Benchmarking and Firm Heterogeneity in Electricity Distribution:
- A Latent Class Analysis of Germany

Astrid Cullmann

25.03.2008

submitted to

EARIE

Astrid Cullmann
DIW Berlin
Department of Industry, Innovation and Service
Mohrenstrasse 58
10117 Berlin

acullmann@diw.de
tel.: +49-30-89789-679
fax.: +49-30-89789-103

Abstract

In January 2009 Germany introduces incentive regulation for the electricity distribution sector based on results obtained from econometric and nonparametric benchmarking analysis. One main problem for the regulator in assigning the relative efficiency scores are unobserved firm-specific factors such as network and technological differences. Comparing the efficiency of different firms usually assumes that they operate under the same production technology, thus unobserved factors might be inappropriately understood as inefficiency. To avoid this type of misspecification in regulatory practice estimation is carried out in two stages: in a first stage observations are classified into two categories according to the size of the network operators. Then separate analyses are conducted for each sub-group. This paper shows how to disentangle the heterogeneity from inefficiency in one step, using a latent class model for stochastic frontiers. As the classification is not based on a priori sample separation criteria it delivers more robust, statistical significant and testable results. Against this backround we analyze the level of technical efficiency of a sample of 200 regional and local German electricity distribution companies for a balanced panel data set (2001-2005). Testing the hypothesis if larger distributors operate under a different technology than smaller ones we assess if a single step latent class model provides new insights to the use of benchmarking approaches within the incentive regulation schemes.

Keywords: stochastic frontiers, latent class model, electricity distribution, incentive regulation

JEL classification: C24, C81, D24, L94

1 Introduction

Electricity distribution and transmission are traditional network industries and are therefore characterized by the typical properties of natural monopolies. In contrast to the generating and supply segments in the electricity market where competition has been introduced, the distribution sector is highly regulated. The last decades of utility regulation were characterized by traditional cost of service regulation schemes by which companies recovered their costs with a risk-free fixed rate of return (Farsi et al., 2007; Joskow, 2006). Hence, distribution companies had little or even no incentive for cost minimization. Therefore, incentive regulation schemes have become increasingly important in Europe with the goals of reducing costs and increasing efficiency. Across European countries price or revenue cap regulation is extensively used in electricity distribution. Within this framework price/revenue caps are set based on the formula $RPI - X$ (see Beesley and Littlechild, 1989). The maximum rate of price increase equals the inflation rate of the retail price index (RPI) less the expected efficiency savings (X). Thus, regulated distribution companies have incentives to increase their profits by improving their productivity at a higher rate than the assigned X-factor (Farsi et al., 2007). In 2009 Germany has begun to implement the framework in the electricity and natural gas distribution sector.

The determination of the X-factors for setting price/revenue caps is usually based on empirical results obtained from benchmarking analysis. The efficiency performance of the companies is therefore evaluated against a reference, best practice, performance. This framework is favored by European regulators concerned about the robustness and reliability of the empirical outcome of individual efficiency estimation.[1] The development and advancement of benchmarking models for a consistent practical application has therefore been an important research aspect in efficiency analysis. The empirical literature (see Jamasb and Pollitt, 2001; Farsi et al., 2007) can be divided into two major groupings: nonparametric and parametric methods. While the nonparametric methods construct the reference technology, the efficiency frontier, by means of linear programming methods, parametric approaches assume a functional form for the underlying production process.

Empirical evidence shows that the individual efficiency estimates are sensitive to the adopted benchmarking models and approaches (Farsi and Filippini, 2004). Thus, the choice of the empirical approach has a strong impact on the price/revenue cap setting (and therefore the economic and financial conditions of the companies). Further, implementation is always affected by the challenge of data availability. Some firm-specific

[1] Shuttleworth (2005) gives a critical summary of the implementation of benchmarking analysis in price cap regulation.

factors are either unobservable, too complex to be accounted for appropriately, or no data is available for the empirical analysis. However, these unobserved characteristics between firms may have an important impact on the underlying production process and therefore the reference or efficiency frontier.[2]

In the traditional framework it is assumed that firms operate under the same production technology. Therefore these unobserved factors might be inappropriately understood as inefficiency. To avoid these types of misspecification estimation is often carried out in two stages in regulatory practice. First, defining different categories based on a priori sampling criteria and then separate analyses for each sub-group. In Germany, under the assumption that the size of the network operators implies different technological and network characteristics, observations are classified ex-ante into two categories: The German Incentive Regulation (ARegV (§24)) determines for the German electricity distributors that network operators with fewer than 30.000 customers connected directly or indirectly to their distribution system can choose to take part in a simplified procedure. Thus, an ex-ante sampling splitting between large and small distributors is present in the German regulation. This shows the implicit assumption that omitted unobserved technological differences between larger and smaller operators might be present and if not taken into account, inappropriately labeled as inefficiency.[3]

Following Kumbhakar and Orea (2004) and Greene (2005b) we suggest to use a single-stage approach to account for heterogeneity within the stochastic frontier analysis using the latent class framework for stochastic frontiers. Here, different technologies and efficiency frontiers for different classes or groups of firms can be identified without the need for a priori sampling separation information (Greene, 2005a,b). This is motivated by the fact that a priori selection might be arbitrary, lacks statistical foundation and is therefore not testable. We apply the latent class model for stochastic frontiers using a multi-input multi-output parametric input distance function for a balanced panel data set (2001-2005) for 200 regional and local German distributors. We first assess if our model confirms that large distributors operate under a different technology than the smaller network operators. This implies that unoserved factors are present and have to be taken into account within the nationwide benchmarking. Secondly, we analyze the level of technical efficiency of our sample of network operators. We derive that a single step latent class model is able to provide new insights to the future use of benchmarking approaches within the incentive regulation schemes and could serve

[2]Alternative econometric approaches have been proposed in the literature to improve benchmarking methodology. These new strategies are mainly based on panel data when companies have been observed over time, attempting to isolate the unobserved firm-specific factors that are not under managerial control from real inefficiencies, see Farsi et al. (2006).

