wish is not granted, and not all variables reflecting the objectives and constraints of the principal and the agents are incorporated into the model, agency and related problems become potential sources of measured (if not actual) inefficiency.

Leibenstein was not writing in a vacuum. His approach fits nicely into the agency literature. The recognition of agency problems goes back at least as far as the pioneering Berle and Means (1932) study of the consequences of the separation of ownership from control, in which owners are the principals and managers are the agents. Leibenstein's notion of X-inefficiency also has much in common with Simon's (1955) belief that in a world of limited information processing ability, managers exhibit "bounded rationality" and engage in "satisficing" behavior. Along similar lines, Williamson (1975, 1985) viewed firms as seeking to economize on transaction costs, which in his view boiled down to economizing on bounded rationality. Bounded rationality and the costs of transacting also become potential sources of measured inefficiency.

It would be desirable, if extraordinarily difficult, to construct and implement Stigler's complete model involving all the complexities mentioned above. We have not seen such a model. What we have seen are simplified (if not simple) models of the firm in which measured performance differentials presumably reflect variation in the ability to deal with the complexities of the real world. Indeed performance measures based on simplified models of the firm are often useful, and sometimes necessary. They are useful when the objectives of producers, or the constraints facing them, are either unknown or unconventional or subject to debate. In this case a popular research strategy has been to model producers as unconstrained optimizers of some conventional objective, and to test the hypothesis that inefficiency in this environment is consistent with efficiency in the constrained environment. The use of such incomplete measures has proved necessary in a number of contexts for lack of relevant data. One example of considerable policy import occurs when the production of desirable (and measured and priced) outputs is recorded, but the generation of undesirable (and frequently unmeasured and more frequently unpriced) byproducts is not. Another occurs when the use of public infrastructure enhances private performance, but its use goes unrecorded. In each case the measure of efficiency or productivity that is obtained may be very different from the measure one would like to have.

Even when all relevant outputs and inputs are included, there remains the formidable second problem of assigning weights to variables. Market prices provide a natural set of weights, but two types of question arise. First, suppose market prices exist. If market prices change through time, or vary across producers, is it possible to disentangle the effects of price changes and quantity changes in a relative performance evaluation? Alternatively, if market prices reflect monopoly or monopsony power, or cross-subsidy, or the determination of a regulator, do they still provide appropriate weights in a relative performance evaluation? Second, suppose some market prices do not exist. In the cases of environmental impacts and public infrastructure mentioned above, the unpriced variables are externalities either generated by or received by market sector producers. How do we value these externalities? However the weighting problem is more pervasive than the case of externalities. The non-market sector is growing relative to the market sector in most advanced economies, and by definition the outputs in this sector are not sold on markets. How then do we value outputs such as law enforcement and fire protection services, or even public education services, each of which is publicly funded rather than privately purchased? Is it possible to develop proxies for missing prices that would provide appropriate weights in a performance evaluation? The presence of distorted or missing prices complicates the problem of determining what is meant by "relevant."

The third problem makes the first two seem easy. It is as difficult for the analyst to determine a producer's potential as it is for the producer to achieve that potential. It is perhaps for this reason that for many years the productivity literature ignored the efficiency component identified by the BLS and the OECD. Only recently, with the development of a separate literature devoted to the study of efficiency in production, has the problem of determining productive potential been seriously addressed. Resolution of this problem makes it possible to integrate the two literatures. Integration is important for policy purposes, since action taken to enhance productivity performance requires an accurate attribution of observed performance to its components.

By way of analogy, we do not know, and cannot know, how fast a human can run 100 meters. But we do observe best practice and its improvement through time, and we do observe variation in actual performance among runners. The world of sport is full of statistics, and we have all-star teams whose members are judged to be the best at what they do. Away from the world of sport, we use multiple criteria to rank cities on the basis of quality of life indicators (Zurich and Geneva are at the top). At the macro level we use multiple criteria to rank countries on the basis of economic freedom (Norway, Sweden and Australia are at the top), environmental sustainability (Finland and Norway are at the top), business risk (Iraq and Zimbabwe pose the most risk) and corruption (Finland and New Zealand are the least corrupt), among many others. The United Nation's Human Development Index is perhaps the best-known and most widely

studied macroeconomic performance indicator (Norway and Sweden are at the top). In each of these cases we face the three problems mentioned at the outset of this section: what indicators to include, how to weight them, and how to define potential. The selection and weighting of indicators are controversial by our standards, although comparisons are appropriately made relative to best practice rather than to some ideal standard.

The same reasoning applies to the evaluation of business performance. We cannot know "true" potential, whatever the economic objective. But we do observe best practice and its change through time, and we also observe variation in performance among producers operating beneath best practice. This leads to the association of "efficient" performance with undominated performance, or operation on a best practice "frontier," and of inefficient performance with dominated performance, or operation on the wrong side of a best practice frontier. Interest naturally focuses on the identification of best practice producers, and of benchmarking the performance of the rest against that of the best. Businesses themselves routinely benchmark their performance against that of their peers, and academic interest in benchmarking is widespread, although potential synergies between the approaches adopted by the two communities have yet to be fully exploited. Davies and Kochhar (2002) offer an interesting academic critique of business benchmarking.

Why the interest in measuring efficiency and productivity? We can think of three reasons. First, only by measuring efficiency and productivity, and by separating their effects from those of the operating environment so as to level the playing field, can we explore hypotheses concerning the sources of efficiency or productivity differentials. Identification and separation of controllable and uncontrollable sources of performance variation is essential to the institution of private practices and public policies designed to improve performance. Zeitsch *et al.* (1994) provide an empirical application showing how important it is to disentangle variation in the operating environment (in this case customer density) from variation in controllable sources of productivity growth in Australian electricity distribution.

Second, macro performance depends on micro performance, and so the same reasoning applies to the study of the growth of nations. Lewis (2004) provides a compelling summary of McKinsey Global Institute (MGI) productivity studies of 13 nations over 12 years, the main findings being that micro performance drives macro performance, and that a host of institutional impediments to strong micro performance can be identified. This book, and the studies on which it is based, make it clear that there are potential synergies, as yet sadly unexploited, between the MGI approach and the academic approach to performance evaluation.

Third, efficiency and productivity measures are success indicators, performance metrics, by which producers are evaluated. However for most

producers the ultimate success indicator is financial performance, and the ultimate metric is the bottom line. Miller's (1984) clever title, "Profitability = Productivity + Price Recovery," encapsulates the relationship between productivity and financial performance. It follows that productivity growth leads to improved financial performance, provided it is not offset by declining price recovery attributable to falling product prices and/or rising input prices. Grifell-Tatjé and Lovell (1999) examine the relationship for Spanish banks facing increasing competition as a consequence of European monetary union. Salerian (2003) explores the relationship for Australian railroads, for which increasing intermodal competition has contributed to declining price recovery that has swamped the financial benefits of impressive productivity gains. This study also demonstrates that, although the bottom line may be paramount in the private sector, it is not irrelevant in the public sector; indeed many governments monitor the financial performance as well as the non-financial performance of their public service providers.

Many other studies, primarily in the business literature, adopt alternative notions of financial performance, such as return on assets or return on equity. These studies typically begin with the "DuPont triangle," which decomposes return on assets as $\pi/A = (\pi/R)(R/A) = $ (return on sales)(investment turnover), where π = profit, A = assets and R = revenue. The next step is to decompose the first leg of the DuPont triangle as $(\pi/R) = [(R-C)/R] = [1 - (R/C)^{-1}]$, where C is cost and R/C is profitability. The final step is to decompose profitability into productivity and price recovery, a multiplicative alternative to Miller's additive relationship. The objective is to trace the contribution of productivity change up the triangle to change in financial performance. Horrigan (1968) provides a short history of the DuPont triangle as an integral part of financial ratio analysis, and Eilon (1984) offers an accessible survey of alternative decomposition strategies. Banker *et al.* (1993) illustrate the decomposition technique with an application to the US telecommunications industry, in which deregulation led to productivity gains that were offset by deteriorating price recovery brought on by increased competition.

In some cases measurement enables us to quantify performance differentials that are predicted qualitatively by economic theory. An example is provided by the effect of market structure on performance. There is a common belief that productive efficiency is a survival condition in a competitive environment, and that its importance diminishes as competitive pressure subsides. Hicks (1935) gave eloquent expression to this belief by asserting that producers possessing market power "…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." (p. 8) Berger and Hannan (1998) provide a test of the quiet life hypothesis in US banking, and find evidence that banks in relatively concentrated markets exhibit relatively low cost efficiency.

Continuing the line of reasoning that firms with market power might not be "pure" profit maximizers, Alchian and Kessel (1962) replaced the narrow profit maximization hypothesis with a broader utility maximization hypothesis, in which case monopolists and competitors might be expected to be equally proficient in the pursuit of utility. The ostensible efficiency differential is then explained by the selection of more (observed) profit by the competitor and more (unobserved) leisure by the monopolist, which of course recalls the analyst's problem of determining the relevant outputs and inputs of the production process. Alchian and Kessel offer an alternative explanation for the apparent superior performance of competitive producers. This is that monopolies are either regulated, and thereby constrained in their pursuit of efficiency, or unregulated but threatened by regulation (or by antitrust action) and consequently similarly constrained. If these producers are capable of earning more than the regulated profit, and if their property rights to the profit are attenuated by the regulatory or antitrust environment, then inefficiency becomes a free good to producers subject to, or threatened by, regulation or antitrust action. As Alchian and Kessel put it, "[t]he cardinal sin of a monopolist…is to be too profitable." (p. 166)

Baumol (1959), Gordon (1961) and Williamson (1964) argued along similar lines. An operating environment characterized by market power and separation of ownership from control leaves room for "managerial discretion." Given the freedom to choose, managers would seek to maximize a utility function in which profit was either one of several arguments or, more likely, a constraint on the pursuit of alternative objectives. This idea, and variants of it, recurs frequently in the agency literature.

Thus competition is expected to enhance performance either because it forces producers to concentrate on "observable" profit-generating activities at the expense of Hicks' quiet life, or because it frees producers from the actual or potential constraints imposed by the regulatory and antitrust processes. One interesting illustration of the market structure hypothesis is the measurement of the impact of international trade barriers on domestic industrial performance. Many years ago Carlsson (1972) used primitive frontier techniques to uncover a statistically significant inverse relationship between the performance of Swedish industries and various measures of their protection from international competition. More recently Tybout and Westbrook (1995), Pavcnik (2002) and Schor (2004) have applied modern frontier techniques to longitudinal micro data in an effort to shed light on the linkage between openness and productivity in Mexico, Chile and Brazil. Specific findings vary, but a general theme emerges. Trade liberalization brings aggregate productivity gains attributable among other factors to improvements in productivity among continuing firms, and to entry of relatively productive firms and exit of relatively unproductive firms.

A second situation in which measurement enables the quantification of efficiency or productivity differentials predicted fairly consistently by theory is in the area of economic regulation. The most commonly cited example is rate of

return regulation, to which many utilities have been subjected for many years, and for which there exists a familiar and tractable analytical paradigm developed by Averch and Johnson (1962). Access to a tractable model and to data supplied by regulatory agencies has spawned numerous empirical studies, virtually all of which have found rate of return regulation to have led to over-capitalization that has had an adverse impact on utility performance and therefore on consumer prices. These findings have motivated a movement toward incentive regulation in which utilities are reimbursed on the basis of a price cap or revenue cap formula RPI - X, with X being a productivity (or efficiency) offset to movements in an appropriate price index RPI. The reimbursement formula allows utilities to pass along any cost increases incorporated in RPI, less any expected performance improvements embodied in the offset X. Since X is a performance indicator, this trend has spawned a huge theoretical and empirical literature using efficiency and productivity measurement techniques to benchmark the performance of regulated utilities. Bogetoft (2000 and references cited therein) has developed the theory within a frontier context, in which X can be interpreted as the outcome of a game played between a principal (the regulator) and multiple agents (the utilities). Netherlands Bureau for Economic Policy Analysis (2000) provides a detailed exposition of the techniques. Kinnunen (2005) reports either declining or stable trends in customer electricity prices in Finland, Norway and Sweden, where variants of incentive regulation have been in place for some time. Since enormous amounts of money are involved, the specification and weighting of relevant variables and the sample selection criteria become important, and frequently contentious, issues in regulatory proceedings.

Another regulatory context in which theoretical predictions have been quantified by empirical investigation is the impact of environmental controls on producer performance. In this context, however, the private cost of reduced efficiency or productivity must be balanced against the social benefits of environmental protection. Of course the standard paradigm that hypothesizes private costs of environmental constraints may be wrong; Porter (1991) has argued that well-designed environmental regulations can stimulate innovation, enhance productivity, and thus be privately profitable. Ambec and Barla (2002) develop a theory that predicts the Porter hypothesis. In any event, the problem of specifying and measuring the relevant variables crops up once again. Färe *et al.* (1989, 1993) have developed the theory within a frontier context. Reinhard *et al.* (1999) examined a panel of Dutch dairy farms that generate surplus manure, the nitrogen content of which contaminates groundwater and surface water and contributes to acid rain. They calculated a mean shadow price of the nitrogen surplus of just over NLG3 per kilogram, slightly higher than a politically constrained levy actually imposed of NLG1.5 per kilogram of surplus. Ball *et al.* (2004) calculated exclusive and inclusive productivity indexes for US agriculture, in which pesticide use causes water pollution. They found that inclusive productivity growth initially lagged behind exclusive productivity growth. However when the Environmental Protection Agency began regulating the manufacture of pesticides, inclusive productivity growth caught up with, and eventually

surpassed, exclusive productivity growth, as would be expected. Consistent with these findings, they found an inverted U shaped pattern of shadow prices, reflecting a period of lax regulation followed by tightened regulation that eventually led to the discovery and use of relatively benign and more effective pesticides.

A third situation in which measurement can quantify theoretical propositions is the effect of ownership on performance. Alchian (1965) noted that the inability of public sector owners to influence performance by trading shares in public sector producers means that public sector managers worry less about bearing the costs of their decisions than do their private sector counterparts. Hence they are contractually constrained in their decision-making latitude, given less freedom to choose, so to speak. "Because of these extra constraints - or because of the 'costs' of them - the public arrangement becomes a higher cost (in the sense of 'less efficient') than that for private property agencies." (p. 828) A literature has developed based on the supposition that public managers have greater freedom to pursue their own objectives, at the expense of conventional objectives. Niskanen (1971) viewed public managers as budget maximizers, de Alessi (1974) viewed public managers as preferring capital-intensive budgets, and Lindsay (1976) viewed public managers as preferring "visible" variables. Each of these hypotheses suggests that measured performance is lower in the public sector than in the private sector. Holmstrom and Tirole (1989) survey much of the theoretical literature, as does Hansmann (1988), who introduces private not-for-profit producers as a third category. Empirical tests of the public/private performance differential hypothesis are numerous. Many of the comparisons have been conducted using regulated utility data, because public and private firms frequently compete in these industries, because of the global trend toward privatization of public utilities, and because regulatory agencies collect and provide data. Jamash and Pollitt (2001) survey the empirical evidence for electricity distribution. Education and health care are two additional areas in which numerous public/private performance comparisons have been conducted.

In any public/private performance comparison one confronts the problem of how to measure their performance. Pestieau and Tulkens (1993) offer a spirited defense of a narrow focus on technical efficiency, so as to level the playing field. They argue that public enterprises have objectives and constraints (e.g., fiscal balance and universal service, uniform price requirements, but at the same time a soft budget constraint) different from those of private enterprises, and the only common ground on which to compare their performance is on the basis of their technical efficiency.

In some cases theory gives no guidance, or provides conflicting signals, concerning the impact on performance of some phenomenon. In such cases empirical measurement provides qualitative as well as quantitative evidence. Four examples illustrate the point. Are profit maximizing firms more efficient than cooperatives? Is one form of sharecropping more efficient than another? Is

slavery an efficient way of organizing production? Is organized crime efficiently organized? The answer to each question seems to be "it depends," and so empirical measurement is called for. Theory and evidence are offered by Pencavel (2001) for cooperatives, by Otsuka *et al.* (1992) and Garrett and Xu (2003) for sharecropping, by Fogel and Engerman (1974) for slavery, and by Fiorentini and Peltzman (1995) for organized crime.

Finally, the ability to quantify efficiency and productivity provides management with a control mechanism with which to monitor the performance of production units under its control. The economics, management science and operations research literatures contain numerous examples of the use of efficiency and productivity measurement techniques for this and related purposes. However interest in these techniques has spread far beyond their origins, as evidenced by the empirical applications referenced in Table 1.1. The recent dates of these studies and the journals in which they appear demonstrate that the techniques are currently in use in fields far removed from their origins. In each of these applications interesting and challenging issues concerning appropriate behavioral objectives and constraints, and the specification of relevant variables and their measurement, arise. These applications also illustrate the rich variety of analytical techniques that can be used in making efficiency and productivity comparisons. It is worth pondering how each of these examples deals with the long list of problems discussed in this Section.

Table 1.1 Empirical Applications of Efficiency and Productivity Analysis	
Accounting, advertising, auditing, insurance and law firms	Banker *et al.* (2005) Luo and Donthu (2005) Dopuch *et al.* (2003) Cummins *et al.* (2005) Wang (2000)
Airports	Oum and Yu (2004) Sarkis and Talluri (2004) Yoshida and Fujimoto (2004) Yu (2004)
Air transport	Coelli *et al.* (2002) Sickles *et al.* (2002) Scheraga (2004) Duke and Torres (2005)
Bank branches	Davis and Albright (2004) Camanho and Dyson (2005) Porembski *et al.* (2005) Silva Portela and Thanassoulis (2005)
Bankruptcy prediction	Wheelock and Wilson (2000) Becchetti and Sierra (2003) Cielen *et al.* (2004)

Benefit – cost analysis	Goldar and Misra (2001)
	Hofler and List (2004)
Community and rural health care	Birman *et al.* (2003)
	Dervaux *et al.* (2003)
	Jiménez *et al.* (2003)
	Kirigia *et al.* (2004)
Correctional facilities	Gyimah – Brempong (2000)
	Nyhan (2002)
Credit risk evaluation	Emel *et al.* (2003)
	Paradi *et al.* (2004)
Dentistry	Buck (2000)
	Grytten and Rongen (2000)
	Linna *et al.* (2003)
	Widstrom *et al.* (2004)
Discrimination	Croppenstedt and Meschi (2000)
	Bowlin *et al.* (2003)
	Mohan and Ruggiero (2003)
Education: primary and secondary	Dolton *et al.* (2003)
	Mayston (2003)
	Ammar *et al.* (2004)
	Dodson and Garrett (2004)
Education: tertiary	Bonaccorsi and Daraio (2003)
	Mensah and Werner (2003)
	Guan and Wang (2004)
	Warning (2004)
Elections	Obata and Ishii (2003)
	Foroughi *et al.* (2005)
Electricity distribution	Ajodhia and Petrov (2004)
	Filippini *et al.* (2004)
	Agrell *et al.* (2005)
	Edvardsen *et al.* (2006)
Electricity generation	Arocena and Waddams Price (2003)
	Korhonen and Luptacik (2004)
	Atkinson and Halabi (2005)
	Cook and Green (2005)
Environment: macro applications	Jeon and Sickles (2004)
	Lindmark (2004)
	Zaim (2004)
	Henderson and Millimet (2005)
Environment: micro applications	Linton (2002)
	Yang *et al.* (2003)
	Koundouri and Xepapadeas (2004)
	Banzhaf (2005)
Financial statement analysis	Chen and Zhu (2003)
	Feroz *et al.* (2003)
Fishing	Chiang *et al.* (2004)

	Herrero (2004)
	Kompas *et al.* (2004)
	Martinez-Cordero and Leung (2004)
Forestry	Otsuki *et al.* (2002)
	Bi (2004)
	Hof *et al.* (2004)
	Liu and Yin (2004)
Gas distribution	Rossi (2001)
	Carrington *et al.* (2002)
	Hammond *et al.* (2002)
	Hawdon (2003)
Hospitals	Chang *et al.* (2004)
	Stanford (2004)
	Ventura *et al.* (2004)
	Gao *et al.* (2006)
Hotels	Hwang and Chang (2003)
	Chiang *et al.* (2004)
	Barros (2005)
	Sigala *et al.* (2005)
Internet commerce	Wen *et al.* (2003)
	Barua *et al.* (2004)
	Chen *et al.* (2004)
	Serrano-Sinca *et al.* (2005)
Labor markets	Sheldon (2003)
	Diaz-Mayans and Sanchez (2004)
	Ibourk *et al.* (2004)
	Millimet (2005)
Libraries	Hammond (2002)
	Shim (2003)
	Kao and Lin (2004)
	Reichmann (2004)
Location	Thomas *et al.* (2002)
	Cook and Green (2003)
	Takamura and Tone (2003)
Macroeconomics	Cherchye *et al.* (2004)
	Grafton *et al.* (2004)
	Despotis (2005)
	Ravallion (2005)
Mergers	Cuesta and Orea (2002)
	Ferrier and Valdmanis (2004)
	Bogetoft and Wang (2005)
	Sherman and Rupert (2006)
Military	Barros (2002)
	Bowlin (2004)
	Brockett *et al.* (2004)
	Sun (2004)

Municipal services	Hughes and Edwards (2000) Moore *et al.* (2001) Prieto and Zofio (2001) Southwick (2005)
Museums	Mairesse and Vanden Eeckaut (2002) Bishop and Brand (2003) Basso and Funari (2004)
Nursing homes	Fried *et al.* (2002) Farsi and Filippini (2004) Hougaard *et al.* (2004) Laine *et al.* (2005)
Physicians and physician practices	Wagner *et al.* (2003) Rosenman and Friesner (2004)
Police	Spottiswoode (2000) Wisniewski and Dickson (2001) Stone (2002) Drake and Simper (2004)
Ports	Sánchez *et al.* (2003) Clark *et al.* (2004) Lawrence and Richards (2004) Turner *et al.* (2004)
Postal services	Pimenta *et al.* (2000) Maruyama and Nakajima (2002) Borenstein *et al.* (2004)
Public infrastructure	Mamatzakis (2003) Martín *et al.* (2004) Paul *et al.* (2004) Salinas-Jiminez (2004)
Rail transport	Baños-Pino *et al.* (2002) Estache *et al.* (2002) Kennedy and Smith (2004) Loizides and Tsionas (2004)
Real estate investment trusts	Lewis *et al.* (2003) Anderson *et al.* (2004)
Refuse collection and recycling	Bosch *et al.* (2000) Worthington and Dollery (2001) Lozano *et al.* (2004)
Sports	Haas (2003) Lins *et al.* (2003) Fried *et al.* (2004) Amos *et al.* (2005)
Stocks, mutual funds and hedge funds	Basso and Funari (2003) Abad *et al.* (2004) Chang (2004) Troutt *et al.* (2005)
Tax administration	Serra (2003)

Telecommunications	Guedes de Avellar *et al.* (2002)
	Pentzaroupoulos and Giokas (2002)
	Façanha and Resende (2004)
	Uri (2004)
Urban transit	De Borger *et al.* (2002)
	Dalen and Gómez-Lobo (2003)
	Jörss *et al.* (2004)
	Odeck and Alkadi (2004)
Water distribution	Thanassoulis (2000)
	Fu and Huang (2002)
	Corton (2003)
	Tupper and Resende (2004)
World Health Organization	Hollingsworth and Wildman (2003)
	Richardson *et al.* (2003)
	Greene (2004)
	Lauer *et al.* (2004)

1.3 Definitions and Measures of Economic Efficiency

Economic efficiency has technical and allocative components. The technical component refers to the ability to avoid waste, either by producing as much output as technology and input usage allow or by using as little input as required by technology and output production. Thus the analysis of technical efficiency can have an output augmenting orientation or an input conserving orientation. The allocative component refers to the ability to combine inputs and/or outputs in optimal proportions in light of prevailing prices. Optimal proportions satisfy the first-order conditions for the optimization problem assigned to the production unit.

Koopmans (1951) provided a formal *definition* of technical efficiency: a producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input, and if a reduction in any input requires an increase in at least one other input or a reduction in at least one output. Thus a technically inefficient producer could produce the same outputs with less of at least one input, or could use the same inputs to produce more of at least one output.

Debreu (1951) and Farrell (1957) introduced a *measure* of technical efficiency. With an input conserving orientation their measure is defined as (one minus) the maximum equiproportionate (i.e., radial) reduction in all inputs that is feasible with given technology and outputs. With an output augmenting orientation their measure is defined as the maximum radial expansion in all outputs that is feasible with given technology and inputs. In both orientations a value of unity indicates technical efficiency because no radial adjustment is

feasible, and a value different from unity indicates the severity of technical inefficiency.

In order to relate the Debreu-Farrell measures to the Koopmans definition, and to relate both to the structure of production technology, it is useful to introduce some notation and terminology. Let producers use inputs $x = (x_1,\ldots,x_N) \in R^N_+$ to produce outputs $y = (y_1,\ldots,y_M) \in R^M_+$. Production technology can be represented by the production set

$$T = \{(y,x): x \text{ can produce } y\}. \tag{1.1}$$

Koopmans' definition of technical efficiency can now be stated formally as $(y,x) \in T$ is technically efficient if, and only if, $(y',x') \notin T$ for $(y',-x') \geq (y,-x)$.

Technology can also be represented by input sets

$$L(y) = \{x: (y,x) \in T\}, \tag{1.2}$$

which for every $y \in R^M_+$ have input isoquants

$$I(y) = \{x: x \in L(y), \lambda x \notin L(y), \lambda < 1\} \tag{1.3}$$

and input efficient subsets

$$E(y) = \{x: x \in L(y), x' \notin L(y), x' \leq x\}, \tag{1.4}$$

and the three sets satisfy $E(y) \subseteq I(y) \subseteq L(y)$.

Shephard (1953) introduced the input distance function to provide a functional representation of production technology. The input distance function is

$$D_I(y,x) = \max \{\lambda: (x/\lambda) \in L(y)\}. \tag{1.5}$$

For $x \in L(y)$, $D_I(y,x) \geqq 1$, and for $x \in I(y)$, $D_I(y,x) = 1$. Given standard assumptions on T, the input distance function $D_I(y,x)$ is nonincreasing in y, and nondecreasing, homogeneous of degree +1 and concave in x.

The Debreu-Farrell input-oriented measure of technical efficiency can now be given a somewhat more formal interpretation as the value of the function

$$TE_I(y,x) = \min \{\theta: \theta x \in L(y)\}, \tag{1.6}$$

and it follows from (1.5) that

$$TE_I(y,x) = 1/D_I(y,x). \tag{1.7}$$

For $x \in L(y)$, $TE_I(y,x) \leqq 1$, and for $x \in I(y)$, $TE_I(y,x) = 1$.

Since so much of efficiency measurement is oriented toward output augmentation, it is useful to replicate the above development in that direction. Production technology can be represented by output sets

$$P(x) = \{y: (x,y) \in T\}, \tag{1.8}$$

which for every $x \in R^N_+$ have output isoquants

$$I(x) = \{y: y \in P(x), \lambda y \notin P(x), \lambda > 1\} \tag{1.9}$$

and output efficient subsets

$$E(x) = \{y: y \in P(x), y' \notin P(x), y' \geq y\}, \tag{1.10}$$

and the three sets satisfy $E(x) \subseteq I(x) \subseteq P(x)$.

Shephard's (1970) output distance function provides another functional representation of production technology. The output distance function is

$$D_o(x,y) = \min \{\lambda: (y/\lambda) \in P(x)\}. \tag{1.11}$$

For $y \in P(x)$, $D_o(x,y) \leqq 1$, and for $y \in I(x)$, $D_o(x,y) = 1$. Given standard assumptions on T, the output distance function $D_o(x,y)$ is nonincreasing in x, and nondecreasing, homogeneous of degree +1 and convex in y.

The Debreu-Farrell output-oriented measure of technical efficiency can now be given a somewhat more formal interpretation as the value of the function

$$TE_o(x,y) = \max \{\phi: \phi y \in P(x)\}, \tag{1.12}$$

and it follows from (1.11) that

$$TE_o(x,y) = [D_o(x,y)]^{-1}. \tag{1.13}$$

For $y \in P(x)$, $TE_o(y,x) \geqq 1$, and for $y \in I(x)$, $TE_o(x,y) = 1$. (Caution: some authors replace (1.12) and (1.13) with $TE_o(x,y) = [\max \{\phi: \phi y \in P(x)\}]^{-1} = D_o(x,y)$, so that $TE_o(x,y) \leqq 1$ just as $TE_I(y,x) \leqq 1$. We follow the convention of defining efficiency of any sort as the ratio of optimal to actual. Consequently $TE_I(y,x) \leqq 1$ and $TE_o(y,x) \geqq 1$.)

