Welfare gains of the poor:

An endogenous Bayesian approach with spatial random effects*

Andrés Ramírez Hassan† Santiago Montoya Blandón‡

May 5, 2016

Abstract

We introduce a Bayesian instrumental variable procedure with spatial random effects that handles endogeneity, and spatial dependence with unobserved heterogeneity. We find through a limited Monte Carlo experiment that our proposal works well in terms of point estimates and prediction. We apply our method to analyze the welfare effects generated by a process of electricity tariff unification on the poorest households. In particular, we deduce an Equivalent Variation measure where there is a budget constraint for a two-tiered pricing scheme, and find that 10% of the poorest municipalities attained welfare gains above 2% of their initial income.

JEL Classification: C11, C15, D11, D12, D60
Keywords: Bayesian Estimation, Endogeneity, Instruments, Simultaneous Equations, Spatial Random Effects, Welfare Analysis

*We thank the participants in the seminar of the research group of *Economía y Empresa* from Universidad EAFIT for comments. All remaining errors are our own.

†Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia, Carrera 49 - 7 S 50; Ph.: (+57)(4)261 95 00, ext. 9402; email: aramir21@eafit.edu.co

‡Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia; email: smontoy4@eafit.edu.co

Introduction

We propose a Bayesian simultaneous equations system with spatial random effects suited to handling spatial dependence and heterogeneity, endogeneity, and statistical inference associated with complicated non-linear functions of the parameter estimates. In particular, we define a system of simultaneous equations where a conditional correlation between the stochastic errors captures the endogeneity, and instrumental variables are used to model the endogenous variables. In addition, we employ a Bayesian hierarchical spatial framework, based on a Conditional Autoregressive (CAR) spatial prior, to structure the unobserved heterogeneity and the spatial dependence. After model specification, we find the conditional posterior distributions of all the parameter sets, thus we can use Gibbs sampling algorithms to draw simulations of all our posterior distributions.

We perform a limited Monte Carlo simulation exercise where we find that our proposal obtains good outcomes regarding point estimation compared with competing alternatives. In addition, prediction is substantially improved introducing spatial effects.

Establishing a Bayesian approach allows performing statistical inference related to functions of the parameter estimates using simple rules of probability theory.

We apply our methodology to evaluate the welfare implications for poor households, measured through the Equivalent Variation, caused by the electricity price changes which took place in the province of Antioquia (Colombia), after *Empresas Públicas de Medellín* (EPM) acquired *Empresa Antioqueña de Energía* (EADE) in 2006. The Equivalent Variation is a non-linear function of parameter estimates of a demand function, which we estimate using data at the municipality level. Therefore, in our empirical exercise, we should take into consideration spatial effects, endogeneity between price and electricity demand, and unobserved heterogeneity due to latent economical, cultural and geographical factors. Finally, we want to perform statistical inference regarding the Equivalent Variation. This application is interesting in itself because electricity services represent a significant share of households' budgets, and this fact is prominent for the poor population (Gomez-Lobo, 1996, You and Lim, 2013). As a consequence, small variations in electricity prices may have relevant impacts on the welfare of households.

Consideration of spatial effects while performing statistical inference based on cross-sectional areal data has a long tradition in spatial statistics (Cressie, 1993, Ripley, 2005), and more recently in spatial econometrics (Anselin, 1988). Most of the methods in spatial econometrics have been based on the Frequentist approach, although there are some remarkable exceptions founded on the Bayesian framework (LeSage, 1997, 2000, Parent and LeSage, 2008, LeSage and Pace, 2009, LeSage and Llano, 2013).

The issue of endogeneity emerges naturally due to the presence of the spatial lag of the dependent variable in Spatial Autoregressive (SAR) models, and spatial econometric estimators have taken this problem into consideration since its beginning (Anselin, 1990, Kelejian and Prucha, 1998, 1999). However, the treatment of endogeneity due to other regressors has only recently been analyzed (Rey and Boarnet, 2004, Kelejian and Prucha, 2004, Fingleton and Le Gallo, 2008, Drukker et al., 2013, Liu and Lee, 2013). Thus, these new kind of spatial estimators take into consideration spatial dependence and feedback endogeneity simultaneously. However, they fail to take into account unobserved heterogeneity, and have to rely on approximations, like the Delta method, to perform statistical inference regarding complicated non-linear functions of the parameter estimates.

Unobserved heterogeneity is another issue that may arise with cross-sectional areal data (Parent and LeSage, 2008). Unfortunately, to the best of our knowledge, there is only limited spatial econometric literature regarding this issue. This fact may be due to the difficulty of introducing unobserved heterogeneity in cross-sectional areal data using Frequentist methods. However, LeSage (2000), Smith and LeSage (2004), LeSage et al. (2007), Parent and LeSage (2008), Seya et al. (2012) and LeSage and Llano (2013) have tackled unobserved heterogeneity using a Bayesian approach, simultaneously including spatial effects and unobserved heterogeneity, but they do not take into consideration recursive endogeneity: an issue that has been considered from a Bayesian perspective in Drèze (1976), Kleibergen and Van Dijk (1998), Zellner (1998), Kleibergen and Zivot (2003).

To the best of our knowledge, we find that few authors have studied welfare effects due to changes within block price systems, and even fewer have introduced spatial random effects within an endogenous framework to analyze these welfare implications. From a microeconomic perspective, there are three main streams: consumer surplus (Acton and Mitchell, 1983), compensating variation (Gomez-Lobo, 1996, Dodonov et al., 2004) and equivalent variation (Dodonov et al., 2004, Lundgren, 2009, Ruijs, 2009, You and Lim, 2013). All of these authors have estimated the welfare implications without consideration of uncertainty due to parameter estimates. From a theoretical standpoint, the main consideration for adopting the Bayesian approach is that it allows us to establish a statistical framework that simultaneously unifies decision theory, statistical inference, and probability theory under a single philosophically and mathematically consistent structure. From an empirical perspective, the Bayesian approach has some advantages in the present setting. In particular, we can easily make statistical inferences associated with the Equivalent Variation, which is a complicated function of the parameter estimates, using simple rules of probability theory. In addition, our econometric approach takes into account the endogeneity issue that is present, allowing us to identify the structural parameters from the reduced form in our empirical exercise. Finally, a Bayesian framework permits us to introduce spatial random effects in our cross-sectional areal data structure, and control the unobserved heterogeneity and autocorrelation that can arise in spatial settings.

Using data at the municipality level for the province of Antioquia, and different spatial contiguity criteria, we find that the posterior mean of the price, income, substitute and urbanization rate demand elasticities are -0.88, 0.30, 0.12 and 0.57, respectively. In addition, the posterior mean of the semi-elasticity of electricity demand associated with a sea level dummy, which is equal to one when the municipality is located less than 1000 meters above sea level, and zero otherwise, is approximately 0.14. With these estimates as inputs, we calculate the posterior distribution of the Equivalent Variation welfare measure as a share of income for each municipality. We deduce this measure using a logarithmic demand function, and taking into account a budget constraint for a two-tiered pricing scheme. We find that the average household enjoys a mean welfare gain of approximately 0.87% of their initial income, which can be considerable when taking their socioeconomic situation into account. However, these results depend heavily on whether the municipalities are part of the Metropolitan Area or not, on their average electricity consumption levels, and on other geographical and economic factors. For example, Medellín, the capital of the province, and its main center of economic activity, presented a mean welfare gain equal to 0.14%, which is approximately equal to the average improvement for all Metropolitan Area municipalities. On the other hand, municipalities located outside of the Metropolitan Area had, in total, mean welfare gains equal to 0.94%. In particular, 11 of the less urban municipalities, which are also the poorest, had welfare gains above 2% of their initial income. Just to serve as a reference, low income households in Colombia expend on average 1.13%, 2.04%, and 4.79% of their income on pension, health care, and education, respectively (DANE, 2015). This illustrates how important are the welfare implications of utility regulation: price changes in this sector may have huge effects on households' welfare, especially for the poorest.

The remainder of this paper is organized as follows. Section 1 outlines the complete endogenous Bayesian modeling strategy, the model formulation, our prior specification, the deduction of the conditional posterior densities, as well as the results of our simulation exercises. Section 2 addresses the generalities of the Colombian energy market that are fundamental to the understanding of our application. Section 3 deals with the microeconomic foundation of the Equivalent Variation welfare measure, its ties to the econometric specification of the system of equations, and derives the measure for the specific case of a logarithmic demand function taking into consideration a budget constraint for a two-tiered pricing scheme. Section 4 is divided into four subsections. The first presents summary statistics for the data used in the econometric exercise. The second presents the specific characteristics of our econometric specification for the application. The third presents a summary of the results of our demand equation estimation, with some robustness checks regarding the spatial structure. The fourth presents the main findings for our application: the analysis of the posterior distribution of the welfare effects and its geographical patterns. Finally, Section 5 presents our conclusions.

1 Econometric Approach

1.1 The model

The formulation of our model is

$$y_i = \delta_0 + \mathbf{z}_{1i}'\boldsymbol{\delta}_1 + \alpha x_i + u_{1i} + v_i \tag{1}$$

where y_i is the variable of interest that dependents on a set of K_1 exogenous controls $\mathbf{z}_{1i}$, and an endogenous regressor x_i such that $\text{E}(x_i u_{1i}) \neq 0$. Omitting this fact would generate biased and inconsistent parameter estimates.

We have two error terms, u_{1i} is an idiosyncratic stochastic shock, which is involved in the endogeneity issue, and v_i is a spatial random effect to control for spatial heterogeneity and spatial dependence between cross-sectional units. This dependence emerges due to clusters and/or spillover effects between neighboring regional units, and allows us to control for unobservable spatial heterogeneity. In particular, we assume that each spatial random effect has an improper (intrinsic) conditionally autoregressive (CAR) structure (Besag et al., 1991):

$$v_i|\mathbf{v}_{i\sim j} \sim \mathcal{N}\left(\sum_{i\sim j}\frac{w_{ij}v_j}{\sum_{i\sim j} w_{ij}}, \frac{\sigma_v^2}{\sum_{i\sim j} w_{ij}}\right) \tag{2}$$

where $\mathbf{v}_{i\sim j}$ is a vector of stochastic error spatial components for the neighbors j of i ($i \sim j$), and w_{ij} are the elements of the contiguity matrix which defines the spatial structure of the model. The joint distribution of the improper CAR is given by $\mathbf{v} \sim \mathcal{N}_N(\underline{\mathbf{v}}, \sigma_v^2(\mathbf{D}_W - \mathbf{W}_N)^{-1})$ where $\mathbf{W}_N$ is the contiguity matrix and $\mathbf{D}_W = \text{diag}(\sum_{i\sim j} w_{ij})$ (Banerjee et al., 2004, Wall, 2004).

The contiguity relation is binary, that is, the ij element is equal to 1 if region i and j are neighbors, and 0 otherwise. Thus, the contiguity matrix is symmetric, which is a requirement of the CAR model. By definition, the elements on the main diagonal of the contiguity matrix are set equal to 0. σ_v^2 is a parameter that defines the conditional variance of the spatial process, where the conditional variance must be inversely proportional to the number of neighbors.

It is well known that the joint distribution of a CAR process is improper. Although we can obtain a proper CAR process by just introducing a single spatial autocorrelation parameter, we decide against it. The parameter's limited range between -1 and 1 can cause theoretical and computational difficulties, as well as limit the set of spatial patterns that the distribution can replicate, becoming much less intuitive (Banerjee et al., 2004).