[3]Other general empirical studies apply the two-stage approaches to decrease the probability of misspeification, see e.g. Grifell and Lovell (1997) for efficiency measurement of banks.

as additional control instrument for the regulator.

The paper is structured as follows: Section 2 presents the econometric specification and Section 3 the data. Section 4 summarizes the main empirical results of the distance functions estimation under different econometric assumption. Section 5 concludes.

2 Model specification

2.1 Distance function approach

Within a technical production setting, the majority of applied parametric efficiency analyses uses the production function to describe the underlying technology of different firms. Single output Cobb-Douglas or translog functional forms are most widely assumed but become critical, when firms produce more than one output (Coelli, 2000). Applied work has managed the issue by either aggregating the different outputs into a single index, or capturing multi-output production via estimating a multi-output cost frontier function.[4]

Another approach to model multi-output production is the concept of parametric distance functions (Coelli, 2000). This approach has been proposed by Shephard (1970) who derives a distance function representation of a multi-output technology as a primal alternative that requires no aggregation, price and cost information and behavioral assumption; see Coelli (2000) for a detailed description on the econometric estimation of the distance function representation.[5] We apply a parametric frontier input distance function to model the customers' supply and the physical amount of electricity delivered to final customers. The input distance function is defined on the input set as

$$d_i(x, y) = max\{\rho : (x/\rho) \, \epsilon \, L(y)\} \tag{1}$$

and considers how much the input vector x may be proportionally contracted by the scalar distance ρ with the output vector held fixed (Coelli, 2000).[6] $d_i(x, y)$ will assume a value greater than or equal to one if the input vector x is an element of the feasible input set $L(y)$. In addition, $d_i(x, y) = 1$ if it is located on the inner boundary of the

[4]Farsi et al. (2006) and Filippini and Wild (2001) analyze the Swiss electricity distribution sector; Burns and Weyman-Jones (1996) consider England and Wales, and Filippini et al. (2004) look at Slovenian distribution companies based on cost frontier functions.

[5]For a discussion on advantages and disadvantages of the use of distance functions see Coelli (2000), and Coelli and Perelman (2000).

[6]It is assumed that the technology satisfies the standard axioms: $d_i(x, y)$ is non-decreasing, positively linearly homogeneous and concave in x and increasing in y (Coelli, 2000; Färe and Primont, 1995).

input set.[7]

The translog functional form is widely used for the distance function approximation in empirical application due to its flexibility for econometric estimation. However, we employ the stricter Cobb Douglas functional form excluding the squared and cross terms of the exogenous regressors. A constant elasticity of substitution and constant scale properties are therefore assumed because we want to capture the parameter heterogeneity to define different technologies. The econometric models for SFA including parameter heterogeneity are already characterized by a sophisticated stochastic component. Moreover, parameter heterogeneity concerning the squared and cross terms of the regressors does not have a real economic interpretation. For the case of M outputs and K inputs the Cobb Douglas input distance function is specified for the i-th firm as

$$\ln d_i = \alpha_0 + \sum_{m=1}^{M} \gamma_m \ln y_m + \sum_{k=1}^{K} \beta_k \ln x_k. \tag{2}$$

To obtain the frontier surface (the transformation function) one would set $d_i = 1$, so the left side equals zero (Coelli and Perelman, 2000). The restriction for homogeneity of degree +1 in inputs is

$$\sum_{k=1}^{K} \beta_k = 1. \tag{3}$$

A convenient approach of imposing homogeneity constraints follows Coelli and Perelman (2000) considering that homogeneity implies that for any $w > 0$

$$d_i(wx, y) = wd_i(x, y). \tag{4}$$

Therefore, one of the inputs may be arbitrarily chosen, such as the K-th input and set $w = 1/x_K$. This yields

$$d_i(x/x_K, y) = d_i(x, y)/x_K \tag{5}$$

and the Cobb Douglas input distance function becomes

$$\ln x_K = \alpha_0 + \sum_{m=1}^{M} \gamma_m \ln y_m + \sum_{k=1}^{K} \beta_k \ln(\frac{x_k}{x_K}) - \ln d_i \tag{6}$$

by dividing equation (2) by an optional input and some rearranging; $\ln d_i$ is a non-negative variable which can be associated with technical inefficiency u_i. Given the

[7] Applications of this concept to estimate parametric distance functions using econometric methods can be found in Coelli and Perelman (2000) for railways, Färe et al. (1993) for electricity generation, Saal et al. (2007) for water and sewerage industry, and Growitsch et al. (forthcoming) for European electricity distribution.

stochastic error v_i this model can be formulated in the common SFA form with the combined error term $v_i - u_i$ (see Section 2.2). Technical efficiency is the ratio of observed output to frontier output. A radial input-oriented measure of technical efficiency is then obtained by

$$TE = \frac{1}{d_i} = exp(-u_i). \tag{7}$$

The distance function provides a promising new solution to the single output restriction of the standard production functions. One concern in the econometric estimation is potential regressor endogeneity which may introduce possible simultaneous equation bias.[8] Some authors have proposed instrumental variables estimation (see Atkinson and Primont, 2002). However, Coelli (2000) found that under an assumption of cost minimization behavior, distance functions do not face such bias and that OLS provides consistent estimates of the parameters of an input distance function. A second issue is that estimated input distance functions often fail to satisfy the concavity and quasi-concavity properties implied by economic theory. This sometimes leads to surprising conclusions regarding the effects of input and output changes on productivity growth and relative efficiency levels. Therefore, the starting point before any interpretation of inefficiencies is to check and to test for the properties.[9]

For the interpretation of the empirical estimates of a distance function it is important to keep in mind the duality between the cost and the input distance functions (Färe and Primont, 1995). For instance, the derivative of an input distance function with respect to a particular input is equal to the cost share of that input. This implies that the expected sign of the coefficients of the inputs should be positive. Moreover, the elasticity of an input distance function with respect to any output is equal to the negative value of the cost elasticity of that output. This implies that the sign of the coefficients of the outputs should be negative. Given that all the variables are in logarithmic form, these coefficients can be directly interpreted as elasticities.

