Cost Efficiency Analysis of Electricity Distribution Sector under Model Uncertainty

Kamil Makieła and Jacek Osiewalski***

ABSTRACT

This paper discusses a Bayesian approach to analyzing cost efficiency of Distribution System Operators when model specification and variable selection are difficult to determine. Bayesian model selection and inference pooling techniques are adopted in a stochastic frontier analysis to mitigate the problem of model uncertainty. Adequacy of a given specification is judged by its posterior probability, which makes the benchmarking process not only more transparent but also much more objective. The proposed methodology is applied to one of Polish Distribution System Operators. We find that variable selection plays an important role and models, which are the best at describing the data, are rather parsimonious. They rely on just a few variables determining the observed cost. However, these models also show relatively high average efficiency scores among analyzed objects.

Keywords: Cost efficiency, Electricity distribution, Bayesian inference, Stochastic frontier analysis

https://doi.org/10.5547/01956574.39.4.kmak

1. INTRODUCTION

Electricity distribution is a key sector of every economy. In most countries, distribution system operators of electricity networks (DSOs hereafter) are natural, usually territorial, monopolies. They are either government-owned or government-controlled in some way. As such they are expected to be held accountable for efficient distribution and provision of electricity for the sake of the public good. However, being a monopoly does not go along with being efficient. As Hicks (1935; p. 8) eloquently pointed out many years ago "the best of all monopoly profits is a quiet life" and efficiency perishes once a company becomes a monopoly. For this reason, national energy regulatory agencies across the world try the "stick and carrot" approach to get DSOs to improve their performances. Electricity market reforms, which establish performance oriented regulations, have been especially active in Europe due to European integration (see, e.g., Jamasb and Pollitt, 2005; Finon and Roques, 2013). For example, in Finland, the Energy Market Authority sets improvement targets for DSOs (see Kuosmanen, 2012); Norwegian Water Resources and Energy Directorate is managing a revenue-cap system for DSOs in Norway (see, e.g., Kumbhakar and Lien, 2017); Swedish Energy Agency, German *Bundesnetzagentur* and Polish Energy Regulatory Office also adopt performance oriented regulations (see, e.g., Agrell et al. 2005; Agrell and Bogetoft, 2007; Energy Regulatory

* Corresponding author. Department of Econometrics and Operations Research, Cracow University of Economics, Rakowicka 27, Kraków, Poland. E-mail: kamil.makiela@uek.krakow.pl.

** Department of Econometrics and Operations Research, Cracow University of Economics, Rakowicka 27, Kraków, Poland. E-mail: eeosiewa@cyf-kr.edu.pl.

Office, 2015b). These institutions regulate most of Central and Northern European electricity distribution markets and their regulations are based on official performance benchmarks.

In order to set up a credible regulatory policy, the regulator first needs to know which DSOs are performance leaders and which need to be "persuaded" to act more efficiently. This calls for an objective and transparent DSO benchmarking framework, and one should stress that due to the "stick and carrot" approach, objectivity is of crucial importance here. For example, 76 DSOs in Finland filed appeals with the Market Court against the framework put forward by the Finish Energy Market Authority (see Finnish Energy Market Authority, 2016; p.22). Cases of failed regulation models have also been documented, e.g., in Sweden (Agrell and Grifell-Tatje, 2016), the Netherlands (Nillesen and Pollitt, 2007) and Belgium (Agrell and Teusch, 2015). This has occurred because performance oriented regulations can translate to considerable financial gains and losses. That is why the DSOs require credible and objective assessment. All relevant aspects of their business conduct should be taken into consideration in the final model. Otherwise, the study may be biased towards certain profiles of a DSO operation, and thus the outcome may be discriminative and even potentially harmful for the energy distribution industry as a whole.

In order to build a credible and objective benchmarking-type study of economic efficiency several aspects need to be considered (Bogetoft, 2012; Coelli et al. 2003). We can summarize them as: (i) the benchmark framework (modelling framework), (ii) estimation technique and (iii) exact model specification, exact choice of variables in particular. All these aspects are intertwined with each other and should not be considered entirely separately or in any particular order. However, for simplicity we can outline the following three steps. First, one needs to set up a general benchmarking framework. For example, in order to study efficiency of Polish DSOs, the Polish Energy Regulatory Office (ERO hereafter) has decided on a cost minimization criterion based on cost functions. These constructs have sound theoretical bases well established in microeconomics and are often used in such studies (see, e.g., Farsi and Filippini, 2004, 2009; Farsi et al., 2006; Kopsakangas-Savolainen and Svento, 2008; Cullmann and von Hirschhausen, 2008b; Osiewalski and Wróbel-Rotter, 2008–9, 2012; Filippini et al., 2016; Dimitropoulos and Yatchew, 2017). The framework sets a common, intuitive goal for all DSOs to achieve, i.e., cost-effectiveness and it is also used it in this paper.[1]

Second, the regulator needs to decide on the estimation approach. In the case of Poland, which is the focus of the empirical analysis in this paper, Bayesian approach to Stochastic Frontier Analysis (SFA) has been chosen by ERO (Polish Energy Regulatory Office, 2015a; Osiewalski and Wróbel-Rotter, 2008–9, 2012). Of course, there are many other approaches. For example, to study Polish electricity distribution sector Cullmann and Hirschhausen (2008ab) also use Data Envelopment Analysis (DEA), Free Disposal Hull and non-Bayesian SFA. Dimitropoulos and Yatchew (2017) use index-based and econometric cost-based methods to study electricity distributors in Ontario. Kumbhakar and Lien (2017) use input distance functions for DSOs in Norway. Kuosmanen (2012) uses StoNED to study DSOs in Finland.[2] An extensive survey of contemporary benchmarking methods (not necessarily related to cost efficiency) can be found in, e.g., Agrell and Bogetoft (2009),

1. Cost minimization is an established microeconomic criterion for analyzing economic performance of units with multiple inputs-outputs. It requires mutually comparable information about costs, which sometimes can be problematic if the DSOs have different accounting standards (e.g., in international studies). Of course the regulator may choose a different criterion. E.g., concurrently with cost analysis ERO also uses the *volume balance difference* framework to set tariffs. Other studies would rely on production function models. The approach presented in the paper is general and may also be used in other studies of efficiency.

2. Other applications of frontier models in energy economics include, e.g., Filippini and Hunt (2011, 2012) and Llorca et al. (2017) who analyze energy efficiency using frontier demand functions and Makridou et al. (2015) who advocate a two-stage approach based on DEA and multi-criteria decision aiding (MCDA).

Bogetoft (2012) and Llorca et al. (2016). Bayesian approach to SFA, however, has some advantages that we exploit in this paper. In particular, it allows us to formally compare (or pool inference from) competing models and obtain exact small sample results. These two features, distinctive to Bayesian inference, are especially appealing in the case of an electricity distribution sector because: (i) model specification, especially variable choice, may have an impact on the results (see, e.g., Jamasb and Pollitt, 2001, 2003; Kuosmanen, 2012); (ii) in some countries, like Poland, the number of DSOs can be small (due to mergers and acquisitions over the years there are currently only five DSOs).

Third, the regulator needs to determine the exact model specification.[3] In particular, the regulator needs to know which factors determine the cost and should be placed in the cost function as explanatory variables. That is, which characteristics of DSOs operation best determine the costs they generate. Since Polish DSOs are assessed via their operational costs (OPEX), we can rely on the theory of a short-run variable cost function to some extent. It points out three sources of variable costs: (i) scale of production, (ii) level of capital assets and (iii) prices of variable production factors if such can be determined.[4] This is, however, as far as the theory can guide us. Exact choice of explanatory variables (i.e., exact model choice) is up to the researcher to decide, and this aspect may have a considerable impact on the benchmarking outcome due to model selection uncertainty.

In the first official cost efficiency studies of Polish DSOs, commissioned by ERO, Osiewalski and Wróbel-Rotter (2008–9, 2012), considered as many as 31 potential explanatory variables delivered by ERO. These variables characterize the scale of economic activity and technical situation of the DSOs. Naturally, it is unreasonable to “throw” all variables into one statistical model and there are more than two billion different model configurations to choose from ($2^K = 2^{31}$, where K is the number of potential explanatory variables). Furthermore, Osiewalski and Wróbel-Rotter (2008–9) have found that different choices of explanatory variables (thus different models) can lead to different results of the cost efficiency benchmark. Eventually, a selection of seven model variants has been chosen with a conclusion that the “third step” requires further research in order to bring more objectivity and transparency. The literature of the subject has addressed many issues relevant to DSO regulation, like possible frontier heterogeneity (e.g., Farsi et al., 2006; Filippini et al. 2016; Kumbhakar and Lien, 2017) or comparison of different modelling techniques (e.g., von Hirschhausen et al. 2006; Cullmann and Hirschhausen, 2008ab; Kopsakangas-Savolainen and Svento, 2008; Filippini et al. 2016). Unfortunately, although the discussion about variable choice and its impact on the results is reoccurring (Jamasb and Pollitt, 2001, 2003; Kuosmanen, 2012; Llorca et al. 2016), the authors are not aware of any DSO study that would formally and objectively address the problem of model selection uncertainty. As a result, model specification, especially regarding the choice of explanatory variables, is usually fixed throughout the study and not disputed or challenged in any formal way. Therefore uncertainty related to model selection is ignored.

3. To some extent this is dictated by the estimation technique. E.g., non-parametric methods require fewer assumptions about the functional form of the model than parametric methods. This, however, comes at the cost of considerably larger data requirements; see Simar and Wilson (2008). Furthermore, Bayesian techniques used in this study require us to specify probability distributions for all unknown quantities of the model, not just latent variables. The reader should know, however, that unless these prior distributions are unrealistic and strongly against information in the data their impact on the results is usually negligible.

4. Prices of variable factors, like price of labor, can be particularly troublesome to obtain if DSOs are territorial monopolies. E.g., for DSOs in Poland, which are territorial monopolies there is no industry-average salary in their respective regions to consider—they are the industry. In such cases the regulator may want to *a priori* exclude price information from the model to prevent potential rent extraction from the management. This way DSOs which are paying too high wages (relative to others) can be penalized by the regulator.

Bayesian inference has long known techniques required to mitigate this problem; see Osiewalski and Steel (1993). Bayesian inference pooling, more commonly known today as Bayesian model averaging (BMA), is a statistical method that allows us to formally pool inference about any quantity of interest (e.g., efficiency scores, model parameters, or their functions) from different models. Alternatively, if one is interested in choosing only one particular model among many, a closely related Bayesian model choice (or selection, BMS) can be performed. Bayesian model selection or averaging, however, requires us to know marginal data density values for all models under consideration. Since Bayesian estimation of SFA models is quite complex (due to latent variables) this is an especially challenging task. Fortunately, recent developments by Pajor (2017) have made it possible to precisely estimate the required marginal data densities in at least some SFA models. This makes Bayesian model comparison (model selection or inference pooling) a tangible option to consider when analyzing cost efficiency of DSOs.

Our contribution to the literature is twofold. First, we propose a fair, credible benchmarking process, in which validity of a given model specification, and thus its contribution to the joint results, is formally judged based on information in the data. This is particularly appealing for the electricity distribution sector, where policy makers (and "takers") have much at stake and possibly conflicting interests. Generality of our approach is also an added value. It can be easily applied to other benchmarking criteria (e.g., technical efficiency) as well as other services utilities where similar regulatory needs and conflicting interests arise (e.g., distribution of natural gas). Second, we contribute to the discussion about the use of SFA for benchmarking DSOs. Cost efficiency analysis in the electricity distribution sector is quite dependent on the model choice, and bad choices may lead to bad policies. Apart from selecting a proper set of relevant explanatory variables, the regulator should decide if DSOs have a control over prices (i.e., "quiet life" of a monopoly) or if they can be treated as price takers (i.e., operate in a fair competitive market). This is likely to be country-specific and have an impact on the benchmark outcome. Furthermore, we find that only a handful of models contribute significantly to the joint results and that the "policy-driven" models (which provide more "desirable" results to particular stakeholders) are marginalized due to very low posterior probabilities.

