UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Paper Economics

Cambridge Working Paper Economics: 1673

Using a spatial econometric approach to mitigate omitted variables in stochastic frontier models: An application to Norwegian electricity distribution networks

Luis Orea

Inmaculada C. Álvarez

Tooraj Jamasb

8 December 2016

An important methodological issue for the use of efficiency analysis in incentive regulation of regulated utilities is how to account for the effect of unobserved cost drivers such as environmental factors. This study combines the spatial econometric approach with stochastic frontier techniques to control for unobserved environmental conditions when measuring firms' efficiency in the electricity distribution sector. Our empirical strategy relies on the geographic location of the firms as a useful source of information that has previously not been explored in the literature. The underlying idea in our empirical proposal is to utilise variables from neighbouring firms that are likely to be spatially correlated as proxies for the unobserved cost drivers. We illustrate our approach using the data of Norwegian distribution utilities for the years 2004 to 2011. We find that the lack of information on weather and geographic conditions can likely be compensated with data from surrounding firms using spatial econometric techniques. Combining efficiency analysis and spatial econometrics methods improve the goodness-of-fit of the estimated models and, hence, more accurate (fair) efficiency scores are obtained. The methodology can also be used in efficiency analysis and regulation of other types of utility sectors.

Using a spatial econometric approach to mitigate omitted variables in stochastic frontier models: An application to Norwegian electricity distribution networks

EPRG Working Paper 1630

Cambridge Working Paper in Economics 1673

Luis Orea, Inmaculada C. Álvarez, and Tooraj Jamasb

Abstract

An important methodological issue for the use of efficiency analysis in incentive regulation of regulated utilities is how to account for the effect of unobserved cost drivers such as environmental factors. This study combines the spatial econometric approach with stochastic frontier techniques to control for unobserved environmental conditions when measuring firms' efficiency in the electricity distribution sector. Our empirical strategy relies on the geographic location of the firms as a useful source of information that has previously not been explored in the literature. The underlying idea in our empirical proposal is to utilise variables from neighbouring firms that are likely to be spatially correlated as proxies for the unobserved cost drivers. We illustrate our approach using the data of Norwegian distribution utilities for the years 2004 to 2011. We find that the lack of information on weather and geographic conditions can likely be compensated with data from surrounding firms using spatial econometric techniques. Combining efficiency analysis and spatial econometrics methods improve the goodness-of-fit of the estimated models and, hence, more accurate (fair) efficiency scores are obtained. The methodology can also be used in efficiency analysis and regulation of other types of utility sectors.

Keywords	Spatial econometrics, stochastic frontier models, environmental conditions, electricity distribution networks.
JEL Classification	D24, L51, L94.

Contact	lorea@uniovi.es
Publication	December 2016
Financial Support	This research was partially funded by the Government of the Principality of Asturias and the European Regional Development Fund (ERDF).

Using a spatial econometric approach to mitigate omitted variables in stochastic frontier models: An application to Norwegian electricity distribution networks *

Luis Orea *
Oviedo Efficiency Group, University of Oviedo

Inmaculada C. Álvarez
Oviedo Efficiency Group, Universidad Autónoma de Madrid

Tooraj Jamasb
Durham University Business School, Durham

Abstract

An important methodological issue for the use of efficiency analysis in incentive regulation of regulated utilities is how to account for the effect of unobserved cost drivers such as environmental factors. This study combines the spatial econometric approach with stochastic frontier techniques to control for unobserved environmental conditions when measuring firms' efficiency in the electricity distribution sector. Our empirical strategy relies on the geographic location of the firms as a useful source of information that has previously not been explored in the literature. The underlying idea in our empirical proposal is to utilise variables from neighbouring firms that are likely to be spatially correlated as proxies for the unobserved cost drivers. We illustrate our approach using the data of Norwegian distribution utilities for the years 2004 to 2011. We find that the lack of information on weather and geographic conditions can likely be compensated with data from surrounding firms using spatial econometric techniques. Combining efficiency analysis and spatial econometrics methods improve the goodness-of-fit of the estimated models and, hence, more accurate (fair) efficiency scores are obtained. The methodology can also be used in efficiency analysis and regulation of other types of utility sectors.

Keywords: spatial econometrics, stochastic frontier models, environmental conditions, electricity distribution networks.

JEL classification: D24, L51, L94.

* Corresponding author. Email: lorea@uniovi.es. This research was partially funded by the Government of the Principality of Asturias and the European Regional Development Fund (ERDF). We would also like to thank the Norwegian Water Resources and Power Directorate (NVE) for providing accurate data and descriptions and detailed advice on regulation of electricity distribution networks in Norway.

1. Introduction

Since the 1990s many network utilities are incentive regulated with the aim of improving their operating and investment efficiency as well as ensuring that consumers benefit from the gains. In many instances, the regulators aim to measure the firms' relative efficiency against those with best practice performance using parametric and non-parametric techniques (see Haney and Pollitt, 2013). As regulators reward or penalise firms using relative efficiency measures, obtaining reliable (and fair) measures of firms' efficiency requires controlling for the different environmental conditions under which each utility operates. This is particularly important in the case of incentive regulation and benchmarking of electricity, gas, and water networks where the results of efficiency analysis have important financial implications for the firms.

However, there are many characteristics of the utilities sector (e.g. geography, climate or network characteristics) that affect production costs but which are unobserved (Farsi and Filippini, 2004). Statistical methods have recently been developed to address this issue. For instance, the True Fixed/Random Effects models introduced by Greene (2005) capture the unobserved heterogeneity through a set of firm-specific intercepts. This approach only uses the temporal (i.e. within) variation contained in the data to estimate the coefficients of other cost drivers. This is quite problematic in our application because many crucial determinants of utility costs such as the energy delivered or number of customers, are persistent or slow changing variables (see Greene et al, 2011). On the other hand, possible differences among utilities associated with their use of different technologies are also often addressed using simple sample selection procedures or by using clustering methods. Recently, conventional latent class stochastic frontier models account for technology heterogeneity among firms belonging to different groups (Llorca et al., 2014).

In this paper we advocate using a new empirical strategy to account for the unobserved differences in environmental conditions among electricity distribution networks based on their geographic location. The latter presents an invaluable source of information that has been ignored in the literature which up to now was dedicated only to estimating network technology or the measurement of their relative inefficiency. Indeed, as many unobservable variables are likely to be spatially correlated, an alternative empirical strategy emerges. Our spatial model is prompted from the fact that any (relevant) unobservable cost driver should be correlated with firms' costs, a variable that is observable by the researcher/regulator. The underlying idea of our empirical proposal is to use (surrounding) firms' costs as proxies of the *unobserved* cost drivers that are likely to be spatially correlated, such as weather and geographic conditions, population structure, electricity demand patterns, input prices, etc. In line with True Fixed/Random Effects models, our approach allows for firm-specific technologies. However, our empirical strategy does not ignore the cross-sectional nature of the data and the inherent information. As with latent class models, the estimated unobserved heterogeneity is allowed to change over time.

