



**Copenhagen
Business School**
HANDELSHØJSKOLEN

Department of Economics

Copenhagen Business School

Working paper 7-2020

Managing Power Supply Interruptions: A
Bottom-Up Spatial (Frontier) Model with an
Application to a Spanish Electricity
Network

Pablo Argüelles
Luis Orea



CBS

DEPARTMENT OF
ECONOMICS

COPENHAGEN
BUSINESS SCHOOL

CSEI Working Paper 2020-07

CBS Department of Economics 7-2020

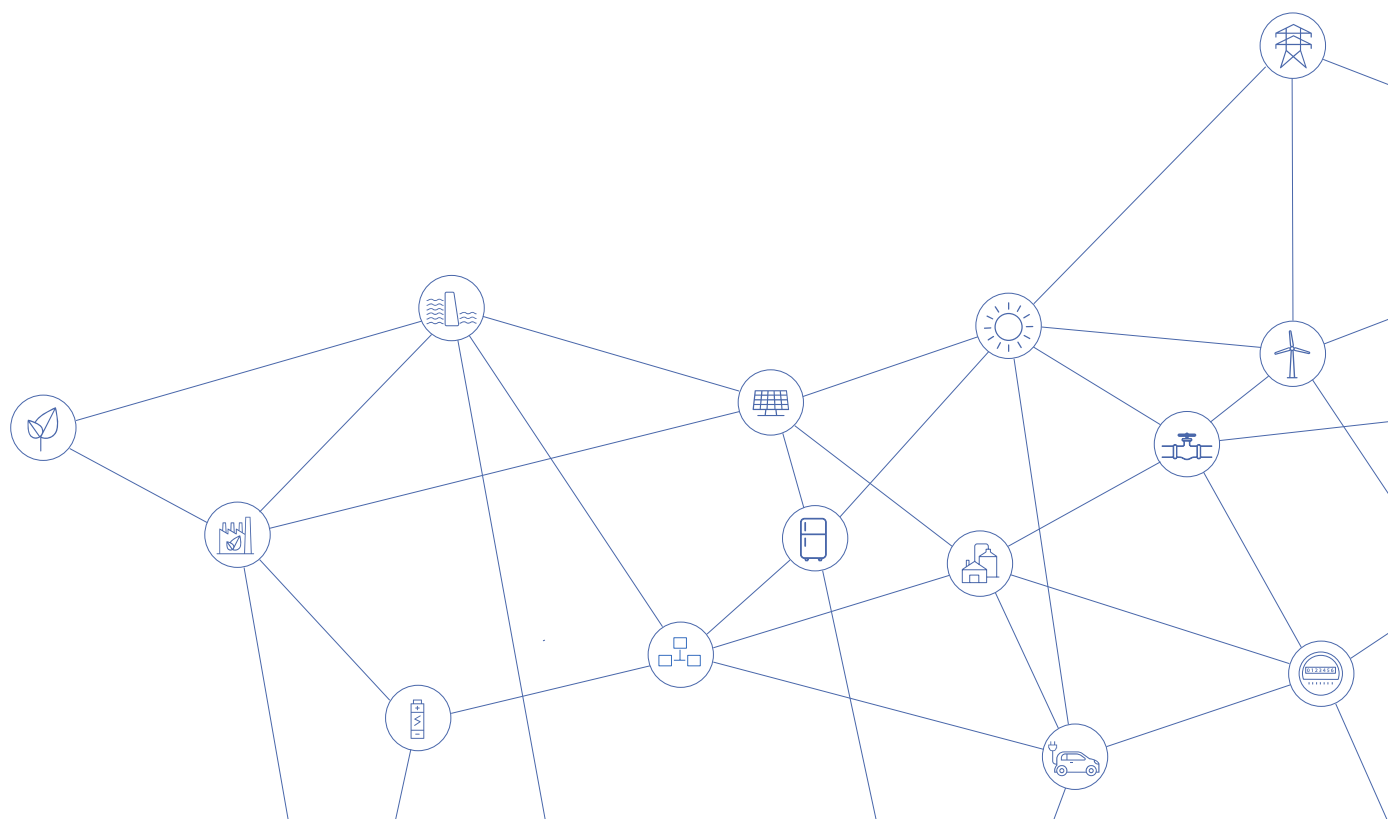
WORKING PAPER

Copenhagen School of Energy Infrastructure | CSEI

Managing Power Supply Interruptions: A Bottom-Up Spatial (Frontier) Model with an Application to a Spanish Electricity Network

Pablo Argüelles

Luis Orea



Managing power supply interruptions: A bottom-up spatial (frontier) model with an application to a Spanish electricity network

Pablo Argüelles
EDP España and University of Oviedo

Luis Orea ^a
University of Oviedo and Oviedo Efficiency Group

Abstract

In December 2013 a new electricity law was approved in Spain as part of an electricity market reform including a new remuneration scheme for distribution companies. This remuneration scheme was updated in December 2019 and the new regulatory framework introduced a series of relevant modifications that aim to encourage the regulated firms to reduce their power supply interruptions using a benchmarking approach. While some managerial decisions can prevent electricity power supply interruptions, other managerial decisions are more oriented to mitigate the consequences of these interruptions. This paper examines the second type of decisions using a unique dataset on the power supply interruptions of a Spanish distribution company network between 2013 and 2019. We focus on the effect of grid automatization on the restoration times, the relative efficiency of the maintenance staff, and the importance of its location. We combine a bottom-up spatial model and a stochastic frontier model to examine respectively external and internal power supply interruptions at municipal level. This model resembles the conventional spatial autoregressive models but differs from them in several important aspects.

Keywords: electricity distribution, power supply interruptions, spatial econometrics, frontier models.

^a Corresponding author. Email: lore@uniovi.es. Address: School of Business and Economics, Department of Economics, Avda. Del Cristo, s/n. 33006, Oviedo, Spain.

Managing power supply interruptions: A bottom-up spatial (frontier) model with an application to a Spanish electricity network

Pablo Argüelles
EDP España and University of Oviedo

Luis Orea
University of Oviedo and Oviedo Efficiency Group

1. Introduction

The National Commission on Markets and Competition (CNMC), the Spanish regulator of the electricity sector approved last year the Circular 6/2019, which was published in the Official State Gazette on 5th December 2019, establishing a new methodology for calculating the remuneration of the electricity distribution companies for the years to come. The new regulatory framework introduced a series of relevant modifications to current regulation that aim to encourage the regulated firms to improve their performance, reducing not only distribution costs but also power supply interruptions. The previous regulatory framework (see Royal Decree 1048/2013), based on intertemporal comparisons, was not able to give the market the appropriate signal to achieve these targets. The new incentive scheme resembles a benchmarking (or yardstick) regulation as the incentives for quality of service improvements rely on the performance of other firms. In particular, the incentive reward/penalty will be computed by comparing the performance (variation) of a standard reliability indicator of each electricity distribution company with average sector performance in comparable urban and rural areas of supply. It is expected that the proposed approach will be more rewarding for the regulated firms if they improve the quality of service in the distribution network related to the interruption time of the electricity supply to their clients. Therefore, they will likely pay more attention to this important topic of the electricity distribution activity.

Although a large percentage of Power Supply Interruptions (hereafter PSI) are beyond the management control of utilities (e.g. due to weather conditions or external human manipulations), managerial decisions such as vegetation management, asset investments and maintenance strategies can contribute to improvements of quality of service. While some of these decisions are more oriented to *prevent* electricity PSI caused by severe environmental conditions (e.g. reducing the number and length of overhead lines or improving grid assets capabilities, etc.), other managerial decisions are more oriented to *mitigate* the consequences of these interruptions (e.g. increasing the number of maintenance crews, improving their location, etc.). While some of these decisions require additional capital costs, other decisions require increasing firm's operation and maintenance costs. It is not clear in the literature which strategy is better. As pointed out by [Giannakis et al. \(2005\)](#) and [Jamasp et al \(2012\)](#), the electricity distribution firms might adopt different strategies to combine capital and operating costs to improve their quality of service.

Other literature has shown that weather conditions influence quality of service in electricity distribution networks. While [Coelho et al. \(2003\)](#), [Domijan et al. \(2003\)](#) and [Zhou et al.\(2006\)](#) find a significant correlation between PSI and rain, wind and temperatures, [Yu et al. \(2009\)](#) find that such factors often do not have a significant economic and statistical effect on the overall performance of the UK utilities. [Wang and Billington \(2002\)](#) show that severe weather conditions do not only increase the frequency of PSI, but also the restoration time.