The paper is structured as follows. In Section 2, we present Bayesian Stochastic Frontier Analysis used in the empirical study. In Section 3, we outline Bayesian model selection and averaging techniques and how the recently developed class of Corrected Arithmetic Mean Estimators of the marginal data density value is adopted in the DSO cost efficiency analysis framework. In Section 4, we describe an empirical study, which analyzes cost models and cost efficiency rankings of business units of a Polish DSO. Section 5 concludes with a discussion.

2. BAYESIAN STOCHASTIC FRONTIER ANALYSIS

The method used to estimate cost efficiency in the Polish electricity distribution sector is based on Bayesian Stochastic Frontier Analysis (BSFA hereafter) for panel data; see Koop et al. (1994b, 1997), Fernández et al. (1997), and Osiewalski and Steel (1998). Let us consider the following basic SFA model for panel data:

$$y_{it} = x_{it}'\beta + \varepsilon_{it} = x_{it}'\beta + u_i + v_{it} \tag{1}$$

where y_{it} is the cost (in logs), x_{it}' is a k-element vector of independent variables (functions of outputs, fixed assets, prices etc.), β is a vector of model parameters, i (i=1,...,n) and t (t=1,...,T) are object and time indices. The composed error ε_{it} contains a "standard" symmetric random disturbance with mean 0 (v_{it}) and one nonnegative, which reflects inefficiency (u_i). This relatively simple model with time invariant inefficiency has been finally accepted and used by Polish ERO because (i) it has performed exemplary at explaining the Polish data given the Bayesian criterion used and (ii) it suits ERO's research needs (DSOs benchmarking). The reader should know, however, that the model can be easily extended to better fit particular needs for regulatory benchmarking and policy-making. First, if the time frame of the analysis is long enough the regulator may also consider inefficiency variation over time ($\varepsilon_{it}=u_{it}+v_{it}$); see, e.g., Osiewalski and Steel, 1998; Makieła, 2017). Second, the regulator may wish to consider other forms of cost functions, which are not necessarily linear in respect to the parameters (Koop et al., 1994a; Wróbel-Rotter, 2007). Third, the regulator may want to incorporate heterogeneity in the frontier, e.g., via individual effects as in true random-effects models (TRE; $\varepsilon_{it}=u_{it}+v_{it}+\alpha_i$; see Greene, 2005ab; Farsi et al., 2006; or Makieła, 2017 for its Bayesian counterpart). Fourth, the regulator may also want to decompose "total" inefficiency (u_{it}) into transient inefficiency (e.g., z_{it}) associated with yardstick regulatory policies and persistent inefficiency (ξ_i), which is more suitable for benchmarking (Filippini et al., 2016; Kumbhakar and Lien, 2017). Models which incorporate four stochastic components ($\varepsilon_{it}=\xi_{it}+z_{it}+\alpha_i+v_{it}$) are currently known as generalized true random-effects (GTRE) models. Bayesian GTRE models are presented in Tsionas and Kumbhakar (2014) and later discussed in Makieła (2017). Finally, it may be important for the regulator to know if it is fair to benchmark DSOs based on differences in their cost efficiency. Restriction $u_i=0$ (thus: $\varepsilon_{it}= v_{it}$) leads to a Bayesian "non-SF" model assuming full relative efficiency of analyzed DSOs. If this simple specification is far more likely than (1) in view of the data then the regulator should either conclude that all DSOs perform equally well or consider criteria for measuring DSOs performance other than cost-effectiveness.

The approach discussed in this paper can be applied to other benchmarking criteria, technical efficiency in particular (for Bayesian models with multiple-output production frontiers see Fernández et al., 2000, 2005). Although technical efficiency is in fact a component of cost efficiency, its measurement uses different data and is based on different variables and exogeneity assumptions.

We now return to the model in (1) and describe its full Bayesian specification. Let y, X and u be the vectors and matrices containing all y_{it}, x_{it} and u_i respectively. Since inefficiencies are by construction unobservable, and thus treated as latent variables, (1) leads to the following Bayesian SF model:

$$p(y,\beta,\sigma_v^{-2},u,\varphi|X) = p(\beta,\sigma_v^{-2},\varphi)\prod_{i=1}^{n}\left(p(u_i\,|\,\varphi)\prod_{t=1}^{T} f_N(y_{it}|x_{it}\beta+u_i,\sigma_v^2)\right) \tag{2}$$

where $f_N(.\,|\,a,s^2)$ denotes the density function of the Normal distribution with mean a and variance s^2; $p(\beta,\sigma_v^{-2},\varphi)=p(\beta)p(\sigma_v^{-2})p(\varphi)$ is the full prior distribution for β (parameters of the cost function), σ_v^{-2} (inverse of variance of v_{it}; precision of v_{it}) and φ (parameter of u_i). For the symmetric disturbance we assume $p(\sigma_v^{-2})=f_G(\sigma_v^{-2}\,|\,0.5n_0,0.5a_0)$, where $f_G(.\,|\,w,z)$ is the density function of the gamma distribution with mean w/z and variance w/z^2. Hyper-parameters (a_0,n_0) are set $a_0=n_0=10^{-4}$, which yields a very flat prior on σ_v^{-2}. For the inefficiency term we set $p(u_i|\varphi)=f_G(u_i\,|\,1,\varphi)$. This reflects a distribution of u_i, conditional on parameters, which is exponential with mean and standard deviation $1/\varphi$. We assume $p(\varphi)=f_G(\varphi\,|\,1,-\ln(r_0))$, where r_0 is the prior median of the efficiency score $exp(-u_i)$, and we set r_0=0.7 throughout the study, which implies a weakly informative prior on φ (see, van den Broeck et al., 1994). This type of model is well known in the SFA literature and usually referred to as the normal-exponential model because the symmetric disturbance is assumed to be normal and inefficiency is exponential. The regulator may choose other types of BSFA models, with different assumptions as regards inefficiency (half-normal, gamma etc.). However, relative differences in efficiency scores are likely to be marginal (van den Broeck et al., 1994; Greene, 2008b; Makieła, 2014). The presented settings for hyper-parameters provide weakly informative priors to allow the data to freely and strongly influence the posterior distribution. Furthermore, since we deal with a microeconomic DSO cost function and wish to perform Bayesian inference pooling we set an informative prior on β, which is $p(\beta)=f_N(\beta|b,\Sigma)$ with k-element vector b of prior mean and a $k\times k$ diagonal prior covariance matrix Σ.[5]

Empirical results discussed in Section 4 are based on the Cobb-Douglas form of the short-run variable cost function. Thus all the coefficients in β, except the intercept, are elasticities of variable cost with respect to its determinants. The Cobb-Douglas function can be treated as either a global cost function (traditional and simplistic), or the first-order local approximation of any smooth cost function expressed in terms of logs of original economic variables. Its simplicity as well as good economic and mathematic foundations make it a reasonable reference function in an empirical analysis. Since we have also considered the translog cost function, the second-order approximation (see Christensen et al.,1971; Koop et al. 1994b, 1997; Greene, 2008a), we address its prior specification as well.[6] Taking all this into account, the prior on β is set as follows:

- All parameters grouped in β are independent.
- Variables that characterize fixed capital (assets) of a DSO—their parameters have prior means 0 and prior standard deviations 0.5.
- Variables that characterize production of a DSO—their parameters have prior means equal to the number of variables that represent production of a DSO to the power of –1 (i.e., the sum of prior means equals 1) and each prior standard deviation is 0.3.
- The intercept—has prior mean 0 and prior standard deviation 10.
- (Optional) variables that are related to observable prices of variable factors of production—their parameters have prior means that are equal to one plus the number of observed prices to the power of –1 and each prior standard deviation is 0.3. We assume that there is one unobserved price, constant over analyzed units. This unobserved price is formally used to impose homogeneity of the cost function with respect to all prices (Marzec and Osiewalski, 2008).
- (Optional) when translog cost functions are considered, parameters of all variables that are related to the second-order terms in the approximation of the unknown cost function have prior mean 0 and prior standard deviation 0.5.

The last two elements are treated additionally (as indicated in brackets "optional"), which means that in the empirical study they have been considered more as possible extensions rather than stand-alone models. The above prior on β allows us to "subtly" account for microeconomic regularity conditions as we do not wish to input strong prior knowledge about the cost function parameters for

5. If all unknown quantities in the model have informative priors then Bayesian inference pooling and model selection can always be performed. Assigning probability distributions to all relevant unknown characteristics of the model (not just to the random disturbance and the inefficiency component, but to the parameters as well) is the "technical"—or rather formal, mathematical –difference between Bayesian and non-Bayesian SFA. This additional complexity is the price we pay for a strict, probabilistic treatment of all relevant model characteristics.

6. The two functions, Cobb-Douglas and translog, have a long history of applications in econometric analyses and they are also used in this study. Of course, the regulator may wish to experiment with other, theoretically important functional forms, which are less popular in the empirical literature (see, e.g., Diewert, 1971; Magnus, 1979; Diewert and Wales, 1987; Koop et al., 1994a; Wróbel-Rotter, 2007).

electricity distribution sector. Such treatment seems particularly appropriate for electricity distribution because the definition of a cost function may not be as simple as in other sectors.

Inference about specific parameters or latent variables (inefficiencies) is made by deriving their marginal posterior distributions. In doing so, we eliminate the remaining unobserved quantities in the model (integrate them out), which are not the subject of our inference. The Bayesian model in (2) is too complex to obtain marginal posterior distributions through analytical integration. The conditional distributions, however, are analytically tractable and we can use Gibbs sampling, a Markov Chain Monte Carlo (MCMC) class algorithm, to simulate samples from the joint posterior and, thus, from the marginal posteriors (see, e.g., Koop et al. 1994b, 1997, 1999; Osiewalski and Steel, 1998; Makiela 2014). For every model considered in the study we run 1 000 000 Gibbs cycles with initial 100 000 discarded (sampler's burn-in phase).

3. MODEL COMPARISON AND INFERENCE POOLING IN BAYESIAN STOCHASTIC FRONTIER MODELS

Consider m sampling models defined over the same data space Y:

$$M_j : p_j(y|\theta_j) = p_j(y|\lambda,\eta_j), \quad j = 1,\dots,m, \tag{3}$$

where $y \in Y$ is the matrix of observations being modelled (observed DSOs costs in logs). The vector of unobserved quantities (i.e., model parameters and inefficiencies) $\theta_j = (\lambda,\eta_j) \in \Theta_j = \Lambda \times H_j$ is made up of vector λ, which groups quantities common to all models (e.g., inefficiencies, the intercept etc.) and vector η_j, which groups quantities specific to model M_j ($j = 1,2,...,m$). By defining m marginal densities $p_j(\theta_j) = p_j(\lambda,\eta_j)$ we get m Bayesian models:

$$p_j\left(y,\theta_j\right) = p_j\left(y|\theta_j\right) p_j\left(\theta_j\right), \quad j = 1,\dots,m. \tag{4}$$

Formal model comparison (given the observed data) in Bayesian inference is based on posterior probabilities; see, e.g., Osiewalski and Steel (1993). To acquire those probabilities we use the well-known Bayes formula:

$$p(M_j|y) = \frac{p(M_j)p(y \mid M_j)}{\sum_{k=1}^{m} p(M_k)p(y \mid M_k)}, \tag{5}$$

where

$$p(y|M_j) = p_j(y) = \int_{\Theta_j} p_j(y|\theta_j) p_j(\theta_j) d\theta_j, \quad j = 1,\dots,m, \tag{6}$$

is the marginal data density (also known as integrated likelihood or marginal likelihood, MDD hereafter) in model M_j and $p(M_j)$ is the prior probability of model M_j. Thus we see that the posterior probability $p(M_j|y)$ of model M_j is determined by the product of the prior model probability and MDD, which is a natural Bayesian measure of model fit. MDD itself is an average of $L_j(\theta_j;y) = p_j(y|\theta_j)$, the usual likelihood values if latent variables are not present, weighted by the marginal (or prior) density of θ_j. Using posterior model probabilities the regulator can:

i) select one particular model for DSOs cost efficiency analysis with the highest posterior probability; this procedure is known as Bayesian model choice (or selection) and it comes intuitive when one model clearly dominates over others;

ii) pool inference about quantities common to all models (in vector λ; e.g., inefficiency scores from different models); this procedure is known as Bayesian inference pooling or Bayesian model averaging; it is especially appealing when, e.g., the regulator wants to make inference about inefficiency scores common to all models and there is no model that clearly dominates; the method amounts to calculating a pooled density which is a weighted average of densities of vector λ from individual models; the weights are equal to each model's posterior probability $p(M_j|y)$, i.e.:

$$p(\lambda|y) = \sum_{j=1}^{m} p(M_j \mid y) p_j(\lambda \mid y) \tag{7}$$

In order to perform Bayesian model selection or inference pooling the regulator needs to know two elements for all analyzed models ($j = 1, 2, \ldots, m$): their prior probabilities—$p(M_j)$—and marginal data density (MDD) values—$p_j(y)$. As far as choosing $p(M_j)$ is concerned, it is up to the researcher, or in this case the regulator, to determine prior probability of each competing model (e.g., rank model adequacy based on theoretical considerations or possible policy implications). If all models are viewed equally likely *a priori* they can be assigned equal probabilities:

$$p(M_j) = \frac{1}{m}\,. \tag{8}$$

Alternatively, if two models have equal explanatory power as measured by $p_j(y)$—MDD—according to the Ockham Razor rule we should prefer the simpler one. Thus, following this principle we can *a priori* penalize over-parametrized models (Osiewalski and Steel, 1993):

$$p(M_j) \propto 2^{-l_j} \tag{9}$$

where l_j denotes the number of parameters in model M_j. Both approaches based on (8) and (9) are discussed in the empirical section of the paper.