The main contribution of this paper is to link efficiency analysis methods addressing unobserved heterogeneity with spatial econometrics methods commonly employed to examine spatial interactions across regions.[1] To the best of our knowledge, our paper is among the first to apply spatial econometrics in efficiency analysis using firm level data. There are no major systemic economic or technical reasons that the conditional cost of a firm (i.e. given its own output and price variables), is affected by those of adjacent firms to any significant degree.[2] In this context, the estimated spillover effects in our model are expected to be spurious, i.e. only caused by omitted variables. This in turn implies that our spatial specification introduces constraints on the parameters, instead of the traditional spatial model. Moreover, the spatial econometric models are used (interpreted) here as a means to control for unobserved heterogeneity in a standard SFA model measuring firms' inefficiency.[3]

The next section presents the spatial econometric model that allows us to use data from surrounding firms as proxies of the omitted, but spatially correlated, cost drivers. Section 3 summarizes the empirical strategy used in this paper to estimate a SFA model that includes a generated variable as an additional regressor. Section 4 dwells on the data

[1] Since the seminar book by Anselin (1988) introducing the existence of spatial effects in econometric models, many authors have developed several spatial econometric models and their estimation methods (see, for instance, Kelejian and Prucha, 1998, 2010 and Baltagi and Liu, 2011). For comprehensive reviews of this literature, see Arbia (2014) and Elhorst (2014). Regarding our spatial approach, Elhorst (2010) provide a detailed discussion of most common spatial econometrics models and highlight the fact that so-called Spatial Durbin Model (SDM) is a good starting point to contrast the correct specification of the potential spatial effects underlying the data generating process. They also point out that a model with spatial autocorrelation in not observable variables (the so-called SEM) can be expressed as an SDM with constraints, which is the idea behind our proposal.

[2] We are thankful to the NVE staff in charge of network regulation who could confirm this point.

[3] See Glass et al (2016) for a recent application with spatial effects in SFA settings.

used in the empirical analysis and its sources. In Section 5 we first estimate a spatial econometric model to compute a proxy variable that will stand in for spatially correlated omitted variables. We then estimate a standard SFA model to estimate firms' inefficiency. A robustness analysis using available environmental data is also provided. Finally, Section 6 presents the conclusions.

2. A cost model with (unobserved) spatially correlated variables

This section develops a micro-level spatial econometric model that allows us to control for unobserved environmental conditions that are likely to be spatially correlated when we use a cost function to estimate the firms' technology. Let us first assume that the firms' cost can be modelled entirely by using the following cost equation:

$$lnC_{it} = \beta X_{it} + Z_{it} + v_{it} + u_{it} \tag{1}$$

where i stands for firms, t stands for periods, C_{it} is a measure of firms' cost, and X_{it} is a vector of k observable cost drivers such as the number of customers, energy delivered, network length, and labour and capital prices and Z_{it} represents the *unobserved* cost drivers. This equation includes two error terms, v_i and u_i. While the former term is a symmetric error term measuring pure random shocks, the latter term is a non-negative error term measuring firms' inefficiency.

As is often the case with observed data,[4] some unobserved cost drivers are also likely to be spatially correlated. In line with the literature on spatial econometrics, the spatial correlation can be modelled as follows:

$$Z_{it} = \lambda W_i Z_t \tag{2}$$

Here Z_t is a vector of Nx1 unobserved cost drivers, W_i is a known 1xN spatial weight vector with elements that are equal to zero if a particular firm j is not a neighbour of firm i and equal to one if the two firms are neighbours – i.e. the service areas of the electricity distribution utilities are adjacent. The term λ is a coefficient that measures the degree of spatial correlation between the unobserved cost drivers.

[4] For illustration purposes, we show several auxiliary regressions in Appendix A where we have used equation (2) to examine the degree of spatial correlation for some of our observed cost drivers. As expected, we find that all variables are spatially correlated to some extent. Therefore, it is reasonable to expect some degree of spatial correlation also in unobserved determinants of firms' costs.

Equation (1) cannot be *directly* estimated as Z_{it} is an omitted variable that, if ignored, will bias our efficiency scores because it will be captured by the noise or inefficiency terms. We thus propose using an *indirect* approach to estimate (1). The underlying idea behind our proposal is that we could use the (purged) costs of surrounding firms as proxies for Z_{it} if the unobserved cost drivers are spatially correlated. Hence, our empirical strategy takes advantage of the spatial proximity of the networks.

First, we proceed to replace Z_{it} in equation (1) with equation (2). Thus, equation (1) can be alternatively rewritten as follows:

$$lnC_{it} = \beta X_{it} + \lambda W_i Z_t + v_{it} + u_{it} \tag{3}$$

This equation again cannot be estimated as the vector Z_t is not observed. However, note that, by rearranging equation (1), we can obtain:

$$Z_{it} = lnC_{it} - \beta X_{it} - v_{it} - u_{it} \tag{4}$$

This equation simply indicates that, if β and both errors terms were observable, Z_{it} should be correlated with a purged cost measure. In this sense, the purged costs can be interpreted as an "observable" counterpart of Z_{it}. We then replace Z_t in equation (3) with its "observable" counterpart, obtaining the following model:

$$lnC_{it} = \beta X_{it} + \lambda W_i lnC_t - \lambda\beta W_i X_t + \varepsilon_{it} \tag{5}$$

where

$$\varepsilon_{it} = h_{it} + v_{it} + u_{it} \tag{6}$$

and

$$h_{it} = -\lambda W_i (v_t + u_t) \tag{7}$$

$C_t = (C_{1t}, C_{2t}, \dots, C_{Nt})$ is an Nx1 vector of observed costs of firms, $X_t = (X_{1t}, X_{2t}, \dots, X_{Nt})$ is an Nx1 vector of firms' explanatory variables, and v_t and u_t are again Nx1 vectors of the firms' random terms.

Several comments are in order with respect to this specification of the firms' cost. First, if we compare the original model in (1) and the new specification in (5)-(7), we notice that:

$$Z_{it} = \hat{Z}_{it} + h_{it} \tag{8}$$

where

$$\hat{Z}_{it} = \lambda W_i lnC_t - \lambda\beta W_i X_t \tag{9}$$

Equation (8) simply shows that the unobserved cost driver Z_{it} can be decomposed into a *predictable* component $\hat{Z}_{it}$ (i.e. the portion of Z_{it} that can be predicted with the data of surrounding firms), and an *unpredictable* component h_{it}. The latter term can in turn be interpreted as a measurement error term. As the inefficiency term is non-negative, h_{it} is negative on average, and hence our predicted $\hat{Z}_{it}$ tends to *overestimate* the true value of the omitted variable Z_{it}.

Second, in contrast to equation (1), equation (5) is a cost model that now includes a set of spatially lagged variables, i.e. $W_i lnC_t$ and $W_i X_t$. Therefore, equation (5) resembles a conventional spatial econometric model. However, in our model, only one additional coefficient is estimated, and the coefficient of the spatially lagged dependent variable should not be interpreted as the effect of neighbours' costs on the cost of a particular firm. Rather, λ is measuring the spatial correlation between the unobserved or omitted variables in our sample. Our empirical strategy relies on the statistical significance of this coefficient as we are unable to use the data of surrounding firms to obtain a proxy for Z_{it} if $\lambda = 0$. Therefore, it is important for our empirical strategy to test whether this parameter is statistically significant.

On the other hand, it is worth mentioning that our spatial specification of firms' costs in equation (5) is similar to the Durbin Stochastic Frontier (SDF) model introduced recently by Glass et al. (2016) in which they propose estimating the following model:

$$lnC_{it} = \beta X_{it} + \lambda W_i lnC_t + \theta W_i X_t + \tilde{\varepsilon}_{it} \tag{10}$$

where $\tilde{\varepsilon}_{it} = v_{it} + u_{it}$. It is easily observable that our spatial model in (5)-(7) and the SDF model differ in two important aspects. First, the set of parameters θ in the SDF model is not restricted to be equal to $-\lambda\beta$. In this sense, our spatial model in (5) is nested in the SDF model. However, no spatially correlated omitted (random) variables are explicitly modelled in the SDF model. Although Glass et al. (2016) state that their approach can be "easily adapted to develop a spatial error stochastic frontier model", they do not include a spatial structure in the error term. In terms of our spatial model, this is equivalent to using a zero h_{it} term. The mentioned differences simply indicate that our spatial model and the SDF model are non-nested. This is because the spatial spillovers in both models are of different nature. While the spatial spillovers in Glass et al. (2016) have an economic or causal interpretation, the spatial spillovers in our spatial model are simply associated with the omitted variables. Hence the spatial effects estimated in our model lack an economic interpretation as they are completely “spurious”.

