Cost Efficiency Analysis of Electricity Distribution Sector under Model Uncertainty

Kamil Makieła, Jacek Osiewalski

(Department of Econometrics and Operations Research, Cracow University of Economics, Poland)

Plan of the presentation:

- motivation, aim and other aspects
- Bayesian stochastic frontier analysis (BSFA) of the Distribution System Operator (DSO) cost efficiency
- Bayesian model selection and inference pooling using Corrected Arithmetic Mean Estimators (CAME) of the marginal data density values
- an empirical study of cost models and cost efficiency of business units of a Polish DSO
- concluding remarks

Motivation

- electricity distribution is a key sector of an economy
- distribution system operators of electricity networks (DSOs) are natural, usually territorial, monopolies
- national energy regulatory agencies try to persuade DSOs to improve their performances
- this calls for an objective and transparent DSO benchmarking framework
- many aspects need to be considered in a benchmark study (Bogetoft, 2012); we summarize them as:

    i) the benchmark (modelling) framework,

    ii) estimation approach,

    iii) exact model specification (choice of variables in particular).

Aim of the study

Within the benchmark framework of cost-effectiveness (used in Poland) and Bayesian stochastic frontier approach (BSFA), we propose a fair model choice process, in which validity of a given specification is formally judged based on information in the data (through posterior model probability distribution)

Other aspects of our cost efficiency analysis in electricity distribution:

- two model scenarios:
    1. DSOs have control over prices,
    2. DSOs can be treated as price takers

Bayesian Stochastic Frontier Analysis used by the Polish Energy Regulatory Office

Basic SFA model for panel data on y_{it} – the log cost, with Cobb-Douglas or translog form:

$$y_{it} = x'_{it}\beta + \varepsilon_{it} = x'_{it}\beta + u_i + v_{it}$$

x'_{it} – a k-element vector of explanatory variables (logs of: outputs, fixed assets, prices),

β – a vector of model parameters,

v_{it} – "standard" symmetric random disturbance with mean 0,

u_i – nonnegative random individual effect, which reflects inefficiency

Bayesian version of the Normal-Exponential SFA model (Koop, Osiewalski and Steel, 1997):

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

where

$$p(\beta,\sigma_v^{-2},\varphi) = p(\beta)p(\sigma_v^{-2})p(\varphi) = f_N(\beta|b,\Sigma)f_G(\sigma_v^{-2}|0.5n_0,0.5a_0)f_G(\varphi|1,-\ln(r_0)),$$

r_0 – prior median of efficiency $exp(-u_i)$, other hyper-parameters reflect weak prior information

$p(\beta,\sigma_v^{-2},u,\varphi|y,X) \propto p(y,\beta,\sigma_v^{-2},u,\varphi|X)$ – non-standard posterior, simulated using MCMC

Model Comparison and Inference Pooling in BSFA

Consider m Bayesian models:

$$M_j\colon\ p_j(y,\theta_j) = p_j(y|\theta_j)p_j(\theta_j); \quad j = 1,\dots,m;$$

$y \in Y$ – matrix of observations being modelled (observed DSOs costs in logs)

θ_j – vector of unobserved quantities (model parameters and inefficiencies) made up of λ (quantities common to all models) and η_j (quantities specific M_j); $\theta_j = (\lambda,\eta_j) \in \Theta_j = \Lambda \times H_j$

Model comparison is based on posterior model probabilities calculated using the Bayes formula:

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

where

$$p(y|M_j) = p_j(y) = \int_{\Theta_j} p_j(y|\theta_j)p_j(\theta_j)d\theta_j$$

is the marginal data density (MDD) value in model M_j.

Using posterior model probabilities

The regulator can:

i) select for DSOs cost efficiency analysis one particular model with the highest posterior probability (Bayesian model choice or selection, 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) – Bayesian inference pooling (or model averaging), appealing when none model dominates:

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

To perform (i) or (ii) the regulator needs to know: prior probabilities – $p(M_j)$ and **MDD** values – $p_j(y)$:

- examples of prior model probability distributions (to be assigned by the regulator):

$p(M_j) = \frac{1}{m}$ – all models are viewed equally likely *a priori*

$p(M_j) \propto 2^{-l_j}$ – parsimonious models are more likely *a priori* (Ockham Razor rule)

- **main task:** to approximate MDD (separately for each model) within posterior MCMC simulation – corrected arithmetic mean estimator (Pajor 2017).