2.2 Stochastic frontiers and latent classes

2.2.1 Panel data approaches for stochastic frontiers

SFA belongs to the parametric benchmarking methods. Contrary to the nonparametric approaches, a functional relationship for the underlying production technology is assumed. Within the benchmarking process we compare some measure of actual per-

[8]Ratios on inputs appear on the right side of the estimating equation which may involve simultaneous feedback problems because these input variables are assumed to be endogenous.

[9]Regularity conditions could also be imposed by estimating the model in a Bayesian framework (O'Donnell and Coelli, 2005).

formance against a reference technology (the stochastic frontier). The distance to the production frontier can be interpreted as a common measure of technical inefficiency. In the SFA framework the error term is divided into two uncorrelated components: a one-sided non-negative disturbance u_i, half-normally distributed, representing the inefficiency; and a symmetric disturbance v_i, assumed to be normally distributed, and capturing random noise in the sample (Greene, 2007). The most general cross-sectional formulation is

$$\begin{aligned} y_i &= \beta' x_i + v_i - u_i \\ u_i &\sim N^+[0, \sigma_u^2] \\ v_i &\sim N[0, \sigma_v^2] \end{aligned} \tag{8}$$

where x_i represents the set of explanatory variables and y_i the observed production of a firm. This model can be estimated using the maximum likelihood approaches.

A number of different stochastic frontier models for panel data have been proposed. The first models define the random or fixed effect as the inefficiency component meaning that the models deduce the efficiency estimates from the individual firm-specific effects.[10] These traditional models assume a common technology/frontier encompassing every sample observation. This may be inappropriate in the sense that the estimated technology is not likely to represent the "true" technology for all observations (Farsi et al., 2006). Thus, the estimate of the underlying technology may be biased. In addition, as unobserved heterogeneity was not accounted for in the econometric models, parameter estimates also may have been biased. Moreover, since all time-invariant heterogeneity was covered by the inefficiency part, these models have a tendency to underestimate firms' performance (Farsi et al., 2007).

European regulators, implementing the efficiency estimates into regulatory practice, are concerned about the robustness and the reliability of the empirical outcome of the individual efficiency estimation. Robust and consistent specification and models are indispensable for a trustable and effective regulation as the choice of the empirical approach has a strong impact on the financial situation of the network operators. Thus over the last decades we observed a research necessity concerning mainly two aspects for parametric stochastic efficiency analysis: first, how unobserved firm-specific factors may influence the underlying production process, and second how to model different technologies for different groups of firms.

With regard to the first aspect a wide range of newer models attempts to separate unobserved heterogeneity from inefficiency. One can model heterogeneity in the

[10]See Pitt and Lee (1981) for the random effects SFA model and Schmidt and Sickles (1984) for the fixed effects SFA model.

stochastic part, in the mean or the variance of the inefficiency distribution u_i.[11] However, it became more important to model both heterogeneity in the stochastic part and firm-specific heterogeneity in the production or cost function of the underlying production process. Kumbhakar (1991) and Greene (2005b) have suggested extending the original stochastic frontier model by adding an individual time-invariant random or fixed effect.[12] The basic underlying assumption is the existence of firm-specific and time-invariant factors that cannot be captured by environmental variables due to the variation of the latter over time and/or omitted variables. With the additional inclusion of heterogeneity terms by means of the random firm-specific effect α_i the model is expected to provide a finer distinction between inefficiency and other unexplained factors (Greene, 2005b).

With regard to the second aspect other formulations of the stochastic frontier model were proposed in the literature that allow not only the constant but the entire function to vary more generally across firms: the random parameter as well as latent class models for stochastic frontiers.[13] Instead of the continuous parameter variation, the latent class formulation can be interpreted as an approximation where the parameter variation is treated as generated by a discrete distribution (Greene, 2007). With these newer models unobserved differences in technologies may be accounted for which previously were inappropriately labeled as inefficiency. The unobserved firm-specific heterogeneity could therefore be applied to marginal products and costs represented by the coefficients of the production cost or distance function.

2.2.2 Latent class specification

The latent class framework for SFA accounts for specific technological characteristics of the observations in the sample; firms are classified into a set of different technologies and efficiency distributions. However the specific classification is à priori unknown (Greene, 2007; Kumbhakar and Orea, 2004), unlike the two-step approach that classifies the sample observations à priori into categories using exogenous sample separation information (Kumbhakar and Orea, 2004). In this model all parameters vary by class standing for the different technologies of the different classes. Others have modeled technology heterogeneity within SFA via the latent class formulation (Kumbhakar and

[11] The literature proposed to define a function of the mean or variance of observed variables (Battese and Coelli, 1995).

[12] These models are called "true" models because they include two stochastic terms for unobserved heterogeneity (one for the time-variant factors and one for the firm-specific constant characteristics (Farsi et al., 2006)).

[13] For applications of the random coefficient model for SFA see Tsionas (2002), Huang (2003).