We next discuss how to estimate our spatial SFA model taking into account that ε_{it} includes two spatially correlated error terms (see equations 6 and 7). If the spatial error correlation involves a one-sided error term, this does not prove to be an easy task. In order to gain an idea of this, we rewrite again our spatial model in equations (5)-(7) as follows:

$$lnC_{it} = [\beta X_{it} + \lambda W_i lnC_t - \lambda\beta W_i X_t] + \Delta v_{it} + \Delta u_{it} \tag{10}$$

where

$$\Delta v_{it} = v_{it} - \lambda W_i v_t$$

$$\Delta u_{it} = u_{it} - \lambda W_i u_t$$

It should be pointed out that while Δv_{it} follows a multivariate normal distribution, the distribution of Δu_{it} (i.e. the difference of, say, two independent one-sided error terms) is not known, and this prevents using a ML estimator (see Wang, 2003; and Wang and Ho, 2010). As a fully ML specification of the model is not feasible in our case, in the next section we propose a procedure that includes $\hat{Z}_{it}$ as an additional regressor, and controls for h_{it} by using a simple linear function of its determinants.

3. Stochastic frontier model with generated regressor

Our estimation strategy uses a two-step procedure, advocated for various models in Kumbhakar and Lovell (2000). In the first step, equation (5) are estimated ignoring the (spatial and frontier) structure of the error term, ε_{it}. The degree of spatial correlation of omitted variables (i.e. parameter λ) and other coefficients of the cost frontier are estimated using the Generalized Method of Moments (GMM) because the spatially lagged dependent variable is endogenous. It is worth noting that, as in previous literature on both spatial and SFA models using two-stage procedures, the first-step GMM residuals are not used here to estimate the *complete* structure of the overall term ε_{it} because its distribution is not known. Instead, the first-step estimates aim to obtain a prediction of Z_{it} that is used in a second regression as an additional explanatory variable.

In the second step, the following specification of firms' cost in equation (1) is estimated:

$$lnC_{it} = \beta X_{it} + \gamma_{it}\hat{Z}_{it} + v_{it} + u_{it} \tag{11}$$

where

$$\gamma_{it} = \frac{Z_{it}}{\hat{Z}_{it}} = \frac{\hat{Z}_{it}+h_{it}}{\hat{Z}_{it}}, \tag{12}$$

In order to obtain (11), we have replaced the original omitted variable in (1) with its predicted counterpart using equation (9). The ratio γ_{it} can be interpreted here as a firm-specific and time-varying coefficient, that tends to be less than unity because h_{it} is on average less than zero. In our empirical application, we will first estimate a common γ value for all firms, so that the final cost model estimated in our paper is:

$$lnC_{it} = \beta X_{it} + \gamma\hat{Z}_{it} + v_{it} + u_{it} \tag{13}$$

where the common γ coefficient can now be interpreted as the average value of γ_{it}. The fact that h_{it} does not appear in (13) does not imply that we are (completely) ignoring the spatial part of the composed error term ε_{it} in (10) because h_{it} is roughly captured (at least its average value) by an estimate of γ that will depart from the theoretical value of unity.

It should be pointed out, however, that γ_{it} is a function of h_{it}, which on average depends on the number of adjacent firms (i.e. W_i) and the inefficiency level of adjacent firms (i.e. the magnitude of u_t). Therefore, more accurate estimates can be obtained if we model γ_{it} as a linear function of the number of adjacent firms (N_i) and, if the SFA model is heteroskedastic, the spatial lags of all determinants of firms' inefficiency (W_iq_i), that is:

$$\gamma_{it} = \gamma_0 + \gamma_1 N_i + \gamma_2 W_i q_i \tag{14}$$

Therefore, our preferred specification of the second-step model is:

$$lnC_{it} = \beta X_{it} + (\gamma_0 + \gamma_1 N_i + \gamma_2 W_i q_i)\hat{Z}_{it} + v_{it} + u_{it} \tag{15}$$

Finally, note that, conditional on $\hat{Z}_{it}$, our new specification of firms' cost has the structure of a conventional SFA model, so it can be estimated using MLE techniques once the distributional assumptions concerning the noise and inefficiency terms are made. As is common in the SFA literature, we will assume that $v_{it} \sim N(0, \sigma_v)$ and the inefficiency term are independently distributed across firms and over time, and follows a half-normal distribution, i.e. $u_{it} \sim N^{+}(0, \sigma_u)$.[5] As anticipated above, this model can accommodate heteroskedastic inefficiency terms simply by making the variance of σ_u functions of some exogenous variables (q_{it}). Regardless of whether the model is homoscedastic or not, efficiency scores are estimated for each firm using the conditional distribution of u_{it} given v_{it}+ u_{it} introduced by Jondrow et al. (1982).

4. Data

We apply our empirical strategy to a balanced set of panel data for the Norwegian distribution utilities over the years 2004 to 2011. The data used in this study was obtained from the sector regulator, the Norwegian Water Resources and Power Directorate (NVE). We specify a simple cost model that uses, in line with the Norwegian benchmarking approach, *social costs* (SCOST) as the dependent variable. In addition to operating expenses (OPEX), capital depreciation and its opportunity cost, the social costs variable also includes the cost of network energy losses, and the cost of energy not supplied (CENS) to different user groups from service interruptions. The cost of network energy losses is obtained by multiplying the units of network energy losses with the average system price in NordPool wholesale market in a given year. CENS is calculated by multiplying the energy not supplied (KWh) during a specific interruption with a unit cost (NOK/KWh) that depends on customer type, duration, and whether the interruption was planned or not.

We follow the previous literature to select the main cost drivers. In particular, all of our estimated cost functions include three outputs (CUS=number of customers; NL=network length; and DE=delivered energy), and three input prices (PK= capital price, regulated return of capital; PE=energy price; PL=labour price).[6] We also use the percentage share of overhead lines (OH) of the total network length as an additional cost driver. This variable is employed to represent the main technical feature in this industry as firms' decisions on, for example, investment and maintenance of overhead and underground lines, are different. Regarding firms' inefficiency, we follow Orea and Jamasb (2017) and use the percentage of overhead lines (OH), the network length variable (NL) and the number of transformer stations (ST) as inefficiency determinants. We

[5] The stochastic frontier model can accommodate heteroskedastic inefficiency terms simply by making the variance of σ_u functions of some exogenous variables.
[6] Energy Price is used to impose linear homogeneity. For this reason, it will not explicitly appear in our set of parameter estimates.

include ST and OH as efficiency determinants to examine whether it is more costly to manage firms with more stations and with higher share of overhead lines. These can also be viewed as measures of complexity of networks something that some regulatory benchmarking models are currently lacking. Finally, NL allows us to know whether larger utilities tend to be more efficient than smaller utilities. The monetary variables finally used in our application are measured in 1000 NOK and have been deflated using the consumer price index to express them in 2004 real terms.

