25 Efficiency measurement in the electricity and gas distribution sectors

Mehdi Farsi and Massimo Filippini

1 Introduction

In the last two decades the electricity and gas distribution sectors have witnessed a wave of regulatory reforms aimed at improving the economic efficiency. In the design of these reforms the information on several efficiency concepts, including scale efficiency, scope efficiency, and cost efficiency has become very important. The first two concepts are directly related to the economies of scale and scope, which are characteristics of the production technology, whereas the concept of cost efficiency is mainly a firm's characteristic related to its economic performance facing market and technological conditions.

Scale efficiency addresses the question of whether, for instance, an electricity distribution company is operating at the minimum of its long-run average cost curve. Any deviation from this level of production could result in inefficiency in terms of scale of operation. Thus, scale efficiency arises when the company cannot lower average costs by changing its output levels. For multiproduct energy companies, that is, a company distributing electricity and gas, scope efficiency focuses on the relative cost of joint production to the cost of producing the same total output in multiple companies. Scope inefficiency exists if the costs can be lowered by changing the output mixes across companies. The concepts of scale and scope efficiency rely on the assumption that the market structure, particularly outputs, can be adjusted to provide the companies with the greatest possibility of exploiting synergies. Cost efficiency measures the ability of energy distribution companies to minimize costs, given specific demand and market conditions. Cost inefficiency, also called 'X-inefficiency', occurs when the company fails to produce with full efficiency at the cost frontier. The performance of a company in minimizing costs can be decomposed into two types of efficiency. The first one is technical efficiency; the extent to which the energy distribution companies could reduce inputs for a given level of outputs (input orientation) or expand outputs for given levels of inputs (output orientation). The distance to an optimal production frontier measures technical efficiency. The second component of cost efficiency is allocative efficiency; that is, the possible reduction in costs by using the different inputs in optimal proportions or equivalently, by operating on the least-cost expansion path.

Due to the economic importance of the regulatory reforms in the electricity and gas distribution sectors, it seems essential that the design of these reforms be based on a clear empirical understanding of the cost structure and efficiency level of the electricity and gas distribution companies. For instance, this understanding is relevant to several regulatory as well as business decision issues. First, it provides information about the validity of the natural monopoly argument in the distribution of gas and electricity. Second, it contributes to an evaluation of the definition of the optimal size of service areas. Third, it gives information on the importance of the potential synergies through 'horizontal' integration, which allows local multi-utility companies to save on costs by exploiting the economies of scope. This information is very important in the assessment of a policy for unbundling multi-utilities.

Finally, performance indicators such as cost efficiency are used to monitor the companies' economic performance by comparing companies with the most-efficient practices. The regulators increasingly use such benchmarking practices in various incentive regulation schemes. For instance, one of the most widely used methods in electricity and gas networks is price-cap regulation (RPI-X). This method sets the maximum rate of increase for the regulated prices equal to the inflation rate of the retail price index (RPI) minus a productivity growth offset referred to as 'X-factors'.[1] In relatively new regulatory regimes (mostly adopted by European regulators), X-factors are set equal to the annual target change in cost efficiency for each individual company. Therefore, the regulator can set differentiated price caps based on the companies' efficiency performance estimated from an empirical analysis. However, the increasing use of efficiency analysis in the electricity industry has raised serious concerns among regulators and companies regarding the reliability of efficiency estimates.[2] In fact the empirical evidence suggests that the estimates are sensitive to the adopted efficiency measurement approach.[3] This implies that the choice of the approach can have important effects on the financial situation of the companies. There are, however, alternative strategies that can be used to improve efficiency measurement methodology regarding the sensitivity issues that we shall shortly discuss in this chapter.

The main goal of this chapter is to present and discuss the empirical measurement of the productive efficiency in the distribution of electricity and gas so as to draw recommendations for regulatory practice as well as business strategic decisions at the company level. An adequate analysis of productive efficiency should consider all the efficiency aspects, including those related to the scale and scope of the distribution networks. From a policy standpoint it is important to distinguish between the three concepts of efficiency. While allowing the companies to exploit the potential economies of scale and scope to the greatest possible extent, the regulators should introduce incentive measures to ensure cost efficiency. However, the measures of firms' cost efficiency should consider the different limitations that various companies face in exploiting the potential synergies, mainly due to their different levels of output, as well as the unobserved heterogeneity in external factors across companies.

The rest of the chapter proceeds as follows. Section 2 reviews some elements of the production theory. Section 3 provides a general overview of the concepts of scale, scope and cost efficiency. Section 4 illustrates the different econometric approaches that can be used to measure the level of efficiency of the companies. A selection of relevant empirical studies of scale, scope and cost efficiency in the electricity and gas distribution sectors are reviewed separately in Sections 5, 6 and 7. Section 0 provides a brief discussion of benchmarking practice in the regulation of electricity distribution networks along with a simple case study from Switzerland's utilities. Section 0 concludes the chapter with a final discussion and policy recommendations.

2 Review of Traditional Production Theory

The microeconomic theory of production is extensively documented in the literature (for example, Shephard, 1953; Chambers, 1988; Jehle and Reny, 1998; Varian, 1992) and will not be repeated here. Instead, this section focuses on some elements of the microeconomic theory of production that are relevant for understanding the measurement of productive and scale efficiency in the electricity and gas distribution industry.

Generally, the empirical studies model the production of firms in an industry which use g inputs $\mathbf{x} = (x_1, x_2, \ldots, x_g)$ to produce m outputs $\mathbf{y} = (y_1, y_2, \ldots, y_m)$. A reasonable way to represent the firm's technology of turning inputs into output in the long run is to specify a transformation function, $T(x_1, \ldots, x_g, y_1, \ldots, y_m) = 0$, in the multiple-output case or as a production function, $y = f(x_1, x_2, \ldots, x_g)$ in the single-output case. These functions represent the border of a set that includes all the production possibilities.

If the firm faces competitive input markets and chooses input bundles to minimize costs in the long run, then the cost-minimizing process can be represented as:

$$\min_x C = \sum_{j=1}^{g} w_j x_j, \text{s.t. } f(\mathbf{x}) \geq \mathbf{y}, \tag{25.1}$$

where C represents long-run total cost, wj is the price of input xj, and f is the production function relating the vector of inputs $\mathbf{x}$ to the output vector $\mathbf{y}$. The solution to (25.1) is of the form $C(\mathbf{y}, \mathbf{w})$, where $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_g)$.

Provided that the transformation function $T(x_1, \ldots, x_g, y_1, \ldots, y_m) = 0$ borders a strictly convex set, McFadden (1978) has shown that the cost function, $C(\mathbf{y}, \mathbf{w})$ has the following properties (regularity conditions):

1. $C(\mathbf{y}, \mathbf{w}) > 0$ for $\mathbf{w} > 0$ and $\mathbf{y} > 0$ (non-negativity);
2. if $\mathbf{w}' > \mathbf{w}$, then $C(\mathbf{y}, \mathbf{w}') \geq C(\mathbf{y}, \mathbf{w})$ (non-decreasing in $\mathbf{w}$);
3. $C(\mathbf{y}, \mathbf{w})$ is concave and continuous in $\mathbf{w}$;
4. $C(\mathbf{y}, \mathbf{w})$ is homogeneous of degree one in input prices: $C(\mathbf{y}, t\mathbf{w}) = t\, C(\mathbf{y}, \mathbf{w})$ for $t > 0$;
5. if $\mathbf{y} > \mathbf{y}'$, then $C(\mathbf{y}, \mathbf{w}) \geq C(\mathbf{y}', \mathbf{w})$ (non-decreasing in $\mathbf{y}$); and
6. $C(0, \mathbf{w}) = 0$ (no fixed costs).

Furthermore, according to Shephard's lemma, the cost-minimizing input demand functions are derivable from the cost functions. If the cost function is differentiable in input prices at the point $(\mathbf{y}^*, \mathbf{w}^*)$ then the following property holds:

7. Shephard's lemma: $\mathbf{x}(\mathbf{y}, \mathbf{w}) = \nabla_w C(\mathbf{y}^*, \mathbf{w}^*)$, where $\mathbf{x}$ denotes the cost-minimizing vector of inputs required to produce the vector $\mathbf{y}^*$ of outputs given input prices $\mathbf{w}^*$.

The long-run total cost methodology shares the stringent behavioral assumption that all inputs are employed at their long-run cost-minimizing level. Some empirical studies present an alternative model where, in the short run, some inputs available to the firms are assumed to be fixed, implying that firms attempt to minimize cost conditional on given quasi-fixed inputs.

There is a dual relationship between the cost function and the transformation function presented above. This implies that if we define a production function and derive its cost function, we can take that cost function to generate a production function. Further, any cost function satisfying properties 1–7 implies some technology for which it represents the minimum costs. Therefore, according to Shephard's (1953) duality theorem, all structural characteristics of production possibilities are contained in the functional specification of the cost function satisfying properties 1–7.

In econometric applications, working with a cost function has many advantages over its dual transformation or production function. For instance, cost functions in comparison to transformation and production functions focus more on economic relationships than on technological relationships, thus facilitating the discussion of economic relationships. Moreover, in production functions typically high correlation between the input variables can result in a multicollinearity problem, which might preclude the estimation of the model or lead to an imprecise interpretation of the estimated coefficients. This problem is less pronounced when a cost function approach is employed. However, two assumptions are required for the cost function to be a valid dual representation of production technology. First, firms should face exogenous input prices and outputs. Second, firms are assumed to adjust input levels to minimize costs. Of course, in certain cases this assumption may be considered too restrictive. It can happen that the firms' optimization strategies do not fully correspond to a perfectly minimal cost function. In such cases, the functions based on cost optimization may still be used as 'behavioral' cost functions and can be helpful in studying the behavior and cost structure of such firms (Bös, 1986, p. 343).