Orea, 2004; Greene, 2005b; Caudill, 2003; Corral and Alvarez, 2008).[14] Figure 1 shows a sample of firms operating under different technologies (Technology A and Technology B). Assuming in this framework a common technology, and therefore frontier, for the companies would result in biased estimates of the distance function and efficiency. The

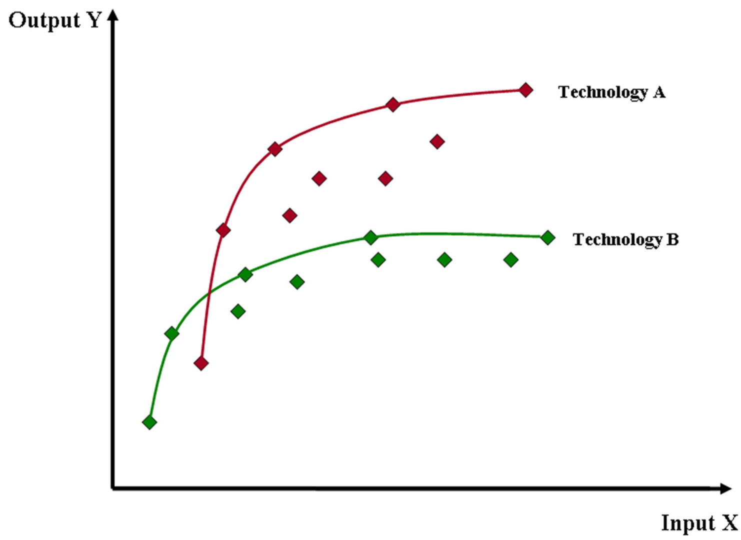

Figure 1: Graphical interpretation of latent classes

latent class model, in general, is a stochastic frontier model of the following form

$$\ln(y_{it}|_j) = f(x_{it}, \beta_j) + v_{it}|_j - u_{it}|_j \tag{9}$$

where j indicates the class or regime and J the total number of classes or regimes. Class membership is unknown. One assumes that there is a latent sorting of the observations in the data resulting in J classes (Greene, 2007). For one specific observation from class j the model is characterized by the conditional density $g(.)$ determined by the class specific parameter vector β_j.

$$g(y_{it}|x_{it}, class_j) = f(\beta_j, y_{it}, x_{it}) \tag{10}$$

The contribution of the company i to the conditional likelihood (conditional on class j) is

$$P(i|j) = \prod_{t=1}^{T} P(i, t|j) \tag{11}$$

[14]Kumbhakar and Orea (2004) extend the latent class model derived by Greene (2002) to the Battese and Coelli (1992, 1995) specification. Greene (2002) models the inefficiency term with a free variation over time; in the Kumbhakar and Orea (2004) specification, the inefficiency term varies systematically over time in a deterministic fashion (Greene, 2007).

The unconditional likelihood for individual i is an average over the J classes. It can be shown that the likelihood function can be expressed by (see Greene, 2005)

$$logLF = \sum_{i=1}^{N} log(\sum P_{ij} \prod_{t=1}^{T} LF_{ijt}) \tag{12}$$

The class probabilities can be parameterized by a multinomial logit model:

$$P_{ij} = \frac{\exp(\delta_j' q_i)}{\sum_{j=1}^{J} \exp(\delta_j' q_i)} \tag{13}$$

where q_i is a vector of firm-specific but time-invariant variables. These variables, called separating or switching variables, are included to identify any regularity in classifying the sample by means of the estimated coefficients of latent class probability functions $\widehat{\delta_j}$ (Greene, 2007). A positive sign of the coefficient suggests that the larger the variable the higher the probability that a firm belongs to this class. Similarly, the significantly negative value of a coefficient indicates that the probability of membership in this class decreases when the variable increases.

Under the maintained assumptions, maximum likelihood techniques will give asymptotically efficient estimates of all parameters. Greene (2002) points out that the technology as well as the probability to belong to a certain class are estimated simultaneously. All observations in the sample are used to estimate the underlying technology for each class. This can be viewed in opposition to the standard two-step approaches, where observations that are allocated to a specific class equal one, and zero for the others, are therefore excluded to estimate other class frontiers (Kumbhakar and Orea, 2004). The estimated parameters can be used to compute the conditional posterior class probabilities. In addition, Greene (2007) suggests that the class probabilities apply unchanged to all years of the observation period.

In standard SFA, the individual efficiency is estimated to the common frontier, since all firms are assumed to operate under the same technology. The latent class specification estimates as many frontiers as the number of classes (see Figure 1 with two different classes). There is no unique technology against which inefficiency is computed. There are different methods to measure the efficiency level of an individual firm (see Kumbhakar and Orea, 2004, for a summary): first, the highest posterior probability for class membership can be taken, and firms' inefficiency is computed using the frontier assigned for that class as its reference technology (most likely frontier is used); second, technologies from every class are taken into account, weighted with the respective probabilities.[15]

An ongoing discussion in the literature concerns the determination of the number of classes (Greene, 2007). For estimation we assume that the number of classes is known, but as Greene (2007) has shown, there is no reason to expect this. Using the likelihood ratio test from a J class model to a $J-1$ class model would lead to an ambiguous number of degrees of freedom. When we test up from $J-1$ to a J class model and the correct model has J classes, the $J-1$ class model is inconsistent. The empirical solution in the literature is to apply information criteria such as the Akaike Information criteria (Greene, 2007). Our empirical model defines two latent classes (see Section 4.1).

3 Data

Two sets of variables are required to estimate the latent class model. First, the variables in the production frontier model have to be defined including appropriate inputs and outputs for the process of electricity distribution. Second, the variables in the class probabilities have to be chosen to determine if observable information helps to classify the companies into different classes with different underlying technologies. The variables to describe the underlying production process are defined in the same manner for the different groups of companies. In the empirical benchmarking literature, a variety of specifications is used depending on what is being investigated. The choice of variables for input and output to describe the underlying production process and technology must account for the international experience with electricity distribution benchmarking (Cullmann and Hirschhausen, 2008a,b). Further, it is constraint by data availability. In this respect, Germany has to be ranked among the least developed for data collection. We therefore depend upon a limited data set of physical and technical data. We must define a simple production process describing the basic input transformation of the German companies. As a result we limit ourselves to the estimation of technical inefficiency which does not require any data on costs or prices; however as shown in Section 2, conclusions regarding cost efficiency can be drawn due to the distance function specification.