The foregoing analysis presumes that M>1, N>1. In the single input case

$$D_I(y,x) = x/g(y) \geqq 1 \Leftrightarrow x \geqq g(y),$$ (1.14)

where $g(y) = \min\{x: x \in L(y)\}$ is an input requirement frontier that defines the minimum amount of scalar input x required to produce output vector y. In this case the input-oriented measure of technical efficiency (1.7) becomes the ratio of minimum to actual input

$$TE_I(y,x) = 1/D_I(y,x) = g(y)/x \leqq 1.$$ (1.15)

In the single output case

$$D_o(x,y) = y/f(x) \leqq 1 \Leftrightarrow y \leqq f(x),$$ (1.16)

where $f(x) = \max\{y: y \in P(x)\}$ is a production frontier that defines the maximum amount of scalar output that can be produced with input vector x. In this case the output-oriented measure of technical efficiency in (1.13) becomes the ratio of maximum to actual output

$$TE_o(x,y) = [D_o(x,y)]^{-1} = f(x)/y \geqq 1.$$ (1.17)

The two technical efficiency measures are illustrated in Figures 1.4 - 1.6. As a preview of things to come, technology is smooth in Figure 1.4 and piecewise linear in Figures 1.5 and 1.6. This reflects different approaches to using data to estimate technology. The econometric approach introduced in Section 1.5 and developed in Chapter 2 estimates smooth parametric frontiers, while the mathematical programming approach introduced in Section 1.6 and developed in Chapter 3 estimates piecewise linear nonparametric frontiers.

In Figure 1.4 producer A is located on the interior of T, and its efficiency can be measured horizontally with an input conserving orientation using (1.6) or vertically with an output augmenting orientation using (1.12). If an input orientation is selected, $TE_I(y^A,x^A) = \theta x^A/x^A \leqq 1$, while if an output orientation is selected, $TE_o(x^A,y^A) = \phi y^A/y^A \geqq 1$.

Figure 1.4 Technical Efficiency
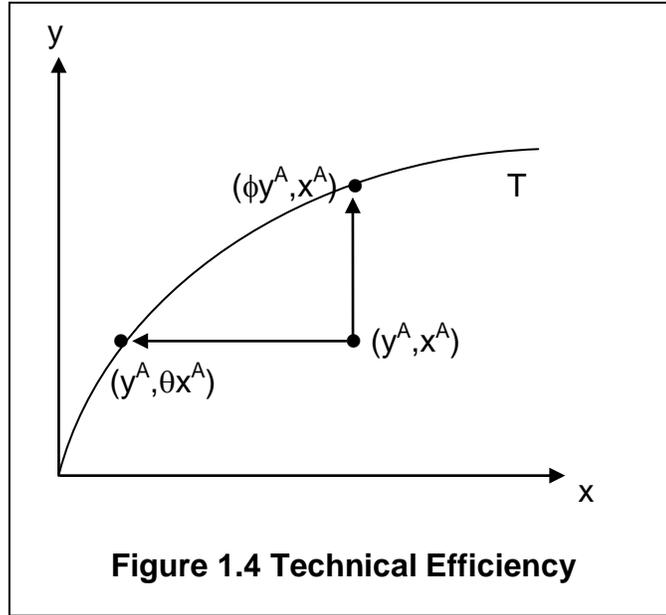
It is also possible to combine the two directions by simultaneously expanding outputs and contracting inputs, either hyperbolically or along a right angle, to arrive at an efficient point on the surface of T between $(y^A,\theta x^A)$ and $(\phi y^A,x^A)$. A hyperbolic measure of technical efficiency is defined as

$$TE_H(y,x) = \max \{\alpha: (\alpha y,x/\alpha) \in T\} \geqq 1, \qquad\qquad (1.18)$$

and $TE_H(y,x)$ is the reciprocal of a hyperbolic distance function $D_H(y,x)$. Under constant returns to scale, $TE_H(y,x) = [TE_o(x,y)]^2 = [TE_I(y,x)]^{-2}$, and $TE_H(y,x)$ is dual to a profitability function. One version of a directional measure of technical efficiency is defined as

$$TE_D(y,x) = \max \{\beta: [(1+\beta)y,(1-\beta)x] \in T\} \geqq 0, \qquad\qquad (1.19)$$

and $TE_D(y,x)$ is equal to a directional distance function $D_D(y,x)$. Even without constant returns to scale, $TE_D(y,x)$ can be related to $TE_o(x,y)$ and $TE_I(y,x)$, and is dual to a profit function. The directional measure and its underlying directional distance function are employed to good advantage in Chapter 5.

In Figure 1.5 input vectors x^A and x^B are on the interior of L(y), and both can be contracted radially and still remain capable of producing output vector y. Input vectors x^C and x^D cannot be contracted radially and still remain capable of producing output vector y because they are located on the input isoquant I(y). Consequently $TE_I(y,x^C) = TE_I(y,x^D) = 1 > \max \{TE_I(y,x^A), TE_I(y,x^B)\}$. Since the

radially scaled input vector $\theta^B x^B$ contains slack in input x_2, there may be some hesitancy in describing input vector $\theta^B x^B$ as being technically efficient in the production of output vector y. No such problem occurs with radially scaled input vector $\theta^A x^A$. Thus $TE_I(y,\theta^A x^A) = TE_I(y,\theta^B x^B) = 1$ even though $\theta^A x^A \in E(y)$ but $\theta^B x^B \notin E(y)$.
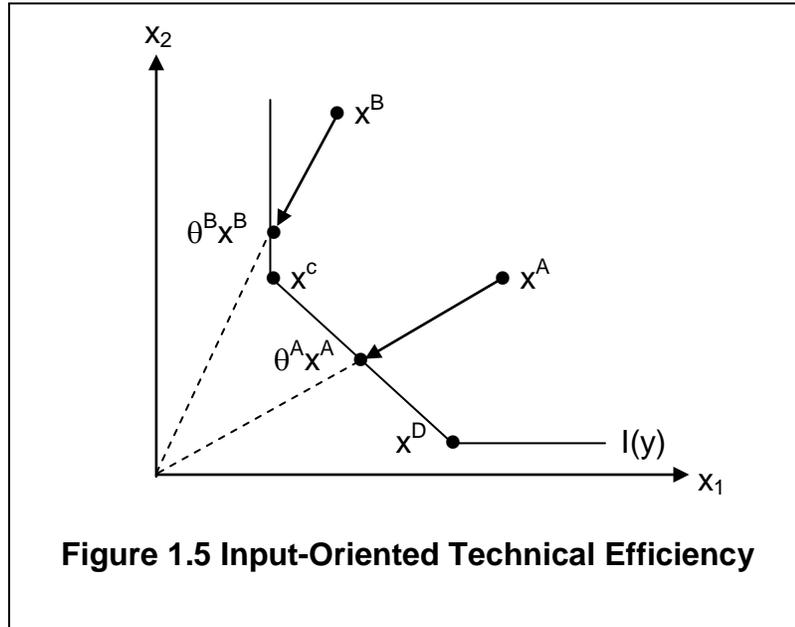

Figure 1.5 Input-Oriented Technical Efficiency

Figure 1.6 tells exactly the same story, but with an output orientation. Output vectors y^C and y^D are technically efficient given input usage x, and output vectors y^A and y^B are not. Radially scaled output vectors $\phi^A y^A$ and $\phi^B y^B$ are technically efficient, even though slack in output y_2 remains at $\phi^B y^B$. Thus $TE_o(x,\phi^A y^A) = TE_o(x,\phi^B y^B) = 1$ even though $\phi^A y^A \in E(x)$ but $\phi^B y^B \notin E(x)$.

Figure 1.6 Output-Oriented Technical Efficiency
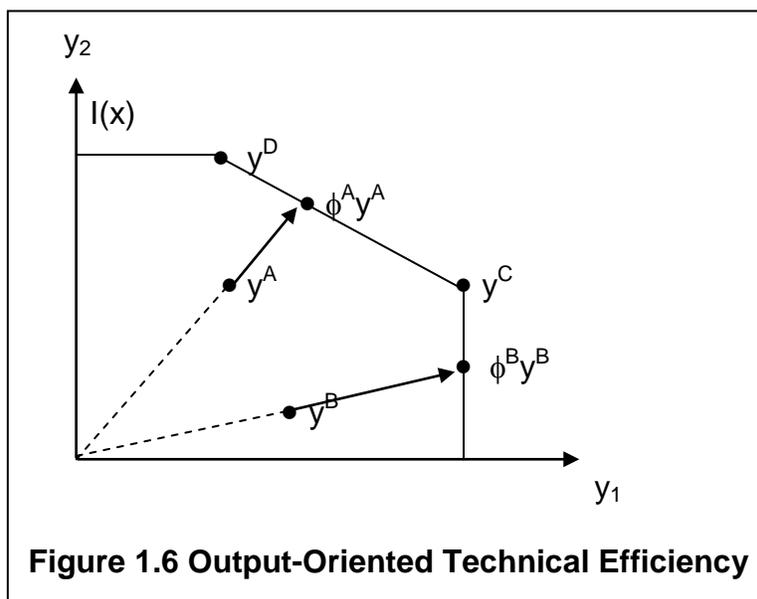
The Debreu-Farrell measures of technical efficiency are widely used. Since they are reciprocals of distance functions, they satisfy several nice properties (as noted first by Shephard (1970), and most thoroughly by Russell (1988, 1990)). Among these properties are

- $TE_I(y,x)$ is homogeneous of degree -1 in inputs, and $TE_o(x,y)$ is homogeneous of degree -1 in outputs
- $TE_I(y,x)$ is weakly monotonically decreasing in inputs, and $TE_o(x,y)$ is weakly monotonically decreasing in outputs
- $TE_I(y,x)$ and $TE_o(x,y)$ are invariant with respect to changes in units of measurement

On the other hand, they are not perfect. A notable feature of the Debreu-Farrell measures of technical efficiency is that they do not coincide with Koopmans' definition of technical efficiency. Koopmans' definition is demanding, requiring the absence of coordinate-wise improvements (simultaneous membership in both efficient subsets), while the Debreu-Farrell measures require only the absence of radial improvements (membership in isoquants). Thus the Debreu-Farrell measures correctly identify all Koopmans-efficient producers as being technically efficient, they also identify as being technically efficient any other producers located on an isoquant outside the efficient subset. Consequently Debreu-Farrell technical efficiency is necessary, but not sufficient, for Koopmans technical efficiency. The possibilities are illustrated in Figures 1.5 and 1.6, where $\theta^B x^B$ and $\phi^B y^B$ satisfy the Debreu-Farrell conditions but not the Koopmans requirement because slacks remain at the optimal radial projections.

Much has been made of this property of the Debreu-Farrell measures, but we think the problem is exaggerated. The practical significance of the problem depends on how many observations lie outside the cone spanned by the relevant efficient subset. Hence the problem disappears in much econometric analysis, in which the parametric form of the function used to estimate production technology (e.g., Cobb-Douglas, but not flexible functional forms such as translog) imposes equality between isoquants and efficient subsets, thereby eliminating slack by assuming it away. The problem assumes greater significance in the mathematical programming approach, in which the nonparametric form of the frontier used to estimate the boundary of the production set imposes slack by a strong (or free) disposability assumption. If the problem is deemed significant in practice, then it is possible to report Debreu-Farrell efficiency scores and slacks separately, side by side. This is rarely done. Instead, much effort has been directed toward finding a "solution" to the problem. Three strategies have been proposed.

- Replace the radial Debreu-Farrell measure with a nonradial measure that projects to efficient subsets (Färe and Lovell (1978)). This guarantees that an observation (or its projection) is technically efficient if, and only if, it is efficient in Koopmans' sense. However nonradial measures gain this "indication" property at the considerable cost of failing the homogeneity property.
- Develop a measure that incorporates slack and the radial component into an inclusive measure of technical efficiency (Cooper *et al.* (1999)). This measure also gains the indication property, but it has its own problems, including the possibility of negative values.
- Eliminate slack altogether by enforcing strictly positive marginal rates of substitution and transformation. We return to this possibility in Section 1.6.4, in a different setting.

Happily, there is no such distinction between definitions and measures of economic efficiency. Defining and measuring economic efficiency requires the specification of an economic objective and information on relevant prices. If the objective of a production unit (or the objective assigned to it by the analyst) is cost minimization, then a measure of cost efficiency is provided by the ratio of minimum feasible cost to actual cost. This measure depends on input prices. It attains a maximum value of unity if the producer is cost efficient, and a value less than unity indicates the degree of cost inefficiency. A measure of input allocative efficiency is obtained residually as the ratio of the measure of cost efficiency to the input-oriented measure of technical efficiency. The modification of this Farrell decomposition of cost efficiency to the output-oriented problem of decomposing revenue efficiency is straightforward. Modifying the procedure to accommodate alternative behavioral objectives is sometimes straightforward and occasionally challenging. So is the incorporation of regulatory and other non-technological constraints that impede the pursuit of some economic objective.

Suppose that producers face input prices $w = (w_1,\ldots,w_N) \in R^N_{++}$ and seek to minimize cost. Then a minimum cost function, or a cost frontier, is defined as

$$c(y,w) = \min_x \{w^T x: D_I(y,x) \geqq 1\}. \tag{1.20}$$

If the input sets $L(y)$ are closed and convex, and if inputs are freely disposable, the cost frontier is dual to the input distance function in the sense of (1.20) and

$$D_I(y,x) = \min_w \{w^T x: c(y,w) \geqq 1\}. \tag{1.21}$$

A measure of cost efficiency is provided by the ratio of minimum cost to actual cost

$$CE(x,y,w) = c(y,w) / w^T x. \tag{1.22}$$

A measure of input allocative efficiency is obtained from (1.6) and (1.22) as

$$AE_I(x,y,w) = CE(x,y,w) / TE_I(y,x). \tag{1.23}$$

$CE(x,y,w)$ and its two components are bounded above by unity, and $CE(x,y,w) = TE_I(y,x) \times AE_I(x,y,w)$.

The measurement and decomposition of cost efficiency is illustrated in Figures 1.7 and 1.8. In Figure 1.7 the input vector x^E minimizes the cost of producing output vector y at input prices w, and so $w^T x^E = c(y,w)$. The cost efficiency of x^A is given by the ratio $w^T x^E / w^T x^A = c(y,w)/w^T x^A$. The Debreu-Farrell measure of the technical efficiency of x^A is given by $\theta^A = \theta^A x^A / x^A = w^T (\theta^A x^A)/w^T x^A$. The allocative efficiency of x^A is determined residually as the ratio of cost efficiency to technical efficiency, or by the ratio $w^T x^E / w^T (\theta^A x^A)$. The magnitudes of technical, allocative and cost inefficiency are all measured by ratios of price-weighted input vectors. The direction of allocative inefficiency is revealed by the input vector difference $(x^E - \theta^A x^A)$. An alternative view of cost efficiency is provided by Figure 1.8, in which $CE(x^A,y^A,w) = c(y^A,w)/w^T x^A$.

Figure 1.7 Cost Efficiency I

Figure 1.8 Cost Efficiency II
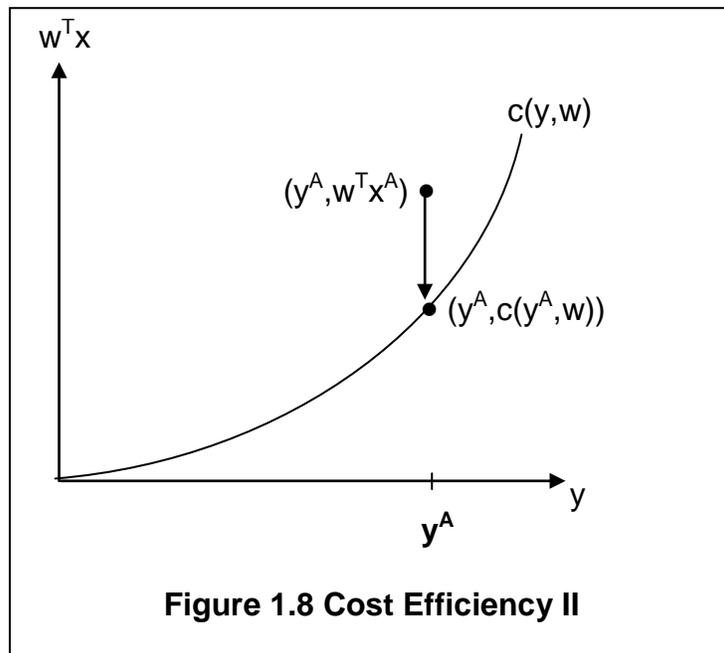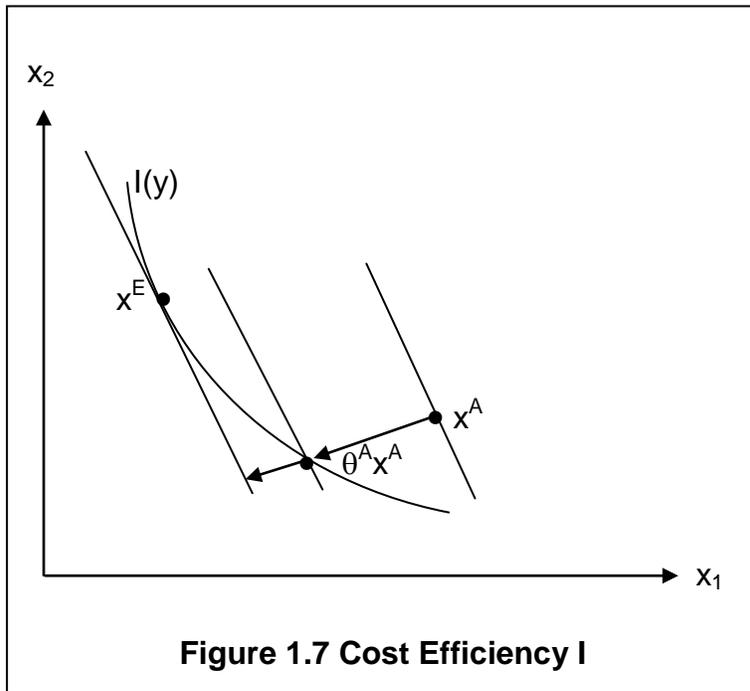
The measurement and decomposition of cost efficiency is illustrated again in Figure 1.9, for the case in which the efficient subset is a proper subset of the isoquant. The analysis proceeds as above, with a twist. The cost efficiency of input vector x^A now has three components, a radial technical component $[w^T(\theta^A x^A)/w^T x^A]$, an input slack component $[w^T x^B/w^T(\theta^A x^A)]$, and an allocative component $[w^T x^E/w^T x^B]$. With input price data all three components can be identified, although they rarely are. The slack component is routinely assigned to the allocative component.

Figure 1.9 Cost Efficiency III
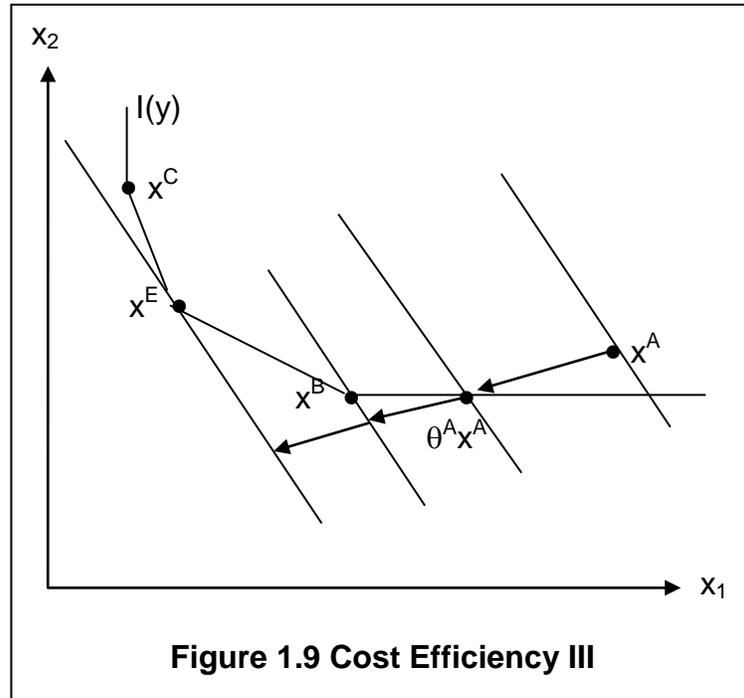
Suppose next that producers face output prices $p = (p_1,\ldots,p_M) \in R^M_{++}$ and seek to maximize revenue. Then a maximum revenue function, or a revenue frontier, is defined as

$$r(x,p) = \max_y \{p^T y: D_o(x,y) \leqq 1\}, \tag{1.24}$$

If the output sets $P(x)$ are closed and convex, and if outputs are freely disposable, the revenue frontier is dual to the output distance function in the sense of (1.24) and

$$D_o(x,y) = \max_p (p^T y: r(x,p) \leqq 1\}. \tag{1.25}$$

29

A measure of revenue efficiency is provided by the ratio of maximum revenue to actual revenue

$$RE(y,x,p) = r(x,p) / p^T y. \qquad\qquad (1.26)$$

A measure of output allocative efficiency is obtained from (1.12) and (1.26) as

$$AE_o(y,x,p) = RE(y,x,p) / TE_o(x,y). \qquad\qquad (1.27)$$

$RE(y,x,p)$ and its two components are bounded below by unity, and $RE(y,x,p) = TE_o(x,y) \times AE_o(y,x,p)$.

The measurement and decomposition of revenue efficiency in Figures 1.10 and 1.11 follows exactly the same steps. The measurement and decomposition of revenue efficiency in the presence of output slack follows along similar lines as in Figure 1.9. Revenue loss attributable to output slack is typically assigned to the output allocative efficiency component of revenue efficiency.

Figure 1.10 Revenue Efficiency I

Figure 1.11 Revenue Efficiency II
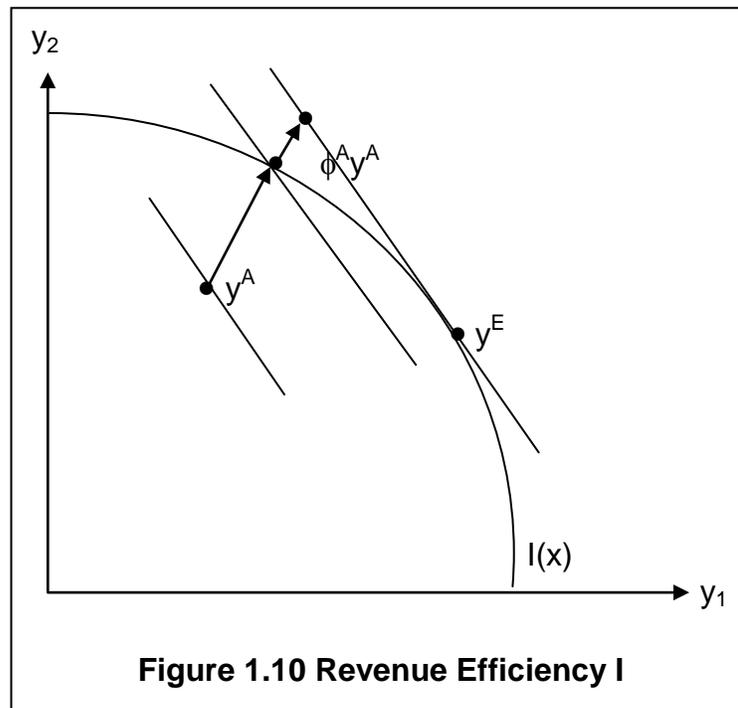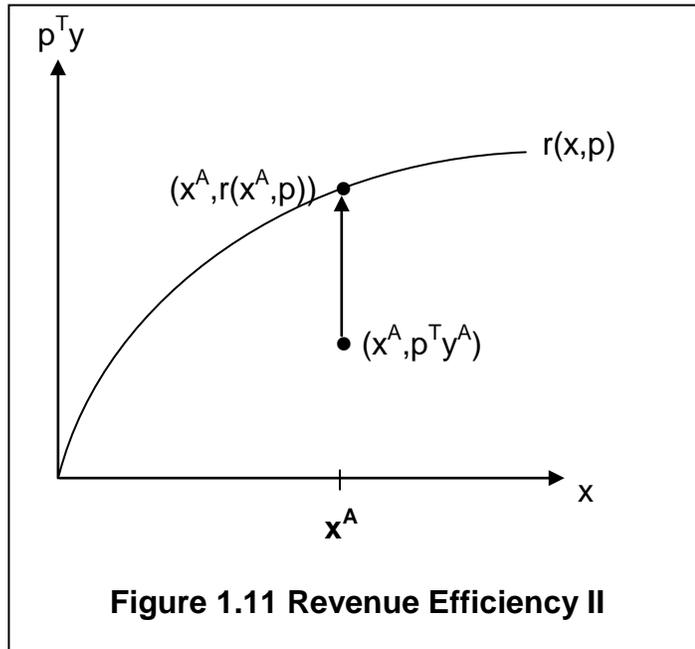
Cost efficiency and revenue efficiency are important performance indicators, but each reflects just one dimension of a firm's overall performance. A measure of profit efficiency captures both dimensions, and relates directly to the bottom line discussed in Section 1.1. Suppose that producers face output prices $p \in R^M_{++}$ and input prices $w \in R^N_{++}$, and seek to maximize profit. The maximum profit function, or profit frontier, is defined as

$$\pi(p,w) = \max_{y,x} \{(p^Ty - w^Tx): (y,x) \in T\}. \tag{1.28}$$

If the production set T is closed and convex, and if outputs and inputs are freely disposable, the profit frontier is dual to T in the sense of (1.28) and

$$T = \{(y,x): (p^Ty - w^Tx) \leqq \pi(p,w) \ \forall \ p \in R^M_{++}, w \in R^N_{++}\}. \tag{1.29}$$

A measure of profit efficiency is provided by the ratio of maximum profit to actual profit

$$\pi E(y,x,p,w) = \pi(p,w)/(p^Ty - w^Tx), \tag{1.30}$$

provided $(p^Ty - w^Tx) > 0$, in which case $\pi E(y,x,p,w)$ is bounded below by unity. The decomposition of profit efficiency is partially illustrated by Figure 1.12, which builds on Figure 1.4. Profit at (y^A,x^A) is less than maximum profit at (y^E,x^E), and two possible decompositions of profit efficiency are illustrated. One takes an input-conserving orientation to the measurement of technical efficiency, and the

31

residual allocative component follows the path from $(y^A, \theta x^A)$ to (y^E, x^E). The other takes an output-augmenting orientation to the measurement of technical efficiency, with residual allocative component following the path from $(\phi y^A, x^A)$ to (y^E, x^E). In both approaches the residual allocative component contains an input allocative efficiency component and an output allocative efficiency component, although the magnitudes of each component can differ in the two approaches. These two components are hidden from view in the two-dimensional Figure 1.12. In both approaches the residual allocative efficiency component also includes a scale component, which is illustrated in Figure 1.12. The direction of the scale component is sensitive to the orientation of the technical efficiency component, which imposes a burden on the analyst to get the orientation right. Because profit efficiency involves adjustments to both outputs and inputs, hyperbolic and directional technical efficiency measures are appealing in this context. Whatever the orientation of the technical efficiency measure, profit inefficiency is attributable to technical inefficiency, to an inappropriate scale of operation, to the production of an inappropriate output mix, and to the selection of an inappropriate input mix.

Figure 1.12 Profit Efficiency
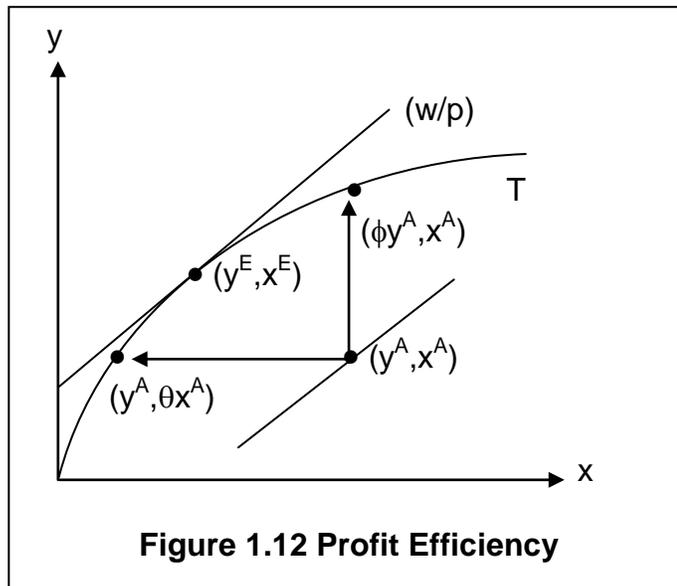
We conclude this Section with a brief discussion of dominance. Producer A dominates all other producers for which $(y^A, -x^A) \geq (y, -x)$. This notion is a direct application of Koopmans' definition of efficiency, in which producer A is "more efficient" than all other producers it dominates. Reversing the definition, producer A is dominated by all other producers for which $(y, -x) \geq (y^A, -x^A)$. In Figure 1.4 producer A is dominated by all producers to the northwest $\in T$ because they use no more input to produce at least as much output. Similar dominance relationships can be constructed in Figures 1.5 and 1.6. In each case dominance

is a physical, or technical, relationship. However dominance can also be given a value interpretation. In Figure 1.8 producer A is dominated (in a cost sense) by all other producers to the southeast on or above c(y,w) because they produce at least as much output at no more cost, and in Figure 1.11 producer A is dominated (in a revenue sense) by all other producers to the northwest on or beneath r(x,p) because they use no more input to generate at least as much revenue.