3 Scale, Scope and Cost Efficiency

A large part of the cost estimation literature focuses on the estimation of the returns to scale. Chambers (1988) defines the returns to scale as the proportional change in output as an input bundle is changed by a scalar. If a proportional increase in all inputs results in a higher output increase, then the production function is said to demonstrate increasing returns to scale. If a proportional increase in all inputs results in a lower output increase, then the production function is said to show decreasing returns to scale. If a proportional increase in all inputs results in an output increase in the same proportion, then the production function is said to have constant returns to scale. Whether a production technology exhibits constant, decreasing, or increasing returns to scale has implications for determining the most efficient structure of the industry. In particular, if an industry were characterized by a single-output production technology with increasing returns to scale for a given output range, the efficient industry structure would entail a natural monopoly within that range.

Returns to scale are usually defined in terms of the relative increase in output resulting from a proportional increase in all inputs. In general, returns to scale need to be defined along a specific input-mix ray that is chosen arbitrarily. Considering this problem, Hanoch (1975) noted that it is more relevant to measure returns to scale by the relationship between total cost and output along the expansion path.

To translate the definition of returns to scale in terms of the cost function, Chambers (1988) proposed the concept of 'cost flexibility' that can be measured with the ratio of marginal cost to average cost, or with the cost elasticity of size. The degree of cost flexibility e_{cy} can then be measured using the following expression:

$$e_{cy} = \sum_{k=1}^{m} \frac{\partial \ln C}{\partial \ln y_k} = \sum_{k=1}^{m} \left(\frac{\partial C}{\partial y_k} / \frac{C}{y_k} \right), \tag{25.2}$$

with m different outputs y_k.

The reciprocal of the cost elasticity of size is then defined as the economies of size or economies of scale (ES):

$$ES = 1 \Bigg/ \left(\sum_{k=1}^{m} \frac{\partial \ln C}{\partial \ln y_k} \right). \tag{25.3}$$

There are economies of scale if ES is greater than 1, and conversely, there are diseconomies of scale if ES is below 1. In the case of $RS = 1$ we do not have economies or diseconomies of scale. Chambers shows that the definition of economies of scale is equivalent to the definition derived from production technology if and only if the production technology is homothetic. Generally, the empirical studies on the estimation of cost functions for electricity and gas distribution companies do not impose such an assumption and, therefore, use the concept of economies of scale.[4] From this discussion it follows that the presence of economies of scale shows that the companies have scale inefficiency.

In neoclassical production and cost theory, all output units are homogeneous. Hence, production of any one unit of output is indistinguishable from the next, and can be interchanged without any impact on costs. In energy distribution, however, an industry is characterized not only by the total output produced but also by the structure of the network served. The motivation for this complex characterization arises from the fact that costs in network industries are influenced not only by the total output produced but also by the network structure and area size. Therefore, for network industries such as gas and electricity distribution, the classical definition of economies of scale can be modified, in order to take into account, for instance, that different service areas and/or network structures have different impacts on the costs.

For these reasons, applied economists have suggested including several output characteristics in the cost model specification for network industries. These characteristics should capture the heterogeneity of the outputs, along with aggregated outputs. For example, in the following cost function, some output characteristics variables are included,

$$C = f(\mathbf{y}, \mathbf{w}, \mathbf{q}_{h+k}), \tag{25.4}$$

where C is total cost, $\mathbf{y}$ is a vector of outputs, $\mathbf{w}$ is a vector of input prices, and $\mathbf{q}_{h+k}$ is a vector of variables reflecting output characteristics. The load factor, the number of customers and the size of the electric utility service area are examples of output characteristic variables for the electric industry.

The inclusion in a cost function of the number of customers and/or the size of the service area allows for the distinction of economies of output density, economies of customer density and economies of scale.

Following Roberts (1986), for the single output electricity and gas distribution companies it is possible to define economies of output density (EOD) as the proportional increase in total costs brought about by a proportional increase in output (y), holding all input prices, and the output characteristic variables such as the number of customers

(CU) and the size of the service territory (AS) fixed. This is equivalent to the inverse of the elasticity of total cost with respect to output:

$$EOD = 1\Big/\left(\frac{\partial \ln C}{\partial \ln y}\right). \tag{25.5}$$

The economies of output density exist if EOD is greater than 1, and the presence of diseconomies of output density is indicated if EOD is below 1. In the case of $EOD = 1$, no economies or diseconomies of output density exist. Economies of output density exist if the average costs of an electricity distribution utility decrease as the volume of electricity sold to a fixed number of customers in a service territory of a given size increases. This measure is relevant to decide whether side-by-side competition or local monopoly is the most efficient form in the electricity distribution industry.

Economies of customer density (ECD) are defined as the proportional increase in total costs brought about by a proportional increase in output and the number of customers, holding all input prices, and the other output characteristics fixed. ECD can thus be defined as:

$$ECD = 1\Big/\left(\frac{\partial \ln C}{\partial \ln y} + \frac{\partial \ln C}{\partial \ln CU}\right). \tag{25.6}$$

Similarly, the economies of customer density are not fully exploited if ECD is greater than 1, and conversely, diseconomies of scale are present if ECD is below 1. In the case of $ECD = 1$, no economies or diseconomies of customer density exist. This measure is relevant for analyzing the cost of distributing more electricity to a fixed service area as it becomes more densely populated.

Finally, the economies of scale (ES) are defined as the proportional increase in total costs brought about by a proportional increase in output, the number of customers and the size of the service area, holding all input prices fixed. ES can thus be defined as:

$$ES = 1\Big/\left(\frac{\partial \ln C}{\partial \ln y} + \frac{\partial \ln C}{\partial \ln CU} + \frac{\partial \ln C}{\partial \ln AS}\right). \tag{25.7}$$

The presence of the economies of scale corresponds to ES values greater than 1, while values smaller than 1 indicate diseconomies of scale. In the case of $ES = 1$, no economies or diseconomies of scale exist. This measure is relevant for analyzing the impact on cost of merging two adjacent electricity or gas distribution companies.

In the energy sector, there is a certain tendency that local utility companies operate in both electricity and gas distribution. Generally, this horizontal integration strategy allows the local multi-utility companies to save on costs by exploiting the economies of scope and to provide customers with an integrated set of services. Baumol et al. (1982) highlight that the economies of scope can result from sharing or joint utilization of inputs such as labor and capital. Economies of scope are present when costs can be reduced by joint production of multiple outputs. Following Baumol et al., the degree of global economies of scope ($ESCO$) across two outputs is defined as the ratio of excess costs of separate production relative to the costs of joint production of all outputs:

$$ESCO = \frac{C(y_1, 0) + C(0, y_2) - C(y_1, y_2)}{C(y_1, y_2)}. \tag{25.8}$$

A positive (negative) value for the above expression implies the existence of global economies (diseconomies) of scope.

A common measure of the technical inefficiency (*TIN*) of the firm which produces output, y, with inputs, x, is given by:

$$TIN = \frac{y}{y^*}, \tag{25.9}$$

where y^* is the frontier output associated with the level of input x. This measure does not include allocative inefficiency, namely, the potential savings by reallocating input factors. Cost inefficiency is an input-oriented measure overall inefficiency, which is defined as the distance from the cost frontier.[5] A measure of the cost inefficiency (*CIN*) of the firm which produces output, y, with cost, C, is given by:

$$CIN = \frac{C}{C^*}, \tag{25.10}$$

where C^* is the frontier cost associated with the level of output y. Each of the above-mentioned measures has their respective advantages and drawbacks. A major advantage of *TIN* is that it does not require any data on costs and prices. These data are usually difficult to obtain but are required for estimating *CIN*. However, *TIN* does not provide any information on the cost minimization process, which is more interesting from an economic standpoint. The cost-inefficiency measure includes both allocative and technical inefficiencies but does not provide an easy way to separate these two components.[6] Another important feature of the cost efficiency approach is that it treats the output as given. This is a realistic assumption in most regulated industries where the level of output is set by the regulator or determined by the demand factors.

4 Econometric Approaches to Efficiency Measurement

Measurement of efficiency in the electricity and gas distribution sectors is done by both cost and production functions in the empirical literature. The discussion that follows, while focusing mainly on cost-based methods, applies equally well to methods based on production function. There are two streams of this empirical literature. First, there are studies that estimate a cost function using, for instance, ordinary least squares (OLS) (or more elaborate econometric methods) without a stochastic component for inefficiency where all the companies are assumed to operate on the cost frontier (so no cost inefficiency is observed). From the estimation of this cost function it is possible to calculate the level of scale and scope inefficiency of the companies. Another group of empirical studies adopt the frontier approach assuming that the full cost efficiency is limited to those companies that are identified as the best-practice producers. All other companies are assumed to operate above the cost frontier, hence just non-zero inefficiency. In this case the econometric estimation of the best-practice cost frontier allows the calculation of the companies' level of cost inefficiency in addition to their scale and scope efficiency. Here we refer to the former approaches as 'non-frontier' methods, in which the estimated cost function is based on average performance, as opposed to 'frontier' analysis based on the best observed practice(s).