The sample includes 200 companies and covers a five-year observation period from 2001 to 2005. We estimate a base model using the traditional input variables (labor and grid size) and the outputs are units sold and the number of customers:

[15] This is the strategy suggested by Greene (2002) to find firm-specific estimates of the parameters of the stochastic frontier model. The magnitude of the difference depends on the relative importance of the posterior probability of the most likely cost frontier; the higher the posterior probability, the smaller the differences.

* Labor input (x_L) is estimated by the number of workers. We are aware of the criticism of this choice of variable due to the potentially distorting effect of outsourcing: a utility can improve its efficiency simply by switching from in-house production to outsourcing. Some of the utilities have their own generating plants and we only dispose of employment data covering all workers in the electricity utility. To get an approximation of workers employed in electricity distribution, we subtract one employee for each 20 GWh electricity produced (following Auer, 2002).

* Capital input (x_{NL}) is approximated by the length of the existing electricity cables and lines. We differentiate between voltage levels (high, medium and low voltage) by introducing a cost factor for each type of line following standard practice used by the German network association (VDN): factor 5 for high voltage, 1.6 for medium voltage, and 1 for low voltage cables.

* Delivery (y_D) is defined as the annual amount of electricity sold to all final customers (household, industrial and others) in MWh.

* Customers (y_C) is defined as the sum of industrial, households and other customers.

* Year dummies ($d1 - d4$) 2001 to 2004.

We also include variables as determinants of the latent class probabilities in order to analyze whether they deliver useful information in classifying the sample:

* Delivery to other distribution companies (dummy)

* Electricity generating activities (dummy)

* Operating high voltage cables (dummy)

* Operating high voltage aerial lines (dummy)

* Annual investment in 1000 Euro

* Annual revenue out of domestic sales in 1000 Euro

* Delivery to households/total delivery in MWh

* Investment per km network in 1000 Euro/km

* Revenue per unit delivery in 1000 Euro/MWh

* Share of cables in total network in km
* Losses in MWh
* Density of inhabitants per km operation area.

For the q variables we use firm average values over the five-year observation period. To summarize, we designed a model to describe the production process, including two input variables, two output variables and year dummies to capture the time dimension. We define labor (x_L) as the *numeraire* input. By dividing the remaining input over the labor input and rearranging we define the following Cobb-Douglas input frontier distance function (see Section 2)

$$\begin{aligned} -\ln x_{it,L} &= \alpha_0 + \alpha_{NW} \ln(\frac{x_{it,NW}}{x_{it,L}}) + \gamma_C \ln(y_{it,C}) + \gamma_D \ln(y_{it,D}) \\ &+ d1 + d2 + d3 + d4 - \ln d_{it} \end{aligned} \quad (14)$$

with

$$\ln d_{it} = \varepsilon_{it} = v_{it} - u_{it} \quad (15)$$

4 Empirical results

4.1 Estimation results for latent classes

To test the hypothesis if larger and smaller network operators in Germany operate under different technologoes we estimate a latent class model with two different classes to model parameter heterogeneity. Within this framework we allow firms to have different underlying production technologies, caused by unobserved technological network differences or variations in the customer structure. Estimating a latent class model requires the a priori determination of the number of classes (see Section 2.2). Following Kumbhakar and Orea (2004) and Greene (2007) we applied the Akaike information criteria which favors the definition of two classes in contrast to one single class. The empirical results with more than two classes (up to four) did not converge to a meaningful solution within the maximum likelihood estimation, therfore we stayed in the framwork of two different classes. This is in our view appropriate to test our hypotheses of different technologies for small and large distributors.

Table 1 shows the regression results of the distance function estimation for the two differnt classes. All variables are median-corrected to avoid outliers in the sample having a large impact on the estimation outcome. The estimated coefficients of the first-order terms have the expected signs and are statistically significant (see Section 2.1). Thus the estimated distance function appears to reasonably fit the observed data. The prior class probabilities show a quite equal latent sorting of the observations into both classes with a slightly higher amount of companies belonging to the first class: 57% in Class 1 and 43% in Class 2. The characteristics of both classes are shown in Table 2.

We start by characterizing both classes and calculate summary statistics of important physical and technical data of the companies differentiated between the two classes (see Table 2). We can derive one clear trend: Class 1 incorporates larger distribution companies with a higher number of employees, higher amount of delivery sold, more final customers and a larger network; Class 2 seems to include the smaller German distributors. With regard to the number of final customers we note that the separation reflects approximately the classification in the ordinance with 30.000 customers connected (AregV §24). This separation already gives a first insight that the size of the network operators matters in identifying the technology.

Next, we want to determine whether the production structures of both classes (larger firms vs. smaller) differ and may be characterized by parameter heterogeneity. Table 1 shows that the year dummies in both classes are insignificant which suggests no important technology shifts within the observation period. In both classes the input and output coefficients have the expected signs and are all significant. The coefficients of first-order output variables represent the cost elasticities with respect to the corresponding outputs. The coefficients of first-order input variables show the cost shares of the respective inputs. However, we note that the coefficients differ significantly for the two classes: Class 1 is characterized by higher capital intensity with a coefficient of 0.81 vs. Class 2 of 0.759. Larger firms operating larger networks and in particular more cost intentsive high voltage networks (up to 110 kV) and are therefore characterized by a more capital intensive distribution.[16]

For an input distance function representation the elasticity of scale (RTS) is measured by the negative of the inverse of the sum of the output elasticities (Färe and Primont, 1995). As the output weights do not sum to unity this can be interpreted as reflecting the effect of non-constant returns to scale (Saal et al., 2007). The sum of the coefficients of the two output variables varies (1.033 for Class 1 and 0.51 for Class 2). This result suggests the presence of important increasing returns to scale for Class 2 while Class 1, the larger companies, operates closely under constant returns to scale. We notice that the output elasticities with respect to customers differ sig-

[16]The homogeneity of degree one assumption involves that the input coefficients sum up to one. Thus we obtain a labor share of 0.19 and 0.24 respectively. The electricity distribution sector is obviously characterized by a high capital cost share.

nificantly between the two classes (-0.458 in Class 1 vs. -0.132 in Class 2) indicating that connecting customers is more cost intesive for Class 1. This can be explained by the fact that larger network operators have a higher share of industrial customers for which connection to the network is more costly than for household customers.

