The effect of restructuring electricity distribution systems on firms' persistent and transient efficiency: The case of Germany ☆

Oleg Badunenko[a], Astrid Cullmann[b,c], Subal C. Kumbhakar[d], Maria Nieswand[e,*]

[a]*Economics and Finance Group, Portsmouth Business School, University of Portsmouth, Portsmouth, PO1 3DE, UK*

[b]*Department of Econometrics and Statistics, TU Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany*

[c]*German Institute for Economic Research, Mohrenstraße 58, 10117 Berlin, Germany*

[d]*Department of Economics, State University of New York, Binghamton, USA*

[e]*School of Business and Economics, Loughborough University, Loughborough, LE11 3TU, UK*

Abstract

We evaluate the efficiency of electricity distribution operators (DSOs) as providers of local public infrastructure. In particular, we consider two types of efficiency, i.e., short-term (transient) and long-term (persistent). We apply the recently developed four-component stochastic frontier model, which allows identifying determinants of the two types of efficiency, after controlling for firm heterogeneity and random noise, to a panel dataset of German DSOs observed during 2006-2012. Those DSOs operating in the eastern parts of Germany have undergone a profound restructuring after the reunification in 1990. We find that this was beneficial for their efficiency as they perform, on average, better in terms of persistent efficiency than DSOs in West Germany. Both eastern and western DSOs perform similarly well in terms of transient efficiency, which is expected as the sector is highly regulated. As such, we provide new insights on identifying the nature and sources of public infrastructure productive inefficiency, which is relevant for public policies.

Keywords: Production, persistent efficiency, transient efficiency, stochastic frontier model, determinants of efficiency

---

☆ The research reported in this paper is supported by KOMIED (Municipal infrastructure companies against the background of energy policy and demographic change) financed by Leibniz Association.

* Corresponding author.

Email addresses: oleg.badunenko@port.ac.uk (Oleg Badunenko), acullmann@diw.de (Astrid Cullmann), kkar@binghamton.edu (Subal C. Kumbhakar), m.nieswand@lboro.ac.uk (Maria Nieswand)

"Germans still debate the process that has been made bringing east and west together. In terms of motorways and other infrastructure the east sparkles today."

— The Economist, 2. October 2015

1. Introduction

The German reunification in 1990 is often referred to as one of the most important political events of the twentieth century and recognized as an unprecedented example of the economic integration of two neighboring regions with different structure and degree of economic development (Burda and Hunt, 2001). Within the shortest period of time, economic and political unity was achieved between the communist German Democratic Republic and the Federal Republic of Germany. Immediately after the reunion, a process of urban restoration and local infrastructure renovation has commenced in the eastern parts of the reunified country (Sinn, 2002). Previous literature refers especially to the telephone, water supply and waste water disposal systems as examples of the ambitious modernization of a desolate public infrastructure (e.g., Burda and Hunt, 2001; Moss, 2008).

Compared to other urban areas in West Germany, the country's division after World War II led to a decline in population growth, a loss of market access and, subsequently, lower economic activities in western regions close to the inner-German border (Redding and Sturm, 2008). However, following the reunification, Bavarian jurisdictions adjacent to the border experienced growth of both population and local economic activities, accompanied by regional congestion and higher costs of living (Jones and Wild, 1994). Despite these immediate effects, the separation of Germany into east and west had long-run impact. Fifteen years after the reunification, regional income in eastern parts of Germany is still lower than that in the western parts due to the economic closeness and the lower level of international market integration of the former communist state (see Buch and Toubal, 2009). While the living standards, consumption behavior and purchasing power widely converged among eastern and western regions (Berlin-Institut, 2015), the labor market has not recovered yet from its poor performance documented, for example by Hunt (2001, 2002); Snower and Merkl (2006). Evidence on the performance of distribution system operators (DSOs) in the context of reuniting the two countries and the associated restructuring process has not yet been provided.

The focus of this paper is the analysis of the persistent impact of the restructuring – that followed the reunification – on the current performance of providers of electricity distribution services, i.e., DSOs. The service of electricity distribution is crucial for regional growth and welfare as it connects residents and local industries to the national power grid and supplies them with electrical power, which is used for business-related and every-day-life activities such as lighting, heating, cooking, communicating, etc.

For the last twelve years of separation (1978-1990), the eastern electricity distribution sector was organized based on the former fifteen energy regions (*Energiebezirke*) of East Germany. This region-based structure followed an organizational but not an economic rationale (Matthes, 2000). At that time, the political and managerial decisions regarding the sector were solely made by the central state in line with the communist paradigm of a centrally planned economy, jurisdictions and private owners of electricity distribution networks have been dispossessed by the government in the 1950s.

Subsequent to the reunification and after initial setbacks, the eastern energy sector was restructured in a rigorous way. At the very beginning, between 1989 and 1995, the transition was especially ineffective due to a large number of stakeholders, property right issues, and the required compatibility of the two technological systems. Integrating both countries' electricity sectors, thus, became the major challenge for the decision-makers who operationalized the reunion. With western structures as a role model, the eastern electricity distribution had undergone a profound restructuring process that caused a complete reshaping and involved radical changes in operating areas, operational management and ownership (Birke, Hensel, Hirschfeld, and Lenk, 2000).

During the process of reunification, the eastern enterprises were transferred to a privatization agency (*Treuhandanstalt*), which was responsible for conveying them into newly structured organizations. The pursued strategy for this reorganization was to broadly adapt the existing structure of the western German electricity distribution sector, i.e., implementing local and regional distribution networks. The privatization agency primarily intended to sell the enterprises to western electricity firms as they were expected to have the necessary capital and knowhow for the imminent restructuring, and hence, requiring no further subsidies. However, those eastern jurisdictions that claimed back their enterprises, which have been previously taken away from them, were successful. As a result of this restructuring, the sector in the East is composed of local and regional DSOs that operated much smaller networks than their region-based predecessors, that are both publicly or privately owned, and obviously, no longer centrally controlled.

Another aftermath of the reunification was a massive input reallocation. On the one hand, labor force moved from East to West, and physical capital and enormous investments, on the other hand, from West to East (Burda, 2006). Between 1992 and 1998, about 72 billion euros were invested into the eastern Germany's infrastructure (Burda and Hunt, 2001) that has been run down under the communist rule. While the impact of these investments on regional growth in eastern Germany is subject to debate (Koetter and Wedow, 2013), they indisputably caused a general modernization of infrastructure. Alongside the massive capital flows, the eastern economy benefited immediately from adopting sound and well-functioning institutions as well as using the best infrastructure available at that time which is considered to be the main reason why the eastern parts of the country is vastly superior compared to the western parts (Snower and Merkl, 2006; Sinn, 2002; Burda and Hunt, 2001).

The question arises whether such immediate and comprehensive restructuring, and in particular the modernization, and the transfer of technology and knowledge, positively influenced the performance of East German DSOs in the long-run. We analyze this by applying a state-of-the art stochastic frontier model that allows identifying determinants of time-variant (transient/short-run) and time-invariant (persistent/log-run) performance. We use a novel panel dataset of eastern and western German DSOs observed between 2006 and 2012. We approximate the restructuring by the location of the DSOs. We find that both eastern and western DSOs are equally efficient with respect to their transient efficiency. We further show that the DSOs operating in the East German areas exhibit, on average, a higher persistent efficiency than those located in the West Germany.

The remainder of the paper is organized as follows: After describing the method to the evaluation of public infrastructure in the next section, Section 3 presents the formal description of the model. In Section 4, we describe the empirical strategy and the data. Section 5 present the results of our analysis, and Section 6 concludes.

2. Efficiency Measurement in Public Sectors

Beginning with the seminal work by Samuelson (1954) and Tiebout (1956) allocative efficiency in local public service provision has been extensively discussed from various angles. More recently, the importance of technical efficiency, defined as the minimum input necessary to produce a certain level of output, has also been acknowledged as an important component in local public economics. Using stochastic frontier (SF) models, which directly incorporate technical inefficiency, some scholars analyze the aggregated public service provision of jurisdictions (e.g., Grossman, Mavros, and

Wassmer, 1999), while others consider particular public services such as schooling (e.g., Grosskopf, Taylor, and Weber, 2001), libraries (e.g., De Witte and Geys, 2011) and multi-utilities (e.g., Farsi, Fetz, and Filippini, 2008).

