Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP01/2008

Title
Econometric Estimation of an Input Distance Function in a System of Equations

Authors
Tim Coelli, Gholamreza Hajargasht and C.A. Knox Lovell

Date: 2008

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Econometric Estimation of an Input Distance Function in a System of Equations

by

Tim Coelli,* Gholamreza Hajargasht and C.A. Knox Lovell

Centre for Efficiency and Productivity Analysis

School of Economics

University of Queensland

Abstract

In this study we propose a methodology that allows one to obtain consistent estimates of a system of equations involving an input distance function along with the first order equations that relate to shadow cost minimizing behaviour. In addition, we show that previously proposed methods are likely to produce inconsistent estimates, even under a fairly weak set of assumptions regarding the data generating process (DGP). Our model is closely related to a random effects shadow prices model recently proposed by Karagiannis et al (2006). However, in our model we express the first order equations in ratio form, which allows us to ensure that our estimates are invariant to the choice of normalizing input. An empirical application of this model involving panel data on US electricity generation firms is presented, where we find that technical inefficiency is the largest contributor to cost inefficiency, and that the majority of allocative mistakes involve under use of fuel relative to the other inputs.

Keywords: Input distance function, system of equations, shadow cost minimization

JEL Codes: C30, D24

* Corresponding Author is Tim Coelli, School of Economics, University of Queensland, St Lucia QLD 4072, Australia, ph +61733656470, fax +61733657299, email t.coelli@uq.edu.au.

1. Introduction

In this paper we describe the outcomes of a research project that began some years ago. The aim of the study is to identify a suitable way of estimating a system of equations involving an input distance function along with the associated first order conditions for shadow cost minimisation. The estimation of a model such as this allows one to obtain information on the structure of production (estimates of production elasticities, economies of scale and scope, etc.) plus firm-specific measures of technical and allocative efficiencies. Given that a handful of authors had already worked on systems estimation of distance functions, we expected that this task would be fairly straight forward. However, we found that it was a very challenging task.

In the following discussion we review a number of previous studies in this area. In doing so we are critical of some aspects of these studies. However, we wish to stress that our aim is not criticism in itself, since we think that these past papers make a number of valuable contributions, but our principal aim is to understand these methods and attempt to refine them if possible.[1]

Why are we interested in estimating a distance function embedded in a system of equations? Why do we not simply estimate the input distance function as a single equation? Systems estimation has a number of potential advantages. First, the inclusion of extra information may result in more efficient econometric estimates of the parameters. Second, it allows one to formally test the hypothesis of systematic deviations from cost minimising behaviour. Third, the issue of potentially endogenous regressors in the distance function could be addressed using these first order equations. Fourth, the systems approach may permit one to obtain firm-specific allocative inefficiency measures as a by-product of estimation. This could allow one to avoid the necessity for the calculation of allocative inefficiency measures in a

---

[1] In fact the reader should note that one of the co-authors of this paper (Lovell) is also the co-author of two of the past papers that we review and find fault with.

second stage, which generally involves the solution of a non-linear optimising problem for each data point when a flexible functional form is used.[2]

Grosskopf and Hayes (1993) appear to be the first authors to have estimated an input distance function system. Their model involves a generalized Leontief functional form, and is estimated using seemingly unrelated regression (SUR) methods. Possible endogeneity in the input variables is dealt with by regressing each input quantity variable on a vector of instruments and then using the input quantity predictions in the SUR estimation. Unfortunately, this method is unlikely to provide consistent estimates when a non-linear model (such as the generalised Leontief) is used.[3]

Technical inefficiency in the input distance function equation in the Grosskopf and Hayes (1993) model is accommodated by using the moment-based estimator proposed in Aigner, Lovell and Schmidt (1977). The specified model motivates allocative mistakes in terms of shadow prices deviating from observed prices. The error terms in the first order equations are designed to capture these mistakes. However, it appears that the model implicitly imposes the restriction that for the average firm observed and shadow prices coincide (because the error terms are assumed to have zero mean), which is born out in the results they obtain. Furthermore, the analysis does not attempt to predict optimal input combinations, nor allocative or cost efficiency scores.

Subsequent papers have proposed models which do not assume that the shadow price deviations must have zero mean. In a series of papers, Baños-Pino, Fernandez-Blanco and Rodriguez-Alvarez (2002), Rodriguez-Alvarez and Lovell (2004) and Rodriguez-Alvarez, Fernandez-Blanco and Lovell (2004) make use of a translog functional form, and allow a non-zero mean for the shadow price deviations. The proposed econometric models are fairly simple to estimate, however our analysis below suggests that the estimators are unlikely to produce consistent estimators because of

---

[2] An alternative way to achieve many of these aims could be to estimate a shadow cost function in a system of equations. However, this is a messy exercise which has a number of challenging aspects. Greene (2007) surveys the issues, and Kumbhakar and Tsionas (2005) propose a Bayesian approach.

[3] For a general discussion of instrumental variables estimation in non-linear models, see Amemiya (1985).

correlation between the error terms and the regressors (unless all firms make identical allocative mistakes).

In a recent working paper, Karagiannis et al (2006) have proposed some alternative models that can be used to estimate a shadow prices system. Their method is designed to address the inconsistency problems in the above models. However, our analysis suggests that their models can be criticised because they are not invariant to the choice of normalising input, and in addition their fixed effects models are likely to produce inconsistent estimates when allocative mistakes differ (in a non-structured manner) across observations.

In addition to these papers that model allocative mistakes using shadow prices, another series of papers – by Atkinson and Primont (2002), Atkinson, Honerkamp and Cornwell (2003), Atkinson, Färe and Primont (2003) and Atkinson and Halabi (2005) – have proposed methods which model the allocative mistakes in terms of shadow input quantities. This approach has the advantage that optimal input combinations, plus allocative or cost efficiency scores are easily obtained as a by-product of the estimation process. However, some challenging (perhaps insurmountable) estimation issues are also encountered in these models when allocative mistakes differ (in a non-structured manner) across observations. These issues are discussed in some detail below.

Overall, from our initial assessment of these two alternatives (the use of shadow input prices versus shadow input quantities) it seemed to us that the use of shadow input quantities was the more natural way to proceed, since the standard underlying economic model (of shadow cost minimisation) assumes that input quantities are endogenous (choice) variables while price information is assumed exogenous (i.e., firms are price takers, such that they are too small to influence the market price by their actions). However, as we argue below, the data generating process (DGP) implied by the use of shadow input quantities produces an econometric model with a number of (apparently intractable) estimation issues. As a result, the model that we propose in this paper involves the use of shadow input prices.

The remainder of this paper is divided into sections. In Section 2 we outline a data generating process that accommodates both management errors and non-management errors. In Section 3 we investigate the viability of some econometric models that involve shadow input quantities, while in Section 4 we evaluate econometric models involving shadow input prices. In Section 5 we describe how efficiency measures can be derived in these two cases. In Section 6 we present our proposed methodology, and then in Section 7 we provide an empirical application of this model involving panel data on US electricity companies. Some concluding comments are made in Section 8.

2. The data generating process

In recent decades, very few econometrics papers carefully outline the assumed data generating process (DGP) which underlies the model that is being estimated.[4] In many papers one will find some hints as to what the author may believe forms part of the DGP, such as some mention of measurement error or endogenous feedback, but rarely is the DGP clearly defined.

The foundation stone of this paper is the DGP. This is required, because without an assumed DGP, it is very difficult for one to discuss the relative merits of alternative approaches to the modelling of a production process using an input distance function that is estimated econometrically in a system of equations, along with the first order conditions for (shadow) cost minimisation.

To begin with we define the following variables:

$\mathbf{x} = (x_1, ..., x_K) \in R_K^+$ is a $K\times 1$ vector of input quantities;

$\mathbf{y} = (y_1, ..., y_M) \in R_M^+$ is $M\times 1$ vector of output quantities; and

$\mathbf{w} = (w_1, ..., w_K) \in R_K^+$ is a $K\times 1$ vector of input prices.

Following Färe and Primont (1995), we define the production technology as

$$T = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \text{ can produce } \mathbf{y}\}. \qquad (2.1)$$

[4] Papers such as those by Simar and Wilson (2000, 2007) which develop bootstrapping methods for non-parametric data envelopment analysis (DEA) models are notable exceptions. However, most recent papers involving econometric analyses of production models have said very little about assumed DGP's, compared to earlier papers such as that by Zellner, Kmenta and Drèze (1966).

Given the assumption of weak disposability in inputs, this production technology can be also represented by an input distance function

$$D(\mathbf{y},\mathbf{x}) = \sup_{\pi}\{\pi : (\mathbf{x}/\pi, \mathbf{y}) \in T\}, \tag{2.2}$$

where π is the (scalar) distance function value, such that $1 < \pi < \infty$. A value of $\pi = 1$ implies no technical inefficiency. That is, the firm is operating on the surface of the production technology.

We assume that the firm faces exogenously determined vectors of input prices and output quantities, and attempts to select a vector of input quantities so as to minimise the cost of producing this vector of outputs. In this situation (and given that the technology is convex and strongly disposable in inputs) the technology can be equivalently described using the cost function

$$C(\mathbf{y},\mathbf{w}) = \min_{\mathbf{x}}\left\{\mathbf{w}'\mathbf{x} \mid D(\mathbf{y},\mathbf{x}) \geq 1\right\}. \tag{2.3}$$

The first order conditions associated with this minimisation problem are:[5]

$$\frac{\partial D(\mathbf{y},\mathbf{x})}{\partial \mathbf{x}} = \frac{\mathbf{w}}{C(\mathbf{y},\mathbf{w})}, \tag{2.4}$$

or equivalently

$$\frac{\partial D(\mathbf{y},\mathbf{x})}{\partial x_i} = \frac{w_i}{C(\mathbf{y},\mathbf{w})}, \quad i = 1,\ldots,K. \tag{2.4a}$$

This relation can also be expressed in terms of the elasticities

$$\frac{\partial \ln D(\mathbf{y},\mathbf{x})}{\partial \ln x_i} = \frac{w_i x_i}{C(\mathbf{y},\mathbf{w})}, \quad i = 1,\ldots,K, \tag{2.4b}$$

which is useful when dealing with logarithmic functional forms, such as the translog.

---

[5] See Färe and Primont (1995).

A world containing no errors

In the first instance we could assume that we have a special *error-free world* where the management of the firm is always perfect. That is, it chooses a technically efficient input vector on the boundary of the production technology, such that $D(\mathbf{y},\mathbf{x})=1$, and the technically efficient input vector is also allocatively efficient, such that $\mathbf{w}'\mathbf{x} = C(\mathbf{y},\mathbf{w})$. Consequently cost is minimised and equation (2.4) is always satisfied.

A world containing management errors

We then relax this strong assumption and introduce errors due to management inefficiency into this model. Following Farrell (1957), we introduce two types of management inefficiency: *technical inefficiency* (producing below the production technology) and *allocative inefficiency* (choosing an input mix which differs from that at the point of cost minimisation).