In practice selecting prior probabilities is trivial and it is relatively easy to trace their impact on posterior model probabilities. What has been the reason for lack of Bayesian model selection and inference pooling in BSFA is the precise calculation of the MDD value—$p_j(y)$. SFA models contain latent variables, which makes their MDD values especially challenging to estimate precisely.[7] Often in such cases, MDD values are acquired using the Harmonic Mean Estimator—HME (Newton and Raftery, 1994)—based on some MCMC scheme, e.g., Gibbs sampler. However, as Lenk (2009) points out, numerical implementation of this estimator is "upwardly pseudo-biased" (Lenk, 2009, p. 952). This means that HME overestimates the real value of MDD, especially in more complex models and when sample size is large; for details see Pajor and Osiewalski (2013–2014). Lenk's proposal to account for this pseudo-bias has been first applied to models with thousands of latent variables by Osiewalski and Osiewalski (2013, 2016). In the first application of *corrected* HME to Bayesian SFA models, Makieła (2014) has shown that differences in MDD estimates between HME

7. There is a notion circulating among some researchers that it is fast and easy to precisely approximate MDD in SF models because they have a closed-form likelihood that one can use. This is not entirely correct. First, these are only "approximately" closed-form solutions, meaning that the likelihood function itself requires some numerical approximation to obtain its values. Second, even if the likelihood can be easily evaluated anywhere in the parameter space, its averaging with respect to the prior density requires numerical integration as the second step of the procedure. Third, we use an approach which is one-step, formally correct, very easy and numerically efficient when the number of latent variables is not very large. If the number of latent variables is considerable then the above-mentioned two-step procedure would be preferable.

and *corrected* HME can be considerable. In this research we have also considered HME and found that the estimator can be very unstable in some models, particularly in the highly dimensional ones.[8] The difference sometimes can be even up to orders of magnitude, which given the fact that some models are close to each other makes the model ranking unstable. Correcting HME does not fully mitigate the problem. The correction factor is relatively stable and as a result the *corrected* HME inherits HME's instability in those models. To sum up, the problem with using (*corrected*) HME in a study such as this one is its stability in some models. If the regulator is fine with excluding those models (on the assumption that they may not play any major role in inference pooling anyway) then HME can be used. However, we want to use all models proposed by the regulator and the DSOs experts, and thus move away from HME.

Also building on Lenk (2009) and Pajor and Osiewalski (2013–2014), Pajor (2017) has shown a way to adjust (or correct) Arithmetic Mean Estimators (AME) of MDD. This relatively straightforward method amounts to "trimming" the space of parameters in order to remove regions of relatively low likelihood. The adjustment procedure usually is based on Monte Carlo—Importance Sampling (MC-IS) and it has proven to perform exceptionally well in comparison to other methods (Pajor 2017). Also, an important feature of so-called CAME estimators is that they can be applied to models with latent variables, such as SFA, especially if the number of latent variables is not very large. We have found this class of estimators to be particularly useful as it provides stable estimates of MDD in the DSOs cost efficiency analysis. In this paper CAME estimator has been implemented as follows:

- We "trim" parameter space based on minimum-maximum values using accepted draws of each parameter (or latent variable) in the MCMC scheme.
- We take the multivariate Normal distribution as the importance sampling distribution in the MC-IS scheme. We set the mean vector and covariance matrix equal to the posterior mean vector and posterior covariance matrix based on accepted draws from the MCMC scheme. The reason behind this is that due to the number of parameters and latent variables (inefficiencies) in SFA models it is important for the sampling mechanism to take into account the covariance structure.

We have found that such an implementation scheme significantly improves the numerical performance of the CAME estimator (i.e., fewer runs and less time required for the simulation to stabilize at the sought-after MDD value). Traditionally one would use multivariate Student distribution as an importance sampling distribution. However, we have found that (dependently on the degrees of freedom) its implementation takes more time to compute than the multivariate Normal case while yielding very similar results (up to the second decimal point). Apart from that, we have also experimented with Gamma distribution for latent variables but this has turned out to be the least effective sampling mechanism in terms of computation speed (also eventually yielding similar results). To our knowledge this is the first application of CAME estimator in BSFA literature.

To sum up, it should be noted that formal model comparison and inference pooling is a distinctive feature of Bayesian inference. No other approach allows us to formally (in a probabilistic sense) make such comparisons or pool inference about quantities common to all models under consideration.

8. It should not be taken as a rule that HME is unstable in more complex models. As it has later turned out these are particularly bad models with very low posterior probabilities as measured by the more precise CAME estimator.

4. COST EFFICIENCY ANALYSIS OF A DSO IN POLAND

The Polish electricity distribution sector has been very dynamic over the last few decades, especially in terms of mergers and acquisitions. There are currently five electricity distribution system operators (DSOs) in Poland: PGE, Tauron, Energa, Enea and RWE. The DSOs have been operating in their current form only since 2008. For the first official benchmarking analysis (2001–2006) there were 14 DSOs and during communism in Poland there were even up to 49 small DSOs, one for each administrative district at that time (*Voivodeship*). The biggest DSO, PGE, operates in eastern and parts of central Poland. It has eight territorial business units which cover almost half of Poland's territory. The second biggest DSO, Tauron, operates in the southern Poland and is made up of 11 business units. It is also the largest electricity retail supplier. Energa is the third biggest DSO covering northern and parts of central Poland. It consists of six business units. The fourth largest is Enea, which covers most of western Poland with its five units. RWE (currently Innogy Stoen, no unit breakdown) distributes electricity in the Warsaw Metropolitan Area. All five DSOs are regulated by the Polish Energy Regulatory Office (ERO) based on government's energy policy described in "Strategy for Regulating Distribution System Operators in 2016–2020" (Polish Energy Regulatory Office, 2015b). Cost efficiency analysis is a significant part of this regulatory policy.

The study presented in this section has been commissioned by one of Polish DSOs in order to analyze cost efficiency of its territorial business units. The study implements methodology described in Sections 2 and 3 and it has been carried out in two stages as requested by the DSO. Stage one uses panel data from the DSO's territorial business units over the period of two subsequent years. Stage two considers three subsequent years of their operation.[9] The separate stages have been done for two reasons. First, the DSO needed to have preliminary analysis before the full dataset (for three years) was made available in order to better prepare its experts. Second, formal model comparison makes little sense for different data (two and three years).

The DSO has provided two sets of data: (i) information on variable costs of its business units (its natural log being the dependent variable) and (ii) a dataset that describes their operation characteristics (their natural logs being the potential explanatory variables). The variable cost modeled in the study is defined as operational cost, which is made up of costs of payroll and social security plus other benefits, costs of external services, costs of consumption of materials and energy, taxes and charges and other costs by type. It has been constructed in accordance with regulations set by the Polish Energy Regulatory Office (2015a); the same definition is used by ERO to study cost efficiency of all Polish DSOs. Furthermore, the DSO has provided information on 34 different characteristics of its operation (see Table 1). These characteristics represent potential explanatory variables and describe differences in its business units' operation in terms of scale, technical properties, quality and observed wages[10] (variable x34). Naturally this list is too long for a reasonable econometric model to be built. Ideally, in econometric modelling we want each explanatory variable to describe a unique portion of information (about the dependent variable), which is not related to information already described by other explanatory variables in the model. Unfortunately, in the case of electricity distribution industry this is next to impossible to achieve. Many of the DSO's characteristics are intertwined with each other, often at a technical level (e.g., number of electric

9. Due to disclosure agreement with the DSO we do not provide its name, exact timeframe of the analysis or the total number of business units that were actually analyzed. Results for six business units are presented in this paper, which is representative.

10. Wages are given as unit's total salary expenses (with overheads) divided by the number of job posts (full time equivalents); outsourced services are centrally managed by the DSO so they are not considered here.

Table 1: List of potential explanatory variables (operational characteristics)

Code		Name of a characteristic
x1	1	Total EHV and HV line length per one circuit
x2	1	MV overhead line length per one circuit
x3	1	MV cable length per one circuit
x4	1	Total LV line length per one circuit
x5	1	Total LV overhead line length per one circuit + the length of LV overhead service wires
x6	1	Total LV cable length per one circuit + the length of LV undeground service wires
x7	1	The number of HV substations
x8	1	The number of MV/MV and MV/LV substations
x9	1	Total power of EHV/HV and HV/MV transformes
x10	1	Total power of MV/MV and MV/LV transformers
x11	1	The number of transformers
x12	2	The number of meters in the HV customer group
x13	2	The number of meters in the MV customer group
x14	2	The number of meters in the LV commercial customer group
x15	2	The number of meters in the LV residential customer group
x16	2	Volume of electricity supplied to the HV customer group
x17	2	Volume of electricity supplied to the MV customer group
x18	2	Volume of electricity supplied to LV commercial customer group
x19	2	Volume of electricity supplied to LV residential customer group
x20	1	Average peak load
x21	2	Energy received from other DSOs (HV, MV, LV) + energy received from TSO (EHV)
x22	1	DSO's area of operations
x23	2	Inverse of System Average Interruption Duration Index (SAIDI ^–1)
x24	2	Inverse of System Average Interruption Frequency Index (SAIFI ^–1)
x25	2	Inverse of time to connect a new customer (^–1)
x26	1	Total MV line length per one circuit
x27	1	Total LV line length per one circuit + the length of LV service wires
x28	1	Total number of substations
x29	1	Total power of transformers
x30	2	The number of meters in the LV customer group
x31	2	Total number of meters
x32	2	Volume of electricity supplied to the LV customer group
x33	2	Total volume of electricity supplied
x34	3	Average salary

Note: EHV is extra high voltage (above 200kV; 220kV or 400kV); HV is high voltage (60kV–200kV; mostly 110kV); MV is medium voltage (1kV–60kV; mostly 15kV); LV is low voltage (below 1kV; mostly 230V); TSO is transmission system operator; second column: *1* means that a variable represents a characteristic of capital assets; *2* that a variable represents a characteristic of production (quantity or quality); *3* that a variable is a price indicator; apart from "average salary" the list of characteristics (definitions and order on the list) is as determined by the Polish Energy Regulatory Office.

energy converters vs. number of electric stations). Table 2 shows a matrix of empirical correlation coefficients between the 34 variables. We see that many characteristics are very highly correlated with each other. This means that supposedly different lists of variables can lead to very similar models because these variables carry similar information. Thus, one of the main goals of this study has been to find (using Bayesian techniques; see Section 3) the sets of explanatory variables that are relevant in describing the variable cost and to make pooled inference about cost efficiency of the DSO's business units based on them.