There is a spatial literature that favors CAR specifications (Banerjee et al., 2004, Parent and LeSage, 2008, Darmofal, 2009, Chakraborty et al., 2013), and other that supports SAR specifications (Smith and LeSage, 2004, LeSage et al., 2007, Ohtsuka et al., 2010, LeSage and Llano, 2013). Our decision to use a CAR specification to model the spatial random effects, instead of a SAR specification, is due to the fact that heteroscedasticity is inherent to the CAR specification, and we achieve a higher level of heterogeneity (Cressie, 1993). In addition, the CAR specification is a Markovian process in space, that is, the spatial heterogeneity is due to local variation, rather than a global spatial pattern, which is present in SAR specification (Anselin, 2003). Our intuition is that the unobserved heterogeneity present in our application, which is related to residential electricity consumption in a municipality, is affected by the first order neighbors (see Section 4, maps 2 and 3 and the comments therein). A SAR specification cannot be used in a two-component disturbances decomposition like the one that we propose (Parent and LeSage, 2008), and parameter estimates do not have an easy interpretation in SAR models due to the presence of the spatial lags (Elhorst, 2014). Finally, a CAR prior offers computational convenience because we just need to work with its conditional distributions, avoiding matrix inversion. On the other hand, SAR models do not have full conditional distributions with a convenient form, and this increases the computational burden (Banerjee et al., 2004).

We set some exclusion restrictions in the main equation to handle endogeneity. These are associated with K_2 instrumental variables $\mathbf{z}_{2i}$ that do not affect y_i if x_i is held constant. Then,

$$x_i = \phi_0 + \mathbf{z}_{1i}'\boldsymbol{\phi}_1 + \mathbf{z}_{2i}'\boldsymbol{\alpha}_s + u_{2i} \tag{3}$$

where u_{2i} is an idiosyncratic stochastic shock such that $(u_{1i}, u_{2i})' \sim \mathcal{N}(\mathbf{0}, \Omega)$, $\Omega = \{\omega_{ij}\}$. Thus, ω_{12} captures the endogeneity of the system (Zellner, 1996).

1.2 A Bayesian estimation approach

We should keep in mind that our final objective is to carry out a statistical inference related to complicated non-linear functions of the parameter estimates, for instance the Equivalent Variation equations (13) and (14) in Section 3. This is an argument in favor of using a Bayesian framework due to obtaining full posterior distributions on all the parameters from Equation (1). Using simple probabilistic rules, we can obtain the posterior distributions of the functions of parameter estimates without any additional computational effort nor assumptions regarding asymptotic outcomes (Bernardo and Smith, 1994). On the other hand, using Frequentist methods would require estimating Equation (1) by means of instrumental variables (Rey and Boarnet, 2004, Kelejian and Prucha, 2004, Fingleton and Le Gallo, 2008, Drukker et al., 2013, Liu and Lee, 2013) or the generalized method of moments (Fingleton and Le Gallo, 2008, Drukker et al., 2013). It would then be necessary to implement spatial resampling algorithms (Lahiri et al., 2006) or the Delta method (Casella and Berger, 2002) to find the standard errors associated with functions of the parameter estimates. These tasks require extra computational and mathematical effort.

In addition, the Bayesian approach works well using weak instruments due to the fact that the likelihood function and its identification are less important for deriving estimates in Bayesian models (Zellner, 1996, Imbens and Rubin, 1997, Zellner, 1998).

Observe that the introduction of the spatial random effects, $v_i, i = 1, 2, \ldots, N$, is another argument in favor of the Bayesian approach. In particular, the additional N parameters cannot be estimated by means of maximum likelihood methods due to the limited number of degrees of freedom (Seya et al., 2012). Thus, we follow a Bayesian hierarchical approach to model the spatial random effects where each unit is associated with a particular v_i, and the conditional distributions of these parameters depend on their neighbors, through a contiguity matrix and a precision parameter that is drawn from a Gamma distribution.

We should bear in mind that the improper CAR is identified only up to an additive constant, thus, to identify the intercepts in our model, it is necessary to add the constraint $\sum_{i=1}^{N} v_i = 0$. As a consequence, it is necessary to use improper uniform priors for the constant terms (δ_0 and ϕ_0) in both equations.

To complete our Bayesian specification, we set the remaining priors as follows: $\boldsymbol{\delta} \sim \mathcal{N}_{K_1+1}(\underline{\boldsymbol{\delta}}, \underline{\boldsymbol{\Delta}})$ and $\boldsymbol{\phi} \sim \mathcal{N}_{K_1+K_2}(\underline{\boldsymbol{\phi}}, \underline{\boldsymbol{\Phi}})$ (see Hoogerheide et al., 2007, for constructing natural conjugate priors for instrumental variables regression in more general settings, but without considering spatial effects). In our application, we set $\underline{\boldsymbol{\delta}} = \mathbf{0}_{K_1+1}$, $\underline{\boldsymbol{\phi}} = \mathbf{0}_{K_1+K_2}$, $\underline{\boldsymbol{\Delta}} = 1000\mathbf{I}_{K_1+1}$ and $\underline{\boldsymbol{\Phi}} = 1000\mathbf{I}_{K_1+K_2}$. This implies vague prior information where there is no effect of each control variable on the dependent variables.

In addition, we assume a Wishart distribution for Ω^{-1}, that is, $\Omega^{-1} \sim \mathcal{W}_2(\underline{\omega}, \underline{\boldsymbol{\Omega}})$. In particular, we set $\underline{\omega} = 3$ and $\underline{\boldsymbol{\Omega}} = \mathbf{I}_2$, where setting the degrees of freedom to $p+1$, where p is the dimension of the covariance matrix, the Wishart form reduces to $\pi(\Omega^{-1}) \propto |\Omega^{-1}|^{-(N+1)/2}$, which is a diffuse prior used by Savage that emerges using Jeffrey's invariance theory (Zellner, 1996). Thus, a priori, there is no endogeneity, and the fat-tailed prior will guarantee the robustness of the outcomes regarding this distribution (Berger, 1985).

To specify the prior distribution of the precision parameter of the CAR component, we must take into account that there are two different sources of stochastic variability in our main equation, u_{1i} and v_i. As a consequence, both sets of hyperparameters of the prior distributions of these random effects cannot imply arbitrarily large variability, since these effects would be unidentifiable. In essence, we would be trying to identify two random effects using a single observation at each spatial unit. Therefore, we cannot use arbitrarily vague prior distributions in our hierarchical approach. We propose a *fair* argument to construct the prior distribution of the precision parameter of the CAR component (Banerjee et al., 2004). Specifically, we posit a priori that the proportion of the variability due to spatial effects is 0.5, that is, we set $\text{Var}(u_{1i}) = \text{Var}(v_i)$. Thus, taking into consideration that $u_{1i} \sim \mathcal{N}(0, \omega_{11})$ where $\omega_{11} \sim \mathcal{IG}((\underline{\omega}-1)/2, 1/2)$ due to our prior assumptions, and $\text{Var}(u_{1i}) = \omega_{11} \approx \frac{\sigma_v^2}{0.7^2\left(\sum_{i\sim j} w_{ij}\right)^{Ave}} \approx \text{Var}(v_i)$ (Bernardinelli et al., 1995) where $\left(\sum_{i\sim j} w_{ij}\right)^{Ave}$ is the average number of neighbors. We then find that the prior distribution of $1/\sigma_v^2$ is approximately proportional to $\mathcal{G}(\frac{\underline{\omega}-1}{2}, 1/2)$. Moreover, we find in our application that the posterior parameter estimates are robust to changes of the hyperparameters of the CAR's precision (available upon request).

Despite the fact that $\mathbf{v}$ has an improper distribution, Theorem 2 in Sun et al. (1999) guarantees that a proper posterior distribution exists if $(\mathbf{D}_W - \mathbf{W}_N)$ is nonnegative definite, the precision parameters have Gamma prior distributions, and the intercepts have diffuse prior distributions. We satisfy all of these criteria.

We assume that the prior distributions are independent, that is,

$$\pi(\Omega, \boldsymbol{\delta}, \boldsymbol{\phi}, \delta_0, \phi_0, \mathbf{v}_i, \sigma_v^2) = \pi(\Omega)\pi(\boldsymbol{\delta})\pi(\boldsymbol{\phi})\pi(\delta_0)\pi(\phi_0)\pi(\mathbf{v}_i|\sigma_v^2)\pi(\sigma_v^2) \tag{4}$$

1.3 Posterior distributions

The full conditional posteriors for all parameters are

$$\begin{gathered}
\Omega|\boldsymbol{\delta}, \boldsymbol{\phi}, \delta_0, \phi_0, \mathbf{v}, Data \sim \mathcal{IW}_2(\bar{\omega}, \bar{\Omega}) \\
\bar{\omega} = \underline{\omega} + N \\
\bar{\Omega} = \left[\underline{\Omega}^{-1} + \sum_{i=1}^{N} \begin{pmatrix} y_i - \delta_0 - \mathbf{w}_i'\boldsymbol{\delta} - v_i \\ x_i - \phi_0 - \mathbf{z}_i'\boldsymbol{\phi} \end{pmatrix} (y_i - \delta_0 - \mathbf{w}_i'\boldsymbol{\delta} - v_i, x_i - \phi_0 - \mathbf{z}_i'\boldsymbol{\phi})\right]
\end{gathered} \tag{5}$$

To sample $\boldsymbol{\delta}$, we use $f(y_i, x_i|\boldsymbol{\Theta}) = f(y_i|x_i, \boldsymbol{\Theta})f(x_i|\boldsymbol{\Theta})$ where $\boldsymbol{\Theta} = (\Omega, \boldsymbol{\delta}, \boldsymbol{\phi}, \delta_0, \phi_0, \mathbf{v})$. In particular, $y_i|x_i, \boldsymbol{\Theta} \sim \mathcal{N}\left(\delta_0 + \mathbf{w}_i'\boldsymbol{\delta} + v_i + \frac{\omega_{12}}{\omega_{22}}(x_i - \phi_0 - \mathbf{z}_i'\boldsymbol{\phi}), \psi_{11}\right)$ where $\psi_{11} = \omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}$. Then,

$$\begin{gathered}
\boldsymbol{\delta}|\Omega, \boldsymbol{\phi}, \delta_0, \phi_0, \mathbf{v}, Data \sim \mathcal{N}_{K_1+1}(\bar{\delta}, \bar{\Delta}) \\
\bar{\Delta} = \left[\underline{\Delta}^{-1} + \psi_{11}^{-1}\mathbf{W}'\mathbf{W}\right]^{-1} \\
\bar{\delta} = \bar{\Delta}\left[\underline{\Delta}^{-1}\underline{\delta} + \psi_{11}^{-1}\mathbf{W}'\mathbf{y}_1\right]
\end{gathered} \tag{6}$$

where $\mathbf{W}$ is an $N \times (K_1 + 1)$ matrix whose rows are $\mathbf{w}_i'$ and $\mathbf{y}_1$ is an $N \times 1$ vector whose elements are $y_i - \frac{\omega_{12}}{\omega_{22}}(x_i - \phi_0 - \mathbf{z}_i'\boldsymbol{\phi}) - \delta_0 - v_i$.

We follow the same procedure to deduce the conditional posterior distribution of $\boldsymbol{\phi}$. That is, we use $f(y_i, x_i|\boldsymbol{\Theta}) = f(x_i|y_i, \boldsymbol{\Theta})f(y_i|\boldsymbol{\Theta})$. Then, by a similar procedure as before, $x_i|y_i, \boldsymbol{\Theta} \sim \mathcal{N}\left(\phi_0 + \mathbf{z}_i'\boldsymbol{\phi} + \frac{\omega_{12}}{\omega_{11}}(y_i - \delta_0 - \mathbf{w}_i'\boldsymbol{\delta} - v_i), \psi_{22}\right)$ where $\psi_{22} = \omega_{22} - \frac{\omega_{12}^2}{\omega_{11}}$. Thus,

$$\begin{aligned}
\boldsymbol{\phi}|\boldsymbol{\delta}, \Omega, \delta_0, \phi_0, \mathbf{v}, Data &\sim \mathcal{N}_{K_1+K_2}(\bar{\phi}, \bar{\Phi}) \\
\bar{\Phi} &= \left[\underline{\Phi}^{-1} + \psi_{22}^{-1}\mathbf{Z}'\mathbf{Z}\right]^{-1} \\
\bar{\phi} &= \bar{\Phi}\left[\underline{\Phi}^{-1}\underline{\phi} + \psi_{22}^{-1}\mathbf{Z}'\mathbf{y}_2\right]
\end{aligned} \tag{7}$$

where $\mathbf{Z}$ is an $N \times (K_1 + K_2)$ matrix whose rows are $\mathbf{z}_i'$ and $\mathbf{y}_2$ is an $N \times 1$ vector whose elements are $x_i - \frac{\omega_{12}}{\omega_{11}}(y_i - \delta_0 - \mathbf{w}_i'\boldsymbol{\delta} - v_i) - \phi_0$.