Dominance is an under-utilized concept in the field of producer performance evaluation, where the emphasis is on efficiency. This neglect is unfortunate, because dominance information offers a potentially useful complement to an efficiency evaluation, as Tulkens and Vanden Eeckaut (1995, 1999) have demonstrated. Inefficient producers can have many dominators, and hence many potential role models from which to learn. To cite one example, Fried *et al.* (1993) report an average of 22 dominators for each of nearly 9,000 US credit unions.

The identification of dominators can constitute the initial step in a benchmarking exercise. It is possible that dominators utilize superior business practices that are transferable to the benchmarking producer. However it is also possible that dominance is due to a more favorable operating environment. Although this may be cold comfort to the benchmarking business, it can be very useful to the analyst who does not want to confuse variation in performance with variation in the operating environment. Incorporating variation in the operating environment is an important part of any performance evaluation exercise, and techniques for doing so are discussed below and in subsequent Chapters.

1.4 Techniques for Efficiency Measurement

Efficiency measurement involves a comparison of actual performance with optimal performance located on the relevant frontier. Since the true frontier is unknown, an empirical approximation is needed. The approximation is frequently dubbed a "best practice" frontier.

The economic theory of production is based on production frontiers and value duals such as cost, revenue and profit frontiers, and on envelope properties yielding cost minimizing input demands, revenue maximizing output supplies, and profit maximizing output supplies and input demands. Emphasis is placed on optimizing behavior subject to constraint. However for over 75 years, at least since Cobb and Douglas started running regressions, the empirical analysis of production has been based on a least squares statistical methodology by which estimated functions of interest pass through the data and estimate mean performance. Thus the frontiers of theory have become the functions of analysis, interest in enveloping data with frontiers has been replaced with the

practice of intersecting data with functions, and unlikely efficient outcomes have been neglected in favor of more likely but less efficient outcomes, all as attention has shifted from extreme values to central tendency.

If econometric analysis is to be brought to bear on the investigation of the structure of economic frontiers, and on the measurement of efficiency relative to these frontiers, then conventional econometric techniques require modification. The modifications that have been developed, improved and implemented in the last three decades run the gamut from trivial to sophisticated. Econometric techniques are introduced in Section 1.5 and developed in detail in Chapter 2.

In sharp contrast to econometric techniques, mathematical programming techniques are inherently enveloping techniques, and so they require little or no modification to be employed in the analysis of efficiency. This makes them appealing, but they went out of favor long ago in the economics profession. Their theoretical appeal has given way to a perceived practical disadvantage, their ostensible failure to incorporate the statistical noise that drives conventional econometric analysis. This apparent shortcoming notwithstanding, they remain popular in the fields of management science and operations research, and they are making a comeback in economics. Programming techniques are introduced in Section 1.6 and developed in detail in Chapter 3.

The econometric approach to the construction of frontiers and the estimation of efficiency relative to the constructed frontiers has similarities and differences with the mathematical programming approach. Both are analytically rigorous benchmarking exercises that exploit the distance functions introduced in Section 1.3 to measure efficiency relative to a frontier. However the two approaches use different techniques to envelop data more or less tightly in different ways. In doing so they make different accommodations for statistical noise, and for flexibility in the structure of production technology. It is these two different accommodations that have generated debate about the relative merits of the two approaches. At the risk of oversimplification, the differences between the two approaches boil down to two essential features.
- The econometric approach is stochastic. This enables it to attempt to distinguish the effects of noise from those of inefficiency, thereby providing the basis for statistical inference.
- The programming approach is nonparametric. This enables it to avoid confounding the effects of misspecification of the functional form (of both technology and inefficiency) with those of inefficiency.

A decade or more ago, the implication drawn from these two features was that the programming approach was non-stochastic and the econometric approach was parametric. This had a disturbing consequence. If efficiency analysis is to be taken seriously, producer performance evaluation must be robust to both statistical noise and specification error. Neither approach was thought to be robust to both.

Happily, knowledge has progressed and distinctions have blurred. To praise one approach as being stochastic is not to deny that the other is stochastic as well, and to praise one approach as being nonparametric is not to damn the other as being rigidly parameterized. Recent explorations into the statistical foundations of the programming approach have provided the basis for statistical inference, and recent applications of flexible functional forms and semi-parametric, non-parametric and Bayesian techniques have freed the econometric approach from its parametric straitjacket. Both techniques are more robust than previously thought. The gap is no longer between one technique and the other, but between best practice knowledge and average practice implementation. The challenge is to narrow the gap.

It is worth asking whether the two techniques tell consistent stories when applied to the same data. The answer seems to be that the higher the quality of the data, the greater the concordance between the two sets of efficiency estimates. Of the many comparisons available in the literature, we recommend Bauer *et al.* (1998), who use US banking data, and Cummins and Zi (1998), who use US life insurance company data. Both studies find strong positive rank correlations of point estimates of efficiency between alternative pairs of econometric models and between alternative pairs of programming models, and weaker but nontheless positive rank correlations of point estimates of efficiency between alternative pairs of econometric and programming models.

Chapters 2 and 3 develop the two approaches, starting with their basic formulations and progressing to more advanced methods. Chapter 4 recasts the parametric econometric approach of Chapter 2 into a nonparametric statistical framework, and explores the statistical foundations of the programming approach of Chapter 3. In addition to these Chapters, we recommend comprehensive treatments of the econometric approach by Kumbhakar and Lovell (2000), and of the programming approach by Cooper *et al.* (2000). Both contain extensive references to analytical developments and empirical applications.

1.5 The Econometric Approach to Efficiency Measurement

Econometric models can be categorized according to the type of data they use (cross-section or panel), the type of variables they use (quantities only, or quantities and prices), and the number of equations in the model. In Section 1.5.1 we discuss the most widely used model, the single equation cross-section model. In Section 1.5.2 we progress to panel data models. In both contexts the efficiency being estimated can be either technical or economic. In Section 1.5.3 we discuss multiple equation models, and in Section 1.5.4 we discuss shadow price models, which typically involve multiple equations. In these two contexts the

efficiency being estimated is economic, with a focus on allocative inefficiency and its cost.

1.5.1 Single equation cross-section models

Suppose producers use inputs $x \in R^N_+$ to produce scalar output $y \in R_+$, with technology

$$y_i \lesseqgtr f(x_i;\beta)exp\{v_i\}, \tag{1.31}$$

where β is a parameter vector characterizing the structure of production technology and $i = 1,\ldots,I$ indexes producers. The deterministic production frontier is $f(x_i;\beta)$. Observed output y_i is bounded above by the stochastic production frontier $[f(x_i;\beta)exp\{v_i\}]$, with the random disturbance term $v_i \gtreqless 0$ included to capture the effects of statistical noise on observed output. The stochastic production frontier reflects what is possible $[f(x_i;\beta)]$ in an environment influenced by external events, favorable and unfavorable, beyond the control of producers $[exp\{v_i\}]$.

The weak inequality in (1.31) can be converted to an equality through the introduction of a second disturbance term to create

$$y_i = f(x_i;\beta)exp\{v_i - u_i\}, \tag{1.32}$$

where the disturbance term $u_i \geqq 0$ is included to capture the effect of technical inefficiency on observed output.

Recall from Section 1.3 that the Debreu-Farrell output-oriented measure of technical efficiency is the ratio of maximum possible output to actual output (and that some authors use the reciprocal of this measure). Applying definition (1.17) to (1.32) yields

$$TE_o(x_i,y_i) = f(x_i;\beta)exp\{v_i\} / y_i = exp\{u_i\} \geqq 1, \tag{1.33}$$

because $u_i \geqq 0$. The problem is to estimate $TE_o(x_i,y_i)$. This requires estimation of (1.32), which is easy and can be accomplished in a number of ways depending on the assumptions one is willing to make. It also requires a decomposition of the residuals into separate estimates of v_i and u_i, which is not so easy.

One approach, first suggested by Winsten (1957) and now known as corrected ordinary least squares (COLS), is to assume that $u_i = 0$, $i = 1,\ldots,I$, and that $v_i \sim N(0,\sigma_v^2)$. In this case (1.32) collapses to a standard regression model

that can be estimated consistently by OLS. The estimated production function, which intersects the data, is then shifted upward by adding the maximum positive residual to the estimated intercept, creating a production frontier that bounds the data from above. The residuals are corrected in the opposite direction, and become $\hat{v}_i = v_i - v_i^{max} \leqq 0$, $i = 1,\ldots,I$. The technical efficiency of each producer is estimated from

$$\hat{TE}_o\,(x_i,y_i) = exp\{-\hat{v}_i\} \geqq 1, \tag{1.34}$$

and $\hat{TE}_o\,(x_i,y_i)$ - 1 $\geqq$ 0 indicates the percent by which output can be expanded, on the assumption that $u_i = 0$, $i = 1,\ldots,I$.

The producer having the largest positive OLS residual supports the COLS production frontier. This makes COLS vulnerable to outliers, although *ad hoc* sensitivity tests have been proposed. In addition, the structure of the COLS frontier is identical to the structure of the OLS function, apart from the shifted intercept. This structural similarity rules out the possibility that efficient producers are efficient precisely because they exploit available economies and substitution possibilities that average producers do not. The assumption that best practice is just like average practice, but better, defies both common sense and much empirical evidence. Finally, it is troubling that efficiency estimates for all producers are obtained by suppressing the inefficiency error component u_i, and are determined exclusively by the single producer having the most favorable noise v_i^{max}. The term $exp\{u_i\}$ in (1.33) is proxied by the term $exp\{-\hat{v}_i\}$ in (1.34). Despite these reservations, and additional concerns raised in Chapters 2 and 4, COLS is widely used, presumably because it is easy.

A second approach, suggested by Aigner and Chu (1968), is to make the opposite assumption that $v_i = 0$, $i = 1,\ldots,I$. In this case (1.32) collapses to a deterministic production frontier that can be estimated by linear or quadratic programming techniques that minimize either $\Sigma_i u_i$ or $\Sigma_i u_i^2$, subject to the constraint that $u_i = \ln[f(x_i;\beta) / y_i] \geqq 0$ for all producers. The technical efficiency of each producer is estimated from

$$\hat{TE}_o\,(x_i,y_i) = exp\{\hat{u}_i\} \geqq 1, \tag{1.35}$$

and $\hat{TE}_o\,(x_i,y_i)$ - 1 $\geqq$ 0 indicates the percent by which output can be expanded, on the alternative assumption that $v_i = 0$, $i = 1,\ldots,I$. The $\hat{u}_i$ are estimated from the slacks in the constraints [$lnf(x_i;\beta)$ - $lny_i \geqq 0$, $i = 1,\ldots,I$] of the program. Although it appears that the term $exp\{\hat{u}_i\}$ in (1.35) coincides with the term $exp\{u_i\}$ in (1.33), the expression in (1.35) is conditioned on the assumption that $v_i = 0$, while the expression in (1.33) is not. In addition, since no distributional assumption is

imposed on $u_i \geqq 0$, statistical inference is precluded and consistency cannot be verified. However Schmidt (1976) showed that the linear programming "estimate" of β is maximum likelihood if the u_i follow an exponential distribution, and that the quadratic programming "estimate" of β is maximum likelihood if the u_i follow a half-normal distribution. Unfortunately we know virtually nothing about the statistical properties of these estimators, even though they are maximum likelihood. However Greene (1980) showed that an assumption that the u_i follow a gamma distribution generates a well-behaved likelihood function that allows statistical inference, although this model does not correspond to any known programming problem. Despite the obvious statistical drawback resulting from its deterministic formulation, the programming approach is also widely used. One reason for its popularity is that it is easy to append monotonicity and curvature constraints to the program, as Hailu and Veeman (2000) have done in their study of water pollution in the Canadian pulp and paper industry.

The third approach, suggested independently by Aigner *et al.* (1977) and Meeusen and van den Broeck (1977), attempts to remedy the shortcomings of the first two approaches, and is known as stochastic frontier analysis (SFA). In this approach it is assumed that $v_i \sim N(0,\sigma_v^2)$, and that $u_i \geqq 0$ follows either a half-normal or an exponential distribution. The motive behind these two distributional assumptions is to parsimoniously parameterize the notion that relatively high efficiency is more likely than relatively low efficiency. After all, the structure of production is parameterized, so we might as well parameterize the inefficiency distribution too. In addition, it is assumed that the v_i and the u_i are distributed independently of each other, and of x_i. OLS can be used to obtain consistent estimates of the slope parameters, but not the intercept because $E(v_i - u_i) = E(-u_i)$ $\leqq 0$. However the OLS residuals can be used to test for negative skewness, which is a test for the presence of variation in technical inefficiency. If evidence of negative skewness is found, OLS slope estimates can be used as starting values in a maximum likelihood routine.

Armed with the distributional and independence assumptions, it is possible to derive the likelihood function, which can be maximized with respect to all parameters (β, σ_v^2 and σ_u^2) to obtain consistent estimates of β. However even with this information neither team was able to estimate $TE_o(x_i,y_i)$ in (1.33) because they were unable to disentangle the separate contributions of v_i and u_i to the residual. Jondrow *et al.* (1982) provided an initial solution, by deriving the conditional distribution of $[-u_i|(v_i - u_i)]$, which contains all the information $(v_i - u_i)$ contains about $-u_i$. This enabled them to derive the expected value of this conditional distribution, from which they proposed to estimate the technical efficiency of each producer from

$$T\hat{E}_o(x_i,y_i) = \{\exp\{E[-\hat{u}_i|(v_i - u_i)]\}\}^{-1} \geqq 1, \qquad (1.36)$$

which is a function of the MLE parameter estimates. Later Battese and Coelli (1988) proposed to estimate the technical efficiency of each producer from

$$T\hat{E}_o\,(x_i,y_i) = \{E[exp\{-\hat{u}_i\}|(v_i - u_i)]\}^{-1} \geqq 1, \tag{1.37}$$

which is a slightly different function of the same MLE parameter estimates, and is preferred because $-\hat{u}_i$ in (1.36) is only the first order term in the power series approximation to $exp\{-\hat{u}_i\}$ in (1.37).

Unlike the first two approaches, which suppress either u_i or v_i, SFA sensibly incorporates both noise and inefficiency into the model specification. The price paid is the need to impose distributional and independence assumptions, the prime benefit being the ability to disentangle the two error components. The single parameter half-normal and exponential distributions can be generalized to more flexible two parameter truncated normal and gamma distributions, as suggested by Stevenson (1980) and Greene (1980), although they rarely are. The independence assumptions seem essential to the MLE procedure. The fact that they can be relaxed in the presence of panel data provides an initial appreciation of the value of panel data, to which we return in Section 1.5.2.

The efficiency estimates obtained from (1.36) and (1.37) are unbiased, but their consistency has been questioned, not because they converge to the wrong values, but because in a cross section we get only one look at each producer, and the number of looks cannot increase. However a new contrary claim of consistency is put forth in Chapter 2. The argument is simple, and runs as follows. The technical efficiency estimates in (1.36) and (1.37) are conditioned on MLEs of $(v_i - u_i) = \ln y_i - \ln f(x_i;\beta)$, and since (is estimated consistently by MLE, so is technical efficiency, even in a cross section.

For over a decade individual efficiencies were estimated using either (1.36) or (1.37). Hypothesis tests frequently were conducted on (, and occasionally on (u2/(v2 (or some variant thereof) to test the statistical significance of efficiency variation. However we did not test hypotheses on either estimator of TEo(xi,yi) because we did not realize that we had enough information to do so. We paid the price of imposing distributions on vi and ui, but we did not reap one of the benefits; we did not exploit the fact that distributions imposed on vi and ui create distributions for [-ui|(vi - ui)] and [exp{-ui}|(vi - ui)], which can be used to construct confidence intervals and to test hypotheses on individual efficiencies. This should have been obvious all along, but Horrace and Schmidt (1996) and Bera and Sharma (1999) were the first to develop confidence intervals for efficiency estimators. The published confidence intervals we have seen are depressingly wide, presumably because estimates of (u2/(v2 are relatively small. In such circumstances the information contained in a ranking of estimated efficiency scores is limited, frequently to the ability to distinguish stars from strugglers.

The preceding discussion has been based on a single output production frontier. However multiple outputs can be incorporated in a number of ways.

(Estimate a stochastic revenue frontier, with pTy replacing y and (x,p) replacing x in (1.32). The one-sided error component provides the basis for a measure of revenue efficiency. Applications are rare.

(Estimate a stochastic profit frontier, with (pTy - wTx) replacing y and (p,w) replacing x in (1.32). The one-sided error component provides the basis for a measure of profit efficiency. Estimation of profit frontiers is popular, especially in the financial institutions literature. Berger and Mester (1997) provide an extensive application to US banks.

(Estimate a stochastic cost frontier, with wTx replacing y and (y,w) replacing x in (1.32). Since $w^T x \geqq c(y,w;\beta)\exp\{v_i\}$, this requires changing the sign of the one-sided error component, which provides the basis for a measure of cost efficiency. Applications are numerous.

- Estimate a stochastic input requirement frontier, with the roles of x and y in (1.32) being reversed. This also requires changing the sign of the one-sided error component, which provides the basis for a measure of input use efficiency. Applications are limited to situations in which labor has a very large (variable?) cost share, or in which other inputs are not reported. Kumbhakar and Hjalmarsson (1995) provide an application to employment in Swedish social insurance offices.

- Estimate a stochastic output distance function $D_o(x,y)\exp\{v_i\} \leqq 1 \Rightarrow D_o(x_i,y_i;\beta)\exp\{v_i - u_i\} = 1$, $u_i \geqq 0$. The one-sided error component provides the basis for an output-oriented measure of technical efficiency. Unlike the models above, a distance function has no natural dependent variable, and at least three alternatives have been proposed. Fuentes *et al.* (2001) and Atkinson *et al.* (2003) illustrate alternative specifications and provide applications to Spanish insurance companies and US railroads, respectively.

- Estimate a stochastic input distance function $D_I(y,x)\exp\{v_i\} \geqq 1 \Rightarrow D_I(y_i,x_i;\beta)\exp\{v_i + u_i\} = 1$, $u_i \geqq 0$. Note the sign change of the one-sided error component, which provides the basis for an input-oriented measure of technical efficiency, and proceed as above.

In the preceding discussion interest has centered on the estimation of efficiency. A second concern, first raised in Section 1.2, involves the incorporation of potential determinants of efficiency. The determinants can include characteristics of the operating environment, and characteristics of the manager such as human capital endowments. The logic is that if efficiency is to be improved, we need to know what factors influence it, and this requires distinguishing the influences of the potential determinants from that of the inputs and outputs themselves. Two approaches have been developed.

Let $z \in R^K$ be a vector of exogenous variables thought to be relevant to the production activity. One approach that has been used within and outside the frontier field is to replace $f(x_i;\beta)$ with $f(x_i,z_i;\beta,\gamma)$. The most popular example involves z serving as a proxy for technical change that shifts the production (or cost) frontier. Another popular example involves the inclusion of stage length and load factor in the analysis of airline performance; both are thought to influence operating cost. Although z is relevant in the sense that it is thought to be an important characteristic of production activity, it does not influence the efficiency of production. The incorporation of potential influences on productive efficiency requires an alternative approach, in which z influences the distance of producers from the relevant frontier.

In the old days it was common practice to adopt a two-stage approach to the incorporation of potential determinants of productive efficiency. In this approach efficiency was estimated in the first stage using either (1.36) or (1.37), and estimated efficiencies were regressed against a vector of potential influences in the second stage. Deprins and Simar (1989) were perhaps the first to question the statistical validity of this two-stage approach. Later Battese and Coelli (1995) proposed a single-stage model of general form

$$y_i = f(x_i;\beta)\exp\{v_i - u_i(z_i;\gamma)\}, \tag{1.38}$$

where $u_i(z_i;\gamma) \geqq 0$ and z is a vector of potential influences with parameter vector γ, and they showed how to estimate the model in SFA format. Later Wang and Schmidt (2002) analyzed alternative specifications for $u_i(z_i;\gamma)$ in the single-stage approach; for example, either the mean or the variance of the distribution being truncated below at zero can be made a function of z_i. They also provided detailed theoretical arguments, supported by compelling Monte Carlo evidence, explaining why both stages of the old two-stage procedure are seriously biased. Hopefully we will see no more two-stage SFA models.

1.5.2 Single equation panel data models

In a cross section each producer is observed once. If each producer is observed over a period of time, panel data techniques can be brought to bear on the problem. At the heart of the approach is the association of a "firm effect" from the panel data literature with a one-sided inefficiency term from the frontier literature. How this association is formulated, and how the model is estimated, are what distinguish one model from another. Whatever the model, the principal advantage of having panel data is the ability to observe each producer more than once. It should be possible to parlay this ability into "better" estimates of efficiency than can be obtained from a single cross section.

Schmidt and Sickles (1984) were among the first to consider the use of conventional panel data techniques in a frontier context. We follow them by writing the panel data version of the production frontier model (1.32) as

$$y_{it} = f(x_{it};\beta)\exp\{v_{it} - u_i\}, \tag{1.39}$$

where a time subscript $t = 1,\ldots,T$ has been added to y, x and v, but not (yet) to u. We begin by assuming that technical efficiency is time-invariant, and not a function of exogenous influences. Four estimation strategies are available.

It is straightforward to adapt the cross-section MLE procedures developed in Section 1.5.1 to the panel data context, as Pitt and Lee (1981) first showed. Allowing u_i to depend on potential influences is also straightforward, as Battese and Coelli (1995) demonstrated. Extending (1.39) by setting $u_{it} = u_{it}(z_{it};\gamma)$ and specifying one of the elements of z_{it} as a time trend or a time dummy allows technical inefficiency to be time-varying, which is especially desirable in long panels. MLE estimators of technical efficiency obtained from (1.36) and (1.37) are consistent in T and I. However MLE requires strong distributional and independence assumptions, and the availability of panel data techniques enables us to relax some of these assumptions.

The fixed effects model is similar to cross-section COLS. It imposes no distributional assumption on u_i, and allows the u_i to be correlated with the v_{it} and the x_{it}. Since the u_i are treated as fixed, they become producer-specific intercepts $\beta_{oi} = (\beta_o - u_i)$ in (1.39), which can be estimated consistently by OLS. After estimation the normalization $\beta_o^* = \beta_{oi}^{max}$ generates estimates of $\hat{u}_i = \beta_o^* - \beta_{oi} \geqq 0$, and estimates of producer-specific technical efficiencies are obtained from

$$T\hat{E}_o(x_i,y_i) = [\exp\{-\hat{u}_i\}]^{-1}. \tag{1.40}$$

These estimates are consistent in T and I, and they have the great virtue of allowing the u_i to be correlated with the regressors. However the desirable property of consistency in T is offset by the undesirability of assuming time-invariance of inefficiency in long panels. In addition, the fixed effects model has a potentially serious drawback. The firm effects are intended to capture variation in technical efficiency, but they also capture the effects of all phenomena that vary across producers but not through time, such as locational characteristics and regulatory regime.

The random effects model makes the opposite assumptions on the u_i, which are allowed to be random, with unspecified distribution having constant mean and variance, but are assumed to be uncorrelated with the v_{it} and the x_{it}. This allows the inclusion of time-invariant regressors in the model. Defining $\beta_o^{**} = \beta_o - E(u_i)$ and $u_i^{**} = u_i - E(u_i)$, (1.39) can be estimated by GLS. After estimation, firm-specific estimates of u_i^{**} are obtained from the temporal means of the

residuals. Finally, these estimates are normalized to obtain estimates of $\hat{u}_i = u_i^{**max} - u_i^{**}$, from which producer-specific estimates of technical efficiency are obtained from

$$T\hat{E}_o(x_i,y_i) = [\exp\{-\hat{u}_i\}]^{-1}. \qquad (1.41)$$

These estimates also are consistent in T and I. The main virtue of GLS is that it allows the inclusion of time-invariant regressors, whose impacts would be confounded with efficiency variation in a fixed effects model.

Finally, a Hausman-Taylor (1981) estimator can be adapted to (1.39). It is a mixture of the fixed effects and random effects estimators that allows the u_i to be correlated with some, but not all, regressors, and can include time-invariant regressors.

We have explored the tip of the proverbial iceberg. Panel data econometrics is expanding rapidly, as is its application to frontier models. Details are provided in Chapter 2.

1.5.3 Multiple equation models

We begin by reproducing a model popularized long ago by Christensen and Greene (1976). The model is

$$\ln(w^T x)_i = c(\ln y_i, \ln w_i; \beta) + v_i$$

$$(w_n x_n / w^T x)_i = s_n(\ln y_i, \ln w_i; \beta) + v_{ni}, \; n = 1,\ldots,N\text{-}1. \qquad (1.42)$$

This system describes the behavior of a cost-minimizing producer, with the first equation being a cost function and the remaining equations exploiting Shephard's (1953) lemma to generate cost-minimizing input cost shares. The errors (v_i, v_{ni}) reflect statistical noise, and are assumed to be distributed multivariate normal with zero means. The original motivation for appending the cost share equations was to increase statistical efficiency in estimation, since they contain no parameters not appearing in the cost function. Variants on this multiple equation theme, applied to flexible functional forms such as translog, appear regularly in production (and consumption) economics.

The pursuit of statistical efficiency is laudable, but it causes difficulties when the objective of the exercise is the estimation of economic efficiency. We do not want to impose the assumption of cost minimization that drives Shephard's lemma, so we transform the Christensen-Greene model (1.42) into a stochastic cost frontier model as follows:

$$\ln(w^T x)_i = c(\ln y_i, \ln w_i; \beta) + v_i + T_i + A_i$$

43

$$(w_n x_n / w^T x)_i = s_n(\ln y_i, \ln w_i; \beta) + v_{ni} + u_{ni}, \quad n = 1, \ldots, N\text{-}1. \tag{1.43}$$

Here v_i and the v_{ni} capture the effects of statistical noise. $T_i \gtreqqless 0$ reflects the cost of technical inefficiency, $A_i \gtreqqless 0$ reflects the cost of input allocative inefficiency, and $(T_i + A_i) \geqq 0$ is the cost of both. Finally, the $u_{ni} \gtreqqless 0$ capture the departures of actual input cost shares from their cost-efficient magnitudes. Since technical inefficiency is measured radially, it maintains the observed input mix and has no impact on input share equations. However allocative inefficiency represents an inappropriate input mix, and so its cost must be linked to the input cost share equations by means of a relationship between A_i and the u_{ni}, $n = 1, \ldots, N\text{-}1$. The linkage must respect the fact that cost is raised by allocative errors in any input in either direction. The formidable problem is to estimate the technology parameters β and the efficiency error components (T_i, A_i and u_{ni}) for each producer.

The problem is both conceptual and statistical. The conceptual challenge is to establish a satisfactory linkage between allocative inefficiency (the u_{ni}) and its cost (A_i). The statistical challenge is to estimate a model with so many error components, each of which requires a distribution. The problem remained unresolved until Kumbhakar (1997) obtained analytical results, which Kumbhakar and Tsionas (2005) extended to estimate the model using Bayesian techniques. This is encouraging, because (1.42) remains a workhorse in the non-frontier literature, and more importantly because its extension (1.43) is capable of estimating and decomposing economic efficiency.

There is an appealing alternative. The solution is to remove the influence of allocative inefficiency from the error terms and parameterize it inside the cost frontier and its input cost shares. We turn to this approach below.

1.5.4 Shadow price models

The econometric techniques described in Sections 1.5.1 - 1.5.3 are enveloping techniques. Each treats technical efficiency in terms of distance to a production frontier, economic efficiency in terms of distance to an appropriate economic frontier, and allocative efficiency as a ratio of economic efficiency to technical efficiency. They are in rough concordance on the fundamental notions of frontiers and distance, in keeping with the theoretical developments in Section 1.3. They differ mainly in the techniques they employ to construct frontiers and to measure distance. However they all convert a weak inequality to an equality by introducing a one-sided error component.

There is a literature that seeks to measure efficiency without explicit recourse to frontiers, and indeed it contains many papers in which the word "frontier" does not appear. In this literature little attempt is made to envelop data

44

or to associate efficiency with distance to an enveloping surface. Unlike most econometric efficiency analysis, the focus is on allocative efficiency. Instead of attempting to model allocative inefficiency by means of error components, as in (1.43), allocative inefficiency is modeled parametrically by means of additional parameters to be estimated.

The literature seems to have originated with Hopper (1965), who found subsistence agriculture in India to attain a high degree of allocative efficiency, supporting the "poor but efficient" hypothesis. He reached this conclusion by using OLS to estimate Cobb-Douglas production functions (not frontiers), then to calculate the value of the marginal product of each input, and then to make two comparisons: the value of an input's marginal product for different outputs, and the values of an input's marginal product with its price. In each comparison equality implies allocative efficiency, and the sign and magnitude of an inequality indicates the direction and severity (and the cost, which can be calculated since the production function parameters have been estimated) of the allocative inefficiency. Hopper's work was heavily criticized, and enormously influential.

In a nutshell, the shadow price models that have followed have simply parameterized Hopper's comparisons, with inequalities being replaced with parameters to be estimated. Thus, assuming M=1 for simplicity and following Lau and Yotopoulos (1971) and Yotopoulos and Lau (1973), the inequality

$$y \leqq f(x;\beta) \tag{1.44}$$

is parameterized as

$$y = \phi f(x;\beta). \tag{1.45}$$

There is no notion of a production frontier here, since in moving from (1.44) to (1.45) the obvious requirement that $\max\{\phi\} \leqq 1$ is ignored. Indeed so far this is just a Hoch (1955)-Mundlak (1961) management bias production function model, in which different intercepts are intended to capture the effects of variation in the (unobserved) management input. But it gets better.