Other researchers have examined the effect of the network characteristics on quality of service. For instance, [Kjølle et al. \(2003\)](#) have found that the number and duration of interruptions are significantly higher in overhead networks compared to cable networks.

From a remedy perspective, the emergence of grid digitalization, which allows faster and more accurate detection of damaged equipment and its location, helps to restore electricity supply and fix faults more quickly. For a matter of time, previous literature has not studied the impact of such technological development on the duration of interruptions. This issue can be now studied as the electricity distribution firms in many countries have been installing remote control on transformers and switching and protection circuits in the last decade. On the other hand, faster restoration in most outages requires efficient and well-located maintenance crews. As the previous literature uses firm level data or aggregate geographical data, they do not allow examining the effect of the location of the maintenance crews on restoration times, as well as the relative efficiency of each maintenance crew.

This paper uses a unique data set on the PSI in 91 Spanish municipalities of a Spanish distribution company network between 2013 and 2019 to identify the main technological and managerial drivers of the duration of power outages. As our dataset is very detailed, we know the *source* municipality of each PSI as well as, if any, the subsequent municipalities affected by the same PSI. In this sense, we are able to distinguish between “internal” PSI where the origin is located in the *source* municipality and “external” PSI caused by outages located in other municipalities that are physically connected through the distribution network. Distinguishing between the two types of supply interruptions is not a semantic issue because while the quick restoration of supply in an “internal” PSI requires one of the maintenance crews to be close and/or the existence of remote controlled switching and protection systems located in the *source* municipality, the restoration time in an “external” PSI has nothing to do with its own factors but with the location and network characteristics of the *source* municipality firstly affected by the outage. For this reason, we use two different approaches to examine “external” and “internal” PSI.

The duration of the external PSI is modelled using a spatial non-frontier econometric model. A spatial non-frontier model is used here because the external PSI only appear if there are PSI in neighboring and connected municipalities. As we know the sequence of the individual PSI across municipalities, we can develop a spatial model from scratch, i.e. using the engineering or physical information of the PSI that occurred in each municipality. Therefore, our model can be viewed as a *Bottom-Up* (BoU) spatial model. A frontier specification is not used here because the managerial decisions aiming to restore the electricity supply in these municipalities fix network faults located in other municipalities (i.e. in the municipality firstly affected by the outages). For this reason, inefficient performance is only examined using internal PSI. The duration of the internal PSI is modelled using a standard non-spatial frontier model. It does not make sense to use here a spatial specification because the faults that have triggered these PSI took place within the municipality.

The paper is structured as follows. [Section 2](#) develops several spatial and frontier models that aim to identify the main technological and managerial drivers of the restoration times of internal and external PSI. The bottom-up nature of our models is not only apparent in that we can distinguish between internal and external PSI, but also in that we aggregate our engineering-based (outage) data on monthly and municipality basis in order to mimic the traditional spatial econometric models. [Section 3](#) discusses the data used in the empirical analysis and its sources. [Section 4](#) provides the parameter estimates and discuss the main results. Finally, [Section 5](#) presents the conclusions.

2. Methodology

2.1. Preliminaries

Let us first explain briefly how we use the individual information of each outage to define our dependent variable(s) and, in particular, to compute an engineering-based spatially lagged dependent variable, which differs from the one often used in the standard spatial econometric models.

As customary in the Spanish electricity distribution system, we use the so-called TIEPI as a reliability indicator to measure (lack of) quality of service. This indicator, which is defined further in [Section 3](#), is a power-adjusted measure of the restoration times as it considers the active power loss due to the outage, relative to the whole system power. As we are interested in the duration and not the frequency of PSI, our final sample does not include months without outages.¹ The duration of *all* power outages occurred in a municipality $m = 1, \dots, M$ during month $t = 1, \dots, T$ will be expressed as Y_{mt} .

In a standard Spatial Autoregressive Model (SAR), it is assumed that the dependent variable Y_{mt} is spatially correlated, and this correlation is modelled as follows:

$$Y_{mt} = \lambda WY_{mt} + \varepsilon_{mt} \quad (1)$$

where ε_{mt} is an error term, $WY_{mt} = W_m Y_t = \sum_{n=1}^M W_{mn} Y_{nt}$ stands for the endogenous spatial lag of the dependent variable, $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{Mt})$ is an $M \times 1$ vector of the dependent variables, and $W_m = (W_{m1}, W_{m2}, \dots, W_{mM})$ is a spatial weight vector where the weights (W_{mn}) equal one for adjacent units and zero for non-bordering units. Finally, the λ parameter is the spatial autoregressive coefficient that measures the degree of spatial correlation between units.

Notice that WY_{mt} can be viewed as a weighted measure of the duration of *all* PSI that occurred in adjacent municipalities to the municipality m in period t , even if the outages in neighboring municipalities have nothing to do with the PSI in municipality m . In some cases, WY_{mt} include common PSI. However, as this spatially lagged dependent variable ignores the *true* sequence of the PSI across municipalities, it surely includes the duration of PSI of subsequent municipalities affected by the same outage. If so, equation (1) would wrongly suggest that the outage in a preceding municipality is caused by the outage of a subsequent municipality. In other words, the traditional spatially lagged dependent variable will provide biased results because it ignores the true physical propagation (i.e. true causality) of PSI in a real electricity distribution network.

Like in a dyadic-type data setting where it is clearly possible to distinguish an origin unit from a destination unit (see e.g. [Neumayer and Plümper, 2010](#)), we know the source municipality of each PSI as well as, if any, the subsequent municipalities affected by the same outage. Therefore, our dataset allows us to model properly contagion, diffusion or spillover effects across municipalities that are physically connected through the distribution network. To achieve this objective, let us decompose the total duration of the PSI (Y_{mt}) into two sets: internal and external PSI. That is:

$$Y_{mt} = YI_{mt} + YE_{mt} = \sum_{i=1}^{I_{mt}} i_{imt} + \sum_{j=1}^{J_{mt}} e_{jmt} \quad (2)$$

¹ Analysing the frequency of PSI is also more challenging as we should cope with an excessive number of zero values in our data. The development and estimation of zero-inflated econometric models in non-frontier settings have become widespread. See [Yang et al. \(2017\)](#) for a comparison of methods. Our models, however, allow examining whether the number of outages matters when estimating the coefficients in which we are interested.

where $i = 1, \dots, I_{mt}$ stands for outages started in municipality m in period t , and $j = 1, \dots, J_{mt}$ stands for PSI in municipality m in period t caused by outages that have started *before* in other municipalities physically connected through the distribution network.² i_{imt} and e_{jmt} are respectively the duration of the *internal* and *external* PSI from the perspective of the municipality m . Therefore, while YI_{mt} measures the duration of all internal PSI occurred in municipality m in period t , YE_{mt} measures the duration of all external PSI that affect municipality m in period t but that have been ‘imported’ from other municipalities. In other words, our dataset allows us to know the *true* outcome (i.e. YE_{mt}) of the spatial spillovers we are able to capture using a spatial econometric model. In a standard spatial application, only the aggregate effect of both *internal* and *external* PSI (i.e. Y_{mt}) is observed by the econometrician. Thus, we have a kind of quasi-natural experiment to test what spatial spillovers are being captured by the most common spatial econometric models.

On the other hand, as we have perfect information about the characteristics of the electricity distribution network and all connections across municipalities, we know the *true* W matrix. This matrix should be computed in our application considering the number of connections between two municipalities and the capacity of these connections. Otherwise, our spatial specification of a physical phenomenon such as the propagation of PSI would not make sense. Moreover, unlike most spatial models in regional economics, contiguity in our case is a necessary but not sufficient condition to be affected by other municipality. It also requires being connected with the preceding municipalities.³ Another attractive feature of our engineering-based W matrix is that its elements can be treated as exogenous spatial weights when estimating the model due to the physical nature of the network connections. This is very useful from an econometric view because it avoids the need to address challenging endogeneity issues associated with W .