There are two alternative ways in which analysts normally introduce technical inefficiency. One option is to append a one-sided error term to the output quantities, to reflect the degree to which the achieved output falls short of the potential output. The second option is to append a one-sided error term to the input quantities, to reflect the degree to which the quantity of input used exceeds the minimum feasible input level. Given that we are considering an input distance function in this paper, it is natural for one to consider the latter option.[6] Thus we allow for the possibility that $\mathbf{x}$ is technically inefficient and define the technically efficient input vector as $\mathbf{x}^{te}=(x_1^{te},...,x_K^{te})$, where

$$x_i^{te} = x_i \exp(-u) \quad ,i=1,...,K , \tag{2.5}$$

[6] When an input-conserving orientation has been selected, $D(\mathbf{y},\mathbf{x})$ and $C(\mathbf{y},\mathbf{w})$ provide primal and dual representations of the technology. In this context modelling technical efficiency with an output-expanding orientation seems contradictory, but several studies have compared input-oriented and output-oriented estimates of technical efficiency in this context, an early and influential example being Atkinson and Cornwell (1994).

and u is a non-negative scalar.[7]

In the case of modelling allocative inefficiency (in inputs) we also have two alternative choices, in terms of how we can introduce the errors. One option is to specify variables that are known as *shadow input prices*, which represent the input price vector that would render the technically efficient scaling of the observed input quantity vector allocatively efficient. The second option is to specify *shadow input quantities*, which represent the technically efficient input quantity vector that is allocatively efficient for the observed input price vector. For the present, we will use the latter option, since the primary interest is generally in determining the inefficient input quantities and hence the efficiency scores. However, we will return to discussion of the former option in the next section.

We define the shadow input vector as the cost efficient input vector $\mathbf{x}^{ce} = (x_1^{ce},...,x_K^{ce})$, where

$$x_i^{ce} / x_i^{te} = \exp(\eta_i) \quad ,i = 1,...,K \ , \tag{2.6}$$

where $\boldsymbol{\eta} = (\eta_1,...,\eta_K)$ is a K×1 vector of scalar input quantity adjustments.[8] In the event that the firm uses the cost-minimising input mix, this vector will be a vector of zeros.

The vector, $\boldsymbol{\eta}$, has K elements, but only K-1 of them are uniquely defined. This is because the production technology is given and hence movement from the point $\mathbf{x}^{te}$ to $\mathbf{x}^{ce}$ can be fully described by the differences in the directions of the two rays which pass through these two points (because the points themselves will be defined by the intersections of these rays and the technology).

The direction of a ray from the origin which passes through the point $\mathbf{x}^{te}$ can be described by deflating the vector by an arbitrarily chosen element:

---

[7] A multiplicative form has been chosen here because it allows one to avoid the possibility that a negative shadow quantity is specified. Note that $\exp(u) = \pi$ = D(y,x).

[8] It follows from equations (2.5) and (2.6) that the cost-efficient input vector can be expressed in terms of the observed input vector as $x_i^{ce} = \exp(\eta_i)\, x_i \ \exp(-u) = x_i \ \exp(\eta_i - u) \quad ,i = 1,...,K$

$\mathbf{x}^{te} / x_K^{te} = (x_1^{te} / x_K^{te}, ..., x_{K-1}^{te} / x_K^{te}, 1)$. Thus, we have a set of *K*-1 ratios which fully describe the direction of this ray. In a similar way, we can define the direction of the $\mathbf{x}^{ce}$ vector as: $\mathbf{x}^{ce} / x_K^{ce} = (x_1^{ce} / x_K^{ce}, ..., x_{K-1}^{ce} / x_K^{ce}, 1)$.

The degree to which these two directions differ can then be fully described by the K-1 ratios

$$\begin{aligned}\frac{\mathbf{x}^{ce} / x_K^{ce}}{\mathbf{x}^{te} / x_K^{te}} &= \left[\frac{x_1^{ce} / x_K^{ce}}{x_1^{te} / x_K^{te}}, ..., \frac{x_{K-1}^{ce} / x_K^{ce}}{x_{K-1}^{te} / x_K^{te}}, 1\right] \\ &= \left[\exp(\eta_1 - \eta_K), ..., \exp(\eta_{K-1} - \eta_K), 1\right], \\ &= \left[\exp(\theta_1), ..., \exp(\theta_{K-1}), 1\right]\end{aligned} \tag{2.7}$$

where, once again, if $\boldsymbol{\theta} = (\theta_1, ..., \theta_{K-1})$ is a vector of zeros, this implies allocative efficiency.

However, it is important to note that the parameterisation in equation (2.7) is not identical to setting one of the η_i equal to 0, such as η_K. This is because this would result in the imposition of the assumption that the *K*-th shadow input quantity is such that $x_K^{ce} / x_K^{te} = 1$. That is, the *K*-th shadow input quantity is equal to the *K*-th technically efficient input quantity. This assumption has been made in Atkinson et al (2002), for example. As we explain later in this paper, this is an unusual restriction which will have the effect of producing incorrect efficiency measures when the assumption is (almost invariably) false.

As a consequence, it is important that the economic model is defined in terms of the ratio variables and the $\boldsymbol{\theta}$ vector from equation (2.7). That is, such that for the *i*-th input we have

$$\frac{x_i^{ce} / x_K^{ce}}{x_i^{te} / x_K^{te}} = \exp(\theta_i), \quad i = 1, ..., K-1. \tag{2.8}$$

This proves to be a natural thing to do since the input distance function must be, by definition, homogenous of degree 1 in input quantities,[9] which is normally achieved via deflation by an arbitrarily chosen input variable.[10]

Another way of writing equation (2.8) is as

$$x_i^{ce} / x_K^{ce} = \exp(\theta_i) x_i^{te} / x_K^{te} = \exp(\theta_i) x_i \ / x_K , \quad i = 1,\ldots,K\text{-}1, \qquad (2.9)$$

where the second equality holds because the radial technical efficiency (from equation 2.5) cancels out in the ratio of technically efficient input quantities.

In Figure 1 we provide a diagrammatic representation of these efficiency concepts using a simple two-input example. We have drawn an isoquant, Isoq(**y**), representing the boundary of the production technology (for a given output vector, **y**), and an iso-cost line, IC(**w**), reflecting the input price ratio. The point of tangency between Isoq(**y**) and IC(**w**) provides the shadow input vector $\mathbf{x}^s$, where cost is minimised, and so the shadow input vector $\mathbf{x}^s$ is the cost minimising input vector $\mathbf{x}^{ce.}$ Furthermore, proportional contraction of the **x** vector by multiplying by $\exp(-u)$, until it reaches the boundary of the technology, produces the technically efficient input vector, $\mathbf{x}^{te} = \mathbf{x}\exp(-u)$.

An iso-cost line corresponding to shadow prices, IC($\mathbf{w}^s$), is also presented in Figure 1. Shadow prices, $\mathbf{w}^s$, are those prices that would ensure that the technically efficient input vector, $\mathbf{x}^{te}$, was also the cost minimising input vector. The shadow price concept provides an alternative way (relative to shadow input quantities) of reflecting deviations from allocative efficiency. This option is discussed in the next section.

---

[9] See Färe and Primont (1995).
[10] This problem is closely related to a parallel debate regarding normalisation procedures in cost function estimation. For more information, see Kumbhakar and Karagiannis (2004) and references cited therein.

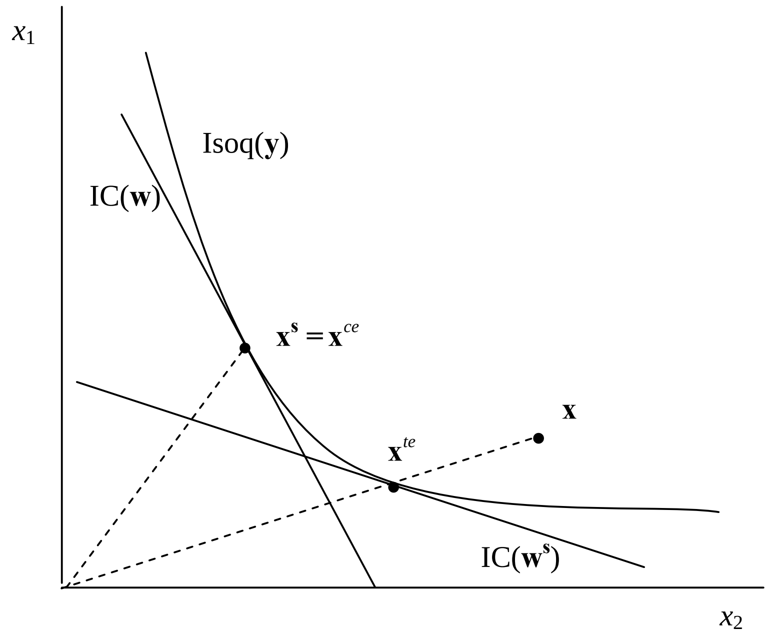

Figure 1: Shadow cost minimisation

<u>A world containing management errors and uncontrollable errors</u>

Next we need to introduce some sources of "noise" into this model. Before doing this we must first carefully ask the question: What factors lead to technical inefficiency and allocative inefficiency? In both cases we assume that it is due to poor management. For example, consider the case of a dairy factory that converts raw milk into cheese, butter, etc. Examples of technical inefficiency could be allowing some of the raw milk to spoil because of not carefully monitoring the vat temperatures, or could be scheduling labour shifts so that in some periods there are idle staff while in other periods there is a shortage of staff.

Examples of allocative inefficiency could be choosing the wrong capital to labour mix, perhaps as a consequence of not being fully aware of all technology options or alternatively incorrectly forecasting the time-path of wage rates and interest rates when choosing a piece of long-lived processing equipment (e.g. with a life of 10-15 years).

The above errors are assumed to be under the control of the managers. Alternatively, those errors which are arguably not under the control of managers could include unanticipated random events such as a drought which may reduce the supply of raw milk; a labour strike or an industrial accident disrupting production; or a sudden change in government policy affecting wage rates or fuel prices, etc.

However, it is not clear where one should draw the line between those factors that the manager should be able to control or reasonably foresee and those that are "random". The issue of bad luck versus bad management is a fuzzy area indeed. Is it reasonable to expect that a manager should be able to forecast wage changes due to macroeconomic cycles but not those due to government policy changes? Could one argue that a good manager should be carefully reading the newspapers so that he/she can anticipate the effects of government policy changes as well?

Anyway, for now we will assume that we are able to conceptually differentiate between errors due to bad luck and those due to bad management. Where do we introduce these new random (uncontrollable) error terms into our model? For items such as industrial accidents we could append error terms to the outputs (cheese and butter), while for labour strikes and droughts we could append error terms to the input quantities. For the case of unanticipated changes in wage rates we could append error terms to the input prices or to the input quantities, to explain why the firm is not operating at the optimal point. To be consistent with our current choice in modelling input allocative efficiency deviations via shadow input quantities, we will put this error term on the input quantities for now.

<u>A world containing management errors, uncontrollable errors and econometrician errors</u>

Up until now we have implicitly assumed that the econometrician is "perfect". That is, perfect in terms of measuring all variables and in specifying the model (functional form, variables to include, etc.). This is unlikely to be true in practice. First we consider the possibility of measurement error. Have we correctly measured the prices faced by the firm? Consider the case of wage rates. In most cases an econometrician will either use wage rates reported by a statistical agency for a particular geographical region (e.g., a country or a region within the country) which are an average for all industries or for a broad industry group (such as the manufacturing sector). They will hence be only an approximate measure of the actual wage rates paid by a particular milk factory. Alternatively we could have data on both wage costs and hours worked for each firm, which will allow us to calculate an implicit wage rate for each firm. This may appear to be a much better source of wage rate data, but if different firms employ different mixes of skilled/unskilled workers then this wage rate measure is likely to contain a lot of measurement error as well.