If producers seek to maximize profit, then the inequalities

$$\partial \phi f(x;\beta)/\partial x_n \gtreqless (w_n/p), \ n = 1,\ldots,N \tag{1.46}$$

are parameterized as

$$\partial \phi f(x;\beta)/\partial x_n = \theta_n(w_n/p), \tag{1.47}$$

where $\theta_n \gtreqless 1$ indicate under- or over-utilization of x_n relative to the profit maximizing values. All that remains is to endow $f(x;\beta)$ with a functional form, and estimation of (β,ϕ,θ_n) provides a more sophisticated framework within which to implement Hopper's procedures. A host of hypotheses can be tested concerning the existence and nature of technical and allocative efficiency, without recourse to the notion of a frontier and error components.

The shadow price approach gained momentum following the popularity of the Averch-Johnson (1962) hypothesis. This hypothesis asserted that regulated utilities allowed to earn a "fair" rate of return on their invested capital would rationally overcapitalize, leading to higher than minimum cost and thus to customer rates that were higher than necessary.

The analysis proceeds roughly as above. A producer's cost

$$w^T x \geqq c(y,w;\beta) \tag{1.48}$$

is parameterized as

$$w^T x = (1/\phi)c(y,\theta w;\beta), \tag{1.49}$$

where θw is a vector of shadow prices. Now $\phi \leqq 1$ reflects technical inefficiency and $\theta_n \gtreqless 1$ reflects allocative inefficiency, and there is an explicit notion of a cost frontier. A producer's input demands

$$x_n \gtreqless x_n(y,w;\beta) \tag{1.50}$$

are parameterized as

$$x_n = (1/\phi)x_n(y,\theta w;\beta). \tag{1.51}$$

Although x_n may be allocatively inefficient for the input prices w a producer actually pays, it is allocatively efficient for the shadow price vector θw.

The Averch-Johnson hypothesis asserts that rate of return regulation lowers the shadow price of capital beneath the cost of capital, leading to rational overcapitalization. The situation is depicted in Figure 1.13. Given exogenous output y and input prices w_K and w_L, the cost minimizing input combination occurs at x^E. The actual input combination occurs at x^A, which is technically efficient but allocatively inefficient, involving overcapitalization. Since the actual input combination must be allocatively efficient for some price ratio, the problem boils down to one of estimating the distortion factor θ along with the technology parameters β. In the two-input case illustrated in Figure 1.13, there is one

distortion parameter, while in the N input case there are N-1 distortion parameters. The hypothesis of interest is that $\theta < 1$, the cost of which is given by the ratio $[c(y,\theta w;\beta) / c(y,w;\beta)] \geqq 1$, which is the reciprocal of the cost efficiency measure (1.22) translated to this analytical framework.

Figure 1.13 The Averch - Johnson Hypothesis
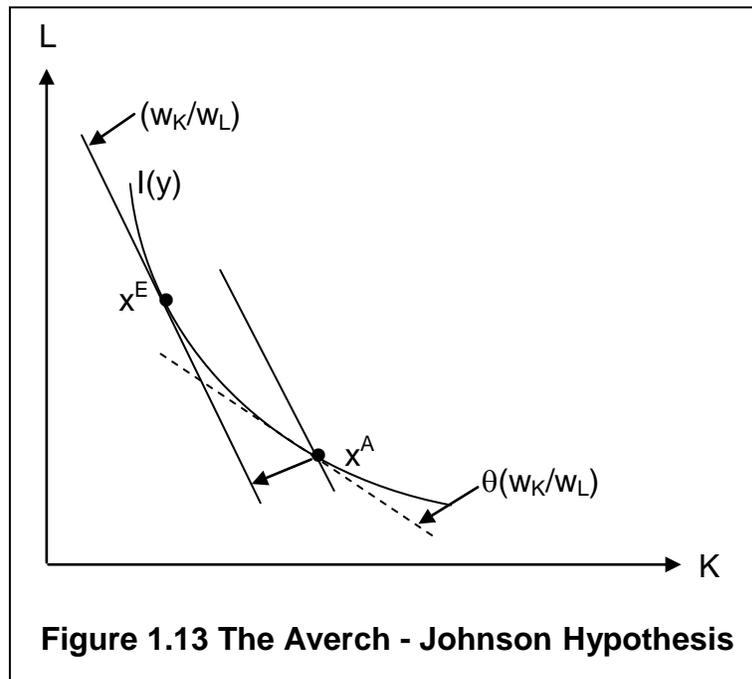
Comparing (1.49) and (1.51) with (1.43) makes it clear that in the shadow price approach both sources of cost inefficiency have been moved from error components to the functions to be estimated. Although the error components approach to estimation and decomposition of economic efficiency has proved intractable so far, the shadow price approach has proved successful and has become very popular. It is also possible to combine the two approaches, by modeling technical efficiency as an error component, and modeling allocative efficiency parametrically. Kumbhakar and Lovell (2000) discuss estimation strategies for the pure shadow price model and the combined model.

When modeling the behavior of producers who are constrained in their pursuit of a conventional objective, or who pursue an unconventional objective, analysts have two choices. The preferred choice is to model objective and constraint(s) correctly, derive the first order conditions, and construct an estimating model based on the assumption that producers are efficient. This can be hard work, as Färe and Logan (1983) have demonstrated for the case of the profit-seeking rate-of-return regulated producer. An easier alternative approach, illustrated above, is to model such producers as being unconstrained in their

pursuit of a conventional objective, allow for failure to satisfy first order conditions, and check to see if the direction of the estimated allocative inefficiency is consistent with what one would expect if in fact the producers were constrained or pursued some other objective. That is, use a model that is inappropriate but familiar, and look for allocative inefficiency by comparing shadow price ratios with actual price ratios.

In a related situation the analyst does not know the constraints or the objective of producers, perhaps because there are competing paradigms at hand. In this case it is feasible to use the familiar model and use estimated shadow prices to provide an indirect test of the competing paradigms.

These are the two purposes that the shadow price approach most frequently serves. Thus, allocative inefficiency in the unconstrained pursuit of cost minimization or profit maximization suggests allocative efficiency in a more complicated environment, and departures of shadow price ratios from actual price ratios provide the basis for hypothesis tests. The model has been used frequently to test the Averch-Johnson hypothesis, and more generally as a framework for testing allocative efficiency hypotheses in a wide variety of contexts. Two other examples come to mind, primarily because they are current and have not yet been subjected to analysis using the shadow price approach. The impact of domestic content legislation could be explored within the shadow price framework. Another popular hypothesis that could be tested within this framework is that of discrimination, against minorities or immigrants or whatever group is of interest.

1.6 The Mathematical Programming Approach to Efficiency Measurement

The mathematical programming approach to the construction of frontiers and the measurement of efficiency relative to the constructed frontiers goes by the descriptive title of data envelopment analysis, with interesting acronym DEA. It truly does envelop a data set; it makes no accommodation for noise, and so does not "nearly" envelop a data set the way the deterministic kernel of a stochastic frontier does. Moreover, subject to certain assumptions about the structure of production technology, it envelops the data as tightly as possible.

Like the econometric approach, the programming approach can be categorized according to the type of data available (cross-section or panel), and according to the types of variables available (quantities only, or quantities and prices). With quantities only, technical efficiency can be estimated, while with quantities and prices economic efficiency can be estimated and decomposed into its technical and allocative components. However DEA was developed in a public sector, not-for-profit environment, in which prices are suspect at best and missing at worst. Consequently the vast majority of DEA studies use quantity data only

and estimate technical efficiency only, despite the fact that the procedures are easily adapted to the estimation of economic efficiency in a setting in which prices are available and reliable.

In Section 1.6.1 we analyze plain vanilla DEA to estimate technical efficiency. In Section 1.6.2 we discuss one of many possible DEA models of economic efficiency. In Section 1.6.3 we discuss the application of DEA to panel data, although the most popular such application occurs in the analysis of productivity change, which we discuss in Section 1.8.3. In Section 1.6.4 we discuss a technical issue, the imposition of weight restrictions, which has important economic implications. Finally in Section 1.6.5 we offer a brief introduction to the statistical foundations of DEA, a subject covered more fully in Chapter 4.

1.6.1 Basic DEA

Producers use inputs $x \in R^N_+$ to produce outputs $y \in R^M_+$. The research objective is to estimate the performance of each producer relative to best observed practice in a sample of $i = 1,\dots,I$ producers. To this end weights are attached to each producer's inputs and outputs so as to solve the problem

$\min_{\upsilon,\mu} \quad \upsilon^T x_o / \mu^T y_o$
subject to $\quad \upsilon^T x_i / \mu^T y_i \geqq 1, \quad i = 1,\dots,o,\dots,I$
$\upsilon,\mu \geqq 0$

(1.52)

Here (x_o, y_o) are the vectors of inputs and outputs of the producer under evaluation, and (x_i, y_i) are the vectors of inputs and outputs of the ith producer in the sample. The problem seeks a set of nonnegative weights, or multipliers, that minimize the weighted input-to-output ratio of the producer under evaluation, subject to the constraints that when these weights are assigned to every producer in the sample, their weighted input-to-output ratios are bounded below by one. Associate the multipliers (υ,μ) with shadow prices, and think of the objective in the problem as one of minimizing the ratio of shadow cost to shadow revenue.

The nonlinear program (1.52) can be converted to a dual pair of linear programs. The first DEA model is known as the CCR model, after Charnes *et al.* (1978). The "multiplier" program appears in the right panel of (1.53), where X is an N×I sample input matrix with columns of producer input vectors x_i and Y is an M×I sample output matrix with columns of producer output vectors y_i. Think of the multiplier program as one of minimizing shadow cost, subject to the constraint that shadow revenue is normalized to one, and subject to the constraints that

when these multipliers are assigned to all producers in the sample, no producer earns positive shadow profit.

CCR Envelopment Program	CCR Multiplier Program
max(,((	min(,((Txo
subject to $X(\leqq xo$	subject to $(Tyo = 1$
$(yo \leqq Y\lambda$	$\upsilon^T X - \mu^T Y \geqq 0$
$\lambda \geqq 0$	$\upsilon, \mu \geqq 0$

$$(1.53)$$

Because the multiplier program is a linear program, it has a dual, which is also a linear program. The dual "envelopment" program appears in the left panel of (1.53), where ϕ is a scalar and λ is an $I \times 1$ intensity vector. In the envelopment program the performance of a producer is evaluated in terms of its ability to expand its output vector subject to the constraints imposed by best practice observed in the sample. If radial expansion is possible for a producer, its optimal $\phi > 1$, while if radial expansion is not possible, its optimal $\phi = 1$. Noting the output orientation of the envelopment program, it follows that ϕ is the DEA estimator of $TE_o(x,y)$ defined in (1.12). Noting that ϕ is a radial efficiency measure, and recalling the divergence between Koopmans' definition of technical efficiency and the Debreu - Farrell measure of technical efficiency, it follows that optimal $\phi = 1$ is necessary, but not sufficient, for technical efficiency since $(\phi y_o, x_o)$ may contain slack in any of its M+N dimensions. At optimum, $\phi = 1$ characterizes technical efficiency in the sense of Debreu and Farrell, while $\{\phi = 1, X\lambda = x_o, \phi y_o = Y\lambda\}$ characterizes technical efficiency in the sense of Koopmans.

The output-oriented CCR model is partly illustrated in Figure 1.14, for the M=2 case. Producer A is technically inefficient, with optimal projection $\phi^A y^A$ occurring at a convex combination of efficient producers D and C on the output isoquant $I^{CCR}(x)$, and so $\lambda^D > 0$, $\lambda^C > 0$ with all other elements of the intensity vector being zero. The efficient role models D and C are similar to, and a linear combination of them is better than, inefficient producer A being evaluated. The envelopment program provides this information. The multiplier program provides information on the trade-off between the two outputs at the optimal projection. The trade-off is given by the optimal shadow price ratio $-(\mu_1/\mu_2)$. The fact that this shadow price ratio might differ from the market price ratio, if one exists, plays a role in the DEA model of economic efficiency in Section 1.6.2. The multiplier program also provides information on input trade-offs $-(\upsilon_n/\upsilon_k)$ and output-input trade-offs (μ_m/υ), although this information is not portrayed in Figure 1.14.

Figure 1.14 The Output-Oriented CCR Model
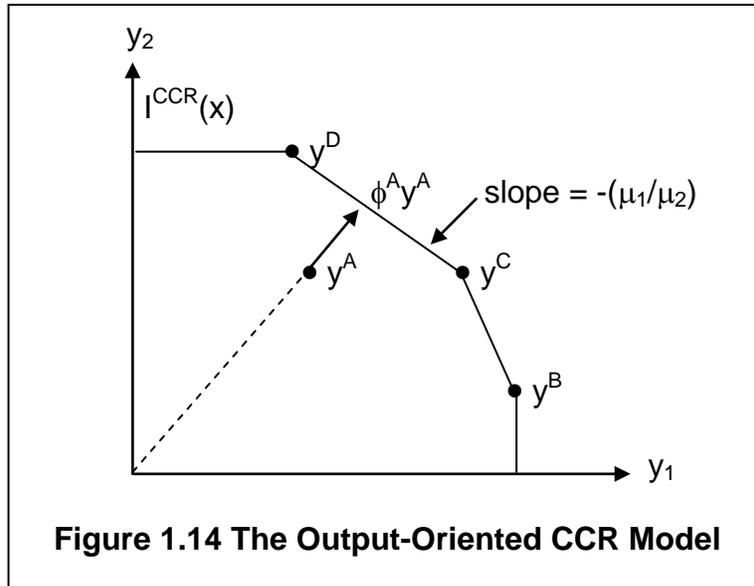
Problem (1.53) is solved I times, once for each producer in the sample, to generate I optimal values of (ϕ, λ) and I optimal values of (υ, μ). It thus provides a wealth of information about the performance of each producer in the sample, and about the structure of production technology.

The CCR production set corresponding to T in (1.1) is obtained from the envelopment problem in (1.53) as $T^{CCR} = \{(y,x): y \leqq Y\lambda, X\lambda \leqq x, \lambda \geqq 0\}$, and imposes three restrictions on the technology. These restrictions are constant returns to scale, strong disposability of outputs and inputs, and convexity. Each of these restrictions can be relaxed.

Constant returns to scale is the restriction that is most commonly relaxed. Variable returns to scale is modeled by adding a free variable υ_o to the multiplier program, which is equivalent to adding a convexity constraint $\Sigma_i\lambda_i = 1$ to the envelopment program. The variable returns to scale model was introduced by Afriat (1972), but is better known as the BCC model after Banker *et al.* (1984). The BCC envelopment and multiplier programs become

BCC Envelopment Program	BCC Multiplier Program
$\max_{\phi,\lambda}\quad \phi$	$\min_{\upsilon,\upsilon o,\mu}\quad \upsilon^T x_o + \upsilon_o$
subject to $\quad X\lambda \leqq x_o$	subject to $\qquad \mu^T y_o = 1$
$\phi y_o \leqq Y\lambda$	$\upsilon^T X + \upsilon_o - \mu^T Y \geqq 0$
$\lambda \geqq 0, \Sigma_i\lambda_i = 1$	$\upsilon,\mu \geqq 0, \upsilon_o$ free

(1.54)

The interpretation of the BCC envelopment and multiplier programs is essentially the same as for the CCR model, but the BCC production set shrinks, becoming $T^{BCC} = \{(y,x): y \leqq Y\lambda, X\lambda \leqq x, \lambda \geqq 0, \Sigma_i\lambda_i = 1\}$. T^{BCC} exhibits variable returns to scale, because only convex combinations of efficient producers form the best practice frontier. For this reason it envelops the data more tightly than T^{CCR} does.

The difference between the two production sets is illustrated in Figure 1.15. Because T^{BCC} envelops the data more tightly than T^{CCR} does, efficiency estimates are generally higher with a BCC specification, and rankings can differ in the two specifications. As in the CCR model, the BCC envelopment program provides efficiency estimates and identifies efficient role models. Also as in the CCR model, the BCC multiplier program estimates optimal shadow price ratios, but it also provides information on the nature of scale economies. The optimal projection to T^{BCC} occurs at $(\phi y_o, x_o)$. At this projection the output-input trade-off is μ/υ. The vertical intercept of the supporting hyperplane $y = \upsilon_o + \upsilon x_o$ at $(\phi y_o, x_o)$ is positive. This indicates decreasing returns to scale at $(\phi y_o, x_o)$, which should be apparent from Figure 1.15. More generally, $\upsilon_o \lesseqgtr 0$ signals that a producer is operating in a region of increasing, constant or decreasing returns to scale.

Figure 1.15 Returns to Scale in DEA
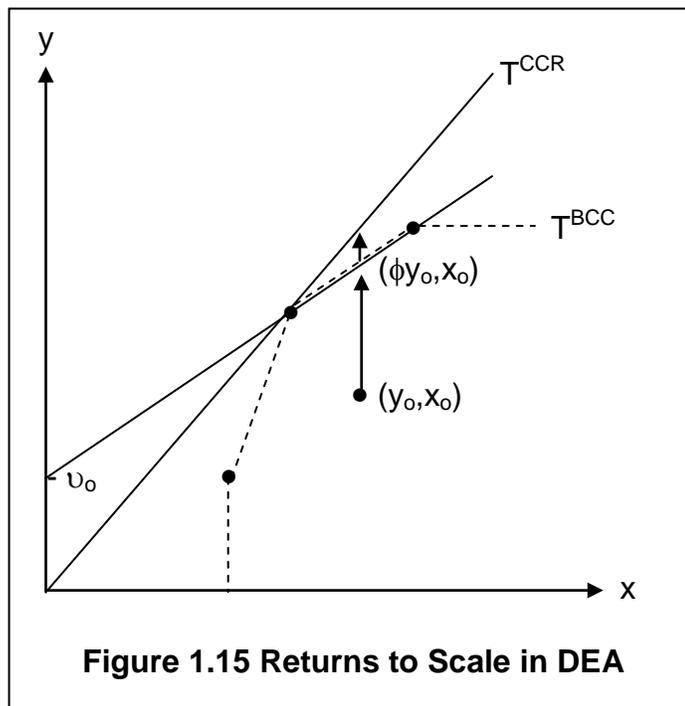
Notice the shape of T^{BCC} in Figure 1.15. Requiring strictly positive input to produce nonzero output is a consequence of not allowing for the possibility of inactivity, and of imposing convexity on T^{BCC}. This creates a somewhat strained notion of variable returns to scale, one that is well removed from the classical S-shaped production frontier that reflects Frisch's (1965) "ultra-passum" law. Petersen (1990) has attempted to introduce more flexibility into the DEA approach to measuring scale economies by dispensing with the assumption of convexity of T, while maintaining the assumption of convexity of L(y) and P(x).

The CCR and BCC models differ in their treatment of scale economies, as reflected by the additional equality constraint $\Sigma_i \lambda_i = 1$ and free variable υ_o in the BCC model. Just as (μ, υ) are shadow prices of outputs and inputs, υ_o is the shadow value of the convexity constraint $\Sigma_i \lambda_i = 1$. It is possible to conduct a test of the null hypothesis that $\upsilon_o = 0$, or that the convexity constraint $\Sigma_i \lambda_i = 1$ is redundant. This is a test for constant returns to scale, and is discussed along with other hypothesis tests in Chapter 4. However a qualification is in order concerning the interpretation of the multipliers. Most efficient producers are located at vertices, and it is possible that some inefficient producers are projected to vertices. At vertices shadow prices of variables (υ, μ) in the CCR and BCC models, and of the convexity constraint (υ_o) in the BCC model, are not unique.

The CCR and BCC envelopment programs are output-oriented, just as the econometric problem (1.32) is. It is a simple matter to obtain analogous input-oriented envelopment programs, by converting the envelopment programs to minimization programs and converting the multiplier problems to maximization programs. Details appear in Chapter 3. The choice between the two orientations depends on the objective assigned to producers. If producers are required to meet market demands, and if they can freely adjust input usage, then an input orientation is appropriate.

The assumption of strong disposability is rarely relaxed, despite the obvious interest in relaxing the free disposability of surplus inputs or unwanted outputs. One popular exception occurs in environmental economics, in which producers use purchased inputs to produce marketed outputs and undesirable byproducts like air or water pollution. In this case the byproducts may or may not be *privately* freely disposable, depending on whether the regulator is watching, but they are surely *socially* weakly or expensively disposable. The value of relaxing the strong output disposability assumption lies in its potential to provide evidence on the marginal private cost of abatement. This evidence can be compared with estimates of the marginal social benefit of abatement to inform public policy.

Without going into details, which are provided by Färe *et al.* (1989, 1993) and a host of subsequent writers, the essence of weak disposability is captured in Figure 1.16. Here y_2 is a marketed output and y_1 is an undesirable byproduct.

A conventional output set exhibiting strong disposability is bounded by the output isoquant $I^S(x)$ with solid line segments. The corresponding output set exhibiting weak disposability of the byproduct is bounded by the output isoquant $I^W(x)$ with dashed line segments. $L^W(x) \subseteq L^S(x)$, and that part of $L^S(x)$ not included in $L^W(x)$ provides an indication of the amount of marketed output foregone if the byproduct is not freely disposable. Disposal is free with technology $L^S(x)$, and abatement is costly with technology $L^W(x)$. For $y_1 < y_1^*$ the conventional strong disposal output set allows abatement of y_1 to be privately free, as indicated by the horizontal solid line segment along which $(\mu_1/\mu_2) = 0$. In contrast, the weak disposal output set makes abatement privately costly, as indicated by the positively sloped dashed line segments to the left of y_1^*. Moreover, increased abatement becomes increasingly costly, since the shadow price ratio $(\mu_1/\mu_2) > 0$ increases with additional abatement.


Figure 1.16 Weak Disposability of y_1
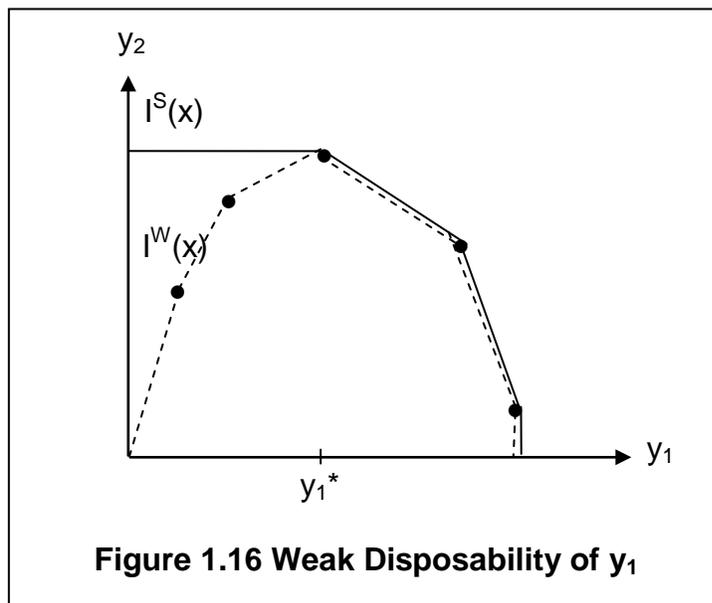
In Figure 1.16 the marginal cost of abatement is reflected in the amount of y_2 (and hence revenue) that must be sacrificed to reduce the byproduct. With given inputs and technology, reducing air pollution requires a reduction in electricity generation. Allowing x or technology to vary would allow the cost of abatement to reflect the additional input or the new technology (and hence cost) required to abate with no loss in marketed output. With given electricity generation, reducing air pollution could be accomplished by installing scrubbers or by upgrading technology.

The assumption of convexity of output sets P(x) and input sets L(y) also is rarely relaxed, despite the belief of many, expressed by McFadden (1978;8-9), that its importance lies more in its analytical convenience than in its technological

realism. In the previous context of scale economies, feasibility of an activity (y,x) does not necessarily imply feasibility of all scaled activities $(\lambda y, \lambda x)$, $\lambda > 0$, which motivates relaxing the assumption of constant returns to scale. In the present context feasibility of two distinct activities (y^A, x^A) and (y^B, x^B) does not necessarily imply feasibility of all convex combinations of them, which motivates relaxing the assumption of convexity.

Deprins *et al.* (1984) were the first to relax convexity. They constructed a "free disposal hull" (FDH) of the data that relaxes convexity while maintaining strong disposability and allowing for variable returns to scale. An FDH output set is contrasted with a BCC output set in Figure 1.17. The BCC output set is bounded by the output isoquant $I^{BCC}(x)$ as indicated by the solid line segments. The FDH output set dispenses with convexity but retains strong disposability, and is bounded by the output isoquant $I^{FDH}(x)$ as indicated by the dashed line segments. The contrast between FDH and DEA input sets and production sets is structurally identical. In each case dispensing with convexity creates frontiers that have a staircase shape. This makes slacks a much more serious problem in FDH than in DEA, and it complicates the FDH multiplier program.

Figure 1.17 An FDH Output Set
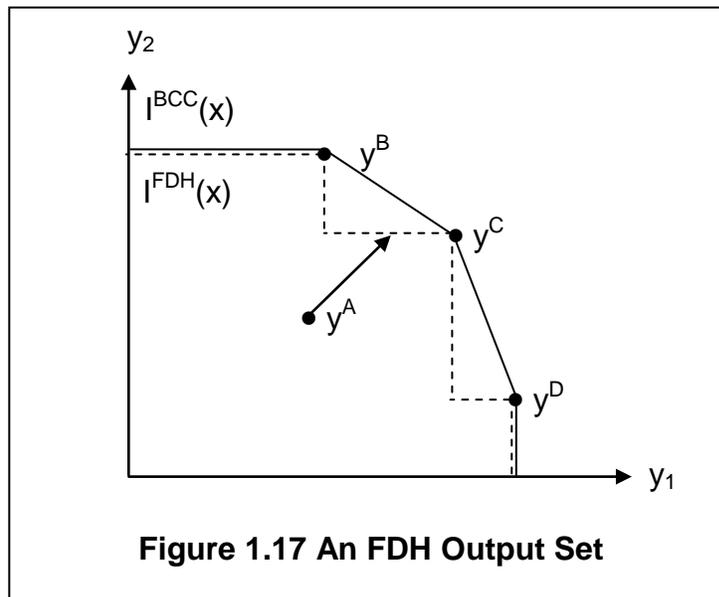
The FDH envelopment program is identical to the BCC envelopment program in (1.54), apart from the addition of an integral constraint $\lambda_i \in \{0,1\}$, i = 1,...,I. Since all intensity variables are assigned values of zero or one, the convexity constraint $\Sigma_i \lambda_i = 1$ implies that exactly one intensity variable has a value of one. Thus FDH identifies exactly one role model for an inefficient producer, and the role model is an actual efficient producer rather than a fictitious convex combination of efficient producers. In Figure 1.17 inefficient producer A

receives an FDH radial efficiency estimate indicated by the arrow, and has efficient role model C rather than a convex combination of C and B as in DEA. The additional constraint in the FDH envelopment program causes $P^{FDH}(x) \subseteq P^{BCC}(x)$, and so FDH efficiency estimates are generally higher than BCC efficiency estimates. Although the addition of an integral constraint converts (1.54) to a more complicated mixed integer program, it actually simplifies the computation of the envelopment program. In fact, programming techniques are not required to obtain FDH efficiency estimates. Tulkens (1993) provides details.

We closed Section 1.3 by bemoaning the neglect of dominance. Although dominance information provides a useful complement to any type of efficiency analysis, it is popular only in FDH efficiency analysis. The vector comparison tools that are used to identify the single efficient role model also serve to identify all dominating producers for each inefficient producer, and all dominated producers for each efficient producer. Identifying dominating producers enhances the likelihood that an inefficient producer can find a useful role model, fully efficient or not. Identifying the number of dominated producers also offers a procedure for ranking ostensibly efficient producers, a problem that has engaged researchers ever since Andersen and Petersen (1993) first raised the issue. The problem is addressed in Chapter 4.

At the end of Section 1.5.1 we discussed procedures for incorporating potential determinants of efficiency in SFA. The same challenge arises in DEA, and at least two approaches have been developed. One approach is to add to the CCR envelopment program (1.53) or the BCC envelopment program (1.54) the additional constraints $Z\lambda \leqq z_o$ or $Z\lambda \geqq z_o$, or a combination of the two, depending on whether the potential determinants enhance or retard output. This of course requires appropriate modification of the dual multiplier programs. This approach is analogous to replacing $f(x_i;\beta)$ with $f(x_i,z_i;\beta,\gamma)$ in SFA. Two difficulties arise. First, unlike SFA, in which we simultaneously estimate both the magnitudes and the signs of the elements of γ, here we must know in advance the direction of the influence of the elements of z in order to set the inequalities. Second, in this formulation elements of z either enhance or retard output, but they do not influence the efficiency with which x produces y. The other approach, far more popular, is to regress estimated efficiency scores against z in a second stage regression. We have already warned of the sins committed by doing so in SFA, and the story is similar, but happily not quite so bleak, in DEA. Chapter 4 provides a detailed analysis of the use of second stage regressions in DEA. The message is of bad news - good news form: (i) statistical inference on the second stage regressions you have seen (or conducted) is invalid; although (ii) it is possible to formulate the model in such a way that it provides a rational basis for regressing efficiency estimates in a second stage analysis, and bootstrapping can provide valid inference.