Clearly, the empirical evidence shows that the two groups operate under different technologies. The latent class specification leads to different technological production frontiers as references for the different companies. Estimating a common frontier without modeling the parameter heterogeneity would produce biased estimates and therefore inconsistent individual efficiency measures. The hypothesis that larger and smaller network operators are characterized by different technologies can be confirmed. The latent class model is able to estimate the technologies and the class probabilities simultaneously. In contrast to the two-stage approach, all observations of the sample are used to determine the underlying technologies for each class. This overcomes the implicit restriction of the two-stage approach which precludes using observations that were allocated to one subgroup to determine the efficiency frontiers of the other groups (Kumbhakar and Orea, 2004). Thus it represents a promising alternative to the two-step procedure.

4.2 Classification of the sample

Table 1 also shows the estimated coefficients of latent class probability functions $\widehat{\delta}$ together with their respective p-value. Thus, the class probabilities are not fixed but dependent on time-invariant observable characteristics of the German distribution companies. Including these variables reveals whether they deliver useful information in classifying the sample, more precisely if they provide information on the probability of a distribution company belonging to a certain class.

The empirical results suggest that the supply to other electricity distribution companies, the generation activity, the operation of high voltage cables, the percentage share of household customers in the total sum and the location in East or West Germany do not have significant impacts on the probability of belonging to a certain class. The p-value of these estimated coefficients is lower than the critical value of 0.05. However, we note that the remaining observable characteristics have an impact on classifying the sample into larger vs. smaller operators. The density has a positive impact on the probability belonging to Class 1. In other words, Class 1 consists of larger companies operating in more densely settled urban areas. Average losses of electricity (a type of quality index) lowers the probability of being in Class 1, which indicates that the larger companies are characterized by higher quality standards, i.e electricity losses.[17]

[17] This may also be explained with the higher voltage levels that seem to prevail in Class 1: electricity

Variables	Coefficient	Standard error	t-stat	p-value
Model parameters for latent class 1				
Constant	0.331	0.038	8.683	0.000
Lnx1 network	0.831	0.015	54.360	0.000
Lny1 delivery	-0.575	0.027	-21.434	0.000
Lny2 customers	-0.458	0.029	-15.899	0.000
Year-Dummy 2001	0.010	0.031	0.323	0.747
Year-Dummy 2002	-0.012	0.031	-0.393	0.695
Year-Dummy 2003	-0.017	0.031	-0.559	0.576
Year-Dummy 2004	-0.009	0.030	-0.285	0.776
Sigma	0.307	0.022	13.946	0.000
Lambda	1.473	0.331	4.448	0.000
Model parameters for latent class 2				
Constant	0.014	0.084	0.160	0.873
Lnx1 network	0.759	0.027	28.443	0.000
Lny1 delivery	-0.606	0.020	-30.119	0.000
Lny2 customers	-0.132	0.022	-6.072	0.000
Year-Dummy 2001	0.000	0.042	-0.009	0.993
Year-Dummy 2002	-0.037	0.042	-0.863	0.388
Year-Dummy 2003	-0.017	0.042	-0.414	0.679
Year-Dummy 2004	-0.006	0.042	-0.135	0.893
Sigma	0.321	0.040	8.072	0.000
Lambda	0.833	0.484	1.719	0.086
Estimated prior probabilities for class membership				
Constant	0.980	0.656	1.493	0.136
Dummy generation	-0.434	0.460	-0.944	0.345
Dummy EDC	-0.186	0.440	-0.422	0.673
Dummy high voltage cable	2.331	1.346	1.732	0.083
Dummy high voltage lines	-2.534	1.048	-2.417	0.016
Dummy West/East	-0.958	0.549	-1.743	0.081
Average Investment	-3.465	0.652	-5.315	0.000
Average Revenue	3.923	0.639	6.137	0.000
Ratio delivery household/total delivery	1.003	0.646	1.553	0.120
Investment per network	3.289	0.653	5.039	0.000
Revenue per delivery	-4.221	0.007	-583.532	0.000
Cable per network	-0.989	0.010	-99.792	0.000
Average losses	-0.572	0.007	-85.981	0.000
Density	0.184	0.008	23.590	0.000
Prior class probabilities at data means for LCM variables				
Class 1	0.571			
Class 2	0.429			
Stochastic frontier model variance parameters				
	Lambda	Sigma	Sigma(u)	Sigma(v)
Class 1	1.473	0.307	0.254	0.172
Class 2	0.833	0.321	0.206	0.247

Table 1: Estimation results of Model 2 (latent class specification)

Variable	Mean	Standard deviation	Confidence	Interval 95%
		Number of employees		
Class 1	126	14	97	154
Class 2	74	8	59	90
		Delivery in MWh		
Class 1	626977	85713	458760	795195
Class 2	312335	50479	213268	411402
		Final customers		
Class 1	54641	6243	42388	66893
Class 2	31751	3142	25584	37917
		Weighted km of lines		
Class 1	822	168	493	1151
Class 2	97	8	81	113
		Weighted km of cables		
Class 1	2188	408	1387	2990
Class 2	752	58	637	866
		Weighted km of network		
Class 1	3010	504	2022	3998
Class 2	848	61	729	968
		Density		
Class 1	3368	63	3244	3491
Class 2	3142	73	2999	3285
		Unweighted km of lines		
Class 1	486	89	312	659
Class 2	75	6	63	87
		Unweighted km of cables		
Class 1	1860	354	1166	2554
Class 2	618	45	530	705
		Unweighted km of network		
Class 1	2345	403	1553	3137
Class 2	693	46	603	783
		Delivery inland in MWh		
Class 1	622306	85534	454440	790171
Class 2	308665	49903	210728	406602
		Delivery to industry in MWh		
Class 1	198341	27888	143610	253073
Class 2	84766	13213	58836	110697
		Delivery to households in MWh		
Class 1	165312	20786	124518	206105
Class 2	72113	6595	59170	85055

Table 2: Sample statistics of the two latent classes

A higher average investment per network length increases the probability of being in the class of larger distributors. Thus, we conclude that larger companies on average invest more per km network.