Regarding the posterior distribution of the constant term δ_0, using as prior an improper uniform distribution and given $f(y_i, x_i|\boldsymbol{\Theta}) = f(y_i|x_i, \boldsymbol{\Theta})f(x_i|\boldsymbol{\Theta})$, we obtain

$$\begin{aligned}
\delta_0|\boldsymbol{\phi}, \boldsymbol{\delta}, \Omega, \phi_0, \mathbf{v}, Data &\sim \mathcal{N}(\bar{\delta}_0, \psi_{11}/N) \\
\bar{\delta}_0 = \frac{1}{N}\sum_{i=1}^{N}&\left\{y_i - \frac{\omega_{12}}{\omega_{22}}(x_i - \phi_0 - \mathbf{z}_i'\boldsymbol{\phi}) - \mathbf{w}_i'\boldsymbol{\delta} - v_i\right\}
\end{aligned} \tag{8}$$

In a similar way, using as prior an improper uniform distribution for ϕ_0, and the fact that $f(y_i, x_i|\boldsymbol{\Theta}) = f(x_i|y_i, \boldsymbol{\Theta})f(y_i|\boldsymbol{\Theta})$, we obtain

$$\begin{aligned}
\phi_0|\delta_0, \boldsymbol{\phi}, \boldsymbol{\delta}, \Omega, \mathbf{v}, Data &\sim \mathcal{N}(\bar{\phi}_0, \psi_{22}/N) \\
\bar{\phi}_0 = \frac{1}{N}\sum_{i=1}^{N}&\left\{x_i - \frac{\omega_{12}}{\omega_{11}}(y_i - \delta_0 - \mathbf{w}_i'\boldsymbol{\delta} - v_i) - \mathbf{z}_i'\boldsymbol{\phi}\right\}
\end{aligned} \tag{9}$$

As we mentioned, we just need to use the conditional prior distribution to obtain the posterior distribution of the spatial random effects. In particular, using the fact that $f(y_i, x_i|\boldsymbol{\Theta}) =$

$f(y_i|x_i,\boldsymbol{\Theta})f(x_i|\boldsymbol{\Theta})$, we find that

$$
\begin{aligned}
v_i|\mathbf{v}_{-i},\phi_0,\delta_0,\boldsymbol{\phi},\boldsymbol{\delta},\Omega,\sigma_v^2, Data &\sim \mathcal{N}(\bar{\xi}_i,\bar{\eta}_i)\\
\bar{\eta}_i &= \left[\psi_{11}^{-1} + \left(\frac{\sigma_v^2}{w_{i+}}\right)^{-1}\right]^{-1}\\
\bar{\xi}_i &= \bar{\eta}_i\left[\left(\frac{\sigma_v^2}{w_{i+}}\right)^{-1}\left(\sum_{i\sim j}\frac{w_{ij}}{w_{i+}}v_j\right) + \psi_{11}^{-1}v_i^0\right]
\end{aligned}
\tag{10}
$$

where $\mathbf{v}_{-i}$ is the set of spatial random effects excluding region i, $w_{i+} = \sum_{i\sim j} w_{ij}$ and $v_i^0 = y_i - \frac{\omega_{12}}{\omega_{22}}(x_i - \phi_0 - \mathbf{z}_i\boldsymbol{\phi}) - \delta_0 - \mathbf{w}_i\boldsymbol{\delta}$. To identify δ_0, we must add the constraint $\sum_{i=1}^N v_i = 0$. Therefore, this constraint must be imposed numerically by recentering each $\mathbf{v}$ vector around its own mean following each Gibbs iteration.

In addition,

$$
\begin{aligned}
\mathbf{v}|\phi_0,\delta_0,\boldsymbol{\phi},\boldsymbol{\delta},\Omega,\sigma_v^2, Data &\sim \mathcal{N}_N(\bar{\mathbf{v}},\bar{\mathbf{V}})\\
\bar{\mathbf{V}} &= \left[\psi_{11}^{-1}\mathbf{I}_N + \sigma_v^{-2}\left(\mathbf{D}_W - \mathbf{W}_N\right)\right]^{-1}\\
\bar{\mathbf{v}} &= \bar{\mathbf{V}}\left[\psi_{11}^{-1}\mathbf{v}^0\right]
\end{aligned}
$$

where $\mathbf{v}^0$ is an $N\times 1$ vector whose elements are v_i^0.

Finally,

$$
\begin{aligned}
1/\sigma_v^2|\mathbf{v} &\sim \mathcal{G}(\bar{\alpha},\bar{\beta})\\
\bar{\alpha} &= \frac{\underline{\omega}-1}{2} + N/2\\
\bar{\beta} &= 1/2 + \frac{\mathbf{v}'(\mathbf{D}_W - \mathbf{W}_N)\mathbf{v}}{2}
\end{aligned}
\tag{11}
$$

An unfortunate consequence of introducing spatial random effects is a reduction of the efficiency of MCMC sampling schemes. This in turn generates poor mixing and slow convergence (Best et al., 1999). To mitigate this problem, we draw multivariate blocks from distributions (5)–(11), whenever possible, using the Gibbs sampler algorithm (Geman and Geman, 1984) (other possible strategy is to use Acceptance-Rejection within a Direct Monte Carlo proposed by Zellner et al., 2014).

1.4 Simulation Exercises

In this subsection we present the results of a limited Monte Carlo experiment comparing Bayesian and Frequentist estimators applied to a very simple two-equation simultaneous model with spatial effects. The main objective of these exercises is to illustrate the consequences of omitting important factors of the estimators' data generating process for parameter estimates and prediction.

In particular, we implement five estimators: our proposed Bayesian Instrumental Variable with Spatial Effects, two instrumental variable approaches without spatial effects (one Bayesian and one Frequentist), a maximum likelihood estimator with conditional autoregressive spatial effects, and ordinary least squares.

Regarding the estimation of the endogenous Bayesian model with spatial effects, we implement the Gibbs sampler algorithm using one million iterations and a burn-in of 500,000. Then, we draw an observation every 50 iterations to have an effective sample size of 10,000. This last step is done to mitigate the autocorrelation of the chains. All the chains seem stable, and different diagnostics indicate that the chains converge to stationary distributions (outcomes available upon request). The same procedure was followed to implement the Bayesian instrumental model without spatial effects, except that in this case it was only necessary to iterate 100,000 times with a burn-in period of 50,000, drawing every 5 iterations, to achieve convergence.

The formulation of our model is

$$y_i = \delta_0 + \delta_1 x_i + v_i + u_{1i}$$

$$x_i = \phi_0 + \phi_1 z_{1i} + \phi_2 z_{2i} + u_{2i}$$

where $\mathbf{v} \sim \mathcal{N}_N(\mathbf{0}, \sigma_v^2(\mathbf{D}_W - \mathbf{W}_N)^{-1})$, $\begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\right)$, $z_{1i} \sim \mathcal{N}(0, \sigma_z^2)$ and $z_{2i} \sim \mathcal{N}(0, \sigma_z^2)$.

We generate four possible scenarios under the conditions summarized in Table 1. In all four scenarios the parameters δ_0, δ_1, ϕ_0 and ϕ_1 were set to 0.7, -1.2, 0.5 and 0.8, respectively. On the other hand, ϕ_2 was set to 0 for runs I and III and -1.0 for runs II and IV. This is to illustrate the consequences on just-identified and over-identified cases. In runs I to IV, the covariance structure of the stochastic shocks is the same as well as the spatial random effects, which are based on a rook binary contiguity criterion. In addition, the variance parameter of the CAR effect is such that 50% of the variance of the main interest variable y_i is explained by the spatial effect. Finally, the variance of the instruments is equal to 0.2 in scenarios I and II and 2 in scenarios III and IV. The main idea is to show the consequences associated with using weak instruments.

Observe that if $\sigma_{12} = 0$ and $\mathbf{v} = \mathbf{0}$, then there are no endogeneity or spatial effects, so we should use OLS, whereas if $\sigma_{12} = 0$ but $\mathbf{v} \neq \mathbf{0}$, the MLE with CAR effects is appropriate. In addition, $\sigma_{12} \neq 0$ and $\mathbf{v} = \mathbf{0}$ requires taking into account endogeneity issues without spatial effects, so IV estimators (Frequentist and Bayesian) are good alternatives. Finally, an IV with spatial effects is required if both σ_{12} and $\mathbf{v}$ are not equal to zero. Therefore, we implement different estimators designed to take into account different nested models.

For all scenarios, we generate samples of size 49, 100 and 144, and 100 repeated trials to assess the estimators' performance, which is the common approach employed in the Frequentist framework, although this methodology is not the most consistent with the Bayesian statistical framework (Zellner, 1996).

We present in Table 2 the Mean Squared Error and the Mean Absolute Error to assess the performance of the point estimators of δ_1. To calculate both measures, we use the median estimates of the Bayesian procedures, and the point estimates of the Frequentist approaches. The main characteristic that we found in this table is that Bayesian estimates obtain the lowest MSE and MAE. The general pattern is that the Bayesian Instrumental Variable with Spatial Effects has the lowest errors in presence of strong instruments using small and large sample sizes, and when there are weak instruments, our proposal obtains the best outcomes using large sample sizes, whereas the Bayesian Instrumental Variable gets the lowest errors using small sample sizes. In addition, it is remarkable that the Frequentist Instrumental Variable estimator has by far the highest errors using weak instruments, in both exactly and over identified cases. We see similar outcomes when comparing ML CAR and OLS. In general, we observe that errors are lower using strong instruments compared to weak instruments.

To assess the forecasting performance of the estimators, we use the Mean Squared Prediction Error and the Mean Absolute Prediction Error. As can be seen in Table 3, the Bayesian Instrumental Variable with Spatial Effects obtains by far the lowest MSPE and MAPE, followed by the maximum likelihood estimator with CAR effects. Taking into account spatial effects substantially improves the prediction performance (Reich et al., 2006).

As a second exercise, we obtain the same measures of performance as before under different data generating process conditions. We will maintain both weak and strong instruments, exactly and over-identified models, as well as all five possible estimation alternatives. This time, however, we will assume that the sample size is fixed at $N = 100$ and we will compare three different specification problems than can arise in applied work: non-normality in the joint error distribution, a misspecified contiguity matrix, and a misspecified spatial effects distribution. In order to treat each specification issue in our simulation exercise, we generate the data and then estimate the parameters under different settings. First, we draw the errors simultaneously from a multivariate-t distribution with three degrees of freedom, but assume normality at the estimation stage. Second, we simulate the data with a contiguity matrix describing the spatial structure using a queen criterion but estimate it assuming a rook criterion matrix. Finally, we generate the spatial effect term from a SAR process but estimate it as a CAR.