The estimation of the cost function in line with the non-frontier approach can be based on the OLS estimation of a parametric cost function, usually expressed in logarithms:

$$\ln C_i = f(\mathbf{y}_i, \mathbf{w}_i) + \varepsilon_i, \tag{25.11}$$

where C is total cost incurred by company i, $f(\cdot)$ is the cost function, y is a vector of outputs, w is a vector of input prices, and ε_i is the stochastic error term. The efficiency of parameter estimates can be improved by combining the cost function with the factor share equations implied by Shephard's lemma. According to cost-minimization theory, certain parameters in the cost function are identical to certain parameters in the share equations. Therefore, additional degrees of freedom are gained without the need to estimate any additional parameters. The input share equations take the following form:

$$S_{ji} = g(\mathbf{y}_i, \mathbf{w}_i) + \varepsilon_{iji}, \tag{25.12}$$

where S_{ji} is share cost incurred by company i for input j, $g(\cdot)$ is the input share function, y is a vector of outputs, w is a vector of input prices, and ε_{ji} is the stochastic error term. The cost system is usually estimated using the iterative Zellner technique (Zellner, 1962) for seemingly unrelated regressions (SUR). This allows for the possibility of correlation between the disturbance terms of the cost and share equations for each observation, while assuming independence across observations within a given equation. As the share equations sum to one, an equation must be omitted from the system to implement SUR. However, using the iterative Zellner technique, the resulting estimates are equivalent to maximum likelihood estimates (MLEs) (Kmenta and Gilbert, 1968), and they are invariant to which share equation is deleted (Barten, 1969).

The frontier analysis approach is well developed with a wide variety of parametric methods. More generally, the estimation of production or cost frontier can also be performed using non-parametric approaches. These approaches, such as data envelopment analysis (DEA) and free disposal hull, use linear programming to determine a company's efficiency frontier. In these approaches, the cost frontier is considered as a deterministic function of the observed variables. These methods are non-parametric in that they do not impose any specific functional form or distribution assumption. Thanks to their relative simplicity and availability, such methods, particularly DEA, are quite popular among both researchers and regulators in energy distribution networks – the advantages and drawbacks have been extensively discussed elsewhere.[7] In this chapter, the focus is upon the econometric approaches with a parametric specification. In Section 8 a brief illustration of the DEA method is given through a simple example.

Figure 25.1 presents a general classification of efficiency measurement approaches base on econometric methods. Apart from a few exceptions, all parametric methods have a stochastic element in their frontier function. Thus, this group of methods is also called 'stochastic frontier analysis' (SFA). The main exception with a deterministic frontier is the COLS method. In this approach the inefficiencies are defined through a constant shift of the OLS residuals (see Greene, 1980). As the entire stochastic term is considered as inefficiency, the frontier remains deterministic.

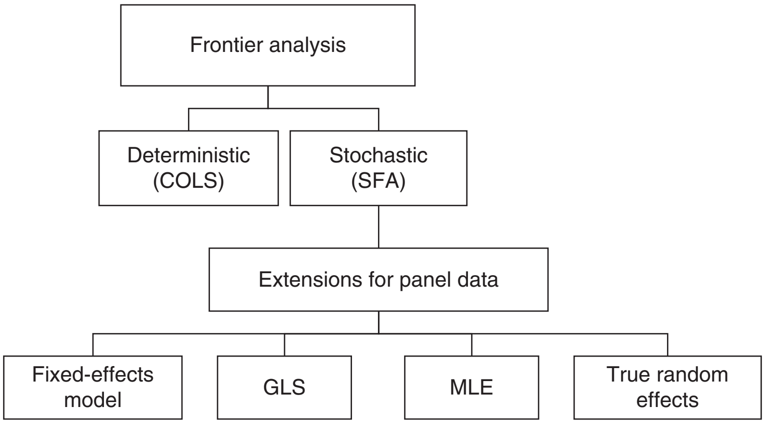

Figure 25.1 Efficiency measurement using econometric methods

The COLS approach is based on the OLS estimation of a parametric cost function, usually expressed in logarithms:

$$\ln C_i = f(\mathbf{y}_i, \mathbf{w}_i) + \varepsilon_i, \tag{25.13}$$

where the parameters are defined above. After correcting the stochastic error term ε_i, by shifting the intercept such that all residuals are positive, the COLS model can be written as:

$$\ln C_i = f(\mathbf{y}_i, \mathbf{w}_i) + \min(\varepsilon_i) + u_i, \text{ with } u_i = \varepsilon_i - \min(\varepsilon_i) \geq 0, \tag{25.14}$$

where u_i is a non-negative term representing the firm's inefficiency. The cost efficiency of firm i is thus given by: $Eff_i = \exp(u_i)$.

The main shortcoming of this method is that it confounds inefficiency with statistical noise: the entire residual is classified as inefficiency. In the stochastic frontier model the error term is divided into two uncorrelated parts: u_i is a one-sided non-negative disturbance reflecting the effect of inefficiency, and v_i is a symmetric disturbance capturing the random noise. Usually the statistical noise is assumed to be normally distributed, while the inefficiency term u_i is assumed to follow a half-normal distribution.[8] A basic SFA model can be written as:

$$\ln C_i = f(\mathbf{y}_i, \mathbf{w}_i) + u_i + v_i. \tag{25.15}$$

This model with a normal-half-normal composite error term can be estimated using the MLE method. Similarly the cost efficiency of firm i is given by: $Eff_i = \exp(u_i)$.

SFA models allow for a random unobserved heterogeneity among different firms (as represented by v_i) but need to specify a functional form for the cost or production function. The main advantage of such methods over deterministic approaches is the separation of the inefficiency effect from the statistical noise due to data errors and omitted variables.

Within the econometric approaches, as represented in Figure 25.1, there exist several models that can be used, and the choice of model is not usually straightforward. Several studies have reported discrepancies in efficiency estimates between different approaches and model specifications. For instance, using a cross-section of 63 power distribution utilities in Europe, Jamasb and Pollitt (2003) show that there are substantial variations in estimated efficiency scores and rank orders across different approaches (parametric and non-parametric) and among different econometric models. Similarly, using data from South America, Estache et al. (2004) provide evidence of 'weak consistency' between parametric and non-parametric methods.[9] These results are supported by two other studies (Farsi and Filippini, 2004, 2005), which show that the efficiency ranking of the companies can differ significantly across econometric models and across different approaches. Such discrepancies are partly due to methodological sensitivity in the estimation of individual efficiency scores and are not limited to a specific network industry.[10] The regulated companies operate in different regions with various environmental and network characteristics that are only partially observed. Given that the unobserved factors are considered differently in each method,[11] the resulting estimates can vary across methods. The magnitude of variation depends on the importance of the unobserved factors, which might change from one case to another.

In most cases, there is no clear criterion for the choice of the model and approach. Thus, it is assumed that the results are valid if they are independently obtained from several models. For instance, Bauer et al. (1998) have proposed a series of criteria that can be used to evaluate if the efficiency estimates from different methods are mutually 'consistent', that is, lead to comparable scores and ranks. However, the empirical results suggest that these criteria are not satisfied in many cases in network industries. The significant uncertainties in efficiency estimates could have important undesired consequences, especially because in many cases the estimated efficiency scores are directly used to reward/punish individual companies through regulation schemes such as price-cap formulas. Given these problems, it is not surprising that the benchmarking models used in electricity and gas networks have frequently been criticized (see, for instance, Irastorza, 2003; Shuttleworth, 2003). The stochastic frontier literature has provided a variety of panel data models that can be used to overcome some of these shortcomings and provide more attractive instruments to use in regulation.[12] In particular, these models can better control for the firm- or network-specific unobserved heterogeneity, which is a source of discrepancy across different benchmarking methods. The use of panel data models is especially interesting as data for several years have become available to an increasing number of regulators in many countries.[13]

The frontier model in (25.13) can be rewritten for panel data using subscripts i and t, respectively, representing the firm and the operation year:

$$\ln C_{it} = f(\mathbf{y}_{it}, \mathbf{w}_{it}) + u_{it} + v_{it}. \tag{25.16}$$

Typically, it is assumed that the heterogeneity term v_{it} is normally distributed and that the inefficiency term u_{it} has a half-normal distribution that is, a normal distribution truncated at zero:

$$u_{it} \sim |N(0, \sigma_u^2)|,\ v_{it} \sim N(0, \sigma_v^2). \tag{25.17}$$

This model is based on the original cost frontier model proposed by Aigner et al. (1977). The firm's inefficiency is estimated using the conditional mean of the inefficiency term as proposed by Jondrow et al. (1982) that is: $E[u_{it}|\hat{\varepsilon}_{it}]$, where $\varepsilon_{it} = u_{it} + v_{it}$.

The first use of panel data models in stochastic frontier models was by Pitt and Lee (1981), who assumed that the inefficiency term u_{it} is constant over time, that is: $u_i \sim |N(0, \sigma_u^2)|$. Pitt and Lee's model is different from the conventional random-effect model in that the individual-specific effects are assumed to follow a half-normal distribution. Important variations of this model were presented by Schmidt and Sickles (1984) who relaxed the distribution assumption and used the generalized least squares (GLS) estimator, and by Battese and Coelli (1988) who assumed a truncated normal distribution. In cases where the individual firm effects (u_i) are correlated with the explanatory variables, the estimated parameters may be biased. Schmidt and Sickles proposed a fixed-effects approach to avoid such biases.