The study groups competing models into two scenarios. All models of the first scenario exclude information about price of labor (wages), while models of the second scenario include it. The reason for such treatment is that accounting (or not) for the observed wage differences across the DSO's business units leads to two fundamentally different types of models as regards cost efficiency interpretation. Models without wage information (first scenario) assume that the business

Table 2: Correlation coefficients between potential explanatory variables

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20	x21	x22	x23	x24	x25	x26	x27	x28	x29	x30	x31	x32	x33	x34
x34	0.51	-0.18	0.59	0.34	0.2	0.47	0.76	0.47	0.62	0.58	0.49	0.55	0.49	0.52	0.62	0.38	0.56	0.53	0.56	0.66	0.5	-0.17	0.37	0.44	0.33	0.36	0.34	0.48	0.61	0.61	0.61	0.55	0.61	**1**
x33	0.78	0.06	**0.92**	0.62	0.33	0.83	0.89	0.75	**0.92**	0.89	0.75	0.72	0.73	0.82	0.86	0.61	0.87	0.85	0.86	**0.97**	0.76	0.07	0.76	0.66	-0.06	0.68	0.57	0.75	**0.91**	0.86	0.86	0.86	**1**	0.61
x32	0.79	0.38	**0.95**	0.89	0.65	**0.97**	0.88	**0.96**	**0.96**	**0.98**	**0.96**	0.45	**0.92**	**0.98**	**0.99**	0.38	**0.94**	**1**	**1**	0.88	0.6	0.35	0.47	0.42	0.05	0.89	0.86	**0.96**	**0.98**	**0.99**	**0.99**	**1**	0.86	0.55
x31	0.78	0.3	**0.94**	0.86	0.62	**0.95**	**0.92**	**0.93**	**0.97**	**0.97**	**0.94**	0.43	**0.93**	**0.97**	**1**	0.33	**0.95**	**0.98**	**0.99**	0.89	0.61	0.28	0.45	0.45	0.08	0.85	0.83	**0.93**	**0.98**	**1**	**1**	**0.99**	0.86	0.61
x30	0.78	0.3	**0.94**	0.86	0.62	**0.95**	**0.92**	**0.93**	**0.97**	**0.97**	**0.94**	0.43	**0.93**	**0.97**	**1**	0.33	**0.95**	**0.98**	**0.99**	0.89	0.61	0.28	0.45	0.45	0.08	0.85	0.83	**0.93**	**0.98**	**1**	**1**	**0.99**	0.86	0.61
x29	0.79	0.22	**0.98**	0.79	0.52	**0.95**	**0.93**	0.89	**0.99**	**0.99**	**0.9**	0.51	0.89	**0.93**	**0.98**	0.42	**0.96**	**0.96**	**0.98**	**0.93**	0.61	0.22	0.54	0.5	0.02	0.82	0.75	0.89	**1**	**0.98**	**0.98**	**0.98**	**0.91**	0.61
x28	0.81	0.59	0.86	**0.96**	0.75	**0.93**	0.79	**1**	0.87	**0.9**	**1**	0.38	0.83	**0.99**	**0.93**	0.34	0.83	**0.97**	**0.95**	0.77	0.55	0.52	0.45	0.36	0.1	**0.96**	**0.92**	**1**	0.89	**0.93**	**0.93**	**0.96**	0.75	0.48
x27	0.64	0.66	0.69	**0.98**	**0.94**	0.77	0.67	**0.92**	0.74	0.76	**0.92**	0.29	0.72	**0.9**	0.82	0.29	0.67	0.87	0.84	0.65	0.43	0.51	0.3	0.24	0.35	0.87	**1**	**0.92**	0.75	0.83	0.83	0.86	0.57	0.34
x26	0.84	0.72	0.82	**0.9**	0.71	0.88	0.72	**0.96**	0.79	0.85	**0.96**	0.34	0.76	**0.93**	0.84	0.36	0.73	**0.9**	0.88	0.68	0.42	0.7	0.46	0.37	0.06	**1**	0.87	**0.96**	0.82	0.85	0.85	0.89	0.68	0.36
x25	0.08	0.13	-0.06	0.22	0.56	-0.15	0.21	0.1	0.05	-0.02	0.12	0.21	-0.05	0.08	0.07	0.21	-0.12	0.04	0.06	0.14	-0.1	-0.02	-0.04	-0.1	**1**	0.06	0.35	0.1	0.02	0.08	0.08	0.05	-0.06	0.33
x24	0.61	0	0.55	0.22	0.14	0.39	0.62	0.36	0.55	0.42	0.35	0.41	0.31	0.43	0.45	0.35	0.44	0.41	0.43	0.65	0.65	0.05	-0.75	**1**	-0.1	0.37	0.24	0.36	0.5	0.45	0.45	0.42	0.66	0.44
x23	0.66	0.09	0.56	0.35	0.15	0.47	0.57	0.45	0.56	0.5	0.44	0.69	0.22	0.49	0.45	0.63	0.43	0.49	0.46	0.71	0.65	0.06	**1**	-0.75	-0.04	0.46	0.3	0.45	0.54	0.45	0.45	0.47	0.76	0.37
x22	0.53	**0.93**	0.27	0.52	0.49	0.38	0.13	0.52	0.17	0.29	0.52	-0.1	0.34	0.43	0.26	0.03	0.16	0.38	0.33	0.03	-0.15	**1**	0.06	0.05	-0.02	0.7	0.51	0.52	0.22	0.28	0.28	0.35	0.07	-0.17
x21	0.45	-0.04	0.61	0.47	0.25	0.6	0.6	0.55	0.65	0.55	0.52	0.45	0.44	0.63	0.61	0.35	0.61	0.62	0.58	0.73	**1**	-0.15	0.65	0.65	-0.1	0.42	0.43	0.55	0.61	0.61	0.61	0.6	0.76	0.5
x20	0.76	0.06	**0.9**	0.67	0.47	0.8	**0.94**	0.76	**0.94**	**0.9**	0.77	0.73	0.73	0.84	0.89	0.63	0.86	0.86	0.88	**1**	0.73	0.03	0.71	0.65	0.14	0.68	0.65	0.77	**0.93**	0.89	0.89	0.88	**0.97**	0.66
x19	0.79	0.35	**0.95**	0.87	0.64	**0.96**	**0.9**	**0.94**	**0.97**	**0.98**	**0.95**	0.45	**0.92**	**0.97**	**0.99**	0.38	**0.94**	**0.99**	**1**	0.88	0.58	0.33	0.46	0.43	0.06	0.88	0.84	**0.95**	**0.98**	**0.99**	**0.99**	**1**	0.86	0.56
x18	0.79	0.43	**0.93**	**0.92**	0.67	**0.97**	0.86	**0.97**	**0.94**	**0.96**	**0.97**	0.45	**0.9**	**0.99**	**0.98**	0.38	**0.92**	**1**	**0.99**	0.86	0.62	0.38	0.49	0.41	0.04	**0.9**	0.87	**0.97**	**0.96**	**0.98**	**0.98**	**1**	0.85	0.53
x17	0.67	0.12	**0.96**	0.72	0.4	**0.95**	0.85	0.82	**0.96**	**0.94**	0.82	0.39	**0.93**	0.88	**0.95**	0.3	**1**	**0.92**	**0.94**	0.86	0.61	0.16	0.43	0.44	-0.12	0.73	0.67	0.83	**0.96**	**0.95**	**0.95**	**0.94**	0.87	0.56
x16	0.41	0.03	0.46	0.27	0.23	0.29	0.47	0.34	0.4	0.44	0.36	**0.94**	0.15	0.38	0.33	**1**	0.3	0.38	0.38	0.63	0.35	0.03	0.63	0.35	0.21	0.36	0.29	0.34	0.42	0.33	0.33	0.38	0.61	0.38
x15	0.78	0.29	**0.94**	0.85	0.61	**0.95**	**0.92**	**0.92**	**0.97**	**0.97**	**0.93**	0.43	**0.93**	**0.96**	**1**	0.33	**0.95**	**0.98**	**0.99**	0.89	0.61	0.26	0.45	0.45	0.07	0.84	0.82	**0.93**	**0.98**	**1**	**1**	**0.99**	0.86	0.62
x14	0.8	0.49	**0.91**	**0.93**	0.71	**0.95**	0.84	**0.99**	**0.92**	**0.93**	**0.98**	0.43	0.87	**1**	**0.96**	0.38	0.88	**0.99**	**0.97**	0.84	0.63	0.43	0.49	0.43	0.08	**0.93**	**0.9**	**0.99**	**0.93**	**0.97**	**0.97**	**0.98**	0.82	0.52
x13	0.65	0.29	0.88	0.75	0.52	**0.91**	0.79	0.83	0.87	**0.91**	0.83	0.24	**1**	0.87	**0.93**	0.15	**0.93**	**0.9**	**0.92**	0.73	0.44	0.34	0.22	0.31	-0.05	0.76	0.72	0.83	0.89	**0.93**	**0.93**	**0.92**	0.73	0.49
x12	0.51	-0.09	0.51	0.3	0.21	0.35	0.6	0.38	0.51	0.51	0.39	**1**	0.24	0.43	0.43	**0.94**	0.39	0.45	0.45	0.73	0.45	-0.1	0.69	0.41	0.21	0.34	0.29	0.38	0.51	0.43	0.43	0.45	0.72	0.55
x11	0.81	0.58	0.86	**0.96**	0.75	**0.93**	0.8	**1**	0.87	**0.91**	**1**	0.39	0.83	**0.98**	**0.93**	0.36	0.82	**0.97**	**0.95**	0.77	0.52	0.52	0.44	0.35	0.12	**0.96**	**0.92**	**1**	**0.9**	**0.94**	**0.94**	**0.96**	0.75	0.49
x10	0.78	0.28	**0.97**	0.8	0.52	**0.96**	0.89	**0.9**	**0.96**	**1**	**0.91**	0.51	**0.91**	**0.93**	**0.97**	0.44	**0.94**	**0.96**	**0.98**	**0.9**	0.55	0.29	0.5	0.42	-0.02	0.85	0.76	**0.9**	**0.99**	**0.97**	**0.97**	**0.98**	0.89	0.58
x9	0.78	0.17	**0.97**	0.77	0.51	**0.92**	**0.94**	0.87	**1**	**0.96**	0.87	0.51	0.87	**0.92**	**0.97**	0.4	**0.96**	**0.94**	**0.97**	**0.94**	0.65	0.17	0.56	0.55	0.05	0.79	0.74	0.87	**0.99**	**0.97**	**0.97**	**0.96**	**0.92**	0.62
x8	0.8	0.59	0.86	**0.96**	0.75	**0.93**	0.79	**1**	0.87	**0.9**	**1**	0.38	0.83	**0.99**	**0.92**	0.34	0.82	**0.97**	**0.94**	0.76	0.55	0.52	0.45	0.36	0.1	**0.96**	**0.92**	**1**	0.89	**0.93**	**0.93**	**0.96**	0.75	0.47
x7	0.84	0.11	**0.9**	0.67	0.5	0.79	**1**	0.79	**0.94**	0.89	0.8	0.6	0.79	0.84	**0.92**	0.47	0.85	0.86	**0.9**	**0.94**	0.6	0.13	0.57	0.62	0.21	0.72	0.67	0.79	**0.93**	**0.92**	**0.92**	0.88	0.89	0.76
x6	0.74	0.38	**0.94**	0.84	0.51	**1**	0.79	**0.93**	**0.92**	**0.96**	**0.93**	0.35	**0.91**	**0.95**	**0.95**	0.29	**0.95**	**0.97**	**0.96**	0.8	0.6	0.38	0.47	0.39	-0.15	0.88	0.77	**0.93**	**0.95**	**0.95**	**0.95**	**0.97**	0.83	0.47
x5	0.5	0.69	0.43	0.86	**1**	0.51	0.5	0.75	0.51	0.52	0.75	0.21	0.52	0.71	0.61	0.23	0.4	0.67	0.64	0.47	0.25	0.49	0.15	0.14	0.56	0.71	**0.94**	0.75	0.52	0.62	0.62	0.65	0.33	0.2
x4	0.69	0.65	0.73	**1**	0.86	0.84	0.67	**0.96**	0.77	0.8	**0.96**	0.3	0.75	**0.93**	0.85	0.27	0.72	**0.92**	0.87	0.67	0.47	0.52	0.35	0.22	0.22	**0.9**	**0.98**	**0.96**	0.79	0.86	0.86	0.89	0.62	0.34
x3	0.78	0.22	**1**	0.73	0.43	**0.94**	**0.9**	0.86	**0.97**	**0.97**	0.86	0.51	0.88	**0.91**	**0.94**	0.46	**0.96**	**0.93**	**0.95**	**0.9**	0.61	0.27	0.56	0.55	-0.06	0.82	0.69	0.86	**0.98**	**0.94**	**0.94**	**0.95**	**0.92**	0.59
x2	0.49	**1**	0.22	0.65	0.69	0.38	0.11	0.59	0.17	0.28	0.58	-0.09	0.29	0.49	0.29	0.03	0.12	0.43	0.35	0.06	-0.04	**0.93**	0.09	0	0.13	0.72	0.66	0.59	0.22	0.3	0.3	0.38	0.06	-0.18
x1	**1**	0.49	0.78	0.69	0.5	0.74	0.84	0.8	0.78	0.78	0.81	0.51	0.65	0.8	0.78	0.41	0.67	0.79	0.79	0.76	0.45	0.53	0.66	0.61	0.08	0.84	0.64	0.81	0.79	0.78	0.78	0.79	0.78	0.51

