Future Research Opportunities in Efficiency and Productivity Analysis

Knox Lovell

Available online at: www.uniovi.es/economia/edp.htm
Abstract
In this paper I speculate on what I consider to be some of the more interesting and fruitful directions for future research opportunities in the area of efficiency and productivity analysis. Most of the topics I discuss have been addressed by a few innovators, but their research has not yet diffused throughout the research community.
1. Introduction
Interest in efficiency and productivity analysis may be said to have originated in the 1950s, with the work of Koopmans (1951), Debreu (1951), Shephard (1953), Farrell (1957) and Solow (1957). However serious analytical and empirical research began twenty years later, with the contributions of Aigner et al. (1977), Meeusen and van den Broeck (1977), Charnes et al. (1978) and Caves et al. (1982a,b). Now, at the dawn of a new millennium, research into efficiency and productivity analysis has spread throughout the globe. Articles, books and dissertations are appearing in increasing numbers, and international workshops proliferate. This book is itself the result of such a workshop, and its Chapters provide a good indication of the current state of research in the field. It also provides me with an opportunity to take stock, not of where we have been but of where we are going.

In this essay I speculate on what I consider to be some of the more interesting and fruitful directions for future research opportunities in the area of efficiency and productivity analysis. Most of the issues I raise have been addressed by a few innovators, but their research has not yet diffused throughout the research community. Some of the issues arise primarily in data envelopment analysis (DEA), some arise primarily in stochastic frontier analysis (SFA), but most are common to both environments. Rather than discuss each issue in detail, I briefly note the significance of each issue and refer the reader to a few relevant sources.

2. The Theoretical Foundations of Efficiency Theory
We devote extensive resources to empirical investigations into the efficiency with which producers allocate resources. We must therefore have some theoretical framework that leads us to hypothesize that some producers are more efficient than others, or else the exercise would be pointless. I believe that the theoretical framework is provided by the extensive literature on agency, incentives and contracts. It is the agency relationship between regulators and owners, between owners and managers, and between managers and employees, that provides agents with limited opportunity to pursue their own objectives. It also provides principals with the incentive to design contracts that more closely align the incentives of agents with their own. This idea clearly underpins the literatures on incentive-compatible regulation and performance-based budgeting, but it has not yet spread to the literature on efficiency measurement. A notable exception is the work of Bogetoft (2000 and other references cited therein).
Bogetoft works within a principal-agent framework, and considers the relationship between a producer (the agent) and an owner or regulator (the principal). This relationship is characterized by asymmetric information favoring the agent, who can exploit this asymmetry by engaging in strategic behavior by claiming high costs associated with the production plan determined by the principal. The consequence of the information asymmetry and strategic behavior is a suboptimal outcome for the principal, and measured inefficiency for the analyst. In such a framework the principal’s objective is to design a production plan that minimizes the information rents the agent can capture, or that minimizes inefficiency in production. Thus Bogetoft’s first contribution is to link the theoretically oriented agency literature with the more practically oriented efficiency measurement literature. But he goes much further. He also demonstrates that if the principal uses cost norms based on DEA when designing future production plans, this limits the information rents the agent can extract, and generates plans that are optimal in the sense that they are biased toward cost-efficient plans. Thus his analysis forges a link between the evaluation of past performance and the creation of future production plans, with DEA providing the link. The potential managerial and regulatory applications of these insights are widespread.

3. Behavioral Objectives and Constraints
The vast majority of published DEA and SFA efficiency studies are based on what I would call “elementary” models of output (or revenue or profit) maximization, or of input (or cost) minimization. However I doubt seriously that many of the producers whose performance is being evaluated are actually confronted by such elementary optimization problems. I suspect they pursue more complex objectives and face constraints that are ignored by the researcher. In my judgement a bit of basic research and modeling creativity would go a long way toward improving the descriptive realism of our empirical models, not to mention enhancing the credibility of our findings. I can think of a number of ways in which our elementary models could be enriched by way of modifying objectives or adding constraints. I mention just one such enriched model that strikes me as being both widely applicable and unfortunately neglected.