The main restriction of these models is that the unobserved factors are random over time and across firms, which implies that the unobserved network and environmental characteristics that are usually time invariant are not considered as heterogeneity. Moreover, with a few recent exceptions such as Sickles (2005), the variation of efficiency over time is deterministic and/or follows the same functional form for all firms. Given that sources of inefficiency depend on technology shocks and other variations in the input markets as well as the managers' abilities to cope with them, it can be argued that inefficiencies are random over time.[14]

The 'true' panel models extend the original stochastic frontier model to panel data by adding an individual time-invariant effect. These models (Kumbhakar, 1991; Polachek and Yoon, 1996; and more recently Greene, 2005a) include two stochastic terms for unobserved heterogeneity, one for the time-variant factors and one for the firm-specific constant characteristics. Assuming that network and environmental characteristics and their effects on production do not vary considerably over time and that the inefficiency is time variant, these models help separate these unobserved effects from efficiency estimates.

Some of these models have been successfully used in electricity distribution networks (Farsi et al., 2006a), as well as other public service sectors (see Farsi et al., 2005, 2006b). The models can be written by adding a firm-specific stochastic term (α_i) in the right-hand side of equation (25.14):

$$\ln C_{it} = f(\mathbf{y}_{it}, \mathbf{w}_{it}) + \alpha_i + u_{it} + v_{it}. \tag{25.18}$$

The term α_i is an iid random component in a random-effects (RE) framework, or a constant parameter in a fixed-effects (FE) approach. The inefficiency term is assumed to be an iid random variable with half-normal distribution. This implies that the inefficiency is not persistent and each period brings about new idiosyncratic elements, thus new sources of inefficiency. This is a reasonable assumption particularly in industries that are constantly facing new technologies. Such models have an important advantage in that they allow for time-variant inefficiency while controlling for firm-level unobserved heterogeneity through fixed or random effects. The main difficulty of these models is that they are numerically cumbersome.

Another problem arises when the firm-specific effects are correlated with the explanatory variables. In such cases, the RE estimators are affected by heterogeneity bias,[15] but the FE model while being consistent regarding the cost-frontier slopes, usually overestimates efficiency variations. Moreover, there is an important practical problem with the FE model in that it requires the estimation of a large number of parameters, which limits its application to reasonably long panels with sufficient within-firm variation. Generally, in short panels the fixed effects are subject to considerable estimation biases, which directly reflect in the inefficiency scores.[16]

When panel data are available, regulators also have the possibility of using panel data parametric methods for the prediction of intervals for companies' costs. The predicted intervals can be used in the application of yardstick competition (Schleifer, 1985) among companies or to assess if the costs reported by the companies and used, for instance in rate-of-return regulation, are reasonable. In these cases, the regulated companies are required to contain the costs within the interval imposed by the regulator, or have to justify any costs beyond the predicted range. A similar approach has been used in Italian water supply regulation (see Antonioli and Filippini, 2001, for more details). In principle this approach is also possible with cross-sectional data; in panel data, the repeated observations of the same companies are used to identify part of their unobserved time-invariant characteristics and adjust the predictions accordingly. Farsi and Filippini (2004) show in an example that panel data frontier models allow a reasonably low prediction error.

To date the use of panel data parametric methods has been limited to academic research. However, despite the sensitivity problems, these methods can provide the regulators with useful information. As the availability of panel data increases, so too should the application of these methods in regulatory practice.

The choice of functional form $f(\cdot)$ is another important aspect of the parametric estimation of cost frontiers. A variety of functional forms have been employed to estimate cost functions in the literature.[17] The main distinction is between traditional and flexible functional forms. Traditional functional forms are first-order approximations to an arbitrary continuous and twice differentiable function, whereas the flexible functional forms provide a second-order approximation.[18] The most important difference between traditional and flexible functional forms is that the former impose restrictions on the values of the first and second partial derivatives whereas the latter do not. Traditional functional forms such as the Cobb–Douglas and the constant clasticity of substitution impose a priori restrictions on technology. For example, the Cobb–Douglas form imposes the restriction of a unitary elasticity of substitution and a value of economies of scale that does not vary with output. Whereas flexible functional forms such as the translog or quadratic forms allow for values of economies of scale, and input price and substitution elasticities to vary with output. Further, flexible functional forms allow the possibility of testing for several technological characteristics, such as homothetical production technology. Moreover, the quadratic functional form has been considered as one of the most relevant options for estimating scope economies because unlike logarithmic forms, it accommodates zero values for outputs thus, allowing a straightforward identification of scope economies.

The first empirical studies focused on the simple relationships between output and costs (see, for example, Nordin, 1947; Johnston, 1952; Huettner and Landon, 1974). These studies employed uncomplicated cost models assumed to be linear or polynomial in output. Later, cost function models were based on the application of duality theory to neoclassical production theory utilizing more elaborate functional forms. Nerlove's (1963) pioneering study on the cost structure of US electric utilities employed the traditional Cobb–Douglas specification. Christensen et al. (1971)'s paper on translog functional form made a significant impact on the empirical literature in applied production analysis. Most of the analyses of cost structures of electric and gas distribution utilities published in the last two decades use either the Cobb–Douglas or the translog functional form. Starting, for example, from the following long-run cost function:

$$C = C(y_1, w_1, w_2), \tag{25.19}$$

where C is total cost, $w1$, $w2$ are the input prices and $y1$, is the output, the following Cobb–Douglas, translog and quadratic cost functions can be written as:

$$\ln C = \alpha_0 + \alpha_{w1}\ln w_1 + \alpha_{w2}\ln w_2 + \alpha_{y1}\ln y_1 \qquad \text{(COBB}-\text{DOUGLAS)}$$

$$\ln C = \alpha_0 + \alpha_{w1}\ln w_1 + \alpha_{w2}\ln w_2 + \alpha_{y1}\ln y_1 + \frac{1}{2}\alpha_{w1w1}(\ln w_1)^2 + \frac{1}{2}\alpha_{w2w2}(\ln w_2)^2$$
$$+ \frac{1}{2}\alpha_{y1y1}(\ln y_1)^2 + \alpha_{w1w2}(\ln w_1)(\ln w_2) + \alpha_{w1y1}(\ln w_1)(\ln y_1) + \alpha_{w2y1}(\ln y_1)(\ln w_2)$$
$$\text{(TRANSLOG)}$$

$$C = \alpha_0 + \alpha_{w1}w_1 + \alpha_{w2}w_2 + \alpha_{y1}y_1 + \frac{1}{2}\alpha_{w1w1}(w_1)^2 + \frac{1}{2}\alpha_{w2w2}(w_2)^2 + \frac{1}{2}\alpha_{y1y1}(y_1)^2$$
$$+ \alpha_{w1w2}(w_1)(w_2) + \alpha_{w1y1}(w_1)(y_1) + \alpha_{w2y1}(y_1)(w_2) \qquad \text{(QUADRATIC)}$$

5 Empirical Studies on the Scale and Cost Efficiency in the Distribution of Gas

Farsi et al. (2007c) review the limited literature on econometric estimation of cost or production functions in gas distribution companies. A few studies that use a cost function approach to analyze the economies of scale and density and/or the level of cost efficiency in a sample of gas distribution companies are presented here.

Kim and Lee (1996) highlight the importance of accounting for output characteristics in estimating a translog cost function for gas distributors. The following cost model specification was adopted:

$$TC = f(y, q_1, q_2, q_3, \mathbf{P}),$$

where TC indicates the total cost of gas distribution, y is the volume of gas served, q_1, q_2, and q_3 are, respectively, the customer density, the average 'customer size' measured as average consumption and the 'supply rate' measured as the number of total customers relative to the number of total potential customers. **P** is a vector of input prices (labor price and the unit price of pipeline). In the estimation of the cost function, a translog functional form was employed to estimate the cost function of seven companies between 1987 and 1992. Almost all the firms were found to be exhausting their scale economies, and the average cost trend can be expressed as a function of output quantity and spatial characteristics.

Fabbri et al. (2000) estimated a total distribution translog cost function for 31 Italian companies observed during two years. Using the long-term cost function:

$$TC = f(V, C, \mathbf{H}, \mathbf{P}, T, O),$$

where TC indicates the total cost of gas distribution, V is the volume of delivered gas, C is the number of customers, **H** is a vector of territorial variables, **P** is a vector of input prices, T is a time-shift variable and O is a shift variable for private firms. The yearly average cost per employee is used as labor price, the book value of equipment divided by the length of the distribution network is used as capital price and the price of material and services is calculated as the residual expenses divided by network length. Output is measured as the volume of delivered gas and the number of customers. The specification also includes the ratio of network length to the number of customers, share of urban population, the average altitude of the service area, and dummy variables for ownership differences and time effects. The results suggest more cost-efficient production in private firms; however, the economies of scale are not significant at the output levels yet economies of density appear to be considerable. These results are in line with most of the findings reported in other studies.

Farsi et al. (2007c) study the cost structure of gas distribution utilities in Switzerland. Three stochastic frontier models are applied to a panel of 26 companies operating from 1996 to 2000. Efficiency is assumed to be constant over time. The output is measured as total volume of natural gas delivered. Input factors consist mainly of the gas purchased from a transmission company, labor and capital. The cost function is specified as:

$$TC = f(Y, P_C, P_L, P_E, LF, TB, CUD, ASIZE),$$

where TC represents total costs; Y is the energy value of the delivered gas measured in MWh; and P_C, P_L and P_E are, respectively, the prices of capital, labor and purchase price of natural gas. In addition to these variables, the load factor LF, the number of terminal blocks TB, the customer density CUD and the area size $ASIZE$ have been introduced in the model as output characteristics variables. Given the small size of the sample and the large number of parameters[19] in the translog model, a Cobb–Douglas form is used.