2.2. Modelling external PSI

As aforementioned, explaining YI_{mt} and YE_{mt} requires a different model because while YI_{mt} emanates from municipality m and the network equipment that should be fixed is located in this municipality, YE_{mt} emanates from other municipalities and thus there is nothing to fix in municipality m . In other words, YE_{mt} is a pure spatial spillover or contagion effect that disappears if municipality m in period t is not involved in multi-municipality PSI that had previously started in another municipality. For this reason, we propose using the following SAR specification to explain the duration of each external interruption:

$$e_{jmt} = \gamma_m \Omega_{jm} Y_j = \gamma_m \sum_{n=1}^{N_j} \Omega_{jmn} Y_{jn} \quad (3)$$

where N_j is the number of municipalities involved in outage j , $Y_j = (Y_{j1}, Y_{j2}, \dots, Y_{jN_j})$ is the $N_j \times 1$ vector of PSI durations involved in outage j , and $\Omega_{jm} = (\Omega_{jm1}, \Omega_{jm2}, \dots, \Omega_{jmN_j})$ is a $N_j \times 1$ spatial weight vector where the weights equal one for *first-order* preceding municipalities that are connected through the electricity distribution network and have been affected *immediately* before than municipality m in outage j . Otherwise, the elements in Ω_{jm} are equal to zero. Finally, the γ_m parameter is the spatial autoregressive coefficient that measures the contagion

² The time lag between outages cannot be perceived by humans as it usually lasts milliseconds.

³ In this sense, our model looks like a multilevel or hierarchical SAR model, which is becoming increasingly popular in social sciences. See [Corrado and Fingleton \(2016\)](#) for a summary of these models. It is assumed in these models that there exist a number of well-defined groups organized within a hierarchical structure, such as class within schools. Much of the multilevel literature assumes that inter-individual interaction is restricted to within group boundaries. From a spatial perspective, this implies that the inter-individual interactions are restricted spatially in a similar fashion that our bottom-up spatial model.

degree of municipality m from preceding municipalities. We expect that this degree depends on the number of connections of municipality m with other municipalities, as well as the capacity of these connections.

If we next plug (3) into $YE_{mt} = \sum_{j=1}^{J_{mt}} e_{jmt}$, and add the traditional noise term, we get the following bottom-up SAR model:⁴

$$YE_{mt} = \gamma_m \Omega Y_{mt} + \omega_{mt} \quad (4)$$

where $\Omega Y_{mt} = \sum_{j=1}^{J_{mt}} \sum_{n=1}^{N_j} \Omega_{jmn} Y_{jn}$, and ω_{mt} is a symmetric and normally distributed noise term. One might be tempted to extend this model with municipality-specific factors, i.e. adding the traditional βX_{mt} term in a conventional SAR model. However, this might yield the unsound result that municipality m has been affected by outages emanated in other municipalities even when municipality m was not involved in common outages, i.e. when either $\Omega_{jmn} = 0$ or $Y_{jn} = 0$.

Equation (4) can be estimated using a simple Ordinary Least Squares (OLS) estimator⁵ given that ΩY_{mt} is by definition an exogenous variable because it has been computed using past values although our notation in equation (4) does not explicitly indicate this recursive nature of ΩY_{mt} . Therefore, equation (4) can be viewed as a *time-space recursive* model where the dependent variable is lagged in both space and time dimensions (see e.g. [Elhorst, 2010](#)).⁶

Instead of using variables in levels, it is also possible to estimate a SAR model in *per outage* terms. That is, we can replace respectively YE_{mt} and ΩY_{mt} with YE_{mt}/J_{mt} and $\Omega Y_{mt}/J_{mt}$ in (4). Compare to a model in levels, the per outage specification reduces heteroskedasticity issues and improves goodness-of-fit.⁷ As it is customary in both spatial and frontier literatures, the SAR model can also be estimated using variables in natural logarithms. The logarithm specification allows us to prevent the existence of extreme values for the heteroskedastic autoregressive parameter in our spatial models. In addition, using logs allows us to reduce convergence issues when estimating our frontier models.

The γ_m parameter that measures the contagion degree of municipality m from preceding municipalities likely depends on the number of connections of municipality m with other municipalities, as well as the average capacity of these connections. Indeed, we expect larger spatial spillovers when the number and capacity of the connections with *preceding*

⁴ Notice that equation (4) cannot be estimated in a conventional spatial econometric application because only, say, the whole GDP in one region is observed, and not the portion of such GDP that actually depends on the GDP of neighbouring regions.

⁵ Equation (4) can also be estimated using Maximum Likelihood (ML) techniques if we assume that ω_{mt} follows a normal distribution, i.e. $\omega_{mt} \sim N(0, \sigma_\omega = e^{\tau_0})$.

⁶ In this sense, [Skevas \(2019\)](#) shows that the endogeneity issues that exist in conventional SAR models such as $Y_t = \gamma \Omega Y_t + \omega_t$ are caused by the fact that the dependent variable of any individual appears both on left and the right-hand side of the spatial autoregressive model, after replacing Y_t on the right-hand side of the above equation with $Y_t = \gamma \Omega Y_t + \omega_t$. A time-space recursive model such as $Y_t = \gamma \Omega Y_{t-1} + \omega_t$ overcomes such a bias because for a particular observation, while Y_t appears on the left-hand-side, the right-hand-side contains Y_{t-2} after replacing Y_{t-1} with $Y_{t-1} = \gamma \Omega Y_{t-2} + \omega_{t-1}$, and the endogeneity issue is wiped out.

⁷ Notice that the total duration of external PSI in (4) can be decomposed into the number of external PSI that affect municipality m in period t (J_{mt}) and the average duration of these interruptions ($\bar{e}_{mt} = \sum_{j=1}^{J_{mt}} e_{jmt} / J_{mt}$). That is, $YE_{mt} = J_{mt} \cdot \bar{e}_{mt}$. This decomposition suggests that using a per outage specification of (4) allows us to focus on outages' duration (i.e. on \bar{e}_{mt}) and not on the number of such outages (i.e. on J_{mt}).

municipalities (i.e. involved in common outages) increases. However, the abovementioned spatial spillovers are likely attenuated if municipality m has other connections.

As it does not make sense to obtain negative PSI from an engineering point of view, we will use non-linear specifications for the SAR model in (4). In order to prevent negative γ_m values, we will parameterize γ_m using an exponential function, that is $\gamma_m = \exp(\lambda q_m)$ where q_m measure municipality m 's connectivity.

2.3. Modelling internal PSI

The duration of the *internal* PSI is modelled as a Standard non-spatial Frontier (SF) model because the managerial decisions aiming to restore power supply in any outage should focus on fixing network faults located in the source municipality, i.e. in the municipality first affected by the outage. Accordingly, we propose modelling YI_{mt} as follows:

$$YI_{mt} = \beta X_{mt} + v_{mt} - u_{mt}(z_{mt}) \quad (5)$$

where X_{mt} is a vector of technological and environmental variables that might determine the length of the PSI. For instance, it might include the length of the network, and the relative importance of aerial and underground lines, as in [Kjølle et al. \(2003\)](#). Another network characteristic that might affect quality of service are the degree of digitalization of the network and the proportion of outdoor transformers. Restoration times might also depend on the distance to the nearest maintenance crew, in particular if the outage is located in the countryside. [Wang and Billington \(2002\)](#) show that severe weather conditions also increase restoration times. For this reason, other good candidates to be included in X_{mt} are the weather conditions. In addition to including seasonal dummy variables we also include the percentage of outages caused by weather-related issues.

Equation (6) also includes two error terms, v_{mt} and u_{mt} . While the former term is a symmetric error term measuring pure random shocks, the latter term is a non-negative error term either measuring managerial inefficiencies associated to restoration times, or extremely large restoration times that cannot be attributed to random shocks. z_{mt} is a vector of exogenous variables that might determine the relative performance of the maintenance crews restoring power supply in municipality m or the frequency of difficult-to-restore PSI. As we do not have information on the maintenance staff characteristics and we have found convergence problems when we included crew-specific dummy variables, we only include seasonal variables aiming to measure within-year variations in u_{mt} and a set of time trends in order to capture changes in u_{mt} over time.

As for the external interruptions, the above frontier model can be estimated in levels, in per outage terms and in logs. Again a per outage specification of (5) allows us to focus on outages' duration (i.e. on $\bar{t}_{mt} = \sum_{j=1}^{I_{mt}} i_{jmt} / I_{mt}$) and not on the number of internal outages (i.e. on I_{mt}). The use of natural logarithms has permitted to get parameter estimates when estimating the model using ML techniques.

The above frontier model can be estimated by ML once we have assumed specific distributions for the noise and inefficiency terms. Hereafter we will assume that $v_{mt} \sim N(0, \sigma_v = e^{\delta_0})$ and that the variable representing inefficiency is the truncation (at zero) of a normally-distributed random variable with mean zero and standard deviation $\sigma_{u_{mt}} = e^{\theta_0 + \theta z_{mt}}$. As is customary in the SFA literature, the error term in any of our spatial stochastic frontier models includes a noise term (v_{mt}) and an inefficiency term (u_{mt}). [Jondrow et al. \(1982\)](#) use the conditional distribution of u_{mt} given the composed error term (i.e. $v_{mt} - u_{mt}$),

to estimate the asymmetric random term u_{mt} . We use the conditional expectation $E(u_{mt}|v_{mt} - u_{mt})$ as a point estimate of u_{mt} .