We could provide a similar discussion of likely measurement errors in all other input prices (in particular, capital will be problematic). Furthermore, there are likely to be measurement errors in input quantities, where measurement of the flow of services from capital will be a large source of headaches, and labour measures such as hours worked will be affected by quality and skill differences across firms. The raw milk input may also be affected by measurement errors. For example, via mistakes in recording quantities collected, or in trying to deal with differences in milk fat contents across different regions – where we could use either litres of milk or kilograms of milk fat as our output measure – neither of which are able to capture all the dimensions of milk quality – and hence are likely to be subject to measurement error.

Output measures are unlikely to escape measurement error issues either. For example, how do we measure the quantity of cheese produced given that a range of different products (cheddar, camembert, etc.) are produced in the different factories? We could use revenue as a proxy, but if different factories face different prices (for a fixed product quality) we will obtain errors. Alternatively if we use physical quantity of cheese in total kilograms, different product mixes will introduce errors. Clearly measurement errors are also a potential problem in output quantities.[11]

An additional issue is that of model specification errors – in functional form and variable selection. If we choose a translog functional form for our econometric model (for example), it will provide a second order approximation to the true functional relationship, and hence some approximation errors are likely to remain. Our model might also suffer from omitted variable bias – for example due to the omission of a variable reflecting the different regulatory environments in which different firms operate, etc. In this paper we will aggregate the errors due to variable measurement

[11] In many of the cases we discuss, the aggregation of different items could be achieved by the use of index number formula, such as the Fisher index, if sufficient data was available. This would reduce measurement error, but is unlikely to remove it completely.

and model specification together into a single category that we will label "econometrician error".

<u>Some notation</u>

We now have identified three different categories of errors, which we label:

1. management error;
2. uncontrollable error (bad weather, strikes, etc.); and
3. econometrician error.

The notation for the management (inefficiency) error terms has already been defined. The notation for the uncontrollable error terms will be τ_i^x for the error associated with the *i*-th input quantity and τ_i^y for the error associated with the *i*-th output quantity. The notation for the econometrician error terms will be e_i^x for the error associated with the *i*-th input quantity, e_i^y for the error associated with the *i*-th output quantity and e_i^w for the error associated with the *i*-th input price.

Note that all error terms could either increase or decrease the value of the associated variable, with the exception of the technical inefficiency error, which can only increase inputs. Also note that all error terms are multiplicative in nature. This is done because measures of technical efficiency (TE) and allocative efficiency (AE) are generally defined in a multiplicative manner, and logarithmic functional forms are generally used in these models (i.e., the Cobb-Douglas and translog forms).[12]

Given these error terms, we can also define notation for various different versions of the quantity and price variables in our model. For example, for the case of input quantities we define the following notation:

1. *observed* input quantities, $\mathbf{x}$;
2. *actual* input quantities, $\mathbf{x}^{ac}$ (observed input quantities with econometrician errors removed);

[12] Multiplicative errors also have the advantages that the error term prevents the possibility of negative price or quantity measures and provides a natural way of mitigating heteroskedastic problems.

3. *planned* input quantities, $\mathbf{x}^{pl}$ (actual input quantities with unpredictable errors removed);
4. *technically efficient* input quantities, $\mathbf{x}^{te}$ (planned input quantities with the technical inefficiency error removed); and
5. *cost efficient* input quantities, $\mathbf{x}^{ce}$ (technically efficient input quantities with the allocative inefficiency errors removed).

Hence, using our newly defined notation we can write

$$x_i^{ce} = x_i \exp(-u_i + \eta_i + v_i^x + e_i^x) . \tag{2.10}$$

That is, the cost efficient value of the *i*-th input is equal to the observed value, combined with technical, allocative, unpredictable and econometrician errors. In a similar manner we can describe the components of observed output quantities and observed input prices as:

$$y_i^{pl} = y_i \exp(v_i^y + e_i^y) \tag{2.11}$$

and

$$w_i^{ac} = w_i \exp(e_i^w) , \tag{2.12}$$

respectively.

For the case of output quantities we can define:

1. *observed* output quantities, $\mathbf{y}$;
2. *actual* output quantities, $\mathbf{y}^{ac}$ (observed output quantities with econometrician errors removed); and
3. *planned* output quantities, $\mathbf{y}^{pl}$ (actual output quantities with unpredictable errors removed).

Furthermore, for the case of input prices we can define:

1. *observed* input prices, $\mathbf{w}$; and
2. *actual* input prices, $\mathbf{w}^{ac}$ (observed input prices with econometrician errors removed).

Can we identify all of these errors?

Observationally, it is practically impossible to distinguish between errors due to "unpredictable events" and those due to "econometrician error". Hence, one normally lumps these two items together into a single category called "random errors" – which we will denote by ε_i^x for the *i*-th input quantity, ε_i^y for the error associated with the *i*-th output quantity and ε_i^w for the error associated with the *i*-th input price.

It is also difficult to distinguish between random errors and inefficiency errors, unless one is willing to make some additional assumptions. For example, if one has access to panel data, one can assume that inefficiency is invariant over time, and hence use a fixed effects or random effects panel data model to disentangle these two sources of error. However, if the panel is longer than a few years, the assumption of time-invariant inefficiency becomes questionable. In this instance some authors instead try to account for time variation by allowing the inefficiency parameters to be a function of a time trend.

If one does not have access to panel data (i.e., one has only a single cross-section of data on a group of firms) then the options narrow. In the case of technical inefficiency, one can use distributional assumptions, such as assuming that the technical inefficiency error term has a half-normal distribution while the noise terms have normal distributions, to allow one to estimate the model using maximum likelihood.[13] For the allocative efficiency side of things one is normally forced to assume that the errors in the first-order equations are either all due to allocative inefficiency or alternatively all due to random noise. In the later case one can introduce a parameter to allow some degree of allocative inefficiency, but this parameter would take a fixed value across all the firms, which would be a strong assumption in most situations.[14]

---

[13] For example, see the stochastic production frontier model described in Aigner *et al* (1977).

[14] For example, see Baños-Pino et al. (2002). It is also possible to weaken this assumption by making the allocative mistakes a systematic function of firm-specific factors that are likely to reflect management ability, such as the age, education and experience levels of the manager. For example, see Ferrier & Lovell (1990).

In the following two sections we investigate the feasibility of econometric estimation of a system of equations that involves both management and non-management errors. In particular, in Section 3 we consider the case of shadow input quantities while in Section 4 we consider shadow input prices.

3. An econometric model involving shadow input quantities

We specify a translog functional form for the production technology

$$\begin{aligned}\ln D(\mathbf{y},\mathbf{x}) = \alpha_0 + \sum_{i=1}^{K}\alpha_i \ln x_i + \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K}\alpha_{ij}\ln x_i \ln x_j + \sum_{i=1}^{M}\beta_i \ln y_i \\ + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\beta_{ij}\ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{M}\gamma_{ij}\ln x_i \ln y_j\end{aligned}, \qquad (3.1)$$

where the associated first order equations are, from (2.4b):

$$\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j} = \frac{\partial \ln D(\mathbf{y},\mathbf{x})}{\partial \ln(x_i)}, \quad i = 1,\ldots,K ,$$

or equivalently

$$S_i = \frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j} = \alpha_i + \sum_{j=1}^{K}\alpha_{ij}\ln x_j + \sum_{j=1}^{M}\gamma_{ij}\ln y_j, \quad i = 1,\ldots,K . \qquad (3.1a)$$

Homogeneity of degree one in inputs implies the parametric restrictions

$$\sum_{i=1}^{K}\alpha_i = 1 \qquad \sum_{j=1}^{K}\alpha_{ij} = 0 \quad \text{and} \quad \sum_{j=1}^{J}\gamma_{ij} = 0, \quad i = 1,\ldots,K , \qquad (3.2)$$

while the symmetry restrictions due to Young's theorem are

$$\alpha_{ij} = \alpha_{ji} \quad \text{and} \quad \beta_{ij} = \beta_{ji}, \quad \forall i, j . \qquad (3.2a)$$

The imposition of the homogeneity restrictions is equivalent to dividing all inputs and the distance term by an arbitrarily chosen input. If we choose the *K*-th input we obtain[15]

$$\begin{aligned}-\ln x_K &= \alpha_0 + \sum_{i=1}^{K-1}\alpha_i \ln\left(x_i / x_K\right) + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_i / x_K\right)\ln\left(x_j / x_K\right) \\ &+ \sum_{i=1}^{M}\beta_i \ln y_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\beta_{ij}\ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{M}\gamma_{ij}\ln\left(x_i / x_K\right)\ln y_j\end{aligned} \tag{3.3}$$

$$\frac{w_i x_i / x_K}{\sum_{j=1}^{K} w_j x_j / x_K} = \alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{ij}\ln y_j, \quad i = 1,\dots,K-1 \tag{3.3a}$$

and the *K*-th share equation will be

$$S_K = 1 - \sum_{j=1}^{K-1} S_j \,. \tag{3.3b}$$

<u>A model with management errors</u>

In the above model we have implicitly assumed a world free of errors. That is, in terms of management errors, we have assumed that

$$x_i = x_i^{te} = x_i^{ce} \,.$$

If we allow the possibility of management errors such as technical inefficiency, we would then have

$$x_i^{te} = x_i \exp(-u)$$

or in logarithms

$$\ln x_i^{te} = \ln x_i - u \,.$$

[15] Note that we multiply the left hand side of the share equations by x_K / x_K so that all input variables are in ratio form.