1.6.2 A DEA model of economic efficiency

The DEA models in Section 1.6.1 use quantity data only, and so capture technical efficiency only. In this Section we show how to extend DEA models to provide measures of economic efficiency. We continue our output orientation by illustrating the extension with a problem of revenue maximization.

Producers are assumed to use inputs $x \in R^N_+$ to produce outputs $y \in R^M_+$ for sale at prices $p \in R^M_{++}$. Their objective is to maximize revenue, subject to the constraints imposed by output prices, input supplies and the structure of production technology, which is allowed to exhibit variable returns to scale. This problem can be expressed in linear programming format as

Revenue Maximization Program
$r(x,p) = \max_{y,\lambda} \quad p^T y$
subject to $\quad X\lambda \leqq x_o$
$y \leqq Y\lambda$
$\lambda \geqq 0, \Sigma_i \lambda_i = 1$

(1.55)

The production set can be recognized as T^{BCC}, and so (1.55) is a straightforward extension of conventional DEA to an economic optimization problem. The problem is illustrated in Figure 1.18. The revenue efficiency of producer A is estimated from (1.55) as $RE(y^A,x,p) = p^T y^{RM}/p^T y^A > 1$. The technical efficiency of A is estimated from (1.54) as $TE_o(x,y^A) = p^T(\phi y^A)/p^T y^A = \phi > 1$. The output allocative efficiency of A is estimated residually as $AE_o(y^A,x,p) = p^T y^{RM}/p^T(\phi y^A) > 1$. Notice that at the optimal projection ϕy^A the estimated shadow price ratio $(\mu_1/\mu_2) < (p_1/p_2)$. This provides an indication of the existence, and the direction, of a misallocation of outputs; the output mix $(y_2/y_1)^A$ is too large, given (p_1/p_2). The cost of this misallocation, in terms of lost revenue, is estimated as $[p^T y^{RM} - p^T(\phi y^A)] > 0$.

Figure 1.18 Revenue Maximization in DEA
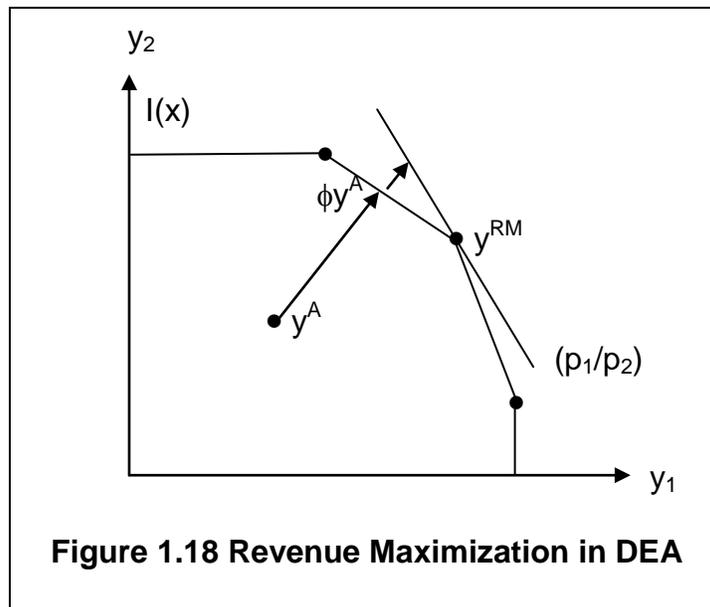
Alternative objectives, and alternative or additional constraints, can be entertained within the same general linear programming format. All that is required is the requisite data and the ability to write down a linear programming problem analogous to (1.55) that captures the objective and the constraints of the economic problem of interest. Färe *et al.* (1985) analyze several economic optimization problems with linear programming techniques.

1.6.3 Panel data

Thus far in Section 6 we have assumed that we have a single cross-section of data with which to evaluate producer performance. Suppose now that we have a panel consisting of T time periods and I producers in each period. How does DEA exploit the ability to observe each producer multiple times? The available techniques are not as sophisticated as panel data econometric techniques, but several options are available.

One option is to pool the data and estimate a single grand frontier. In doing so this option assumes an unvarying best practice technology, which may be untenable in long panels. It does, however, generate T efficiency estimates for each producer, all against the same standard, and trends in efficiency estimates of individual producers may be of interest.

At the other extreme, it is possible to estimate T separate frontiers, one for each period. This allows for technical progress and regress. It also allows for intersecting frontiers, which would signal local progress in a region of output-

input space and local regress in another region. A danger of this approach is the possibility of excessive volatility in efficiency scores resulting from excessive variation in temporally independent period frontiers.

An intermediate option is to estimate a sequence of overlapping pooled panels, each consisting of a few time periods of arbitrary length. Known as "window analysis," this option tracks efficiency trends through successive overlapping windows. One purpose of window analysis is to relieve degrees of freedom pressure when M+N is large relative to I. As such it provides a compromise between running DEA once on one large I×T pooled panel and running DEA T times on T small cross sections. Another objective of window analysis is to alleviate volatility in efficiency estimates.

A second intermediate option is to estimate a sequential frontier by continuously adding data from successive time periods. In the end, this procedure constructs a grand frontier, but prior to the terminal period frontiers are estimated sequentially from current and all previous (but not subsequent) data. This option rules out the possibility of technical regress, presumably in the belief that techniques once known are not forgotten, and remain available for adoption. A drawback of this option is that sample sizes increase sequentially, which complicates statistical inference.

A final option is to use two adjacent periods of data at a time, beginning with periods 1 and 2, continuing with periods 2 and 3, and so on. This option may look like two-period window analysis, but it is very different. It is used to estimate and decompose Malmquist indexes of productivity change, and we defer discussion of this option to Section 1.7.

1.6.4 Weight restrictions

The multipliers (υ,μ) in the CCR problem (1.53) and the BCC problem (1.54) are not market prices. Indeed a frequently touted virtue of DEA is that it is practical in situations in which market prices are missing, as in the environmental context illustrating weak disposability in Section 1.6.1. The multipliers are in fact endogenously determined *shadow prices* revealed by individual producers in their effort to maximize their relative efficiency. The great Russian mathematician (and, together with Koopmans, the 1975 recipient of the Nobel Prize in Economic Sciences) Kantorovich referred to them as "resolving multipliers," ostensibly because they solve the dual linear programs. As Figure 1.14 illustrates, different producers can choose different sets of shadow price ratios, and the freedom to choose is limited only by the nonnegativity constraints $\upsilon,\mu \geqq 0$. Consequently the range of multipliers chosen by producers might differ markedly from market prices (when they exist), or might offend expert judgement on the relative values of the variables (when market prices are missing). This opens up the possibility of limiting the freedom to choose.

At the other extreme, many comparisons are based on fixed weights that give no freedom to choose. A recent example is provided by the World Health Organization (2000), which somewhat controversially evaluated the ability of 191 member countries to provide health care to their citizens. WHO used five health care indicators y_m, to which they assigned fixed weights $\mu_m > 0$, $\Sigma_m\mu_m = 1$. These fixed weights were based on expert opinion, but they were common to all countries, regardless of their development status. To require Mali and Canada, for example, to assign equal importance to each indicator seems undesirably restrictive. Why not retain the five indicators, but give countries the freedom to choose their own weights, subject to the requirements that $\mu_m \geqq 0$, $\Sigma_m\mu_m = 1$?

This is exactly what DEA does. Lauer *et al.* (2004) ran the output-oriented DEA program

$\max_{\phi,\lambda} \quad \phi$	$\min_{\mu,\omega} \quad \omega$
subject to	subject to $\quad \mu^T y_o = 1$
$\phi y_o \leqq Y\lambda$	$-\mu^T Y + \omega \geqq 0$
$\lambda \geqq 0, \Sigma_i\lambda_i = 1$	$\mu \geqq 0, \Sigma_m\mu_m = 1, \omega$ free

$$(1.56)$$

which is the BCC model (1.54) with scalar input with unit value for each country (each country is "itself"). Each country is allowed to select its own nonnegative health care indicator weights.

The results were unacceptable. Over one-third of the countries assigned a zero weight to four of five indicators, and nearly 90% of the countries assigned a zero weight to y_1 = population health, the defining goal of any health system. Only three countries assigned positive weights to all five indicators. The fixed positive weights common to all countries used by WHO are unappealing, but so is the excessive variability of self-assigned nonnegative weights allowed by DEA. A frequently touted virtue of DEA, that it is value-free in its selection of weights, can turn out to be a vice.

This is not an isolated incident. Imposing only the restrictions that output and input weights be nonnegative can, and frequently does, generate silly weights and implausible weight variability, both of which offend common sense. Fortunately a remedy exists. It is possible to allow weight flexibility, and at the same time to restrict weight flexibility. This was the fundamental insight of Thompson *et al.* (1986). They were forced to figure out how to impose weight restrictions in their DEA study of identifying an optimal site to place a high-energy physics facility. Their motivation was that sites have characteristics, and in the opinion of experts no characteristic could be ignored by assigning it zero weight. Necessity is the mother of invention.

The DEA literature on imposing weight restrictions has come a long way since 1986, and there exist many ways of restricting weights. One appealing procedure is to append to the multiplier program of (1.56) the restrictions

$$\gamma_m \geq \mu_m y_m / \mu^T y \geq \beta_m, \quad m = 1,\ldots,5, \tag{1.57}$$

which place lower and upper bounds on the relative importance of each indicator in the evaluation of health care performance. Although these bounds are common to all countries, they do allow limited freedom to choose.

More generally it is possible to impose restrictions on output weights, on input weights, and on the ratio of output weights to input weights, in the BCC model (1.54). The appeal of the procedure is that it offers a compromise between the arbitrary imposition of common weights and the excessively flexible DEA weights. Of course we still need experts to set the bounds, and experts frequently disagree. A fascinating example, related by Takamura and Tone (2003), is occurring in Japan, whose government plans to move several agencies out of congested Tokyo. Ten candidate sites have been identified, and 18 criteria have been specified. Criteria vary in their importance, and a committee of wise men has been named to establish bounds of the form (1.57) on each criterion, and these bounds must reflect differences of opinion among the wise men. DEA with weight restrictions is being used to solve a 12 trillion yen problem!

Of course the problem of unreasonable shadow prices is not limited to DEA; it can arise in SFA as well. The problem has received far more attention in DEA, where it is arguably easier to resolve. For more on weight restrictions in DEA, see Chapter 3.

1.6.5 Statistical foundations of DEA

A distinguishing feature of the DEA models discussed above is that they do not contain a random error term that would incorporate the impacts of statistical noise; the DEA frontier is not stochastic as it is in SFA. This has led to two separate strands of research.

Land *et al.* (1993) and Olesen and Petersen (1995) sought to make DEA stochastic by introducing a chance that the constraints (on either the envelopment problem or the multiplier problem) in either (1.53) or (1.54) might be violated with some probability. This approach is an extension to DEA of chance-constrained programming developed by Charnes *et al.* (1958) and Charnes and Cooper (1959), and is known as "chance-constrained DEA."

We follow Land *et al.* (1993) by writing the CCR envelopment problem in (1.53) as

61

Chance-Constrained CCR Envelopment Program
$\max_{\phi\lambda} \quad \phi$
subject to $\quad \Pr[X\lambda \leqq x_o] \geqq 0.95$
$\Pr[(y_o \leqq Y\lambda] \geqq 0.95$
$\lambda \geqq 0$

$$(1.58)$$

where the pre-specified probabilities of satisfying each constraint are assumed equal at the popular 95% level to simplify the exposition. Program (1.58) asks producers to radially expand their output vector as far as possible, subject to the constraint that ($\phi y_o, x_o$) "probably" is feasible.

Program (1.58) is not operational. It can be made operational by making assumptions on the distributions of the sample data. If it is assumed that each output y_{im} is a normally distributed random variable with expected value Ey_{im} and variance-covariance matrix $Vy_{im}y_{jm}$, and that each input is a normally distributed random variable with expected value Ex_{in} and variance-covariance matrix $Vx_{in}x_{jn}$, then (1.58) can be expressed in modified certainty-equivalent form as

Chance-Constrained CCR Envelopment Program: Certainty Equivalent Form
$\max_{\phi,\lambda} \quad \phi$
subject to $\sum_i x_{in}\lambda_i + \sum_i (Ex_{in} - x_{in})\lambda_i + 1.645[\sum_i\sum_j \lambda_i\lambda_j Vx_{in}x_{jn}]^{1/2} \leqq x_{on}$, n=1,…,N
$\phi y_{om} \leqq \sum_i y_{im}\lambda_i + \sum_i (Ey_{im} - y_{im})\lambda_i - 1.645[\sum_i\sum_j \lambda_i\lambda_j Vy_{im}y_{jm}]1/2$, m=1,…,M
$\lambda_I \geqq 0$, i=1,…,I

$$(1.59)$$

where 1.645 = F-1(0.95) is the pre-specified value of the distribution function of a standard normal variate. If x_{in} - Ex_{in} = y_{im} - Ey_{im} = $Vx_{in}x_{jn}$ = $Vy_{im}y_{jm}$ = 0 for all producers i and j and for all variables m and n, the nonlinear program (1.59) collapses to the linear program (1.53). However if we have reason to believe that a sample data point departs from its expected value, perhaps due to unusually good weather or unexpected supply disruption, then this information is fed into the chance-constrained program. The desired outcome is that, unlike program (1.53), good or bad fortune does not distort efficiency measures for any producer. Similarly, if we have reason to believe that any pair of inputs or outputs is correlated across producers, perhaps because farmers in the same region experience similar weather patterns, this information is also fed into the program.

The data requirements of chance-constrained efficiency measurement are severe. In addition to the data matrices X and Y, we require information on expected values of all variables for all producers, and variance-covariance

matrices for each variable across all producers. The idea is neat, and developments continue, but serious applications are few and far between.

There is another way of dealing with the absence of an explicit random error term in DEA. This is to acknowledge at the outset that DEA efficiency scores are estimators of true, but unknown, efficiencies. The properties of these estimators depend on the structure of the true, but unknown, technology, and also on the process by which the sample data have been generated, the DGP.

We know the DEA assumptions on the structure of the true technology. In a series of papers, Simar and Wilson and their colleagues have introduced assumptions on the DGP. This enables them to interpret the DEA efficiency measure as an estimator with statistical properties, thus endowing DEA with statistical foundations. In addition to convexity and strong disposability of the true, but unknown, technology, they make the following assumptions on the DGP:
- the sample data (x_i,y_i), i=1,…,I, are realizations of iid random variables with probability density function f(x,y);
- the probability of observing an efficient unit $[\phi(x,y) = 1]$ approaches unity as the sample size increases;
- for all (x,y) in the interior of T, $\phi(x,y)$ is differentiable in (x,y).

Armed with these assumptions on the DGP, it is possible to prove that
- the DEA efficiency estimator $\phi^{DEA}(x,y)$ is biased toward unity;
- but $\phi^{DEA}(x,y)$ is a consistent estimator;
- although convergence is slow, reflecting the curse of dimensionality.

A closed form for the density of $\phi^{DEA}(x,y)$ has yet to be derived. Consequently bootstrapping techniques must be used to approximate it in order to conduct statistical inference. A sobering message emerges from the bootstrapping exercises we have seen. DEA efficiency estimates are frequently used to compare the performance of one producer, or one group of producers, to another. However bootstrapping tends to generate confidence intervals that are sufficiently wide to question the reliability of inferences drawn from such comparisons. This message mirrors that of the relatively wide confidence intervals surrounding SFA efficiency estimates.

In Chapter 4 Simar and Wilson provide the analytical details, they explain why and how to bootstrap, and they discuss hypothesis testing.


1.7 Malmquist Productivity Indexes

Throughout this Chapter, and particularly in Section 1.3, we have associated distance functions with efficiency measures. We now show how distance

functions also constitute the building blocks for a measure of productivity change. The story begins with Malmquist (1953), who introduced the input distance function in the context of consumption analysis. His objective was to compare alternative consumption bundles. He did so by developing a standard of living (or consumption quantity) index as the ratio of a pair of input distance functions. In the context of production analysis, Malmquist's standard of living index becomes an input quantity index. An analogous output quantity index is expressed as the ratio of a pair of output distance functions.

An obvious extension is to define a productivity index based on distance functions. Two such indexes have been developed, both bearing Malmquist's name even though he proposed neither one. One index is is defined as the ratio of an output quantity index to an input quantity index. The output quantity index is a ratio of output distance functions, and the input quantity index is a ratio of input distance functions. It provides a rigorous extension to multiple outputs and multiple inputs of the fundamental notion of productivity as the ratio of output to input discussed in Section 1.2. Caves *et al.* (1982b) mentioned and dismissed this index, which subsequently was introduced by Bjurek (1996). The other index uses only output distance functions or only input distance functions. In its output-oriented form it defines a productivity index as the ratio of a pair of output distance functions, and in its input-oriented form it defines a productivity index as the ratio of a pair of input distance functions. Caves *et al.* introduced this version of the Malmquist productivity index, and it is the subject of this Section because it is more popular than the Bjurek version.

Intuition is provided by Figure 1.19, in which a producer's input and output are depicted in two adjacent periods. It is obvious that productivity has increased, since $(y^{t+1}/x^{t+1}) > (y^t/x^t)$ or, equivalently, $(y^{t+1}/y^t) > (x^{t+1}/x^t)$. The challenge is to quantify productivity growth. A Malmquist productivity index does so by introducing the period t technology T_c^t as a benchmark, and by comparing the distances of (y^{t+1},x^{t+1}) and (y^t,x^t) to T_c^t. Distance can be measured vertically, with an output expanding orientation, or horizontally, with an input conserving orientation, depending on the orientation of producers. The ratio of these two distances provides a quantitative measure of productivity change, which in Figure 1.19 is greater than unity with either orientation.

Figure 1.19 The Malmquist Productivity Index I
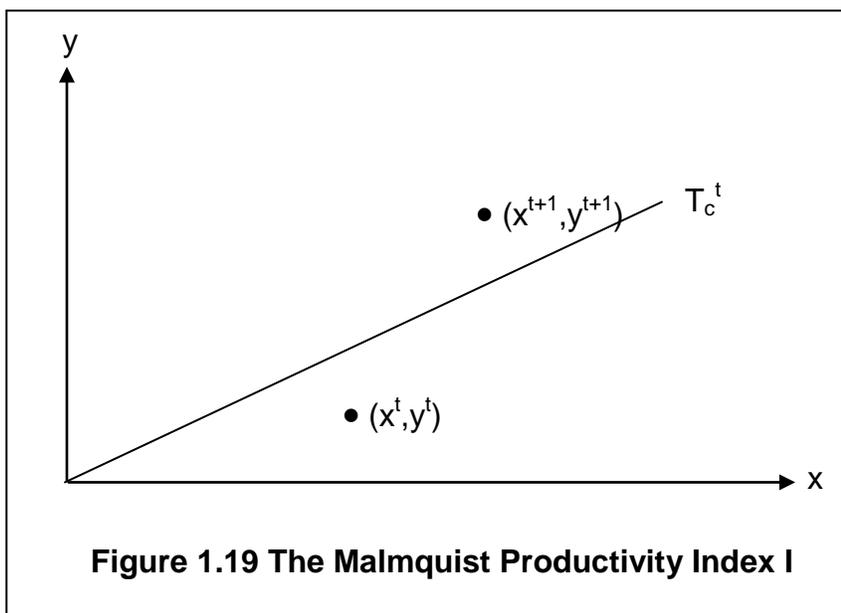
This raises the question of how to specify the period t technology. Caves *et al.* defined their index on a technology that allowed for varying returns to scale. However Grifell-Tatjé and Lovell (1995) showed, by way of a simple numerical example, that this convention creates an index that ignores the contribution of scale economies to productivity growth. Färe and Grosskopf (1996) proved that if M=N=1, the Caves *et al.* index provides an accurate measure of productivity change in the sense that it equals $(y^{t+1}/y^t)/(x^{t+1}/x^t)$ if, and only if, the index is defined on a technology exhibiting constant returns to scale. In light of these results we follow what is now common practice by defining the Caves *et al.* index on a *benchmark* technology satisfying constant returns to scale, which is to be distinguished from a *best practice* technology allowing for variable returns to scale. This convention enables the Malmquist productivity index to incorporate the influence of scale economies on productivity change, as a departure of the best practice technology from the benchmark technology. In the general M>1, N>1 case the influence of scale economies can be broadened to include the influence of changes in the output mix and changes in the input mix.

1.7.1 Definitions and properties

As in Section 1.3, let inputs $x \in R^N_+$ be used to produce outputs $y \in R^M_+$. The benchmark technology $T_c = \{(y,x): x \text{ can produce } y\}$ is the set of all technologically feasible output-input combinations, and is assumed to satisfy global constant returns to scale. The output set $P_c(x) = \{y: (x,y) \in T_c\}$ is the set of all technologically feasible output vectors given inputs x, with outer boundary

given by the output isoquant $I_c(x) = \{y \in P_c(x), \lambda y \notin P_c(x) \; \forall \; \lambda > 1\}$. The output distance function is defined on $P_c(x)$ as $D_{oc}(x,y) = \min\{\lambda: (y/\lambda) \in P_c(x)\}$.

Using the period t benchmark technology, the period t output-oriented Malmquist productivity index is written as

$$M_{oc}{}^t(x^t,y^t,x^{t+1},y^{t+1}) = \frac{D_{oc}{}^t\left(x^{t+1},y^{t+1}\right)}{D_{oc}{}^t\left(x^t,y^t\right)}. \tag{1.60}$$

$M_{oc}{}^t(x^t,y^t,x^{t+1},y^{t+1})$ compares (x^{t+1},y^{t+1}) to (x^t,y^t) by comparing their distances to the benchmark technology $T_c{}^t$. Although $D_{oc}{}^t(x^t,y^t) \leqq 1$ because (x^t,y^t) must be feasible for $T_c{}^t$, $D_{oc}{}^t(x^{t+1},y^{t+1}) \gtreqless 1$ because (x^{t+1},y^{t+1}) may or may not be feasible for $T_c{}^t$. Hence $M_{oc}{}^t(x^t,y^t,x^{t+1},y^{t+1}) \gtreqless 1$ according as productivity growth, stagnation or decline has occurred between periods t and t+1, from the forward-looking perspective of period t benchmark technology.

Using the period t+1 benchmark technology, the period t+1 output-oriented Malmquist productivity index is written as

$$M_{oc}{}^{t+1}(x^t,y^t,x^{t+1},y^{t+1}) = \frac{D_{oc}{}^{t+1}\left(x^{t+1},y^{t+1}\right)}{D_{oc}{}^{t+1}\left(x^t,y^t\right)}. \tag{1.61}$$

$M_{oc}{}^{t+1}(x^t,y^t,x^{t+1},y^{t+1})$ compares (x^{t+1},y^{t+1}) to (x^t,y^t) by comparing their distances to the benchmark technology $T_c{}^{t+1}$. Although $D_{oc}{}^{t+1}(x^{t+1},y^{t+1}) \leqq 1$ because (x^{t+1},y^{t+1}) must be feasible for $T_c{}^{t+1}$, $D_{oc}{}^{t+1}(x^t,y^t) \gtreqless 1$ because (x^t,y^t) may or may not be feasible for $T_c{}^{t+1}$. Hence $M_{oc}{}^{t+1}(x^t,y^t,x^{t+1},y^{t+1}) \gtreqless 1$ according as productivity growth, stagnation or decline has occurred between periods t and t+1, from the backward-looking perspective of period t+1 benchmark technology.

Both indexes compare (x^{t+1},y^{t+1}) to (x^t,y^t), but they use benchmark technologies from different periods. The choice of benchmark technology is arbitrary, and the two indexes are not necessarily equal except under restrictive neutrality conditions on technical change. Indeed one index may signal productivity growth and the other productivity decline. Consequently it is conventional to define the Malmquist productivity index as the geometric mean of the two, and to write it as

$$M_{oc}(x^t,y^t,x^{t+1},y^{t+1}) = \{[M_{oc}{}^t(x^t,y^t,x^{t+1},y^{t+1}) \times M_{oc}{}^{t+1}(x^t,y^t,x^{t+1},y^{t+1})]\}^{1/2}$$

$$= \left[\frac{D_{oc}{}^t\left(x^{t+1}, y^{t+1}\right)}{D_{oc}{}^t\left(x^t, y^t\right)} \times \frac{D_{oc}{}^{t+1}\left(x^{t+1}, y^{t+1}\right)}{D_{oc}{}^{t+1}\left(x^t, y^t\right)} \right]^{\frac{1}{2}}. \qquad (1.62)$$

$M_{oc}(x^t, y^t, x^{t+1}, y^{t+1}) \gtreqless 1$ according as productivity growth, stagnation or decline has occurred between periods t and t+1.

The two Malmquist productivity indexes are illustrated in Figure 1.20, with M=N=1. $M_{oc}{}^t(x^t, y^t, x^{t+1}, y^{t+1}) > 1$ (indicated by the solid arrows) because output has increased faster than input relative to the period t benchmark technology. This shows up in (1.60) as $D_{oc}{}^t(x^t, y^t) < 1$ and $D_{oc}{}^t(x^{t+1}, y^{t+1}) > 1$. $M_{oc}{}^{t+1}(x^t, y^t, x^{t+1}, y^{t+1}) > 1$ (indicated by the dotted arrows) because output has increased faster than input relative to the period t+1 benchmark technology. This shows up in (1.61) as $D_{oc}{}^{t+1}(x^t, y^t) < D_{oc}{}^{t+1}(x^{t+1}, y^{t+1}) < 1$. Consequently $M_{oc}(x^t, y^t, x^{t+1}, y^{t+1}) > 1$.

Figure 1.20 The Malmquist Productivity Index II
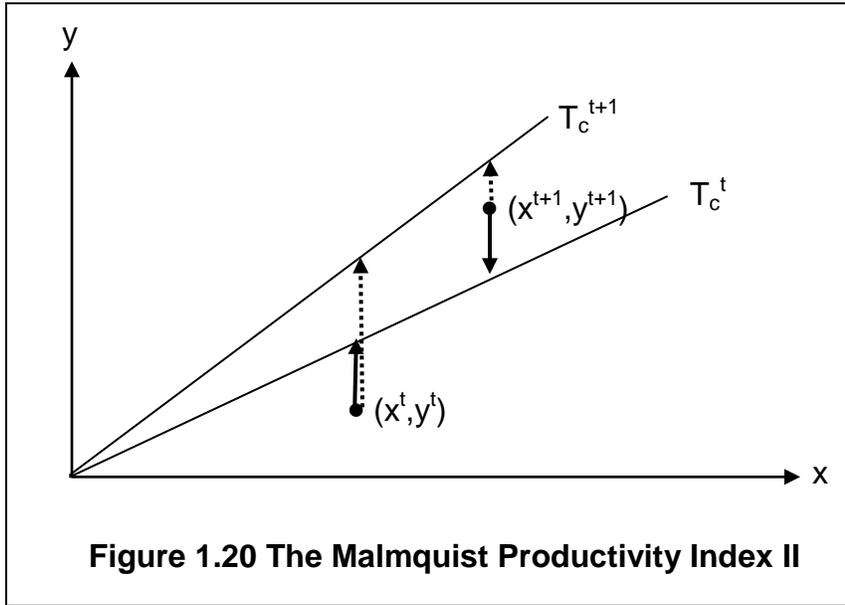
Because it is based on output distance functions, which satisfy a number of desirable properties, $M_{oc}(x^t, y^t, x^{t+1}, y^{t+1})$ also satisfies a number of properties. The Malmquist productivity index satisfies most of the following desirable properties, with failure indicated by an inequality:

M1: Weak Monotonicity

$$y'' \geq y' \Rightarrow M_{oc}(x^t, y^t, x^{t+1}, y'') \gtreqless M_{oc}(x^t, y^t, x^{t+1}, y')$$

$$y'' \geq y' \Rightarrow M_{oc}(x^t, y'', x^{t+1}, y^{t+1}) \lesseqgtr M_{oc}(x^t, y', x^{t+1}, y^{t+1})$$

$$x'' \geq x' \Rightarrow M_{oc}(x^t,y^t,x'',y^{t+1}) \lneqq M_{oc}(x^t,y^t,x',y^{t+1})$$

$$x'' \geq x' \Rightarrow M_{oc}(x'',y^t,x^{t+1},y^{t+1}) \gneqq M_{oc}(x',y^t,x^{t+1},y^{t+1})$$

M2: Homogeneity

$$M_{oc}(x^t,y^t,x^{t+1},\lambda y^{t+1}) = \lambda M_{oc}(x^t,y^t,x^{t+1},y^{t+1}), \ \lambda > 0$$
$$M_{oc}(x^t,\lambda y^t,x^{t+1},y^{t+1}) = \lambda^{-1} M_{oc}(x^t,y^t,x^{t+1},y^{t+1}), \ \lambda > 0$$
$$M_{oc}(x^t,\lambda y^t,x^{t+1},\lambda y^{t+1}) = M_{oc}(x^t,y^t,x^{t+1},y^{t+1}), \ \lambda > 0$$
$$M_{oc}(x^t,y^t,\lambda x^{t+1},y^{t+1}) = \lambda^{-1} M_{oc}(x^t,y^t,x^{t+1},y^{t+1}), \ \lambda > 0$$
$$M_{oc}(\lambda x^t,y^t,x^{t+1},y^{t+1}) = \lambda M_{oc}(x^t,y^t,x^{t+1},y^{t+1}), \ \lambda > 0$$
$$M_{oc}(\lambda x^t,y^t,\lambda x^{t+1},y^{t+1}) = M_{oc}(x^t,y^t,x^{t+1},y^{t+1}), \ \lambda > 0$$

M3: Proportionality

$$M_{oc}(x^t,y^t,x^{t+1},\mu y^t) \neq \mu, \ \mu > 0$$
$$M_{oc}(x^t,y^t,\lambda x^t,y^{t+1}) \neq \lambda^{-1}, \ \lambda > 0$$
$$M_{oc}(x^t,y^t,\lambda x^t,\mu y^t) = \mu/\lambda, \ \mu,\lambda > 0$$

M4: Identity

$$M_{oc}(x,y,x,y) = 1$$

M5: Commensurability (independence of units of measurement)

$$M_{oc}(\mu_1 x_1^t,...,\mu_N x_N^t,\lambda_1 y_1^t,...,\lambda_M y_M^t,\mu_1 x_1^{t+1},...,\mu_N x_N^{t+1},\lambda_1 y_1^{t+1},...,\lambda_M y_M^{t+1}) =$$
$$M_{oc}(x^t,y^t,x^{t+1},y^{t+1}),$$

$$\lambda_m > 0, \ m=1,...,M, \ \mu_n > 0, \ n=1,...,N$$

M6: Circularity

$$M_{oc}(x^t,y^t,x^{t+1},y^{t+1}) \bullet M_{oc}(x^{t+1},y^{t+1},x^{t+2},y^{t+2}) \neq M_{oc}(x^t,y^t,x^{t+2},y^{t+2})$$

M7: Time Reversal

$$M_{oc}(x^t,y^t,x^{t+1},y^{t+1}) = [M_{oc}(x^{t+1},y^{t+1},x^t,y^t)]^{-1}$$

Although the Malmquist productivity index does not satisfy the proportionality test in either outputs or inputs separately, it does satisfy the proportionality test in outputs and inputs simultaneously. In addition, it is not circular except under restrictive neutrality conditions on technical change. The seriousness of the failure to satisfy the circularity test depends on the persuasiveness of the arguments of Fisher (1922), who rejected the test, and of Frisch (1936), who endorsed it. We leave this evaluation to the reader, who may seek guidance from Samuelson & Swamy (1974).