For robustness analyses, we will extend the above set of cost drivers to include several geographic and weather (W&G) variables. In particular, in our extended models we include six environmental variables: WIND=average reference wind from measuring stations; WINDEX=expected extreme wind exposure; and DIS=average distance to coast; FOREST=a measure of forest density in the service areas of networks; AVESLOPE=average slope of terrain; and MAXSLOPE=maximum slope of terrain.

The above geographic and weather variables were obtained from the Norwegian regulator. The regulator has access to more than 60 weather and geographic condition variables that can potentially affect the performance of networks. However, for practical reasons only a few of these variables can be included in parametric efficiency analysis models. Most of our selected environmental variables are considered as relevant by the Norwegian regulator. For instance, the regulator uses the ratio of squared wind speed over distance to coast in order to reflect the effect of coastal climate and corrosion caused by a combination of wind and salt water on the networks. Similar comments deserve our variables measuring the slope of terrain. Moreover, the regulator considers a range of variables in pre-benchmarking analysis to account for the effect of types of forestation in the service area, as fast-growing forest may represent a cost disadvantage due to the added cost of forest clearing. We use here an aggregate measure of forestation (FOREST) that has been computed using principal component analysis as we encountered convergence problems in Orea and Jamasb (2017) when we included the whole set of available variables to account for forest conditions.

In our study we follow the common approach in the literature for capturing and measuring the spatial interdependence using a physical contiguity matrix, *W*, whose elements are one for two bordering areas, and zero otherwise. As a result, the diagonal elements of *W* are null, while its off diagonal entries take a value of 1 for the areas that are adjacent and 0 otherwise. Therefore, WX should be interpreted as the sum of the X variables for the adjacent areas. The same applies for the WY product. In order to include the spatial interactions, we consider the map showing the distribution of service areas provided by NVE in October, 2015 (see Figure 1). This map is georeferenced using the ArcGIS data system. We have used this georeferenced information to identify the adjacent distribution areas.

[Insert Figure 1 here]

Finally, it is worth mentioning that our observations are the distribution areas of distribution utilities. Both the data on firms' costs and the map provided by the Norwegian regulator include the name of the distribution utilities. This information allowed us to match the distribution areas with the data of the firms operating in those service areas. The data for each distribution area normally coincides with the data of a single firm. However, the data for a small number of distribution areas involves more than one firm because they were involved in merger processes from 2004 onwards and their individual distribution areas are not available due to the map provided by the Norwegian regulator only shows the distribution of service areas many years later. We only have data on the overall distribution area of these merged firms in 2015. This forced us to aggregate the data of merged firms from 2004 onwards.[7]

Table 1 provides a descriptive summary of the variables used in this study. As the number of distribution areas in 2015 with available data is 129, the total number of observations used in our analysis is 1032.

[Insert Table 1 here]

5. Results

5.1. First-stage GMM regression and predicted values of the omitted cost drivers

We first estimate equation (5) using GMM in order to control for the endogeneity of the spatial lagged dependent variable. Table 2 shows the estimated coefficient of this variable. We do not provide the other coefficients of the model in this table as they are

[7] In previous specifications of our models, we included a merger dummy variable to control for possible aggregation biases. As expected, the coefficient of this variable was not statistically significant due to the small number of observations involved in merger processes during the period analysed in this paper.

similar to those obtained in the next section, mainly focused on the technological characteristics of firms' cost frontier.

[Insert Table 2 here]

We observe that the coefficient of spatial correlation λ is positive and significant. Hence, we conclude that the *unobserved* cost drivers are, at least to some extent, spatially correlated. This result also indicates that weather and geographic conditions, and other spatially unobserved cost drivers (such as the population structure, electricity demand patterns, input prices) matter and that they should be included as cost determinants.[8]

The fact that the coefficient of spatial correlation λ is statistically different from zero implies that we can use equation (9) and the data of surrounding firms to compute a proxy variable for the omitted cost drivers. The predicted values of the omitted cost drivers are summarized in Figure 2, where we plot kernel density functions of the percentage of cost attributable to (unfavourable) environmental conditions, measured in relation to the "average" firm. Figure 2 thus suggests the existence of remarkable cost differences between utilities attributable to different environmental conditions. This is most probably what regulators wish to control for.

[Insert Figure 2 here]

The firm with the most unfavourable omitted conditions has 33.5% higher costs than the representative firm. On the other hand, the firm with the most favourable omitted costs has 22.5% less costs than the representative firm. Orea et al. (2015) have found similar results using supervised environmental composite variables. For instance, their preferred model predicts up to 35% higher costs for utilities operating in areas with unfavourable environmental conditions. For utilities operating in good environmental conditions, their preferred model predicts up to 44% lower costs.

Table 3 shows the between and within standard deviations of the predicted values of the omitted variables and the main observed drivers of firms' costs. It is worth mentioning that the within-variation of $\hat{Z}_{it}$ is only slightly lower than the between-variation. Thus, our approach based on a spatial econometric model to capture unobserved heterogeneity uses both the between and within-variation contained in the data of neighbouring firms. In contrast, a FE-type estimator only uses the within-variation

[8] Growitsch et al. (2012) have found a similar conclusion using a different approach to control for unobserved and observed environmental conditions.

contained in the data to estimate the coefficients of the other cost drivers. If we use one of these estimators we will obtain negative and statistically non-significant coefficients for delivered energy, number of customers, network length and other crucial determinants of utility costs. The low precision of a FE-type estimator in the present application is caused by the fact that the within-variations of most of these variables tend to be much lower than the between-variation (see Table 3).

[Insert Table 3 here]

5.2. Second-stage MLE parameter estimates

Once we have generated a proxy variable for the omitted cost drivers, we proceed to estimate the stochastic cost frontier in equation (15) without the W&G variables. The results adding environmental variables are discussed later on.

In Table 4 we show four alternative specifications of the stochastic cost frontier. The simple-SFA model does not include the estimated values for Z_{it}, and it is only included for comparison purposes. The next three models include the generated variable $\hat{Z}_{it}$ as a proxy for the omitted variable Z_{it}. In this sense, they are labelled as "spatial" models. The spatial-SFA1 model only includes the generated variable $\hat{Z}_{it}$. The subsequent model (spatial-SFA2) adds the number of adjacent firms (N_i) to the specification of γ_{it}. Finally, as the inefficiency term is heteroskedastic, the spatial-SFA3 model extends the previous one by adding the spatial lags of all determinants of firms' inefficiency.

[Insert Table 4 here]

It should be noted that, compared to the simple-SFA model, the simplest spatial model that only adds the estimated values for Z_{it} improves the joint significance considerably, based on the likelihood function value. The estimated value of γ_0 is smaller than unity, an expected result as $\hat{Z}_{it}$ tends to overestimate the true values of Z_{it}. The next two spatial SFA models allow for firm-specific values of γ_{it}. In this case, as all variables are mean-centred, γ_0 can be interpreted as the sample mean value of γ_{it}. It is worth mentioning that the new spatial models again improve the likelihood function values. Interestingly, the estimated value for γ_0 in both models is now not statistically different from unity. This seems to indicate that only controlling for the number of adjacent firms is enough to obtain the unbiased value of γ_{it}, at least evaluated at the sample mean. This supports our empirical strategy based on a linear specification of γ_{it} that takes into account that h_{it} is the sum of several inefficiency terms, so its expected value depends on the number of adjacent firms (and their average inefficiency levels, which in turn depends on their efficiency determinants).