The analysis highlights the importance of output characteristics such as customer density and network size. The application of three cost frontier models suggests an average inefficiency of about 7 percent in the sector. This result is robust across all the models. The individual efficiency scores and ranks estimated from different models show a strong correlation. However, the companies identified as 'best' and 'worst' practices change across models. Therefore, the individual efficiency estimates cannot be directly used as X-factors in the price-cap formulas. As for the scale and density economies, the results are more or less consistent with the findings of studies performed in other countries, in that they provide evidence of considerable density economies but insignificant or weak scale economies. This implies that distributors could decrease their average costs by increasing the output as long as they use the same network, but the extension of networks does not result in any significant economies.

The empirical evidence reported suggests that franchised monopolies, rather than side-by-side competition, is the most efficient form of production organization in the gas distribution industry. Further, consolidation of small utilities whose service territories are adjacent is likely to reduce costs. However, for large and medium-sized companies consolidation would not bring about large savings.

6 Empirical studies on the scale and cost efficiency in the distribution of electricity

The literature on econometric estimation of cost functions in electricity distribution companies is relatively large. In this section we present a selection of studies that have used a cost function approach. The emphasis has been placed on the studies that analyze the economies of scale and density and/or the level of cost efficiency.[20]

Roberts (1986) was the first attempt to estimate a cost function for electricity production and distribution using the duality theory and a flexible functional form. Previous cost studies on electricity production and distribution had focused on simpler relationships between cost, output and service area characteristics. The specified cost model assumes that firms follow a two-stage production optimization process. In the first stage they make decisions regarding electricity generation, while in the second they make decisions concerning electricity distribution. Thus, in the second stage, firms are viewed as buyers of electricity either from their own plants or from other suppliers.

A cross-section of 65 privately owned vertically integrated electric utilities in 1978 is considered by Roberts. The electricity generation activity of these utilities was very high. Approximately 80 percent of kWh input was self-generated, while the remaining 20 percent was purchased from other electric generation utilities. The cost structure of this vertically integrated industry is specified in the following total cost function:

$$TC = g(Y_L, Y_H, AS, CU, P_L, P_E, P_C),$$

where TC is the total cost, P_L, P_C and P_E are input prices of labor, capital and purchased electricity. Y_L, Y_H are, respectively, the low- and high-voltage deliveries. To control for other determinants of production and delivery costs, the services area, AS, and the total number of customers, CU, were also included. Therefore, Roberts tried to capture the heterogeneous nature of kWh deliveries by including two outputs as well as two variables characterizing these outputs, as explanatory variables in the cost function. In this way, the economies of density, customer density, and size may be distinguished from each other. The empirical results showed the existence of substantial economies of density and slight economies of customers and size in the production and distribution of electricity; indicating no strong evidence that larger service areas result in any economies of power delivery.

Salvanes and Tjøota's (1994) made the first attempt to estimate a cost function only for distribution companies using a cross-section of 100 Norwegian public electricity distribution utilities in 1988. These companies behave as local natural monopolies, whose activities include transportation of electricity to end users, installation and maintenance of equipment and administration. The following short-run cost function is used to analyze the cost structure:

$$VC = g(Y, N, P_L, P_E, F),$$

where VC is the variable cost specified as a translog functional form, and P_L and P_E are input prices of labor and purchased electricity. The wage rate is defined as the annual labor expenses divided by the total labor force. The aggregate measure of output is given by Y. To control for other determinants of delivery cost, the total number of customers, N, and the length of the distribution lines F were also included.

Like Roberts (1986), Salvanes and Tjøota distinguish between economies of density and economies of scale and identify considerable economies of density but only moderate economies of scale. The coefficient of the capital stock variable is positive, implying variable costs are increasing in fixed input (capital), violating the non-increasing regularity condition at the median of the data.

Filippini (1998) is the first paper with a total cost function only for distribution companies. The firm's total cost of distributing electricity has been represented by a translog cost function:

$$TC = C(y, P_c, P_p, P_l, LF, CU, ST, T),$$

where TC represents total cost and y is the output represented by the total number of kWh delivered, and Pc, Pp and Pl are the prices of capital, kWh input and labor, respectively. LF is the load factor, ST the size of the service territory of the distribution utility measured in squares kilometers and CU the number of customers. These variables are introduced in the model as output characteristics. The load factor has been included in order to capture the impact on cost of the intensity of use of the plant. T is a time variable which captures the shift in technology representing change in technical efficiency. The cost model has been estimated for cross-sectional samples of publicly owned electricity distribution utilities operating in Swiss cities. The dataset is composed of a sample of 39 city electricity distribution utilities observed for four years, from 1988 to 1991. The results indicate the existence of economies of output and customer density as well as economies of scale for most output levels.

The cost efficiency of a sample of Swiss electricity distribution utilities is estimated by Farsi et al. (2006a) using three alternative frontier models to estimate the firm's total cost function:

$$C = C(Y, P_K, P_L, P_P, LF, CU, AS, HGRID, DOT, DW, T),$$

where C represents total cost; Y is the output in kWh; P_K, P_L and P_P are, respectively, the prices of capital, labor and input power; and T is a time variable representing the linear trend in technological progress.

Also included in a cost function model are the six output (and network) characteristics: LF is the 'load factor' defined as the ratio of utility's average load on its peak load; AS the size of the service area served by the distribution utility; CU is the number of customers; $HGRID$ is a binary indicator to distinguish the utilities that operate a high-voltage transmission network in addition to their distribution network; DOT is a dummy variable representing the utilities whose share of auxiliary revenues is more than 25 percent of total revenues; and DW is an indicator for the cases in which more than 40 percent of the service area is covered by forests to capture the relatively high maintenance costs and risk of damage to power lines in forests.

Table 25.1 Efficiency scores of Swiss distribution utilities

	GLS	MLE	TRE
Average	0.868	0.887	0.957
Minimum	0.723	0.735	0.861
Maximum	1	0.993	0.996
Correlation coefficients			
with GLS	1	0.970	0.042
with MLE	0.970	1	0.055

Source: Farsi et al. (2006a).

A Cobb–Douglas functional form has been adopted. The authors excluded the translog form to avoid the potential risk of multicollinearity among second-order terms due to strong correlation between output and the larger number of characteristics variables. An unbalanced panel of 59 Swiss distribution utilities over a nine-year period (1988 to 1996) included 380 observations with a minimum of four observations per company. Three cost-frontier models, namely GLS (Schmidt and Sickles, 1984), MLE (Pitt and Lee, 1981), and the true random effects (TRE) model (Greene, 2005a) were estimated. A descriptive summary of the efficiency estimates from different models and their correlation matrix is presented in Table 25.1. While GLS and MLE models give similar results (12 to 15 percent excess costs), the TRE model predicts a much higher average efficiency rate, implying only about 4 percent excess in costs. There is a high correlation between GLS and MLE estimates whereas the TRE estimates show a weak correlation with the conventional models. These results generally confirm the existence of discrepancies in efficiency estimates between different cost-frontier models.

The TRE model assumes a time-variant inefficiency term and a separate stochastic term for firm-specific unobserved heterogeneity the results suggest that the other models might overestimate the inefficiency. This conclusion is valid to the extent that inefficiencies do not remain constant over time and unobserved network effects remain constant.[21] Unfortunately these relatively new models can only give a partial solution to the sensitivity problems.

In general, the empirical evidence reported in these outlined studies suggests that franchised monopolies, rather than side-by-side competition, is the most efficient form of production organization in the electricity distribution industry. Further, consolidation of small utilities whose service areas are adjacent is likely to reduce costs. However, similar large and medium-sized gas distribution companies, a consolidation would not imply large savings. The concept of small, medium and large companies is relative to the size of the companies included in each sample.

7 Empirical studies on the scale and scope efficiency in the distribution of Electricity and gas

There are only a few studies on the economies of scope in multi-utilities: Mayo (1984) and Sing (1987) in electricity and gas distribution, and Fraquelli et al. (2004) and Farsi et al. (2007a) in the electricity, gas and water sectors. Mayo (1984) estimates a quadratic cost function for two cross-sectional data sets from the US electricity and gas distribution sectors and reports scope economies only for small companies. Sing (1987), also using a cross-sectional dataset including electricity and gas distributors, estimates a generalized translog cost function with a Box–Cox transformation for outputs. In addition to the factor prices of labor, capital and fuel, the customer density is included as an output characteristic. While reporting diseconomies of scope for the sample mean, Sing finds scope synergies for certain output combinations, without any clear pattern regarding the outputs magnitude.

Fraquelli et al. (2004) use data from 90 Italian electricity, gas and water distributors over three years. However, the data are pooled across the years and no panel data models are applied. They compare different functional forms such as the translog cost function with a small value transformation, the generalized translog, the separable quadratic and the composite cost function introduced by Pulley and Braunstein (1992). They conclude that economies of scope exist but their statistical significance can only be asserted over small outputs.