Table 4 presents the results associated with errors in the estimation of the parameter of interest. As expected, errors for any configuration increase when using over-identified models but decrease in the presence of strong instruments. Additionally, having a misspecified contiguity matrix appears to increase errors the most in any setting. It can be seen that our Bayesian IV CAR specification continues to perform very well against other alternatives. Our approach ranks first in any possible scenario when the normal distribution is replaced by a multivariate t-distribution. In the case of a misspecification in the contiguity criterion, rook rather than queen, our Bayesian IV CAR procedure ranks first using strong instruments, although all IV procedures, including ours, are the worst using weak instruments. Our Bayesian IV CAR appears to deal well when disturbances are spatially autocorrelated through a SAR structure instead of the assumed CAR distribution. In particular, IV procedures (Bayesian and Frequentist) rank first in the presence of exactly identified conditions followed by our proposal, whereas the former procedures are the worst in the case of over identified conditions, and our Bayesian IV CAR ranks first in the case of strong instruments.

Table 5 deals with the prediction errors for the same configurations previously mentioned. Overall, we have that our Bayesian IV CAR specification obtains the best outcomes when there is a misspecification in the spatial process, although the ML CAR obtains the best outcomes in the case of misspecification in the disturbances distribution and contiguity criteria followed most of the times by our specification.

2 The Colombian energy market

To better understand our application and its microeconomic foundation, there are some characteristics of the Colombian electricity market that must be taken into consideration. In particular, we must explain the price changes, and thus, their welfare implications. First, Colombian law divides its population into socioeconomic strata. This segmentation is defined as "an instrument that allows a municipality or district to classify its population in distinct groups or strata with similar social and economic characteristics." (Bushnell and Hudson, 1996). This classification was actually initiated to establish cross-subsidies that would help the lower socioeconomic classes to pay for utilities such as electricity. Housing characteristics are the main criteria used for classifying the population into six strata: one represents lower-low, two is low, three is upper-low, four is medium, five is medium-high, and six is high. Second, the Colombian energy regulator establishes a subsistence electricity consumption that is subsidized for strata one, two and three. The regulator determines the maximum subsidy percentage, and each municipality defines its own measure within this limit. In addition, the subsistence consumption level depends on whether the altitude of the municipality exceeds one thousand meters above sea level or not, due to weather conditions that may affect electricity consumption. Municipalities located near the sea level have higher temperatures, and as a consequence they present a higher electricity consumption. Specifically, the subsistence consumption is 173 kWh a month per household for the municipalities below this threshold and 130 kWh for the ones above it.[1] Third, the Colombian energy regulator stipulates that each electric company must have the same reference tariff throughout its entire market, which involves many municipalities. Fourth, there are basically four components to establish the reference electricity tariff for each company: electricity generation, transport at the country level, distribution at the market level, and commercialization. As a consequence of this regulation framework, we should bear in mind that although there is just one reference tariff for each electric company, there are different average electricity prices among the consumers of different strata and municipalities.

The acquisition of EADE by EPM led to a tariff unification process that generated welfare effects, especially for the households belonging to stratum one. Such a household's electricity consumption is approximately 5% of its income in the province of Antioquia, whereas this share is less than 1% for stratum six. In particular, there is EPM, whose market was characterized as an urban region with high population density, and on the other hand there is EADE, whose market was a rural area with low population density. Under the Colombian electricity regulation framework, *ceteris paribus*, these market structure differences imply a higher reference tariff in the latter company than in the former. This is because of the third and fourth components of the reference tariff: distribution at the market level and commercialization. Thus, the acquisition of EADE by EPM implied that the stratum one electricity consumers of the former company experienced a huge decrease in their electric bills, while the consumers of the latter company faced a slight increase.[2] As a consequence, these changes generated considerable welfare impacts on the poorest inhabitants of the province of Antioquia, who live in the rural areas.

[1] Resolution 0355 of the Mining/Energy Planning Unit (UPME).

[2] Regulations stipulate that strata one and two cannot have a tariff increase higher than the inflation rate.

3 Microeconomic Foundations: Equivalent Variation

We apply our methodology to analyze the welfare changes arising from the tariff unification in the municipalities of Antioquia using an Equivalent Variation (EV) approach. The Equivalent Variation measures the "amount that the consumer would be indifferent about accepting in lieu of the price change; that is, it is the change in her wealth that would be equivalent to the price change in terms of its welfare impact" (Mas-Colell et al., 1995). The EV presents several advantages over other welfare measures used in applied economic work, such as the Compensating Variation (CV) and consumer surplus (CS). In particular, Chipman and Moore (1980) and Mas-Colell et al. (1995) show that the EV is the appropriate measure to correctly order different pricing policies in welfare analysis. The CV orders alternatives correctly only when consumers exhibit homothetic preferences and income remains unchanged. However, in our empirical application, tariff unification translates into implied subsidies for some consumers, and therefore, changes in income. Another argument in favor of the EV is that, by definition, it is an ex-post measure of welfare based on the Hicksian demand function. It takes into account income effects associated with price changes, which are ignored by the Marshallian demand function on which the CS is based on. Furthermore, Hausman (1981) showed that it was possible to derive EV as a product of observable Marshallian demand functions. His method can be applied to the case of linear budget constraints, and Ruijs (2009) extends it to the case of budget constraints generated by block-pricing systems using linear demand functions.

To build our application, we consider the two-good case in which a representative agent consumes a good x, say electricity, and an aggregate good as a numeraire (x_a). We note that, for our application, the representative agent assumption is not as restrictive as it may seem. In particular, given that we work with the stratum one population at the municipality level, a fairly homogeneous group within each polygon, the assumption that agents with similar preferences can be aggregated into a single agent per municipality is not unthinkable. This could be thought of as a case of dispersion in preferences and income where, although individuals may present erratic utility functions, the aggregate demand for the commodities of interest are well-behaved (Trockel, 1987). In addition, representative agent models dominate microfounded macroeconomics due to their simplicity and tractability (Acemoglu, 2008). One final argument is the impossibility of obtaining data at the micro-level to correct for the bias raised by agent heterogeneity. Therefore, although we are aware of the disadvantages of the representative agent (Kirman, 1992, Reiss and White, 2006), we will continue to work under this assumption.

Throughout this paper we will indicate a situation before or after tariff unification with the subscripts 0 or 1, respectively. Subsistence consumption will be denoted by $\bar{x}$. This subsistence consumption divides demand into two possible tiers, denoted by superscript 1 when the consumer demands a quantity less than $\bar{x}$ and 2 when it is greater than $\bar{x}$. Call x_1 the new demand at prices $\mathbf{p}_1$ and expenditure $e(\mathbf{p}_1, u_1) = y_0$. Tangency between u_1 and the budget curve characterized by $\mathbf{p}_0 = (p_0^1, p_0^2, 1)$ and expenditure $e(\mathbf{p}_0, u_1)$ is referred to as *virtual* consumption (x_e, see Figure 1).

Using the following demand function,

$$x(p, y) = p^{\alpha} y^{\delta_1} e^{z'\boldsymbol{\delta}} \tag{12}$$

The Equivalent Variation associated with the first block is

$$EV(\mathbf{p}_0, \mathbf{p}_1, y_0) = \left[\frac{1-\delta_1}{1+\alpha} \left(p_0^{1(1+\alpha)} - p_1^{1(1+\alpha)} \right) e^{z'\boldsymbol{\delta}} + y_0^{1-\delta_1} \right]^{\frac{1}{1-\delta_1}} - y_0 \tag{13}$$

For the second block, the Equivalent Variation is

$$\begin{aligned} EV(\mathbf{p}_0, \mathbf{p}_1, y_0) = {} & \left[\frac{1-\delta_1}{1+\alpha} \left(p_0^{2(1+\alpha)} - p_1^{2(1+\alpha)} \right) e^{z'\boldsymbol{\delta}} + \left(y_0 + (p_1^2 - p_1^1)\bar{x} \right)^{1-\delta_1} \right]^{\frac{1}{1-\delta_1}} - \\ & (p_0^2 - p_0^1)\bar{x} - y_0 \end{aligned} \tag{14}$$

We can see from Equations (13) and (14) that a price decrease of an inelastic necessity good produces welfare gains that can be quantified through a positive Equivalent Variation. Equation (14) takes into consideration that the subsidy has effects on the expenditure function as well as on the agent's income.

4 Results

4.1 Data

Data was collected for the average individual of stratum one in all 125 municipalities of the department of Antioquia (Colombia) in 2005. Table A.1 in the Appendix lists all the variables, their measurement units, and sources. We standardized both electricity and substitute good prices to US$/kWh by taking their calorific power into account. For municipalities in the Metropolitan Area, the substitute good was natural gas. For the other municipalities, it was liquefied petroleum gas due to absence of natural gas. In addition, we have to mention that by construction, the average price is affected by electricity consumption because the average electricity tariff is a weighted average between the tariffs in the first and second tiers, where the weights depend on the observed and subsistence consumption levels. This generates the endogeneity problem in our application.

We present descriptive statistics in Table 6. The average annual electricity consumption is 234.87 kWh with a standard deviation of 117.81 kWh. The prices for electricity and the substitute good averaged 6.10 and 3.00 cents a kWh, respectively. Additionally, the average annual per capita income was US$397, with a standard deviation of US$95.24. Approximately 29% of the municipalities in the province of Antioquia are located less than 1000 meters above sea level, the average urbanization rate is 45.8%, and 77.4% of the municipalities used to be covered by EADE prior to its acquisition by EPM.

We can observe the geographical distribution of the electricity consumption in Map 2. In particular, the average consumption of electricity tends to be higher in regions that are located less than one thousand meters above sea level (the Northern and Eastern regions). Consumption is also exceptionally high in the Metropolitan Area of Antioquia (South-Central region), where most of the population and economic activity of the province are focused.

Map 3 shows that most of the spatial autocorrelation is due to local clusters. This obeys unobserved social, cultural, economical and geographical restrictions, like the limited and bad roads between municipalities or constrained budgets that poor households face in these municipalities. Avoiding a global spatial effect regarding electricity consumption for the inhabitants of the province appears to be the most natural approach, as is provided by the CAR specification.

4.2 Model Specification

We need to estimate the electricity demand function to perform statistical inference of the Equivalent Variation. However, it is necessary to take into account the endogeneity issue between price and consumption to avoid biased and inconsistent parameter estimates. It is also necessary to introduce spatial effects in order to have a good municipality electricity prediction. We realize that both elements are crucial to obtaining a reliable Equivalent Variation measure in light of Equations (13) and (14).

As instrument, we use a dummy variable that is equal to 1 if the municipality was serviced by EADE, and 0 otherwise. The argument behind this instrument is that the national electricity regulations generate restrictions that imply that the only effect of the electricity supplier on average consumption in each municipality is through price. However, the regional market reference tariff, and as a consequence the average electricity price of the low strata in each municipality, is drastically affected by each supplier.

The structural specification of our system is

$$\ln x_i = \delta_0 + \mathbf{z}'_{1i}\boldsymbol{\delta}_1 + \alpha \ln p_i + u_{1i} + v_i \tag{15}$$

$$\ln p_i = \phi_0 + \mathbf{z}'_{1i}\boldsymbol{\phi}_1 + \alpha_s z_{2i} + u_{2i} \tag{16}$$

where x_i and p_i are the electricity consumption and price, $\mathbf{z}'_{1i} = (\ln y_i, \ln p_i^s, alt_i, \ln urb_i)$ is a vector of exogenous covariates that affects the system (income, substitute price, sea level dummy, and urbanization rate) and $z_{2i} = EADE_i$ is our instrument. Additionally, δ_0, $\boldsymbol{\delta}'_1 = (\delta_1, \delta_2, \delta_3, \delta_4)$, ϕ_0, $\boldsymbol{\phi}'_1 = (\phi_1, \phi_2, \phi_3, \phi_4)$, α and α_s are parameters to be estimated. Finally, u_{1i} and u_{2i} are the idiosyncratic error terms associated with the demand and price of each municipality, and v_i are spatial random effects to control for spatial heterogeneity and spatial dependence between cross-sectional units that is present in our application (see Map 3). These emerge due to clusters and/or spillover effects between neighboring municipalities, and allow us to control for unobservable spatial heterogeneity. Omitting this last component can cause a loss of good statistical properties of estimators (Anselin, 1988).