The interest in determining the (in)efficiency of DSOs stems from welfare considerations (e.g., Shleifer, 1985; Laffont and Tirole, 1993; Armstrong and Sappington, 2007) and the importance of energy supply for local economic activities. The latter is emphasized by the fact that the volume of electricity delivered are often used as proxies for regional gross domestic product (GDP) where direct measures of GDP are missing (Henderson, Storeygard, and Weil, 2012). At least in developed countries where almost the entire population is supplied with electric power, the efficiency of DSOs is directly related to consumer prices. Abstracting from the complexity associated with prices, an efficient DSO uses the resources in such a way that the cost of supplying a given amount of electricity is minimized. In turn, it is expected to lower prices paid by the consumer. In developing countries where only a small share of the population is supplied with electric power and other public infrastructures, the implementation of efficient networks could help increasing the benefits from urban concentration.[1]

There are ample empirical works estimating the extent of inefficiency of DSOs for multiple regions; e.g., Kumbhakar, Amundsvee, Kville, and Lien (2015a) and Bjørndal, Bjørndal, Cullmann, and Nieswand (2018) analyze DSOs in Norway, Kumbhakar and Hjalmarsson (1998) in Sweden, Filippini, Hrovatin, and Zorič (2004) in Slovenia, Cullmann (2012) and Hirschhausen, Cullmann, and Kappeler (2006) in Germany, Filippini and Wetzel (2014) in New Zealand, Giannakis, Jamasb, and Pollitt (2005) in the UK, Farsi and Filippini (2004) in Switzerland, and Bağdadioğlu, Price, and Weyman-Jones (1996) in Turkey. A shortcoming of these studies is that they rely on the assumption

[1] A sufficiently high access to improved public infrastructure can yield in growth-enhancing benefits of urban concentration prevailing agglomeration costs (e.g., Castells-Quintana, 2016).

that inefficiency is either time-variant (transient) or time-invariant (persistent). Conceptually transient inefficiency relates to non-systematic management problems that are controllable and, therefore, can be reduced in a short time period (Filippini and Greene, 2016). Persistent inefficiency, on the contrary, is linked to differences between DSOs that are systematic in their operation environments or managerial capabilities. The distinction between transient and persistent is important for the assessment of public infrastructure since for public service provision, and in particular, for investment and capital-intensive public sectors such as electricity distribution, it is unreasonable to assume that the short- or long-run inefficiency could be managed in the same way. On the one hand, electricity distribution can be adapted relatively soon to minor changes in the operational environment of DSOs such as variations in population density. Also, non-systematic managerial shortcomings can be addressed quickly. On the other hand, severe economic shocks and decisions made with respect to the organization of the distribution networks are likely to have persistent effects on costs and consumer prices since the network is quasi-fixed and not easily reversible. Each type of inefficiency is different in nature and, therefore, requires different improvement strategies. For example, improvement of the long-term efficiency requires systematic changes to the production process, which can only be done to a limited extent due to technological constraints.

Recently developed SF models allow estimating the production technology of firms while decomposing the error term into noise, DSO-specific effects and the persistent and transient inefficiency components (Colombi, Kumbhakar, Martini, and Vittadini, 2014; Tsionas and Kumbhakar, 2014a; Filippini and Greene, 2016). This model is referred to as the generalized true random effects model (GTRE) and can be readily applied to analyze performance of DSOs. Filippini et al. (2016b) disentangle persistent and transient efficiency for network infrastructures, i.e., DSOs in New Zealand to investigate the implication of disentangling the two types of efficiency for price cap regulation. Distinction between transient and persistent efficiency has been accounted for in the analysis of the US residential sector of electricity demand (see Alberini and Filippini, 2015), Swiss hydro power systems (Filippini et al., 2016a), and nonprofit nursing homes (see Di Giorgio, Filippini, and Masiero, 2015).

The common drawback of all these studies is the failure to control for determinants of two types of inefficiency. Badunenko and Kumbhakar (2017) have extended the GTRE model by introducing determinants of inefficiency which they refer to as a heteroscedastic GTRE primarily because the determinants are modeled via the variance of inefficiency components. The major benefit of this extension is that factors explaining variations in persistent and transient inefficiency can be identified, and their impact (marginal effects) on output (cost) can be estimated. Additionally, the variance of the noise term that is often viewed as production risk can also be made heteroscedastic to allow for factors to explain production risk.

By applying this model, our approach to analyze the performance of DSOs differs notably from previous studies that applied homoscedastic GTRE model. We not only identify structural differences among DSOs, but we also look at the determinants of structural differences, while explicitly controlling for systematic and non-systematic differences in the provision of public services. The structural differences in our model are introduced via persistent and transient inefficiency.

In many cases, structural differences caused by changes in the operational environments can affect the performance of public service provision. De Witte and Geys (2011), for example, show that the productive efficiency of public libraries is lower when right-wing councils are majority whereas it is higher in jurisdictions with higher income of the population and higher levels of urbanization. Likewise, since DSOs operate locally, the structural differences between jurisdictions, and more generally between the DSOs operation areas, are expected to influence the performance of service providers. Moreover, in the case of DSOs, the operational environment also include economic and market conditions because, at least in Europe, electricity distribution is a public service that local governments delegate to private or publicly-owned firms.

3. Methodology

3.1. Multi-input multi-output production technology

We express the multi-input multi-output production technology in terms of a transformation function which, in implicit form, can be expressed as $F(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = A$ where $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are vectors of outputs, inputs, and environmental variables. In general $A = 1$ but to make the transformation function stochastic, we assume $A = \exp(v)$ where v is a stochastic noise term that can take both positive and negative values. The above transformation function assumes that the production process is fully efficient. If there is inefficiency in the use of inputs, which is what we are assuming because of our application, the above transformation function can be written as

$$A\, F(\theta \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\beta}) = 1$$

where $\theta \leq 1$ is input efficiency (a scalar) and $\boldsymbol{\beta}$ is the vector of the parameters of a parametrically specified technology F. In general, an input-oriented (IO) inefficiency model is chosen when inputs are endogenous (choice) variables and the outputs (mostly services) are exogenous (demand determined). The $\boldsymbol{z}$ variables are always exogenous.

For identification purpose, it is standard to assume the transformation function to be homogeneous of degree one in $\boldsymbol{x}$, which implies

$$A\, F(\lambda\, \theta\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\beta}) = \lambda,\ \forall \lambda > 0. \tag{1}$$

Setting $\lambda^{-1} = x_1\theta$, (1) can be rewritten as

$$\theta^{-1} x_1^{-1} A^{-1} = f(\tilde{\boldsymbol{x}}_{-1}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\beta}), \tag{2}$$

where $\tilde{\boldsymbol{x}}_{-1} = (x_2/x_1, \ldots x_N/x_1)$, $F(1, \tilde{\boldsymbol{x}}_{-1}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\beta}) = f(\tilde{\boldsymbol{x}}_{-1}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\beta})$ and N is the number of inputs. Taking the logs of both sides of (2) we obtain

$$-\log x_1 = \log f(\tilde{\boldsymbol{x}}_{-1}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\beta}) + \log\theta + v. \tag{3}$$

Denoting $\log\theta = -u, u \geq 0$, we obtain a typical composite error IO transformation (popularly known as stochastic input distance) function

$$-\log x_1 = \log f(\tilde{\boldsymbol{x}}_{-1}, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\beta}) - u + v. \tag{4}$$

This fits into the SF function introduced by Aigner et al. (1977) and Meeusen and Broeck (1977). In the efficiency literature it goes by the name input distance function (IDF).

3.2. Stochastic IDF with panel data

The SF model originally proposed by Aigner et al. (1977) and Meeusen and Broeck (1977) has traveled a long way since its inception. The panel version of the standard 1977 SF model (without any amendments) can be written as

$$-\log x_{1,it} = \log f(\tilde{\boldsymbol{x}}_{-1,it}, \boldsymbol{y}_{it}, \boldsymbol{z}_{it}; \boldsymbol{\beta}) - u_{it} + v_{it}, \tag{5a}$$

where $i = 1, \cdots, n$ denotes the ith DSO and $t = 1, \cdots, T_i$ denotes the time period in which DSO i is observed, v_{it} is the noise term and $u_{it} \geq 0$ is time-varying technical inefficiency. See Kumbhakar et al. (2015b) for a detailed discussion of various models with different specifications of the time-varying inefficiency term, u_{it}.

Although panel data are extensively used in the literature, only a few papers use models that make use of the panel nature of the data by including DSO heterogeneity. Kumbhakar (1991), Kumbhakar and Heshmati (1995), Kumbhakar and Hjalmarsson (1993), and Kumbhakar and Hjalmarsson (1995) introduced heterogeneity but they interpreted it as persistent inefficiency, viz.,

$$-\log x_{1,it} = \log f(\tilde{\boldsymbol{x}}_{-1,it}, \boldsymbol{y}_{it}, \boldsymbol{z}_{it}; \boldsymbol{\beta}) - u_{0i} - u_{it} + v_{it}, \tag{6}$$

where u_{0i} is assumed to be persistent inefficiency not firm heterogeneity. Greene (2005) used the same specification but interpreted the time-invariant term u_{0i} as firm-effects instead of persistent inefficiency. Since the time-invariant component can include both firm effects (heterogeneity) Colombi, Kumbhakar, Martini, and Vittadini (2014); Kumbhakar, Lien, and Hardaker (2014) and Tsionas and Kumbhakar (2014b) introduced a model that split the error term into four components. The first component captures firms' latent heterogeneity (see Greene, 2005) and the second component captures long-run/persistent/time-invariant inefficiency as in Kumbhakar and Hjalmarsson (1993); Kumbhakar and Heshmati (1995) and Kumbhakar and Hjalmarsson (1995), both of which are time-invariant. The third component captures time-varying inefficiency (see Kumbhakar, 1987), while the last component captures random shocks. Both the third and fourth components are observation-specific (i.e., vary across firms and over time). Thus the model that captures all the four components can be formally expressed as

$$-\log x_{1,it} = \log f(\tilde{\boldsymbol{x}}_{-1,it}, \boldsymbol{y}_{it}, \boldsymbol{z}_{it}; \boldsymbol{\beta}) + v_{0i} - u_{0i} - u_{it} + v_{it}, \tag{7}$$

where $u_{0i} \geq 0$ and $u_{it} \geq 0$ represent persistent and time-varying inefficiency, respectively, while v_{0i} captures latent firm heterogeneity and v_{it} is the classical random noise. We call this homoscedastic four-component model. The four-component homoscedastic model has been applied to analyze the efficiency in health care, agriculture, transportation (Colombi et al., 2014; Kumbhakar et al., 2014) and US banks (Tsionas and Kumbhakar, 2014b). Kumbhakar and Lai (2016) further extended the model by considering a system of revenue share equations each having four-components.