It is convenient to point out that model (5) is estimated using a different set of observations than model (4). Indeed, while the SAR model in (4) is estimated using the monthly observations with non-zero external PSI (i.e. when $YE_{mt} > 0$), the estimation of the frontier model in (5) requires instead using the monthly observations with non-zero internal PSI (i.e. when $YI_{mt} > 0$). As most municipal interruptions emanate from themselves in our application, the number of observations to estimate (5) is much larger than that to estimate (4). Notice in this sense that the total number of observations in our application (hereafter N) can be split into three non-overlapped subsets: i) observations with positive internal PSI but with no external PSI (hereafter N_I); ii) observations with positive external PSI but with no internal PSI (hereafter N_E); and iii) observations with both positive external and internal PSI (hereafter N_{I+E}). Therefore, while model (4) is estimated using $N_E^* = N_E + N_{I+E}$ observations, model (5) is estimated using $N_I^* = N_I + N_{I+E}$ observations. Typically, $N_I > N_E$. Thus, N_I^* is much larger than N_E^* in our application.

To conclude this section, it is germane to point out that both models, although estimated separately, can be viewed as a unique but more comprehensive BoU spatial frontier models. Indeed, notice that if we model the sum of internal and external PSI and combine appropriately equations (4) and (5), we get the following spatial frontier model:

$$Y_{mt} = \gamma_m \Omega Y_{mt}^* + \beta X_{mt}^* + v_{mt}^* - u_{mt}^* \quad (6)$$

where $\Omega Y_{mt}^* = \Omega Y_{mt} D_{E>0}$, $X_{mt}^* = X_{mt} D_{I>0}$, $v_{mt}^* = v_{mt} D_{I>0} + \omega_{mt} D_{E>0}$, and $u_{mt}^* = u_{mt}(z_{mt}) D_{I>0}$. $D_{E>0}$ and $D_{I>0}$ are two dummy variables identifying the observations with respectively non-zero external and internal interruptions. Notice that this model should be estimated using $N = N_E + N_I + N_{I+E}$ observations, i.e. with less observations than the sum of observations in (4) and (5). On the other hand, it is worth mentioning that the above spatial frontier model is similar to other spatial and frontier models proposed in the literature. For instance, the inefficiency term u_{mt}^* resembles the inefficiency term of the Zero-Inefficiency Stochastic Frontier (ZISF) model introduced by [Kumbhakar et al. \(2013\)](#) but with perfect information about the allocation of the observations between the so-called fully-efficient group (observations with $D_{I>0} = 0$) and inefficiency group (observations with $D_{I>0} = 1$). The above model also resembles the Generalized Spatial Autoregressive Frontier (GSARF) model introduced by [Gude et al. \(2018\)](#) that also allows the degree of spatial interaction to be observation specific. However, while the spatial lagged dependent variable in this paper is endogenous, ΩY_{mt}^* above is exogenous due to its (implicit) recursive nature.

3. Sample and data

We apply our SAR and SF models to unique data set on the PSI in 91 Spanish municipalities of a Spanish distribution company network between 2013 and 2019. The data used in this study is confidential as it was obtained directly from this private utility. The variables used in our empirical application have been constructed considering the geographical structure of the electricity distribution network of such company. This allows us to know the real propagation effects of the outages from one municipality to the next.

Our dependent variables are computed using the standard reliability indicator (TIEPI) used in Spain to measure (lack of) quality of service. The TIEPI reliability indicator is the equivalent PSI time found in medium voltage (1 kV to 36 kV) of the installed capacity affected by the outage. This index has the following expression:

$$TIEPI = \frac{\sum_{i=1}^k (PI_i * H_i)}{\sum PI} \quad (7)$$

where PI_i is the sum of the capacity of the transformers (kVA) installed in the substations located in the distribution grid and affected by the outage i , H_i is the supply interruption time affecting the capacity P_i , K is the total number of interruptions during the considered lapse time, and $\sum PI$ is the sum of the capacity of the transformers (kVA) installed in the substations located in the distribution grid. Equivalent measures in other countries are the well-known SAIDI and CAIDI indicators, which measure the duration of the interruptions in per consumer terms.

The data set that is being used in the study come from the Distributed Control System (DCS) of the network of the Spanish distribution company. The time resolution of the DCS allows us to know which is the first event that triggers the outage and, therefore, to know which is the “source” municipality where the outage takes place. The data set considers 168,705 events (distribution stations or step-down transformers affected during an outage) and 23,816 outages that took place between 2013 and 2019. The complete outage dataset obtained from the DCS gives information about all the grid components (distribution stations and step-down transformers) that are affected on each outage: the outage number of each outage, the *exact* outage and restored times, the substations or step-down transformers that are affected during the outage, the capacity of each one of the substations or step-down transformers that are affected during the outage, the so-called *event category* which gives information about the cause of the incident (weather related, not foreseen, scheduled outage or third-party grid origin), and name of the municipality where is placed each one of the grid components affected during the outage.

Other different data sources allow us to improve the data quality giving information about the urban vs. rural nature of the municipalities affected by the outages,⁸ the dates when a remote-control system was installed in a distribution station or step-down transformer, the neighboring municipalities (identified following both a geographical and electrical approach), the municipality cross border grid segments,⁹ the distribution stations and step-down transformers characteristics,¹⁰ the length of the network,¹¹ and finally the location of each Operational and Maintenance (O&M) crew.

The above detailed information has allowed us to compute the municipality-level variables finally used in our empirical application. For each municipality and month, we know the number and TIEPI of outages that affected it, the number and TIEPI of outages that were triggered by it, the number and TIEPI of outages that progress to neighboring municipalities, and the number and TIEPI of outages that come from neighboring municipalities. These

⁸ The regulator distinguishes between urban areas (municipalities with more than 20,000 supplies, including provincial capitals, even if they do not reach the previous figure), semi-urban areas (municipalities with a number of supplies between 2,000 and 20,000, excluding provincial capital), rural areas (municipalities with less than 2,000 supplies).

⁹ More than 3,000 grid segments that cross the border between two geographically adjacent municipalities have been identified. We identified the municipalities involved in each interconnection, the number of connections, capacities (kVA), sections (mm²), voltage (kV), location (overhead / underground) of all of them.

¹⁰ For each one of them, we have identified the name, identification code, municipality name where is installed, municipality identification code, UMTS (latitudinal and longitudinal location), rated capacity (kVA), and outdoor, indoor or underground location.

¹¹ We have identified more than 400 thousand grid segments. For all of them number (#), capacity (kVA), section (mm²), voltage (kV) and location (overhead / underground) of the interconnection have been identified.

variables allow us to compute the duration of internal and external PSI used as *dependent* variables in our BoU spatial and SF models.

We also know the number and name of geographical neighboring municipalities, the number and name of electrical neighboring municipalities, and the number, capacity and voltage level of electrical connections between electrical neighboring municipalities. We have used this information to compute a set of determinants of the autoregressive parameter that appears in our (heteroskedastic) BoU spatial models. In this model we use the number of connections with *preceding* municipalities involved in common outages (*PNU*) and the average capacity of these connections (*PCA*), as well as the total number of connections of municipality *m* with other municipalities (*NU*) and their average capacity (*CA*).

We have also computed for each municipality the following information: the distribution grid length (km's of medium voltage grid), the proportion of overhead and underground lines, the proportion of outdoor, indoor and underground step-down transformers, the proportion of remote controlled distribution station or step-down transformer on monthly basis, the average geographical position of all the distribution stations and step-down transformers that are placed in each municipality, and the minimum geographical distance between each municipality and the closest one with operation and maintenance crews located on it.

The above information has been used to compute the technological and environmental drivers included in our SF models. As in previous literature, we include a set of network characteristics to explain restoration times. We first include the logarithm of the length of the network (*lnNL*) as it might capture accessibility problems to outages. We have also included the natural logarithm of installed capacity (*lnCAP*) to control for differences in power capacities between municipalities. This variable is highly correlated with the number of transformers and it can also be interpreted as a measure of network complexity. We also consider the proportion of distribution stations or step-down transformers fitted with a remote-control system (*DIGT*) because they help to restore supply more quickly. This variable is used here to measure the effect of grid automatization on the restoration times. The proportion of underground (*UT*) and outdoor transformers (*OT*), and the proportion of overhead lines (*OL*) are also included as explanatory variables due to differences in accessibility to the damaged equipment or broken lines. For instance, access physically to outdoor transformers or overhead lines is easier than in the case of underground transformers and cables.