We find just one available instrument, however, we can use this situation to illustrate that our econometric framework encompasses the more simple technique of multivariate regression models when there is an exactly identified system (Zellner et al., 2014, warn about using improper priors in exactly identified systems). In particular, we estimate the reduced model that results from substituting (16) into (15) in our application.

$$\begin{aligned}\ln x_i &= \pi_0 + \pi_1 \ln y_i + \pi_2 \ln p_i^s + \pi_3 alt_i + \pi_4 \ln urb_i + \gamma EADE_i + \mu_{1i} + v_i \\ \ln p_i &= \phi_0 + \phi_1 \ln y_i + \phi_2 \ln p_i^s + \phi_3 alt_i + \phi_4 \ln urb_i + \alpha_s EADE_i + \mu_{2i}\end{aligned} \tag{17}$$

where $\mu_{1i} = u_{1i} + \alpha u_{2i}$ and $\mu_{2i} = u_{2i}$, such that $(\mu_{1i}, \mu_{2i})' \sim \mathcal{N}(\mathbf{0}, \Sigma)$, $\Sigma = \{\sigma_{ij}\}$. The structural parameters can be recovered using $\alpha = \gamma/\alpha_s$ and $\delta_l = \pi_l - \phi_l \alpha$, $l = \{0, 1, \ldots, 4\}$.

Setting $\mathbf{z}_i' = (\ln y_i, \ln p_i^s, alt_i, \ln urb_i, EADE_i)$, $\boldsymbol{\pi}' = (\pi_1, \pi_2, \pi_3, \pi_4, \gamma)$, $\boldsymbol{\phi}' = (\phi_1, \phi_2, \phi_3, \phi_4, \alpha_s)$ and $\mathbf{v}' = (v_1, v_2, \ldots, v_n)$, and taking into consideration that the determinant of the Jacobian matrix of the transformation is 1, the likelihood function of the system is

$$f(\ln \mathbf{x}, \ln \mathbf{p} | \mathbf{z}; \Sigma, \boldsymbol{\pi}, \boldsymbol{\phi}, \pi_0, \phi_0, \mathbf{v}) = \frac{|\Sigma|^{-N/2}}{(2\pi)^{N/2}} \times \tag{18}$$
$$\exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} (\ln x_i - \mathbf{z}_i'\boldsymbol{\pi} - \pi_0 - v_i, \ln p_i - \mathbf{z}_i'\boldsymbol{\phi} - \phi_0) \Sigma^{-1} \begin{pmatrix} \ln x_i - \mathbf{z}_i'\boldsymbol{\pi} - \pi_0 - v_i \\ \ln p_i - \mathbf{z}_i'\boldsymbol{\phi} - \phi_0 \end{pmatrix} \right\}$$

We can estimate our reduced form model with spatial random effects using this likelihood, and prior independent distributions such that

$$\pi(\Sigma, \boldsymbol{\pi}, \boldsymbol{\phi}, \pi_0, \phi_0, \mathbf{v}_i, \sigma_v^2) = \pi(\Sigma)\pi(\boldsymbol{\pi})\pi(\boldsymbol{\phi})\pi(\pi_0)\pi(\phi_0)\pi(\mathbf{v}_i|\sigma_v^2)\pi(\sigma_v^2) \tag{19}$$

where $\Sigma^{-1} \sim \mathcal{W}_2(3, \mathbf{I}_2)$, $\boldsymbol{\pi} \sim \mathcal{N}_5(\mathbf{0}, 1000\mathbf{I}_5)$, $\boldsymbol{\phi} \sim \mathcal{N}_5(\mathbf{0}, 1000\mathbf{I}_5)$, $\pi(\pi_0) \propto 1$, $\pi(\phi_0) \propto 1$,

$v_i|\mathbf{v}_{i\sim j} \sim \mathcal{N}\left(\sum_{i\sim j}\frac{w_{ij}v_j}{\sum_{i\sim j}w_{ij}}, \frac{\sigma_v^2}{\sum_{i\sim j}w_{ij}}\right)$ and $1/\sigma_v^2 \sim \mathcal{G}(1, 1/2)$. As we can see, most of the priors are vague or diffuse, except the Gamma distribution, which was set to reflect the prior belief that 50% of the variability of the reduced model are subject to spatial effects. However, we find in our application that the posterior parameter estimates are robust to changes in the hyperparameters of the CAR's precision (available upon request).

We should bear in mind that the posterior framework that was deduced in Section 1 applies to our empirical exercise. We just need to take into consideration that in this subsection, $\mathbf{w}_i = \mathbf{z}_i$ and we treat $\boldsymbol{\pi}$ in a similar way as $\boldsymbol{\delta}$.

We must bear in mind two aspects that are important in our application. First, a just-identified model allows $\boldsymbol{\pi}$ to have independent variation, a situation that is not possible with over-identified models in reduced form. Second, under independent prior distributions, we can express the posterior as $\pi(\boldsymbol{\delta}|\boldsymbol{\pi}, Data) \propto \pi(\boldsymbol{\delta}|\boldsymbol{\pi})$, that is, the conditional posterior of the structural parameters is unaffected by the observations once the reduced form parameters are taken into account (Zellner, 1996).

4.3 Estimation Results

Since the solution for the model depends on the selection of the contiguity matrix, we will test our specification under three different matrices. The first one uses the road lengths between each municipality, regarding two regions to be neighbors if the roads connecting them are less than 300 kilometers long, which ensures that each region has at least one neighbor. The second uses the queen criterion, where two regions are considered as neighbors if they share at least a single border point. The third one uses the rook criterion, where regions are considered neighbors if they share more than one border point.

We estimate each of our models using Markov chain Monte Carlo techniques (MCMC, see Robert and Casella, 2004, for details). In particular, we use the Gibbs sampling algorithm, due to the availability of all the conditional posterior distributions (Geman and Geman, 1984). After running the chains for 10 million iterations, we discard the first 5 million and draw an observation every 500 iterations to get an effective sample size of 10,000. We compute several diagnostics to assess the convergence and stationarity of the chains. In particular, we employ the method due to Heidelberger and Welch (1983), the mean difference test proposed by Geweke (1992), and the diagnostic from Raftery et al. (1992). We show that in general all the chains under different contiguity criteria achieve convergence and stationarity in the Table B.1 in the Appendix, subsection B.[3]

The correlation of the instrument with the logarithm of the price is approximately -0.36, its variability is very low due to its being a dummy variable, its standard deviation is equal to 0.42,[4] and its 95% probability highest density interval in the price equation is $(-0.49, -0.23)$. However, even with this weak instrument, the Bayesian approach works well in this context using proper priors due to the fact that the likelihood function and its identification are less important for deriving estimates in Bayesian models (Zellner, 1996, Imbens and Rubin, 1997, Zellner, 1998, Crespo-Tenorio and Montgomery, 2013).

We present in Table 7 the main outcomes associated with the structural parameters of our estimation exercises, using different contiguity criteria to check their robustness. As we can see, all contiguity criteria yield similar results. In particular, we present the posterior mean and median that minimize the quadratic and absolute value loss functions under a decision theory framework. In addition, in order to describe the inferential content of the posterior distributions of the parameters, we present the 95% highest probability density credible intervals for each parameter of interest. Finally, to test whether the microeconomic restrictions are compatible with the observed data, we calculate the odds ratio in favor of the null hypothesis $H_0 : \theta \in (0, \infty)$ versus $H_1 : \theta \in (-\infty, 0]$, using 0.5 as the prior probability for each of these hypotheses. This procedure is consistent with a symmetric loss function, for instance a zero–one loss function (Berger, 1985, Zellner, 1996). Testing microeconomic restrictions is very important in this setting because our main objective is to carry out statistical inference regarding Equivalent Variation, and so there are some implicit restrictions placed on the parameter estimates. Thus, we follow a statistical decision theory framework, where an action regarding the domain of the posterior densities must be made. These actions are based on

[3]All the simulation exercises and posterior analyses were performed using R (R Core Team, 2014).

[4]As this instrument is a Bernoulli random variable, its possible maximum standard deviation is equal to 0.5. This would be a limitation of using dummy variables as instruments in a Frequentist approach.

prior and sample information. Kleit and Terrell (2001) reiterate the importance of placing restrictions on Bayesian models and priors based on microeconomic theory.

Regarding endogeneity in our application, we find that the posterior median estimates of σ_{12} are approximately -0.06 using different contiguity criteria, and the highest probability interval at 95% of credibility is $(-0.096, -0.040)$. This evidence suggests that there is endogeneity between electricity consumption and price.

Given that we obtain robust outcomes regarding the contiguity criteria, we discuss the results associated with the road length criterion. This criterion better illustrates the connectivity between municipalities in a province that is characterized by irregular geographical conditions and poor roads. Thus, when we observe the posterior mean and median, we see that all the point estimates have the expected signs. Electricity behaves as both an ordinary and a normal good, given the negative price-demand elasticity and positive income-demand elasticity. For instance, the average as well as the median price demand elasticity is -0.88, that is, an increase of 1% in the price of electricity implies a reduction of 0.88% in consumption. In addition, the average and median income elasticity is approximately 0.30, which implies that a 1% income increase means a 0.30% increase in electricity consumption. Regarding the 95% highest probability density credible intervals of these parameters, we have that these are $(-1.58, -0.13)$ and $(-0.10, 0.73)$ for the price elasticity and income elasticity, respectively. In addition, we calculate the inverse odds ratio in favor of the null hypothesis $H_0 : \alpha \in (-\infty, 0]$ to check the microeconomic restriction of a negative price elasticity. This is equal to 0.014, which means an odds ratio supporting H_0 equal to 71.42, which implies that $\log_{10}(R_{01}) = 1.85$. Thus, we have very strong evidence for H_0 following Jeffreys's guidelines (Greenberg, 2008). Regarding the null hypothesis of a positive income elasticity, $H_0 : \delta_1 \in (0, \infty)$, which is suggested by most of the literature on electricity demand (Hsiao and Mountain, 1985, Dergiades and Tsoulfidis, 2008), we have $\log_{10}(R_{01}) = 1.07$, indicating strong evidence for H_0.

Regarding cross elasticity with the substitute good, a 1% increase in the price of the substitute implies a 0.12% increase in electricity demand. Although this is positive on average, there is weak evidence for $H_0 : \delta_2 \in (0, \infty)$ due to the fact that $\log_{10}(R_{01}) = 0.40$. This result probably holds because of the lack of electricity substitutes in rural areas, or the fact that the demand for electricity is derived for most household appliances which cannot function with anything but electricity. For this parameter, we observe a HPD credible interval between -0.29 and 0.52. The mean altitude semi-elasticity is equal to 0.14, which means that municipalities located at lower altitude demand approximately 14% more electricity, *ceteris paribus*. In this case, we have $\log_{10}(R_{01}) = 0.97$, which is substantial evidence for $H_0 : \delta_3 \in (0, \infty)$. Finally, there is the urbanization rate, which has a strong positive effect on electricity consumption, as one would expect: $\log_{10}(R_{01}) = 2.96$ for $H_0 : \delta_4 \in (0, \infty)$, which is decisive support for H_0. The median and mean urbanization rate elasticity is approximately 0.57 with a highest probability density credible interval equal to $(0.38, 0.76)$.

Despite the fact that our prior assumption regarding the participation of the spatial effects on electricity consumption variability is 50%, we find that the posterior mean proportion is 4% with a standard deviation equal to 9.45% and the HPD at 95% equal to (0.008%, 26.610%). This outcome is robust to hyperparameter combinations of the prior distribution of the precision parameter of the CAR component (available upon request).