If we replace every occurrence of x_i in the system of equations in (3.3) with $x_i \exp(-u)$ we find that the *u* term only appears on the end of the distance equation, because the homogeneity condition ensures that it cancels out in all the other ratio terms. That is, we obtain:[16]

$$\begin{aligned}-\ln x_K &= \alpha_0 + \sum_{i=1}^{K-1}\alpha_i \ln\left(x_i / x_K\right) + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_i / x_K\right)\ln\left(x_j / x_K\right) \\ &\quad + \sum_{i=1}^{M}\beta_i \ln y_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\beta_{ij}\ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{M}\gamma_{ij}\ln\left(x_i / x_K\right)\ln y_j + u\end{aligned}, \quad (3.4)$$

$$\frac{w_i x_i / x_K}{\sum_{j=1}^{K} w_j x_j / x_K} = \alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{ij}\ln y_j, \quad i = 1,\ldots,K-1. \qquad (3.4a)$$

Furthermore, if we permit allocative inefficiency we have

$$x_i^{ce} / x_K^{ce} = \exp(\theta_i) x_i / x_K$$

or in logarithms

$$\ln\left(x_i^{ce} / x_K^{ce}\right) = \ln\left(x_i / x_K\right) + \theta_i .$$

If we now replace every occurrence of x_i / x_K in the first order equations in (3.4a) with $\exp(\theta_i) x_i / x_K$ we obtain the following set of equations[17]

$$\begin{aligned}-\ln x_K &= \alpha_0 + \sum_{i=1}^{K-1}\alpha_i \ln\left(x_i / x_K\right) + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_i / x_K\right)\ln\left(x_j / x_K\right) \\ &\quad + \sum_{i=1}^{M}\beta_i \ln y_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\beta_{ij}\ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{M}\gamma_{ij}\ln\left(x_i / x_K\right)\ln y_j + u\end{aligned}, \quad (3.5)$$

---

[16] Since the *K*-th share equation needs to be omitted when SUR estimation is applied (because of singularity in the covariance matrix), we do not include it here.
[17] Note that there is no need for one to adjust the distance function equation with this allocative inefficiency error because the observed data point will by definition differ from the frontier surface by the amount of technical inefficiency.

$$\frac{w_i x_i / x_K \exp\theta_i}{\sum_{j=1}^{K} w_j x_j / x_K \exp\theta_j} = \alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\left[\ln\left(x_i / x_K\right) + \theta_i\right] + \sum_{j=1}^{M}\gamma_{ij}\ln y_j, \quad i = 1,...,K-1. \tag{3.5a}$$

In this set of equations and in many other equations in this paper we suppress the notation for the observational unit – so as to reduce notational clutter. If this was not our practice, in the case of cross-sectional data we would normally put a *n* subscript on the inputs, outputs and error terms to indicate that these vary across firms (n=1,2,…,N),[18] while the values of the α, β and γ parameters are assumed to be fixed across observations. Thus we should emphasise that the values of the *u* and θ_i management errors in these models vary across observations.

The above point is important because in some papers (e.g., Atkinson and Primont, 2002), the authors argue that estimating a model in which the θ_i are random variables (perhaps with a non-zero mean) is too difficult, and hence suggest that it is easier to model them parametrically. However, since one is unable to estimate a model with more parameters than observations, this requires the imposition of restrictions – such as the assumption that the θ_i take a fixed value over all firms. As we shall explain shortly, assumptions such as this are in conflict with observed data and are hence likely to produce questionable results.[19]

Estimation

Now let us consider the estimation of this system of equations using econometric methods. It is evident that one can safely estimate the distance function in equation (3.5) using a single equation method, such as the COLS method,[20] even though the observed data that is being used is not allocatively efficient.[21] However the share equations are more problematic. The θ_i terms appear in both a linear and an exponential form. It appears to be impossible for one to isolate these terms so that the

---

[18] If we had panel data we would also use a *t* subscript to denote time periods (t=1,2,…,T).
[19] This comment applies equally to models in which the allocative efficiency parameters are made a systematic function of a time trend variable or some other exogenous factors.
[20] See Lovell et al. (1994) or Coelli and Perelman (1999) for discussion of the COLS method.
[21] This is essentially the result provided in Coelli (2000).

likelihood function can be derived. Thus MLE does not appear to be a feasible option here.

Perhaps we could alternatively use GMM? To this end, let us rearrange equation (3.5a) so that all random terms are included in an aggregate error term.[22] The new system we obtain is:

$$
\begin{aligned}
-\ln x_K = \alpha_0 + \sum_{i=1}^{K-1} \alpha_i \ln\left(x_i / x_K\right) + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_i / x_K\right) \ln\left(x_j / x_K\right) \\
+ \sum_{i=1}^{M} \beta_i \ln y_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \beta_{ij} \ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{M} \gamma_{ij} \ln\left(x_i / x_K\right) \ln y_j + \xi_d
\end{aligned}
\tag{3.6}
$$

and

$$
\frac{w_i x_i / x_K}{\sum_{j=1}^{K} w_j x_j / x_K} = \alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_j / x_K\right) + \sum_{j=1}^{M} \gamma_{ij} \ln y_{ji} + \xi_i, \quad i = 1,\ldots,K-1. \tag{3.6a}
$$

The error terms in these equations have the form

$$
\xi_d = u \tag{3.7}
$$

and

$$
\begin{aligned}
\xi_i = \sum_{j=1}^{K-1} \alpha_{ij}\theta_j + \frac{w_i x_i / x_K}{\sum_{j=1}^{K} w_j x_j / x_K} - \frac{w_i x_i / x_K \exp\theta_i}{\sum_{j=1}^{K} w_j x_j / x_K \exp\theta_j}, \\
i = 1,\ldots,K-1.
\end{aligned}
\tag{3.7a}
$$

It is not clear how one could use GMM to estimate the share equations in equation (3.6a) because the error term in (3.7a) is a function of the input quantities and input prices. Some past studies have used input prices as instruments in this type of model. However, our derivations suggest that this may not be wise – unless the allocative

[22] Here we have assumed that the θ_i have zero means to simplify the discussion. If this was not the case we would have *K*-1 extra parameters to estimate, but the conclusions regarding the viability of GMM would not be changed.

inefficiency term *truly* follows the restricted parametric structure implied by the model.

A model with management and uncontrollable errors

Now let us consider the case where we consider the possibility of random errors (i.e., the ε's) in our model. In this case we take the system of equations in (3.5) that contains management errors and replace every occurrence of x_i, y_i and w_i with $x_i \exp(\varepsilon_i^x)$, $y_i \exp(\varepsilon_i^y)$ and $w_i \exp(\varepsilon_i^w)$, respectively.

In this case the model becomes

$$\begin{aligned}
-\ln x_K &= \alpha_0 + \sum_{i=1}^{K-1} \alpha_i \left[\ln\left(x_i / x_K\right) + \varepsilon_i^x - \varepsilon_K^x \right] \\
&+ \frac{1}{2} \sum_{i=1}^{K-1} \sum_{j=1}^{K-1} \alpha_{ij} \left[\ln\left(x_i / x_K\right) + \varepsilon_i^x - \varepsilon_K^x \right] \left[\ln\left(x_j / x_K\right) + \varepsilon_j^x - \varepsilon_K^x \right] \\
&+ \sum_{i=1}^{M} \beta_i \left[\ln y_i + \varepsilon_i^y \right] + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{ij} \left[\ln y_i + \varepsilon_i^y \right] \left[\ln y_j + \varepsilon_j^y \right] \\
&+ \frac{1}{2} \sum_{i=1}^{K-1} \sum_{j=1}^{M} \gamma_{ij} \left[\ln\left(x_i / x_K\right) + \varepsilon_i^x - \varepsilon_K^x \right] \left[\ln y_j + \varepsilon_j^y \right] + u
\end{aligned}, \tag{3.8}$$

$$\begin{aligned}
S_i = \frac{w_i x_i / x_K \exp\left(\theta_i \varepsilon_i^x \varepsilon_i^w\right)}{\sum_{j=1}^{K} w_j x_j / x_K \exp\left(\theta_j \varepsilon_j^x \varepsilon_j^w\right)} &= \alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \left[\ln\left(x_j / x_K\right) + \theta_j + \varepsilon_j^x - \varepsilon_K^x \right] \\
&+ \sum_{j=1}^{M} \gamma_{ij} \left[\ln y_j + \varepsilon_j^y \right], \quad i = 1, \ldots, K-1
\end{aligned} \tag{3.8a}$$

In this case the corresponding error terms in equations (3.6-3.6a) will now contain the following more complicated expressions

$$\begin{aligned}
\xi_d &= \sum_{i=1}^{K-1} \alpha_{ii} (\varepsilon_i^x - \varepsilon_K^x) + \sum_{i=1}^{K-1} \sum_{j=1}^{K-1} \alpha_{ij} (\varepsilon_i^x - \varepsilon_K^x) \ln\left(x_j / x_K\right) \\
&+ \frac{1}{2} \sum_{i=1}^{K-1} \sum_{j=1}^{K-1} \alpha_{ij} (\varepsilon_i^x - \varepsilon_K^x)(\varepsilon_j^x - \varepsilon_K^x) \\
&+ \sum_{i=1}^{M} \beta_i \varepsilon_j^y + \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{ij} \varepsilon_i^y \varepsilon_j^y + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \beta_{ij} \varepsilon_i^y \ln y_j \\
&+ \frac{1}{2} \sum_{i=1}^{K-1} \sum_{j=1}^{M} \gamma_{ij} (\varepsilon_i^x - \varepsilon_K^x)(\ln y_j + \varepsilon_j^y) + u
\end{aligned} \tag{3.9}$$

$$\xi_i = \sum_{j=1}^{K-1} \alpha_{ij}(\theta_j + \varepsilon_j^x - \varepsilon_K^x) + \sum_{j=1}^{M} \gamma_{ij}\varepsilon_j^y + \frac{w_i x_i / x_K}{\sum_{j=1}^{K} w_j x_j / x_K} - \frac{w_i x_i / x_K \exp(\theta_i \varepsilon_i^w \varepsilon_i^x)}{\sum_{j=1}^{K} w_j x_j / x_K \exp(\theta_j \varepsilon_j^w \varepsilon_j^x)}, \quad i = 1, \ldots, K-1 \tag{3.9a}$$

With these new random error terms included, estimation becomes even more complicated. First, it is clear that all equations in the system are likely to be affected by an errors-in-variables problem if one attempted to use OLS or SUR. One could alternatively attempt to use GMM – however the choice of instruments seems to be even more limited than before, given that the error terms are functions of the inputs, outputs and input prices.[23] Finally, the MLE option is again infeasible because of the non-linear way in which the error terms enter the first order equations.

Past papers

Atkinson and Primont (2002) estimate a translog input distance function system using panel data on US electric utilities.[24] Their model is motivated in terms of shadow input quantities. In their econometric model they model technical inefficiency as a firm-specific parametric (quadratic) function of time, while they model allocative inefficiency deviations as non-firm-specific parametric (cubic) functions of time.

Using their notation, shadow input quantities, x_i^*, are defined as those variables which solve the cost minimisation problem (that, is the problem defined in our equation 2.3). They permit observed input quantities to differ from these shadow input quantities via the notation $x_i^* = x_i k_i$ (Atkinson and Primont, 2002, p206). In estimation they add the restriction that for one input variable (they select last one listed in the input vector) the k_i value is restricted to be equal to one for all firms and all time periods (Atkinson and Primont, 2002, p212). This restriction imposes the assumption that for one input

[23] Also, even if one was able to identify good instruments, the model is non-linear and hence IV estimation will be generally inconsistent (Amemiya, 1985).

[24] Note that this paper uses similar methods to those used in Atkinson, Honerkamp and Cornwell (2003) and Atkinson, Färe and Primont (2003).

variable the observed quantity and the shadow quantity must be equal. This case is illustrated in Figure 2, where the imposition of the constraint that $x_1^* = x_1$ means that the shadow input vector, $\mathbf{x}*$, does not correspond to the cost minimising input vector, $\mathbf{x}^{ce}$, except by chance. This in turn implies that estimates of allocative and cost efficiency will almost invariably be biased.

Atkinson and Primont (2002) argue that it is more convenient to specify the distance function in a form similar to equation (3.1) and impose the required homogeneity restrictions during estimation, rather than expressing it in the deflated form as in equation (3.3). As they correctly note, this has no effect on the estimates obtained. Furthermore, they choose to specify the first order conditions with input prices used as dependent variables[25] and they also append to each equation an "error term with zero mean". However there is no discussion of the likely sources of these errors.