The restoration times in some cases might be larger than originally expected because all firm's maintenance crews are far from the outages, in particular if the outage is located in the countryside. To capture this location effect, we first include a dummy variable (*RURAL*) indicating the rural nature of the municipality. We next have added the logarithm of the distance to the nearest maintenance crews (*lnDIS*) as explanatory variable. This variable is iterated with the previous *RURAL* dummy variable as well as its urban counterpart (*URBAN*) to capture differences between rural and urban distances.

Finally, we have included the proportion of outages caused by weather-related issues (*WEATHER*) in order to see if weather conditions not only might increase the frequency of outages (as shown in previous literature) but also increase restoration times. To capture seasonal variations in weather conditions we include a set of dummy variables identifying the *WINTER*, *SPRING* and *SUMMER* seasons of the year. We ultimately include a set of regional dummy variables in our SF models to control for unobserved differences in the electricity distribution network between Spanish provinces (i.e. *ASTURIAS*, *HUESCA*, *ZARAGOZA*, *ALICANTE*, *VALENCIA* and *MADRID*).

Regarding the inefficiency term, we only include the above set of regional variables to capture seasonal variations in the relative performance of each maintenance crew. Such differences can likely be attributed to a concentration of extremely difficult-to-manage outages in specific seasons, for instance due to extremely bad weather conditions or faults located in very difficult-to-access places. The previous set of variables captures within-year changes in maintenance crews' inefficiency. In order to capture general trends in maintenance inefficiency over time (i.e. between-year changes), we have also included a third-order polynomial function of monthly time trends.

Tables 1 and 2 provide the descriptive statistics of the variables used in our BoU spatial and SF models, respectively.

[Table 1 here]

[Table 2 here]

4. Empirical specifications and results

4.1. BoU and standard SAR models

As aforementioned, we try to explain the duration of the external PSI using spatial autoregressive models. For comparison grounds, we estimate several specifications of our bottom-up spatial model. The first BoU spatial model uses exactly the specification that appears in (4). In this model, all variables are in levels. In the second model we estimate (4) but in per outage terms. In this case, our dependent variable is $\bar{e}_{mt} = \sum_{j=1}^{J_{mt}} e_{jmt} / J_{mt}$ and $\Omega Y_{mt} / J_{mt}$ is used as spatial lagged variable. In the third BoU spatial model, we again use the original restoration times, but now in logs. In this model, the dependent variable is $\ln Y E_{mt}$ and the spatially lagged variable is $\ln \Omega Y_{mt}$. Our final BoU spatial model uses the log of the per outage variables, i.e. $\ln(Y E_{mt} / J_{mt})$ and $\ln(\Omega Y_{mt} / J_{mt})$.

It is germane to recall here that the above spatially lagged variables do not coincide with the spatially lagged variable used in a standard SAR specification. While the lagged variable in a standard SAR is the spatial lag of the dependent variable (e.g. if the dependent variable is $Y E_{mt}$, the spatially lagged variable is $W Y E_{mt}$), in our BoU spatial model the spatially lagged variables is computed using both neighboring external and internal PSI (i.e. we use ΩY_{mt} and not $\Omega Y E_{mt}$). A second difference has to do with the spatial matrix. While the W matrix in a standard SAR model aggregates the duration of *all* PSI occurred in adjacent municipalities to the municipality m in period t , the Ω matrix takes into account the true sequence of the PSI across municipalities, and thus it only includes *common* PSI. Moreover, the Ω matrix only includes preceding municipalities involved in common outages, in contrast to the standard SAR model where the spatially lagged variable is computed using contemporaneous values.

As the standard SAR model ignores the true physical propagation of PSI in a real electricity distribution network, it likely provides biased results. To examine how severe this issue is, we also compare our BoU results with those obtained using standard SAR model. This robustness analysis is carried out using a simple homoscedastic (linear) specification for the autoregressive parameter as it is customary in the spatial literature.

Table 3 summarizes the main results. Regarding the BoU spatial models, we find clear improvements in goodness-of-fit either using logs or per-outage values. In this sense, as expected, the best specification is the model that uses the log of the per outage variables. The estimated autoregressive parameters in all BoU models are sound as they are positive, and quite

large (between 0.74 and 1.02). This is an expected result given that we have assumed that the restoration times of preceding municipalities involved in common outages are the main factor explaining external PSI. On the other hand, it is worth remarking on the poor goodness-of-fit of all standard SAR models, especially if the variables are not expressed in natural logarithms. The low values of the estimated autoregressive parameters in all standard SAR models are also worth mentioning, in particular in the specifications where the variables are not expressed in natural logarithms. Although they are statistically significant, they are close to zero. As the external interruptions only appear if $\Omega Y_{mt} > 0$, we do not expect significant intercept values. This happens in all BoU models, whether we use levels, per outage values or logs. In all standard SAR models, the estimated intercept is statistically significant, and even negative. All these results can be taken as evidence of the biases caused by using ad-hoc specifications of the spatial weight matrix.¹²

[Table 3 here]

In [Table 4](#) we show the results of two heteroskedastic specifications of our BoU spatial model. As the best goodness-of-fit were found using logged variables, we only show the results using variables in logs and per-outage variables in logs. [Table 3](#) shows that the heteroskedastic specification in both models improves the R-squared statistics and that most of the determinants of the autoregressive parameter or function have a significant effect on γ_m . The average value of the autoregressive parameter is slightly less than the estimated valued using its homoskedastic counterpart in [Table 3](#). However, the autoregressive parameter does vary across municipalities.

[Table 4 here]

The remarkable variation in γ_m can be clearly appreciated in [Figure 1](#) that presents the kernel density of the autoregressive parameter estimates. Both kernel densities are similar, indicating that our results are robust to using or not per-outage variables. This figure also shows that the autoregressive parameter estimates range from 0.46 and 0.93. Note, however, that the modal value is close to 0.55 in both specifications. This means that if there were 100% increases in restoration times in preceding municipalities, the restoration times in a subsequent municipality will increase by 55% if the preceding outages have the same nature. As all autoregressive parameter estimates are less than unity in [Figure 1](#), there are not one-to-one contagion effects in most municipalities if restoration time (changes) is measured in logs (growth rates). We have found values much larger than unity when logs are not used to estimate model (4). This is also a sound result because the power-adjusted restoration time in one municipality might be larger than in the preceding municipality simply due to the installed capacity in the subsequent municipality being larger than in the preceding municipality.¹³

[Figure 1 here]

[Table 4](#) also shows the parameter estimates of the autoregressive function $\gamma_m = \exp(\lambda q_m)$. As mentioned in [Section 2](#), we expect that the contagion degree of municipality m from preceding municipalities depends on its connectivity with not only these municipalities, but also with other municipalities that might attenuate the lack of supply from the preceding municipalities. Conditional on the number of connections with *preceding* municipalities involved in common outages ($\ln PNU$) and their average capacity ($\ln PCA$), the coefficients of

¹² In both cases we have used spatial lags of Y_{mt} . Larger biases are expected if the standard SAR models are estimated using YE_{mt} , i.e. using a wrong spatially dependent variable.

¹³ Therefore, the use of logged variables reduces the importance of the differences in capacity between preceding and subsequent municipalities when estimating spatial spillovers.

the number of connections of municipality m with other municipalities ($lnNU$) and their average capacity ($lnCA$) capture the effect of municipality m 's connectivity with *non-preceding* municipalities, i.e. subsequent municipalities involved in common outages plus municipalities that are connected with municipality m but not involved in common outages. The parameter estimates in Table 4 confirm our expectations. The contagion degree increases with the average capacity of the connections with preceding municipalities, as expected. The relationship with the number of connections with preceding municipalities is not statistically significant. We also find the expected result that γ_m decreases with $lnNU$ and $lnCA$ because other sources of supply tend to attenuate the impact of PSI imported from preceding municipalities.

4.2. Frontier models

The duration of the *internal* PSI is modelled using standard non-spatial stochastic frontier (SF) models, where the dependent variable is now the duration of all *internal* PSI occurred in municipality m in period t . As we have found convergence problems when the dependent variable is not logged, we discuss the frontier results using $lnYI_{mt}$ and $ln(YI_{mt}/I_{mt})$.