Before concluding this section, we present one final table, that compares the results of our Bayesian methodology with those obtained by Frequentist IV and OLS. As we see in Table 8, estimates from our Bayesian methodology that take into account spatial effects in electricity consumption are very close to those where there is no spatial autocorrelation. While the estimates for regressors without endogeneity problems in OLS are also very similar to the IV methods, the coefficient for the endogenous price variable is overestimated in absolute terms. This magnitude for the price coefficient would characterize electricity as an elastic good in the long-run for low income households, which does not seem likely. Regarding interval estimation, we note that the credible intervals for exogenous variables are a little wider than their confidence counterparts. For the endogenous regressor, however, the credible interval at 95% probability is significantly narrower than the 95% confidence interval. Furthermore, while both bounds for the credible interval are negative, the confidence interval includes zero in the case of IV estimation, which signifies that the price coefficient is not statistically significant at a 5% level.

4.4 Welfare Implications

The tariff unification procedure brought about by the acquisition of EADE by EPM created tier price variations that depended on whether the municipality was part of the Metropolitan Area or not. In particular, by the end of this process, p_1^1 and p_1^2 changed according to the values in Table 9 with respect to their pre-unification values. We expect to see that the municipalities which consumed less than the subsistence consumption and are not part of the Metropolitan Area have the largest welfare gains, followed by those that are not part of the Metropolitan Area and had average consumption higher than the subsistence consumption. The welfare effects in the municipalities that belong to the Metropolitan Area are not clear and will depend on whether they consumed more than the subsistence consumption or not, and how much of their consumption was above this threshold, among other factors (Ramírez and Londoño, 2009). Here, we note that subsistence consumption is measured in kilowatts/hour a month per household. Therefore, in order to make it comparable with our measure of income, we work with an annual per capita consumption for each altitude.[5]

To compute the posterior distribution of the Equivalent Variation, we follow the guidelines of the Bayes theorem, and renormalize the unrestricted posterior distribution of each parameter according to the outcomes of the microeconomic restrictions in Table 7, where the statistical evidence suggests the fulfillment of those restrictions (Berger, 1985, Bernardo and Smith, 1994). This allows us to obtain sensible results, based on a statistical decision theory framework, regarding the Equivalent Variation, which is calculated for each municipality at each observation of these new chains, through Equations (13) and (14). This procedure leaves an effective sample size of 5,410 with which to make the computations.

Table 10 lists the mean, median, and 95% highest probability density interval for the total, for the Metropolitan Area, and for the rest of Antioquia, as a share of original income y_0.

[5] The original levels were multiplied by 12 to obtain an annual measure and then divided by an average of 4.04 people per household in stratum one to get our variable of interest.

As can be observed, the median Equivalent Variation in the whole province is approximately 0.63%, and its standard error is approximately 0.67%. This welfare gain is lower in the Metropolitan Area (0.14%) and greater in the rest of the province (0.65%). The 95% HPD interval is equal to (0.0%, 2.3%), and the posteriors tend to be skewed to the right, as the mean is greater than the median. For stratum one households, this impact can be very substantial, especially for those who are located in regions other than the Metropolitan Area of Antioquia.

Map 4 presents the median Equivalent Variation as a share of income in the province. The spatial distribution is low around the Metropolitan Area (South-Central region) and high in the more rural areas, especially Eastern region, which received the greatest improvement and benefits from the tariff unification.

5 Concluding Remarks

In this paper, we introduced spatial random effects into an endogenous Bayesian framework with simultaneous equations and deduced the complete conditional posterior distributions. Thus, we were able to draw observations from the model using a Gibbs sampler algorithm. This approach allows dealing simultaneously with three shortcomings, which would be quite difficult to manage simultaneously with a Frequentist approach. First, it permits taking into account the endogeneity issues in our estimation procedure. Second, we can carry out statistical inferences of complicated non-linear functions of the parameter estimates in our application. Third, it allows controlling for the non-observable heterogeneity and spatial autocorrelation present in cross-sectional data.

We performed simple Monte Carlo simulation exercises which show that our econometric approach handles the endogeneity and spatial effects well. The posterior point estimates are sensible, and the prediction is significantly improved by introducing the spatial effects.

Using these features of the Bayesian framework to our advantage, we estimated the Equivalent Variation welfare measure, as a share of mean income, that stemmed from a process of electricity tariff unification in the province of Antioquia (Colombia), with data at the municipality level. We estimated a demand function for electricity and found the average price, income, substitute, and urbanization rate demand elasticities to be, respectively, -0.88, 0.30, 0.12, and 0.57. The semi-elasticity associated with a dummy for the altitude of the municipalities was approximately 0.14. Using this information as input, we found the Equivalent Variation for the province as a whole to be 0.87% on average and with a median of 0.63%. When taking into account the welfare gains of the municipalities of the Metropolitan Area, these amount only to 0.13%. However, the municipalities that are not part of the Metropolitan Area gained on average 0.94%, while the 10% of the municipalities with the least urbanization and least income increased their welfare by an amount well above 2% of their initial income. Comparing these figures with the the amount that low income households expend on pensions (1.13%), health care (2.04%), and education (4.79%) illustrates the huge effect of electricity regulation on the welfare of the poor.

References

Acemoglu, D. (2008). *Introduction to Modern Economic Growth.* Princeton University Press.

Acton, J. P. and Mitchell, B. M. (1983). *Welfare analysis of electricity rate changes.* RAND Corporation.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Springer-Verlag.

Anselin, L. (1990). Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics*, 20(2):141–163.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):153–166.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall / CRC Press.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag.

Bernardinelli, L., Clayton, D., and Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, 14(21-22):2411–2431.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory.* John Wiley & Sons Ltd.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Best, N. G., Waller, L. A., Thomas, A., Conlon, E. M., and Arnold, R. A. (1999). Bayesian models for spatially correlated diseases and exposure data. In Bernardo, J. M., editor, *Bayesian Statistics*, pages 131–156. Oxford University Press.

Bushnell, D. and Hudson, R. A. (1996). The society and its enviroment. In *Colombia: A Country Study*, pages 63–139. Government Printing Office, Washington D.C.

Casella, G. and Berger, R. L. (2002). *Statistical Inference.* Thompson Publishing.

Chakraborty, A., Beamonte, M., Gelfand, A. E., Alonso, M., Gargallo, P., and Salvador, M. (2013). Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets. *Computational Statistics & Data Analysis*, 58:292–307.

Chipman, J. S. and Moore, J. C. (1980). Compensating variation, consumer's surplus, and welfare. *American Economic Review*, 70(5):933–49.

Crespo-Tenorio, A. and Montgomery, J. (2013). A Bayesian approach to inference with instrumental variables—Improving estimation of treatment effects with weak instruments and small samples. Technical report, Washington University, St. Louis.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* revised ed. Wiley.

DANE (2015). Index numbers and expenditure weights for low income households, 2009–2015. Technical report, Departamento Administrativo Nacional de Estadística, Colombia.

Darmofal, D. (2009). Bayesian spatial survival models for political event processes. *American Journal of Political Science*, 53(1):241–257.

Dergiades, T. and Tsoulfidis, L. (2008). Estimating Residential Demand for Electricity in the United States, 1965–2006. *Energy Economics*, Vol. 30:2722–2730.

Dodonov, B., Opitz, P., and Pfaffenberger, W. (2004). How much do electricity tariff increases in Ukraine hurt the poor? *Energy Policy*, 32(7):855–863.

Drukker, D. M., Egger, P., and Prucha, I. R. (2013). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews*, 32(5-6):686–733.

Drèze, J. H. (1976). Bayesian limited information analysis of the simultaneous equations model. *Econometrica*, 44(5):1045–1075.

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer-Verlag.

Fingleton, B. and Le Gallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science*, 87(3):319–339.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, pages 169–193. Oxford University Press.

Gomez-Lobo, A. (1996). The welfare consequences of tariff rebalancing in the domestic gas market. *Fiscal Studies*, 17(4):49–65.

Greenberg, E. (2008). *Introduction to Bayesian Econometrics*. Cambridge University Press.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. *American Economic Review*, 71(4):662–676.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144.

Hoogerheide, L., Kleibergen, F., and Van Dijk, H. K. (2007). Natural conjugate priors for the instrumental variables regression model applied to Angrist-Krueger data. *Journal of Econometrics*, 138:63–103.

Hsiao, C. and Mountain, D. (1985). Estimating the short-run income elasticity of demand for electricity by using cross-sectional categorized data. *Journal of the American Statistical Association*, 80(390):259–265.

Imbens, G. and Rubin, D. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25(1):305–327.

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. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2):509–533.

Kelejian, H. H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, 118(1):27–50.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *The Journal of Economic Perspectives*, 6(2):117–136.

Kleibergen, F. and Van Dijk, H. K. (1998). Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory*, 14(0):701–743.

Kleibergen, F. and Zivot, E. (2003). Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics*, 114(1):29–72.

Kleit, A. N. and Terrell, D. (2001). Measuring potential efficiency gains from deregulation of electricity generation: A Bayesian approach. *Review of Economics and Statistics*, 83(3):523–530.

Lahiri, S., Zhu, J., et al. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *The Annals of Statistics*, 34(4):1774–1813.

LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall / CRC Press.

LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20(1-2):113–129.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32(1):19–35.

LeSage, J. P., Fischer, M. M., and Scherngell, T. (2007). Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science*, 86(3):393–421.

LeSage, J. P. and Llano, C. (2013). A spatial interaction model with spatially structured origin and destination effects. *Journal of Geographical Systems*, 15(3):265–289.

Liu, X. and Lee, L.-F. (2013). Two-stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments. *Econometric Reviews*, 32(5-6):734–753.

Lundgren, J. (2009). Consumer welfare in the deregulated Swedish electricity market. *Frontiers in Finance and Economics*, 6(2):101–119.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

Ohtsuka, Y., Oga, T., and Kakamu, K. (2010). Forecasting electricity demand in Japan: A Bayesian spatial autoregressive ARMA approach. *Computational Statistics & Data Analysis*, 54(11):2721–2735.

Parent, O. and LeSage, J. P. (2008). Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers. *Journal of Applied Econometrics*, 23(2):235–256.

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

Raftery, A. E., Lewis, S., et al. (1992). How many iterations in the Gibbs sampler? *Bayesian Statistics*, 4(2):763–773.

Ramírez, A. and Londoño, E. (2009). Implicaciones de bienestar en el sector residencial de la unificación tarifaria en el servicio de electricidad en el departamento de Antioquia. *Ecos de Economía*, 13(28):7–52.

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects disease-mapping models. *Biometrics*, 62(4):1197–1206.

Reiss, P. C. and White, M. W. (2006). Evaluating welfare with nonlinear prices. Technical Report 12370, National Bureau of Economic Research.

Rey, S. J. and Boarnet, M. G. (2004). A taxonomy of spatial econometric models for simultaneous equations systems. In *Advances in Spatial Econometrics*, pages 99–119. Springer-Verlag.

Ripley, B. D. (2005). *Spatial Statistics.* Wiley.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer-Verlag.

Ruijs, A. (2009). Welfare and distribution effects of water pricing policies. *Environmental and Resource Economics*, 43(2):161–182.

Seya, H., Tsutsumi, M., and Yamagata, Y. (2012). Income convergence in Japan: A Bayesian spatial Durbin model approach. *Economic Modelling*, 29(1):60–71.

Smith, T. E. and LeSage, J. P. (2004). A Bayesian probit model with spatial dependencies. *Advances in Econometrics*, 18:127–160.

Sun, D., Tsutakawa, R. K., and Speckman, P. L. (1999). Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika*, 86(2):341–350.

Trockel, W. (1987). Market demand by non-convex preferences. *Milan Journal of Mathematics*, 57(1):311–320.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2):311–324.

You, J. S. and Lim, S. (2013). Welfare effects of nonlinear electricity pricing. *Submission for the 37th International Association for Energy Economics International Conference.*

Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics.* Wiley.