In the homoscedastic four-component model, all the components are independently and identically distributed (i.i.d.) random variables. Thus, the model can not be useful for policy purposes unless efficiency levels can be systematically changed by changing the policy variables. In other words, we need a model in which inefficiency is systematically related to some firm characteristics. For example, if the regulators want the firms they regulate to move faster to the frontier (increase the catch-up rate) by giving them incentives (carrots), the time-varying inefficiency has to be related to some policy variables that the regulators can change. Similarly, to talk about reducing production risk, the model has to allow the variances of the time-invariant firm-effects and/or the noise term to depend on some exogenous factors. In summary, although the homoscedastic four-component model can give us estimates of persistent and time-varying efficiency, the model cannot explain the determinants of inefficiency, and therefore, cannot be used for prescribing policies to increase efficiency. Similarly, the model can not explain differences in risks within and between firms.

3.3. Determinants of inefficiency

Since our application focuses on the role of being located in East Germany (i.e., having undergone restructuring) as well as some other firm-specific characteristics that are outside DSO's influence on efficiency, we argue that both the regulator and firms are interested in knowing what determines persistent and time-varying inefficiency, and what their marginal effects are. For example, the regulator might be more interested in the drivers of persistent inefficiency, while firms presumably strive to eliminate the long-run inefficiency.

In our specification we use the determinants of persistent inefficiency to appear in the (pre-truncated) variance of u_{0i}, which is time-invariant, viz.,

$$u_{0i} \sim N^{+}(0, \sigma^2_{u0i}) \text{ where } \sigma^2_{u0i} = \sigma^2_{u0} \exp\left(\boldsymbol{z}_{u0i}\boldsymbol{\gamma}_{u0}\right), \quad i = 1, \cdots, n, \tag{8}$$

where σ^2_{u0} is a constant and $\boldsymbol{z}_{u0i}$ is the vector of covariates that determines the heteroscedasticity function of persistent inefficiency and is by definition time-invariant. Since $E(u_{0i}) = \sqrt{(2/\pi)}\sigma_{u0i} = \sqrt{(2/\pi)} \exp\left(\frac{1}{2}\boldsymbol{z}_{u0i}\boldsymbol{\gamma}_{u0}\right)$, the $\boldsymbol{z}_{u0i}$ variables can be viewed as determinants of persistent inefficiency.[2] Variables in $\boldsymbol{z}_{u0i}$ may vary by firms, but not over time within firms. This means that σ^2_{u0i} is explained only by time-invariant covariates.

In a similar fashion, we introduce determinants of time-varying inefficiency via the pre-truncated variance of u_{it}. More specifically, we assume

$$u_{it} \sim N^{+}(0, \sigma^2_{uit}) \;\text{ where }\; \sigma^2_{uit} = \sigma^2_{u} \exp\left(\boldsymbol{z}_{uit}\boldsymbol{\gamma}_{u}\right), \quad i = 1, \cdots, n, \quad t = 1, \cdots, T_i, \tag{9}$$

where σ^2_u is a constant and $\boldsymbol{z}_{uit}$ denotes the vector of covariates that explains time-varying inefficiency. Since u_{it} is half-normal, $E(u_{it}) = \sqrt{(2/\pi)}\sigma_{uit} = \sqrt{(2/\pi)} \exp\left(\frac{1}{2}\boldsymbol{z}_{uit}\boldsymbol{\gamma}_{u}\right)$, and therefore, anything that affects σ_{uit} also affects time-varying inefficiency.[3]

The model presented above can be estimated using either the classical ML method proposed by Colombi et al. (2014) or the simulated ML method advocated by Filippini and Greene (2016). Details on these can be found in Colombi et al. (2014), Filippini

[2] Persistent inefficiency can also be modeled assuming the pre-truncation mean of u_{0i} to be a function of the z_{ui0} variables.

[3] Similar to persistent inefficieny, time-varying inefficiency can also be modeled assuming the pre-truncation mean of u_{it} to be a function of the z_{uit} variables.

and Greene (2016) and Badunenko and Kumbhakar (2017). Colombi et al. (2014) also provide the formula for computing persistent and transient inefficiency. For completeness we discuss these in the appendix.

4. Empirical Model and Data

Based on the input distance function approach, we estimate a translog input distance function for three inputs (x_1, x_2, x_3), two outputs (y_1, y_2), and R time-varying external factors combined in vectors $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, respectively. In our application we consider one time-varying external factor, i.e. $R = 1$, represented by z_1. Further, we include a linear time trend t, its square t^2 and interactions with inputs and outputs to accommodate (non-neutral and non-monotonic) technological change:[4]

$$
\begin{aligned}
-\log x_{1it} = {} & \beta_0 + \beta_{x_2} \log\left(x_{2,it}/x_{1,it}\right) + \beta_{x_3} \log\left(x_{3,it}/x_{1,it}\right) \\
& + \beta_{y_1} \log\left(y_{1,it}\right) + \beta_{y_2} \log\left(y_{2,it}\right) \\
& + 0.5\left(\beta_{x_{22}} [\log\left(x_{2,it}/x_{1,it}\right)]^2 + \beta_{x_{32}} [\log\left(x_{3,it}/x_{1,it}\right)]^2\right) \\
& + 0.5\left(\beta_{y_{12}} [\log\left(y_{1,it}\right)]^2 + \beta_{y_{22}} [\log\left(y_{2,it}\right)]^2\right) \\
& + \beta_{x_2 x_3} \log\left(x_{2,it}/x_{1,it}\right) \log\left(x_{3,it}/x_{1,it}\right) \\
& + \beta_{x_2 y_1} \log\left(x_{2,it}/x_{1,it}\right) \log\left(y_{1,it}\right) + \beta_{x_2 y_2} \log\left(x_{2,it}/x_{1,it}\right) \log\left(y_{2,it}\right) \\
& + \beta_{x_3 y_1} \log\left(x_{3,it}/x_{1,it}\right) \log\left(y_{1,it}\right) + \beta_{x_3 y_2} \log\left(x_{3,it}/x_{1,it}\right) \log\left(y_{2,it}\right) \\
& + \beta_{y_1 y_2} \log\left(y_{1,it}\right) \log\left(y_{2,it}\right) \\
& + \beta_{x_2 t} \log\left(x_{2,it}/x_{1,it}\right) t + \beta_{x_3 t} \log\left(x_{3,it}/x_{1,it}\right) t
\end{aligned}
\tag{10}
$$

[4] We thank an anonymous reviewer for suggesting to extend our specification to accommodate non-neutral technical change.

$$+ \beta_{y_1 t} \log(y_{1,it}) t + \beta_{y_2 t} \log(y_{2,it}) t$$
$$+ \beta_{z_1} \log(z_{1,it}) + \beta_t t + 0.5 \left(\beta_{tt} t^2\right) + v_{0i} - u_{0i} + v_{it} - u_{it}.$$

The data for 242 German DSOs is gathered and combined from two sources, the German Federal Statistical Office and 'ene't', a professional data provider (RDC, 2006-2012; ene't, 2015). The merged dataset is unique and it allows us to model the production process while controlling for structural characteristics such as population density and the location of the operation areas. Our sample is an unbalanced panel observed over 7 years (2006-2012) with a total of 1370 observations of which 442 refer to the 71 DSOs in our sample that are located in eastern parts of Germany and, therefore, subject to restructuring after 1989. The remaining 946 observations belong to 171 DSOs located in western parts of Germany. Table 1 shows the characteristics of our data.[5]

We model the production process using three inputs and two outputs.[6] The input variables are labor input (x_L) measured in total number of hours worked,[7] network length (x_N) in kilometers (km),[8] and transformer capacity (x_C) in megawatt hours (MWh).[9] Outputs are the annual amount of electricity delivered (y_E) in MWh and the

---

[5] Due to non-disclosure requirements of the Federal Statistical Office the minimum and maximum values of the data cannot be presented. This requirement extends to presentation of regression results. We, therefore, present the 1 percent and 99 percent quantiles as lowest and highest values.

[6] Our modelling of the technology closely corresponds to previous academic and regulatory specifications related to DSOs in Germany. In addition to the literature mentioned in section 2, the reader is referred to, e.g., BNetzA (2006) for further reading about applied approaches of technology modelling.

[7] We are aware of the criticism of this choice due to the potentially distorting effect of outsourcing: a utility can improve its efficiency simply by switching from in-house production to outsourcing. However, there is no data available that would provide a closer approximation

[8] Network length is the sum of cables and overhead power lines.

[9] Capacity gives the installed power of the trafo stations in MVA.

Table 1: Descriptive statistics

Variable	Unit	Name	Q1	Q25	Med	Q75	Q99	SD
Customers	number	y_C	$1,287$	$7,919$	$16,456$	$29,469$	$1,124,662$	$163,351$
Electricity delivered	MWh	y_E	$13,968$	$100,736$	$212,064$	$431,549$	$16,051,738$	$3,443,759$
Labor	hours	x_L	$2,799$	$44,721$	$94,267$	$186,940$	$2,725,095$	$400,250$
Network length	km	x_N	70	259	444	845	$54,182$	$8,885$
Capacity	MVA	x_C	7	40	75	155	$13,155$	$1,858$
Population density	inhabitants/km^2	z_D	20	123	240	467	$1,381$	302
Location (East)	dummy	z_{East}	0	0	0	1	1	0

Notes: tnobs = 1370, years = 2006 - 2012, firms = 242, source: Federal Statistical Office and 'e'net.'
Each firm is observed at least 4 years and the average time span firms are observed is 5.661.

number of connected customers (y_C). Since the technology of electricity distribution is subject to the characteristics of its operational environment (Bjørndal et al., 2018; Nieswand and Seifert, 2018), we incorporate population density (z_D), defined as the number of connection points per squared km in the operational area, to capture exogenously induced neutral technology shift. Neglecting such shift variables from the production process is likely to bias estimates of inefficiency.[10]

The variances of the transient and persistent inefficiency, $\sigma^2_{u_{it}}$ and $\sigma^2_{u_{0i}}$, are modeled to depend population density and location, respectively

$$\log\left(\sigma^2_{u_{it}}\right) = \delta_0 + \delta_1 z_{r,it} \tag{11}$$

$$\log\left(\sigma^2_{u_{0i}}\right) = \gamma_0 + \gamma_1 z_{s,0i} \tag{12}$$

Additional to the production process, we let population density influence transient inefficiency by letting $\log\left(\sigma^2_{u_{it}}\right)$ vary with z_D, i.e. $z_r = z_{1it} = z_D$. Given that z_D also appears in the frontier function (10), we separate the twin effects of z_D, viz., separate the frontier shift from transient efficiency change due to this external factor. Omitting one from the model is likely to affect the other.