Table 5 shows the parameter estimates of our two SF models. In general, both specifications of equation (5) yield very similar results, except for $lnNL$. The coefficient of this variable is positive and statistically significant when restoration times are in logs, but not when they are expressed in per-outage terms. Both results together seem to suggest that the monthly interruptions increase with both network length. On the other hand, the fact that we do not find a significant effect on per-outage restoration times indicates that, in general, the differences in accessibility between municipalities do not matter to restoring power supply.¹⁴ The coefficient of $lnCAP$ is positive in both specifications of the SF model. This is an expected result because our dependent variable is a power-adjusted measure of the restoration times in each municipality and the installed capacity varies notably across municipalities. However, the positive effect found in both specifications might also indicate that the severity of the PSI increases with the number of transformers of the network.

[Table 5 here]

The percentage of distribution stations or step-down transformers fitted with a remote-control system ($DIGT$) allows us to measure the effect of grid automatization on the restoration times. We find a negative and statistically significant effect on restoration times in both specifications. Therefore, grid automatization does reduce the duration of all PSI and the restoration times of each outage. If the percentage of digital transformers increases in 10% points, the restoration times will be reduced about 8%. On the other hand, we do not find a significant effect of underground transformers (UT) on restoration times. This is an expected result because the reference transformers are located indoor, and there are no differences in restoration times with the underground transformers from an engineering point of view.

The proportion of outdoor transformers (OT) has a negative and significant effect on restoration times when the restoration times are not measured in per-outage terms. We originally expected a positive effect because firm' engineers believe that fixing an outdoor transformer normally requires more time than an indoor or underground transformer. Therefore, the negative coefficient is likely to do with the fact that most of the outdoor

¹⁴ If the lack of accessibility has been extraordinarily relevant in occasional outages, its effect should be captured by the corresponding inefficiency score.

transformers are located in rural areas where the installed capacity tends to be small. In this way, the percentage of outdoor transformers might have a negative effect on a power-adjusted measure of restoration times. Similar comments can be made for the negative coefficient found for the proportion of overhead lines (*OL*). The overhead lines are mainly located in rural areas. Therefore, the overhead lines tend to supply power to a smaller number of customers per network kilometer. However, the coefficient of the *RURAL* dummy variable is positive and statistically significant in both specifications. As the proportion of overhead lines in rural municipalities (85%) is higher than in more urban municipalities (78%), this dummy variable thus might be capturing part of the effect of *OL* on restoration times.

The two variables included in the model to measure the effect of the distance to the nearest maintenance crews (*lnDIS*) on outage durations have positive coefficients and in general are statistically significant. This result seems to confirm that the location of maintenance crews is a relevant factor to managing PSI, especially if the municipality affected for the outage is more urbanized. The larger effect for urban municipalities has likely to do with the fact that the installed capacity in these municipalities tend to be larger than in urban areas. In addition, we do not find an increasingly effect of the distance on restoration times given the lack of significance of the coefficient of the quadratic distance term.

We have tried to capture the effect of weather conditions on restoration times through the *WEATHER* variable, that measures the proportion of outages caused by weather-related issues, plus a set of seasonal dummy variables. The larger the proportion of outages caused by weather-related issues, the larger the restoration times are in both specifications. The estimated coefficient is however much smaller when we use a per-outage measure for the duration of the PSI. This result therefore seems to indicate that the weather conditions not only increase restoration time of each outage, but also the frequency of outages, as found in previous literature. Unlike the *WEATHER* variable, the three seasonal dummy variables included as frontier determinants do not provide additional information about the severity of the PSI.

Finally, and regarding the set of regional dummy variables, we have found highly significant coefficients, indicating that the above set of variables were not able to control for all differences in the electricity distribution network between Spanish provinces. It is worth highlighting the large values of the coefficients of *HUESCA* and *ZARAGOZA*, i.e. the two provinces of Aragón region. However, both coefficients are quite different because the network of both provinces differ notably.¹⁵ Similar comments apply to the two coefficients estimated for Alicante and Valencia provinces.

Figure 2 shows the average efficiency scores by provinces. They have been computed using the per-outage specification of our model because the goodness-of-fit is much better in this model, and in addition because the estimated efficiency scores do not depend on the number of power interruption, which is far from the control of the maintenance crews. Our results show a relatively good performance of the maintenance crews located in each province because the average value in all of them is larger than 92%. This figure also provides a very interesting result regarding the performance of the maintenance crews. The distribution of the inefficiency scores in Asturias and Huesca provinces are much more skewed than in other provinces. This is again a somewhat expected outcome because the municipalities of these two provinces are more rural than in other provinces and their population is widely scattered in

¹⁵ For instance, while 95% of the network in Huesca province is made up with overhead lines and its population is widely scattered in a wide area, the overhead lines in Zaragoza province only represents a 35% in Zaragoza province due it is a much more urbanized area.

wider areas. The large number of extremely low values for the efficiency scores in Asturias and Huesca reveals the existence of frequent difficult-to-restore PSI in these two provinces.

[Figure 2 here]

We depict the annual evolution of the efficiency scores in [Figure 3](#). Our SF model yields much lower efficiency scores in 2019 than in previous years. This is something that the distribution company must examine using the outage-level information. The Spanish regulator (CNMC) justifies the new regulatory framework on the poor quality of services performance of many Spanish electricity distribution utilities. The observed deterioration in quality of service seems thus to corroborate such decision.

[Figure 3 here]

5. Conclusions and final remarks

This paper uses a unique dataset to identify the main technological and managerial drivers of the duration of power outages. Unlike previous literature we focus our analysis on two issues: the emergence of grid digitalization and the location (and inefficiency) of maintenance staff. As we know the sequence of the individual PSI across municipalities, we use two different approaches to examine external and internal PSI. In a standard spatial application, only the aggregate effect of both PSI is observed by the econometrician.

The duration of the external PSI is modelled using a bottom-up SAR model that is developed using the engineering information of the PSI occurred in each municipality. Unlike the standard SAR model, our bottom-up model can be estimated using a simple OLS estimator because of the recursive nature of the spatially lagged variable. We show that the standard SAR model is seriously biased because it ignores the true sequence of the PSI across municipalities. In contrast, the duration of the internal PSI is modelled using a standard frontier model because the equipment that must be fixed is in the municipality initially affected by the outages. For this reason, the inefficient performance of the maintenance crews is only examined using internal PSI.

The estimated standard SAR models provide very poor goodness-of-fit as well as unsound autoregressive parameters. The logged specifications of our heteroskedastic BoU spatial models show that there are not one-to-one contagion effects in most municipalities, and that the degree of municipality contagion increases with the number of connections, and their average capacity, with preceding municipalities involved in common outages. However, the analysis also confirms that radial grid areas with few alternative routes to supply energy to end customers, often located in rural municipalities, are more affected by external outages than more interconnected areas, often in urban municipalities.

Regarding the internal PSI, our SF models show that both the weather conditions and the network characteristics influence quality of service in electricity distribution networks, as in previous literature. Unlike previous papers, we find a negative effect of transformers fitted with a remote-control system on restoration times in both specifications. Therefore, grid automatization does reduce the duration of all PSI and the restoration times of each outage.

We finally find that the performance of the maintenance crews is quite good because their average efficiency is larger than 92%. However, our efficiency analysis also reveals that there are numerous difficult-to-restore PSI in Asturias and Huesca provinces, an outcome that the firm might examine in detail in order to improve its quality performance in these two Spanish provinces. Another result that the firm should pay attention to is the observed deterioration in the estimated efficiency scores in 2019.

From a distribution company point of view, the analyzes carried out in this paper are useful to reinforce the idea that grid digitalization or remote control of grid assets is a must in order to reduce the impact of outages in terms of active power loss. Our results seem to confirm that the location of maintenance crews is also a relevant factor to managing PSI. Whether increasing capital expenditure by installing remote control systems in distribution stations or step-down transformers is more efficient than increasing operational costs in maintenance crews is an issue that we will try to explore in the future. We will also examine in the future whether the installation of a second line to increase the supply capacity to a neighborhood or a new commercial area is better than reinforcing the capacity of the existing line, or it is preferable to install underground assets.

Acknowledgements

The authors thank Charles Howell for providing helpful comments and suggestions on an earlier version of this paper.

Compliance with Ethical Standards

Funding:

Luis Orea thanks the “Salvador de Madariaga” grant obtained from the Spanish Ministry of Science, Innovation and Universities (Grant PRX19/00596) and the financial support from the Ministry of Education and Research of Madrid Region and FEDER funds (Grant H2019JHUM-5761).

