"MEASURING SCHOOL DEMAND IN THE PRESENCE OF SPATIAL DEPENDENCE. A CONDITIONAL APPROACH"

Laura López-Torres[1]
Diego Prior
Universitat Autònoma de Barcelona

Abstract

Improving educational quality is an important public policy goal. However, its success requires identifying factors associated with student achievement. At the core of these proposals lies the principle that increased public school quality can make school system more efficient, resulting in correspondingly stronger performance by students. Nevertheless, the public educational system is not devoid of competition which arises, among other factors, through the efficiency of management and the geographical location of schools. Moreover, families in Spain appear to choose a school on the grounds of location. In this environment, the objective of this paper is to analyze whether geographical space has an impact on the relationship between the level of technical quality of public schools (measured by the efficiency score) and the school demand index. To do this, an empirical application is performed on a sample of 1,695 public schools in the region of Catalonia (Spain). This application shows the effects of spatial autocorrelation on the estimation of the parameters and how these problems are addressed through spatial econometrics models. The results confirm that space has a moderating effect on the relationship between efficiency and school demand, although only in urban municipalities.

Keywords: school efficiency, school demand, spatial econometrics, spatial dependence.

JEL Codes: C14, C21, C61, C67, I21.

[1] Corresponding author: Laura.Lopez.Torres@uab.es. Department of Business (Universitat Autònoma de Barcelona). Building B – 08193 Bellaterra (Barcelona – Spain). Phone: +34 93 581 1209.

1. INTRODUCTION

The quality of education system and the factors that may be associated with better student achievement is attracting growing academic interest in the 21st century (Ngware et al. 2011). On the one hand, investments in education affect numerous individual behaviors throughout the life course (Hanushek and Kimko, 2000). On the other hand, expanding school choice can improve the efficiency of public schools through heightened competition which arises, among other factors, through the geographical location of schools (Hoxby, 2000). This location can affect the choice of school families make. Parents decide on a particular school based on their personal judgments about the quality of teaching it provides. In this decision, location is an essential factor. One implication of this finding is that public schools already face some competition from other public schools in the area (Barrow, 2002). Understanding the strength of the competitive forces emanating from alternative public schools in neighboring areas may shed light on the value added from additional demand that may be induced through expanded school choice.

Consequently, the main purpose of this paper is to analyze whether school location has an impact on the relationship between the level of technical quality of public schools (measured by the efficiency score) and the school demand index. Using data for Catalonia (Spain) over the academic year 2009/2010, we apply a specific methodology scarcely seen in the education literature, namely spatial econometrics (SE) (Anselin, 1988a), and combine it with the use of robust non-parametric techniques. This process allows us to study, in a first step, school efficiency taking into account not only the internal inputs that affect school efficiency, but also non-discretionary variables such as the complexity inside the school or the school environment. In a second step, we estimate a specific regression model that introduces the spatial problems detected, thereby providing a better approximation to the school demand index. Ignoring spatial effects in the estimation of models can lead to inefficient or even biased estimators. At the same time, including the spatial dimension in the analysis contributes new information that can improve the research and shed light on the phenomenon studied.

In addition, we want to test whether or not the type of municipality (rural *vs* urban) changes the effect of space on the relationship between demand and school efficiency. This specific objective concerns the choices available to parents depending on the location of schools and the type of municipality. In some towns with very small populations only one public school is available. In these cases, the school operates in isolation and parents have no option but to send their children to this school. In these rural municipalities, location would not be an indicator of competition. In these cases we do not expect space to be relevant or significant. In contrast, many public schools are available in cities with large populations, therefore increasing the choices available to parents. They may make better decisions and use more decision variables in choosing the most suitable school for their children. Schools in these locations operate in a situation of increased competition compared to other schools. This line of inquiry is not new, although empirical examinations are relatively sparse. Hoxby (2000) examines the impact of competition (measured by number of school districts within a metropolitan area) on student achievement, finding positive effects on achievement. Similarly, Marlow (2000) finds positive effects of competition (measured using either a Herfindahl index or number of neighboring school districts) on achievement. Moreover, Zanzig (1997) finds that greater competition is irrelevant once a certain competitive threshold is attained.

Our results are striking. We find strong support for the notion that location is determinant in explaining the relationship between the level of technical efficiency of public schools and the school demand index. Space reduces the negative effect of inefficiency on school demand, proving there is a spatial spillover effect among neighboring schools. Specifically, we find this effect is stronger in urban zones and insignificant in rural areas, thus supporting our idea about availability of choice. Finally, while perhaps initially surprising, the results support the hypothesis that some (negative) variables related to the school environment positively affect schools' potential outcomes, in contrast to what the literature has revealed so far (Muñiz, 2002; Corman, 2003; Cordero et al. 2010).