Note: Based on empirical Pearson correlation coefficients between natural logs of characteristics x1 to x34. High and very high coefficient levels have been marked.

units have some control over the price of hired labor and that wages should not be treated as exogenously determining cost. Any variation in observed labor costs, which are due to differences in mean wages between units, are exhibited in their cost efficiency scores. This is because the DSO's business units are assumed to have control over wages, and thus those that pay more for their labor are deemed less cost-effective. This scenario is also often the case when a fair market price is difficult to determine due to DSOs territorial monopoly. Models that include information on wages (second scenario) assume that business units are "price takers", i.e., they have no influence on the observed wages. Hence, labor price is regarded as a justified determinant of their costs—one that they have no control over (i.e., it is beyond the decision making process of the management). We do not wish to comment on which scenario (with wages or not) is more suitable to describe situation on an energy distribution market. This may be very country-specific and it is up to the regulator to decide. As a rule of thumb, however, one may assume that the more centralized an energy market is, the more the first scenario seems a better fit. In this study a comparative analysis of the two scenarios allows us to show what share of the observed cost generated by the DSO's business units would be justified if wages were equal among all objects and time (i.e., what would happen to the cost efficiency ranking if salary levels where fixed and centralized by the DSO management).

The models finally used are based on a Bayesian stochastic cost frontier model with Cobb-Douglas (CD) cost function. Initially, in stage one we have also considered translog specifications, time trend and more flexible SFA assumptions as regards inefficiency modelling. Models with translog functions have been dropped due to considerably lower marginal data density (MDD) values in comparison to their simpler Cobb-Douglas counterparts. We have considered time trend as a possible explanatory variable to account for neutral technical change or price inflation ($\beta_{time} = \partial \ln y / \partial t,\ t = 1,2$). Again, however, the analysis has revealed that simpler, "static" models with no time trend are significantly better at explaining the data due to considerably higher MDD levels. The analysis has also indicated that the BSFA models with (in)efficiency components constant over time significantly outperform "richer" models with (in)efficiencies varying over time, as measured by the MDD value. Furthermore, since we are only interested in benchmarking and we analyze units of one particular DSO there is no reason to impose any heterogeneity in the frontier, e.g., via standard individual effects or decompose inefficiency into transient and persistent components as in Kumbhakar and Lien (2017). For the above reasons, all models reported are based on specification given in (2). We should note that the outcome of preliminary analyses performed in stage one seems intuitive especially due to small number of observations over time (T=2 in stage one T= 3 in stage two).

4.1. Stage One of the Cost Efficiency Analysis

As mentioned, the study has been performed in two stages. Stage one has been mostly designed as an introductory study, one which would allow the authors and the DSO experts to better prepare for the second stage once the full dataset (for three years) is made available. It has also allowed the authors to perform a series of preliminary analyses briefly described in the previous paragraph. The final number of models considered at this stage was 37: (i) 10 models for the first scenario and (ii) 27 models for the second scenario (see Table 3). These models were proposed by various participants involved in the study—the authors as well as the DSO experts—and they were based on various sources, e.g., (i) authors experience, (ii) DSO's managerial expertise, (iii) desk research or (iv) by means of purely statistical analysis of the data (e.g., Principal Component Analysis, statistical tests and reductions of the "full" model—model 0). The variety of sources was intentional

Table 3: Model variants considered in the analysis

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
x1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1			1		1	1	1	1	1	1	1	1	1	1			1
x2	1	1	1	1		1	1	1		1				1	1		1		1		1							1	1	1	1	1			1				
x3	1	1	1	1		1	1		1			1		1			1		1		1							1	1	1									1
x4	1	1	1											1			1	1									1					1							
x5	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1		1		1	1			1			1	1	1	1				1	1			1
x6	1	1	1	1	1	1	1		1			1		1	1	1	1		1	1	1		1	1		1		1	1	1							1		1
x7	1	1	1	1		1	1	1	1	1	1		1	1	1		1		1	1		1		1	1			1	1	1	1	1			1	1			
x8	1	1	1	1		1	1		1					1					1									1	1	1									
x9	1	1	1	1										1																									
x10	1	1	1	1										1																		1							
x11	1	1	1	1										1		1		1			1													1					
x12	2	2		2	2									2	2	2			2	2			2	2		2											2	2	
x13	2	2		2										2		2			2																				
x14	2	2		2	2									2	2	2			2				2			2													
x15	2	2		2	2									2	2	2			2	2			2	2		2											2		
x16	2	2	2	2	2									2	2		2			2	2	2	2	2	2	2											2	2	
x17	2	2	2	2										2		2	2				2	2			2														
x18	2	2	2	2										2	2	2					2																		
x19	2	2	2	2										2	2	2					2																		
x20	1	1	1											1			1	1																					
x21	2	2		2	2									2	2		2	2		2		2	2	2	2														
x22	1				1	1	1	1		1				1	1		1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	
x23	2							2					2	2			2	2									2		2		2								
x24	2				2						2			2	2		2			2		2	2	2	2	2	2									2	2	2	
x25	2													2			2					2					2												
x26	1	1	1													1		1									1						1	1					
x27	1	1	1												1	1		1															1	1					
x28	1	1	1													1		1															1						
x29	1	1	1													1		1									1												
x30	2					2			2			2								2				2				2											2
x31	2				2										2			2		2			2	2		2							2					2	
x32	2	2	2					2		2	2		2		2		2					2			2		2				2	2			2	2			
x33	2	2			2										2	2		2		2			2			2								2					
x34	3													3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
St.2	Y					Y		Y	Y		Y		Y		Y								Y			Y		Y		Y	Y				Y	Y			Y

Note: Each column labelled from 0 to 38 represents a different model specification as regards the list of explanatory variables. Value "*1*" in the table indicates that a given variable represents a characteristic of capital assets; "*2*" that a variable represents a characteristic of production; "*3*" that a variable is a price indicator; Label "Y" (Yes) in row "St.2" indicates that a model has been used in stage two; models 8 and 12 have been used only in stage two .

for two reasons. First, the authors wanted to make the model proposition process as inclusive as possible for the DSO experts. Second, for the proposed BSFA framework it is really not that relevant how a given model is proposed as long as it does not violate general principles of a short-run microeconomic variable cost function, i.e., it contains at least one variable that represents output and one that is a measure of capital assets (fixed inputs). This is because Bayesian model selection and inference pooling techniques marginalize (via MDD values) models that are bad at describing the data anyway.

Table 4 shows MDD values (in decimal logs), mean efficiency scores and the total number of explanatory variables in each model from stage one. We see that the data clearly prefer only a few

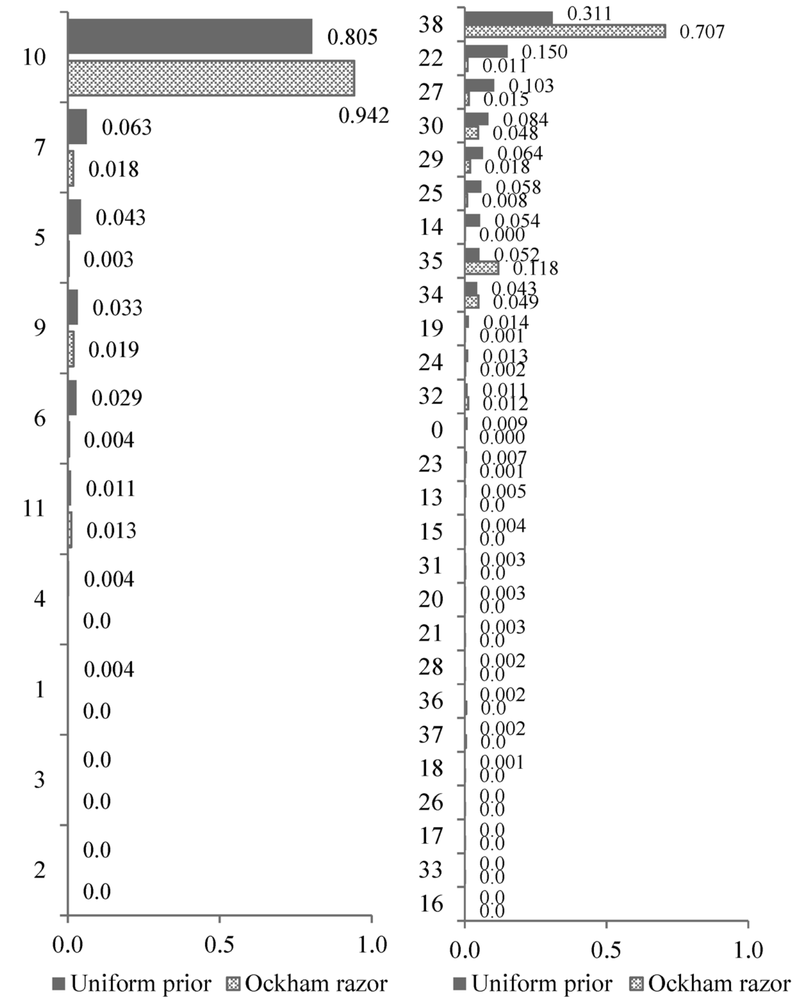
Table 4: Marginal data density values, mean efficiency scores and the number of explanatory variables in each model

Model code	MDD	Mean efficiency	No. of variables
10	3.290	0.888	5
38 (p)	2.840	0.901	6
22 (p)	2.524	0.920	11
27 (p)	2.361	0.899	10
30 (p)	2.270	0.899	8
7	2.182	0.891	7
29 (p)	2.155	0.908	9
25 (p)	2.111	0.925	10
14 (p)	2.079	0.897	19
35 (p)	2.063	0.896	6
5	2.020	0.893	9
34 (p)	1.979	0.892	7
9	1.900	0.885	6
6	1.847	0.901	8
19 (p)	1.507	0.909	12
24(p)	1.446	0.873	10
11	1.415	0.902	5
32 (p)	1.376	0.855	7
0 (p)	1.287	0.866	34
23 (p)	1.198	0.880	11
13 (p)	1.054	0.864	26
4	1.027	0.919	10
1	0.966	0.856	27
15 (p)	0.935	0.869	17
31 (p)	0.882	0.861	8
20 (p)	0.872	0.870	12
21 (p)	0.809	0.874	11
28 (p)	0.702	0.892	10
36 (p)	0.648	0.854	6
37 (p)	0.587	0.807	6
18 (p)	0.466	0.886	13
26 (p)	0.079	0.845	10
17 (p)	0.006	0.872	13
3	-0.059	0.868	19
33 (p)	-0.182	0.812	7
2	-0.921	0.859	21
16 (p)	-1.308	0.854	17

Note: Results obtained in stage one; MDD stands for marginal data density in decimal log; ”(p)” next to model code indicates that this model contains information about wages.

parsimonious models with no more than 5–6 explanatory variables. This seems particularly interesting since previous studies about Polish electricity distribution sector would either use more models with much larger list of explanatory variables (Osiewalski and Wróbel-Rotter, 2008–9, 2012), or take combinations of several pre-selected explanatory variables and use various modelling techniques (e.g., Cullman and von Hirschhausen, 2008b). We also find that the best models show very high, relative to other models, average cost efficiency scores despite having a small number of variables and parameters.