Regarding the parameters of the cost frontiers, generally all the first-order coefficients have the expected sign and their magnitudes are also reasonable from a theoretical standpoint. The first-order coefficients of all three outputs are positive and statistically different from zero. A similar observation can be made with respect to the coefficients of input prices, which are also positive and statistically significant. The frontier coefficient of OH is negative and statistically significant in all models, indicating that the larger the percentage of overhead lines, the smaller is the total cost. This result indicates that, although underground cables are probably negatively correlated with CENS and reduce OPEX, they are costlier and therefore increase the total costs.

The sum of the first-order coefficients of customer numbers and energy delivered allows us to measure density economies, associated with *vertical* output, i.e. output expansions that do not require additional network in the existing service areas. We find that the elasticity of density evaluated at the sample mean is quite similar in all models, i.e. 0.48. The estimated coefficients for these two outputs in Table 4 indicate that electricity distribution networks have strong natural monopoly characteristics. In contrast, scale economies are associated with *horizontal* output expansions that require enlarging the existing network. These economies can be measured by the sum of cost elasticities with respect to customer numbers, energy delivered and network length. The elasticity of scale evaluated at the sample mean in both models is about 0.94. This value suggests that Norwegian electricity distribution networks still exhibit natural monopoly characteristics when the network is expanded to meet new demand.[9]

In addition to the frontier parameters, Table 4 displays the coefficients of the variables that are related to the inefficiency term. The lack of significance of the coefficient of OH seems to indicate that managing firms with a relatively large proportion of overhead lines (more likely to be serving rural areas), have been managed similarly to those firms with more underground lines (more likely to be serving urban areas).

[9] These results are in line with the actual features of the Norwegian electricity distribution networks. While Norway has one of the highest levels of per capita energy consumption in the world, with the exception of a few cities, the number of network utilities is large relative to the population and, on the whole, the customer density across the networks is generally low.

Following Orea and Jamasb (2017), in addition to the percentage of overhead lines, we have included the logs of the network length (NL) in order to capture the size effects on firms' inefficiency, and the number of substations (ST) as a proxy for network complexity. As mentioned in our previous paper, we obtain a negative and statistically significant coefficient for NL, indicating that larger utilities tend to be more efficient than smaller utilities. In contrast, the positive coefficients of ST indicate that it is costlier to manage firms with more stations.

5.3. Efficiency Scores

Table 5 presents the summary statistics of the efficiency scores. Our efficiency estimates are high, on average about 92% using our preferred model (Spatial SFA 3). The high level of efficiency of this industry is most probably attributable to the maturity of Norway as a regulator that has consistently been supervising and incentivizing the Norwegian utilities to perform efficiently. Similar figures are obtained in Orea et al. (2015) using a SFA approach for the period 2004 to 2011, Miguéis et al. (2012) using a DEA method for the period 2004 to 2007, and in Growitsch et al. (2012) using a SFA approach for the 2001-2004 period.

[Insert Table 5 here]

On the other hand, it should be pointed out that the estimated efficiency levels in the models with spatial interactions (about 92.5%), are slightly higher than those obtained using the single SFA model (on average 91.6%), indicating that ignoring the omitted variables of a spatially correlated nature tends to *underestimate* the firms' efficiency scores. However, the small difference found between the single and the spatial SFA models might be suggesting that this bias is not severe. We observe in the next subsection that this is not the case.

5.4. Robustness analysis using weather and geographic data

One advantage of the present application is that the Norwegian energy regulator (NVE) has systematically examined the effects of several environmental factors such as geographic and weather conditions on cost and service quality performance of the utilities and it has reflected these in the cost efficiency benchmarking models used in incentive regulation of these utilities (see, e.g., Growitsch et al., 2012; Orea et al., 2015). This information is often not available in most other countries because collecting all the relevant environmental data requires a substantial effort and financial resources as well as considerable time. Therefore, our results in previous subsections –that, on purpose, ignore any weather and geographic information- are the likely outcomes that one could find in another application on electricity distribution networks, or indeed in other network utilities such as gas and water, where the regulator does not have access to W&G data.

However, as some environmental information is available in our application, we can carry out a (quasi) natural experimental exercise to examine the robustness of our empirical strategy based on spatial econometric techniques to utilise the effect of omitted variables on costs of neighbouring utilities. Our natural experiment exercise only attempts to compare the estimated spatial SFA models in previous subsections with a simple SFA model that now includes a set of W&G variables. This model (hereafter W&G SFA model), is used here as a benchmark as it is able to control for both economic and environmental cost drivers. As many of the W&G variables are spatially correlated (see Appendix A), we expect similar efficiency scores using a (non-spatial) model that includes W&G cost drivers and a spatial model that "replace" the W&G data (often not available) with data from surrounding firms using spatial econometric techniques.

The parameter estimates of the W&G SFA model are shown in Appendix B. In our W&G SFA model, we extend the previous set of cost drivers with several W&G variables. In particular, we include three weather variables (WIND, WINDEX, and DIS),[10] and three geographical variables (FOREST, AVESLOPE, and MAXSLOPE). This appendix also includes the results of an extended version of our previous spatial SFA3 model where we have now added W&G variables. This model (hereafter W&G spatial-SFA3 model), will allow us to examine whether omitted variables that are spatially correlated are still present.

Overall, our new results indicate that weather is an important factor in determining cost efficiency in this industry as the estimated coefficients for the weather variables are always significant. For instance, we find that a higher exposure to wind conditions implies larger costs for the distribution networks. On the other hand, the coefficient of the distance

[10] We use the geographic variable (DIS) in order to capture the effect of coastal climate on the networks. In Norway, this effect is related to problems with corrosion on network components normally caused by a combination of wind and salt water.

to the coast is negative as expected because inland weather conditions are likely to be less severe than coastal weather conditions. Our results also indicate that some geographic features of the terrain on which the networks are supported (i.e. forestry and maximum terrain slope), are also important determinants of cost efficiency. Finally, it is worth mentioning that all coefficients associated to $\hat{Z}_{it}$, are not statistically significant, except for $\hat{Z}_{it}$ alone whose coefficient is slightly larger than unity.

Figure 3 compares the individual efficiency scores that are obtained using the four models in Table 4 that do not include any environmental variable (see "dot" observations), with the scores that are obtained using the W&G SFA model in Appendix B (see "cross" observations), which serves as a benchmark model because it includes relevant environmental variables. This figure relates several interesting stories.

[Insert Figure 3 here]

First, most observations in Figure 3 are above the bisecting line, indicating that the efficiency scores of a simple SFA Model tend to be downward biased if either spatial effects or W&G variables are ignored. This result has been partially highlighted in the previous subsection. However, Figure 3 now shows that the bias is much larger when the efficiency scores are small. This implies that the most inefficient firms in a simple SFA specification of firms' cost would be wrongly penalized in an incentive regulated framework.

The second story has to do with the evolution of firms' efficiency scores when we move from simpler to more comprehensive models. Indeed, it is apparent in Figure 3 that we move closer to the benchmark efficiency scores when we add spatially generated variables as cost determinants. Moreover, the efficiency scores of the Spatial SFA3 model (the yellow dots) are quite close to the efficiency scores of the W&G SFA model (see the cross observations). This implies that we have been able to (almost) reproduce the same results as a SFA model that includes a set of relevant environmental variables that are not available in many cases. This result thus suggests that when W&G data are not available, this lack of information can likely be compensated by using data from surrounding firms using spatial econometric techniques.