After this introduction, the remainder of the paper is organized as follows. The next section outlines the conceptual framework, establishing the relationship between SE and regional science, and also offers a brief literature review about SE applications. In section three we explain the methodology and data used. We analyze the results in section four. The main conclusions, limitations and future research lines are shown in the section five.

2. SPATIAL ECONOMETRICS AND REGIONAL SCIENCE

Conventional economic analysis has traditionally given more importance to the role of time as a key dimension of study, rather than the spatial factor. During the 90s authors such as Krugman (1991a, 1991b, 1998) renewed interest in this issue by taking into account space as a variable of analysis. Thus, the re-emergence of regional science through the reconsideration of space has led to the emergence of a new theoretical field known as spatial econometrics (SE). SE is a separate discipline from conventional econometrics due to the need to work with the special nature of cross-sectional data and the importance of location in the estimation of economic models (Anselin 1988a, Anselin and Rey, 1997).

When we use cross-sectional data two spatial effects can appear: spatial heterogeneity and spatial dependence[2]. On the one hand, spatial heterogeneity refers to the variation of relations in space. It can lead to problems such as heteroskedasticity or structural instability, which can be solved by existing econometric techniques for time series[3]. On the other hand, spatial dependence occurs as a consequence of the existence of a functional relationship between what happens at one point in the space and what happens elsewhere (Cliff and Ord, 1973; Paelink and Klaassen, 1979; Anselin, 1988a). It can be positive or negative. For example, if the presence of a phenomenon in one school (e.g., installation of a computer room or a laboratory) causes the same phenomenon to spread to other schools nearby, we have a case of positive spatial autocorrelation. Otherwise, there will be negative spatial autocorrelation. When the variable analyzed is randomly distributed, there is no spatial autocorrelation.

There are several causes that lead to the emergence of spatial dependence (Anselin, 1988a, 11-13) such as the existence of measurement errors and spatial interaction phenomena, spillover effects and spatial hierarchies. However, it cannot be dealt with by standard econometrics because of the relations of multidirectional interdependence between spatial units. In order to solve these problems SE provides the contrasting and estimation techniques required to work with data that present problems of heterogeneity and/or spatial dependence[4].

[2] We refer to spatial dependence and spatial autocorrelation synonymously in this paper.
[3] Because spatial heterogeneity can be solved with traditional econometric techniques we do not analyze the problem in this paper.
[4] For a detailed analysis of SE and spatial effects, see Anselin (1988a) and Elhorst (2012).

A number of branches of economics have incorporated SE in their analyses, including urban economics, regional economics and macroeconomics (Moreno and Vayá, 2000). However, the poor dissemination of SE is evident in the education field, especially in Spain, thus revealing a need to bring SE techniques to researchers in the area of economics of education.

During the last decade, several empirical articles have dealt with problems associated with constructing econometric models in a spatial context. Arbia's (2011) paper provides an excellent and extensive literature review[5] of the theoretical and empirical contributions to SE from 2007 to 2012 and the main journals that have published papers related to SE. The author considers more than 230 papers that appeared in this period in several scientific journals. Therefore, SE is a discipline with an increasing number of applications in very diverse scientific fields. Consequently, the contributions are wide-ranging and distributed across many different scientific journals, suggesting that SE is becoming more robust.

Finally, it is important to highlight the work of Arbia (2001) in motivating the empirical specification of our study. In the paper, the author defends the need for a microeconomic approach in spatial analysis, as opposed to the usual meso-approach (based on regional aggregates). In recent years, there has been a growing demand for information on small spatial units (urban districts, municipalities, regions). This question, together with the scant number of previous studies in the education field (e.g. Zanzig, 1997; Marlow, 2000; Hoxby; 2000; Millimet and Rangaprasad, 2006, Gu, 2012a, 2012b), justifies the need to apply such SE to smaller units like schools. Our aim is to better explain the determinants of school demand through a new methodology with few applications in the literature so far.

3. METHODOLOGICAL ISSUES

Our objectives suggest the need for a multi-stage methodology to solve them. Thus, the methodological approach is developed in two parts. First, we conduct an efficiency analysis, and second we develop a spatial study through a specific regression model.

[5] Anselin (2007, 2009) and Pinkse and Slade (2010) also provide a comprehensive review of the subject.

3.1. Robust Non-Parametric Efficiency Estimations

To perform this part of the analysis we use a specific non-parametric and robust approach, the conditional order-*m* model (introduced by Cazals et al. 2002 and Daraio and Simar, 2005). Order-*m* frontier estimators are known to be more robust to outliers and extreme values than the full frontier estimates (Data Envelopment Analysis (DEA) or Free Disposal Hull (FDH)). The basic ideas of the algorithms developed are taken from Daraio and Simar (2005). We therefore use the same notation as these authors to avoid possible confusions.