Figure 1: Model posterior probabilities in stage one



Note: Left chart: models without labor prices (scenario one); right chart: models with labor prices (scenario two); solid bars: posterior probability based on a uniform prior; doted bars: based on Ockham Razor.

Figure 1 shows models' posterior probability rankings for the two scenarios in stage one. As mentioned in Section 3, prior model probability has a straightforward impact on the posterior model probability. For this reason in stage one the authors and the DSO experts have agreed on two rankings based on priors in (8) and in (9). Solid bars present posterior model probabilities based on a uniform prior (i.e., equal prior probabilities for all models); doted bars present posterior model probabilities according to Ockham Razor as in (9). Based on this we can say that:

- In the first scenario posterior model probability ranking is dominated by only one model—model 10. Dependently on the assumed prior model probability structure, the model has 0.8 or 0.94 of the posterior probability mass. The model has not only a relatively high MDD but is also quite parsimonious. Thus if we were to follow the rule of Ockham Razor we could make inference solely based on this model.
- In the second scenario posterior model probability ranking is no longer that dominated by one particular model. Of course, model 38 does take a large portion of the posterior

probability mass, but there are several other models in the ranking (22, 25, 27, 29, 30, 34, 35) that cannot be entirely ignored. The model ranking changes dependently on the assumed prior model probability. If we assume a uniform prior (solid bar) the ranking is relatively even. If we consider Ockham Razor (doted bar) there are only two (38, 35), maybe four (38, 35, 34, 30) models we should be really concerned with in Bayesian inference pooling (i.e., efficiency estimates from those models will significantly contribute to the final results).

We find that inference pooling based on the two scenarios leads to similar cost efficiency rankings and indicates a relatively high average cost efficiency of DSO's business units, which is 0.889 (with an average posterior standard deviation 0.072) for models without wages (first scenario) and 0.902 (0.079) for models with wages (second scenario). However, in models without wages the gap between the leader and the least efficient unit is almost twice the size as in models with wages. On the one hand, looking at this result from the first scenario's perspective wages can be viewed as a potentially good determinant of efficiency differences. On the other hand, given the second scenario wages can be viewed as a relevant determinant of variable cost differences.

4.2. Stage Two of the Cost Efficiency Analysis

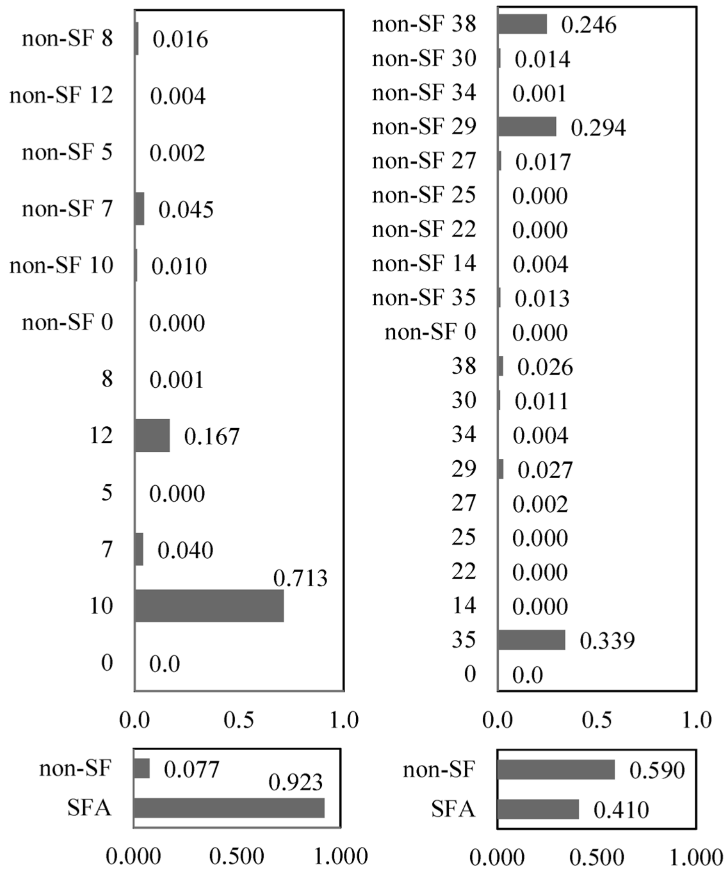
Stage two of the analysis builds on the results from stage one. Since many models analyzed in stage one have turned out to be highly inadequate in view of the data, the following simple criteria have been used to "trim" the list of models under consideration in stage two:

- In the second scenario we take those models that have increased their posterior probability in relation to the uniform prior. Since we have 27 models in this scenario, for the second stage we consider those, for which posterior probability is above 1/27 in stage one. These are models: 38, 22, 27, 30, 29, 25, 14, 35, 34.
- In the first scenario if we take the above mentioned criterion for models without wages we end up with only one model—model 10. For this reason the authors and the DSO have decided to take the best three models, i.e., 10, 7, and 5. Since models 7 and 5 are quite "parameter-rich" their simplified versions have been constructed and introduced into stage two of the study: model 12, which is a more parsimonious version of model 7, and model 8, which is a more parsimonious version of model 5 (see Table 3).
- We also retain model 0 as a baseline—a model that indiscriminately contains all 34 characteristics. This, however, should not be viewed as a justification for building such a model. Apart from the fact that the model is numerically challenging to estimate it is also among the worst models in the study.

This gives a total of 15 models: five models in the first scenario, nine models in the second scenario and model 0. These models have been treated in stage two as *a priori* equally probable, regardless of the number of parameters (i.e., we have used a uniform prior). Details regarding each model specification are provided in Table 3 (label "Y").

Stage two of the study also extends the analytical framework. The first stage has concluded that pooled inferences from both scenarios give similar cost efficiency rankings. However, models with wages provide cost efficiency scores, which are much more even (less variation). So, once wages are accounted for, is there any heterogeneity left, as regards cost-effectiveness of those objects? Since SFA models assume inefficiency terms by construction (via latent variables) simple statistical testing cannot provide a satisfactory answer. Fortunately, we can use exactly the same

Figure 2: Model posterior probabilities in stage two



Note: top-left chart: models without labor prices (scenario one); top-right chart: models with labor prices (scenario two); bottom charts: comparison between SFA and non-SF models in scenario one (left) and scenario two (right).

Bayesian techniques to extend the current framework in a way to provide a straightforward, probabilistic answer to this research question. Each SFA model considered in stage two has been assigned its simplified "non-SF" counterpart with no latent variables representing inefficiency differences; see Section 2. In those models $u_i = 0$, which means that cost efficiency $exp(-u_i)$ is equal one by default for all objects. Hence, these models assume, by construction, that all objects are relatively equally efficient and thus there is no inefficiency component, i.e., no heterogeneity as regards cost-effectiveness. This yields additional 15 models to consider. Fortunately, since these Bayesian models are relatively simple, their marginal data density computation is trivial to calculate in comparison to Bayesian SFA models.

The idea behind this procedure is quite simple. If under equal prior model probabilities non-SF models obtain significantly higher posterior probability mass in comparison to SFA models, then indeed differentiating cost-effectiveness of the business units may, or even should be questioned. Furthermore, we can pool inference on cost efficiency also from non-SF models, which assume that all objects are equally (and thus relatively fully) efficient. The final cost efficiency ranking built this way not only takes into account uncertainty of the model as regards the choice of explanatory variables but also uncertainty of modelling efficiency variation in general.

Figure 2 shows posterior probabilities of each model in the first scenario (no wages, top-left chart). Model 10 is still the most likely one (its posterior probability being 0.713) with a new

Table 5: Posterior means and standard deviations of cost efficiency of DSO's business units based on inference pooling

Unit	Scenario one: without labor price information				
	Only SFA models		SFA and non-SF		Rank
	mean	std	mean	std	
A	0.961	0.034	0.964	0.035	1
B	0.960	0.036	0.964	0.036	2
C	0.948	0.047	0.952	0.048	3
D	0.855	0.060	0.866	0.069	4
E	0.843	0.090	0.855	0.096	5
F	0.840	0.058	0.852	0.070	6
Average	*0.903*	*0.059*	*0.910*	*0.063*	
Range	*0.121*		*0.112*		
Unit	**Scenario two: with labor price information**				
	Only SFA models		Only SFA models		Rank
	mean	std	mean	std	
A	0.965	0.033	0.986	0.027	1
B	0.954	0.042	0.981	0.035	2
C	0.947	0.048	0.978	0.041	3
D	0.875	0.058	0.949	0.072	4
E	0.849	0.090	0.938	0.094	6
F	0.858	0.062	0.942	0.080	5
Average	*0.906*	*0.061*	*0.962*	*0.061*	
Range	*0.116*		*0.047*		

Note: Results obtained in stage two; "mean" is the posterior mean, "std" is posterior standard deviation; "Only SFA models" are results from inference pooling only based on SFA models; "SFA and non-SF" are final results based on inference pooling that takes both specifications (SFA and non-SF) into account.

SFA model 12 (simplification of model 7) taking most of the remaining probability mass. Simpler non-SF Bayesian models are clearly not favored by the data in this scenario. The sum of their probability mass is just 0.077 whereas SFA models take as much as 0.923; see Figure 2, bottom-left chart (scenario one). As a result non-SF models have little effect on inference pooling and the resulting efficiency scores remain virtually unchanged (Table 5).

Inclusion of non-SF specification makes a difference in the second scenario, in which information on observed wages has been used. Posterior probability is 0.59 for non-SF models vs. 0.41 in favor of SFA specifications; see Figure 2, bottom-right chart (scenario two). In this case three models clearly dominate the ranking, two of which are non-SF (Figure 2, top-right chart). Table 5 shows the final efficiency ranking in the second scenario that takes into account both SFA and non-SF models by weighting the scores based on posterior model probabilities. We see that as a result the final scores are significantly closer to one (i.e., full relative efficiency) and the difference between the efficiency leader and the least efficient unit is even smaller than it has been previously anticipated in stage one.

Conceptual differences between the two scenarios—models with and without wages—reveal that taking the observed wages as a justified determinant of variable cost significantly decreases differences in cost efficiency between DSO's business units (variation of cost (in)efficiency). At the same time it also diminishes the importance of SFA specification as its posterior probability declines from 0.92 to 0.41. However, including wages has not marginalized SFA models entirely and it does not impact the order of efficiency ranking.

Table 6: Comparison of cost efficiency estimates in four models

	Models			
	10	**35**	**2**	**3**
10	1	0.994	0.743	0.729
35	0.994	1	0.740	0.715
2	0.743	0.740	1	0.977
3	0.729	0.715	0.977	1
mean	0.888	0.897	0.859	0.869
p.std	0.037	0.037	0.063	0.059
max	0.953	0.955	0.917	0.910
min	0.798	0.819	0.766	0.814
std	0.056	0.048	0.043	0.029

Note: Upper part contains empirical correlation coefficients between efficiency estimates; mean is the posterior mean of average efficiency $exp(-\varphi^{-1})$; p.std is the posterior standard deviation of average efficiency; max/min values denote the highest and the lowest posterior mean of efficiency in the sample; std is an empirical standard deviation of the posterior means in a given model.