Zellner, A. (1998). The finite sample properties of simultaneous equations' estimates and estimators Bayesian and non-Bayesian approaches. *Journal of Econometrics*, 83(1):185–212.

Zellner, A., Ando, T., Baştürk, N., Hoogerheide, L., and van Dijk, H. (2014). Bayesian analysis of instrumental variable models: Acceptance-rejection within Direct Monte Carlo. *Econometric Reviews*, 33(1–4):3–35.

5.1 Tables

Table 1: Conditions under which data were generated

Run I	Run II	Run III	Run IV
$\delta_0 = 0.7$	$\delta_0 = 0.7$	$\delta_0 = 0.7$	$\delta_0 = 0.7$
$\delta_1 = -1.2$	$\delta_1 = -1.2$	$\delta_1 = -1.2$	$\delta_1 = -1.2$
$\phi_0 = 0.5$	$\phi_0 = 0.5$	$\phi_0 = 0.5$	$\phi_0 = 0.5$
$\phi_1 = 0.8$	$\phi_1 = 0.8$	$\phi_1 = 0.8$	$\phi_1 = 0.8$
$\phi_2 = 0.0$	$\phi_2 = -1.0$	$\phi_2 = 0.0$	$\phi_2 = -1.0$
$\sigma_{11} = 1$	$\sigma_{11} = 1$	$\sigma_{11} = 1$	$\sigma_{11} = 1$
$\sigma_{22} = 1$	$\sigma_{22} = 1$	$\sigma_{22} = 1$	$\sigma_{22} = 1$
$\sigma_{12} = -0.5$	$\sigma_{12} = -0.5$	$\sigma_{12} = -0.5$	$\sigma_{12} = -0.5$
$\mathbf{W}_N$ is rook $\{0,1\}$	$\mathbf{W}_N$ is rook $\{0,1\}$	$\mathbf{W}_N$ is rook $\{0,1\}$	$\mathbf{W}_N$ is rook $\{0,1\}$
$\sigma_v^2 = 0.7^2 \left(\sum_{i\sim j} w_{ij}\right)^{Ave} \sigma_{11}$	$\sigma_v^2 = 0.7^2 \left(\sum_{i\sim j} w_{ij}\right)^{Ave} \sigma_{11}$	$\sigma_v^2 = 0.7^2 \left(\sum_{i\sim j} w_{ij}\right)^{Ave} \sigma_{11}$	$\sigma_v^2 = 0.7^2 \left(\sum_{i\sim j} w_{ij}\right)^{Ave} \sigma_{11}$
$\sigma_z^2 = 0.2$	$\sigma_z^2 = 0.2$	$\sigma_z^2 = 2$	$\sigma_z^2 = 2$

Table 2: Simulation results: Mean Squared Error and Mean Absolute Error for the parameter of interest

Instrument Type	Sample Size	Bayesian IV CAR MSE	Bayesian IV CAR MAE	Bayesian IV MSE	Bayesian IV MAE	Frequentist OLS MSE	Frequentist OLS MAE	Frequentist IV MSE	Frequentist IV MAE	Frequentist ML CAR MSE	Frequentist ML CAR MAE
Exactly Identified											
Weak Instruments	49	0.08	0.27	0.01	0.09	0.30	0.50	70.4	4.20	0.29	0.51
	100	0.03	0.18	0.04	0.19	0.23	0.46	9.62	1.15	0.23	0.47
	144	0.16	0.40	0.29	0.54	0.25	0.49	19.04	1.39	0.25	0.49
Strong Instruments	49	4.19e-04	0.02	0.01	0.12	0.01	0.10	0.03	0.14	0.03	0.15
	100	5.69e-04	0.02	0.01	0.07	0.04	0.18	0.01	0.08	0.02	0.13
	144	1.86E-03	0.04	4.72E-03	0.07	0.04	0.18	0.01	0.07	0.03	0.17
Over Identified											
Weak Instruments	49	0.14	0.38	0.12	0.34	0.21	0.43	5.02	1.11	0.21	0.44
	100	0.36	0.60	0.33	0.57	0.26	0.50	27.37	2.66	0.25	0.50
	144	0.14	0.38	0.47	0.69	0.32	0.54	2.55	1.18	0.25	0.49
Strong Instruments	49	2.29e-03	0.05	0.01	0.08	2.87e-03	0.04	0.01	0.07	2.70e-03	0.04
	100	1.99e-05	3.56e-03	6.94e-04	0.03	2.35e-03	0.04	0.01	0.07	3.45e-03	0.05
	144	9.71e-04	0.03	0.01	0.07	0.02	0.13	0.01	0.09	0.01	0.09

Source: Author's calculations

Table 3: Simulation results: Mean Squared Prediction Error and Mean Absolute Prediction Error

Exactly Identified											
		Bayesian				Frequentist					
Instrument Type	Sample Size	IV CAR		IV		OLS		IV		ML CAR	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weak Instruments	49	0.75	0.66	2.43	1.24	1.99	1.13	63.6	3.81	1.41	0.95
	100	0.89	0.74	1.66	1.03	1.52	0.99	10.61	1.63	1.30	0.91
	144	0.80	0.7	1.74	1.05	1.67	1.03	20.53	1.64	1.29	0.92
Strong Instruments	49	0.69	0.64	2.51	1.27	2.37	1.24	2.51	1.27	1.52	0.97
	100	0.85	0.72	2.34	1.20	2.29	1.19	2.34	1.20	1.44	0.96
	144	0.92	0.75	1.85	1.08	1.80	1.07	1.85	1.08	1.44	0.97
Over Identified											
		Bayesian				Frequentist					
Instrument Type	Sample Size	IV CAR		IV		OLS		IV		ML CAR	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weak Instruments	49	0.86	0.73	1.18	0.87	1.09	0.84	7.83	1.41	1.06	0.82
	100	0.67	0.63	2.05	1.14	1.74	1.06	31.13	2.91	1.27	0.92
	144	0.56	0.59	3.65	1.54	3.57	1.53	5.27	1.74	1.60	1.01
Strong Instruments	49	0.98	0.78	1.48	0.97	1.44	0.96	1.48	0.97	1.33	0.92
	100	0.87	0.73	1.92	1.10	1.89	1.09	1.98	1.12	1.58	1.01
	144	0.71	0.67	3.80	1.58	3.77	1.57	3.80	1.57	1.81	1.08

Source: Author's calculations

Table 4: Simulation results: Mean Squared Error and Mean Absolute Error for the parameter of interest under different types of misspecification

Exactly Identified

Instrument Type	Specification Problem	Bayesian IV CAR MSE	Bayesian IV CAR MAE	Bayesian IV MSE	Bayesian IV MAE	Frequentist OLS MSE	Frequentist OLS MAE	Frequentist IV MSE	Frequentist IV MAE	Frequentist ML CAR MSE	Frequentist ML CAR MAE
Weak Instruments	t-Student	0.14	0.31	0.15	0.32	0.31	0.50	0.28	0.37	0.32	0.51
	Queen	0.61	0.74	0.60	0.73	0.35	0.50	0.79	0.82	0.36	0.52
	SAR	0.15	0.33	0.08	0.22	0.29	0.47	0.13	0.27	0.26	0.46
Strong Instruments	t-Student	0.02	0.11	0.02	0.11	0.11	0.29	0.02	0.11	0.13	0.32
	Queen	0.06	0.24	0.06	0.24	0.15	0.34	0.06	0.24	0.17	0.37
	SAR	0.02	0.10	0.01	0.08	0.09	0.26	0.01	0.08	0.08	0.26

Over Identified

Instrument Type	Specification Problem	Bayesian IV CAR MSE	Bayesian IV CAR MAE	Bayesian IV MSE	Bayesian IV MAE	Frequentist OLS MSE	Frequentist OLS MAE	Frequentist IV MSE	Frequentist IV MAE	Frequentist ML CAR MSE	Frequentist ML CAR MAE
Weak Instruments	t-Student	0.24	0.45	0.24	0.45	0.25	0.46	1.07	0.98	0.27	0.48
	Queen	0.86	0.90	0.87	0.90	0.55	0.68	5.89	2.24	0.56	0.69
	SAR	0.45	0.66	1.35	1.14	0.52	0.68	4.06	1.87	0.40	0.60
Strong Instruments	t-Student	0.02	0.13	0.02	0.13	0.05	0.21	0.11	0.32	0.06	0.23
	Queen	0.08	0.28	0.08	0.28	0.14	0.36	0.44	0.66	0.16	0.38
	SAR	0.07	0.25	0.13	0.36	0.17	0.41	0.31	0.55	0.12	0.34

Source: Author's calculations

Table 5: Simulation results: Mean Squared Prediction Error and Mean Absolute Prediction Error for different types of misspecification

		Exactly Identified									
Instrument Type	Specification Problem	Bayesian				Frequentist					
		IV CAR		IV		OLS		IV		ML CAR	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weak Instruments	t-Student	6.42	1.96	6.65	2.00	6.44	1.98	6.93	2.02	5.59	1.83
	Queen	22.29	3.64	22.60	3.66	22.32	3.63	22.72	3.68	21.65	3.56
	SAR	0.04	0.14	14.85	3.15	14.54	3.11	14.92	3.16	7.34	2.20
Strong Instruments	t-Student	6.84	1.98	7.07	2.01	6.56	2.01	6.97	2.01	5.68	1.85
	Queen	22.19	3.63	22.50	3.65	22.39	3.66	22.50	3.65	21.67	3.58
	SAR	0.13	0.27	14.85	3.15	14.67	3.13	14.84	3.15	7.47	2.21
		Over Identified									
Instrument Type	Specification Problem	Bayesian				Frequentist					
		IV CAR		IV		OLS		IV		ML CAR	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weak Instruments	t-Student	6.07	1.95	6.27	1.98	6.11	1.98	6.61	2.03	4.65	1.71
	Queen	19.65	3.52	19.97	3.56	19.72	3.54	24.17	3.85	17.41	3.31
	SAR	0.04	0.15	9.64	2.48	9.24	2.41	12.00	2.73	5.17	1.81
Strong Instruments	t-Student	6.12	1.96	6.32	1.99	6.23	1.98	6.36	1.98	4.76	1.72
	Queen	19.63	3.53	19.94	3.56	19.87	3.54	20.28	3.52	17.44	3.30
	SAR	0.10	0.22	9.25	2.43	9.22	2.43	9.35	2.45	5.35	1.85

Source: Author's calculations

Table 6: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
Consumption (kWh)	234.874	117.811	26.595	588.937
Electricity Price (US$)	0.061	0.024	0.039	0.240
Income (US$)	397.085	95.242	230.514	619.227
Substitute Price (US$)	0.030	0.006	0.016	0.056
Sea level	29.032%	45.575%	0.000	1.000
Urbanization	45.876%	19.917%	10.700%	98.247%
Coverage (EADE)	77.419%	41.981%	0.000	1.000

Source: Author's calculations

Table 7: Summary of structural parameter posterior estimates

Road Length Contiguity					
Parameter	Mean	Median	95% HPD Interval Lower	95% HPD Interval Upper	$R_{01} = \frac{P(\theta\in(0,\infty))}{P(\theta\in(-\infty,0]))}$
Constant	1.913	1.940	-1.969	6.515	4.663
Price	-0.886	-0.882	-1.577	-0.134	0.014
Income	0.301	0.297	-0.100	0.728	11.920
Subs. Price	0.123	0.120	-0.286	0.523	2.560
Altitude	0.139	0.137	-0.078	0.342	9.235
Urbanization	0.571	0.566	0.376	0.755	908.091
Queen Contiguity					
Parameter	Mean	Median	95% HPD Interval Lower	95% HPD Interval Upper	$R_{01} = \frac{P(\theta\in(0,\infty))}{P(\theta\in(-\infty,0]))}$
Constant	1.873	1.961	-2.992	6.795	4.038
Price	-0.877	-0.876	-1.741	0.069	0.027
Income	0.308	0.298	-0.182	0.786	9.941
Subs. Price	0.117	0.117	-0.321	0.639	2.331
Altitude	0.130	0.135	-0.221	0.530	5.826
Urbanization	0.575	0.566	0.322	0.810	139.840
Rook Contiguity					
Parameter	Mean	Median	95% HPD Interval Lower	95% HPD Interval Upper	$R_{01} = \frac{P(\theta\in(0,\infty))}{P(\theta\in(-\infty,0]))}$
Constant	1.956	1.965	-3.004	6.838	4.061
Price	-0.818	-0.876	-1.790	0.035	0.027
Income	0.297	0.298	-0.168	0.801	9.834
Subs. Price	0.084	0.117	-0.355	0.609	2.328
Altitude	0.178	0.135	-0.258	0.502	5.775
Urbanization	0.575	0.565	0.326	0.817	124.000