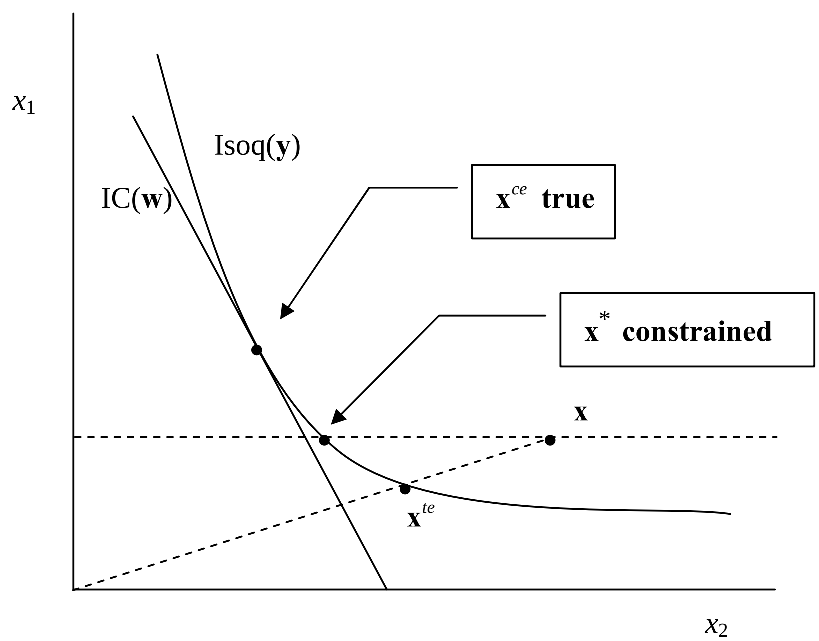

Figure 2: Constraining one input to be optimal

[25] This seems a bit unusual, since their theoretical model has input prices as exogenous variables.

Thus, it appears that their share equations would be (using our notation)

$$w_i = \frac{\sum_{j=1}^{K} w_j x_j \exp\theta_j}{x_i \exp\theta_i} \left\{ \alpha_i + \sum_{j=1}^{K} \alpha_{ij} \left[\ln x_i + \theta_i \right] + \sum_{j=1}^{M} \gamma_{ij} \ln y_j \right\} + \lambda_i, \qquad (3.10)$$
$$i = 1, ..., K,$$

where λ_i is an error term with zero mean.

The system of equations involving the distance function and these K input price equations is estimated using GMM involving the use of the Newey and West (1987) heteroskedasticity and autocorrelation adjustments. The authors refer to possible endogeneity in right-hand side variables, but are not explicit about the likely sources of these problems, and hence it is difficult to judge the quality of their chosen instruments from a theoretical perspective. They estimate their model using instrument sets that include output quantities and input prices, plus dummy variables and time trends.[26] They test for the validity of over-identifying restrictions using the Hansen (1982) *J*-test and conclude that their instrument set is valid.

It is important to emphasise that this particular model involves the brave assumption that all firms in the sample have *exactly the same allocative mistake attributes in any one particular year*. That is, (using our notation) one must have $\theta_{int} = \theta_{it}$ for all n=1,...,N. This is very unlikely given that different managers tend to have different skill sets.[27]

Given that the assumption is incorrect, we have $\theta_{int} = \theta_{it} + \tau_{int}$, where τ_{int} is a zero mean random variable. Substituting this into equation (3.10), rearranging and including firm and time subscripts we obtain

---

[26] This suggests that they believe that the input price and output quantity variables are exogenous, as would be implied from their theoretical model, yet they do not explicitly state this.
[27] Note that the imposition of similar parametric restrictions (relating to allocative mistakes) have been widespread in both cost and input distance function systems.

$$w_{nti} = \frac{\sum_{j=1}^{K} w_{ntj} x_{ntj} \exp\theta_{tj}}{x_{nti} \exp\theta_{ti}} \left\{ \alpha_i + \sum_{j=1}^{K} \alpha_{ij} \left[\ln x_{ntj} + \theta_{tj} \right] + \sum_{j=1}^{M} \gamma_{ij} \ln y_{ntj} \right\} + \xi_{nti}, \quad i = 1,...,K-1 \tag{3.11}$$

where

$$\xi_{nti} = \frac{\sum_{j=1}^{K} w_{ntj} x_{ntj} \exp(\theta_{tj} + \tau_{ntj})}{x_{nti} \exp(\theta_{tj} + \tau_{ntj})} \left\{ \alpha_i + \sum_{j=1}^{K} \alpha_{ij} \left[\ln x_{ntj} + (\theta_{tj} + \tau_{ntj}) \right] + \sum_{j=1}^{M} \gamma_{ij} \ln y_{ntj} \right\}$$
$$- \frac{\sum_{j=1}^{K} w_{ntj} x_{ntj} \exp\theta_{tj}}{x_{nti} \exp\theta_{ti}} \left\{ \alpha_i + \sum_{j=1}^{K} \alpha_{ij} \left[\ln x_{ntj} + \theta_{tj} \right] + \sum_{j=1}^{M} \gamma_{ij} \ln y_{ntj} \right\} + \lambda_i, \quad i = 1,...,K-1. \tag{3.11}$$

Thus we see that the disturbance term contains input quantities, input prices and output quantities. Hence, it is apparent that GMM involving any of these instruments will produce inconsistent estimates – if the $\theta_{int} = \theta_{it}$ assumption is incorrect.

4. An econometric model involving shadow input prices

Given that the estimation of a shadow input quantity model appears to be very challenging, let us now instead consider the shadow input prices option. In this case we replace equation (2.6) with an equation that describes the relationship between the input prices and the shadow input prices, $\mathbf{w}^s = (w_1^s,...,w_K^s)$

$$w_i^s / w_i = \exp(\kappa_i) \quad ,i = 1,...,K, \tag{4.1}$$

where $\boldsymbol{\kappa} = (\kappa_1,...,\kappa_K)$ is a K×1 vector of scalar input price adjustments. In the event that the firm uses the cost-minimising input mix, this vector will be a vector of zeros.

With this set of shadow prices we now take equations (3.5-3.5a) and do two things. First, we set the θ_i to zero, since we are not modelling shadow input quantities. Second, given allocative inefficiency of the form

$$w_i^s = \exp(\kappa_i) w_i \ ,$$

we replace every occurrence of w_i with $\exp(\kappa_i)w_i$ to obtain

$$\begin{aligned}-\ln x_K = \alpha_0 + \sum_{i=1}^{K-1}\alpha_i \ln\left(x_i / x_K\right) + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_i / x_K\right)\ln\left(x_j / x_K\right) \\ + \sum_{i=1}^{M}\beta_i \ln y_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\beta_{ij}\ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{M}\gamma_{ij}\ln\left(x_i / x_K\right)\ln y_j + u\end{aligned} \quad ,(4.3)$$

$$\frac{x_i\, w_i\, \exp(\kappa_i)}{\sum_{j=1}^{K} x_j w_j \exp(\kappa_j)} = \alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{ij}\ln y_j , \qquad i = 1, \ldots, K-1 \quad (4.3a)$$

The main difference between equations (4.3a) and (3.5a) is that the allocative efficiency parameter no longer appears on the right hand side of the share equations. One would expect that this would make estimation easier,[28] but this is not the case.

Past papers

Baños-Pino *et al* (2002) estimate a translog input distance function along with the set of share equations, using time-series data on Spanish railways.[29] They assume that the firms attempt to minimise cost, and hence that input quantities are endogenous, while output quantities and input prices are exogenous. They motivate the possibility of allocative inefficiency via a discussion of deviations between observed price ratios and shadow price ratios.[30]

To estimate a system of equations involving the input distance function and the above share equations they appear to be implicitly rearranging equation (4.3a) to obtain (using our notation)

[28] Note that the value of κ_K can be obtained from the homogeneity adding up conditions if it is needed.

[29] Note that this paper uses similar methods to those used in Rodriguez-Alvarez and Lovell (2004) and Rodriguez-Alvarez, Fernandez-Blanco and Lovell (2004)

[30] Their model uses the notation k_i instead of the $\exp(\kappa_i)$ notation we use above.

$$\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j} = \left[\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j} - \frac{\exp(\kappa_i) w_i x_i}{\sum_{j=1}^{K} \exp(\kappa_j) w_j x_j} \right] + \alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_j / x_K \right) + \sum_{j=1}^{M} \gamma_{ij} \ln y_j, \qquad (4.4)$$

$$i = 1,\ldots,K-1.$$

They then appear to replace the messy term in the square brackets with a constant parameter A_i and append an error term λ_i to capture "the effects of random noise" (Baños-Pino *et al*, 2002, p.197).[31] There is no discussion of the likely sources of this random noise. Thus the share equations become

$$\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j} = A_i + \alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_j / x_K \right) + \sum_{j=1}^{M} \gamma_{ij} \ln y_j + \lambda_i, \quad i = 1,\ldots,K-1. \qquad (4.5)$$

Given that the term in the square brackets in equation (4.4) is unlikely to be identical across all observations, we need to rewrite equation (4.5) as

$$\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j} = A_i + \alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_j / x_K \right) + \sum_{j=1}^{M} \gamma_{ij} \ln y_j + \xi_i, \quad i = 1,\ldots,K-1. \qquad (4.6)$$

where

$$\xi_i = \frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j} - \frac{\exp(\kappa_i) w_i x_i}{\sum_{j=1}^{K} \exp(\kappa_j) w_j x_j} - A_i + \lambda_i, \quad i = 1,\ldots,K-1. \qquad (4.6a)$$

This error term involves both input quantities and input prices. Thus OLS methods are likely to be inconsistent and GMM/IV methods are likely to suffer from a lack of valid instruments.

They state that since the input variables are assumed to be endogenous, they will be correlated with the errors in their model (though it is not clear how this conclusion is

[31] In their distance function equation they indicate that they assume that the error term has two components in the spirit of Aigner *et al* (1977). The technical inefficiency error term is assumed to have an *i.i.d.* half-normal distribution while the random noise error term is assumed to have an *i.i.d.* normal distribution.

reached). They hence use instrumental variables to overcome this perceived problem. The instruments they choose are various Spanish macro economic variables: fixed capital, employees in agriculture and automobile gasoline consumption. There is no discussion as to why these variables are likely to be valid instruments. That is, why they should be correlated with the input quantities but uncorrelated with the disturbance terms.

They indicate that they estimate the model using iterated seemingly unrelated regressions (ITSUR), with a correction for autocorrelation. It is not clear how the composed error structure in the distance function, nor the instrumental variables, are incorporated into this particular estimation technique. In fact, it is not clear to us how this technique could be implemented. Perhaps, some form of maximum likelihood estimation (MLE) is used – however the likelihood function (which would be very complicated) is not presented.

After econometric estimation, they then calculate (for each observation in the sample) measures of the degree of allocative inefficiency by calculating ratios of observed price ratios over predicted shadow price ratios (see their equation 22). This calculation implicitly assumes that the error terms in their share equations are attributed to allocative mistakes.[32] This appears to be in conflict with the description of the error terms in the share equations as being due to "the effects of random noise" (Baños-Pino *et al,* 2002, p.197).