Finally, in Figure 4 we compare the individual efficiency scores obtained using the non-spatial W&G SFA model and the W&G spatial-SFA3 model that extends our previous spatial SFA3 model by including W&G variables. We find that both efficiency scores are quite similar. This result indicates that, once we have controlled for W&G variables, the remainder of the spatially correlated omitted variables are of little importance, In other words, most of the omitted information that is spatially correlated has to do with environmental conditions. In summary, we have shown that the spatial econometric techniques can offer an effective and efficient possibility to control for this issue without recurring to the collection of costly weather and environmental data.

[Insert Figure 4 here]

6. Conclusions

This study combines stochastic frontier and spatial econometric techniques to evaluate a firm's efficiency in the Norwegian electricity distribution sector, taking into account spatially correlated omitted variables. In doing so, first we propose estimating a spatial econometric model to obtain a proxy for this type of variable by means of the available data for neighbouring utilities. Next we plug the variable generated into a standard SFA model.

We illustrate our approach using panel data for the Norwegian distribution utilities for the years 2004 to 2011. In order to implement our empirical strategy, we have matched the information on concession areas of distribution utilities with the data provided by the Norwegian regulator on firms' costs. We are not aware of other studies that have carried out a similar spatial matching exercise.

We find that the coefficient of the spatial correlation is significant in our auxiliary regression, indicating that the unobserved cost drivers are correlated. This thereby justifies the use of neighbouring firm data in order to control for unobserved cost drivers in our application. Next, the estimated stochastic cost frontier that includes our generated variable outperforms the model that excludes the omitted cost drivers. In this sense, as expected, the firm efficiency scores are larger when we include our proxy for the omitted variables, especially for firms that are more inefficient. In an incentive regulation framework, the upshot is that the latter types of firms are likely to be more severely penalized when the effect of this variable is not taken into account.

One advantage of the present application is that the Norwegian energy regulator has collected data on a set of W&G variables. This information is often not readily available in many other countries. As some environmental information is available in our application, we have been able to carry out a quasi-natural experimental exercise to examine the robustness of our empirical strategy. We have found that our spatial SFA model is able to roughly reproduce the efficiency scores of a more comprehensive model that includes the W&G variables that are not available in many applications. That is, we find that this lack of information can likely be compensated with data from surrounding firms using spatial econometric techniques. Finally, we have detected that most of the omitted information that is spatially correlated has to do with environmental conditions.

As a final point, we have found that combining efficiency analysis and spatial econometrics methods always improve the goodness-of-fit of the estimated models and, hence, more accurate and fair efficiency scores are obtained. Our approach is particularly useful in utilities sectors with a large number of distribution service areas, and where collecting and updating environmental data requires substantial amount of investment in human or financial resources as well as time.

References

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Dordrecht, Kluwer.

Arbia, G. (2014). *A primer for spatial econometrics: with applications in R*. Palgrave texts in Econometrics. Palgrave Macmillan.

Baltagi, B. H., and Liu, L. (2011), "Instrumental variable estimation of a spatial autoregressive panel model with random effects", *Economics Letters*, 111(2), 135–137.

Elhorst, J.P. (2010), "Applied Spatial Econometrics: raising the bar". *Spatial Economic Analysis*, 5, 9-28.

Elhorst, J.P. (2014), "Spatial Econometrics: from cross-sectional data to spatial panels", *Springer briefs in Regional Science*. Springer.

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

Glass, A., Kenjegalieva, K., and Sickles, R. (2016). "Spatial autoregressive and spatial Durbin stochastic frontier models for panel data." *Journal of Econometrics/Elsevier* 190, 289-300.

Greene, W. (2005). "Reconsidering heterogeneity in panel data estimators of the stochastic frontier model", *Journal of Econometrics*, 126, 269-303.

Greene, W., Orea, L., and Wall, A. (2011). *A one-stage random effect counterpart of the fixed-effect vector decomposition model with an application to UK electricity distribution utilities*, Efficiency Series Papers 01/2011, Oviedo Efficiency Group, University of Oviedo.

Growitsch, C., Jamasb, T., and Wetzel, H. (2012). "Efficiency effects of observed and unobserved heterogeneity: Evidence from Norwegian electricity distribution networks", *Energy Economics* 34, 542-548.

Haney, A.B. and Pollitt, M. (2013). "International benchmarking of electricity transmission by regulators: A contrast between theory and practice?", *Energy Policy* 62, 267-281.

Jondrow, J., Lovell, K., Materov, I., and Schmidt, P. (1982). "On the estimation of technical inefficiency in the stochastic frontier production function model.' *Journal of Econometrics* 19, 233-238.

Kelejian, H.H., and Prucha, I. R. (1998), “A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances”, *The Journal of Real Estate Finance and Economics*, 17(1), 99–121.

Kelejian, H.H., and Prucha, I. R. (2010), “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”, *Journal of Econometrics*, 157(1), 53–67.

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

Llorca, M., Orea, L., and Pollitt, M. (2014). "Using the latent class approach to cluster firms in benchmarking: An application to the US electricity transmission industry", *Operations Research Perspectives* 1(1), 6-17.

Miguéis, V.L., Camanho, A.S., Bjørndal, E., and Bjørndal, M. (2012). “Productivity change and innovation in Norwegian electricity distribution companies”, *Journal of the Operational Research Society* 63, 982-990.

Orea, L. and Jamasb, T. (2017). “Regulating heterogeneous utilities: A new latent class approach with application to the Norwegian electricity distribution networks”, *The Energy Journal* 38(4). forthcoming.

Orea, L., Jamasb, T., and Growitsch, C. (2015). “Using supervised environmental composites in production and efficiency analyses: An application to Norwegian electricity networks”, *Competition and Regulation in Network Industries* 16(3), 260-288.

Wang, H.-J. (2003). “A stochastic frontier analysis of financing constraints on investment: The case of financial liberalization in Taiwan*”, Journal of Business and Economic Statistics* 21, 406-419.

Wang, H.-J. and Ho, C.W. (2010), “Estimating fixed-effect panel stochastic frontier models by model transformation”, *Journal of Econometrics* 157(2), 286-296.

Table 1.

Descriptive statistics of the data

		Mean	*St.Dev.*	*Min*	*Max*
SCOST	1000 NOK	92899.69	192397.59	793.35	1797173.24
CUS	Number	21118.54	56320.29	14	552342
DE	1000 MWh	570990.65	1570418.93	3979	17000000
NL	Km	752.47	1290.72	9	8648
OH	%	0.66	0.20	0.00	0.97
PK	%	0.06	0.01	0.05	0.1
PL	Index	163.86	16.89	139	189.5
PE	NOK/MWh	331.01	73.93	234.6	436.3
ST	Number	948.06	1828.73	8	13525
WIND	m/s	25.5	2.48	22	31
WINDEX	m/s	5.28	1.02	2.71	8.13
DIS	km	53455.32	54567.00	191	19637
FOREST	Index	0	2.45	-3.21	22.51
AveSLOPE	%	10.13	3.74	2.86	22.22
MaxSLOPE	%	51.09	11.91	19	75

Table 2.

First-stage GMM parameter estimates.

	Coefficient	Robust-t
Intercept	10.5699	378.74
Spatial lag of the dependent variable (W·lnC)	0.1660	5.69
Cost drivers:		
Output variables		Yes
Input prices		Yes
Overhead variable		Yes
Hansen Chi-squared test (df)		0.1332 (1)
Weak instruments F-test (df in parenthesis)		47.21 (24,1007)
R-squared	0.9870	

Notes:
(a) For more details about the cost drivers and the functional form of the cost function, see Table 4.
(b) Instruments= all exogenous explanatory variables plus the spatial lag of lnCUS and $lnCUS^2$.

Table 3. Between and within standard deviations of the main cost drivers.