Shephard (1974) introduced the notion of a budget-constrained output maximizing producer. Central to this framework is the cost-indirect output set. This is the set of output vectors that is both technologically and economically feasible, in the sense that
the set of output vectors is technologically feasible with any input vector satisfying an operating budget constraint. In this framework producers are free to choose both outputs and inputs, provided only that the desired input vector is affordable. Within this framework it is possible to evaluate producer performance in terms of their ability to maximize output, revenue or profit, subject to the constraints imposed by technology and by the input prices and operating budgets they face, but not constrained by exogenously determined input vectors.

Many producers in both the private sector and the public sector are assigned an annual operating budget, within which they are relatively free to select an economically feasible input vector in an effort to maximize output, revenue or profit. Consequently Shephard's framework is relevant to a large variety of situations, although it remains a sadly under-utilized framework. An illustration of the insights that can be achieved within the cost-indirect framework is provided by the Glass et al. (1998) study of productivity change in UK universities.

4. Endogeneity and Exogeneity

The issue of whether certain variables are endogenous or exogenous arises in at least four situations, and it is not always dealt with in a thoughtful manner.

Consider an elementary input-(output-) oriented DEA model. Here the question is: are all inputs (outputs) really endogenous? Or are some inputs (outputs) quasi-fixed, perhaps through contractual obligations? We have analytical models of subvector efficiency, and we have DEA software capable of allowing for non-controllable variables, but in my judgement these models and software options are under-utilized because researchers fail to think before formulating their models and conducting their empirical analyses.

Consider next the use of SFA to estimate a distance function, a currently popular exercise. Here the question is reversed: are all regressors, including both inputs and outputs, really exogenous? This seems unlikely, since then there is no economic optimization problem to be solved. Exogeneity tests exist, but they are rarely applied in this context. If a test is applied, and the exogeneity hypothesis is rejected, then conventional maximum likelihood is an inappropriate estimation procedure, and an instrumental variables procedure is called for, although it is rarely used in this context.
A pair of thoughtful investigations into procedures available for the estimation of distance functions is provided by Coelli (2001) and Rodriguez (2001).

Consider next the use of SFA to estimate a cost or profit frontier. Here the question becomes: are all included prices really exogenous? The exogeneity assumption unthinkingly permeates virtually the entire empirical literature on the estimation of cost or profit frontiers. However if producers enjoy either monopsony or monopoly power in some markets, the corresponding prices are endogenous rather than exogenous, and we confront the same problem as occurs with the estimation of distance functions. Perhaps more seriously, an inappropriate assumption of price exogeneity leads to the confusion of the appearance of relatively efficient behavior with the exploitation of market power.

Consider finally the construction of a productivity index, using prices to weight quantities. Fisher and Törnqvist index procedures are preferred, both because they satisfy long lists of desirable axioms and because they are superlative indexes, consistent with flexible approximations to the true but unknown underlying technology. However these indexes are superlative only under restrictive conditions involving the structure of technology and optimizing behavior in the face of exogenous market-determined prices. These restrictive conditions raise two issues. First, if prices are endogenous, perhaps due to the exploitation of market power, these indexes lose their superlative characterization because they are biased. Caves et al. (1980) encountered such a situation, in which railroad output prices were distorted by cross subsidization allowed by market power. They then developed a creative combination of indexing procedures on the input side and econometric estimation of shadow prices on the output side, to construct an unbiased (or less biased) Törnqvist railroad productivity index. Unfortunately their novel approach remains neglected. Second, if producers are inefficient and fail to optimize, then even if prices are exogenous, the superlative characterization of Fisher and Törnqvist productivity indexes needs to be reassessed.

5. Partial Adjustment Models
This issue is relevant to both DEA and SFA, although it originated within an econometric framework. It arises in an intertemporal context, with either time series data or panel data. I am unaware of any study in the efficiency measurement literature that has addressed this issue.
Assume that a producer is producing current output in a cost-efficient manner. Assume also that the producer knows what input vector will produce future output in a cost-efficient manner. However due to time delays, delivery lags and installation costs, the adjustment from current input use to desired future input use is imperfect. Because some inputs can be adjusted only slowly, other inputs must be over-adjusted to compensate. This partial adjustment process continues through a number of periods, and leads to input demand equations having lagged values of input use as regressors, unlike conventional input demand equations based on an implicit assumption of complete adjustment. Such partial adjustment models of producer behavior were developed over three decades ago; Fair (1969) and Nadiri and Rosen (1973) provide good examples.