Considering the average investment separately, without relating it to the network length, we obtain a negative sign indicating that the probability of being in Class 1 decreases. This would appear to contradict the previous result. However, the relation to the capital input defined in our analysis by the length of the network indicates more reliable results than considering it separately. The same argument applies to revenue, included first as a separate variable and then in relation to the units delivered. We argue that the revenue per unit delivered shows a more reliable picture; here we obtain a negative coefficient.[18] This empirical result is unexpected as it suggests that higher revenue per unit electricity delivered decreases the probability of being in Class 1. Smaller distribution operators therefore are characterized by higher revenues per units. However we can argue that small local distributors (Stadtwerke) might be characterized by a higher cross-subsidization.[19]

The latent class estimation provides empirical evidence that on the one hand we have to consider different technologies for different classes. We can also explain the classification of companies by observable characteristics that provide more sophisticated information about the groups. These are important for correctly estimating the true technology frontier for efficiency analysis.

4.3 Efficiency analysis

The sample statistics for the estimated efficiencies for the whole sample and for each estimated class are shown in Table 3. The values lie between 0 and 1, with no company showing full efficiency. The values of the efficiency vary from 0.647 to 0.978. The values of the mean technical efficiency indices are relatively high: 0.91. The high average efficiency is conform with the mean efficiency calcualted by the German regulator for the German network operators. We observe a difference of the performance levels in the latent classes (0.90 vs. 0.92). In our sample it appears that the smaller distribution companies are operating under increasing returns to scale, but from a pure technical efficiency perspective show a higher performance compared to larger distributors.

losses are inversely related to voltage levels.

[18]The correlation of the four variables is very low (0.161 for average investment and investment per network; -0.277 for average revenue and revenue per unit delivered); therefore we can explain the different coefficients. This also ensures that we do not have any muti-collinearity problems including all variables as explanatory factors.

[19]When we consider the average revenue separately it indicates that higher revenues are related to larger operators.

	Class	Number of observation	Mean	Std.dev	Min	Max
Latent class model		1000	0.910	0.041	0.647	0.978
Latent class model	Class 1	535	0.9	0.002	0.647	0.978
	Class 2	465	0.921	0.001	0.791	0.967

Table 3: Descriptive statistics of efficiency estimates ordered by classes

			Model 2	Statistically more efficient class
Class 1 535 observations	vs	Class 2 465 observations	0.0001	can be confirmed

Table 4: Kruskal-Wallis test

This empirical evidence shows that accounting for the true frontier in each group is important for the benchmarking process.[20]

The difference between the classes can be confirmed statistically by means of the Kruskal-Wallis Test, testing the hypothesis that several samples are from the same population. The results of the Kruskal-Wallis test are shown in Table 4. The p-value is 0.0001. These results indicate that we can reject the hypotheses of equal distribution. This again leads us to conclude that when assuming different technologies by capturing parameter heterogeneity in the econometric model we obtain more robust results for the individual efficiency estimates. This is due to the fact that we can adapt better the technology and therefore the production frontier to different classes of firms with different characteristics.

[20]For comparison reasons we also estimated the true random effects model (see section 2.2). Within this framework all companies are benchmarked against the same technology (apart from the technology neutral shift captured by the individual specific randomly distributed constant). For this model specification we do not note any significant difference in the average efficiency in both classes. From a descriptive perspective we observe that within the latent class model specification some of the technological heterogeneity captured is labeled as inefficiency in the true random effects model. Accounting for unobserved factors not only in the intercept (technological neutral shift) but in different technologies and therefore frontiers produces other conclusions about firms' performance.

5 Conclusions

In this paper we analyzed the technical efficiency level for a sample of 200 German electricity distribution companies subject to incentive regulation since January 2009. The new regulatory instruments are based on benchmarking procedures to determine the revenue caps of the individual companies. In the empirical application of benchmarking, regulators and researchers are always faced with the problem of a high degree of heterogeneity for environmental or network characteristics. Only some of this heterogeneity is observed and can therefore be accounted for in the econometric model. Another part will be unobserved. These unobserved characteristics between firms, that are not measured in the sample might influence the underlying production process. Therefore, the problem becomes one of modeling unobserved heterogeneity.

Comparing the efficiency of different firms usually assumes that they operate under the same production technology and therfore these unobserved factors might be understood as inefficiency. To avoid such types of misspecifications we observe that in regulatory practice estimation is often carried out in two stages. First, observations are classified into several groups assuming a priori that they operate under different technologies. Then in a second step, separate analyses are conducted for each sub-group of the sample. This paper shows how to disentangle the heterogeneity from inefficiency in one step, using a latent class model for stochastic frontiers. Within this framework the classification is not based on a priori sample separation criteria and therfore delivers more robust and statistical significant and testable results. In the latent class model the unobserved firm-specific heterogeneity is accounted for by parameter heterogeneity, identifying different technologies for companies. We show that this model will partially solve the unobserved heterogeneity problem in measuring the technical efficiency. The empirical results suggest that our proposed model is able to account for the fact that larger distributors operate under a different technology than smaller companies and that different frontiers are necessary to obtain more robust and reliable efficiency estimates. It represents a promising alternative to the traditional two step procedures.

We find that the estimated Cobb-Douglas distance function is a reasonable fit to the observed data and that the estimated input and output elasticities have the correct sign and magnitude for both classes. Determining the returns to scale we observe that the latent class specification can differentiate between large and smaller distribution companies: in the latent class model the estimated coefficients indicate that the larger distributors operate under constant returns to scale and the smaller firms under increasing returns to scale.