Variable	*Between*	*Within*	*B/W ratio*
$\boldsymbol{\hat{Z}_i}$	0.065	0.045	1.45
lnCUS	1.454	0.160	9.09
lnNL	1.162	0.035	32.76
lnDE	1.404	0.100	13.99
OH	0.201	0.026	7.83

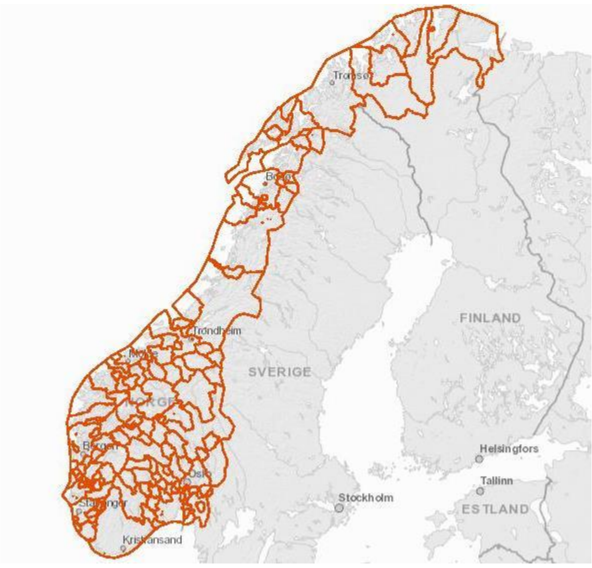
Table 4. Second stage parameter estimates. Cost frontier function.

	Single SFA		**Spatial SFA 1**		**Spatial SFA 2**		**Spatial SFA 3**	
Parameters	Estimates	t-ratio	Estimates	t-ratio	Estimates	t-ratio	Estimates	t-ratio
Intercept	10.511	677.7	10.518	665.1	10.518	636.1	10.517	625.0
lnCUS	0.291	10.81	0.273	10.72	0.276	10.74	0.271	10.31
lnNL	0.549	25.53	0.564	28.06	0.560	27.53	0.561	27.54
lnDE	0.142	6.01	0.146	6.45	0.147	6.51	0.148	6.53
OH	-0.312	-4.83	-0.298	-4.91	-0.294	-4.80	-0.285	-4.69
$0.5 \cdot \ln CUS^2$	0.130	6.46	0.124	5.85	0.120	5.75	0.121	5.66
$0.5 \cdot \ln NL^2$	-0.007	-0.08	-0.041	-0.50	-0.036	-0.45	-0.049	-0.61
$0.5 \cdot \ln DE^2$	0.196	4.99	0.202	5.24	0.199	5.26	0.204	5.36
$0.5 \cdot OH^2$	0.227	0.40	0.349	0.64	0.445	0.81	0.465	0.86
lnCUS·lnNL	-0.007	-0.18	0.003	0.07	0.000	0.00	0.006	0.16
lnCUS·lnDE	-0.109	-4.42	-0.114	-4.48	-0.110	-4.36	-0.113	-4.51
LnCUS·OH	-0.127	-1.07	-0.145	-1.20	-0.113	-0.94	-0.136	-1.10
lnNL·lnDE	-0.056	-1.22	-0.045	-1.00	-0.046	-1.03	-0.045	-1.02
LnNL·OH	-0.390	-1.87	-0.370	-1.82	-0.395	-1.96	-0.391	-1.94
LnDE·OH	0.483	3.34	0.492	3.37	0.483	3.34	0.505	3.46
lnPK	0.277	14.19	0.263	13.99	0.264	13.93	0.263	13.81
lnPL	0.662	16.89	0.664	17.98	0.663	17.88	0.661	17.78
Z			0.894	11.19	1.034	11.40	1.007	11.18
Z·N					-0.100	-2.74	-0.151	-3.83
Z·WlnNL							-0.302	-2.22
Z·WOH							0.111	0.59
Z·WlnST							0.230	1.82
$\ln\sigma_v$	-2.136	-51.02	-2.182	-51.69	-2.184	-49.28	-2.181	-48.56
$\ln\sigma_u$	-2.376	-11.15	-2.447	-10.79	-2.436	-10.34	-2.462	-10.10
lnNL	-1.623	-3.55	-1.621	-3.56	-1.485	-3.28	-1.504	-3.19
OH	0.659	1.85	0.064	0.17	-0.004	-0.01	-0.167	-0.44
lnST	1.012	2.64	1.085	2.84	0.971	2.56	1.014	2.55
Mean log-likelihood	0.553		0.612		0.616		0.621	
Observations	1032		1032		1032		1032	
LF	570.514		631.716		635.365		640.414	

Table 5. Efficiency scores

	Mean	**Std. Dev.**	**Min**	**Max**
Single SFA	0.916	0.064	0.535	0.990
Spatial SFA 1	0.923	0.058	0.498	0.987
Spatial SFA 2	0.923	0.057	0.498	0.985
Spatial SFA 3	0.925	0.055	0.485	0.985

Figure 1: Norwegian electricity distribution service areas



Source: Norwegian Water Resources and Power Directorate (NVE)

Figure 2. Histograms and Kernel density plots of estimated environmental cost differences.