In Section 1.3 we noted that distance to a production frontier can be measured hyperbolically or directionally. The two distance functions $D_H(y,x)$ and $D_D(y,x)$ can be used to derive hyperbolic and directional Malmquist productivity indexes analogous to the output-oriented Malmquist productivity index discussed in this Section. A Malmquist productivity index based on directional distance functions is used extensively in Chapter 5.

1.7.2 Decomposing the Malmquist productivity index

In Section 1.2 we noted that the BLS and the OECD attribute productivity change to technical change, efficiency change, scale economies and changes in the operating environment in which production occurs. It is possible to decompose the Malmquist productivity index (1.62) into the first three of these sources. Thus this index is capable not just of quantifying productivity change, but also of quantifying its three principal sources.

Färe *et al.* (1992) obtained an initial decomposition of (1.62). The mathematics is straightforward, and the economics is enlightening. Extracting the term $[D_{oc}^{t+1}(x^{t+1},y^{t+1})/D_{oc}^{t}(x^{t},y^{t})]$ from the right side of (1.62) yields

$$M_{oc}(x^t,y^t,x^{t+1},y^{t+1}) = \frac{D_{oc}^{t+1}\left(x^{t+1},y^{t+1}\right)}{D_{oc}^{t}\left(x^t,y^t\right)} \times \left[\frac{D_{oc}^{t}\left(x^t,y^t\right)}{D_{oc}^{t+1}\left(x^t,y^t\right)} \times \frac{D_{oc}^{t}\left(x^{t+1},y^{t+1}\right)}{D_{oc}^{t+1}\left(x^{t+1},y^{t+1}\right)} \right]^{\frac{1}{2}}$$

$$= TE\Delta_{oc}(x^t,y^t,x^{t+1},y^{t+1}) \times T\Delta_{oc}(x^t,y^t,x^{t+1},y^{t+1}). \qquad (1.63)$$

Recalling from Section 1.3 that $TE_o(x,y) = [D_o(x,y)]^{-1}$, the first term on the right side of (1.63) measures the contribution of technical efficiency change to productivity change. $TE\Delta_{oc}(x^t,y^t,x^{t+1},y^{t+1}) \gtreqless 1$ according as technical efficiency improves, remains unchanged or deteriorates between periods t and t+1. The second term on the right side of (1.63) measures the contribution of technical change to productivity change. It is the geometric mean of two terms, one comparing period t technology to period t+1 technology from the perspective of period t data, and the other comparing the two technologies from the perspective of period t+1 data. $T\Delta_{oc}(x^t,y^t,x^{t+1},y^{t+1}) \gtreqless 1$ according as technical progress, stagnation, or regress has occurred between periods t and t+1. In Figure 1.20 it is apparent that productivity growth has occurred between periods t and t+1 because technical efficiency has improved, and because technical progress has occurred.

There is, however, a problem with decomposition (1.63), which is why we refer to it as an initial decomposition. Productivity change is properly measured relative to the benchmark technologies T_c^t and T_c^{t+1}. Unfortunately so are its technical efficiency change and technical change components. They should be measured relative to the best practice technologies T^t and T^{t+1} that are not constrained to satisfy global constant returns to scale. In addition, (1.63) attributes productivity change exclusively to technical efficiency change and technical change. Introducing a term capturing the contribution of scale economies requires introducing the best practice technologies.

Figure 1.21 illustrates a subsequent decomposition. The middle row corresponds to the initial decomposition in (1.63). The bottom row describes a generic decomposition of productivity change into a technical efficiency change component $TE\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ measured relative to the best practice technologies, a technical change component $T\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ characterizing the shift in the best practice technologies, and a third component $S\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ measuring the contribution of scale economies to productivity change. However there is more than one way to implement this subsequent decomposition mathematically, and different mathematical decompositions have different economic interpretations. All appear to agree that a subsequent decomposition is needed, but disagreement over the nature of the subsequent decomposition persists. Grosskopf (2003) and Lovell (2003) survey the landscape, and the decomposition issue is revisited in Chapter 5.

Figure 1.21 Decomposing the Malmquist Productivity Index
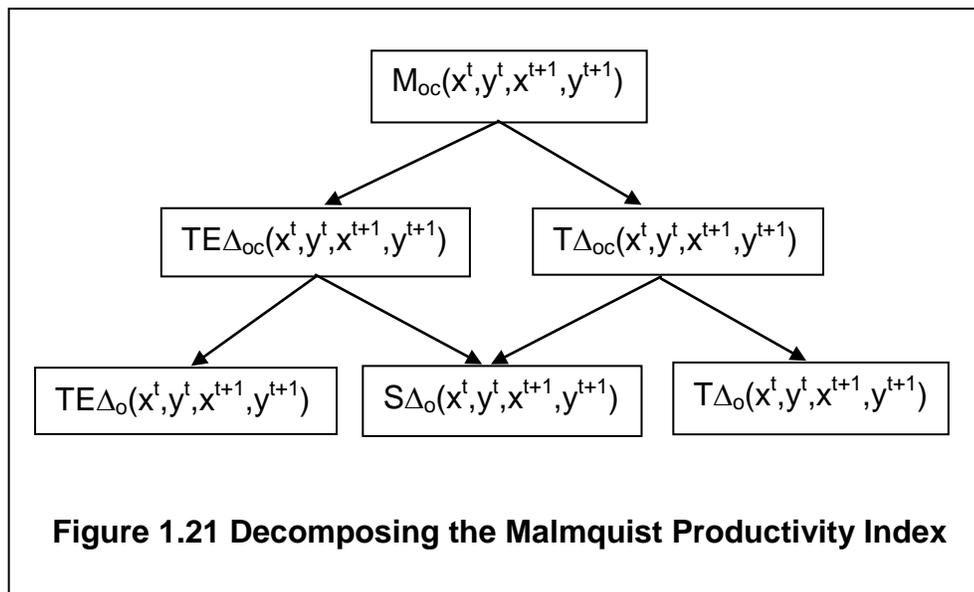
1.7.3 Evaluating Malmquist

The Malmquist productivity index has several nice theoretical features. Because it is based on distance functions, it inherits several desirable properties from them. Again because it is based on distance functions, it readily accommodates multiple outputs as well as multiple inputs. The output expanding orientation can be reversed to generate an input oriented Malmquist productivity index based on input distance functions, and nothing of substance would change.

The Malmquist productivity index also has a very nice practical feature. Once again because it is based on distance functions, it requires information on

quantities, but not prices. This makes it suitable for productivity measurement in situations in which prices are distorted or missing. We mentioned several such situations in Section 1.2, and we revisit the issue in Section 1.8.

The Malmquist productivity index can be decomposed into economically meaningful sources of productivity change, as Figure 1.21 suggests. However its decomposition requires a number of producers sufficiently large to enable one to construct benchmark and best practice technologies for each period. Construction can be based on either SFA techniques introduced in Section 1.5 or DEA techniques introduced in Section 1.6, as we indicate in Section 1.8. Particularly in the widely used DEA approach, however, the statistical significance of the contributions of the "economically meaningful" components is rarely investigated.

One potential source of productivity change is the vector z of exogenous variables previously discussed in the context of SFA and DEA. The pros and cons of alternative approaches to incorporating z in SFA and DEA efficiency analysis apply with equal force to the use of SFA and DEA to implement a Malmquist productivity analysis, a topic to which we now turn.

1.8 Approximating Malmquist

The Malmquist productivity index is a theoretical index, expressed in terms of distance functions defined on the true, but unknown, technology. If the index is to be implemented empirically, it must be approximated. Two philosophically different approaches have emerged. The older approach, which is far more popular, uses price information in place of technology information to *compute* productivity index numbers that provide empirical approximations to the theoretical Malmquist productivity index. This is the approach adopted by government statistical agencies around the world. The newer approach eschews price information, and uses either econometric or mathematical programming techniques to *estimate* the theoretical Malmquist productivity index itself, by estimating its component distance functions that characterize the structure of the underlying technology. Balk (1998), Diewert (1981, 1987) and Diewert and Nakamura (2003, 2006) survey the literature.

1.8.1 Superlative index numbers: Fisher and Törnqvist

Suppose that producers use inputs $x \in R^N_+$ available at prices $w \in R^N_{++}$ to produce outputs $y \in R^M_+$ for sale at prices $p \in R^M_{++}$.

Laspeyres output quantity and input quantity indexes use base period prices to weight quantity changes, and so

$$Y_L = p^{tT}y^{t+1} / p^{tT}y^t, \qquad X_L = w^{tT}x^{t+1} / w^{tT}x^t. \qquad (1.64)$$

Paasche output quantity and input quantity indexes use comparison period prices to weight quantity changes, and so

$$Y_P = p^{t+1T}y^{t+1} / p^{t+1T}y^t, \qquad X_P = w^{t+1T}x^{t+1} / w^{t+1T}x^t. \qquad (1.65)$$

Fisher (1922) output quantity and input quantity indexes are geometric means of Laspeyres and Paasche indexes, and so

$$Y_F = (Y_L \times Y_P)^{1/2} = [(p^{tT}y^{t+1} / p^{tT}y^t) \times (p^{t+1T}y^{t+1} / p^{t+1T}y^t)]^{1/2}$$

$$X_F = (X_L \times X_P)^{1/2} = [(w^{tT}x^{t+1} / w^{tT}x^t) \times (w^{t+1T}x^{t+1} / w^{t+1T}x^t)]^{1/2}. \qquad (1.66)$$

Fisher quantity indexes use both base period and comparison period prices to weight quantity changes. A Fisher productivity index is defined as

$$\Pi_F = \frac{Y_F}{X_F} = \frac{[(p^{tT}y^{t+1} / p^{tT}y^t) \times (p^{t+1T}y^{t+1} / p^{t+1T}y^t)]^{1/2}}{[(w^{tT}x^{t+1} / w^{tT}x^t) \times (w^{t+1T}x^{t+1} / w^{t+1T}x^t)]^{1/2}}. \qquad (1.67)$$

Π_F makes no use of the true but unobserved technology, and does not estimate it. It is computed from observable information on prices and quantities in base and comparison periods. What sort of approximation to the truth does it provide?

Diewert (1992) proved that, under certain conditions, $\Pi_F = M_{oc}(x^t,y^t,x^{t+1},y^{t+1})$, so that there is no approximation error at all. However these conditions are restrictive, collectively if not individually, and require that
- the output distance functions must be defined on the benchmark technologies exhibiting constant returns to scale
- the output distance functions must have a flexible functional form that is not reproduced here
- the period t and period t+1 output distance functions must have certain coefficients identical, which limits the extent to which technology can differ from one period to the next
- production in both periods must be allocatively efficient in competitive output markets and competitive input markets.

The first three requirements are not terribly restrictive, but the final requirement is, since it precludes precisely what this book is concerned with, a failure to optimize. Unfortunately we do not know the extent to which the performance of Π_F deteriorates as allocative inefficiency increases.

Törnqvist (1936) output and input quantity indexes are given (in logarithmic form) by

$$\ln Y_T = (1/2)\Sigma_m[(p_m^t y_m^t/\Sigma_m p_m^t y_m^t) + (p_m^{t+1} y_m^{t+1}/\Sigma_m p_m^{t+1} y_m^{t+1})]\ln(y_m^{t+1}/y_m^t)$$

$$\ln X_T = (1/2)\Sigma_n[(w_n^t x_n^t/\Sigma_n w_n^t x_n^t) + (w_n^{t+1} x_n^{t+1}/\Sigma_n w_n^{t+1} x_n^{t+1})]\ln(x_n^{t+1}/x_n^t). \qquad (1.68)$$

The output quantity index uses the arithmetic mean of adjacent period revenue shares to weight output quantity changes, and the input quantity index uses the arithmetic mean of adjacent period cost shares to weight input quantity changes. A Törnqvist productivity index is defined as

$$\Pi_T = \frac{Y_T}{X_T} = \exp\{\ln Y_T - \ln X_T\}$$

$$= \exp\{(1/2)\Sigma_m[(p_m^t y_m^t/\Sigma_m p_m^t y_m^t) + (p_m^{t+1} y_m^{t+1}/\Sigma_m p_m^{t+1} y_m^{t+1})]\ln(y_m^{t+1}/y_m^t)$$

$$- (1/2)\Sigma_n[(w_n^t x_n^t/\Sigma_n w_n^t x_n^t) + (w_n^{t+1} x_n^{t+1}/\Sigma_n w_n^{t+1} x_n^{t+1})]\ln(x_n^{t+1}/x_n^t)\}. \qquad (1.69)$$

Like the Fisher productivity index, the Törnqvist productivity index makes no use of the true but unobserved technology, and does not estimate it. It is computed from observable information on shares and quantities in base and comparison periods. What sort of approximation to the truth does it provide?

Caves *et al.* (1982b) proved that, under certain conditions, $\Pi_T = M_{oc}(x^t,y^t,x^{t+1},y^{t+1})$, so that there is no approximation error at all. However these conditions are restrictive, collectively if not individually, and require that
- all output quantities and all input quantities must be strictly positive
- the output distance functions must be defined on the benchmark technologies exhibiting constant returns to scale
- the output distance functions must have flexible translog functional form
- the period t and period t+1 output distance functions must have identical second order coefficients, which limits the extent to which technology can differ from one period to the next
- production in both periods must be allocatively efficient in competitive output markets and competitive input markets.

Our evaluation of the Törnqvist productivity index parallels our evaluation of the Fisher productivity index. The first four requirements are not terribly restrictive, although the first does rule out corner solutions. However the final requirement is restrictive, since it precludes a failure to optimize. Unfortunately we do not know the extent to which the performance of Π_T deteriorates as allocative inefficiency increases.

In the economic approach to index numbers Fisher and Törnqvist productivity indexes are called *superlative* because, under the conditions stated above, each provides a close approximation to the truth as given by the theoretical Malmquist productivity index. If production technology is characterized by a flexible functional form (either Diewert or translog output distance functions), and if producers are allocatively efficient in competitive markets, then subject to some provisos $\Pi_F = M_{oc}(x^t, y^t, x^{t+1}, y^{t+1}) = \Pi_T$. However we do not yet have a good sense of the performance of either Π_F or Π_T in the presence of scale economies, market power or allocative inefficiency. In addition, like the Malmquist productivity index itself, Π_F and Π_T are bilateral indexes that fail the circularity test. Both can be converted to circular multilateral indexes, but at a cost; the weights applied to quantity changes depend on the data of all producers, not just on the data of the producer whose productivity change is being measured. Caves *et al.* (1982a) provide the details for Π_T and references for Π_F.

1.8.2 An econometric approach

The econometric tools summarized in Section 1.5 can be adapted to the estimation and decomposition of a Malmquist productivity index. We summarize an approach suggested by Orea (2002). This approach extends previous analyses of Denny *et al.* (1981) and Nishimizu and Page (1982), and exploits the Caves *et al.* (1982b) analysis of the relationship between a Malmquist productivity index and a translog specification of the underlying distance functions. Suppose the output distance functions in the Malmquist productivity index have translog functional form in (x,y,t), so that

$$\ln D_o(x,y,t) = \alpha_o + \Sigma_n \alpha_n \ln x_n + \Sigma_m \beta_m \ln y_m + (\tfrac{1}{2}) \Sigma_n \Sigma_k \alpha_{nk} \ln x_n \ln x_k$$

$$+ (\tfrac{1}{2}) \Sigma_m \Sigma_q \beta_{mq} \ln y_m \ln y_q + \Sigma_n \Sigma_m \gamma_{nm} \ln x_n \ln y_m + \delta_t t + (\tfrac{1}{2}) \delta_{tt} t^2$$

$$+ \Sigma_n \delta_{tn} t \ln x_n + \Sigma_m \delta_{tm} t \ln y_m. \tag{1.70}$$

Since this function is quadratic in (x,y,t), the change in the value of the distance function from period t to period t+1 can be decomposed into the impacts of changes in outputs, changes in inputs, and the passage of time by means of

$\ln D_o(x^{t+1}, y^{t+1}, t+1) - \ln D_o(x^t, y^t, t) =$

$$(\tfrac{1}{2})\Sigma_m\left[\frac{\partial \ln D_o(x^{t+1}, y^{t+1}, t+1)}{\partial \ln y_m} + \frac{\partial \ln D_o(x^t, y^t, t)}{\partial \ln y_m}\right] \bullet \ln(y_m^{t+1}/y_m^t)$$

$$+ (\tfrac{1}{2})\Sigma_n\left[\frac{\partial \ln D_o(x^{t+1}, y^{t+1}, t+1)}{\partial \ln x_n} + \frac{\partial \ln D_o(x^t, y^t, t)}{\partial \ln x_n}\right] \bullet \ln(x_n^{t+1}/x_n^t)$$

$$+ (\tfrac{1}{2})\left[\frac{\partial \ln D_o(x^{t+1}, y^{t+1}, t+1)}{\partial t} + \frac{\partial \ln D_o(x^t, y^t, t)}{\partial t}\right]. \qquad (1.71)$$

If we define a logarithmic Malmquist productivity index $M_o(x,y,t)$ as the difference between weighted average rates of growth of outputs and inputs, with distance function elasticities as weights, (1.70) and (1.71) yield

$$\ln M_o(x,y,t) = (\tfrac{1}{2})\Sigma_m\left[\frac{\partial \ln D_o(x^{t+1}, y^{t+1}, t+1)}{\partial \ln y_m} + \frac{\partial \ln D_o(x^t, y^t, t)}{\partial \ln y_m}\right] \bullet \ln(y_m^{t+1}/y_m^t)$$

$$- (\tfrac{1}{2})\Sigma_n\left[\frac{-\partial \ln D_o(x^{t+1}, y^{t+1}, t+1)}{\partial \ln x_n} + \frac{-\partial \ln D_o(x^t, y^t, t)}{\partial \ln x_n}\right] \bullet \ln(x_n^{t+1}/x_n^t), \quad (1.72)$$

from which it follows that

$$\ln M_o(x,y,t) = [\ln D_o(x^{t+1}, y^{t+1}, t+1) - \ln D_o(x^t, y^t, t)]$$

$$- (\tfrac{1}{2})\left[\frac{\partial \ln D_o(x^{t+1}, y^{t+1}, t+1)}{\partial t} + \frac{\partial \ln D_o(x^t, y^t, t)}{\partial t}\right]. \qquad (1.73)$$

Expression (1.73) decomposes the logarithmic Malmquist productivity index $\ln M_o(x,y,t)$ into a term capturing the impact of technical efficiency change and a term capturing the impact of technical change. However because we did not impose constant returns to scale on $\ln D_o(x,y,t)$ in (1.70), the input weights in (1.72) do not necessarily sum to unity, and consequently $\ln M_o(x,y,t)$ in (1.73) ignores the contribution of scale economies to productivity change. The two terms on the right side of (1.73) are correct, but $\ln M_o(x,y,t)$ is not a proper productivity index, as we noted in Section 1.7. The two components on the right side of (1.73) correspond to $TE\Delta_o(x^t, y^t, x^{t+1}, y^{t+1})$ and $T\Delta_o(x^t, y^t, x^{t+1}, y^{t+1})$ in Figure 1.20, but a scale economies component corresponding to $S\Delta_o(x^t, y^t, x^{t+1}, y^{t+1})$ is missing.

Expression (1.72) decomposes $M_o(x,y,t)$ by aggregating outputs and inputs using distance function elasticities. Decomposing $M_o(x,y,t)$ by aggregating outputs and inputs using distance function elasticity *shares* instead gives

$$\ln M_o(x,y,t) = (\tfrac{1}{2})\sum_m[\varepsilon_m(x^{t+1},y^{t+1},t+1) + \varepsilon_m(x^t,y^t,t)]\bullet\ln(y_m^{t+1}/y_m^t)$$

$$- (\tfrac{1}{2})\sum_n[\varepsilon_n(x^{t+1},y^{t+1},t+1) + \varepsilon_n(x^t,y^t,t)]\bullet\ln(x_n^{t+1}/x_n^t), \qquad (1.74)$$

where

$$\varepsilon_m(x^s,y^s,s) = \frac{\partial \ln D_o(x^s,y^s,s)}{\partial \ln y_m}$$

$$\varepsilon_n(x^s,y^s,s) = \frac{\partial \ln D_o(x^s,y^s,s)/\partial \ln x_n}{\sum_n \partial \ln D_o(x^s,y^s,s)/\partial \ln x_n}$$

for $s = t, t+1$. $\ln M_{oc}(x,y,t)$ in (1.74) is a proper productivity index because its input weights sum to unity. Consequently it corresponds to a benchmark technology satisfying constant returns to scale, as is required if it is to provide an accurate measure of productivity change. Finally substituting (1.73) into (1.74) yields

$$\ln M_{oc}(x,y,t) = [\ln D_o(x^{t+1},y^{t+1},t+1) - \ln D_o(x^t,y^t,t)]$$

$$- (\tfrac{1}{2})\left[\frac{\partial \ln D_o(x^{t+1},y^{t+1},t+1)}{\partial t} + \frac{\partial \ln D_o(x^t,y^t,t)}{\partial t}\right]$$

$$+ (\tfrac{1}{2})\sum_n\left[\begin{array}{l}\left(-\sum\dfrac{\partial \ln D_o(x^{t+1},y^{t+1},t+1)}{\partial \ln x_n}-1\right)\bullet\varepsilon_n(x^{t+1},y^{t+1},t+1)\\[2mm]+\left(-\sum\dfrac{\partial \ln D_o(x^t,y^t,t)}{\partial \ln x_n}-1\right)\bullet\varepsilon_n(x^t,y^t,t)\end{array}\right]\bullet\ln(x_n^{t+1}/x_n^t). \quad (1.75)$$

Expression (1.75) attributes productivity change to technical efficiency change and technical change, both from $\ln M_o(x,y,t)$, and to scale economies, expressed as the logarithmic difference between $\ln M_{oc}(x,y,t)$ and $\ln M_o(x,y,t)$. The three terms on the right side of (1.75) provide empirical approximations to the components $TE\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$, $T\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ and $S\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ in Figure 1.20, and so their sum provides an empirical approximation to $M_{oc}(x^t,y^t,x^{t+1},y^{t+1})$.

All that is required to implement (1.75) is to estimate the translog output distance function (1.70), imposing linear homogeneity in outputs and making an assumption about the error structure. After estimation, parameter estimates can be used to estimate the elasticities involved in the second and third components

of the right side of (1.75). Estimation of the first component requires frontier techniques described in Section 1.5 and employed by Orea (2002).

1.8.3 A Mathematical programming approach

The mathematical programming tools summarized in Section 1.6 also can be adapted to the estimation and decomposition of a Malmquist productivity index. We summarize an approach that originated with Färe *et al.* (1992), and that has been refined by many authors since.

The Malmquist productivity index given by (1.62) in Section 1.7 contains four output distance functions, each defined on a benchmark technology satisfying constant returns to scale. The within-period distance functions are estimated using the CCR DEA envelopment program given by (1.53) in Section 1.6

$D_{oc}{}^{s}(x^{s},y^{s})$, s=t,t+1
$\max_{\phi,\lambda} \quad \phi$
subject to $\quad X^{s}\lambda \leqq x_{o}{}^{s}$
$\phi y_{o}{}^{s} \leqq Y^{s}\lambda$
$\lambda \geqq 0$

(1.76)

and the adjacent-period distance functions are estimated using similar CCR DEA programs

$D_{oc}{}^{s}(x^{r},y^{r})$, s,r=t,t+1, s≠r
$\max_{\phi,\lambda} \quad \phi$
subject to $\quad X^{s}\lambda \leqq x_{o}{}^{r}$
$\phi y_{o}{}^{r} \leqq Y^{s}\lambda$
$\lambda \geqq 0$

(1.77)

Substituting the solutions to these four programs into (1.62) generates a Malmquist productivity index estimated using mathematical programming techniques.

Decomposing the Malmquist productivity index requires the estimation of distance functions defined on a best practice technology allowing for variable returns to scale. This requires use of the BCC DEA envelopment program given by (1.54) in Section 1.6. $TE\Delta_{o}(x^{t},y^{t},x^{t+1},y^{t+1})$ is estimated as the ratio of the following distance functions

$$
\begin{array}{|l|}
\hline
D_o{}^s(x^s,y^s),\ s=t,t+1 \\
\hline
\max_{\phi,\lambda}\quad \phi \\
\hline
\text{subject to}\quad X^s\lambda \leqq x_o{}^s \\
\hline
\phi y_o{}^s \leqq Y^s\lambda \\
\hline
\lambda \geqq 0,\ \Sigma_i\lambda_i = 1 \\
\hline
\end{array}
$$

(1.78)

$T\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ involves these two distance functions and the two following distance functions as well

$$
\begin{array}{|l|}
\hline
D_{oc}{}^s(x^r,y^r),\ s,r=t,t+1,\ s\neq r \\
\hline
\max_{\phi,\lambda}\quad \phi \\
\hline
\text{subject to}\quad X^s\lambda \leqq x_o{}^r \\
\hline
\phi y_o{}^r \leqq Y^s\lambda \\
\hline
\lambda \geqq 0,\ \Sigma_i\lambda_i = 1 \\
\hline
\end{array}
$$

(1.79)

Programs (1.79) evaluate the performance of producers in period r against best practice technology prevailing in adjacent period s. Because best practice technologies allow for variable returns to scale, it is possible that not all programs have feasible solutions, as Ray and Desli (1997) discovered. This possibility notwithstanding, once $M_{oc}(x^t,y^t,x^{t+1},y^{t+1})$ has been estimated using the CCR programs, and $TE\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ and $T\Delta_o(x^t,y^t,x^{t+1},y^{t+1})$ have been estimated using the BCC programs, the contribution of scale economies to productivity change is estimated residually by means of

$$S\Delta_o(x^t,y^t,x^{t+1},y^{t+1}) = M_{oc}(x^t,y^t,x^{t+1},y^{t+1})\ /\ [TE\Delta_o(x^t,y^t,x^{t+1},y^{t+1})\bullet T\Delta_o(x^t,y^t,x^{t+1},y^{t+1})].$$
(1.80)

1.8.4 Evaluating the approximations

Since the truth is unknown, it is difficult to judge the accuracy of the econometric estimate of the Malmquist productivity index discussed in Section 1.8.2 and the mathematical programming estimate of the Malmquist productivity index discussed in Section 1.8.3. It is similarly difficult to judge whether an empirical Malmquist productivity index, estimated from either econometric or mathematical programming techniques, provides a better or worse approximation to the truth than a computed Fisher or Törnqvist productivity index does. In both cases statistical inference is required.

We can, however, make some relevant observations.

- First and foremost, Fisher and Törnqvist indexes are superlative only under restrictive assumptions. Among them are allocative efficiency, which we believe should be a hypothesis to be tested rather than a maintained assumption. The econometric and mathematical programming approaches do not asume allocative efficiency. They are capable of testing the hypothesis, by comparing market price ratios with estimated shadow price ratios.