4.3. Cost efficiency comparison

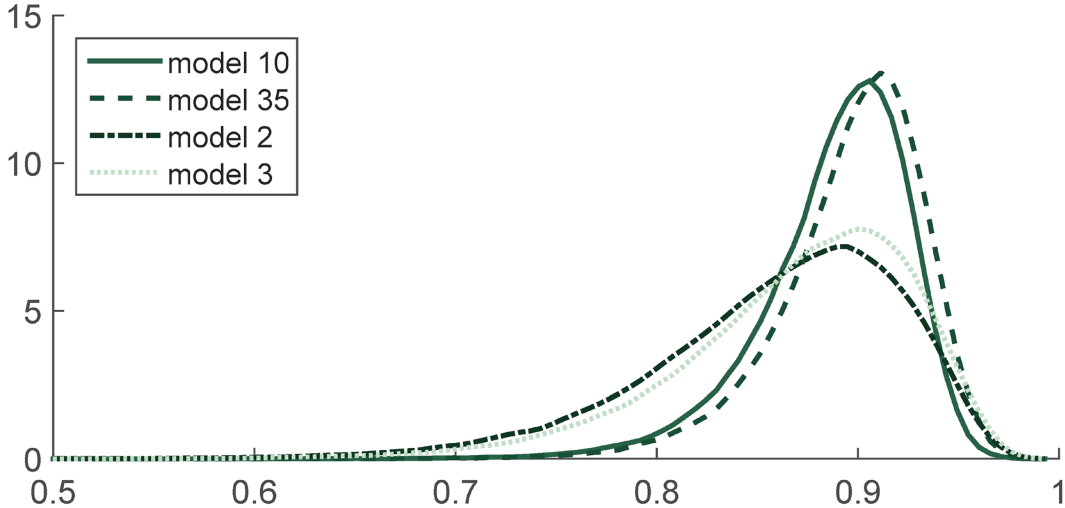
The choice of explanatory variables is likely to have an impact on efficiency analysis. In this section we compare differences between estimated cost efficiencies for the best two SFA models from stage two (model 10—scenario one, model 35—scenario two) and models used previously by Polish ERO (models 2, 3; see Osiewalski and Wróbel-Rotter, 2008–9, 2012).

In a non-Bayesian literature it is customary to compare estimates of efficiency from different models in terms of their averages, empirical standard deviations and correlations (Table 6). If we were to do that as well, we would conclude that the average efficiency is not that much higher in the "best" models and that empirical standard deviation is smaller in models 2 and 3. Furthermore, Table 6 also shows empirical correlation coefficients between point estimates of individual efficiency scores (posterior means) from models 2, 3, 10 and 35. We see that efficiencies are well (0.729) or even perfectly (0.994) correlated. Clearly models that are similar in terms of explanatory variables lead to highly correlated efficiency estimates; compare pairs of models (2, 3) and (10, 35).

Given the above, one may conclude that the previously used models (2, 3) are as good or even better at explaining the data because (i) they leave less cost variation to be explained by latent variables and (ii) their efficiency estimates are relatively well correlated with others. This, however, is not entirely correct. Figure 3 shows posterior distributions of $exp(-\varphi^{-1})$, which can be interpreted as average efficiency in models 2, 3, 10 and 35. Indeed, posterior means of this parameter are not that far away from each other. However, the posterior distributions in models 2 and 3 are much more distorted and the tails are pulled significantly more towards lower values. This shows that even though these models use more explanatory variables there is considerably more uncertainty as regards inference about efficiency.

To sum up, we should outline two important aspects. First, efficiencies may differ substantially between models. This is not directly evident when analyzing efficiency averages, spread or empirical correlations but it becomes clear once we examine the posterior distributions of efficiency scores. Second, although the previously used models (2, 3) have turned out to be highly inadequate

Figure 3: Posterior distributions of average efficiency *exp*($-\varphi^{-1}$) in models 2, 3, 10 and 35



in this application (very low posterior probabilities) it should not be viewed as a mistake of previous Polish DSO studies. These studies were based on different objects and different time frames. We only wish to point out that model "copy-pasting" does not seem to work in the case of electricity distribution sector and that appropriate list of cost determinants is likely to be dependent on a particular application.

5. CONCLUDING REMARKS

Efficiency estimates in SFA are not only dependent on the data, but also, to large extent, on model specification. That is why the focal point of this research has been to develop a framework in which the model specification itself is also determined based on information in the data. The cost efficiency benchmarking framework presented in the paper is based on a Bayesian approach to stochastic frontier analysis—BSFA. This allows us to incorporate formal Bayesian techniques of model selection and inference pooling to mitigate the problem of model selection uncertainty when analyzing cost efficiency of Distribution System Operators (DSOs). This is an especially important aspect for electricity distribution sector where proper model selection can be particularly difficult. However, the proposed approach can also be applied to other energy services utilities whenever performance oriented regulations are set as well as to benchmarking and productivity analysis in other energy-related issues (e.g., standard production function-based technical efficiency or energy efficiency using demand SFA as in Filippini and Hunt, 2011, 2012; Llorca et al., 2017).

As shown in the empirical section, model choice can have a considerable impact on efficiency ranking. If the regulator makes a bad choice, it may inappropriately penalize efficient DSOs. Moreover, it may very well be that efficiency analysis is not warranted by the data if the hypothesis of equal (full) relative efficiency cannot be rejected. The presented approach makes it possible to compare numerous competing model specifications not only in terms of optimal choice of cost determinants for the cost function, but also to formally judge uncertainty of a given SFA specification (time dynamics, inefficiency constant or varying over time—as in, e.g., Das, 2015; frontier heterogeneity etc.). All of these aspects can be incorporated into the final, pooled inference.

The two scenarios discussed in Section 4 have allowed us to investigate the consequence of including price of labor as a justified determinant of variable cost differences. We have found that although information on average wages in the DSO's business units does not significantly impact the order of the efficiency ranking, it does decrease differences between cost efficiency scores of the analyzed objects and makes the SF specifications less likely. We should stress, however, that the two scenarios are fundamentally different as regards efficiency interpretation. The first scenario assumes that DSOs are territorial monopolies and the management has, at least theoretically, control over price of labor creation. Of course, the regulator may still want to analyze the impact of wage differences on DSO's cost-effectiveness, but then wages should be treated as potential efficiency determinants. The regulator can use Bayesian Varying Efficiency Distribution models (Koop et al., 1997), which generalize the normal-exponential SF model used in this paper or other, non-Bayesian, approaches summarized in Llorca et al. (2016). The second scenario assumes that DSOs pay a fair "market price" in their region and thus wage differences should be treated as a justified determinant of variable costs, not manifest itself in cost efficiency scores. Which scenario is more suitable for a given energy distribution sector is up to the regulator to decide. We do note, however, that the choice is likely to have an impact on the posterior probability of SFA specification, as well as the spread of efficiency ranking.

The presented approach makes the benchmarking process not only more transparent but also much more inclusive. All participants of the benchmarking study, the regulator, the DSOs, as well as the benchmarking experts can propose models they think best describe the data. All these models can be included to obtain the final results through inference pooling and the contribution each model makes is determined by its posterior probability. A given model's posterior probability answers a simple, yet important question: how likely the model is in view of the data. And of course, the best models contribute the most. It takes the model building process away from the stakeholders (the regulator, DSOs, the experts), which may have different agendas and "push" their preferences. Theoretically, should there be such a need, every possible model specification can be considered in a study. Thus, from a technical viewpoint, there is no need for any model proposals or "stages" to be made by the DSOs or the regulator. Considering model proposition, however, is much more practical due to the numerical effort needed to estimate all possible scenarios. For example, in this study, one would need to estimate 17,179,869,184 Bayesian SF models along with their marginal data density values. It is important to note that operational techniques of Bayesian model comparison and inference pooling have been already elaborated for regression models with dozens of potential explanatory variables; such models and techniques appear in empirical studies of economic growth (Fernández et al., 2001a,b; Ley and Steel, 2009). However, these techniques are based on a particular structure of the prior distribution for the parameters and on the Normality assumption for the error term of the regression equation. While the form of the prior distribution is relatively easy to change in our cost efficiency framework, the assumption of Normality is simply inadequate as it excludes SF models from the start (SF models have composed errors). Further research is needed to generalize the BMA techniques proposed in the literature.

It is worth noting that Bayesian model selection and inference pooling have also turned out to be very useful because they vividly show one important aspect—quality trumps quantity. We have found that no matter how many models experts from the regulator or the DSOs propose, no more than a few significantly contribute to the final results. This effectively discourages the participants from researching models, which would provide more "desirable" results (in their view) because such policy-driven models practically marginalize themselves through very low posterior probabilities.

ACKNOWLEDGMENTS

This paper is a theoretical advancement of an earlier, practical-oriented study commissioned by one of Distribution System Operators in Poland, based on a new methodology adopted by the Polish Energy Regulatory Office in 2015 (see Polish Energy Regulatory Office, 2015ab). We would like to thank the DSO's management for allowing us to partially disclose the main empirical findings and its team of experts for their valuable contribution. We are grateful to the editor and four anonymous referees for their valuable guidelines and suggestions. Our thanks also go to Mark Hoffman for his help in preparing the final version of the paper and to Per Agrell for his encouraging comments. The authors acknowledge support from research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

REFERENCES

Agrell, P. and P. Bogetoft (2007). *Development of benchmarking models for German electricity and gas distribution. SUMICSID Consultation Report.* http://www.sumicsid.se/web-files/docs/projects/SUMICSID_Report_Distribution_Regulation_2007.pdf.

Agrell, P. and P. Bogetoft (2009). *International Benchmarking of Electricity Transmission System Operators*. *The e3GRID project final report*. http://www.acm.nl/nl/download/bijlage/?id=8813.

Agrell, P., P. Bogetoft and J. Tind (2005). "DEA and Dynamic Yardstick Competition in Scandinavian Electricity Distribution." *Journal of Productivity Analysis* 23(2): 173–201. https://doi.org/10.1007/s11123-005-1327-6.

Agrell, P. and E. Grifell-Tatje (2016). "A dynamic model for firm-response to non-credible incentive regulation regimes." *Energy Policy* 90: 287–299. https://doi.org/10.1016/j.enpol.2015.12.029.

Agrell, P. and J. Teusch (2015). "Making the Belgian distribution system fit for the energy transition." *Reflets et Perspectives* 54(1): 157–174. https://doi.org/10.3917/rpve.541.0157.

Bogetoft, P. (2012). *Performance Benchmarking: Measuring and Managing Performance*. New York: Springer. https://doi.org/10.1007/978-1-4614-6043-5.

Broeck (van den), J., G. Koop, J. Osiewalski and M.F.J. Steel (1994). "Stochastic frontier models: A Bayesian perspective." *Journal of Econometrics* 61(2): 273–303. https://doi.org/10.1016/0304-4076(94)90087-6.

Christensen, L.R., D.W. Jorgenson and L.J. Lau (1971). "Conjugate duality and the transcendental logarithmic production function." *Econometrica* 39(4): 255–256.

Coelli, T., A. Estache, S. Perelman and L. Trujillo (2003). *A Primer on Efficiency Measurement for Utilities and Transport Regulators.* Washington, D.C.: World Bank Publications. http://hdl.handle.net/10986/15149. https://doi.org/10.1596/0-8213-5379-9.

Cullmann, A. and C. von Hirschhausen (2008a). "Efficiency analysis of East European electricity distribution in transition: legacy of the past?" *Journal of Productivity Analysis* 29: 155–167. https://doi.org/10.1007/s11123-007-0075-1.

Cullmann, A. and C. von Hirschhausen (2008b). "From transition to competition. Dynamic efficiency analysis of Polish electricity distribution companies." *Economics of Transition* 16(2): 335–357. https://doi.org/10.1111/j.1468-0351.2008.00312.x.

Das, A. (2015). "Copula-based stochastic frontier model with autocorrelated inefficiency." *Central European Journal of Economic Modelling and Econometrics* 7(2): 111–126.

Diewert, W.E. (1971). "An application of the Shephard duality theorem: a Generalized Leontief production function." *Journal of Political Economy* 79(3): 481–507. https://doi.org/10.1086/259764.

Diewert, W.E. and T.J. Wales (1987). "Flexible functional forms and global curvature conditions." *Econometrica* 55(1): 43–68. https://doi.org/10.2307/1911156.

Dimitropoulos, D. and A. Yatchew (2017). "Is Productivity Growth in Electricity Distribution Negative? An Empirical Analysis Using Ontario Data." *The Energy Journal* 38(2): 175–200. https://doi.org/10.5547/01956574.38.2.ddim.

Farsi, M. and M. Filippini (2004). "Regulation and measuring cost efficiency with panel data models application to electricity distribution utilities." *The Review of Industrial Organization* 25(1): 1–19. https://doi.org/10.1023/B:REIO.0000040474.83556.54.