In a recent paper, Karagiannis et al (2006) make note of some of the above estimation issues and propose some alternative methods. Their models involve the use of the Balk (1997) normalisation

$$\sum_{j=1}^{K} x_j w_j = \sum_{j=1}^{K} x_j w_j^s \left(= \sum_{j=1}^{K} x_j w_j \exp(\kappa_j) \right), \qquad (4.7)$$

which normalises shadow prices so that shadow cost must equal actual cost.

By inserting this into equation (4.3a), taking logs and rearranging one obtains

[32] This result can be easily illustrated by noting that the observed values of the $w_i x_i$ in their equation (22) will be equal to the predictions from their equation (19') <u>plus</u> the estimated residuals.

$$\ln\left(\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j}\right) = \ln\left[\alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{ij}\ln y_j\right] - \kappa_i, \quad i = 1,...,K-1 \tag{4.8}$$

Karagiannis et al (2006) estimate this model using MLE.[33] In the distance function equation they assume *u* has a half normal distribution, and append a "stochastic noise term", *v*. They provide no discussion of the sources of this noise. They assume the κ_i in equation (4.7) have a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$. It is assumed that *v* and *u* are distributed independently of each other and of the $\boldsymbol{\kappa}$. The $\ln x_i$ are assumed to be the endogenous variables, and hence the likelihood function involves a Jacobian term to reflect this.

The above model does require one to assume that there are no non-management errors in the first-order equations, but otherwise seems promising. However, one fault we have noted is that these authors have chosen to arbitrarily drop one of the log-share equations when constructing the ML estimator. In our assessment this is not required because the log-shares need not add to one (nor to any other constant) and hence the system need not be singular. Furthermore, the estimates obtained will not be invariant to the choice of which log-share equation is dropped. In Section 6 we propose an adjustment to this model which deals with these issues.

Karagiannis et al (2006) also propose another model that involves fixed effects. They note that if one has access to panel data, that one could choose to not model the κ_i as error components (as described above), and instead model them as parameters which are fixed over time for each firm (or alternatively are a firm-specific function of a time-trend variable). They note that the advantage of this is that they "can append statistical noise terms in all equations in the system". Thus in the share equations they no longer need to assume that all errors are due to allocative inefficiency, and are able

---

[33] They also derive additional models with an additive allocative mistake formulation of $w_i^s = w_i + k_i$. However this model seems less attractive since the MLE methods are complicated by the necessity to ensure that the predicted shadow prices remain non-negative.

to "account for the possibility of omitted variables and measurement error in the dependent variables".

Unfortunately, in our assessment, this fixed effects model will face additional problems because the Balk normalisation is incompatible with this model. Consider the case where we have panel data on N firms in T time periods. If one assumes that the n-th firm has fixed values of κ_i over time, it is easy to show that, in general, equation (4.7) cannot be satisfied if $T > 1$ (with the exception of some trivial cases). This is because the number of equations exceeds the number of unknowns.[34] Hence econometric estimation will suffer from problems similar to those seen in equation (4.6). That is, since the Balk normalisation cannot be assumed, equation (4.8) will essentially become

$$\ln\left(\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j}\right) = \ln\left[\alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{ij}\ln y_j\right] - \kappa_i + \xi_i, \quad i = 1,\ldots,K-1. \quad (4.9)$$

where

$$\xi_i = \ln\left(\frac{w_i x_i}{\sum_{j=1}^{K} w_j x_j}\right) - \ln\left(\frac{\exp(\kappa_i) w_i x_i}{\sum_{j=1}^{K}\exp(\kappa_j) w_j x_j}\right) + \kappa_i + \lambda_i, \quad i = 1,\ldots,K-1. \quad (4.9a)$$

This suggests that the fixed effects MLE methods used in Karagiannis et al (2006) will produce inconsistent estimates. It also suggests that any attempt at using OLS or GMM/IV will also be problematic for the reasons discussed above.

---

[34] See Appendix 1 for further explanation.

5. Calculation of efficiency scores

Shadow input quantities model

If the (substantial) estimation problems described in Section 3 can be overcome, the calculation of firm-specific efficiency scores is an easy process when a shadow input quantities model has been estimated (as we have noted in the introduction). First one uses the estimated value of θ_i to calculate

$$x_i^{ce} / x_K^{ce} = \exp(\theta_i) x_i \ / x_K \qquad , i = 1, ..., K-1, \qquad (5.1)$$

and then uses these ratios plus the fact that $D(\mathbf{y}, \mathbf{x}^{ce}) = 1$ to calculate

$$\begin{aligned} -\ln x_K^{ce} = \alpha_0 + \sum_{i=1}^{K-1} \alpha_i \ln\left(x_i^{ce} / x_K^{ce}\right) + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_i^{ce} / x_K^{ce}\right) \ln\left(x_j^{ce} / x_K^{ce}\right) \\ + \sum_{i=1}^{M} \beta_i \ln y_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \beta_{ij} \ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{M} \gamma_{ij} \ln\left(x_i^{ce} /_K^{ce}\right) \ln y_j \end{aligned} . \qquad (5.2)$$

One can then use equation (5.1) to solve for the other *K*-1 x_i^{ce} .

Then one calculates cost efficiency (*CE*) as

$$CE = \mathbf{w}'\mathbf{x}^{ce} / \mathbf{w}'\mathbf{x} ,$$

technical efficiency (*TE*) as

$$TE = \mathbf{w}'\mathbf{x}^{te} / \mathbf{w}'\mathbf{x} ,$$

where $\mathbf{x}^{te} = \mathbf{x}\exp(-u)$, and allocative efficiency (*AE*) as

$$AE = \mathbf{w}'\mathbf{x}^{te} / \mathbf{w}'\mathbf{x}^{ce} ,$$

where *CE*=*AE*×*AE*.

This process is very simple. There is no need for one to solve a set of non-linear equations for each observation, as is needed when a shadow input prices model is used.

Shadow input prices model

Given that the input distance function can be correctly estimated, the calculation of observation-specific efficiency scores is a complicated process when a shadow input prices model has been estimated. The difficulty is associated with the identification of the cost minimising input ratios. To identify these one must solve a set of non-linear equations for each observation in the sample. The process is as follows. First, one observes that the cost efficient data point must satisfy the first order condition set out in equation (3.4a). That is

$$\frac{w_i x_i / x_K}{\sum_{j=1}^{K} w_j x_j / x_K} = \alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_j / x_K\right) + \sum_{j=1}^{M} \gamma_{ij} \ln y_j, \quad i = 1,\ldots,K . \tag{5.3}$$

If we divide each of the first K-1 equations by the K-th equation we obtain

$$\frac{w_i x_i / x_K}{w_K} = \frac{\alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_j / x_K\right) + \sum_{j=1}^{M} \gamma_{ij} \ln y_j}{\alpha_K + \sum_{j=1}^{K-1} \alpha_{Kj} \ln\left(x_j / x_K\right) + \sum_{j=1}^{M} \gamma_{Kj} \ln y_j}, \quad i = 1,\ldots,K-1 \tag{5.4}$$

or equivalently

$$\begin{aligned} w_i \,(x_i / x_K) &\left[\alpha_K + \sum_{j=1}^{K-1} \alpha_{Kj} \ln\left(x_j / x_K\right) + \sum_{j=1}^{M} \gamma_{Kj} \ln y_j \right] = \\ & w_K \left[\alpha_i + \sum_{j=1}^{K-1} \alpha_{ij} \ln\left(x_j / x_K\right) + \sum_{j=1}^{M} \gamma_{ij} \ln y_j \right], \quad i = 1,\ldots,K-1 \end{aligned} \tag{5.5}$$

This provides K-1 non-linear equations in K-1 unknowns (i.e., the optimal x_i^{ce} / x_K^{ce} ratios).[35] This system of equations can be solved using Newton-type methods – once

[35] This approach is related to that used by Kopp and Diewert (1982) and Zieschang (1983) in a cost function context.

for each observation in the sample.[36] Once these optimal input ratios are obtained, one then follows the steps outlined earlier for the shadow input quantities model.[37]

6. A feasible model

In earlier sections of this paper we carefully describe a range of possible DGPs and review a number of past methods. Unfortunately, the conclusions are not encouraging. We began with a DGP where allocative mistakes were modelled using shadow input quantities. The model derived from this DGP appears to be such that econometric estimation is not feasible because of the non-linear manner in which the error terms enter the share equations. Atkinson and Primont (2002) estimate a model involving shadow input quantities using non-linear GMM methods in which efficiency is modelled parametrically, however their approach is open to criticism since one must assume that one input quantity is always used efficiently and one must also assume that all firms share the same allocative efficiency parameters in any one particular year. If either of these (rather implausible) assumptions do not hold, the estimators will be inconsistent.

As a consequence we also considered an alternative DGP involving shadow input prices (instead of quantities). We reviewed past studies that involve models containing shadow price constructs – by Banos-Pino et al (2002) and Karagiannis et al (2006) – and found that these methods also face a number of econometric estimation problems as well.

However, below we propose a model which is closely related to the error components model proposed by Karagiannis et al (2006). The key difference is that we specify first order equations which are in ratio form, which allows us to avoid the invariance violation problem in that model. The parameters of this model can be consistently estimated using MLE.

---

[36] A similar procedure has been used in some past in studies. For example, see Karagiannis et al (2004) and Alvarez et al. (2004).

[37] It is interesting to note that this procedure does not involve the explicit use of the κ_i parameter estimates. Hence, the fact that κ_K is not estimated is not a concern.

We begin with equations (4.3-4.3a) and divide the *i*-th equation (4.3a) by the *K*-th equation to obtain the system of equations

$$\begin{aligned}-\ln x_K &= \alpha_0 + \sum_{i=1}^{K-1}\alpha_i \ln\left(x_i / x_K\right) + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_i / x_K\right)\ln\left(x_j / x_K\right) \\ &\quad + \sum_{i=1}^{M}\beta_i \ln y_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\beta_{ij}\ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{K-1}\sum_{j=1}^{M}\gamma_{ij}\ln\left(x_i / x_K\right)\ln y_j + u\end{aligned} \quad ,(6.1)$$

$$\frac{x_i w_i \exp(\kappa_i)}{x_K w_K \exp(\kappa_K)} = \frac{\alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{ij}\ln y_j}{\alpha_K + \sum_{j=1}^{K-1}\alpha_{Kj}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{Kj}\ln y_j}, \quad i = 1,\ldots,K-1. \qquad (6.1a)$$

We then take logs of equation (6.1a) and rearrange to obtain

$$\ln\left(\frac{x_i w_i}{x_K w_K}\right) = \ln\left[\frac{\alpha_i + \sum_{j=1}^{K-1}\alpha_{ij}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{ij}\ln y_j}{\alpha_K + \sum_{j=1}^{K-1}\alpha_{Kj}\ln\left(x_j / x_K\right) + \sum_{j=1}^{M}\gamma_{Kj}\ln y_j}\right] + \phi_i, \quad i = 1,\ldots,K-1 \qquad (6.2)$$

where $\phi_i = \kappa_K / \kappa_i$.

The parameters of the model in equations (6.1-6.2) can be estimated using MLE and the estimates obtained will be consistent and invariant to the choice of normalising input used. This is thus an improvement over the other methods we have considered. However, the model is far from perfect. It has the disadvantage that one must attribute all errors in the share equations to allocative mistakes (i.e., one must assume no other sources of noise such as unanticipated events, specification error and measurement error). Furthermore, calculation of allocative efficiency scores requires the solution of a set of non-linear equations for each observation in the sample.[38]

This model is applied to US electricity industry data in the next section.