The significance of the partial adjustment literature for the efficiency measurement literature is little appreciated. A producer can be following an optimal intertemporal path that incorporates costly and partial adjustment from one equilibrium to another. The failure to incorporate such costly and partial adjustment into the model can lead to the inappropriate labeling of an intertemporally efficient producer as being statically inefficient during the adjustment period. Two alternative research strategies would appear to be available. One would begin with a conventional partial adjustment model and add the possibility of inefficient behavior. The other would begin with a conventional frontier model and build in partial adjustment. I have not experimented with either alternative, but it is easier to envision such an exercise within an SFA framework than within a DEA framework.

6. Accounting for Variation in the Operating Environment

Many studies employ input and output quantity (and perhaps price) data to measure the technical (and perhaps economic) efficiency of producers. Such an exercise generates either of two possible outcomes: (i) producers are correctly labeled as being relatively efficient or inefficient; and (ii) producers are incorrectly labeled as being relatively efficient or inefficient because they enjoy relatively favorable, or endure relatively unfavorable, operating environments. It is of course desirable to incorporate features of producers' operating environments into the analysis. Only by doing so is it possible to distinguish controllable variation in managerial efficiency from largely uncontrollable variation in the operating environment. Flexible models have been
developed to incorporate environmental variables into efficiency analyses, but some of them are new and all of them remain under-utilized.

In DEA this can be accomplished in either of two ways. If the direction of the impact of the environmental variables is known in advance, these variables can be incorporated directly into the DEA model. However this procedure creates a degrees of freedom problem, since producers are evaluated only relative to other producers having more (or less) favorable operating environments. This approach is discussed by Muñiz (2001). Alternatively, regardless of whether the direction of the impact is known in advance, a multi-stage procedure is available. The first stage consists of the calculation of DEA efficiency scores, ignoring the environmental variables. In the second stage the initial DEA efficiency scores are regressed against the vector of environmental variables. Parameter estimates provide information on directions and magnitudes of environmental impacts on performance. If the second stage regression analysis is based on SFA, an environmental efficiency frontier is obtained, and the efficiency scores generated in the second stage are thereby adjusted for the impacts of the environmental variables. A good illustration of what can be accomplished with a multi-stage DEA/SFA approach to incorporating environmental variables is provided by Fried et al. (1999).

In SFA this can be accomplished in either of two ways as well. A standard approach is to include the environmental variables among the regressors in the deterministic kernel of the stochastic frontier. In this approach the environment is assumed to influence technology, but it is assumed not to influence efficiency. A common example is the incorporation of network characteristics in transportation studies. A second approach is to incorporate the environmental variables as determinants of efficiency in the one-sided error component. In this approach the environment is assumed to influence efficiency. A common example is the incorporation of various characteristics of farm managers in agricultural studies. The two approaches are discussed in Kumbhakar and Lovell (2000).

7. Statistical Properties of DEA and FDH
SFA has a random noise error component, and so is explicitly stochastic. This has two virtues. It allows for the incorporation of the effects of statistical noise, which permeates all economic data. It also permits the conduct of statistical inference, the
rigorous testing of various hypotheses concerning the structure of technology and the existence and statistical significance of inefficiency. DEA and FDH (the monotonic but non-convex free disposal hull approach to efficiency analysis) have no such random noise error term, and so both are ostensibly deterministic. This raises two questions. Can DEA and/or FDH nonetheless support rigorous statistical inference? Is it possible to develop an explicitly stochastic version of DEA and/or FDH?

Simar and Wilson (2000 and additional references cited therein) have answered the first question in the affirmative. They begin by carefully modeling the data generating process. This essential first step enables them to derive a statistical model that allows them to determine the statistical properties of the nonparametric DEA and FDH estimators. They obtain bias, consistency and speed of convergence results for both estimators. They also obtain asymptotic sampling distributions of the FDH estimator in a general multivariate setting, and of the DEA estimator in a bivariate setting. However since these results are asymptotic, and most data sets are "small," they devote special attention to modifications that must be made to familiar bootstrapping techniques in order to conduct statistical inference in small samples. The practical implication of their findings is that statistical inference based on either estimator is feasible. Support for their findings is provided by an extensive set of Monte Carlo simulations of DEA efficiency scores and hypothesis tests reported by Kittelsen (1999).