- Second, Fisher and Törnqvist indexes require price or share information in their computation. But prices (and therefore shares) can be distorted by market power, by cross-subsidy, and by regulation. In addition, prices are missing in large parts of the non-market sector of the economy. The econometric and mathematical programming approaches do not require price information.

- Third, the econometric and mathematical programming approaches generate the same structural decomposition of the Malmquist productivity index, enabling one to attribute productivity change to technical efficiency change, technical change and the contribution of scale economies. These are the sources identified by the US BLS and the OECD. Ray and Mukherjee (1996) and Kuosmanen and Sipiläinen (2004) have obtained similar decompositions of the Fisher index, but the Törnqvist index has resisted similar efforts. The Fisher and Törnqvist productivity indexes have much more natural decompositions in terms of identifying the contributions of individual variables. Balk (2004) surveys the literature, and Salerian (2003) provides an application to Australian railroads.

1.9 Concluding Remarks

We began this Chapter with an investigation into the recent variation in the economic and financial performance of US airlines. Variation in efficiency and productivity is commonplace, and examples are reported regularly in the business press. Since business performance variation exists, it is incumbent on the profession to develop the analytical tools and the empirical techniques needed to study it. If we can quantify it, and if we can identify its sources, we have a chance of adopting private practices and public policies designed to improve it.

We have provided motivation for the study of efficiency and productivity, and we have referred to a wide range of empirical applications. We have laid out the basics of the underlying theory and the empirical techniques. The reader is now properly motivated and adequately prepared to continue on to the more extensive analyses provided in subsequent Chapters.

References

Abad, C., S. A. Thore and J. Laffarga (2004), "Fundamental Analysis of Stocks by Two-Stage DEA," *Managerial and Decision Economics* 25:5 (July-August), 231-41.

Abramovitz, M. (1956), "Resource and Output Trends in the United States Since 1870," *American Economic Review* 46 (May), 5-23.

Afriat, S. N. (1972), "Efficiency Estimation of Production Functions," *International Economic Review* 13:3 (October), 568-98.

Agrell, P., P. Bogetoft and J. Tind (2005), "DEA and Dynamic Yardstick Competition in Scandinavian Electricity Distribution," *Journal of Productivity Analysis* 23:2 (March), 173-201.

Aigner, D. J., and S.-F. Chu (1968), "On Estimating the Industry Production Function," *American Economic Review* 58:4 (September), 826-39.

Aigner, D. J., C. A. K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics* 6:1 (July), 21-37.

Ajodhia, V., and K. Petrov (2004), "Establishment of Cap Regulation and Determination of Efficiency Factor X for Electricity Networks," *European Transactions on Electrical Power* 14:2 (March-April), 97-109.

Alchian, A. A. (1965), "Some Economics of Property Rights," *Il Politico* 30:4 (December), 816-29.

Alchian, A. A., and R. A. Kessel (1962), "Competition, Monopoly, and the Pursuit of Money," in *Aspects of Labor Economics.* Princeton: Princeton University Press for the National Bureau of Economic Research.

Ambec, S., and P. Barla (2002), "A Theoretical Foundation of the Porter Hypothesis," *Economics Letters* 75:3 (May), 355-60.

Ammar, S. W. Duncombe, B. Jump, *et al.* (2004), "Constructing a Fuzzy-Knowledge-Based System: An Application for Assessing the Financial Condition of Public Schools," *Expert Systems with Applications* 27:3 (October), 349-64.

Amos, D., T. R. Beard and S. B. Caudill (2005), "A Statistical Analysis of the Handling Characteristics of Certain Sporting Arms: Frontier Regression, the Moment of Inertia, and the Radius of Gyration," *Journal of Applied Statistics* 32:1, 3-16

Andersen, P., and Petersen, N. C. (1993), "A Procedure for Ranking Efficient Units in DEA," *Management Science* 39:10 (October), 1261-64.

Anderson, R. I., C. M. Brockman, C. Giannikos and R. W. McLeod (2004), "A Non-Parametric Examination of Real Estate Mutual Fund Efficiency," *International Journal of Business and Economics* 3:3 (December), 225-38.

Arocena, P., and C. Waddams Price (2003), "Generating Efficiency: Economic and Environmental Regulation of Public and Private Electricity Generators in Spain," *International Journal of Industrial Organization* 20:1 (January), 41-69.

Atkinson, S. E., R. Färe and D. Primont (2003), "Stochastic Estimation of Firm Inefficiency Using Distance Functions," *Southern Economic Journal* 69, 596-611.

Atkinson, S. E., and C. E. Halabi (2005), "Economic Efficiency and Productivity Growth in the Post-Privatization Chilean Hydroelectric Industry," *Journal of Productivity Analysis* 23:2 (March), 245-73.

Averch, H., and L. L. Johnson (1962), "Behavior of the Firm Under Regulatory Constraint," *American Economic Review* 52:5 (December), 1052-69.

Balk, B. M. (1998), *Industrial Price, Quantity and Productivity Indexes*. Boston: Kluwer Academic Publishers.

Balk, B. M. (2004), "Decompositions of Fisher Indexes," *Economics Letters* 82:1 (January), 107-13.

Ball, V. E., C. A. K. Lovell, R. Nehring and H. Luu (2004), "Incorporating Environmental Impacts in the Measurement of Agricultural Productivity Growth," *Journal of Agricultural and Resource Economics* 29:3 (December), 436-60.

Banker, R. D., H.-H. Chang and S. K. Majumdar (1993), "Analyzing the Underlying Dimensions of Firm Profitability," *Managerial and Decision Economics* 14:1 (January-February), 25-36.

Banker, R. D., H. Chang and R. Natarajan (2005), "Productivity Change, Technical Progress, and Relative Efficiency Change in the Public Accounting Industry," *Management Science* 51:2 (February), 291-304.

Banker, R. D., A. Charnes and W. W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* 30:9 (September), 1078-92.

Baños-Pino, J., V. Fernández-Blanco and A. Rodríguez-Alvarez (2002), "The Allocative Efficiency Measure by Means of a Distance Function: The Case of

Spanish Public Railways," *European Journal of Operational Research* 137, 191-205.

Banzhaf, H. S. (2005), "Green Price Indices," *Journal of Environmental Economics and Management* 49:2 (March), 262-80.

Barros, C. P. (2002), "Development and Conflict in the Balkans: Catch-Up and Military Expenditure," *Defence and Peace Economics* 13:5, 353-63.

Barros, C. P. (2005), "Measuring Efficiency in the Hotel Industry," *Annals of Tourism Research* 32:2 (April), 456-77.

Barua, A., P. L. Brockett, W. W. Cooper, H. Deng, B. R. Parker, T. W. Ruefli and A. Whinston (2004), "DEA Evaluations of Long- and Short-Run Efficiencies of Digital vs. Physical Product 'dot com' Companies," *Socio-Economic Planning Sciences* 38:4 (December), 233-53.

Basso, A., and S. Funari (2003), "Measuring the Performance of Ethical Mutual Funds: A DEA Approach," *Journal of the Operational Research Society* 54:5 (May), 521-31.

Basso, A., and S. Funari (2004), "A Quantitative Approach to Evaluate the Relative Efficiency of Museums," *Journal of Cultural Economics* 28:3 (August), 195-216.

Battese, G. E., and T. Coelli (1988), "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data," *Journal of Econometrics* 38, 387-99.

Battese, G. E., and T. J. Coelli (1995), "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," *Empirical Economics* 20, 325-32.

Bauer, P. W., A. N. Berger, G. D. Ferrier and D. B. Humphrey (1998), "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Methods," *Journal of Economics and Business* 50:2 (March/April), 85-114.

Baumol, W. J. (1959), *Business Behavior, Value and Growth.* New York: Macmillan.

Becchetti, L., and J. Sierra (2003), "Bankruptcy Risk and Productive Efficiency in Manufacturing Firms," *Journal of Banking & Finance* 27:11 (November), 2099-2120.

Bera, A. K., and S. C. Sharma (1999), "Estimating Production Uncertainty in Stochastic Frontier Production Function Models," *Journal of Productivity Analysis* 12:3 (November), 187-210.

Berger, A. N., and T. H. Hannan (1998), "The Efficiency Cost of Market Power in the Banking Industry: A Test of the 'Quiet Life' and Related Hypotheses," *Review of Economics and Statistics* 80:3 (August), 454-65.

Berger, A. N., and L. J. Mester (1997), "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance* 21:7 (July), 895-947.

Berle, A. A., and G. C. Means (1932), *The Modern Corporation and Private Property*. New York: Macmillan.

Bi, H. Q. (2004), "Stochastic Frrontier Analysis of a Classic Self-Thinning Experiment," *Austral Ecology* 29:4 (August), 408-17.

Birman, S. V., P. E. Pirondi and E. Y. Rodin (2003), "Application of DEA to Medical Clinics," *Mathematical and Computer Modeling* 37:9-10 (May), 923-36.

Bishop, P., and S. Brand (2003), "The Efficiency of Museums: A Stochastic Frontier Production Function Approach," *Applied Economics* 35:17 (November), 1853-58.

Bjurek, H. (1996), "The Malmquist Total Factor Productivity Index," *Scandinavian Journal of Economics* 98:2, 303-13.

Bogetoft, P. (2000), "DEA and Activity Planning Under Asymmetric Information," *Journal of Productivity Analysis* 13:1 (January), 7-48.

Bogetoft, P., and D. Wang (2005), "Estimating the Potential Gains from Mergers," *Journal of Productivity Analysis* 23:2 (March), 145-71.

Bonaccorsi, A., and C. Daraio (2003), "A Robust Nonparametric Approach to the Analysis of Scientific Productivity," *Research Evaluation* 12:1 (April), 47-69.

Borenstein, D., J. L. Becker and V. J. do Prado (2004), "Measuring the Efficiency of Brazilian Post Office Stores Using Data Envelopment Analysis," *International Journal of Operations & Production Management* 24:9-10, 1055-78.

Bosch, N., F. Pedraja and J. Suarez-Pandiello (2000), "Measuring the Efficiency of Spanish Municipal Refuse Collection Services," *Local Government Studies* 26:3 (Fall), 71-90.

Bowlin, W. F. (2004), "Financial Analysis of Civil Reserve Air Fleet Participants Using Data Envelopment Analysis," *European Journal of Operational Research* 154, 691-709.

Bowlin, W. F., C. J. Renner and J. M. Reeves (2003), "A DEA Study of Gender Equity in Executive Compensation," *Journal of the Operational Research Society* 54:7 (July), 751-57.

Brockett, P. L., W. W. Cooper, S. C. Kumbhakar, M. J. Kwinn, Jr., and D. McCarthy (2004), "Alternative Statistical Regression Studies of the Effects of Joint and Service Specific Advertising on Military Recruitment," *Journal of the Operational Research Society* 55:10 (October), 1039-48.

Buck, D. (2000), "The Efficiency of the Community Dental Service in England: A Data Envelopment Analysis," *Community Dentistry and Oral Epidemiology* 26:4 (August), 274-80.

Camanho, A. S., and R. G. Dyson (2005), "Cost Efficiency, Production and Value-Added Models in the Analysis of Bank Branch Performance," *Journal of the Operational Research Society* 56, 483-94.

Carlsson, B. (1972), "The Measurement of Efficiency in Production: An Application to Swedish Manufacturing Industries, 1968," *Swedish Journal of Economics* 74:4 (December), 468-85.

Carrington, R., T. Coelli and E. Groom (2002), "International Benchmarking for Monopoly Price Regulation: The Case of Australian Gas Distribution," *Journal of Regulatory Economics* 21:2 (March), 191-216.

Caves, D. W., L. R. Christensen and W. E. Diewert (1982a), "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers," *Economic Journal* 92 (March), 73-86.

Caves, D. W., L. R. Christensen and W. E. Diewert (1982b), "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity," *Econometrica* 50:6 (November), 1393-1414.

Chang, H., W.-J. Chang, S. Das and S.-H. Li (2004), "Health Care Regulation and the Operating Efficiency of Hospitals: Evidence from Taiwan," *Journal of Accounting and Public Policy* 23:6 (November-December), 483-510.

Chang, K.-P. (2004), "Evaluating Mutual Fund Performance: An Application of Minimum Convex Input Requirement Set Approach," *Computers & Operations Research* 31:6 (May), 929-40.

Charnes, A., and W. W. Cooper (1959), "Chance Constrained Programming," *Management Science* 6:1 (October), 73-79.

Charnes, A., W. W. Cooper and E. Rhodes (1978), "Measuring the Efficiency of Decision-Making Units," *European Journal of Operational Research* 2:6 (November), 429-44.

Charnes, A., W. W. Cooper and G. H. Symonds (1958), "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil," *Management Science* 4:3 (April), 235-63.

Chen, Y., L. Motiwalla and M. R. Khan (2004), "Using Super-Efficiency DEA to Evaluate Financial Performance of E-Business Initiative in the Retail Industry," *International Journal of Information Technology & Decision Making* 3:2 (June), 337-51.

Chen, Y., and J. Zhu (2003), "DEA Models for Identifying Critical Performance Measures," *Annals of Operations Research* 124, 225-44.

Cherchye, L., W. Moesen and T. Van Puyenbroeck (2004), "Legitimately Diverse, Yet Comparable: On Synthesizing Social Inclusion Performance in the EU," *Journal of Common Market Studies* 42:5 (December), 919-55.

Chiang, F. S., C. H. Sun and J. M. Yu (2004), "Technical Efficiency Analysis of Milkfish (Chanos chanos) Production in Taiwan - An Application of the Stochastic Frontier Production Function," *Aquaculture* 230:1-4 (February), 99-116.

Chiang, W.-E., M.-H. Tsai and L. S.-M. Wang (2004), "A DEA Evaluation of Taipei Hotels," *Annals of Tourism Research* 31:3 (July), 712-15.

Christensen, L. R., and W. H. Greene (1976), "Economies of Scale in US Electric Power Generation," *Journal of Political Economy* 84:4, 655-76.

Cielen, A., L. Peeters and K. Vanhoof (2004), "Bankruptcy Prediction Using a Data Envelopment Analysis," *European Journal of Operational Research* 154:2 (April), 526-32.

Clark, X., D. Dollar and A. Micco (2004), "Port Efficiency, Maritime Transport Costs, and Bilateral Trade," *Journal of Development Economics* 75:2 (December), 417-50.

Coelli, T., E. Grifell-Tatjé and S. Perelman (2002), "Capacity Utilisation and Profitability: A Decomposition of Short-Run Profit Efficiency," *International Journal of Production Economics* 79, 261-78.

Cook, W. D., and R. H. Green (2003), "Selecting Sites for New Facilities Using Data Envelopment Analysis," *Journal of Productivity Analysis* 19:1 (January), 77-91.

Cook, W. D., and R. H. Green (2005), "Evaluating Power Plant Efficiency: A Hierarchical Model," *Computers & Operations Research* 32:4 (April), 813-23.

Cooper, W. W., K. S. Park and J. T. Pastor (1999), "RAM: A Range Adjusted Measure of Inefficiency for Use with Additive Models, and Relations to Other Models and Measures in DEA," *Journal of Productivity Analysis* 11:1 (February), 5-42.

Cooper, W. W., L. M. Seiford and K. Tone (2000), *Data Envelopment Analysis*. Boston: Kluwer Academic Publishers.

Corton, M. L. (2003), "Benchmarking in the Latin American Water Sector: The Case of Peru," *Utilities Policy* 11:3 (September), 133-42.

Croppenstedt, A., and M. M. Meschi (2000), "Measuring Wage Discrimination in Italy: A Random-Coefficient Approach," *Applied Economics* 32:8 (June), 1019-28.

Cuesta, R. A., and L. Orea (2002), "Mergers and Technical Efficiency in Spanish Savings Banks: A Stochastic Distance Function Approach," *Journal of Banking and Finance* 26, 2231-2247.

Cummins, J. D., M. Rubio-Misas and H. Zi (2005), "The Effect of Organizational Structure on Efficiency: Evidence from the Spanish Insurance Industry," *Journal of Banking & Finance* 28:12 (December), 3113-3150.

Cummins, J. D., and H. Zi (1998), "Comparison of Frontier Efficiency Methods: An Application to the U.S. Life Insurance Industry," *Journal of Productivity Analysis* 10:2 (October), 131-52.

Dalen, D. M., and A. Gómez-Lobo (2003), "Yardsticks on the Road: Regulatory Contracts and Cost Efficiency in the Norwegian Bus Industry," *Transportation* 30:4 (November), 371-86.

Davies, A. J., and A. K. Kochhar (2002), "Manufacturing Best Practice and Performance Studies: A Critique," *International Journal of Operations & Production Management* 22:3, 289-305.

Davis, S., and T. Albright (2004), "An Investigation of the Effect of Balanced Scorecard Implementation on Financial Performance," *Management Accounting Research* 15:2 (June), 135-53.

de Alessi, L. (1974), "An Economic Analysis of Government Ownership and Regulation: Theory and the Evidence from the Electric Power Industry," *Public Choice* 19:1, 1-42.

De Borger, B., K. Kerstens and A. Costa (2002), "Public Transit Performance: What Does One Learn From Frontier Studies?" *Transport Reviews* 22:1 (January), 1-38.

Debreu, G. (1951), "The Coefficient of Resource Utilization," *Econometrica* 19:3 (July), 273-92.

Denny, M., M. Fuss and L. Waverman (1981), "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications," in T. G. Cowing and R. Stevenson, eds., *Productivity Measurement in Regulated Industries.* New York: Academic Press.

Deprins, D., and L. Simar (1989), "Estimation de Frontières Déterministes avec Facteurs Exogènes d'Inefficacité," *Annales d'Economie et de Statistique* 14, 117-50.

Deprins, D., L. Simar and H. Tulkens (1984), "Measuring Labor-Efficiency in Post Offices," in M. Marchand, P. Pestieau and H. Tulkens, eds., *The Performance of Public Enterprises: Concepts and Measurement.* Amsterdam: North-Holland.

Dervaux, B., H. Leleu, V. Valdmanis and D. Walker (2003), "Parameters of Control when Facing Stochastic Demand: A DEA Approach Applied to Bangladeshi Vaccination Sites," *International Journal of Health Care Finance and Economics* 3:4 (December), 287-99.

Despotis, D. K. (2005), "Measuring Human Development via Data Envelopment Analysis: The Case of Asia and the Pacific," *Omega* 33:5 (October), 385-90.

Diaz-Mayans, M. A., and R. Sanchez (2004), "Temporary Employment and Technical Efficiency in Spain," *International Journal of Manpower* 25:2, 181-94.

Diewert, W. E. (1981), "The Economic Theory of Index Numbers: A Survey," Chapter 7 in A. Deaton, ed., *Essays in the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone*. New York: Cambridge University Press.

Diewert, W. E. (1987), "Index Numbers," in J. Eatwell, M. Milgate and P. Newman, eds., *The New Palgrave: A Dictionary of Economics, Volume 2.* New York: The Macmillan Press.

Diewert, W. E. (1992), "Fisher Ideal Output, Input and Productivity Indexes Revisited," *Journal of Productivity Analysis* 3:3 (September), 211-48.

Diewert, W. E., and A. O. Nakamura (2003), "Index Number Concepts, Measures and Decompositions of Productivity Growth," *Journal of Productivity Analysis* 19:2/3 (April), 127-59.

Diewert, W. E., and A. O. Nakamura (2006), "The Measurement of Aggregate Total Factor Productivity Growth," in J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics, Volume 6.* Amsterdam: Elsevier.

Dodson, M. E., and T. A. Garrett (2004), "Inefficient Education Spending in Public School Districts: A Case for Consolidation?" *Contemporary Economic Policy* 22:2 (April), 270-80.

Dolton, P., O. D. Marcenaro and L. Navarro (2003), "The Effective Use of Student Time: A Stochastic Frontier Production Function Case Study," *Economics of Education Review* 22:6 (December), 547-60.

Dopuch, N., M. Gupta, D. A. Simunic *et al.* (2003), "Production Efficiency and the Pricing of Audit Services," *Contemporary Accounting Research* 20:1 (Spring), 47-77.

Drake, L. M., and R. Simper (2004), "The Economics of Managerialism and the Drive for Efficiency in Policing," *Managerial and Decision Economics* 25:8 (December), 509-23.

Duke, J., and V. Torres (2005), "Multifactor Productivity Change in the Air Transportation Industry," *Monthly Labor Review* 128:3 (March), 32-45.

Edvardsen, D. F., F. R. Førsund, W. Hansen, S. A. C. Kittelsen and T. Neurauter (2006), "Productivity and Deregulation of Norwegian Electricity Distribution Utilities," in T. J. Coelli and D. Lawrence, eds., *Performance Measurement and Regulation of Network Utilities.* Cheltenham, UK: Edward Elgar.

Eilon, S. (1984), *The Art of Reckoning - Analysis of Performance Criteria.* London: Academic Press.

Emel, A. B., M. Oral, A. Reisman and R. Yolalan (2003), "A Credit Scoring Approach for the Commercial Banking Sector," *Socio-Economic Planning Sciences* 37:2 (June), 103-23.

Estache, A., M. González and L. Trujillo (2002), "What Does 'Privatization' Do For Efficiency? Evidence from Argentina's and Brazil's Railways," *World Development* 30:11 (November), 1885-97.

Façanha, L. O., and M. Resende (2004), "Price Cap Regulation, Incentives and Quality: The Case of Brazilian Telecommunications," *International Journal of Production Economics* 92, 133-44.

Färe, R., and S. Grosskopf (1996), *Intertemporal Production Frontiers: With Dynamic DEA*. Boston: Kluwer Academic Publishers.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1992), "Productivity Changes in Swedish Pharmacies 1980-1989: A Nonparametric Malmquist Approach," *Journal of Productivity Analysis* 3:1/2 (June), 85-101.

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

Färe, R., S. Grosskopf, C. A. K. Lovell and C. Pasurka (1989), "Multilateral Productivity Comparisons When Some Outputs are Undesirable: A Nonparametric Approach," *Review of Economics and Statistics* 71:1 (February), 90-98.

Färe, R., S. Grosskopf, C. A. K. Lovell and S. Yaisawarng (1993), "Derivation of Shadow Prices for Undesirable Outputs: A Distance Function Approach," *Review of Economics and Statistics* 75:2 (May), 374-80.

Färe, R., and J. Logan (1983), "The Rate-of-Return Regulated Firm: Cost and Production Duality," *Bell Journal of Economics* 14:2 (Autumn), 405-14.

Färe, R., and C. A. K. Lovell (1978), "Measuring the Technical Efficiency of Production," *Journal of Economic Theory* 19:1 (October), 150-162.

Farrell, M. J. (1957), "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society, Series A, General,* 120:3, 253-82.

Farsi, M., and M. Filippini (2004), "An Empirical Analysis of Cost Efficiency in Non-Profit and Public Nursing Homes," *Annals of Public and Cooperative Economics* 75:3 (September), 339-65.

Feroz, E. H., S. Kim and R. L. Raab (2003), "Financial Statement Analysis: A Data Envelopment Analysis Approach," *Journal of the Operational Research Society* 54, 48-58.

Ferrier, G. D., and V. G. Valdmanis (2004), "Do Mergers Improve Hospital Productivity?" *Journal of the Operational Research Society* 55:10 (October), 1071-80.

Filippini, M., N. Hrovatin and J. Zorič (2004), "Efficiency and Regulation of the Slovenian Electricity Distribution Companies," *Energy Policy* 32:3 (February), 335-44.

Fiorentini, G., and S. Peltzman, eds. (1995), *The Economics of Organized Crime*. New York: Cambridge University Press.

Fisher, I. (1922), *The Making of Index Numbers*. Boston: Houghton Mifflin.

Fogel, R. W., and S. L. Engerman (1974), *Time on the Cross*. Boston: Little, Brown and Company.

Foroughi, A. A., D. F. Jones and M. Tamiz (2005), "A Selection Method for a Preferential Election," *Applied Mathematics and Computation* 163:1 (April), 107-16.

Fried, H. O., J. Lambrinos and J. Tyner (2004), "Evaluating the Performance of Professional Golfers on the PGA, LPGA and SPGA Tours," *European Journal of Operational Research* 154:2 (April), 548-61.

Fried, H. O., C. A. K. Lovell, S. S. Schmidt and S. Yaisawarng (2002), "Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis," *Journal of Productivity Analysis* 17:1-2 (January), 157-74.

Fried, H. O., C. A. K. Lovell and P. Vanden Eeckaut (1993), "Evaluating the Performance of U.S. Credit Unions," *Journal of Banking and Finance* 17:2/3 (April), 251-65.

Frisch, R. (1936), "Annual Survey of General Economic Theory: The Problem of Index Numbers," *Econometrica* 4:1 (January), 1-38.

Frisch, R. (1965), *Theory of Production*. Chicago: Rand McNally.

Fu, T.-T., and C.-H. Huang (2002), "Efficiency Measurement and Multi-Output Pricing of Waste Water Treatment in Taiwan," in T.-T. Fu, C. J. Huang and C. A. K. Lovell, eds., *Productivity and Economic Performance in the Asia-Pacific Region*. Cheltenham, UK: Edward Elgar.

Fuentes, H. J., E. Grifell-Tatjé and S. Perelman (2001), "A Parametric Distance Function Approach for Malmquist Productivity Index Estimation," *Journal of Productivity Analysis* 15:2 (March), 79-94.

Gao, J., J. Campbell and C. A. K. Lovell (2006), "Equitable Resource Allocation and Operational Efficiency Evaluation," *International Journal of Healthcare Technology and Management* 7:1/2, 143-67.

Garrett, M. A. Jr., and Z. Xu (2003), "The Efficiency of Sharecropping: Evidence from the Postbellum South," *Southern Economic Journal* 69:3 (January), 578-95.

Goldar, B., and S. Misra (2001), "Valuation of Environmental Goods: Correcting for Bias in Contingent Valuation Studies Based on Willingness-to-Accept," *American Journal of Agricultural Economics* 83:1 (February), 150-56.

Gordon, R. A. (1961), *Business Leadership in the Large Corporation*. Berkeley: University of California Press.

Grafton, R. Q., S. Knowles and P. D. Owen (2004), "Total Factor Productivity, Per Capita Income and Social Divergence," *Economic Record* 80:250 (September), 302-13.

Greene, W. H. (1980), "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics* 13:1 (May), 27-56.

Greene, W. H. (2004), "Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems," *Health Economics* 13:10 (October), 959-80.

Grifell-Tatjé, E., and C. A. K. Lovell (1995), "A Note on the Malmquist Productivity Index," *Economics Letters* 47:2 (February), 169-75.

Grifell-Tatjé, E., and C. A. K. Lovell (1999), "Profits and Productivity," *Management Science* 45:9 (September), 1177-93.

Griliches, Z. (1996), "The Discovery of the Residual: A Historical Note," *Journal of Economic Literature* 34:3 (September), 1324-1330.

Grosskopf, S. (2003), "Some Remarks on Productivity and its Decomposition," *Journal of Productivity Analysis* 20:3 (November), 459-74.

Grytten, J., and G. Rongen (2000), "Efficiency in Provision of Public Dental Services in Norway," *Community Dentistry and Oral Epidemiology* 28:3 (June), 170-76.

Guan, J. C., and J. X. Wang (2004), "Evaluation and Interpretation of Knowledge Production Efficiency," *Scientometrics* 59:1, 131-55.

Guedes de Avellar, J. V., A. O. D. Polezzi and A. Z. Milioni (2002), "On the Evaluation of Brazilian Landline Telephone Services Companies," *Pesquisa Operacional* 22:2 (July-December), 231-46.

Gymah - Brempong, K. (2000), "Cost Efficiency in Florida Prisons," in J. L. T. Blank, ed., *Public Provision and Performance*. Amsterdam: North - Holland.

Haas, D. J. (2003), "Productive Efficiency of English Football Teams – A Data Envelopment Analysis Approach," *Managerial and Decision Economics* 24, 403-410.

Hailu, A., and T. S. Veeman (2000), "Environmentally Sensitive Productivity Analysis of the Canadian Pulp and Paper Industry, 1959-1994: An Input Distance Function Approach," *Journal of Environmental Economics and Management* 40:3 (November), 251-74.

Hammond, C. (2002), "Efficiency in the Provision of Public Services: A Data Envelopment Analysis of UK Public Library Systems," *Applied Economics* 34:5 (March), 649-57.

Hammond, C. J., G. Johnes and T. Robinson (2002), "Technical Efficiency Under Alternative Regulatory Regimes: Evidence from the Inter-War British Gas Industry," *Journal of Regulatory Economics* 22:3 (November), 251-70.

Hansmann, H. (1988), "Ownership of the Firm," *Journal of Law, Economics and Organization* 4:2 (Fall), 267-304.

Hausman, J. A., and W. E. Taylor (1981), "Panel Data and Unobservable Individual Effects," *Econometrica* 49:6 (November), 1377-98.

Hawdon, D. (2003), "Efficiency, Performance and Regulation of the International Gas Industry---A Bootstrap Approach," *Energy Policy* 31, 1167-78.

Henderson, D. J., and D. L. Millimet (2005), "Environmental Regulation and US State-Level Production," *Economics Letters* 87:1 (April), 47-53.