---

[38] Schmidt and Lovell (1979) have also used this type of ratio approach in a Cobb-Douglas production context.

7. Empirical application

The empirical analysis in this study involves panel data on the fossil fuel steam electric power generation activities of 61 US electric utilities during 1986–1998. The primary sources of data are obtained from the Energy Information Administration (EIA), the Federal Energy Regulatory Commission (FERC), the Bureau of Labor Statistics (BLS), and the Federal Reserve Board (FRB). The data set used to obtain the econometric estimates contains information on one output quantity variable: electricity, and quantity and price information on three input variables: fuel, labor and maintenance, and capital. These variables are now briefly described.[39]

The output variable, y, is represented by net steam electric power generation in megawatt-hours, which is defined as the amount of power produced using fossil-fuel fired boilers to produce steam for turbine generators during a given period of time.

The price of fuel aggregate, w_1, is a multilateral Törnqvist price index of the three fuels used (coal, oil and gas), derived from firm-level price and quantity data. The quantity of fuel, x_1, is calculated as the steam power production fuel costs divided by the multilateral Törnqvist price index for fuels.

The price of labor and maintenance aggregate, w_2, is a multilateral Törnqvist price index for labor and maintenance.[40] The price of labor is a firm-level average wage rate. The price of maintenance and other supplies is an industry-level price index of electrical supplies. The quantity of labor and maintenance, x_2, is measured as the aggregate costs of labor and maintenance divided by the multilateral Törnqvist price index for labor and maintenance.

The price of capital, w_3, is the yield of the firm's latest issue of long-term debt adjusted for appreciation and depreciation of the capital good using the Christensen and Jorgenson (1970) cost of capital formula

---

[39] For more detail on the data set, see Rungsuriyawiboon and Coelli (2006). We are grateful to Supawat Rungsuriyawiboon for allowing us to use his data in this study.

[40] These costs were not separated into labor and non labor costs because the widespread use of outsourcing has made such distinctions rather arbitrary.

$$w_{3it} = p_{kt}\left[i_{dit} + s_{it}\left(r_{eit} - i_{dit}\right) + d - f_t\right] \qquad (7.1)$$

where p_{kt} is a price index for electrical generating plant and equipment; i_{dit} is the adjusted corporate bond rate by firm based upon its bond ratings by Moody's Investor Service; s_{it} is the equity share of total capital defined as total proprietary capital (TPC) divided by the sum of total proprietary capital and total long-term debt (TOTB); r_{eit} is the equity rate of return defined as the ratio of net income to total proprietary capital; d is a depreciation rate assuming 30 years straight line depreciation; and f_t the inflation rate.

The values of capital stocks are calculated by the valuation of base and peak load capacity at replacement cost to estimate capital stocks in a base year and then updating it in the subsequent years based upon the value of additions and retirements to steam power plant as discussed in Considine (2000)

$$x_{3it} = \frac{(1-\upsilon)\left(x_{3it-1}\right) p_{kit}}{p_{kit-1}} + A_{it} - R_{it}\,, \qquad (7.2)$$

where υ denotes the depreciation rate; x_{3it} is equal to the nominal stock divided by the price index for electrical generating plant and equipment, p_{kit}; A_{it} and R_{it} denote additions and retirements to steam power plant.

Table 1 represents a summary of the data used in this study. The average expenses of aggregate fuels, aggregate labor and maintenance, and capital are calculated to be 258.79, 66.66, and 97.43 million dollars, respectively. The mean cost shares of fuel, labor and maintenance, and capital are approximately 59, 18, and 23 per cent, respectively.

The model that is estimated involves three equations

$$-\ln x_3 = \alpha_0 + \sum_{i=1}^{4}\alpha_i z_i + \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4}\alpha_{ij} z_i z_j + u + v\,, \qquad (7.3)$$

$$\ln\left(\frac{w_i x_i}{w_3 x_3}\right) = \ln\left(\frac{\alpha_i + \sum_{j=1}^{4}\alpha_{ij} z_j}{\alpha_3 + \sum_{j=1}^{4}\alpha_{3j} z_j}\right) + \phi_i, \qquad i = 1,2. \qquad (7.3a)$$

Note that the notation $\mathbf{z} = [z_1, z_2, z_3, z_4] = [\ln y, \ln(x_1 / x_3), \ln(x_2 / x_3), t]$ represents a netput vector (where t is a time trend variable), $v \sim N(0, \sigma_v^2)$, $u \sim |N(0, \sigma_u^2)|$ and $\boldsymbol{\varphi} = (\phi_1, \phi_2) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = (\mu_1, \mu_2) \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix},$$

and v, u and ϕ_i are independent. The σ_v^2 and σ_u^2 parameters are replaced with $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$ and $\delta = \sqrt{\sigma_u / \sigma_v}$. The square root transforms are used to ensure that non-negative variances are not selected during the iterative estimation phase.[41]

This model is estimated using MLE, where the likelihood function has be concentrated with respect to the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ parameters and involves a Jacobian term to reflect the fact that the endogenous variables are the $\ln x_i$.[42]

Table 1: Data summary for 61 US electric utilities, 1986–98

Variable	Units*	Mean	St.Dev.	Min.	Max.
Quantities:					
Output, y	($\times 10^6$ MWh)	13.709	12.561	0.499	79.723
Fuel, x_1	($\times 10^6$ dollars)	300.568	351.842	12.823	2,522.324
Labor and Maint., x_2	($\times 10^6$ dollars)	61.776	53.366	1.810	444.453
Capital, x_3	($\times 10^6$ dollars)	955.225	877.403	9.070	3,878.295
Prices:					
Fuel, w_1	(index)	0.861	0.208	0.306	1.338
Labor and Maint., w_2	(index)	1.079	0.255	0.443	1.928
Capital, w_3	(index)	0.102	0.019	0.009	0.203

* These are 1993 dollar values.

[41] The Davidon-Fletcher-Powell Quasi-Newton routine is used to maximise the likelihood function.
[42] See the appendix for details of the structure of the likelihood function.

Two sets of ML estimates along with asymptotic standard errors are listed in Table 2. The first set of estimates relate to the full model while the second set relate to a restricted model where we impose the restriction that $\boldsymbol{\mu} = \mathbf{0}$. A likelihood ratio (LR) test provides a calculated value of 8.28 which is greater than the 5% chi-square critical value of 5.99, suggesting that the ϕ_i have means which are significantly different to zero.

The estimates of the input elasticities[43] in Table 2 are 0.610, 0.162 and 0.228 for fuel, labor and maintenance, and capital, respectively. These are similar to the average observed shares in this data set of 0.59, 0.18 and 0.23, respectively. The estimated output elasticity of minus 1.008 indicates that the average firm is operating in a region of constant returns to scale.[44] This result is not surprising given the results reported in past studies (for example see Christensen and Greene, 1976). Finally, the first order coefficients of the time trend variable provides an estimate of the average annual rate of technical change of 1.2 % per year. Again, this figure is within expectations, as most studies of technical change in utilities tend to report technical change estimates of between 1 and 2% per annum.

Technical efficiency scores are calculated using the conditional expectation measures described in Battese and Coelli (1988). Allocative efficiency scores are calculated using the methods described in Section 5. This involves the solution of a set of non-linear equations for each of the 793 observations in the sample. This is achieved using the Davidon-Fletcher-Powell Quasi-Newton optimisation routine.

---

[43] The data variables used in the model estimation were each transformed by division by their respective geometric means, as is common practice. This transformation does not alter the performance measures obtained, but does allow one to interpret the estimated first-order parameters as elasticities, evaluated at the sample means.

[44] The scale economies measure is equal to $-\left(\partial \ln D / \partial \ln y\right)^{-1}$ in this model.

Table 2: MLE parameter estimates of input distance functions

	Coefficient	S-Error	t-ratio	Coefficient	S-Error	t-ratio
Intercept	0.3088	0.0105	29.50	0.2987	0.0116	25.79
Elec	-1.0126	0.0109	-92.47	-1.0049	0.0108	-93.24
Fuel	0.6843	0.0083	82.26	0.5990	0.0060	100.62
L&M	0.1227	0.0044	28.14	0.1723	0.0040	42.99
T	0.0085	0.0017	4.98	0.0111	0.0019	5.71
Elec*Elec/2	-0.0636	0.0141	-4.52	-0.0531	0.0138	-3.85
Elec*Fuel	0.0368	0.0038	9.79	0.0341	0.0058	5.86
Elec*L&M	-0.0355	0.0033	-10.67	-0.0392	0.0046	-8.55
Elec*t	0.0022	0.0017	1.30	0.0024	0.0018	1.34
Fuel*Fuel/2	-0.1431	0.0076	-18.82	-0.1805	0.0125	-14.45
Fuel*L&M	0.0704	0.0013	52.92	0.0994	0.0068	14.60
Fuel*t	0.0013	0.0009	1.32	0.0033	0.0014	2.38
L&M*L&M/2	-0.0841	0.0035	-23.80	-0.1126	0.0081	-13.91
L&M*t	-0.0027	0.0006	-4.94	-0.0038	0.0009	-4.00
t*t/2	0.0020	0.0010	1.96	0.0021	0.0011	1.92
Sigma	0.3252	0.0110	29.58	0.3212	0.0097	32.96
Delta	1.7071	0.0814	20.97	1.6694	0.0741	22.53
LLF:	1482.57			1478.43		
Mu1	-0.1577			-		
Mu2	0.3203			-		
Sigma11	0.0605			0.0770		
Sigma22	0.3689			0.3615		
Sigma12	-0.1026			-0.1217		

Summary statistics for these efficiency scores are reported in Table 3. The average cost efficiency score is 0.763, indicating that the average firm could reduce costs by 23.7% and still produce the same output. Technical inefficiency is the main contributor, with a mean score of 0.803 versus a mean allocative efficiency score of 0.950. The small contribution of allocative inefficiency is not surprising given the observation above that shadow shares and market shares are similar (at the sample mean).[45]

Table 3: Summary of efficiency scores

	TE	AE	CE
Mean	0.803	0.950	0.763
Median	0.836	0.971	0.792
Standard Deviation	0.121	0.067	0.126
Minimum	0.353	0.453	0.308
Maximum	0.987	1.000	0.950

In calculating the above efficiency scores, vectors of cost-minimising input quantities and technically efficient input quantities were obtained for each firm in the sample. The latter were divided by the former to produce ratio measures which provide information on the degree to which the different firms selected sub-optimal input mixes. Table 4 contains summary statistics on these ratios. We observe that the median ratio for Fuel reflects a degree of under use, while those for L&M and Capital reflect some overuse.

[45] A LR test of the null hypothesis that $\delta = 0$ was also conducted. This produced a statistic of 97.58, which is substantially large than the 5% critical value of 1.96 (see Kodde and Palm, 1986). Thus the technical inefficiency error term, u, is a significant addition to the model.