Cost efficiency analysis of a DSO in Poland

Problems with explanatory variables for our short-run variable cost model:

- DSO provided information on 34 characteristics describing differences in its TBUs operation in terms of scale, technical properties, quality and observed wages;
- many characteristics are highly correlated with each other for obvious technical reasons (e.g., number of electric energy converters and number of electric stations) – variables carry similar information;
- main goal of this study – to find models (i.e. sets of explanatory variables) relevant in describing VC and to make pooled inference about short-run cost efficiency of DSO's business units.

Cost efficiency analysis of a DSO in Poland

Two modeling scenarios:

1. Information about wages is excluded – TBUs have control over the price of hired labor so wages do not exogenously determine cost; market price is difficult to determine due to territorial monopoly;
2. Information about wages is included – TBUs are “price takers”, wage is a justified determinant of costs.

Comparative analysis of 1. and 2. also 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).

Stage One of BSFA

37 models, 10 for scenario 1 (no wages) and 27 for scenario 2 (wages included)

Design of Stage One:

- models proposed by various participants involved in the study – authors and the DSO experts;
- based on various sources, (i) authors experience, (ii) DSO’s managerial expertise, (iii) desk research, (iv) purely statistical analysis of the data (e.g., PCA, statistical tests and reductions of the full model);
- <u>variety of sources intentional</u> – to make the model proposition process inclusive for the DSO experts;
- all models represent short-run variable cost function (with at least one variable for output and capital assets);

Stage One of BSFA

37 models, 10 for scenario 1 (no wages) and 27 for scenario 2 (wages included)

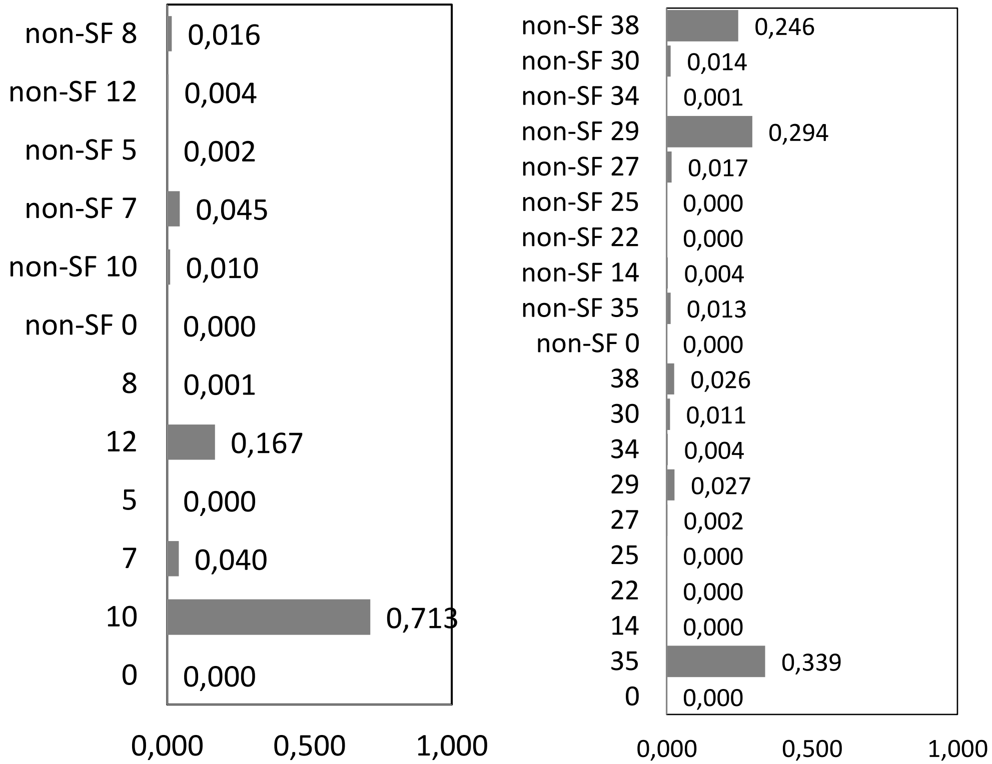
Results from Stage One:

- high MDD values only for a few parsimonious models with 5-6 explanatory variables (!!!);
- best models show high cost efficiency scores despite small number of variables and parameters (!!!);
- scenario 1: posterior model probability ranking dominated by model 10 with 5 variables: total EHV and HV line length per one circuit (x1), total LV overhead line length per one circuit + length of LV overhead service wires (x5), number of HV substations (x7), inverse of System Average Interruption Frequency Index (x24), volume of electricity supplied to the LV customer group (x32);
- scenario 2: model 38 (5 variables + average wage) takes a large portion of the posterior probability mass, but there are other models (22, 25, 27, 29, 30, 34, 35) that cannot be entirely ignored;
- model extensions (translog cost function, trend variable, time-varying efficiency) – unnecessary, lead to lower MDD and posterior model probability.

Stage Two of BSFA

15 models, 5 for scenario 1 (no wages) and 10 for scenario 2 (wages included)

- Stage Two builds on results from Stage One – inadequate models removed, new ones added;
- Stage Two extends analytical framework – also full efficiency (non-SF) models;

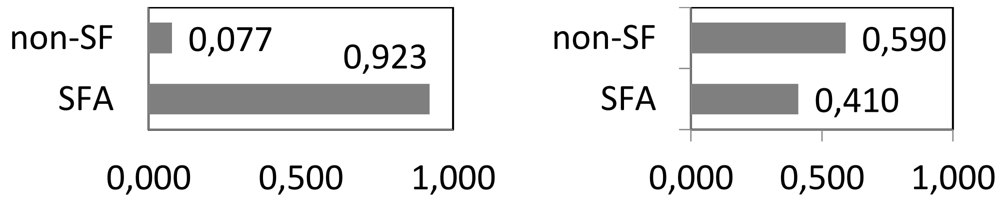


Posterior model probabilities
left box: no labor prices (scenario 1), right box: with labor prices (scenario 2)

Posterior means and standard deviations of cost efficiency

Based on inference pooling, Stage Two

	Scenario one: without labor price information					Scenario two: with labor price information				
DSO Unit	Only SFA models		SFA and non-SF		Rank	Only SFA models		SFA and non-SF		Rank
	mean	std	mean	std		mean	std	mean	std	
Average	*0.903*	*0.059*	*0.910*	*0.063*		*0.906*	*0.061*	*0.962*	*0.061*	
Range	*0.121*		*0.112*			*0.116*		*0.047*		



Posterior probabilities for SF and non-SF model groups scenario 1 (left) and scenario 2 (right)

Concluding remarks

- efficiency estimates in SFA depend not only on data, but also on model specification – thus we develop a framework in which model specification itself is determined based on data information
- our cost efficiency benchmarking framework uses BSFA, which incorporates formal Bayesian techniques of model selection and inference pooling (important in electricity distribution sector where model specification can be very difficult)
- the proposed approach can be used in other energy distribution sectors (whenever performance-based regulatory policy applies) as well as to other benchmarking criteria
- two scenarios show important consequences of including/excluding wages in a cost function
- our approach makes benchmarking process much more credible, transparent and inclusive; all participants of the study (regulator, DSOs, experts ...) can propose their models
- our approach does not consider all models (given possible explanatory variables); such Bayesian operational techniques exist, but for regression models with normally distributed error terms; further research is needed to extend such techniques to non-normal (composed error) SF models
- Bayesian model selection and inference pooling show that only a few models contribute to final results – (model) quality trumps quantity (measured by the number of models)

Thank you. Questions?