Note that the external factors that influence persistent inefficiency can not change over time and may include dummy variables. We create a location dummy (z_{East}) that takes value 1 for the DSOs operating in the eastern parts of Germany and 0 otherwise. Our main interest is to identify the potential influence of the restructuring process after 1989 on the persistent inefficiency of the eastern German DSOs. More specifically, we would like to know whether the restructuring of the electricity distribution sector in East Germany is associated with a higher persistent efficiency. For this purpose, we specify the variance of persistent inefficiency $\left(\sigma^2_{u_{0i}}\right)$ as a function of the location

[10] Their effect would not be captured by the noise and DSO-specific fixed effects.

dummy z_{East}, thus $z_s = z_{20i} = z_{East}$ in (12). Since the dummy variable takes a value of 1 for the DSOs operating in East Germany, a negative (positive) coefficient would imply a lower variance of persistent inefficiency, and hence, higher (lower) average persistent efficiency.

5. Estimation Results

5.1. Distance function estimation

Table 2 shows the estimated coefficients of the IDF in (10), which represents the production technology of the German DSOs. The IDF is estimated under the assumption of homoscedastic noise components (random effect and random noise) and heteroscedastic inefficiency components (transient and persistent) as in (11) and (12).

Since the IDF is dual to the cost function (see Färe and Primont, 1995), $-\partial \log x_1 / \partial \log(x_j/x_1)$ measures the cost elasticity of input x_j ($j = N, C$). Scaling our variables by the respective medians before estimation, allows interpreting the coefficient of the first order term of $\log(x_j/x_1)$ in the translog representation of production technology in (10) directly as an estimate of cost elasticity of input x_j at the median values of all inputs and outputs.[11] Thus, our results displayed in Table 2 suggest that, e.g., increasing its network input by 1 percent, increases costs by 0.52 percent at the median as the point estimate of the coefficient of x_D equals 0.52. Table 2 further indicates that the cost elasticity of network is by far the largest (0.52), followed by the

[11] The derivative $-\partial \log x_1 / \partial \log(x_j/x_1)$ is equal to the coefficient at the first order term plus coefficients at the second order terms multiplied by log of either median scaled inputs or outputs. This derivative depends on the specific values of all inputs and outputs, however if they are all set equal to their respective medians, expression under *log* is equal to 1, making the terms beyond the first order coefficients all zero.

Table 2: Estimates of the input distance function

Dependent var.:	$-\log(x_1)$	Estimation results		
Parameter	Variable	Coefficient	SE	p-value
β_0	$Intercept$	0.2516	0.0199	0.0000
β_{x_2}	x_N	0.5234	0.0134	0.0000
β_{x_3}	x_C	0.2537	0.0142	0.0000
β_{y_1}	y_E	-0.0829	0.0108	0.0000
β_{y_2}	y_C	-0.3141	0.0114	0.0000
$\beta_{z_{1it}}$	z_D	0.1798	0.0099	0.0000
β_t	t	0.0034	0.0041	0.4048
β_{tt}	t^2	-0.0038	0.0012	0.0016
$\beta_{x_{22}}$	$(x_N)^2$	0.2060	0.0170	0.0000
$\beta_{x_{32}}$	$(x_C)^2$	0.0782	0.0048	0.0000
$\beta_{y_{12}}$	$(y_E)^2$	-0.0213	0.0075	0.0044
$\beta_{y_{22}}$	$(y_C)^2$	-0.0480	0.0044	0.0000
$\beta_{x_2x_3}$	$x_N \cdot x_C$	-0.1029	0.0103	0.0000
$\beta_{x_2y_1}$	$x_N \cdot y_E$	0.0127	0.0097	0.1896
$\beta_{x_2y_2}$	$x_N \cdot y_C$	-0.0718	0.0088	0.0000
$\beta_{x_3y_1}$	$x_C \cdot y_E$	-0.0287	0.0092	0.0017
$\beta_{x_3y_2}$	$x_C \cdot y_C$	0.0511	0.0086	0.0000
$\beta_{y_1y_2}$	$y_E \cdot y_C$	-0.0009	0.0045	0.8437
β_{x_2t}	$x_N \cdot t$	-0.0011	0.0022	0.6114
β_{x_3t}	$x_C \cdot t$	0.0008	0.0024	0.7293
β_{y_1t}	$y_E \cdot t$	-0.0023	0.0016	0.1484
β_{y_2t}	$y_C \cdot t$	0.0047	0.0017	0.0051

Note: All variables, except for t, are median-corrected and are in logs. The left hand side variable x_1 is labor, while x_N and x_C are log ratios of network length (x_2) and capacity (x_3) to labor, i.e., $x_N = \log(x_2/x_1)$ and $x_C = \log(x_3/x_1)$. Total number of observations is 1370, the number of DSOs is 242, the minimum, average and maximum number of time periods a DSO is observed are 4, 5.67, and 7, respectively.

cost elasticity of capacity (0.25) and labor (0.23).[12] Given that electricity distribution

[12] Due to homogeneity of degree 1 of the cost function, the cost elasticity of labor at median values of all variables is obtained as $1 - 0.52 - 0.25 = 0.23$.

is a network-intensive sector, these are reasonable estimates and comparable to other empirical work on the German DSOs (e.g., Cullmann, 2012).

Duality results further imply that $-\partial \log x_1/\partial \log y_m$ is the cost elasticity of output y_m $(m = C, E)$. The point estimates of coefficients at the output variables are negative, which indicates that an increase in electricity delivered by 1 percent is associated with an increase in the use of all the inputs and, hence, costs by about 0.08 percent.[13] Similarly, cost is increased by 0.31 percent when the number of connected customers is increased by 1 percent. Therefore, adding new connections is more costly than increasing the supply of electricity delivered using the existing infrastructure.

As our IDF is a full translog in t, technical change (TC) contains both neutral and non-neutral components. Technical change is neutral if $\beta_{x_2t} = \beta_{x_3t} = \beta_{y_1t} = \beta_{y_2t} = 0$. The LR test statistic of this hypothesis equals 215.34, which exceeds the critical value of the mixed chi-squared distribution at the 1 percent level, 12.4827. The technical change is therefore decidedly non-neutral. The speed of the technical change is decreasing evidenced by the coefficient at the squared time variable -0.0038.

The coefficient associated with the environmental variable z_D is used to capture cost differences due to the characteristics of the operational area. In our model, population density influences total cost through two channels that must be jointly considered to capture the overall effect of population density on production and costs, respectively. The first channel is the production processes, which is captured by the technology: The coefficient of population density (0.18) has the usual constant elasticity interpretation and shows that if population density increases by 1 percent, total costs would reduce by 0.18 percent. The second channel is through inefficiency: Population density is mod-

[13] Again, the coefficient at the first order term of $\log y_m$ in the translog representation of production technology in (10) gives an estimate of cost elasticity of output y_m at the median values of all inputs and outputs.

eled further to determine transient inefficiency (in which it appears in non-logarithmic form and is further discussed in section 5.3). To interpret the marginal effect of population density on total costs in a meaningful way, we visualize its joint effect on costs through technology and inefficiency in Figure 1.[14] The Kernel density of the overall marginal effect shows that increasing z_D always reduces total costs for our sample DSOs. However, the magnitude of this effect varies roughly between 11 and 18 percent with the marginal effect mainly ranging between 14 and 18 percent. Thus, increasing the population density by 1 percent, decreases total costs by 11 to 18 meaning that total cost of electricity distribution declines with population density.[15]

5.2. *Overall, persistent and transient efficiency*

Table 3 shows selected statistics (first column) of the DSO-specific overall efficiency values (second column) and its decomposition into persistent (third column) and transient efficiency (fourth column) components. An efficiency score of less than 100 percent is associated with over-use of inputs, which implies observed costs higher than potential costs. The overall efficiency is the product of persistent and transient efficiency.

From Table 3, we observe fairly large differences across the DSOs in terms of their efficiency. The overall efficiency of our sample DSOs ranges from about 25 to 87 percent while its median value is 67 percent. This finding suggests that, on average, at the given level of produced output, inputs could be reduced by roughly one third

---

[14] Figure 1 shows the estimated kernel density of elasticity of total costs with respect to population density. More specifically, it is calculated as $\partial \ln x_1 / \partial \ln z_d = -0.18 + \partial u / \partial \ln z_d$, where u is replaced by $E(u)$.

[15] We also tested for the potential impact of other control variables on the technology, such as the share of overhead cables and the location of DSOs. We did not find those variables to be statistically significant and, therefore, excluded them from our analysis.