Table 4: Summary of x_i^{te} / x_i^{ce} ratios

	Fuel	L&M	Capital
Mean	0.951	1.304	1.209
Median	0.930	1.228	1.166
Standard Deviation	0.130	0.552	0.449
Minimum	0.549	0.391	0.154
Maximum	2.420	4.980	5.638

8. Concluding comments

In this study our aim was to identify the best way to estimate a system of equations involving an input distance function along with the first order equations that relate to shadow cost minimising behaviour. We began with a detailed analysis of the DGP, discussing various types of both management and non-management errors. We then conducted a review of past studies which led us to the conclusion that there is no model available that can capture both types of errors in a reliable manner. In fact, even if one is willing to assume that non-management errors do not exist, we were still unable to identify a model that was in our view appropriate.

The least problematic model that we could identify was the error components model proposed by Karagiannis et al (2006). The principal problem with this model is that it was not invariant to the choice of normalising input. We hence propose an adjusted version of this model which involves re-expressing the first-order equations in ratio form so as to avoid the invariance problem.

An empirical application of this model involving panel data on US electricity generation firms is presented, where we find that technical inefficiency is the largest contributor to cost inefficiency, and that the majority of allocative mistakes involve under use of fuel relative to the other inputs.

References

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

Amemiya, Y. 1985. "Instrumental variable estimator for the nonlinear errors-in-variables model," *Journal of Econometrics* 28, 273-289.

Atkinson, S. E., and C. Cornwell (1994), “Parametric Estimation of Technical and Allocative Inefficiency with Panel Data,” *International Economic Review* 35, 231-44.

Atkinson, S. E., and D. Primont (2002), “Stochastic Estimation of Firm Technology, Inefficiency and Productivity Growth Using Shadow Cost and Distance Functions,” *Journal of Econometrics* 108, 203-25.

Atkinson, S.E., O. Honerkamp and C. Cornwell (2003), “Measuring and Decomposing Productivity Change: Stochastic Distance Function Estimation vs. DEA”, *Journal of Business and Economic Statistics* 21, 284-94.

Atkinson, S.E., R. Färe and D. Primont (2003), “Stochastic Measurement 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, 245-73.

Balk, B.M. (1997), “The decomposition of cost efficiency and the canonical form of cost function and cost share equations”, *Economics Letters* 55, 45–51.

Baños-Pino, J., Fernandez-Blanco, V. and A. Rodriguez-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.

Battese, G.E. and T.J. Coelli (1988), ‘Prediction of firm-level technical efficiencies with a genneralized frontier production function and panel data’, *Journal of Econometrics* 38, 387–99.

Christensen, L.R. and D.W. Jorgenson (1970), ‘US real product and real factor input, 1928–1967’, *Review of Income and Wealth* 16, 19–50.

Christensen, L.R. and W.H. Greene (1976), 'Economies of scale in US electric power generation', *Journal of Political Economy* 84, 655–76.

Coelli, T.J. (2000), "On the Econometric Estimation of the Distance Function Representation of a Production Technology", CORE Discussion Paper 2000/42, Universite Catholique de Louvain.

Coelli, T.J. and S. Perelman (1999), "A Comparison of Parametric and Non-parametric Distance Functions: With Application to European Railways", *European Journal of Operations Research* 117, 326-339.

Considine, T.J. (2000), 'Cost structures for fossil fuel-fired electric power generation', *The Energy Journal* 21, 83–104.

Färe, R., and D. Primont (1995) *Multi-Output Production and Duality: Theory and Applications*, Boston: Kluwer Academic Publishers.

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, 374-380.

Farrell, M.J. (1957), "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society* Series A, CXX, Part 3, 253-290.

Ferrier, G. D., and Lovell, C. A. K. (1990), "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence," *Journal of Econometrics* 46, 229-45.

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, 79-94.

Greene, W. H. (2007), "The Econometric Approach to Efficiency Analysis," Chapter 2 in Fried, H.O., C.A.K. Lovell and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press, New York.

Grosskopf, S. and K. Hayes (1993), "Local Public Sector Bureaucrats and Their Input Choices", *Journal of Urban Economics* 33, 151-66.

Hansen, L.P. (1982), "Large sample properties of generalized method of moments estimation", *Econometrica* 50, 1029–1054.

Karagiannis, G., E.G. Tsionas and S.C. Kumbhakar, (2006), "Estimation of Input Distance Functions in the Presence of Technical and Allocative Inefficiency: A System Approach", mimeo, State University of New York at Binghamton.

Karagiannis, G., Kumbhakar, S.C. and E. Tsionas, (2004), "A Distance Function Approach for Estimating Technical and Allocative Inefficiency", *Indian Economic Review* 39, 19-30.

Kodde, D.A., and F.C. Palm (1986), "Wald Criteria for Jointly Testing Equality and Inequality Restrictions" *Econometrica*, 54, 1243-1248.

Kopp, R. J., and W. E. Diewert (1982), "The Decomposition of Frontier Cost Function Deviations into Measures of Technical and Allocative Efficiency," *Journal of Econometrics* 19, 319-31.

Kumbhakar, S.C., and G. Karagiannis (2004), "On the equivalence of two normalizations in estimating shadow cost functions", *Economics Letters* 82, 15–19.

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, 355–384.

Lovell, C.A.K., S. Richardson, P. Travers and L.L. Wood (1994), "Resources and Functionings: A New View of Inequality in Australia", W. Eichhorn (ed.), *Models and Measurement of Welfare and Inequality*, Berlin: Springer-Verlag.

Maietta, O.W. (2002), "The effect of the normalization of the shadow price vector on the cost function estimation", *Economics Letters* 77, 381–385.

Newey, W.K. and K.D. West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703–706.

Rodriguez-Alvarez, A. and C.A.K. Lovell (2004), "Excess Capacity and Expense Preference Behavior in National Health Systems: An Application to the Spanish Public Hospitals", *Health Economics* 13, 157-69.

Rodriguez-Alvarez, A., V. Fernandez-Blanco and C.A.K. Lovell (2004), "Allocative Efficiency and its Cost: The Case of Spanish Public Hospitals", *International Journal of Production Economics* 92, 99-111.

Rungsuriyawiboon, S. and T.J. Coelli (2006), "Regulatory reform and economic performance in US electricity generation" in Coelli, T.J. and D. Lawrence (eds.), *Performance Measurement and Regulation of Network Utilities,* Edward Elgar Publishing, Cheltenham, 267-296.

Schmidt, P. and C.A.K. Lovell, (1979), “Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers”, *Journal of Econometrics* 9, 343–366.

Simar, L. and P.W. Wilson (2007), “Estimation and inference in two-stage, semi-parametric models of production processes”, *Journal of Econometrics* 136, 31–64.

Simar, L. and P.W. Wilson (2000), “Statistical inference in nonparametric frontier models: the state of the art”, *Journal of Productivity Analysis* 13, 49–78.

Zellner, A., J. Kmenta and J. Drèze (1966), “Specification and Estimation of Cobb-Douglas Production Function Models”, *Econometrica* 34, 784-795.

Zieschang, K. D. (1983), “A Note on the Decomposition of Cost Efficiency into Technical and Allocative Components,” *Journal of Econometrics* 23, 401-05.

Appendix 1: The Balk normalisation and fixed effects panel data models

In this appendix we show that the Balk (1997) normalisation is inconsistent with the assumption that all firms possess the same vector of "allocative efficiency parameters".

Using the notation in Karagiannis et al (2006) the Balk normalisation is

$$\sum_{j=1}^{K} x_j w_j = \sum_{j=1}^{K} x_j w_j k_i . \qquad \text{(A1.1)}$$

For each data point, the K-1 ratios k_i / k_K are determined by the gradient of the distance function at that point. Hence we can write $k_i / k_K = d_i$ or equivalently

$$k_i = d_i k_K \qquad , i = 1,..., K-1, \qquad \text{(A1.2)}$$

where the $d_i = D_i / D_K$ represent the gradient information. Thus we have K equations in K unknowns and a solution is possible. In fact, if we substitute equation (A1.2) into equation (A1.1) and rearrange we obtain[46]

$$k_K = \sum_{j=1}^{K} x_j w_j \Bigg/ \sum_{j=1}^{K} x_j w_j d_i , \qquad \text{(A1.3)}$$

and then equation (A1.2) can be used to obtain the remaining k_i.

This normalisation is acceptable if each data point is permitted to have a unique set of k_i (as is the case in the error components model). However, if one specifies a model (such as a fixed effects model) where the same set of k_i must apply over $T > 1$ observations we will have a set of KT equations in K unknowns, which has no solution.

Note also that allowing the k_i to be a polynomial function of time will not solve this problem, except in the case where the polynomial function is of order T-1. However, in this case the number of parameters in the econometric model will exceed the number of observations, and hence estimation is not feasible.

---

[46] Note that $d_K = 1$ by definition.

Appendix 2: Derivation of Likelihood function

The system of equations is[47]

$$\ln D(\mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\alpha}) = \varepsilon_n, \quad n = 1, 2, ..., N , \qquad \text{(A2.1)}$$

$$\ln\left(\frac{w_{in}x_{in}}{w_{Kn}x_{Kn}}\right) = \ln\left(\frac{D_{in}'}{D_{Kn}'}\right) + \phi_{in}, \qquad i = 1, K-1,\ n = 1, 2, ..., N, \qquad \text{(A2.1a)}$$

where $D'_{in} = \partial \ln D_n / \partial \ln x_{in}$, $\varepsilon_n = u_n + v_n$ $v_n \sim N(0, \sigma_v^2)$, $u_n \sim |N(0, \sigma_u^2)|$, $\boldsymbol{\varphi}_n = (\phi_{1n}, ..., \phi_{K-1,n}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and v_n, u_n and ϕ_{in} are independent.

Since the vector of logged input quantities are the endogenous variables, the likelihood function involves a Jacobian term

$$J_n(\alpha) = \left\| \frac{\partial(\varepsilon_n, \boldsymbol{\varphi}_n)}{\partial \ln \mathbf{x}_n} \right\| = \left\| \frac{\partial \boldsymbol{\varphi}_n}{\partial \ln \mathbf{x}_n} \right\|$$

where it is easy to show that

$$\frac{\partial \phi_{ni}}{\partial \ln x_{nj}} = \Delta_{ij} - D_{ijn}'' / D_{in}' + D_{iKn}'' / D_{Kn}'$$

where $D_{ijn}'' = \partial D_{in}' / \partial \ln x_{jn}$ and Δ_{ij} is the Kronecker delta.

Given the above distributional assumptions, it is easy to show that the concentrated log likelihood function (with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ concentrated out) is[48]

$$\ln L = \frac{N}{2}\ln(2\pi/\sigma^2) + \frac{1}{2\sigma^2}\sum_{n=1}^{N}\hat{\varepsilon}_n^2 - \sum_{n=1}^{N}\ln\Phi(\delta^2\hat{\varepsilon}_n/\sigma) - \frac{N}{2}\ln\left|\hat{\boldsymbol{\Sigma}}\right| + \sum_{n=1}^{N}\ln J_n(\alpha)$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N}(\hat{\phi}_n - \hat{\mu})(\hat{\phi}_n - \hat{\mu})'$, $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N}\hat{\phi}_n$, $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$ and $\delta = \sqrt{\sigma_u / \sigma_v}$.

---

[47] The following derivation follows a similar structure to that in Karagiannis et al (2006).

[48] Note that in the empirical application in this paper we treat the panel data as if it is a "single cross-section" and hence *N*=793.