Conflict of Interest:

Luis Orea and Pablo Argüelles declare that they have no conflict of interest.

References

- Coelho, J., Nassar, S. M., Gauche, E., Ricardo, V. W., Queiroz, H. L., de Lima, M., and Lourenço, M. C. (2003, June). Reliability diagnosis of distribution system under adverse weather conditions. In 2003 IEEE Bologna Power Tech Conference Proceedings, (Vol. 4, pp. 5-pp). IEEE.
- Corrado, L., & Fingleton, B. (2016). The W Matrix in Network and Spatial Econometrics: Issues Relating to Specification and Estimation. CEIS Tor Vergata Research Paper Series, 14(3), No. 369 - February 2016.
- Domijan Jr, A., Matavalam, R. K., Montenegro, A., Willcox, W. S., Diaz, J., Davis, L., and D'Agostini, J. (2003). Analysis of rain, wind and temperature effects on power distribution outages. In *Proceedings of the IASTED International Conference, Power-Con-Special Theme: Blackout* (pp. 44-48).
- Elhorst, J. P. (2010) Applied Spatial Econometrics: Raising the Bar, *Spatial Economic Analysis*, 5:1, 9-28,
- Giannakis, D., Jamasb, T., Pollitt, M., (2005). Benchmarking and incentive regulation of quality of service: an application to the UK electricity distribution networks. *Energy Policy* 33 (1), 2256–2271.
- Glass, A. J., Kenjegalieva, K., and Sickles, R. C. (2016). A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers. *Journal of Econometrics*, 190(2), 289-300.
- Gude, A. Álvarez, I., and Orea, L., (2018), “Heterogeneous spillovers among Spanish provinces: A generalized spatial stochastic frontier model”, *Journal of Productivity Analysis*, 50(3), 155-173.
- Jamasb, T., L. Orea and M. Pollitt (2012). Estimating marginal cost of quality improvements: the case of the UK electricity distribution companies. *Energy Economics* 34(5), 1498-1506.
- Jondrow, J., Lovell, C. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of econometrics*, 19(2-3), 233-238.
- Kjølle, G. H., Seljeseth, H., Heggset, J., and Trengereid, F. (2003). Quality of supply management by means of interruption statistics and voltage quality measurements. *European Transactions on Electrical Power*, 13(6), 373-379.
- Kumbhakar, S. C., Parmeter, C. F., and Tsionas, E. G. (2013). A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172(1), 66-76.
- Neumayer, E. and Plümper, T. (2010). Spatial effects in dyadic data. *International Organization*, 64(1), 145-166.
- Skevas, I. (2020). Inference in the spatial autoregressive efficiency model with an application to Dutch dairy farms. *European Journal of Operational Research*, 283(1), 356-364.
- Wang, P., and Billinton, R. (2002). Reliability cost/worth assessment of distribution systems incorporating time-varying weather conditions and restoration resources. *IEEE Transactions on Power Delivery*, 17(1), 260-265.

- Yang, S., Harlow, L. I., Puggioni, G., & Redding, C. A. (2017). A comparison of different methods of zero-inflated data analysis and an application in health surveys. *Journal of Modern Applied Statistical Methods*, 16(1), 518-543.
- Yang, Si., Puggioni, G., Harlow, L.L., and Redding, C.A. (2017). A Comparison of Different Methods of Zero-Inflated Data Analysis and an Application in Health Surveys. *Journal of Modern Applied Statistical Methods*, 16(1), 518-543.
- Yu, W., Jamasb, T., and Pollitt, M. (2009). Does weather explain cost and quality performance? An analysis of UK electricity distribution companies. *Energy Policy*, 37(11), 4177-4188.
- Zhou, Y., Pahwa, A., and Yang, S. S. (2006). Modeling weather-related failures of overhead distribution lines. *IEEE Transactions on power systems*, 21(4), 1683-1690.

Table 1. Descriptive statistics. SAR models

Variable	Definition	Units	Obs.	Mean	Std. Dev.	Min	Max
YE	Restoration times of external PSI	seconds	1516	62.876	487.467	0.001	10252
J	Number of internal PSI	number	1516	6.658	8.463	1	77
WY	Restoration times of preceding municipalities	seconds	1516	69.711	420.295	0.001	10570
WY	Restoration times of connected municipalities	seconds	1516	1025.473	9953.262	0.001	215927
UN	Number of connections with neighboring municipalities	number	1516	21.090	20.138	1	88
CA	Capacity of connections with neighboring municipalities	KW	1516	931.677	1092.167	2	3962
PNU	Number of connections with preceding municipalities	number	1516	6.660	7.879	1	41
PCA	Capacity of connections with preceding municipalities	KW	1516	269.725	463.876	2	2530

Table 2. Descriptive statistics. SF models

Variable	Definition	Units	Obs.	Mean	Std. Dev.	Min	Max
YI	Restoration times of internal PSI	seconds	3739	217.349	3958.901	0	215154
I	Number of internal PSI	number	3739	6.126	7.880	1	82
NL	Network length	km	3739	395	431	2	2125
OL	Proportion of overhead lines	proportion	3739	0.807	0.259	0	1.00
DIGT	Proportion of digital transformers	proportion	3739	0.073	0.082	0	0.50
UT	Proportion of underground transformers	proportion	3739	0.025	0.095	0	0.79
OT	Proportion of outdoor transformers	proportion	3739	0.562	0.296	0	1.00
WEATHER	Proportion of outages caused by weather issues	proportion	3739	0.013	0.092	0	1.00
DIS	Distance to the nearest maintenance crew	km	3739	13.902	10.895	0	51
RURAL	Rural municipality	dummy	3739	0.362	0.481	0	1
URBAN	Urban municipality	dummy	3739	0.638	0.481	0	1
AUTUMN	Season of the year	dummy	3739	0.252	0.434	0	1
WINTER	Season of the year	dummy	3739	0.244	0.430	0	1
SPRING	Season of the year	dummy	3739	0.249	0.432	0	1
SUMMER	Season of the year	dummy	3739	0.255	0.436	0	1
ASTURIAS	Spanish province	dummy	3739	0.857	0.350	0	1
HUESCA	Spanish province	dummy	3739	0.059	0.236	0	1
ZARAGOZA	Spanish province	dummy	3739	0.010	0.100	0	1
ALICANTE	Spanish province	dummy	3739	0.014	0.116	0	1
PALENCIA	Spanish province	dummy	3739	0.043	0.203	0	1
MADRID	Spanish province	dummy	3739	0.017	0.128	0	1
2013	Year	dummy	3739	0.145	0.352	0	1
2014	Year	dummy	3739	0.148	0.355	0	1
2015	Year	dummy	3739	0.153	0.360	0	1
2016	Year	dummy	3739	0.144	0.351	0	1
2017	Year	dummy	3739	0.146	0.353	0	1
2018	Year	dummy	3739	0.136	0.343	0	1
2019	Year	dummy	3739	0.128	0.334	0	1

Table 3. Linear homoscedastic SAR models

	BoU		Standard	
	Coef.	t-ratio	Coef.	t-ratio
In levels				
Intercept	8.995	0.95	58.885 ***	4.69
Autoregressive parameter (γ)	0.773 ***	34.78	0.004 ***	3.10
R-squared	0.444		0.006	
Per outage				
Intercept	-4.578	-0.89	24.678 ***	3.17
Autoregressive parameter (γ)	1.029 ***	44.73	0.003 ***	2.73
R-squared	0.569		0.005	
In logs				
Intercept	0.037	0.89	-0.680 ***	-7.24
Autoregressive parameter (γ)	0.724 ***	43.08	0.462 ***	23.59
R-squared	0.551		0.269	
Per outage in logs				
Intercept	0.003	0.10	-0.511 ***	-9.21
Autoregressive parameter (γ)	0.745 ***	49.98	0.437 ***	29.46
R-squared	0.623		0.364	
Obs.	1516		1516	

Table 4. BoU heteroscedastic SAR models

	In logs			Per outage in logs		
	Coef.	s.e.	t-ratio	Coef.	s.e.	t-ratio
α	0.149 ***	0.043	3.51	0.055 *	0.029	1.88
λ_0	-0.483 ***	0.034	-14.37	-0.453 ***	0.036	-12.76
$\ln NU (\lambda_1)$	-0.125 ***	0.031	-4.05	-0.066 **	0.031	-2.17
$\ln CA (\lambda_2)$	-0.108 ***	0.035	-3.07	-0.114 ***	0.036	-3.17
$\ln PNU (\lambda_3)$	0.013	0.037	0.34	-0.059	0.038	-1.56
$\ln PCA (\lambda_4)$	0.050 *	0.028	1.76	0.077 ***	0.029	2.68
R-squared	0.582			0.640		
Average SAR parameter	0.626			0.643		
Obs.	1516			1516		