Source: Author's calculations

Table 8: Comparison of structural parameter estimates using different techniques

Bayesian IV CAR					
Parameter	Mean	Std. Error[a]	$R_{01} = \frac{P(\theta \in (0,\infty))}{P(\theta \in (-\infty,0]))}$	95% HPD Interval	
				Lower	Upper
Constant	1.913	0.063	4.663	-1.969	6.515
Price	-0.886	0.021	0.014	-1.577	-0.134
Income	0.301	0.005	11.920	-0.100	0.728
Subs. Price	0.123	0.004	2.560	-0.286	0.523
Altitude	0.139	0.004	9.235	-0.078	0.342
Urbanization	0.571	0.007	908.091	0.376	0.755
Frequentist IV					
Parameter	Point Est.	Std. Error[b]	z-Statistic	95% Conf. Interval	
				Lower	Upper
Constant	1.884	2.230	0.840	-2.487	6.255
Price	-0.887	0.470	-1.890	-1.809	0.034
Income	0.302	0.183	1.660	-0.056	0.660
Subs. Price	0.121	0.190	0.640	-0.251	0.493
Altitude	0.136	0.090	1.520	-0.040	0.312
Urbanization	0.566	0.088	6.430	0.394	0.739
Frequentist OLS					
Parameter	Point Est.	Std. Error[b]	t-Statistic	95% Conf. Interval	
				Lower	Upper
Constant	0.594	1.348	0.440	-2.077	3.264
Price	-1.165	0.185	-6.290	-1.532	-0.798
Income	0.377	0.166	2.270	0.025	0.705
Subs. Price	0.107	0.196	0.550	-0.281	0.494
Altitude	0.165	0.099	1.680	-0.030	0.361
Urbanization	0.566	0.090	6.280	0.388	0.744

Notes: [a] Standard error of the mean, taking autocorrelation into account by using the spectral density at 0. [b] Heteroskedasticity-robust standard errors

Source: Author's calculations

Table 9: Tariff variations due to unification

Location	p_1^1	p_1^2
Metropolitan Area	-0.33%	8.12%
Rest	-17.53%	-0.95%

Source: Author's calculations

5.2 Figures

Figure 1: Example of Equivalent Variation with price changes on both tiers, and new and virtual consumption on the second tier

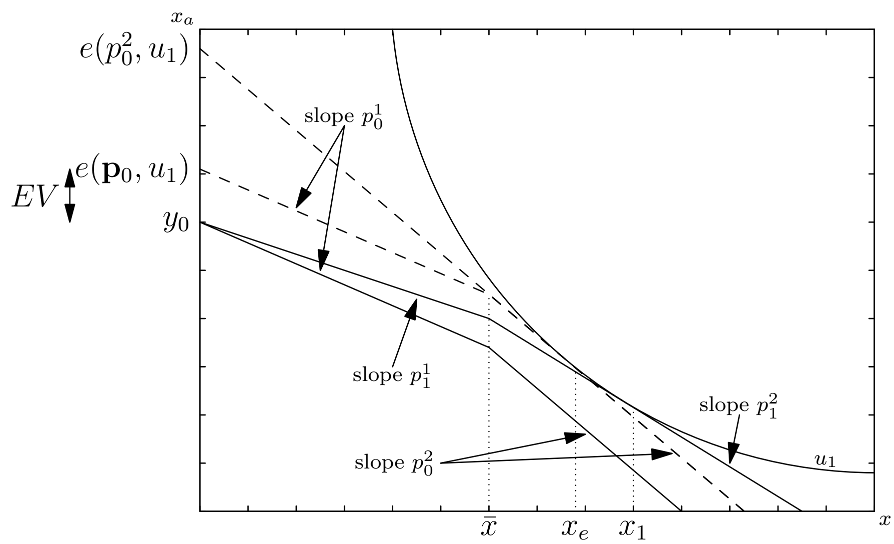
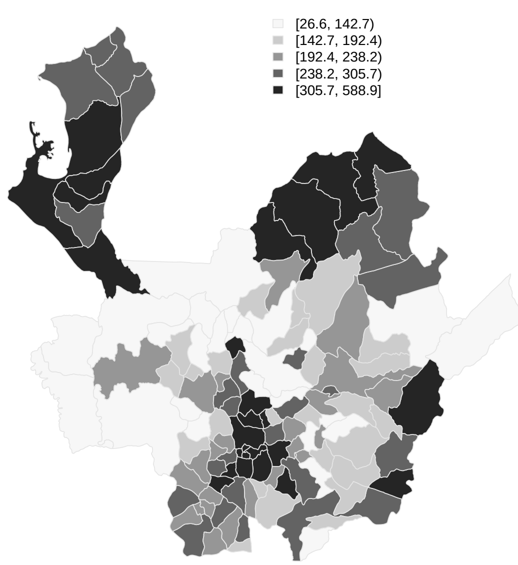


Figure 2: Average Annual Electricity Consumption per Household (kWh): Province of Antioquia (Colombia) in 2005, Stratum One



Source: Empresas Públicas de Medellín

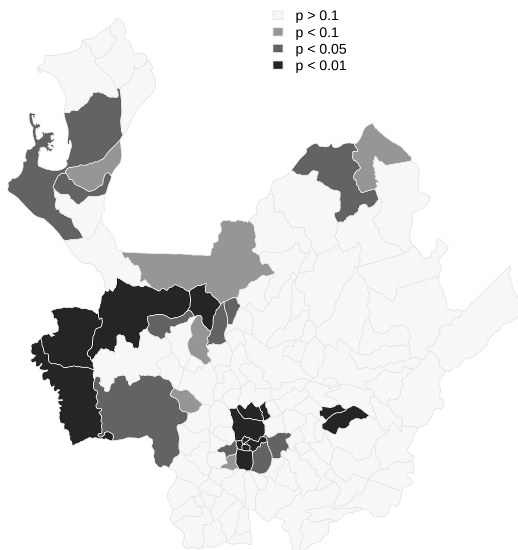
Table 10: Equivalent Variation as share of income by Total, Metropolitan Area and Rest

Road Length Contiguity				
Equivalent Variation	Mean	Median	95% HPD Interval	
			Lower	Upper
Metropolitan Area	0.126%	0.141%	0.005%	0.223%
Rest	0.940%	0.655%	0.257%	2.412%
Total	0.874%	0.630%	0.005%	2.358%
Queen Contiguity				
Equivalent Variation	Mean	Median	95% HPD Interval	
			Lower	Upper
Metropolitan Area	0.127%	0.141%	0.005%	0.229%
Rest	0.937%	0.653%	0.256%	2.416%
Total	0.872%	0.628%	0.004%	2.353%
Rook Contiguity				
Equivalent Variation	Mean	Median	95% HPD Interval	
			Lower	Upper
Metropolitan Area	0.127%	0.141%	0.005%	0.229%
Rest	0.936%	0.653%	0.257%	2.412%
Total	0.871%	0.627%	0.005%	2.350%

Source: Author's calculations

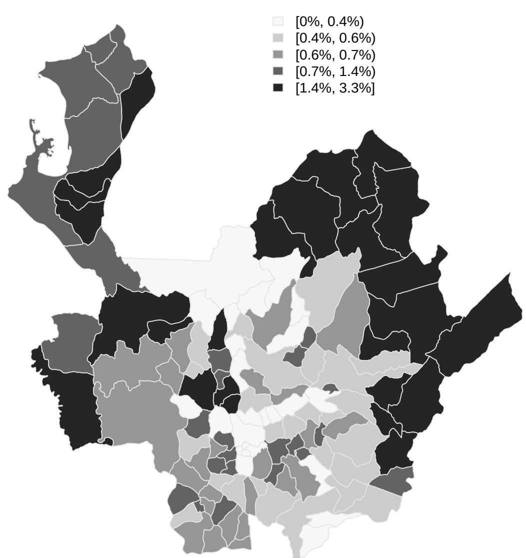
Appendices

Figure 3: Local Moran's I test p-values of Average Annual Electricity Consumption per Household (kWh): Province of Antioquia (Colombia) in 2005, Stratum One



Source: Authors' calculations

Figure 4: Median Equivalent Variation by Municipality



Source: Authors' calculations

A Variable Definitions

Table A.1: Variable definitions and sources

Variable	Definition	Source
Consumption (x)	Average annual electricity consumption per household in kilowatts hour (kWh)	EPM[a]
Price (p)	Average annual electricity price in US$ by kilowatt hour (US$/kWh)	EPM
Income (y)	Average annual per capita income in US$	Ramírez and Londoño (2009)
Substitute price (p^s)	Average annual price of the substitute good in US$ by kilowatt hour (US$/kWh)	CREG[b]
Urbanization (urb)	Ratio of urban to total population	DANE[c]
Altitude (alt)	Dummy variable taking on 1 when the municipality is located less than 1000ms above sea level	Anuario Estadístico de Antioquia[d]
Coverage ($EADE$)	Dummy variable taking on 1 when municipality used to be covered by EADE and 0 otherwise	SUI[e]

Notes: [a] Empresas Públicas de Medellín, [b] Comisión de Regulación de Energía y Gas, [c] Departamento Administrativo Nacional de Estadística, [d] Antioquia's Statistical Yearbook compiled by the Government of Antioquia, [e] Sistema Único de Información

B Diagnostics

Table B.1: Stationarity and Convergence diagnostics

Parameter	Heidelberger (1st Part/p-value)[a]	Heidelberger (2nd Part)[b]	Geweke[c]	Raftery[d]
Road Length Contiguity				
Constant	0.887	0.064	0.758	1.46
Price	0.923	-0.047	0.820	1.10
Income	0.414	0.035	-0.740	2.74
Subs. Price	0.909	0.067	-0.311	1.11
Altitude	0.581	0.052	-0.515	1.08
Urbanization	0.871	0.024	-0.407	1.05
Queen Contiguity				
Parameter	Heidelberger (1st Part/p-value)[a]	Heidelberger (2nd Part)[b]	Geweke[c]	Raftery[d]
Constant	0.830	0.085	-0.079	2.37
Price	0.142	-0.054	-0.238	1.12
Income	0.578	0.037	-0.399	2.71
Subs. Price	0.530	0.111	-0.450	1.16
Altitude	0.266	0.216	1.083	1.34
Urbanization	0.126	0.013	0.550	1.21
Rook Contiguity				
Parameter	Heidelberger (1st Part/p-value)[a]	Heidelberger (2nd Part)[b]	Geweke[c]	Raftery[d]
Constant	0.604	0.208	-0.930	2.39
Price	0.280	-0.155	-0.967	1.16
Income	0.226	0.095	0.866	2.72
Subs. Price	0.634	0.455	-0.745	1.14
Altitude	0.935	0.356	0.541	1.33
Urbanization	0.894	0.034	0.034	1.24

Notes: [a] Null hypothesis is stationarity of the chain, [b] Half-width to mean ratio (threshold of 0.1), [c] Mean difference test z-score, [d] Dependence factor (threshold of 5)