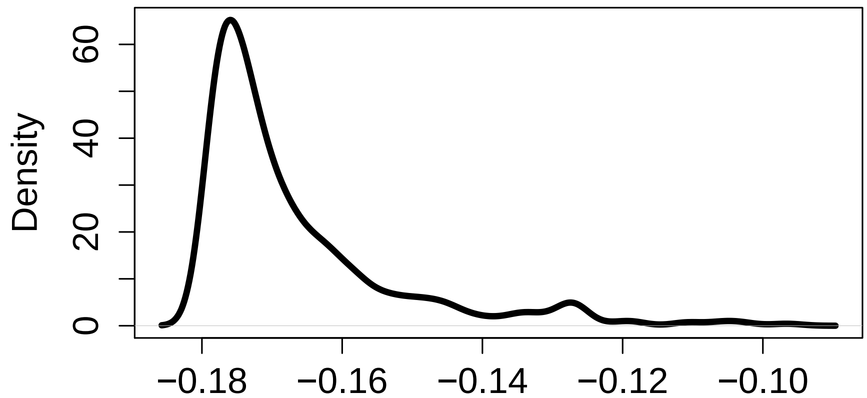

Figure 1: Kernel estimated density of the elasticity of costs with respect to population density z_D.

(33 percent). Compared to other studies on German DSOs and Germany's regulatory benchmarking results (Cullmann, 2012; Swiss economics and Sumicsid, 2014; Swiss Economics and Sumicsid, 2018), the minimum figures seem to be rather low. However, as revealed by the remaining two columns in Table 3, this is mainly due to the low values in persistent efficiency, which has not been considered and captured before. Our results on transient efficiency (see column four in Table 3) are indeed in line with previous findings.[16]

Examining the components of overall efficiency in more detail, we find that transient efficiency is between 83 and 98 percent with a median of 95 percent while persistent efficiency is much lower and ranges from 25 to 91 percent with a median value of

[16] In our sample, the mean value of transient efficiency is 94 percent. Cullmann (2012) estimated an average efficiency of roughly 88 percent while the German regulator identified mean values of 92.2 percent, 94.7 percent and 94.1 percent for the regulatory periods 2009 − 2013, 2014 − 2018, and 2019 − 2023, respectively (Swiss Economics and Sumicsid, 2018).

Table 3: Descriptive statistics of efficiency estimates

	Efficiency scores in percent		
Statistic	overall	persistent	transient
p1	25.18	27.19	82.84
p25	56.59	59.76	93.20
p50	67.71	71.56	94.75
p75	83.10	89.55	95.73
p99	88.27	91.43	98.17

Note: p# denotes the #th sample percentile.

72 percent. Consequently, we conclude that inefficiency in the operations of German DSOs is mainly driven by persistent, hence, structural reasons rather than short-term managerial inefficiency.

We emphasize that our results regarding the transient efficiency are not only comparable to previous studies but also indicate that the currently employed regulation scheme successfully incentivizes the German DSOs to reduce inefficiency and operate in a relatively cost efficient manner. It is worth noting that, however, even small values of inefficiency translate into notable monetary amounts of revenue caps and we would likely observe much lower transient efficiency in the absence of this regulatory practice.[17]

[17] In 2009, Germany introduced the so-called *Anreizregulierung* (incentive regulation) based on benchmarking exercises. This regulation specifically aimed to incentivize efficient cost structures and used frontier models for determining parts of the revenue allowances. We argue, that, even though our sample starts in 2006, our analysis is valid and comprehensive because firms already started working towards efficient cost structures throughout our observed time period since the introduction of this regulatory scheme was announced in 2005 and data collection for the first regulatory period already conducted in 2006.

Table 4: Estimates of the parameters of the error components

Parameter	Variable	Coefficient	SE	p-value
Random effect	$\sigma_{v_{0i}}$	-1.2902	0.0516	0.0000
Random noise	$\sigma_{v_{it}}$	-6.0397	0.1750	0.0000
Persistent inefficiency (u_{0i})				
γ_0	Intercept of $\log \sigma^2_{u_{0i}}$	-0.2336	0.0613	0.0001
γ_1	z_{East}	-3.7889	0.2866	0.0000
Transient inefficiency (u_{it})				
δ_0	Intercept of $\log \sigma^2_{u_{it}}$	-5.3809	0.2539	0.0000
δ_1	z_D	0.2593	0.0697	0.0002

5.3. Determinants of persistent and transient inefficiency

In this section we examine the determinants of inefficiency. More specifically, we are interested in examining whether population density and the location of the DSOs explain their persistent and transient efficiency, respectively.

Table 4 presents the point estimates of the parameters associated with the error components. Focusing first on the variance of persistent inefficiency, we find a significant and negative coefficient for our location dummy z_{East} (i.e., $\gamma_{z20i} = -3.79$), which means that the variance of inefficiency is smaller for the DSOs in Eastern Germany. Thus, on average, East German DSOs perform better than their West German counterparts in terms of persistent efficiency. This result confirms our previous discussion and the literature (e.g., Sinn, 2002) and must be attributed to the restructuring of the East German electricity distribution sector, that followed the German reunification.

Figure 2 provides more details on the persistent efficiency estimates by displaying its distribution for eastern and western DSOs separately. The gray, solid line represents

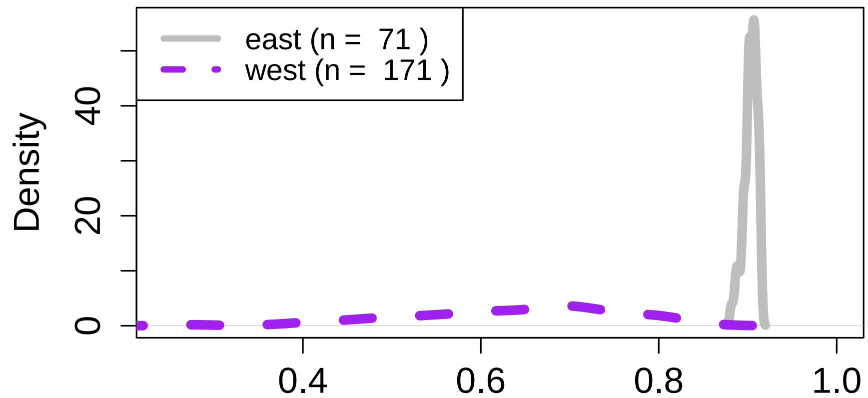

Figure 2: Kernel estimated density of the persistent efficiency

the distribution of persistent efficiency for the East German DSOs while the purple, dashed line represents it for the West German DSOs. Two observations are worth noting. First, eastern DSOs perform uniformly well and better than most of the western DSOs. Second, the mass of the solid distribution is at the level of efficiency where dashed distribution tails. This suggests that most of the eastern DSOs are at par with best practice western counterparts. These two observations seem to confirm that the best operating western structures served as a role model for the eastern DSOs (Birke et al., 2000).

Table 5: Descriptive statistics of elasticity of transient efficiency with respect to population density

Statistic	Marginal effect
p1	-0.0108
p25	-0.0667
p50	-0.1297
p75	-0.2525
p99	-0.7463

Notes: p# denotes the #th sample percentile.

Table 4 further suggests that the variance of transient inefficiency is positively influenced by population density (i.e. $\gamma_{z1it} = 0.2593$). Thus, DSOs operating in areas with higher population densities are, on average, more inefficient (less efficient) than those operating in less densely populated areas. Note that the marginal effect of population density on transient efficiency (TE_{it}) is observation-specific and vary with the respective values of z_D. Since $TE_{it} \approx 1 - u_{it}$ for small values of u_{it}, $\partial TE_{it}/\partial z_{Dit} = -\partial u_{it}/\partial z_{Dit}$. Table 5 presents the elasticity of transient efficiency with respect to population density. Since elasticity is negative for all observed DSOs, increasing population density by 1%, decreases efficiency (increases inefficiency) by between 0.0108 and 0.7463%.

Comparing the performance of eastern and western DSOs in terms of transient efficiency, we find that there are virtually no difference between the DSOs in the two groups. Figure 3 illustrates that using Kernel densities of transient efficiency scores separately for eastern and westerns DSOs. Consequently, we find that the location of DSOs matters in terms of persistent (structural) but not in terms of transient (non-systematic) performance.

The equivalence of transient efficiency between eastern and western DSOs is expected since all DSOs are subject to the same regulatory scheme, which incentivizes

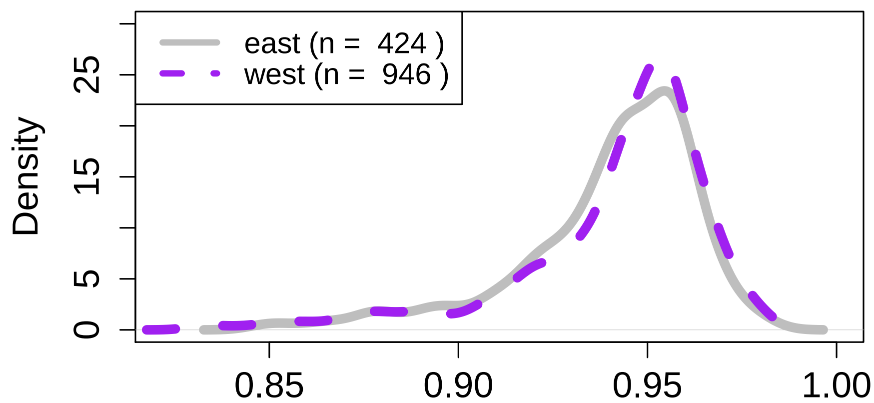

Figure 3: Kernel estimated density of the transient efficiency

efficient operations as discussed in section 5.2. There is no obvious reason why DSOs in different locations would not have the same ability and incentive to reduce their transient inefficiencies when it is under their control.