Cooper et al. (1996,1998) and Olesen and Petersen (1995,1999) have addressed the second question, and have provided a cautiously affirmative answer. They are developing a chance-constrained approach to DEA, in which either the envelopment constraints or the multiplier constraints are expressed in probabilistic form. The intuition behind the approach is that constraints are "probably" satisfied because data are noisy, having been drawn from some unknown probability distribution. Once distributional assumptions are made, chance constraints can be converted to certainty equivalents, the resulting model can be implemented, and chance-constrained efficiency scores can be calculated. However implementation requires that the researcher specify all parameters of the probability distribution from which the data are drawn, as well as a set of probabilities with which the constraints are required to be satisfied. Much work remains to be done on this important topic.
An intriguing alternative to the chance-constrained programming approach to DEA is provided by a fuzzy programming approach to both DEA and FDH. The programming notion that data are fuzzy is analogous to the statistical notion that data are noisy. The application of fuzzy programming to DEA and FDH proceeds in three stages. In the first stage fuzzy inputs and outputs are assigned to a membership set having "risk-free" and "impossible" bounds. Where in this set to assign each input and output obviously requires expert judgement, an element missing from most efficiency studies. In the second stage conventional DEA and FDH models are modified, taking into account the membership function and the bounds on each fuzzy input and output. In the final stage efficiency scores are calculated for alternative values of the membership function. This provides an analysis of the sensitivity of individual producer efficiency scores to variations in the fuzziness of individual inputs and outputs. A good introduction to this approach, with an illuminating empirical illustration, is provided by Triantis and Girod (1998).

8. Risk and Uncertainty
Most efficiency studies ignore risk and uncertainty. But the producers they study certainly face uncertainty, about technology reliability, about input performance, and about price trends. In the face of these uncertainties, they make decisions based on the degree of risk they are willing to bear. Economic models of producer behavior under uncertainty are decades old, an influential study being Just and Pope (1978), but they are barely beginning to be incorporated into efficiency studies.

The significance of incorporating risk into efficiency models is that inferences concerning the structure of technology and the existence and magnitude of inefficiency are sensitive the treatment of risk. Hughes and Mester (1998) accounted for variation in risk preference of banks by including financial capital in their SFA analyses, and found estimates of scale economies to be sensitive to the inclusion or exclusion of this risk preference indicator. Battese et al. (1997) provide an empirical application comparing the results of three models: a production function incorporating risk, a production frontier ignoring risk, and a production frontier incorporating risk. They found efficiency estimates to be sensitive to the treatment of risk.
9. **Weight Restrictions**

The use of weight restrictions is fairly common in DEA, having originated with the pioneering work of Thompson et al. (1986). The SFA analogue would be to impose parameter restrictions that would restrict, for example, marginal products or marginal rates of substitution. The use of such restrictions is virtually nonexistent in SFA, where standard practice is to test monotonicity and curvature properties after estimation rather than to impose restrictions prior to estimation.

Without weight restrictions, the envelopment procedures employed by both DEA and SFA (minimum extrapolation and maximum likelihood, respectively) are designed to make producers appear to be relatively efficient, perhaps more efficient than they actually are. They accomplish this by forcing technology to envelop the data as closely as possible, possibly at the cost of imposing an unlikely structure on technology. This can happen because the envelopment procedures are mathematical and statistical rather than economic. The dual implication of these envelopment procedures is that they enable producers to attach relatively high (low) implicit weights to those outputs in which they do (do not) specialize, and to those inputs which they use sparingly (heavily).

The problem with these envelopment procedures is that they can generate results in which the implicit weights fail to conform to engineering or economic reality, and make little or no sense to "experts." Accordingly, it is often desirable to impose prior constraints on weights, for example so as not to allow producers to assign unduly low weights to valuable resources. Admittedly, doing so imposes value judgements, and this makes it desirable that these judgements be imposed only after consultation with the experts. A good illustration of what can be done with weight restrictions in DEA is provided by Pedraja-Chaparro et al. (1997).