In addition, we test the differences in inefficiency scores between the two classes via a Kruskal-Wallis test. The results underline the importance of modeling and estimating two classes. The latent class model can be helpful in distinguishing unobserved heterogeneity in technologies from inefficiency estimates. The results can be used as an additional instrument to reduce the information asymmetry between the regulator and regulated companies.

6 Acknowledgments

I am indebted to Christian von Hirschhausen, Matthias Walter, Borge Hess and the DIW IO seminar for their helpful comments and support.

References

Atkinson, S. E. and Primont, D. (2002). Stochastic estimation of firm technology, inefficiency and productivity growth using shadow cost and distance functions. *Journal of Econometrics*, 108(2):203–225.

Auer, H. (2002). Benchmarking und Regulierung elektrischer Netze in liberalisierten Strommärkten: Grundlagen, internationale Erfahrungen und Anwendung auf Österreich. Technische Universität Wien, Institut für Elektrische Anlagen und Energiewirtschaft.

Battese, G. E. and Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis*, 3(1):153–169.

Battese, G. E. and Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20:325–332.

Beesley, M. and Littlechild, S. (1989). The regulation of privatized monpolies in the United Kingdom. *Rand Journal of Economics*, 20(3):454–472.

Burns, P. and Weyman-Jones, T. G. (1996). Cost functions and cost efficiency in electricity distribution: a Stochastic Frontier approach. *Bulletin of Economics Research*, 48(1):41–64.

Caudill, S. B. (2003). Estimating a mixture of stochastic frontier regression models via the EM algorithm: A multiproduct cost function application. *Empirical Economics*, 28(3):581–598.

Coelli, T. J. (2000). On the econometric estimation of the distance function representation of a production technology. Working Papers, Université Catholique de Louvain - Center for Operations Research and Economics, Louvain-la-Neuve, Belgium.

Coelli, T. J. and Perelman, S. (2000). Technical efficiency of European railways: A distance function approach. *Applied Economics*, 32(15):1967–76.

Corral, J. and Alvarez, A. (2008). Estimation of different technologies using a latent class model. Working Paper, Universitiy of Oviedo, Economics Departement, Oviedo, Spain.

Cullmann, A. and Hirschhausen, C. v. (2008a). Efficiency analysis of East European electricity distribution in transition: legacy of the past? *Journal of Productivity Analysis*, 29(2):155–167.

Cullmann, A. and Hirschhausen, C. v. (2008b). From transition to competition - dynamic efficiency analysis of Polish electricity distribution. *Economics of Transition*, 16(2):335–357.

Färe, R., Grosskopf, S., Lovell, C., and Yaisawarng, S. (1993). Derivation of shadow prices for undesirable outputs: a distance function approach. *The Review of Economics and Statistics*, 75(2):374–380.

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

Farsi, M., Fetz, A., and Filippini, M. (2007). Benchmarking and regulation in the electricity distribution sector. CEPE Working Paper No. 54, ETH Zürich, Zurich, Switzerland.

Farsi, M. and Filippini, M. (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., Filippini, M., and Greene, W. H. (2006). Application of panel data models in benchmarking analysis of the electricity distribution sector. *Annals of Public and Cooperative Economics*, 77(3):271–290.

Filippini, M., Hrovatin, N., and Zoric, J. (2004). Regulation of the Slovenian electricity distribution companies. *Energy Policy*, 32:335–344.

Filippini, M. and Wild, J. (2001). Regional differences in electricity distribution costs and their consequences for yardstick regulation of access prices. *Energy Economics*, 23(4):477–88.

Greene, W. H. (2002). Alternative panel data estimators for stochastic frontier models. Working Paper, Stern School of Business, New York University.

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

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

Greene, W. H. (2007). The econometric approach to efficiency measurement. In Fried, H., Lovell, C. K., and Schmidt, S., editors, *The Measurement of Productive Efficiency*. Oxford University Press, Oxford.

Grifell, E. and Lovell, C. A. K. (1997). The sources of productivity change in Spanish banking. *European Journal of Operations Research*, 98:364–380.

Growitsch, C., Jamasb, T., and Pollit, M. G. Quality of service, efficiency, and scale in network industries: an analysis of European electricity distribution. *Applied Economics*. forthcoming.

Huang, H. C. (2003). Estimation of technical inefficiencies with heterogeneous technologies. *Journal of Productivity Analysis*, 21(3):277–296.

Jamasb, T. and Pollitt, M. G. (2001). Benchmarking and regulation: international electricity experience. *Utilities Policy*, 9(3):107–130.

Joskow, P. J. (2006). Regulation of natural monopolies. In Polinsky, A. M. and Shavell, S., editors, *Handbook of Law and Economics*. North-Holland, Amsterdam.

Kumbhakar, S. C. (1991). Estimation of technical inefficiency in panel data models with firm and time-specific effects. *Economics Letters*, 36(1):43–48.

Kumbhakar, S. C. and Orea, L. (2004). Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics*, 29(1):169–183.

O'Donnell, C. J. and Coelli, T. J. (2005). A Bayesian approach to imposing curvature on distance functions. *Journal of Econometrics*, 126(2):493–523.

Pitt, M. and Lee, L. (1981). The measurement of sources of technical inefficiency in Indonesian weaving industry. *Journal of Development Economics*, 9(1):43–64.

Saal, D. S., Parker, D., and Weyman-Jones, T. G. (2007). Determining the contribution of technical change, efficiency change and scale change to productivity growth in the privatized English and Welsh water and sewerage industry: 1985-2000. *Journal of Productivity Analysis*, 28(1):127–139.

Schmidt, P. and Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business and Economics Statistics*, 2(4):367–374.

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

Shuttleworth, G. (2005). Benchmarking of electricity networks: practical problems with its use for regulation. *Utilities Policy*, 13(3):310–317.

Tsionas, E. (2002). Stochastic frontier models with random coefficients. *Journal of Applied Econometrics*, 17(2):127–147.