Herrero, I. (2004), "Risk and Strategy of Fishers Alternatively Exploiting Sea Bream and Tuna in the Gibraltar Strait from an Efficiency Perspective," *ICES Journal of Marine Science* 61:2 (April), 211-17.

Hicks, J. R. (1935), "The Theory of Monopoly: A Survey," *Econometrica* 3:1 (January), 1-20.

Hoch, I. (1955), "Estimation of Production Function Parameters and Testing for Efficiency," *Econometrica* 23:3 (July), 325-26.

Hof, J., C. Flather, T. Baltic and R. King (2004), "Forest and Rangeland Ecosystem Condition Indicators: Identifying National Areas of Opportunity Using Data Envelopment Analysis," *Forest Science* 50:4 (August), 473-94.

Hofler, R. A., and J. A. List (2004), "Valuation on the Frontier: Calibrating Actual and Hypothetical Statements of Value," *American Journal of Agricultural Economics* 86:1 (February), 213-21.

Hollingsworth, B., and J. Wildman (2003), "The Efficiency of Health Production: Re-Estimating the WHO Panel Data Using Parametric and Non-Parametric Approaches to Provide Additional Information," *Health Economics* 12, 493-504.

Holmstrom, B. R., and J. Tirole (1989), "The Theory of the Firm," in R. Schmalensee and R. D. Willig, eds., *Handbook of Industrial Organization, Volume I*. Amsterdam: Elsevier-Science Publishers.

Hopper, W. D. (1965), "Allocation Efficiency in a Traditional Indian Agriculture," *Journal of Farm Economics* 47:3 (August), 611-24.

Horrace, W. C., and P. Schmidt (1996), "Confidence Statements for Efficiency Estimates From Stochastic Frontier Models," *Journal of Productivity Analysis* 7:2/3 (July), 257-82.

Horrigan, J. O. (1968), "A Short History of Financial Ratio Analysis," *Accounting Review* 43:2 (April), 284-94.

Hougaard, J. L., D. Kronborg and C. Overgård (2004), "Improvement Potential in Danish Elderly Care," *Health Care Management Science* 7:3 (August), 225-35.

Hughes, P. A. N., and M. E. Edwards (2000), "Leviathan vs. Lilliputian: A Data Envelopment Analysis of Government Efficiency," *Journal of Regional Science* 40:4 (November), 649-69.

Hwang, S.-N., and T.-Y. Chang (2003), "Using Data Envelopment Analysis to Measure Hotel Managerial Efficiency Change in Taiwan," *Tourism Management* 24:4 (August), 357-69.

Ibourk, A., B. Maillard, S. Perelman and H. R. Sneesens (2004), "Aggregate Matching Efficiency: A Stochastic Production Frontier Approach, France 1990-1994," *Empirica* 31:1, 1-25.

Jamash, T., and M. Pollitt (2001), "Benchmarking and Regulation: International Electricity Experience," *Utilities Policy* 9:3 (September), 107-30.

Jeon, B. M., and R. C. Sickles (2004), "The Role of Environmental Factors in Growth Accounting: A Nonparametric Analysis," *Journal of Applied Econometrics* 19:5 (September-October), 567-91.

Jiménez, J. S., F. P. Chaparro and P. C. Smith (2003), "Evaluating the Introduction of a Quasi-Market in Community Care," *Socio-Economic Planning Sciences* 37, 1-13.

Jondrow, J., C. A. K. Lovell, I. Materov and P. Schmidt (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Model," *Journal of Econometrics* 19:2/3 (August), 233-38.

Jörss, M., D. E. Powell and C. Wolff (2004), "A **Streetcar** Named Productivity," *McKinsey Quarterly* Number 3.

Kao, C., and Y.-C. Lin (2004), "Evaluation of the University Libraries in Taiwan: Total Measure *versus* Ratio Measure," *Journal of the Operational Research Society* 55:12 (December), 1256-65.

Kennedy, J., and A. S. J. Smith (2004), "Assessing the Efficient Cost of Sustaining Britain's Rail Network - Perspectives Based on Zonal Comparisons," *Journal of Transport Economics and Policy* 38:2 (May), 157-90.

Kinnunen, K. (2005), "Pricing of Electricity Distribution: An Empirical Efficiency Study of Finland, Norway and Sweden," *Utilities Policy* 13:1 (March), 15-25.

Kirigia, J. M., A. Emrouznejad, L. G. Sambo, N. Munguti and W. Liambila (2004), "Using Data Envelopment Analysis to Measure the Technical Efficiency of Public Health Centers in Kenya," *Journal of Medical Systems* 28:2 (April), 155-66.

Knight, F. H. (1933 [1965]), *The Economic Organization*. New York: Harper & Row.

Kompas, T., T. N. Che and R. Q. Grafton (2004), "Technical Efficiency Effects of Input Controls: Evidence from Australia's Banana Prawn Fishery," *Applied Economics* 36:15, 1631-41.

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

Korhonen, P. J., and M. Luptacik (2004), "Eco-Efficiency Analysis of Power Plants: An Extension of Data Envelopment Analysis," *European Journal of Operational Research* 154:2 (April), 437-46.

Koundouri, P., and A. Xepapadeas (2004), "Estimating Accounting Prices for Common Pool Natural Resources: A Distance Function Approach," *Water Resources Research* 40:6 (April), Article No. W06S17.

Kumbhakar, S. C. (1997), "Modeling Allocative Inefficiency in a Translog Cost Function and Cost Share Equations: An Exact Relationship," *Journal of Econometrics* 76:1/2 (January/February), 351-56.

Kumbhakar, S. C., and L. Hjalmarsson (1995), "Labour-Use Efficiency in Swedish Social Insurance Offices," *Journal of Applied Econometrics* 10:1 (January-March), 33-47.

Kumbhakar, S. C., and C. A. K. Lovell (2000), *Stochastic Frontier Analysis*. New York: Cambridge University Press.

Kumbhakar, S. C., and E. G. Tsionas (2005), "Measuring Technical and Allocative Inefficiency in the Translog Cost System: A Bayesian Approach," *Journal of Econometrics* 126:2 (June), 355-84.

Kuosmanen, T., and T. Sipiläinen (2004), "On the Anatomy of Productivity Growth: A Decomposition of the Fisher Ideal TFP Index," Discussion Paper 2004:17, MTT Economic Research, Agrifood Research Finland, Helsinki, Finland. http://mtt.fi

Laine, J., M. Linna, A. Noro and U. Häkkinen (2005), "The Cost Efficiency and Clinical Quality of Institutional Long-Term Care for the Elderly," *Health Care Management Science* 8:2 (May), 149-56.

Land, K. C., C. A. K. Lovell and S. Thore (1993), "Chance-Constrained Data Envelopment Analysis," *Managerial and Decision Economics* 14:6 (November/December), 541-54.

Lau, L. J., and P. A. Yotopoulos (1971), "A Test for Relative Efficiency and Application to Indian Agriculture," *American Economic Review* 61:1 (March), 94-109.

Lauer, J. A., C. A. K. Lovell, D. B. Evans and C. J. L. Murray (2004), "World Health System Performance Revisited: The Impact of Varying the Relative Importance of Health System Goals," *BMC Health Services Research* 4:19 (July).

Lawrence, D., and A. Richards (2004), "Distributing the Gains from Waterfront Productivity Improvements," *Economic Record* 80 (September), S43-S52.

Leibenstein, H. (1966), "Allocative Efficiency vs. 'X-Efficiency'," *American Economic Review* 56:3 (June), 392-415.

Leibenstein, H. (1976), *Beyond Economic Man*. Cambridge, MA: Harvard University Press.

Lewis, D., T. M. Springer and R. I. Anderson (2003), "The Cost Efficiency of Real Estate Investment Trusts: An Analysis with a Bayesian Stochastic Frontier Model," *Journal of Real Estate Finance and Economics* 26:1 (January), 65-80.

Lewis, W. W. (2004), *The Power of Productivity*. Chicago: University of Chicago Press.

Lindmark, M. (2004), "Patterns of Historical CO_2 Intensity Transitions Among High- and Low-Income Countries," *Explorations in Economic History* 41:4 (October), 426-47.

Lindsay, C. M. (1976), "A Theory of Government Enterprise," *Journal of Political Economy* 84:5, 1061-77.

Linna, M., A. Nordblad and M. Koivu (2003), "Technical and Cost Efficiency of Oral Health Care Provision in Finnish Health Centres," *Social Science & Medicine* 56:2 (January), 343-53.

Lins, M. P. E., E. G. Gomes, J. C. C. B. Soares de Mello and A. J. R. Soares de Mello (2003), "Olympic Ranking Based on a Zero Sum Gains DEA Model," *European Journal of Operational Research* 148, 312-22.

Linton, J. D. (2002), "DEA: A Method for Ranking the Greenness of Design Decisions," *Journal of Mechanical Design* 124:2 (June), 145-50.

Liu, C., and R. S. Yin (2004), "Poverty Dynamics Revealed in Production Performance and Forestry in Improving Livelihoods: The Case of West Anhui, China," *Forest Policy and Economics* 6:3-4 (June), 391-401.

Loizides, J., and E. Tsionas (2004), "Productivity Growth in European Railways," *Journal of Transport Economics and Policy* 38:1 (January), 45-76.

Lovell, C. A. K. (2003), "The Decomposition of Malmquist Productivity Indexes," *Journal of Productivity Analysis* 20:3 (November), 437-58.

Lozano, S., G. Villa and B. Adenzo-Diaz (2004), "Centralised Target Setting for Regional Recycling Operations Using DEA," *Omega* 32:2, 101-10.

Luo, X. M., and N. Donthu (2005), "Assessing Advertising Media Spending Inefficiencies in Generating Sales," *Journal of Business Research* 58:1 (January), 28-36.

Mairesse, F., and P. Vanden Eeckaut (2002), "Museum Assessment and FDH Technology: Towards a Global Approach," *Journal of Cultural Economics* 26:4 (November), 261-86.

Malmquist, S. (1953), "Index Numbers and Indifference Surfaces," *Trabajos de Estadistica* 4, 209-42.

Mamatzakis, E. C. (2003), "Public Infrastructure and Productivity Growth in Greek Agriculture," *Agricultural Economics* 29, 169-80.

Martín, J. C., J. Gutiérrez and C. Román (2004), "Data Envelopment Analysis (DEA) Index to Measure the Accessibility Impacts of New Infrastructure Investments: The Case of the High-speed Train Corridor Madrid-Barcelona-French Border," *Regional Studies* 38:6 (August), 697-712.

Martinez-Cordero, F. J., and P. S. Leung (2004), "Sustainable Aquaculture and Producer Performance: Measurement of Environmentally Adjusted Productivity and Efficiency of a Sample of Shrimp Farms in Mexico," *Aquaculture* 241:1-4 (November), 249-68.

Maruyama, S., and T. Nakajima (2002), "The Productivity Analysis of Postal Services," Chapter 7 in M. A. Crew and P. R. Kleindorfer, eds., *Postal and Delivery Services: Delivering on Competition.* Boston: Kluwer Academic Publishers.

Mayston, D. J. (2003), "Measuring and Managing Educational Performance," *Journal of the Operational Research Society* 54:7 (July), 679-91.

McFadden, D. (1978), "Cost, Revenue and Profit Functions," in M. Fuss and D. McFadden, eds., *Production Economics: A Dual Approach to Theory and Applications, Volume 1.* Amsterdam: North-Holland.

Meeusen, W., and J. van den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review* 18:2 (June), 435-44.

Mensah, Y. M., and R. Werner (2003), "Cost Efficiency and Financial Flexibility in Institutions of Higher Education," *Journal of Accounting and Public Policy* 22:4 (July-August), 293-323.

Miller, D. M. (1984), "Profitability = Productivity + Price Recovery," *Harvard Business Review* 62:3 (May/June), 145-53.

Millimet, D. L. (2005), "Job Search Skills, Employer Size and Wages," *Applied Economics Letters* 12:2 (February), 95-100.

Mohan, N., and J. Ruggiero (2003), "Compensation Differences Between Male and Female CEOs for Publicly Traded Firms: A Nonparametric Analysis," *Journal of the Operational Research Society* 54:12 (December), 1242-48.

Moore, A. T., J. Nolan, G. F. Segal and M. Taylor (2001), *Competitive Cities: A Report Card on Efficiency in Service Delivery in America's Largest Cities.* http://www. Reason Public Policy Institute.org/compcity

Mundlak, Y. (1961), "Empirical Production Function Free of Management Bias," *Journal of Farm Economics* 43:1 (February), 44-56.

Netherlands Bureau for Economic Policy Analysis (2000), "Yardstick Competition: Theory, Design and Practice," Working Paper No. 133. http://www.cpb.nl/eng/pub/werkdoc/133

Nishimizu, M., and J. M. Page (1982), "Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia 1965-78," *Economic Journal* 92:368 (December), 920-36.

Niskanen, W. A. (1971), *Bureaucracy and Representative Government.* Chicago: Aldine Publishing Co.

Nyhan, R. C. (2002), "Benchmarking Tools: An Application to Juvenile Justice Facility Performance," *Prison Journal* 82:4 (December), 423-39.

Obata, T., and H. Ishii (2003), "A Method for Discriminating Efficient Candidates with Ranked Voting Data," *European Journal of Operational Research* 151:1 (November), 233-37.

Odeck, J., and A. Alkadi (2004), "The Performance of Subsidized Urban and Rural Public Bus Operators: Empirical Evidence from Norway," *Annals of Regional Science* 38:3 (September), 413-31.

OECD (2001), *OECD Productivity Manual: A Guide to the Measurement of Industry-Level and Aggregate Productivity Growth.* Paris: OECD. http://www.oecd.org/subject/growth/prod-manual.pdf

Olesen, O. B., and N. C. Petersen (1995), "Chance Constrained Efficiency Evaluation," *Management Science* 41, 442-57.

Orea, L. (2002), "Parametric Decomposition of a Generalized Malmquist Productivity Index," *Journal of Productivity Analysis* 18:1 (July), 5-22.

Otsuka, K., H. Chuma and Y. Hayami (1992), "Land and Labor Contracts in Agrarian Economies: Theories and Facts," *Journal of Economic Literature* 30:4 (December), 1965-2018.

Otsuki, T., I. W. Hardie and E. J. Reis (2002), "The Implication of Property Rights for Joint Agriculture - Timber Productivity in the Brazilian Amazon," *Environment and Development Economics* 7:2 (May), 299-323.

Oum, T. H., and C. Yu (2004), "Measuring Airports' Operating Efficiency: A Summary of the 2003 ATRS Global Airport Benchmarking Report," *Transportation Research Part E* 40:6 (November), 515-32.

Paradi, J. C., M. Asmild and P. C. Simak (2004), "Using DEA and Worst Practice DEA in Credit Risk Evaluation," *Journal of Productivity Analysis* 21:2 (March), 153-65.

Paul, S., B. S. Sahni and B. Biswal (2004), "Public Infrastructure and the Performance of Canadian Manufacturing Industries," *Southern Economic Journal* 70:4, 998-1011.

Pavcnik, N. (2002), "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies* 69:1 (January), 245-76.

Pencavel, J. (2001), *Worker Participation: Lessons from the Worker Co-ops of the Pacific Northwest.* New York: Russell Sage Foundation.

Pentzaroupoulos, G. C., and D. I. Giokas (2002), "Comparing the Operational Efficiency of the Main European Telecommunications Organizations: A Quantitative Analysis," *Telecommunications Policy* 26:11 (December), 595-606.

Pestieau, P., and H. Tulkens (1993), "Assessing and Explaining the Performance of Public Enterprises," *Finanz Archiv* 50:3, 293-323.

Petersen, N. C. (1990), "Data Envelopment Analysis on a Relaxed Set of Assumptions," *Management Science* 36:3 (March), 305-14.

Pimenta, A. A., R. G. Santos and S. C. Lagoa (2000), "Technical Efficiency in CTT - Correios de Portugal," Chapter 12 in M. A. Crew and P. R. Kleindorfer, eds., *Current Directions in Postal Reform*. Boston: Kluwer Academic Publishers.

Pitt, M., and L.-F. Lee (1981), "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry," *Journal of Development Economics* 9, 43-64.

Porembski, M., K. Breitenstein, and P. Alpar (2005), "Visualizing Efficiency and Reference Relations in Data Envelopment Analysis with an Application to the Branches of a German Bank," *Journal of Productivity Analysis* 23:2 (March), 203-21.

Porter, M. E. (1991), "America's Green Strategy," *Scientific American* April.

Prieto, A. M., and J. L. Zofio (2001), "Evaluating Effectiveness in Public Provision of Infrastructure and Equipment: The Case of Spanish Municipalities," *Journal of Productivity Analysis* 15:1 (January), 41-58.

Ravallion, M. (2005), "On Measuring Aggregate 'Social Efficiency'," *Economic Development and Cultural Change* 53:2 (January), 273-92.

Ray, S. C., and E. Desli (1997), "Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries: Comment," *American Economic Review* 87:5 (December), 1033-39.

Ray, S. C., and K. Mukherjee (1996), "Decomposition of the Fisher Ideal Index of Productivity: A Non-Parametric Dual Analysis of US Airlines Data," *Economic Journal* 106:439 (November), 1659-78.

Reichmann, G. (2004), "Measuring University Library Efficiency Using Data Envelopment Analysis," *Libri* 54:2 (June), 136-46.

Reinhard, S., C. A. K. Lovell and G. Thijssen (1999), "Econometric Estimation of Technical and Environmental Efficiency: An Application to Dutch Dairy Farms," *American Journal of Agricultural Economics* 81:1 (February), 44-60.

Richardson, J., J. Wildman and I. K. Robertson (2003), "A Critique of the World Health Organization's Evaluation of Health System Performance," *Health Economics* 12, 355-66.

Rosenman, R., and D. Friesner (2004), "Scope and Scale Inefficiencies in Physician Practices," *Health Economics* 13:11 (November), 1091-1116.

Rossi, M. A. (2001), "Technical Change and Efficiency Measures: The Post-Privatization in the Gas Distribution Sector in Argentina," *Energy Economics* 23, 295-304.

Russell, R. R. (1988), "On the Axiomatic Approach to the Measurement of Technical Efficiency," in W. Eichhorn, ed., *Measurement in Economics: Theory and Applications of Economic Indices.* Heidelberg: Physica-Verlag.

Russell, R. R. (1990), "Continuity of Measures of Technical Efficiency," *Journal of Economic Theory* 51:2 (August), 255-67.

Salerian, J. (2003), "Analysing the Performance of Firms Using a Decomposable Ideal Index Number to Link Profit, Prices and Productivity," *Australian Economic Review* 36:2 (June), 143-55.

Salinas-Jiminez, M. D. (2004), "Public Infrastructure and Private Productivity in the Spanish Regions," *Journal of Policy Modeling* 26:1 (January), 47-64.

Samuelson, P. A., and S. Swamy (1974), "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *American Economic Review* 64:4 (September), 566-93.

Sánchez, R. J., J. Hoffmann, A. Micco, G. V. Pizzolitto, M. Sgut and G. Wilmsmeier (2003), "Port Efficiency and International Trade: Port Efficiency as a Determinant of Maritime Transport Costs," *Maritime Economics & Logistics* 5:2 (June), 199-218.

Sarkis, J., and S. Talluri (2004), "Performance Based Clustering for Benchmarking of US Airports," *Transportation Research Part A*, 38:5 (June), 329-46.

Scheraga, C. A. (2004), "Operational Efficiency versus Financial Mobility in the Global Airline Industry: A Data Envelopment and Tobit Analysis," *Transportation Research Part A* 38:5 (June), 383-404.

Schmidt, P. (1976), "On the Statistical Estimation of Parametric Frontier Production Functions," *Review of Economics and Statistics* 58:2 (May), 238-39.

Schmidt, P., and R. Sickles (1984), "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics* 2:4 (October), 367-74.

Schor, A. (2004), "Heterogeneous Productivity Response to Tariff Reduction: Evidence from Brazilian Manufacturing Firms," *Journal of Development Economics* 75:2 (December), 373-96.

Serra, P. (2003), "Measuring the Performance of Chile's Tax Administration," *National Tax Journal* 56:2 (June), 373-83.

Serrano-Cinca, C., Y. Fuertes-Callén and C. Mar-Molinero (2005), "Measuring DEA Efficiency in Internet Companies," *Decision Support Systems* 38:4 (January), 557-73.

Sheldon, G. M. (2003), "The Efficiency of Public Employment Services: A Nonparametric Matching Function Analysis for Switzerland," *Journal of Productivity Analysis* 20:1 (July), 49-70.

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

Shephard, R. W. (1970), *Theory of Cost and Production Functions*. Princeton: Princeton University Press.

Sherman, H. D., and T. J. Rupert (2006), "Do Bank Mergers Have Hidden or Foregone Value? Realized and Unrealized Operating Synergies in One Bank Merger," *European Journal of Operational Research* 168:1 (January), 253-68.

Shim, W. (2003), "Applying DEA Technique to Library Evaluation in Academic Research Libraries," *Library Trends* 51:3 (Winter), 312-32.

Sickles, R. C., D. H. Good and L. Getachew (2002), "Specification of Distance Functions Using Semi- and Non-parametric Methods with an Application to the Dynamic Performance of Eastern and Western European Air Carriers," *Journal of Productivity Analysis* 17:1/2 (January), 133-55.

Sigala, M., P. Jones, A. Lockwood and D. Airey (2005), "Productivity in Hotels: A Stepwise Data Envelopment Analysis of Hotels' Rooms Division Processes," *Service Industries Journal* 25:1 (January), 61-82.

Silva Portela, M. C. A., and E. Thanassoulis (2005), "Profitability of a Sample of Portuguese Bank Branches and its Decomposition into Technical and Allocative Components," *European Journal of Operational Research* 162:3 (May), 850-66.

Simon, H. A. (1955), "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69:1 (February), 99-118.

Solow, R. M. (1957), "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39:3 (August), 312-20.

Southwick, Jr., L. (2005), "Sewer Plant Operating Efficiency, Patronage and Competition," *Managerial and Decision Economics* 26:1, 1-13.

Spottiswoode, C. (2000), *Improving Police Performance: A New Approach to Measuring Police Efficiency*. London: HM Treasury. http://www.hm-treasury.gov.uk

Stanford, R. E. (2004), "A Frontier Analysis Approach for Benchmarking Hospital Performance in the Treatment of Acute Myocardial Infarction," *Health Care Management Science* 7:2 (May), 145-54.

Stevenson, R. E. (1980), "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics* 13:1 (May), 58-66.

Stigler, G. J. (1976), "The Xistence of X-Efficiency," *American Economic Review* 66:1 (March), 213-16.

Stone, M. (2002), "How Not to Measure the Efficiency of Public Services (and How One Might)," *Journal of the Royal Statistical Society* Series A, 165, Part 3, 405-34.

Stone, R. (1980), "Whittling Away at the Residual: Some Thoughts on Denison's Growth Accounting," *Journal of Economic Literature* 18:4 (December), 1539-43.

Sun, S. (2004), "Assessing Joint Maintenance Shops in the Taiwanese Army Using Data Envelopment Analysis," *Journal of Operations Management* 22:3 (June), 233-45.

Takamura, Y., and K. Tone (2003), "A Comparative Site Evaluation Study for Relocating Japanese Government Agencies Out of Tokyo," *Socio-Economic Planning Sciences* 37, 85-102.

Thanassoulis, E. (2000), "DEA and its Use in the Regulation of Water Companies," *European Journal of Operational Research* 127, 1-13.

Thomas, P., Y. P. Chan, L. Lehmkuhl and W. Nixon (2002), "Obnoxious-Facility Location and Data Envelopment Analysis: A Combined Distance-Based Formulation," *European Journal of Operational Research* 141:3 (September), 494-514.

Thompson, R. G., F. Singleton, R. Thrall and B. Smith (1986), "Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas," *Interfaces* 16, 35-49.

Törnqvist, L. (1936), "The Bank of Finland's Consumption Price Index," *Bank of Finland Monthly Bulletin* 10, 1-8.

Troutt, M. D., M. Y. Hu and M. S. Shanker (2005), "A Distribution-Free Approach to Estimating Best Response Values with Application to Mutual Fund Performance Modeling," *European Journal of Operational Research* 166:2 (October), 520-27.

Tulkens, H. (1993), "On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit," *Journal of Productivity Analysis* 4:1/2 (June), 183-210.

Tulkens, H., and P. Vanden Eeckaut (1995), "Non-Parametric Efficiency, Progress and Regress Measures for Panel Data: Methodological Aspects," *European Journal of Operational Research* 80, 474-99.

Tulkens, H., and P. Vanden Eeckaut (1999), "Mesurer l'efficacité: avec ou sans frontières?," Chapter 3 in P.-Y. Badillo and J. C. Paradi, eds., *La méthode DEA: analyse des performances.* Paris: Hermes Science Publications

Tupper, H. C., and M. Resende (2004), "Efficiency and Regulatory Issues in the Brazilian Water and Sewage Sector: An Empirical Study," *Utilities Policy* 12:1 (March), 29-40.

Turner, H., R. Windle and M. Dresner (2004), "North American Containerport Productivity: 1984-1997," *Transportation Research Part E* 40:4 (July), 339-56.

Tybout, J. R., and M. D. Westbrook (1995), "Trade Liberalization and the Dimensions of Efficiency Change in Mexican Manufacturing Industries," *Journal of International Economics* 39:1-2 (August), 53-78.

Uri, N. D. (2004), "Measuring the Impact of Incentive Regulation on Technical Efficiency in Telecommunications in the United States," *Applied Mathematical Modeling* 28:3 (March), 255-71.

US Department of Labor, Bureau of Labor Statistics (2005), "Multifactor Productivity Trends, 2002". http://www.bls.gov/news.release/pdf/prod3.pdfm

Ventura, J., E. González and A. Cárcaba (2004), "Efficiency and Program-Contract Bargaining in Spanish Public Hospitals," *Annals of Public and Cooperative Economics* 75:4 (December), 549-73.

Wagner, J. M., D. G. Shimshak and M. A. Novak (2003), "Advances in Physician Profiling: The Use of DEA," *Socio-Economic Planning Sciences* 37, 141-63.

Wang, H.-J., and P. Schmidt (2002), "One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels," *Journal of Productivity Analysis* 18:2 (September), 129-44.

Wang, W. R. (2000), "Evaluating the Technical Efficiency of Large US Law Firms," *Applied Economics* 32:6 (May), 689-95.

Warning, S. (2004), "Performance Differences in German Higher Education: Empirical Analysis of Strategic Groups," *Review of Industrial Organization* 24:4 (June), 393-408.

Wen, H. J., B. Lim and H. L. Huang (2003), "Measuring E-Commerce Efficiency: A Data Envelopment Analysis (DEA) Approach," *Industrial Management & Data Systems* 103:9, 703-10.

Wheelock, D. C., and P. W. Wilson (2000), "Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions," *Review of Economics and Statistics* 82:1 (February), 127-38.

Widstrom, E., M. Linna and T. Niskanen (2004), "Productive Efficiency and its Determinants in the Finnish Public Dental Service," *Community Dentistry and Oral Epidemiology* 32:1 (February), 31-40.

Williamson, O. E. (1964), *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Williamson, O. E. (1975), *Markets and Hierarchies*. New York: The Free Press.

Williamson, O. E. (1985), *The Economic Institutions of Capitalism*. New York: The Free Press.

Winsten, C. B. (1957), "Discussion of Mr. Farrell's Paper," *Journal of the Royal Statistical Society* Series A, General, 120:3, 282-84.

Wisniewski, M., and A. Dickson (2001), "Measuring Performance in Dumfries and Galloway Constabulary with the Balanced Scorecard," *Journal of the Operational Research Society* 52:10, 1057-66.

World Health Organization (2000), *The World Health Report 2000. Health Systems: Improving Performance.* Geneva: WHO.

Worthington, A. C., and B. E. Dollery (2001), "Measuring Efficiency in Local Government: An Analysis of New South Wales Municipalities' Domestic Waste Management Function," *Policy Studies Journal* 29:2, 232-49.

Yang, Y. S., G. H. Lu, X. Y. Gu and R. Yamamoto (2003), "Greenness Assessment of Products in PLCA by DEA Approach," *Materials Transactions* 44:4 (April), 645-48.

Yoshida, Y., and H. Fujimoto (2004), "Japanese-Airport Benchmarking with the DEA and Endogenous-Weight TFP Methods: Testing the Criticism of Overinvestment in Japanese Regional Airports," *Transportation Research Part E* 40:6 (November), 533-46.

Yotopoulos, P. A., and L. J. Lau (1973), "A Test for Relative Economic Efficiency: Some Further Results," *American Economic Review* 63:1 (March), 214-23.

Yu, M.-M. (2004), "Measuring Physical Efficiency of Domestic Airports in Taiwan with Undesirable Outputs and Environmental Factors," *Journal of Air Transport Management* 10:5 (September), 295-303.

Zaim, O. (2004), "Measuring Environmental Performance of State Manufacturing Through Changes in Pollution Intensities: A DEA Framework," *Ecological Economics* 48, 37-47.

Zeitsch, J., D. Lawrence and J. Salerian (1994), "Comparing Like with Like in Productivity Studies: Apples, Oranges and Electricity," *Economic Record* 70:209 (June), 162-70.