10. **DEA and MOLP**

Conventional DEA is value-free, in the sense that producers are allowed to assign weights to variables that put them in the most favorable light. As I mentioned in the previous Section, it is possible to restrict these weights to "sensible" ranges. Alternatively, it is possible to impose preferences over the range of possible DEA outcomes. The essence of multiple objective linear programming (MOLP) is that a principal has preferences over the range of outcomes generated by the agents under
evaluation. These preferences might lead the principal to select a preferred DEA-efficient outcome, or even to prefer a DEA-inefficient outcome to a DEA-efficient outcome if the former placed sufficiently greater emphasis on a variable preferred by the principal.

In a pair of recent papers, Joro et al. (1998) and Halme et al. (1999) have explored the relationship between conventional DEA and MOLP, and the possibility of merging the two techniques. Essentially they superimpose a principal's value function (which must be approximated as part of the analysis) on a modified DEA model. They assume only that the principal's value function is strictly increasing in outputs, strictly decreasing in inputs, and pseudoconcave (because then a local optimum over a convex DEA technology set is also a global optimum). They also provide simple empirical applications comparing DEA efficiency with what they call "value" efficiency, which is analogous to the economist's notion of economic efficiency. These initial applications suggest a potentially widespread use of MOLP to solve augmented DEA problems that incorporate the preferences of a principal (a home office manager or a regulator, for example).

11. Bayesian Econometrics
The use of Bayesian techniques to estimate stochastic frontiers is relatively new, having originated in the work of van den Broeck et al. (1994), and it does not yet have many practitioners and it has not yet exerted much of an influence. The question is: will it become influential? I do not have an answer to this question, but Kim and Schmidt (2000) provide the first credible empirical evidence I have seen.

Kim and Schmidt compare a pair of classical models, a fixed effects model that imposes no distributional assumption on inefficiency and a maximum likelihood model that does, with an analogous pair of Bayesian models, one with an uninformative prior and the other with an informative prior. Based on findings from three panel data sets, Kim and Schmidt draw two conclusions. First, they find large gains, in terms of tightness of confidence intervals and tightness of posterior distributions, from the imposition of distributional assumptions and informative priors. Second, they find little difference between results obtained from the preferred classical and Bayesian models that do impose distributional assumptions and informative priors.
12. Undesirable Outputs

In conventional efficiency studies a form of monotonicity is assumed, so that all variables are strongly, or freely, disposable. Consider, however, a producer using inputs to produce two outputs, a desirable output that is marketed at a positive price, and an undesirable byproduct that is not marketed. The mill using pulp, chemicals and fresh water from a stream to produce paper and downstream chemical pollution provides a good example. The paper output is marketed at a price, and so is strongly disposable. Years ago, when the environmental regulators were not watching, the chemical pollution byproduct was also freely disposable, since it could be discharged at no (private) cost. But since there is an obvious social cost of the stream pollution, the environmental regulators are now watching. The imposition of penalties such as effluent fees means that the undesirable byproduct is now weakly disposable, since either the undertaking of abatement activities or the payment of effluent fees makes disposability costly.

It is desirable to extend conventional efficiency analysis to the situation in which not all outputs are freely disposable. Such an extension would provide two benefits. It would provide a framework for the measurement of the efficiency of producers in an environment in which either abatement or disposal of byproducts is costly. It would also provide a framework for establishing how costly abatement is, or for calculating marginal costs of abatement, for various types of producers.

All the tools are in place. A theoretical framework embodying weak disposability exists. A nonparametric DEA model capable of incorporating undesirable byproducts has been developed by Färe et al. (1989). More recently a parametric SFA model capable of incorporating undesirable byproducts has been developed by Reinhard et al. (1999). Both models are capable of calculating technical efficiency, productivity change and marginal abatement costs for each producer. Both models are also capable of measuring the environmental efficiency of each producer, essentially as a subvector efficiency measure. The measurement of environmental efficiency is also discussed by Hernández (2001).

What is missing is empirical applications. This is doubtless a consequence of the difficulty of obtaining reliable data on environmental emissions, although the widespread public interest in the detrimental environmental impacts of various
production activities is presumably relaxing the data constraint. The empirical applications I have in mind would address three issues: (i) the identification of the environmental efficiency laggards most in need of government attention; (ii) the calculation of marginal abatement costs to compare with "willingness to pay" estimates; and (iii) investigation into the merits of the Porter (1991) hypothesis, which asserts that environmental regulation is likely to be beneficial to regulated firms. The current paucity of empirical applications is to be lamented, since the next environmental efficiency study will add far more value than the next hundred banking efficiency studies will, where the data constraint is not a problem.