This finding is further in line with the relevant theoretical and empirical literature, which provides evidence for the fact that (time-varying) efficiency of firms does not vary among different types of DSOs when they are highly regulated. In highly regulated sectors, such as it is the case in Germany,[18] private and state-owned firms are expected to perform equally well, at least under complete regulatory contracts (e.g., Laffont and

[18] Beginning in 1998, the EU Directives 96/92/EC and 2003/54/EG initiated a gradual liberalization of the German electricity market. This process opened end-consumer markets and involved the unbundling of the distribution networks from other parts of the value-added chain of electricity provision. Further, in 2009, the regulatory scheme changed from a cost-plus regulation to a revenue-cap-based incentive regulation, which involved cost benchmarking and aimed to foster cost efficiency. Due to the liberalization and the newly implemented regulatory approach, the once monopolistic electricity distribution market was transformed into a much more competition-oriented environment.

Tirole, 1993). The theoretical prediction is further widely supported by empirical studies on electricity companies. For example, Atkinson and Halvorsen (1986) show that the private and public electricity utilities in the USA exhibit similar levels of relative efficiency.

6. Concluding Remarks

Efficiency of public service provision has experienced an increasing attention in local public economics. Providing these services efficiently is relevant for maximizing welfare and assuring economic activities in both developed and developing countries, regions and municipalities. Many core public infrastructures, e.g., telecommunication, sanitary, and electricity distribution, are network-based sectors for which it is reasonable to assume that inefficiency has both transient and persistent components.

Using a sample of German electricity distribution companies, we estimate an input distance function accounting for both persistent and transient inefficiency. We find that overall inefficiency is mainly driven by the persistent component, which is of structural and long-term nature. Our findings further show that DSOs located in East Germany exhibit, on average, lower persistent inefficiency, i.e., higher persistent efficiency, induced by the restructuring process that took place after the reunification of Germany. The transient component contributes to overall inefficiency to a much smaller extend as we observe relatively high transient efficiency among all DSOs, irrespective of their location. From this, we conclude that the regulatory scheme in place is successfully addressing the aim of incentivizing efficient production and cost structures. This further suggests that persistent inefficiency is not yet reduced by implemented regulatory instruments but offers large potential for further improvements in the sector. Given our results, further restructuring could be considered, especially of the western DSOs. We leave, however, identifying effective regulatory or policy instruments targeting structural inefficiency for further research.

Our analysis shows that disentangling both types of inefficiency is an important exercise because it identifies improvement potentials, it can explain which factors actually drive short-term and long-term efficiency and, thereby, helps identifying appropriate strategies to achieve this improvement. Short- and long-term inefficiencies are likely to be issues in most public infrastructures due to the network-based technologies. Thus, we consider the applied methodology as being also relevant to other public sectors, e.g., gas distribution, water supply, sewerage, and local transportation.

References

Aigner, D., Lovell, C. A. K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6 (1), 21–37.

Alberini, A., Filippini, M., 2015. Transient and persistent energy efficiency in the US residential sector: Evidence from household-level data. Working Paper 15/220, Center of Economic Research at ETH Zurich.

Armstrong, M., Sappington, D. E. M., 2007. Recent developments in the theory of regulation. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization. Vol. 3. North Holland/ Elsevier: Amsterdam, The Netherlands, Ch. 27, pp. 1557–1700.

Atkinson, S. E., Halvorsen, R., 1986. The relative efficiency of public and private firms in a regulated environment: The case of U.S. electricity utilities. Journal of Public Economics 29 (3), 281–294.

Badunenko, O., Kumbhakar, S. C., July 2017. Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter? European Journal of Operational Research 260 (2), 789–803.
URL https://doi.org/10.1016/j.ejor.2017.01.025

Bağdadioğlu, N., Price, C., Weyman-Jones, T., 1996. Efficiency and ownership in electricity distribution: a non-parametric model of the Turkish experience. Energy Economics 18, 1–23.

Berlin-Institut, 2015. So geht Einheit. techreport, Berlin-Institut für Bevölkerung und Entwicklung.

Birke, A., Hensel, V., Hirschfeld, O., Lenk, T., 2000. Die ostdeutsche Elektrizitätwirtschaft zwischen Volkseigentum und Wettbewerb. resreport 22, University of Leipzig, Institute for Public Finance and Public Managment.

Bjørndal, E., Bjørndal, M., Cullmann, A., Nieswand, M., 2018. Finding the right yardstick: Regulation of electricity networks under heterogeneous environments. European Journal of Operational Research 265 (2), 710–722.

BNetzA, 2006. Bericht der Bundesneztagentur nach §112a EnWG zur Einführung der Anreizregulierung nach §21a. Technical report, Bundesnetzagentur, Bonn, Germany.

Buch, C. M., Toubal, F., 2009. Openess and growth: The long sahdow of the Berlin wall. Jounal of Macroeconomics 31 (3), 409–422.

Burda, M. C., 2006. Factor reallocation in Eastern Germany after reunification. The American Economic Review 96 (2), 368–374.

Burda, M. C., Hunt, J., 2001. From reunification to economic integration: Productivity and the labor market in Eastern Germany. resreport 2, Brookings Papers on Economic Activity.

Butler, J. S., Moffitt, R., 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. Econometrica 50 (3), 761–764.

Castells-Quintana, D., 2016. Malthus living in a slum: urban cocentration, infrastructures and economic growth. Journal of Urban Economics in press.

Colombi, R., Kumbhakar, S., Martini, G., Vittadini, G., 2014. Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. Journal of Productivity Analysis 42 (2), 123–136.

Cullmann, A., 2012. Benchmarking and firm heterogeneity in electricity distribution: A latent class analysis of Germany. Empirical Economics 42 (1), 147–169.

De Witte, K., Geys, B., 2011. Evaluating efficient public good provision: Theory and evidence from a generalised conditional efficiency model for public libraries. Journal of Urban Economics 69 (3), 319–327.

Di Giorgio, L., Filippini, M., Masiero, G., 2015. Structural and managerial cost differences in nonprofit nursing homes. Economic Modelling 51, 289–298.

ene't, 2015. Beschreibung Datenbeschreibung Netznutzung Strom. Tech. Rep. Stand 01.10.2014, ene't GmbH.

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

Farsi, M., Fetz, A., Filippini, M., 2008. Economies of scale and scope in multi-utilities. Energy Journal 29 (4), 123–143.

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

Filippini, M., Geissmann, T., Greene, W., 2016a. Persistent and transient cost efficiency - an application to the Swiss hydropower systems. resreport CER-ETH Economics Working Paper Series 16/251, ETH Zurich.

Filippini, M., Greene, W., 2016. Persistent and transient productive inefficiency: a maximum simulated likelihood approach. Journal of Productivity Analysis 45 (2), 187–196.

Filippini, M., Greene, W., Masiero, G., 2016b. Persistent and transient productive inefficiency in a regulated industry: Electricity distribution in New Zealand. resreport IdEP Economic Papers 2016/03, Universitá della Svizzera Italiana.

Filippini, M., Hrovatin, N., Zorič, J., 2004. Efficiency and regulation of the Slovenian electricity distribution companies. Energy Policy 32 (3), 335–344.

Filippini, M., Wetzel, H., 2014. The impact of ownership unbundling on cost efficiency: Empirical evidence from the New Zealand electricity distribution sector. Energy Economics 45, 412–418.

Giannakis, D., Jamasb, T., Pollitt, M., 2271 2005. Benchmarking and incentive regulation of quality of service: an application to the UK electricity distribution utilities. Energy Policy 33 (17), 2256–.

Greene, H. W., June 2005. Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. Journal of Econometrics 126 (2), 269–303.

Grosskopf, S., Taylor, L. L., Weber, W. L., 2001. On the determinants of school district efficiency: Competition and monitoring. Journal of Urban Economics 49, 453–478.

Grossman, P., Mavros, P., Wassmer, R. W., 1999. Public sector technical inefficiency in large U.S. cities. Journal of Urban Economics 46, 278–299.

Henderson, J. V., Storeygard, A., Weil, D. N., 2012. Measuring economic growth from outer space. American Economic Review 102 (2), 994–1028.

Hirschhausen, C., Cullmann, A., Kappeler, A., 2006. Efficiency analysis of german electricity distribution utilities – non-parametric and parametric tests. Applied Economics 38 (21), 2553–2566.

Hunt, J., 2001. Post-unification wage growth in East Germany. The Review of Economics and Statistics 83 (1), 190–195.

Hunt, J., 2002. The transition in East Germany: When is a ten-point fall in the gender wage gap bad news? Journal of Labor Economics 20 (1), 148–169.

Jondrow, J., Lovell, C. A. K., Materov, I. S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19 (2/3), 233–38.

Jones, P. N., Wild, T., 1994. Opening the frontier: Recent spatial impacts in the former inner-German border zone. Regional Studies 28 (3), 259–273.

Koetter, M., Wedow, M., 2013. Transfer payments without growth: Evidence from German regions, 1992-2005. International Journal of Urban and Regional Research 37 (4), 1438–1455.

Kumbhakar, S., Hjalmarsson, L., 1998. Relative performance of public and private ownership under yardstick competition: electricity retail distribution. European Economic Review 42 (1), 97–122.

Kumbhakar, S. C., 1987. The specification of technical and allocative inefficiency in stochastic production and profit frontiers. Journal of Econometrics 34 (3), 335–348.

Kumbhakar, S. C., 1991. The measurement and decomposition of cost-inefficiency: the translog cost system. Oxford Economic Papers 43 (4), 667–683.