Farsi et al. (2007a) analyzed the cost structure of a panel dataset from 87 electricity, gas and water utilities in Switzerland. Assuming firms minimized cost and used a quadratic cost function specified as:

$$C = C(q^{(1)}, q^{(2)}, q^{(3)}, w^{(1)}, w^{(2)}, w^{(3)}, w^{(4)}, r, \tau^{(1)}, \tau^{(2)}, \tau^{(3)}),$$

where C represents total costs; $q^{(1)}$, $q^{(2)}$ and $q^{(3)}$ are, respectively, the distributed electricity, gas and water during the year, $w^{(1)}$, $w^{(2)}$, $w^{(3)}$ and $w^{(4)}$ are, respectively, the input factor prices for labor and capital services and the purchased electricity and gas; r is the customer density measured by the number of customers divided by the size of the service area measured in square kilometers; and the sector-specific linear trends are represented by $\tau^{(1)}$, $\tau^{(2)}$ and $\tau^{(3)}$, respectively, for the electricity, gas and network sectors.

The econometric analysis based on a random effect GLS model and a random coefficient specification uses an unbalanced panel dataset containing financial and technical information from 87 companies observed during the nine-year period between 1997 and 2005. The results suggest the presence of scope and scale economies at most output levels with a well-behaved variation of the synergies, output indicated a fall (rise) in both scale and scope economies as outputs increase (decrease).

8 Case study: benchmarking and regulation in the electricity distribution sector

The use of benchmarking in the regulation of electricity distribution operators is a prominent example of a direct policy application of efficiency measurement methods. In many countries regulatory reforms are relatively advanced in the electricity industry. A characterizing feature of these reforms is that the introduction of competition in the generation and retail sectors is often combined with the implementation of an incentive regulation scheme, and increasingly benchmarking practices, in the distribution sector. The use of parametric methods in the estimation of efficiency is however limited to a few cases.[22] An interesting example is the application of COLS by the UK electricity regulator (Ofgem) in setting the prices based on individual X-factors estimated from a sample of 14 observations (Pollitt, 2005). The validity of the adopted approach can be questioned mainly because of the limited size of the sample and the deterministic nature of the frontier that does not distinguish between inefficiency and statistical noise. The robustness of the efficiency estimates could be improved using panel data and a stochastic frontier approach.

In this section we study a similar example of benchmarking of power distribution utilities to illustrate the potential differences across models and the resulting sensitivity problems in estimating efficiency, encountered by the regulators. These problems are particularly important in cases such as Ofgem, in which the benchmarking analysis is based on cross-sectional data (CEPA, 2003). Here, the example has been chosen from the Swiss power distribution sector. The sample consists of 52 utilities operating in 1994. The cost efficiency of these companies has been analyzed using two parametric approaches (COLS and SFA) and one non-parametric method (DEA). This choice provides the opportunity for comparing the results among different approaches to draw conclusions for the regulators.

A three-input single-output production function has been considered. The output is measured as the total delivered electricity in kWh, and the three input factors are capital, labor and the input power purchased from the generator. Capital price is measured as the ratio of capital expenses (depreciation plus interest) to the total installed capacity of the utility's transformers in kVA (kilo volt-ampere). The capital costs are approximated by the residual costs that is, total costs minus labor and purchased power costs. Labor price is defined as the average annual salary of the firm's employees.

The costs of distribution utilities consist of two main parts: the costs of the purchased power and the network costs including labor and capital costs. There are therefore two alternatives for measuring cost efficiency: total costs approach and network costs approach. The network costs approach has a practical advantage in that the estimated average costs can be directly used in a price-cap formula.[23] However, this approach neglects the potential inefficiencies in the choice of the generator and also in the possibility of substitution between capital and input energy. In this example we use the first approach based on the total costs.

In addition to input prices and output, three output characteristics are included. The resulting specification of the cost function can be written as:

$$C = C(Y, PK, PL, PP, LF, CU, AS), \tag{25.20}$$

where C represents total cost; Y is the output in kWh; P_K, P_L and P_P are, respectively, the prices of capital, labor and input power; LF is the 'load factor' defined as the ratio of utility's average load on its peak load; CU is the number of customers; and AS the size of the service area served by the distribution utility.

For the parametric models we have chosen a Cobb–Douglas functional form. The condition of linear homogeneity in input prices is imposed by dividing the input prices by the price of purchased electricity. The cost function can therefore be formulated as:

$$\begin{aligned}\ln\left(\frac{\mathrm{C}}{P_P}\right)_i = {} & \beta_0 + \beta_Y \ln Y_{it} + \beta_K \ln\left(\frac{P_K}{P_P}\right)_i + \beta_L \ln\left(\frac{P_L}{P_P}\right)_i \\ & + \gamma_1 \ln LF_i + \gamma_2 \ln AS_i + \gamma_3 \ln CU_i + r_i, \end{aligned} \tag{25.21}$$

with $i = 1, 2, \ldots, N$, where r_i represents the residuals, namely, a mean-zero iid error term for COLS and a composite normal-half-normal iid term for the SFA model.

The specification given in (25.20) can be readily used in the DEA method. In this method there is no need to specify any functional form. In a sample of N companies with a k-input–m-output production function with variable returns to scale (VRS), the measurement of cost efficiency using DEA method reduces to the following minimization problem:

$$\min_{\lambda, x_i} w_i' x_i$$

$$\text{s.t.:} \ -y_i + Y\lambda \geq 0, x_i - X\lambda \geq 0, \mathbf{N}'\lambda = 1, \lambda \geq 0, \qquad (25.22)$$

where w_i and x_i are $k \times 1$ vectors, respectively, representing input prices and quantities for firm i ($i = 1, 2, \ldots, N$); y_i is an $m \times 1$ vector representing the given output bundle; X and Y are, respectively, input and output matrices namely, a $k \times N$ and an $m \times N$ matrix consisting of the observed input and output bundles for all the companies in the sample; $\mathbf{N}$ is an $N \times 1$ vector of ones; and λ is an $N \times 1$ vector of non-negative constants to be estimated. The VRS property is satisfied through the convexity constraint ($\mathbf{N}\lambda = 1$) that ensures companies are benchmarked against companies with similar size.

The minimization problem given in (25.22) can be solved by linear programming (LP) methods. The LP algorithm finds a piece-wise linear isoquant in the input space, which corresponds to the minimum costs of producing the given output at any given point. The solution gives the minimum feasible costs for each company namely, $w_i' x_i^*$, where x_i^* is the optimal input bundle for firm i. The cost efficiency of each production plan is then estimated as its distance to the envelope. Namely, firm i's cost efficiency is therefore obtained by: $Eff_i = (w_i' x_i^*)/(w_i' x_i^o)$ where x_i^o is the observed input bundle used by company i.

The quantities of labor, capital stock and the amount of input energy are considered as input. Labor and capital inputs are, respectively, measured as the number of full-time equivalent employees and the installed capacity of the transformers.[24] The output (Y) and the three output characteristics are consiered as output. With the exception of the load factor (LF) all these characteristics take resources, thus can be considered as an output. As for the load factor, since a higher LF implies a smoother demand, thus lower costs, the corresponding output characteristics in the DEA model is defined as the inverse of LF. Therefore, the DEA model can be considered as a production with three inputs and four outputs. We assume VRS for the DEA model.[25]

The three models have been applied to the cross-sectional data (from 52 companies in 1994). Summary statistics of the estimated efficiency scores are given in ~~Error! Reference source not found.~~. The efficiency scores are normalized to a scale between 0 and 1, where the highest value (1) implies a perfectly efficient company and the difference with 1 approximates the percentage of the total costs that the company can potentially save. The results in ~~Error! Reference source not found.~~ suggest, the companies are on average about 86 to 92 percent efficient. The COLS efficiency scores are lower than the other models by an average of 6 percent. COLS and DEA methods are similar in that neither accounts for stochastic variation in the frontier. However, the DEA model has a non-parametric frontier, which allows certain flexibility for a better adjustment with different

Table 25.2 Efficiency scores of Swiss distribution utilities (52 utilities in 1994)

	DEA	SFA	COLS
Average	0.917	0.920	0.858
Minimum	0.734	0.819	0.727
Maximum	1	0.977	1
Correlation coefficients			
with DEA	1	0.563	0.603
with SFA	0.563	1	0.961

companies. The average efficiency estimate is quite similar between the SFA and DEA models, suggesting that a rigid model like COLS can underestimate the efficiency. These results also suggest that in our example, allowing for stochastic variation or removing parametric restrictions have at least on average, a similar effect on efficiency estimates.

The correlation coefficients between the efficiency scores obtained from different models are also listed in ~~Error! Reference source not found.~~ Although the COLS and SFA estimates show quite a high correlation, their correlation with the DEA estimates is relatively low. These results suggest that the efficiency ranking of the companies could change considerably, depending on the adopted model.

In order to see the differences in ranking individual companies, the rank status of the 10 most-efficient and least-efficient companies according to the SFA method is considered. ~~Error! Reference source not found.~~ lists the efficiency ranks of these 20 companies based on the two other models. The results indicate a similar ranking across the two parametric methods (SFA and COLS), which is considerably different from that of DEA. However, the differences are less important for the first 10 companies. In fact, the DEA model predicts a higher than 98 percent efficiency for all these companies. According to this model, 19 companies are perfectly efficient and 24 companies have an efficiency of higher than 95 percent. But for the 10 companies at the bottom of the list, the differences are quite considerable. For instance, two of these companies are evaluated as perfectly efficient by the DEA model. Ten of the 19 companies evaluated as 100 percent efficient by DEA are less than 95 percent efficient, and three are less than 90 percent efficient.

Overall, our comparison shows that the DEA model predicts perfect efficiency more often than SFA. This might be due to the fact this model has no parametric restriction, thus provides more flexibility to account for unobserved differences among companies. However, such perfect efficiency scores might be due to the sensitivity of the DEA model to outlier values and/or to the 'curse of dimensionality', a general problem in non-parametric methods with a large number of variables[26] Unfortunately, there is no simple method to identify the extent of such problems, especially for individual companies.