13. Labor Markets
Earnings (wage rates or incomes) vary widely across individuals, perhaps more so now than ever before. Ever since the notion of human capital was formalized in the 1950s, economists have been estimating earnings functions, with the aim of explaining variation in individual earnings. The idea is that individual earnings should depend on human capital, as proxied by education, training and experience, and may also depend on other individual characteristics as well, such as gender and race. A worthwhile extension to this literature would be to use SFA or DEA techniques to construct an earnings frontier, the best that can be achieved with the human capital at hand, conditional on observable individual characteristics. Such an exercise would shed new light on the maximum returns to various human capital indicators, and this information might be more reliable than previous information since it would incorporate the possibility of inefficiency. The exercise would also support the calculation of the efficiency with which individuals or groups approach the earnings frontier. Measured inefficiency could then be attributed to wage ignorance on the part of individuals, or to discrimination on the part of employers.

Once again the theory and the empirical models are in place. Lovell (1995) surveyed the extant literature, which was scarce, and asked for more research into labor market inefficiency and its causes. Recently Ibourk and Perelman (1999) have reported on an innovative study using SFA and DEA to construct a labor market matching frontier against which to evaluate the efficiency with which vacancies and unemployment are used to create new employment. Aside from this study, I am still waiting.
14. Efficiency, Productivity and Financial Performance

Economists are interested in the efficiency with which businesses operate. Business people are concerned with their profitability. There must be a cause-and-effect relationship between efficiency and profitability. Other things being equal, the more efficient a business is, the more profitable it should be, and the faster is its rate of productivity growth, the more rapidly its profitability should increase. The problem is that other things are rarely equal. This makes it desirable to decompose profit variation into variation in efficiency or productivity, and variation in other sources.

Years ago in the managerial accounting literature variation in business profit was decomposed into variation in prices (called "price recovery," reflecting the ability to recover input price increases through output price increases), variation in the scale of operation (the "activity effect") and variation in productivity (the "productivity effect"). A standard reference is Miller (1984). More recently economists have exploited the tools of modern production theory to enrich this profit decomposition literature. Recent contributions include Grifell-Tatjé and Lovell (1999) and Han and Hughes (1999). The two decompositions differ in their details, but both decompose profit variation into a price effect, a scale effect, a technical change effect, and an operating efficiency change effect. Grifell-Tatjé and Lovell use DEA to implement their decomposition, while Han and Hughes use index number techniques to implement their decomposition.

As with the issues I raised in the two previous Sections, what is needed now is new empirical applications of the decomposition models. This is an important linkage that needs to be established and quantified across a range of sectors. Han and Hughes note, for example, that their framework can be used by regulators to cap prices or revenues in a way that fairly distributes the financial benefits of productivity gains to producers and consumers. Waters and Street (1998) apply a simplified version of the profit decomposition to the related problem of monitoring the productive and financial performance of public enterprises.

15. Conclusions

In this paper have endeavored to identify a variety of important topics in need of greater research effort in the future. The list of topics can easily be expanded. Indeed this is an abbreviated version of a list I have presented in recent forums around the
world. At these forums other scholars have suggested additional topics. Rather than list, much less discuss, these additional topics, I wish to conclude with two pleas.

My first plea concerns the relationship we have with the sector whose performance we are evaluating. Either (i) we are using the data because the data are available, and we need the data only as a vehicle for showing off our fancy techniques so that we can get another article published, or (ii) we are using our fancy techniques in an earnest effort to learn something about a sector in which we have a serious interest. All too often it is the former; what else could explain the proliferation of efficiency studies of banking and electricity? I plea for greater interaction between researchers and representatives of the sector we are evaluating. They can provide us with valuable insights that may lead us to modify our preconceived models. In this way they can provide us with free, if sometimes painful, referee reports that may make our studies institutionally credible as well as academically acceptable.

My second plea concerns the data constraint. We all complain about the inadequacies of the data to which we apply our wonderful models. Unfortunately the data constraint is getting worse, because the reasonably measurable ("factory") sector of advanced economies is shrinking relative to the less easily measurable (service) sector. In this regard I can do no better than to urge readers to read Griliches (1994).
References