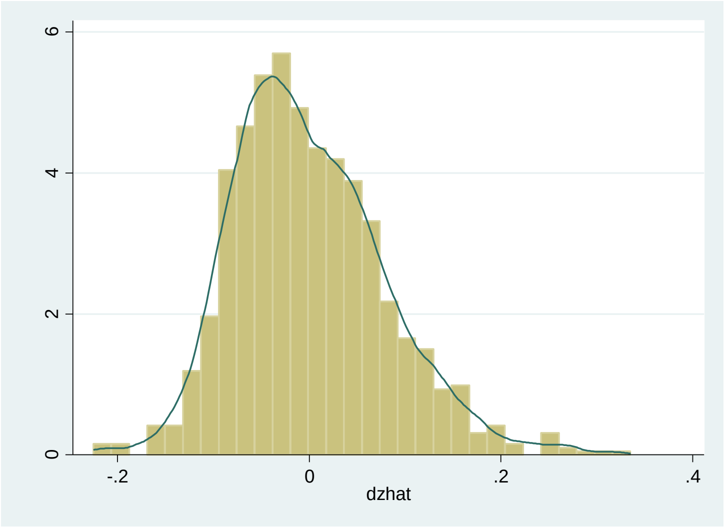
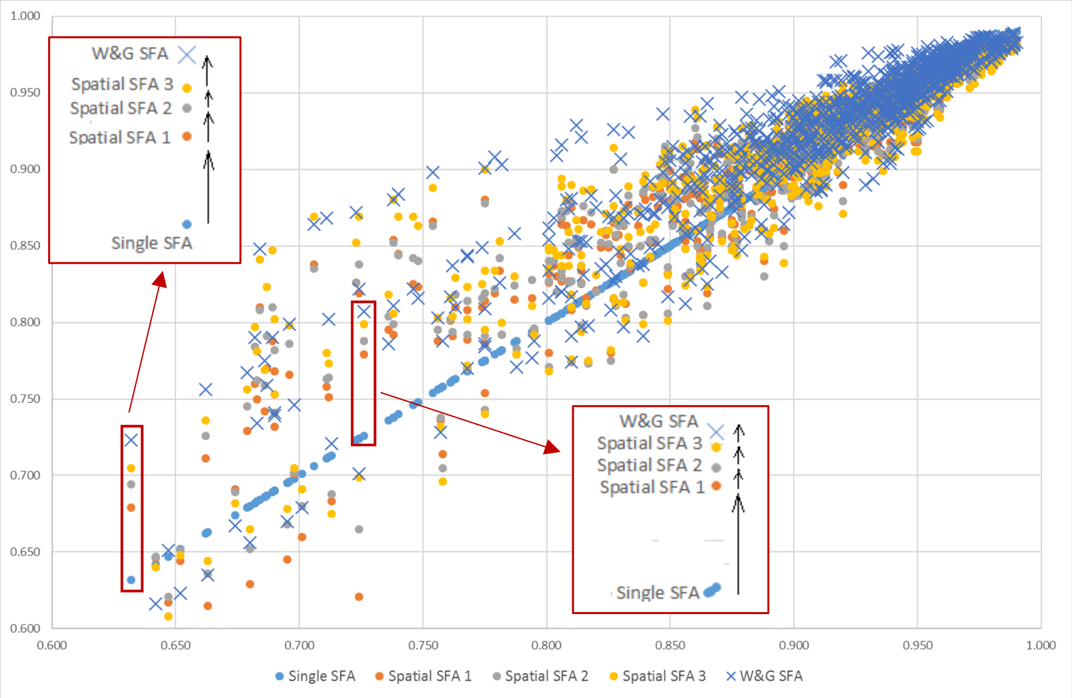
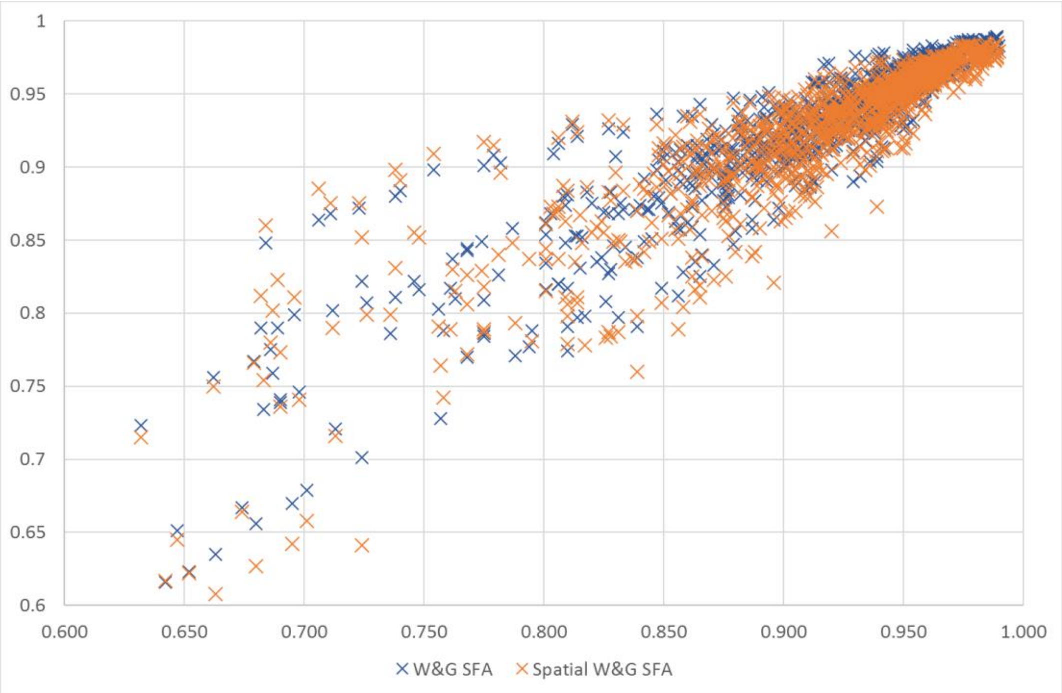


Figure 3. Efficiency scores.



Note: Efficiency scores of the Simple SFA model in the horizontal axis.

Figure 4. Efficiency scores using W&G data



Note: Efficiency scores of the Simple SFA model in the horizontal axis.

Appendix A

Spatial correlations of the main cost drivers. OLS auxiliary regressions.

Regression	Coef.	t-ratio
Customer numbers		
Intercept	-0.1356***	-2.75
Spatial lag	0.0816***	6.48
R^2	0.0392	
Network Length		
Intercept	-0.0571	-1.44
Spatial lag	0.0443***	3.41
R^2	0.0111	
Delivered Energy		
Intercept	-0.0861*	-1.89
Spatial lag	0.0677***	5.92
R^2	0.0329	
Overhead lines (%)		
Intercept	-0.0048	-0.83
Spatial lag	0.1482***	14.43
R^2	0.1683	
Wind		
Intercept	26.2482***	140.30
Spatial lag	-0.0061***	-4.37
R^2	0.0183	
Wind Exposure		
Intercept	5.4931***	70.78
Spatial lag	-0.0086***	-3.04
R^2	0.0089	
Distance to coast (in logs)		
Intercept	8.3434***	79.73
Spatial lag	0.0328***	17.60
R^2	0.2314	
Forest		
Intercept	-0.0367	-0.47
Spatial lag	0.0448***	3.26
R^2	0.0103	
AveSlope		
Intercept	6.7957***	33.00
Spatial lag	0.0657***	18.58
R^2	0.2512	
MaxSlope		
Intercept	37.9739***	50.63
Spatial lag	0.0505***	19.31
R^2	0.2659	

Appendix B

SFA models with W&G variables.

	W&G SFA		Spatial W&G SFA	
Parameters	Estimates	t-ratio	Estimates	t-ratio
Intercept	10.668	101.334	10.595	107.749
lnCUS	0.295	10.979	0.285	10.954
lnNL	0.523	22.291	0.539	24.433
lnDE	0.148	6.143	0.142	6.169
OH	-0.181	-2.923	-0.259	-4.483
$0.5 \cdot \ln CUS^2$	0.108	5.117	0.101	4.999
$0.5 \cdot \ln NL^2$	-0.108	-1.150	-0.154	-1.730
$0.5 \cdot \ln DE^2$	0.193	4.861	0.168	4.659
$0.5 \cdot OH^2$	0.822	1.388	0.548	0.979
lnCUS·lnNL	0.040	0.980	0.058	1.523
lnCUS·lnDE	-0.123	-4.591	-0.128	-4.902
LnCUS·OH	-0.142	-1.124	-0.194	-1.538
lnNL·lnDE	-0.028	-0.583	-0.002	-0.048
LnNL·OH	-0.365	-1.625	-0.244	-1.163
LnDE·OH	0.500	3.324	0.449	3.141
lnPK	0.273	14.427	0.268	14.470
lnPL	0.667	17.661	0.663	18.500
Z			1.195	10.512
Z·N			-0.080	-1.615
Z·WlnNL			0.050	0.298
Z·WOH			-0.153	-0.694
Z·WlnST			-0.051	-0.343
WIND	-0.015	-4.667	-0.014	-4.766
WINDEX	0.041	4.536	0.045	5.250
lnDIS	-0.017	-4.191	-0.016	-3.965
Forrest	0.008	2.745	0.008	2.866
AveSlope	0.002	0.854	0.001	0.224
MaxSlope	0.003	3.706	0.004	4.493
$\ln\sigma_v$	-2.147	-52.673	-2.235	-52.423
$\ln\sigma_u$	-2.600	-10.736	-2.497	-12.658
lnNL	-2.138	-3.770	-1.847	-3.967
OH	0.017	0.039	-0.114	-0.300
lnST	1.602	3.278	1.371	3.383
Mean log-likelihood	0.608		0.668	
Observations	1032		1032	
LF	627.159		689.086	