Farsi, M. and M. Filippini (2009). "An analysis of cost efficiency in Swiss multi-utilities." *Energy Economics* 31(2): 306–315. https://doi.org/10.1016/j.eneco.2008.11.009.

Farsi, M., M. Filippini and W.H. Greene (2006). "Application of panel data models in benchmarking analysis of the electricity distribution sector." *Annals of Public and Cooperative Economics* 77(3): 271–290. https://doi.org/10.1111/j.1467-8292.2006.00306.x.

Fernández, C., G. Koop and M.F.J Steel (2000). "A Bayesian analysis of multiple-output production frontiers." *Journal of Econometrics* 98(1): 47–79. https://doi.org/10.1016/S0304-4076(99)00074-3.

Fernández, C., G. Koop and M.F.J Steel (2005). "Alternative efficiency measures for multiple-output production." *Journal of Econometrics* 126(2): 411–444. https://doi.org/10.1016/j.jeconom.2004.05.008.

Fernández, C., E. Ley and M.F.J. Steel (2001a). "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* 100(2): 381–427. https://doi.org/10.1016/S0304-4076(00)00076-2.

Fernández, C., E. Ley and M.F.J. Steel (2001b). "Model uncertainty in cross-country growth regressions." *Journal of Applied Econometrics* 16(5): 563–576. https://doi.org/10.1002/jae.623.

Fernández, C., J. Osiewalski and M.F.J. Steel (1997). "On the use of panel data in stochastic frontier models with improper priors." *Journal of Econometrics* 79(1): 169–193. https://doi.org/10.1016/S0304-4076(97)88050-5.

Filippini, M., W.H. Greene and G. Masiero (2016). "Persistent and transient productive inefficiency in a regulated industry: electricity distribution." *Energy Economics* 69: 325—334. https://doi.org/10.1016/j.eneco.2017.11.016.

Filippini, M. and L. Hunt (2011). "Energy demand and energy efficiency in the OECD countries: A stochastic demand frontier approach." *The Energy Journal* 32(2): 59–80. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol32-No2-3.

Filippini, M. and L. Hunt (2012). "US residential energy demand and energy efficiency: A stochastic demand frontier approach." *Energy Economics* 34(5): 1484–1491. https://doi.org/10.1016/j.eneco.2012.06.013.

Finnish Energy Market Authority (2016). *National Report 2016 to the Agency for the Cooperation of Energy Regulators and to the European Commission.* Ref: 1518/601/2016 8.7.2016.

Finon, D. and F. Roques (2013). "European Electricity Market Reforms: The 'Visible Hand' of Public Coordination." *Economics of Energy & Environmental Policy* 2(2): 107–124. https://doi.org/10.5547/2160-5890.2.2.6.

Greene, W.H. (2005a). "Reconsidering heterogeneity in panel data estimators of the stochastic frontier model." *Journal of Econometrics* 126(2): 269–303. https://doi.org/10.1016/j.jeconom.2004.05.003.

Greene, W.H. (2005b). "Fixed and random effects in stochastic frontier models." *Journal of Productivity Analysis* 23(1): 7–32. https://doi.org/10.1007/s11123-004-8545-1.

Greene, W.H. (2008a). *Econometric Analysis* (Sixth Edition). Upper Saddle River, N.J.: Pearson-Prentice Hall.

Greene, W.H. (2008b). "The Econometric Approach to Efficiency Analysis." In H.O. Fried et al., eds., *The Measurement of Productive Efficiency and Productivity Growth*. New York: Oxford University Press. pp. 92–159. https://doi.org/10.1093/acprof:oso/9780195183528.003.0002.

Hicks, J.R. (1935). "Annual survey of economic theory: The theory of monopoly." *Econometrica* 3(1): 1–20. https://doi.org/10.2307/1907343.

Hirschhausen (von), C., A. Cullmann and A. Kappeler (2006). "Efficiency analysis of German electricity distribution utilities—non-parametric and parametric tests." *Applied Economics* 38(21): 2553–2566. https://doi.org/10.1080/00036840500427650.

Jamasb, T. and M. Pollitt (2001). "Benchmarking and Regulation: International Electricity Experience." *Utilities Policy* 9(3): 107–130. https://doi.org/10.1016/S0957-1787(01)00010-8.

Jamasb, T. and M. Pollitt (2003). "International Benchmarking and Yardstick Regulation: An Application to European Electricity Distribution Utilities." *Energy Policy* 31(15): 1609–1622. https://doi.org/10.1016/S0301-4215(02)00226-4.

Jamasb, T. and M. Pollitt (2005). "Electricity Market Reform in the European Union: Review of Progress toward Liberalization & Integration." *The Energy Journal* 26: 11–41. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol26-NoSI-2.

Koop, G., J. Osiewalski and M.F.J. Steel (1994a). "Bayesian efficiency analysis with a flexible form: The AIM cost function." *Journal of Business and Economic Statistics* 12(3): 339–346.

Koop, G., Osiewalski J. and M.F.J. Steel (1994b). "Hospital efficiency analysis through individual effects: A Bayesian approach." *CentER Discussion Paper 9447*.

Koop, G., J. Osiewalski J. and M.F.J. Steel (1997). "Bayesian efficiency analysis through individual effects: Hospital cost frontiers." *Journal of Econometrics* 76(1–2): 77–105. https://doi.org/10.1016/0304-4076(95)01783-6.

Koop, G., J. Osiewalski and M.F.J. Steel (1999). "The components of output growth: A stochastic frontier analysis." *Oxford Bulletin of Economics and Statistics* 61(4): 455–487. https://doi.org/10.1111/1468-0084.00139.

Kopsakangas-Savolainen, M. and R. Svento (2008). "Estimation of cost-effectiveness of the Finnish electricity distribution utilities." *Energy Economics* 30(2): 212–229. https://doi.org/10.1016/j.eneco.2007.07.004.

Kumbhakar, S. and G. Lien (2017). "Yardstick Regulation of Electricity Distribution—Disentangling Short-run and Long-run Inefficiencies." *The Energy Journal* 38(5): 17–37. https://doi.org/10.5547/01956574.38.5.skum.

Kuosmanen, T. (2012). "Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model." *Energy Economics* 34(6): 2189–2199. https://doi.org/10.1016/j.eneco.2012.03.005.

Lenk, P. (2009). "Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods." *Journal of Computational and Graphical Statistics* 18(4): 941–960. https://doi.org/10.1198/jcgs.2009.08022.

Ley, E. and M.F.J. Steel (2009). "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions." *Journal of Applied Econometrics* 24(4): 651–674. https://doi.org/10.1002/jae.1057.

Llorca, M., J. Banos, J. Somoza and P. Arbues (2017). "A Stochastic Frontier Analysis Approach for Estimating Energy Demand and Efficiency in the Transport Sector of Latin America and the Caribbean." *The Energy Journal* 38(5): 153–174. https://doi.org/10.5547/01956574.38.5.mllo.

Llorca, M., L. Orea and M. Pollitt (2016). "Efficiency and environmental factors in the US electricity transmission industry." *Energy Economics* 55: 234–246. https://doi.org/10.1016/j.eneco.2016.02.004.

Magnus, J.R. (1979). "Substitution between energy and non-energy inputs in the Netherlands 1950-1976." *International Economic Review* 20(2): 465–484. https://doi.org/10.2307/2526494.

Makieła, K. (2014). "Bayesian stochastic frontier analysis of economic growth and productivity change in the EU, USA, Japan and Switzerland." *Central European Journal of Economic Modelling and Econometrics* 6(3): 193–216.

Makieła, K. (2017). "Bayesian Inference and Gibbs Sampling in Generalized True Random-Effects Models." *Central European Journal of Economic Modelling and Econometrics* 9(1): 69–95.

Makridoua, G., K. Andriosopoulosb, M. Doumpos and C. Zopounidis (2015). "A Two-Stage Approach For Energy Efficiency Analysis In European Union Countries." *The Energy Journal* 36(2): 47–69. https://doi.org/10.5547/01956574.36.2.3.

Marzec, J. and J. Osiewalski (2008). "Bayesian Inference on Technology and Cost Efficiency of Bank Branches." *Bank i Kredyt* 39(9): 29–43.

Newton, M. and A. Raftery (1994). "Approximate Bayesian inference with the weighted likelihood bootstrap." *Journal of the Royal Statistical Society (Series B)* 56(1): 3–48.

Nillesen, P. and M. Pollitt (2007). "The 2001–3 electricity distribution price control review in the Netherlands: regulatory process and consumer welfare." *Journal of Regulatory Economics* 31(3): 261–287. https://doi.org/10.1007/s11149-006-9014-8.

Osiewalski, J. and K. Osiewalski (2016). "Hybrid MSV-MGARCH models—general remarks and the GMSF-SBEKK specification." *Central European Journal of Economic Modelling and Econometrics* 8(4): 241–271.

Osiewalski, J. and M.F.J. Steel (1993). "Una perspectiva bayesiana en selección de modelos." *Cuadernos Economicos* 55(3): 327–351. http://www.cyfronet.krakow.pl/~eeosiewa/pubo.htm.

Osiewalski, J. and M.F.J. Steel (1998). "Numerical tools for the Bayesian analysis of stochastic frontier models." *Journal of Productivity Analysis* 10(1): 103–117. https://doi.org/10.1023/A:1018302600587.

Osiewalski, J. and R. Wróbel-Rotter (2008–2009). "Bayesowskie graniczne funkcje kosztu dla sektora dystrybucji energii [Bayesian frontier cost functions for the electricity distribution sector; in Polish]." *Folia Oeconomica Cracoviensia* 49–50: 47–69.

Osiewalski, J. and R. Wróbel-Rotter (2012). "Model ekonometryczny—analiza kosztów operacyjnych w sektorze dystrybucji energii [Econometric model—operational costs analysis in energy distribution sector; in Polish]." *Bulletin of Energy Regulations Office* 79: 9–23.

Osiewalski, K. and J. Osiewalski (2013). "A long-run relationship between daily prices on two markets: the Bayesian VAR(2)–MSF-SBEKK model." *Central European Journal of Economic Modelling and Econometrics* 5(1): 65–83.

Pajor, A. (2017). "Estimating the marginal likelihood using the arithmetic mean identity." *Bayesian Analysis* 12(1): 261–287. https://doi.org/10.1214/16-BA1001.

Pajor, A. and J. Osiewalski (2013–2014). "A note on Lenk's correction of the harmonic mean estimator." *Central European Journal of Economic Modelling and Econometrics* 5(4): 271–275; correction: 6(1): 69.

Polish Energy Regulatory Office (2015a). *Koszty operacyjne dla Operatorów Systemów Dystrybucyjnych na lata 2016–2020 [Operational Costs for Distribution System Operators in 2016–2020; in Polish]. Staff Report.* http://bip.ure.gov.pl/download/3/6432/OPEX2016-2020OSD.pdf.

Polish Energy Regulatory Office (2015b). *Strategia Regulacji Operatorów Systemów Dystrybucyjnych na lata 2016–2020* [*Strategy for Regulating Distribution System Operators in 2016–2020; in Polish*]. *Staff Report.* http://bip.ure.gov.pl/download/3/6435/STRATEGIAREGULACJIOSDostateczna.pdf.

Simar, L. and P. Wilson (2008). "Statistical Inference in Nonparametric Frontier Models." In H.O. Fried et al., eds., *The Measurement of Productive Efficiency and Productivity Growth*. New York: Oxford University Press. pp. 421–521.

Tsionas, M. and S.C. Kumbhakar (2014). “Firm Heterogeneity, Persistent And Transient Technical Inefficiency: A Generalized True Random-Effects model.” *Journal of Applied Econometrics* 29(1): 110–132. https://doi.org/10.1002/jae.2300.

Wróbel-Rotter, R. (2007). “Bayesowskie porównywanie giętkich form funkcyjnych jako granicznych funkcji kosztu [The Bayesian comparison of flexible functional forms as stochastic cost frontiers, in Polish].” *Przegląd Statystyczny [Statistical Review]* 54(2): 5–27.