This example illustrates a main problem in benchmarking analysis that is, the discrepancy of the results across different methods. In some cases, the sensitivity of individual efficiency estimates is so high that a slight change in the model's assumptions or including an additional variable might change the results considerably. Given the extremely large variety of models and specifications, this problem does not appear to have a clear solution. However, as our example suggests, the sensitivity problems are less severe if the efficiency is estimated at the sector level rather than for individual companies.

Table 25.3 Efficiency ranks for the 'best' and 'worst' practices (1994)

Companies ordered according to SFA	DEA*	COLS
1	22	1
2	1–19	2
3	1–19	4
4	1–19	3
5	1–19	5
6	1–19	6
7	20	7
8	1–19	8
9	1–19	9
10	1–19	10
.	.	.
.	.	.
.	.	.
43	1–19	43
44	47	44
45	41	46
46	39	45
47	45	47
48	46	49
49	1–19	48
50	34	50
51	52	51
52	38	52

Note: * According to DEA method, 19 companies are 100% efficient.

9 Concluding remarks

We presented an overview of the different methods of efficiency measurement in the electricity and gas distribution sectors. Asserting that the productive efficiency should be considered in the design and organization of the sector as well as in the incentive regulation of individual companies, we presented the efficiency concept with regard to three aspects: scale, scope and cost efficiency. The scale and scope efficiency are mainly related to the market structure, particularly the company's output level and mix, whereas cost efficiency is mainly due to the specific company's economic performance facing the external market conditions. An effective regulatory reform should consider all the dimensions of productive inefficiency and its patterns across companies. An adequate understanding of these patterns requires a clear distinction between the three aspects of efficiency.

Among these three concepts, the estimation of inefficiencies related to the scale and scope of the distribution utilities is relatively straightforward. In fact technological characteristics such as the economies of scale and scope can be estimated using the coefficients of a cost function. These estimates help determine the optimal size and scope of the production and the extent of natural monopoly in gas and electricity distribution.

Regulatory reforms should aim at an optimal organization of the distribution networks that allows companies the greatest possibility of using various synergies to reduce their costs. In many cases the empirical evidence reported in the literature suggests that an optimal organization might require relatively large distribution utilities that operate in both the gas and the electricity sectors.

Moreover, the empirical findings generally confirm the natural monopoly characteristic of energy distribution networks favoring local monopoly on side-by-side competition models. In such situations, an effective incentive regulation system is especially important because of the absence of direct competition among companies. Hence, the measurement of companies' economic performance in terms of cost efficiency or similar indicators has an essential role in ensuring the productive efficiency as well as restricting the monopolists' market power.

This chapter illustrates the measurement of cost efficiency to be the contentious issue that it is, especially if the performance of the individual companies is of interest. We argue that an effective regulation system requires appropriate models for the measurement of cost efficiency. Such models need to account for unobserved heterogeneity across companies, which can be provided by certain panel data models. Moreover, since different models posit various assumptions on the variation of efficiency across companies and over time, it is important that these assumptions are clearly specified and the efficiency estimates are interpreted accordingly. In many cases, it is important to view the efficiency from several angles, which requires the application of several models with different assumptions.

Notes

1. In addition to inflation, the changes beyond companies' control may include changes in input factor prices and exogenous changes in demand and network characteristics, generally referred to as 'Z-factors'.
2. Shuttleworth (2005) provides a critical overview of the use of benchmarking in the regulation of electricity networks.
3. See Jamasb and Pollitt (2003), Estache et al. (2004) and Farsi and Filippini (2004) for examples.
4. In the literature on the cost structure of electricity distributors, some authors have estimated a variable cost function. In this case it is possible to calculate a measure of utilization economies. For instance, Caves and Christensen (1988) define economies of utilization as unity divided by a proportional increase in variable cost resulting from a proportional increase in output holding the capital stock constant. For an empirical analysis on the electricity distribution sector, see Filippini (1996).
5. 'Production frontier' represents the maximum output produced by a given set of inputs, whereas 'cost frontier' defines the minimum costs of producing an output level with given input prices.
6. The only way to disentangle allocative and technical inefficiencies in a cost-frontier framework is through input factor demand equations. Because of the complexity of the resulting error structure, a satisfactory econometric solution remains to be developed (Greene, 1997; Kumbhakar and Lovell, 2000).
7. See for instance, Murillo-Zamorano (2004) for a general presentation of different methodologies and Coelli et al. (2003) for more details on DEA.
8. Other extensions of the basic frontier model have also considered exponential and truncated normal distributions for the inefficiency term. See, for instance, Battese and Coelli (1992).
9. Other authors such as Horrace and Schmidt (1996), Jensen (2000) and Street (2003) reported substantial errors and inconsistency problems in the estimation of individual efficiency scores in cross-sectional data.
10. Horrace and Schmidt (1996), Jensen (2000) and Street (2003) reported substantial errors and inconsistency problems in the estimation of individual efficiency scores in cross-sectional data. The first paper shows that such problems persist in a panel dataset with six periods.
11. While the econometric approach uses additive stochastic terms, the linear programming method allows heterogeneity in production by relaxing the restrictions imposed by a specific functional form.
12. Some of these alternative models have been used in a few recent studies (Alvarez et al., 2004; Greene, 2005a; Farsi and Filippini, 2004; Farsi et al., 2006a).
13. Panel datasets are characterized by repeated observations for a sample of units over several periods.

14. Alvarez et al. (2004) show that even in cases where inefficiency is due to time-invariant factors such as constant managers' capability, the resulting inefficiencies can vary over time because it depends on a host of time-variant factors that have an interacting effect with the manager's skills.
15. The term 'heterogeneity bias' has been used by Chamberlain (1982) to refer to the bias induced by the correlation between individual effects and explanatory variables in a general RE model.
16. See Greene (2005b) for more details. This author considers a panel of 5 years as a short panel.
17. Griffin et al. (1987) and Chambers (1988) provide insightful overviews of various functional forms in applied production analysis.
18. Within a group of flexible functional forms we can differentiate those functional forms derived from second-order Taylor series approximations, such as the translog, the generalized Leontief and the quadratic or the functional forms derived from Fourier or Laurent series approximations. The former are characterized by local flexibility, which implies a perfect approximation for an arbitrary function and its first two derivatives at a particular point. The latter are distinguished by their global flexibility. In applied work, however, the global approximation in Fourier or Laurent form is not commonly used, mainly because they usually need a considerably higher number of parameters than are required for a functional form derived from second-order Taylor approximations.
19. The number of coefficients in a translog model would have been 36, which results in a relatively small number of degrees of freedom in a sample of 26 companies with 129 observations.
20. For a review of a larger number of studies, see Ramos-Real (2005).
21. If instead it is assumed that inefficiencies are persistent and do not change considerably over time, then the results obtained from conventional panel models provide better estimates of the inefficiencies.
22. See Jamasb and Pollitt (2001) and Farsi et al. (2007b) for surveys of different regulation practices in electricity markets around the world.
23. Price cap is generally applied to the network access.
24. The measurement unit of input factors is not relevant, as long as the prices are defined such that the resulting costs have the same unit.
25. The alternative assumption would be constant returns to scale. This assumption is too restrictive because it implies that all companies operate at the optimal scale. See Coelli et al. (2005) for more details.
26. See Simar and Wilson (2000) for a discussion of 'curse of dimensionality' and Simar (2003) for the outliers issue.

References

Aigner, D., C.A.K. Lovell and P. Schmidt (1977), 'Formulation and estimation of stochastic frontier production function models', *Journal of Econometrics*, **6**, 21–37.

Alvarez, A., C. Arias and W.H. Greene (2004), 'Accounting for unobservables in production models: management and inefficiency', Working Paper E2004/72, Centro de Estudios Andaluces, Seville, Spain.

Antonioli, B. and M. Filippini (2001), 'The use of a variable cost function in the regulation of the Italian water industry', *Utilities Policy*, **10** (3–4), 181–7.

Barten, A.P. (1969), 'Maximum likelihood estimation of a complete system of demand equations', *European Economic Review*, **1**, 25–9.

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

Battese, G.E. and T. Coelli (1992), 'Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India', *Journal of Productivity Analysis*, **3**, 153–69.

Bauer, P.W., A.N. Berger, G.D. Ferrier and D.B. Humphrey (1998), 'Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods', *Journal of Economics and Business*, **50**, 85–114.

Baumol, W.J., J.C. Panzar and R.D. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, New York: Harcourt Brace Jovanovich.

Bös, D. (1986), *Public Enterprise Economics*, Amsterdam: North-Holland.

Caves, D.W. and L.R. Christensen (1988), 'The importance of economies of scale, capacity utilization, and density in explaining interindustry differences in productivity growth', *Logistics and Transportation Review*, **24**, 3–31.

CEPA (2003), *Background to work on assessing efficiency for the 2005 distribution price control review*, Scoping study, Final report, Prepared for the UK Office of Gas and Electricity Markets (Ofgem), Cambridge Economic Policy Associates, available at: www.ofgem.gov.uk (accessed 16 February 2009).

Chamberlain, G. (1982), 'Multivariate regression models for panel data', *Journal of Econometrics*, **18**, 5–46.

Chambers, R.G. (1988), *Applied Production Analysis: A Dual Approach*, Cambridge: Cambridge University Press.