Kumbhakar, S. C., Amundsvee, R., Kville, H. M., Lien, G., 2015a. Scale economies, technical change and efficiciency in Norwegian electricity distribution, 1998-2010. Journal of Productivity Analysis 43, 295–305.

Kumbhakar, S. C., Heshmati, A., 1995. Efficiency measurement in Swedish dariy farms: An application of rotating panel data. American Journal of Agricultural Economics 77 (3), 660–674.

Kumbhakar, S. C., Hjalmarsson, L., 1993. Technical efficiency and technical progress in swedish dairy farms. In: Fried, H. O., Lovell, C. A. K., Schmidt, S. S. (Eds.), The Measurement of Productive Efficiency. Techniques and Applications. Oxford University Press, Oxford, pp. 256–270.

Kumbhakar, S. C., Hjalmarsson, L., Jan.-Marc 1995. Labour-Use Efficiency in Swedish Social Insurance Offices. Journal of Applied Econometrics 10 (1), 33–47.

Kumbhakar, S. C., Lai, H., 2016. Maximum-likelihood estimation of the revenue function system with output-specific technical efficiency. Economics Letters 138, 42–45.

Kumbhakar, S. C., Lien, G., Hardaker, J. B., 2014. Technical efficiency in competing panel data models: A study of Norwegian grain farming. Journal of Productivity Analysis 41 (2), 321–337.

Kumbhakar, S. C., Wang, H.-J., Horncastle, A. P., 2015b. A Practitioner's Guide to Stochastic Frontier Analysis Using Stata. Cambridge University Press.

Laffont, J.-J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. The MIT Press.

Matthes, F. C., 2000. Stromwirtschaft und deutsche Einheit: eine Fallstudie zur Transformation der Elektrizitätswirtschaft in Ost-Deutschland. edition umwelt + energie.

Meeusen, W., Broeck, J. v. d., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18 (2), 435–444.

Moss, A., 2008. 'cold spots' of urban infrastructure: 'shrinking' processes in Eastern Germany and the modern infrastructural ideal. International Jounal of Urban Research and Regional Research 32 (2), 436–451.

Nieswand, M., Seifert, S., 2018. Environmental factors in frontier estimation–a monte carlo analysis. European Journal of Operational Research 265 (1), 133–148.

RDC, 2006-2012. Tech. Rep. 2006-2012, RDC of the Federal Statistical Office and Statistical Offices of the Länder.

Redding, S. J., Sturm, D. M., 2008. The costs of remoteness: Evidence from German division and reunification. The American Economic Review 98 (5), 1766–1797.

Samuelson, P. A., 1954. The pure theory of public expenditure. Review of Economics and Statistics 36 (4), 387–389.

Shleifer, A., 1985. A theroy of yardstick competition. The RAND Journal of Economics 16 (3), 319–327.

Sinn, H. W., 2002. Germany's economic unification: An assessment after ten years. Review of International Economics 10 (1), 113–128.

Snower, D. J., Merkl, C., 2006. The caring hand that cripples: The East German labor market after reunification. The American Economic Review 96 (2), 375–382.

Swiss economics and Sumicsid, 2014. Effizienzvergleich für Verteilernetzbetreiber Strom 2013. Tech. rep., Bundesnetzagentur, Bonn, Germany. URL https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Strom/Effizienzvergleich_VNB/2Regulierungsperiode/Gutachten_EVS2_geschw.pdf?__blob=publicationFile&v=2

Swiss Economics and Sumicsid, 2018. Effizienzvergleich Verteilernetzbetreiber Strom der dritten Regulierungsperiode. Tech. rep., Bundesnetzagentur, Bonn, Germany. URL https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Strom/Effizienzvergleich_VNB/3RegPer/Gutachten_Effizienzvergleich.pdf?__blob=publicationFile&v=2

Tiebout, C. M., 1956. A pure theory of local expenditures. The Journal of Political Economy 64 (5), 416–424.

Tsionas, E. G., Kumbhakar, S. C., 2014a. Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random effects model. Journal of Applied Econometrics 29 (1), 110–132.

Tsionas, E. G., Kumbhakar, S. C., 01 2014b. Firm Heterogeneity, Persistent and Transient Technical Inefficiency: A Generalized True Random-Effects Model. Journal of Applied Econometrics 29 (1), 110–132.

Appendix

A1. Methodology

A1.1. Full maximum likelihood method

Rewrite model (7) as

$$-\log x_{1,it} = \log f(\tilde{\boldsymbol{x}}_{-1,it}, \boldsymbol{y}_{it}; \boldsymbol{\beta}) + \epsilon_{0i} + \epsilon_{it}, \tag{13}$$

where $\epsilon_{it} = v_{it} - u_{it}$ and $\epsilon_{0i} = v_{0i} - u_{0i}$ decompose the error term into two 'composed error' terms (both of which contain inefficiency and noise terms). This decomposition will be useful later when we discuss the estimation of the model.

To obtain a tractable likelihood function, we follow Colombi et al. (2014) and draw results from skew normal and closed skew normal (CSN) distributions. Assuming v_{it} is an independent [in probability] random normal variable and u_{it} is an independent random half normal variable, ϵ_{it} in (13) has a skew normal distribution. Using the same argument, ϵ_{0i} in (13) has a skew normal distribution when v_{0i} is an independent random normal variable and u_{0i} is an independent random half normal variable. Thus, the composed error term $\epsilon_{0i} + \epsilon_{it}$ in (13) has a CSN distribution (being the sum of two independent skew normal distributions), which has a well defined pdf that is used to define the log-likelihood function, the maximization of which gives the MLE of all the parameters.

The model in (13) can be rewritten in a compact form, viz.,

$$-\log x_{1,i} = \log f(\tilde{\boldsymbol{x}}_{-1,i}, \boldsymbol{y}_i; \boldsymbol{\beta}) + \boldsymbol{1}_{T_i} v_{0i} + \boldsymbol{A}\boldsymbol{u}_i + \boldsymbol{v}_i, \quad \forall i \tag{14}$$

where bold symbols denote vectors for DSO i, $\boldsymbol{u}_i = (u_{0i}, u_{i1}, \ldots, u_{iT_i})'$, $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iT_i})'$, $\boldsymbol{A} = -[\boldsymbol{1}_{T_i} \boldsymbol{I}_{T_i}]$, $\boldsymbol{1}_{T_i}$ is the column vector of length T_i and $\boldsymbol{I}_{T_i}$ is the identity matrix of dimension T_i. Since the composed error term $\boldsymbol{\epsilon}_i = \mathbf{1}_{T_i} v_{0i} + \boldsymbol{A}\boldsymbol{u}_i + \boldsymbol{v}_i$ follows a CSN distribution, its joint density can be derived from the definition of a CSN probability density function, and the resulting panel level log-likelihood function of the four component model is given by:[19]

$$\log L_i\left(\boldsymbol{\beta}, \boldsymbol{\gamma}_{u0}, \boldsymbol{\gamma}_{v0}, \boldsymbol{\gamma}_u, \boldsymbol{\gamma}_v\right) = (T_i + 1)\log 2 + \log \phi_{T_i}\left(\boldsymbol{r}_i, \mathbf{0}, \boldsymbol{\Sigma}_i + \boldsymbol{A}\boldsymbol{V}_i\boldsymbol{A}'\right) + \log \overline{\Phi}_{T_i+1}\left(\boldsymbol{R}_i\boldsymbol{r}_i, \boldsymbol{\Lambda}_i\right), \tag{17}$$

where $\boldsymbol{r}_i = -\log x_{1,i} - \log f(\tilde{\boldsymbol{x}}_{-1,i}, \boldsymbol{y}_i; \boldsymbol{\beta})$, the diagonal elements of $\boldsymbol{V}_i$ are $[\exp\left(\boldsymbol{z}_{u0i}\boldsymbol{\gamma}_{u0}\right)\ \exp\left(\boldsymbol{z}_{uit}\boldsymbol{\gamma}_u\right)]$, $\boldsymbol{\Sigma}_i = \exp\left(\boldsymbol{z}_{vit}\boldsymbol{\gamma}_v\right)\boldsymbol{I}_{T_i} + \exp\left(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0}\right)\mathbf{1}_{T_i}\mathbf{1}'_{T_i}$,[20] $\boldsymbol{\Lambda}_i = \boldsymbol{V}_i - \boldsymbol{V}_i\boldsymbol{A}'\left(\boldsymbol{\Sigma}_i + \boldsymbol{A}\boldsymbol{V}_i\boldsymbol{A}'\right)^{-1}\boldsymbol{A}\boldsymbol{V}_i = \left(\boldsymbol{V}_i^{-1} + \boldsymbol{A}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{A}\right)^{-1}$, $\boldsymbol{R}_i = \boldsymbol{V}_i\boldsymbol{A}'\left(\boldsymbol{\Sigma}_i + \boldsymbol{A}\boldsymbol{V}_i\boldsymbol{A}'\right)^{-1} = \boldsymbol{\Lambda}_i\boldsymbol{A}'\boldsymbol{\Sigma}_i^{-1}$, $\phi_q\left(x, \boldsymbol{\mu}, \boldsymbol{\Omega}\right)$ is the density function of a q-dimensional normal variable with expected value $\boldsymbol{\mu}$ and variance $\boldsymbol{\Omega}$ and $\overline{\Phi}_q\left(\boldsymbol{\mu}, \boldsymbol{\Omega}\right)$ is the probability that a q-variate normal variable of expected value $\boldsymbol{\mu}$ and variance $\boldsymbol{\Omega}$ belongs to the positive orthant.