Table 5. Parameter estimates of the SF models

	In logs			Per outage in logs		
	Coef.	s.e.	t-ratio	Coef.	s.e.	t-ratio
<i>Frontier</i>						
Intercept	1.472 ***	0.065	22.79	0.856	0.058	14.87
<i>ln</i> NL	0.262 ***	0.067	3.89	-0.092 *	0.048	-1.91
<i>ln</i> CAP	0.374 ***	0.062	6.00	0.229 ***	0.044	5.15
DIGT	-0.809 *	0.483	-1.68	-0.750 **	0.345	-2.17
UT	-0.149	0.519	-0.29	0.457	0.369	1.24
OT	-0.710 ***	0.246	-2.89	-0.282	0.175	-1.61
OL	-1.905 ***	0.482	-3.95	-1.086 ***	0.343	-3.17
RURAL	0.455 ***	0.085	5.34	0.153 **	0.061	2.53
<i>ln</i> DIS·RURAL	0.175 ***	0.040	4.40	0.032	0.028	1.11
<i>ln</i> DIS·URBAN	0.195 ***	0.058	3.39	0.100 **	0.041	2.45
$1/2$ <i>ln</i> DIS ²	0.022	0.038	0.56	-0.010	0.027	-0.36
WEATHER	1.954 ***	0.234	8.34	0.832 ***	0.166	5.00
WINTER	-0.110 *	0.064	-1.74	-0.061	0.054	-1.13
SPRING	-0.097	0.062	-1.55	-0.038	0.069	-0.56
SUMMER	-0.094	0.062	-1.52	-0.087	0.058	-1.49
HUESCA	5.461 ***	0.196	27.89	4.218 ***	0.139	30.25
ZARAGOZA	3.706 ***	0.303	12.25	3.570 ***	0.216	16.56
ALICANTE	1.913 ***	0.318	6.02	2.647 ***	0.226	11.72
VALENCIA	0.623 *	0.324	1.92	1.257 ***	0.230	5.46
MADRID	1.272 ***	0.336	3.78	1.921 ***	0.239	8.03
<i>Noise term</i>						
Intercept	0.506 ***	0.023	21.75	-0.173 ***	0.023	-7.42
<i>Inefficiency term</i>						
Intercept	-10.227	7.955	-1.29	0.494	0.305	1.62
t	0.531 *	0.315	1.69	-0.028 *	0.015	-1.80
$1/2 \cdot t^2$	-0.005 *	0.003	-1.82	0.000 *	0.000	1.92
$1/3 \cdot t^3$	1.229	1.659	0.74	-0.001	1.367	0.00
WINTER	-1.121	0.979	-1.15	2.176	1.728	1.26
SPRING	-1.561 **	0.788	-1.98	1.174	1.393	0.84
SUMMER	-10.871	147.2	-0.07	-10.591 **	4.804	-2.20
Obs	3739			3739		
Log likelihood	-6260.1			-4988.9		

Figure 1. Kernel densities of spatial autoregressive values.

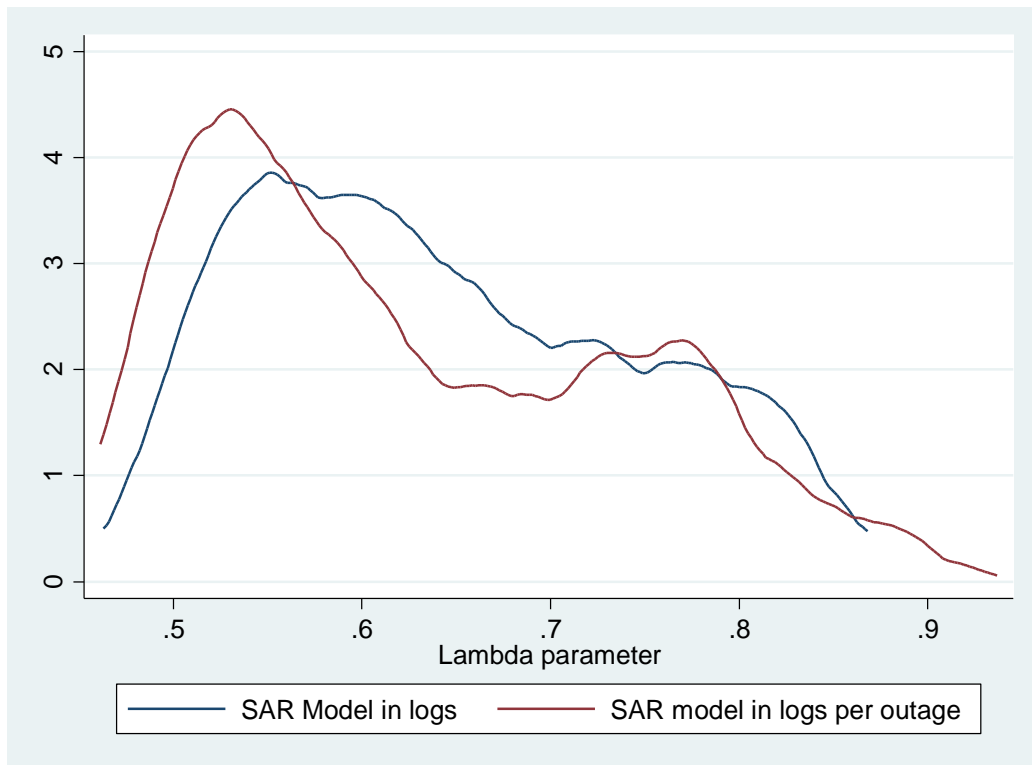


Figure 2. Average efficiency scores by provinces.

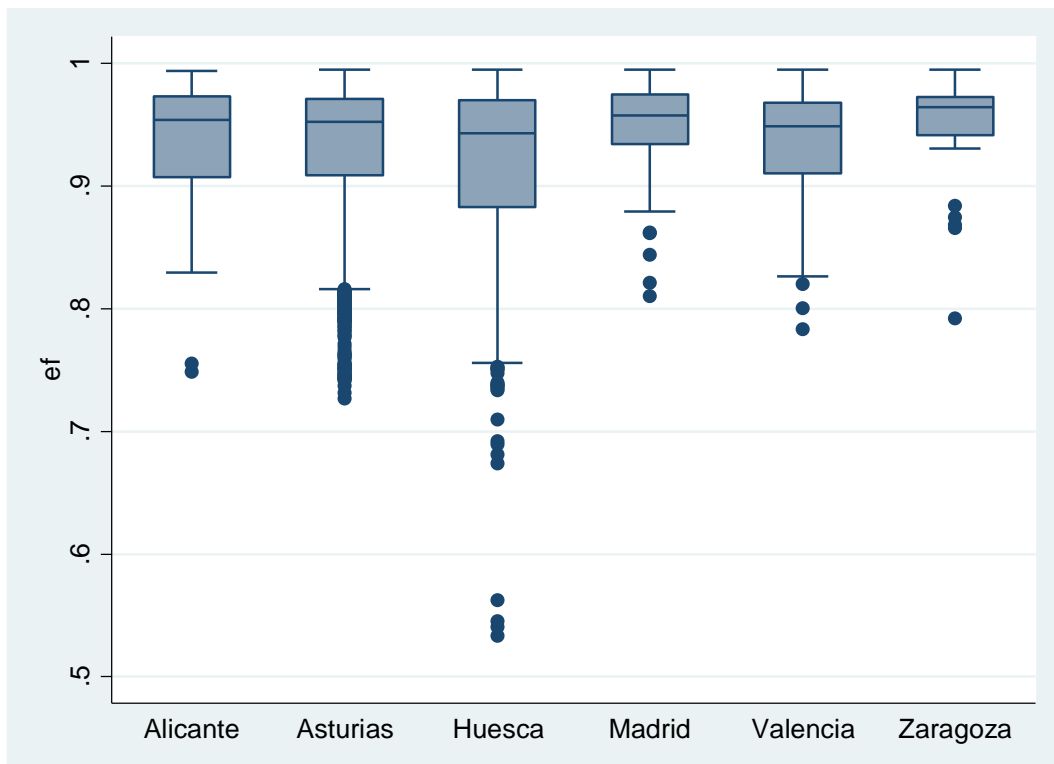


Figure 3. Average efficiency scores over time.