Christensen, L.R., D. Jorgenson and L.J. Lau (1971), 'Trascendental logarithmic production frontiers', *Review of Economics and Statistics*, **55**, 28–45.

Coelli, T., A. Estache, S. Perelman and L. Trujillo (2003), *A Primer on Efficiency Measurement for Utilities and Transport Regulators*, World Bank Institute of Development Studies, Washington, DC: World Bank.

Coelli, T., D.S.P. Rao and G.E. Battese (2005), *An Introduction to Efficiency and Productivity Analysis*, 2nd edn, New York: Springer Verlag.

Estache, A., M.A. Rossi and C.A. Ruzzier (2004), 'The case for international coordination of electricity regulation: evidence from the measurement of efficiency in South America', *Journal of Regulatory Economics*, **25** (3), 271–95.

Fabbri, P., G. Fraquelli and R. Giandrone (2000), 'Costs, technology and ownership of gas distribution in Italy', *Managerial and Decision Economics*, **21**, 71–81.

Farsi, M., A. Fetz and M. Filippini (2007a), 'Economies of scale and scope in the Swiss multi-utilities sector', CEPE Working Paper 59, Centre for Energy Policy and Economics, ETH Zurich, September.

Farsi, M., A. Fetz and M. Filippini (2007b), 'Benchmarking and regulation in the electricity distribution sector', in M. Marrelli, F. Padovano and I. Rizzo (eds), *Servizi Pubblici: Nuovo tendenze nella regolametazione nella produzione et nel finanziamento*, Milano: Franco Angeli, pp. 159–76.

Farsi, M. and M. Filippini (2004), 'Regulation and measuring cost efficiency with panel data models: application to electricity distribution utilities', *Review of Industrial Organization*, **25** (1), 1–19.

Farsi, M. and M. Filippini (2005), 'A Benchmarking analysis of electricity distribution utilities in Switzerland', Working Paper No. 43, Centre for Energy Policy and Economics, Swiss Federal Institute of Technology, Zurich, Switzerland.

Farsi, M., M. Filippini and W.H. Greene (2005), 'Efficiency measurement in network industries: application to the Swiss railway companies', *Journal of Regulatory Economics*, **28** (1), 69–90.

Farsi, M., M. Filippini and M. Kuenzle (2007c), 'Cost efficiency in the Swiss gas distribution sector', *Energy Economics*, **29** (1), 64–78.

Farsi, M., M. Filippini and W.H. Greene (2006a), 'Application of panel data models in benchmarking analysis of the electricity distribution sector', *Annals of Public and Cooperative Economics*, **77** (3), 271–90.

Farsi, M., M. Filippini and M. Kuenzle (2006b), 'Cost efficiency in regional bus companies: an application of new stochastic frontier models', *Journal of Transport Economics and Policy*, **40** (1), 95–118.

Filippini, M. (1996), 'Economies of scale and utilization in the Swiss electric power distribution industry', *Applied Economics*, **28**, 543–50.

Filippini, M. (1998), 'Are municipal electricity distribution utilities natural monopolies?', *Annals of Public and Cooperative Economics*, **2**, 157–74.

Fraquelli, G., M. Piacenza and D. Vannoni (2004), 'Scope and scale economies in multi-utilities: evidence from gas, water and electricity combinations', *Applied Economics*, **36** (18), 2045–57.

Greene, W.H. (1997), 'Frontier production functions', in M.H. Pesaran and P. Schmidt (eds), *Handbook of Applied Econometrics*, *Vol. II: Microeconomics*, Oxford: Blackwell, ch. 3, pp. 81–166.

Greene, W.H. (1980), 'Maximum likelihood estimation of econometric frontier functions', *Journal of Econometrics*, **13**, 27–56.

Greene, W.H. (2005a), 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics*, **126** (2), 269–303.

Greene, W.H. (2005b), 'Fixed and random effects in stochastic frontier models', *Journal of Productivity Analysis*, **23** (1), 7–32.

Griffin, R.C., J.M. Montgomery and M.E. Rister (1987), 'Selecting functional form in production function analysis', *Western Journal of Agricultural Economics*, December, 216–27.

Hanoch, G. (1975), 'The elasticity of scale and the shape of average costs', *American Economic Review*, **65**, 956–65.

Horrace, W.C. and P. Schmidt (1996), 'Confidence statements for efficiency estimates from stochastic frontier models', *Journal of Productivity Analysis*, **7**, 257–82.

Huettner, D.A. and J.H. Landon (1974), 'Electric utilities: economies and diseconomies of scale', Working Paper No. 55, Research Program in Industrial Economics, Case Western Reserve University, Cleveland, OH.

Irastorza, V. (2003), 'Benchmarking for distribution utilities: a problematic approach to defining efficiency', *Electricity Journal*, **16** (10), 30–38.

Jamasb, T. and M. Pollitt (2001), 'Benchmarking and regulation: international electricity experience', *Utilities Policy*, **9** (3), 107–30.

Jamasb, T. and M. Pollitt (2003), 'International benchmarking and regulation: an application to European electricity distribution utilities', *Energy Policy*, **31** 1609–22.

Jehle, G.A. and Ph.J. Reny (1998), *Advanced Microeconomic Theory*, Englewood Cliffs, NJ: Prentice-Hall International.

Jensen, U. (2000), 'Is it efficient to analyse efficiency rankings?', *Empirical Economics*, **25**, 189–208.

Johnston, J. (1952), 'Statistical cost functions in electricity supply', *Oxford Economic Papers*, **7**, 68–105.

Jondrow, J., I. Materov, K. Lovell and P. Schmidt (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics*, **19** (2/3), August, 233–8.

Kim, T.-Y. and J.-D. Lee (1996), 'Cost analysis of gas distribution industry with spatial variables', *Journal of Energy and Development*, **20** (2), 247–67.

Kmenta, J. and R.F. Gilbert (1968), 'Small sample properties of alternative estimators of seemingly unrelated regressions', *Journal of the American Statistical Association*, **63**, 1180–200.

Kumbhakar, S.C. (1991), 'Estimation of technical inefficiency in panel data models with firm- and time-specific effects', *Economics Letters*, **36**, 43–8.

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

Mayo, J.W. (1984), 'Multiproduct monopoly, regulation, and firm costs', *Southern Economic Journal*, **51** (1), 208–18.

McFadden, D. (1978), 'Cost, revenue, and profit functions', in Melvyn Fuss and Daniel L. McFadden (eds), Production Economics: A Dual Approach to Theory and Applications Volume I: The Theory of Production, Amsterdam: North-Holland, pp. 3–109.

Murillo-Zamorano, L.R. (2004), 'Economic efficiency and frontier techniques', *Journal of Economic Surveys*, **18** (1), 33–77.

Nerlove, M. (1963), 'Return to scale in electricity supply', in C.F. Christ (ed.), *Measurement in Economic Studies in Honor of Yehuda Grunfeld*, Stanford, CA: Stanford University Press, pp. 167–98.

Nordin, J. (1947), 'Note on a light plant's cost curves', *Econometrica*, **15**, 321–35.

Pitt, M. and L. Lee (1981), 'The measurement and sources of technical inefficiency in the Indonesian weaving industry', *Journal of Development Economics*, **9**, 43–64.

Polachek, S. and B. Yoon (1996), 'Panel estimates of a two-tiered earnings frontier', *Journal of Applied Econometrics*, **11**, 169–78.

Pollitt, M. (2005), 'The role of efficiency estimates in regulatory price reviews: Ofgem's approach to benchmarking electricity networks', *Utilities Policy*, **13**, 279–88.

Pulley, L.B. and Y.M. Braunstein (1992), 'A composite cost function for multiproduct firms with an application to economies of scope in banking', *Review of Economics and Statistics*, **74** (2), 221–30.

Ramos-Real, F.J. (2005), 'Cost functions and the electric utility industry: a contribution to the debate on deregulation', *Energy Policy*, **33** (1), 69–87.

Roberts, M.J. (1986), 'Economies of density and size in the transmission and distribution of electric power', *Land Economics*, **62**, 378–87.

Salvanes, K.G. and S. Tjøota (1994), 'Productivity differences in multiple output industries: an empirical application to electricity distribution', *Journal of Productivity Analysis*, **5**, 23–43.

Schleifer, A. (1985), 'A theory of yardstick competition', *Rand Journal of Economics*, **16** (3), 319–27.

Schmidt, P. and R.E. Sickles (1984), ,Production frontiers and panel data', *Journal of Business and Economic Statistics*, **2**, 367–74.

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

Shuttleworth, G. (2003), 'Firm-specific productive efficiency: a response', *Electricity Journal*, **16** (3), 42–50.

Shuttleworth, G. (2005), 'Benchmarking of electricity networks: practical problems with its use for regulation', *Utilities Policy*, **13**, 310–17.

Sickles, R. (2005), 'Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings', *Journal of Econometrics*, **126** (2), 305–34.

Simar, L. (2003), 'Detecting outliers in frontier models: a simple approach', *Journal of Productivity Analysis*, **20**, 391–424.

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.

Sing, M. (1987), 'Are combination gas and electric utilities multiproduct natural monopolies?', *Review of Economics and Statistics*, **69** (3), 392–8.

Street, A. (2003), 'How much confidence should we place in efficiency estimates?', *Health Economics*, **12** (11), 895–907.

Varian, H.R. (1992), *Microeconomic Analysis*, New York: Norton.

Zellner, A. (1962), 'An efficient method of estimating seemingly unrelated regressions and test for aggregation bias', *Journal of the American Statistical Association*, **58**, 348–68.