---

[19] Note, this model considers possible heteroscedasticity functions of noise and random effects

$$v_{0i} \sim N(0, \sigma^2_{v0i}) \text{ where } \sigma^2_{v0i} = \sigma^2_{v0}\exp\left(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0}\right), \quad i = 1, \cdots, n, \tag{15}$$

$$v_{it} \sim N(0, \sigma^2_{vit}) \text{ where } \sigma^2_{vit} = \sigma^2_v\exp\left(\boldsymbol{z}_{vit}\boldsymbol{\gamma}_v\right), \quad i = 1, \cdots, n, \quad t = 1, \cdots, T_i, \tag{16}$$

where σ^2_{v0} and σ^2_v are constants and $\boldsymbol{z}_{v0i}$ denotes the vector of time-invariant covariates that determine variance of random firm-effects. Similarly, $\boldsymbol{z}_{vit}$ denotes the vector of covariates that determine variance of both the firm-specific and time-varying random noise.

[20] Note that $\exp\left(\boldsymbol{z}_{u0i}\boldsymbol{\gamma}_{u0}\right)$ and $\exp\left(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0}\right)$ are both scalars, whereas $\exp\left(\boldsymbol{z}_{uit}\boldsymbol{\gamma}_u\right)$ and $\exp\left(\boldsymbol{z}_{vit}\boldsymbol{\gamma}_v\right)$ are both vectors of length T_i. Changing the notation of (8), (9), (15), and (16), $\boldsymbol{\gamma}_{u0}$, $\boldsymbol{\gamma}_{v0}$, $\boldsymbol{\gamma}_u$, and $\boldsymbol{\gamma}_v$ include an intercept, while the variance functions do not contain σ^2_{u0}, $\sigma^2_{v_0}$, σ^2_u, and σ^2_v, respectively. Thus, for example, if the random effects component is constant, $\boldsymbol{Z}_{v0i}$ is a constant and $\sigma^2_{v_{0i}} = \exp\left(\gamma_0\right)$ for all i.

A1.2. A simulated maximum likelihood estimator

Although the CSN framework gives a closed form expression of the log-likelihood function, implementing it in practice is a daunting task. Using the insights of Butler and Moffitt (1982), Filippini and Greene (2016) note that the density can be greatly simplified by conditioning on ϵ_{0i}. In this case, the conditional density is simply the product over time of T_i univariate skew normal densities. Thus, only a single integral, as opposed to T_i integrals, needs to be evaluated.

Recall that for each i, ϵ_{it} is a skew normal variate with parameters $\lambda_{it} = [\exp(\boldsymbol{z}_{uit}\boldsymbol{\gamma}_u)/\exp(\boldsymbol{z}_{vit}\boldsymbol{\gamma}_v)]^{1/2}$ and $\sigma_{it} = [\exp(\boldsymbol{z}_{uit}\boldsymbol{\gamma}_u) + \exp(\boldsymbol{z}_{vit}\boldsymbol{\gamma}_v)]^{1/2}$. Similarly, ϵ_{0i} is a skew normal variate with parameters $\lambda_{0i} = [\exp(\boldsymbol{z}_{u0i}\boldsymbol{\gamma}_{u0})/\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}$ and $\sigma_{0i} = [\exp(\boldsymbol{z}_{u0i}\boldsymbol{\gamma}_{u0}) + \exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}$. Thus, the conditional density of $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT_i})$ is given by

$$f(\boldsymbol{\epsilon}_i|\epsilon_{0i}) = \prod_{t=1}^{T_i} \frac{2}{\sigma_{it}}\phi\left(\frac{\epsilon_{it}}{\sigma_{it}}\right)\Phi\left(\frac{\epsilon_{it}\lambda_{it}}{\sigma_{it}}\right). \tag{18}$$

Integrate ϵ_{0i} (the distribution of which we know) out to get the unconditional density of $\boldsymbol{\epsilon}_i$

$$f(\boldsymbol{\epsilon}_i) = \int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i} \frac{2}{\sigma_{it}}\phi\left(\frac{\epsilon_{it}}{\sigma_{it}}\right)\Phi\left(\frac{\epsilon_{it}\lambda_{it}}{\sigma_{it}}\right)\right] \times \frac{2}{\sigma_{0i}}\phi\left(\frac{\epsilon_{0i}}{\sigma_{0i}}\right)\Phi\left(\frac{\epsilon_{0i}\lambda_{0i}}{\sigma_{0i}}\right)d\epsilon_{0i}. \tag{19}$$

The log-likelihood function for the i-th observation of model (13) is therefore given by

$$\begin{aligned}&\log L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}_{u0}, \boldsymbol{\gamma}_{v0}, \boldsymbol{\gamma}_u, \boldsymbol{\gamma}_v)\\ &= \log\left[\int_{-\infty}^{+\infty}\left(\prod_{t=1}^{T_i}\left\{\begin{array}{l}\frac{2}{\sigma_{it}}\phi\left(\frac{r_{it}-\epsilon_{0i}}{\sigma_{it}}\right)\\ \times\Phi\left(\frac{(r_{it}-\epsilon_{0i})\lambda_{it}}{\sigma_{it}}\right)\end{array}\right\}\right)\frac{2}{\sigma_{0i}}\phi\left(\frac{\epsilon_{0i}}{\sigma_{0i}}\right)\Phi\left(\frac{\epsilon_{0i}\lambda_{0i}}{\sigma_{0i}}\right)\mathrm{d}\epsilon_{0i}\right]\end{aligned}$$

$$= \log\left[\int_{-\infty}^{+\infty}\left(\prod_{t=1}^{T_i}\left\{\frac{2}{\sigma_{it}}\phi\left(\frac{\epsilon_{it}}{\sigma_{it}}\right)\Phi\left(\frac{\epsilon_{it}\lambda_{it}}{\sigma_{it}}\right)\right\}\right)\times\frac{2}{\sigma_{0i}}\phi\left(\frac{\epsilon_{0i}}{\sigma_{0i}}\right)\Phi\left(\frac{\epsilon_{0i}\lambda_{0i}}{\sigma_{0i}}\right)\mathrm{d}\epsilon_{0i}\right], \tag{20}$$

where $\epsilon_{it} = r_{it} - (v_{0i} + u_{0i})$. Although, following CSN, one can derive the likelihood function in closed form, we approximate the log-likelihood function and avoid using the classical ML method, which is quite complicated for the reasons mentioned above. We rely on Monte-Carlo integration as a method to approximate the integral in (20).[21] For estimation purposes, we write $\epsilon_{0i} = [\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}V_i + [\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}|U_i|$, where both V_i and U_i are standard normal random variables. The resulting simulated log-likelihood function for the i-th observation is

$$\begin{aligned}
&\log L_i^S(\boldsymbol{\beta}, \boldsymbol{\gamma}_{u0}, \boldsymbol{\gamma}_{v0}, \boldsymbol{\gamma}_u, \boldsymbol{\gamma}_v) \\
&= \log\left[\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i}\left\{\begin{array}{l}\dfrac{2}{\sigma_{it}}\phi\left(\dfrac{r_{it} - ([\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}V_{ir} + [\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}|U_{ir}|)}{\sigma_{it}}\right) \\ \times\Phi\left(\dfrac{[r_{it} - ([\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}V_{ir} + [\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}|U_{ir}|)]\lambda}{\sigma_{it}}\right)\end{array}\right\}\right)\right] \\
&= \log\left[\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i}\left\{\frac{2}{\sigma}\phi\left(\frac{\epsilon_{itr}}{\sigma}\right)\Phi\left(\frac{\epsilon_{itr}\lambda}{\sigma}\right)\right\}\right)\right],
\end{aligned} \tag{21}$$

where V_{ir} and U_{ir} are R random deviates from the standard normal distribution, and $\epsilon_{itr} = r_{it} - ([\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}V_{ir} + [\exp(\boldsymbol{z}_{v0i}\boldsymbol{\gamma}_{v0})]^{1/2}|U_{ir}|)$. R is the number of draws for approximating the log-likelihood function. The full log-likelihood is the sum of panel-i specific log-likelihoods given in (21).

We use the results of Colombi, Kumbhakar, Martini, and Vittadini (2014) to estimate persistent and time-varying cost efficiencies. Using the moment generating function of the CSN distribution, the conditional means of $u_{0i}, u_{i1}, \cdots, u_{iT_i}$ which are, in

[21] Note that another approximation of (20) can be achieved by using the M-point Gauss-Hermite quadrature method.

principle, similar to the Jondrow et al. (1982) estimator, are given by:

$$E(\exp\{\boldsymbol{t}'\boldsymbol{u}_i\}|\boldsymbol{y}_i) = \frac{\overline{\Phi}_{T_i+1}\left(\boldsymbol{R}_i\boldsymbol{r}_i + \boldsymbol{\Lambda}_i\boldsymbol{t}, \boldsymbol{\Lambda}_i\right)}{\overline{\Phi}_{T_i+1}\left(\boldsymbol{R}_i\boldsymbol{r}_i, \boldsymbol{\Lambda}_i\right)} \times \exp\left(\boldsymbol{t}'\boldsymbol{R}_i\boldsymbol{r}_i \ + 0.5\boldsymbol{t}'\boldsymbol{\Lambda}_i\boldsymbol{t}\right), \quad (22)$$

where $\boldsymbol{u}_i = (u_{0i}, u_{i1}, \ldots, u_{iT_i})'$ and $-\boldsymbol{t}$ is a row of the identity matrix of dimension $(T_i + 1)$. If $-\boldsymbol{t}$ is the τ-th row, Eq (22) provides the conditional expected value of the τ-th component of the cost efficiency vector $\exp\left(-\boldsymbol{u}_i\right)$. In particular, for $\tau = 1$, we get the conditional expected value of the persistent technical efficiency.