Let us define our working variables. Pupils transform a set of inputs $\mathrm{x} \in \mathrm{R}_{+}^{p}$ into heterogeneous outputs $\mathrm{y} \in \mathrm{R}_{+}^{q}$. In this framework, the production set is defined as:

$$\Psi = \{\, (x, y) \in \mathrm{R}_{+}^{p+q} \mid \text{x can produce y}\} \quad (1)$$

We also have several non-discretionary factors denoted as $Z \in \mathbb{R}^{r}$ that affect the efficiency estimations. The efficiency analysis should take these variables into account[6].

The order-*m* approach creates a partial frontier that envelops only *m*[7] observations randomly drawn from the sample. This procedure is repeated B times[8] resulting in multiples efficiency scores $(\hat{\theta}_{m}^{1}, \dots, \hat{\theta}_{m}^{B})$ from which the final order-*m* efficiency measure is computed as the simple mean $(\hat{\theta}_{m})$. This estimator allows us to compare the efficiency of an observation with the *m* potential DMUs that have a production larger or equal to y. The production set could be as follow:

$$\Psi_{m}(x) = \left\{(x', y) \in \mathbb{R}_{+}^{p+q} \mid x' \leq x, Y_{i} \leq y, i = 1, \dots, m\right\} \quad (2)$$

[6] Our purpose here is to achieve a final conditional order-*m* efficiency model in which we have the strictly necessary non-discretionary factors (non-separables). To achieve it, first we ran an unconditional order-*m* model only taking into account the inputs and outputs. Then we ran a conditional order-*m* model for each $z_i \in Z$. Finally, we conducted a separability test by applying an extension of the method proposed in Daraio et al. (2010). In this case, we test the null hypothesis $H_0 = \Psi_{m}^{z} = \Psi_{m}\ \forall\, z \in Z$ versus $H_1\ \Psi_{m}^{z} \neq \Psi_{m}$ for some $z \in Z$.

To do so we apply the following test: $\widehat{D}_{order-m,i} = (Y_i\ \tilde{\theta}_{m}^{b}(X_i, Y_i) - Y_i \tilde{\theta}_{m}^{b,z}(X_i, Y_i \mid Z_i)) \neq 0$. When we reject the H_0, we will qualify the environmental (or non-discretionary) factor as a non-separable variable, so it has to be part of the efficiency model. When the efficiency score does not significantly change with the inclusion of one z_i (H_0 is not reject) we qualify the variable as a separable environmental factor that will be excluded from the analysis. Once we obtain the rating for each environmental variable, we can run the final efficiency assessment model, the conditional robust order-*m* estimation with the strictly necessary environmental factors, namely the non-separables.

[7] According to Daraio and Simar (2005) we use value of *m* for which the decrease in super-efficient observations stabilizes. We therefore fix *m* = 100.

[8] Here we are following Simar (2003) and we fix *B* = 200. This level of repetition seems to be a reasonable choice.

We also control for the inclusion of non-discretionary factors $Z \in \mathbb{R}^r$. Although these variables are exogenous to the production process, they play an important role. The literature reports different approaches on how to introduce them (for an overview see Simar and Wilson (2007) and De Witte and Kortelainen, (2013)). In this study we apply a conditional order-m model for introducing environmental variables (Cazals et al. 2002; Daraio and Simar, 2005). The conditional model works with probabilistic formulation and incorporates the environmental effect, conditioning the characteristics of the non-discretionary factors. It constructs a boundary representing the reference set in which each unit is compared. This method also avoids the separability condition of two-stage methods and does not require specification of the influence of each environmental variable on the efficiency. To estimate the conditional model, smoothing techniques are needed such that in the reference samples of size m observations with comparable z-values have a higher probability of being chosen. To do this we apply the method first proposed by Badin et al. (2010) and then modified by De Witte and Kortelainen (2013)[9]. Therefore, the estimator for the conditional survivor function of Y can be expressed as (expression 16 in De Witte and Kortelainen, 2013, 2405):

$$S'_{Y,n}\,(y \mid x, z) = \frac{\sum_{i=1}^{n} I(x_i \le x, y_i \ge y) K_h(z, z_i)}{\sum_{i=1}^{n} I(x_i \le x) K_h(z, z_i)} \qquad (3)$$

Where $K_h(\cdot)$ represents the multivariate kernel function, $I\ (\cdot)$ is an indicator function and h is an appropriate bandwidth parameter for this kernel. This leads to the conditional order-m output efficiency estimator derived from this algorithm (Daraio and Simar, 2005):

1. Compute equation (3)

$$\tilde{\theta}_m^z\,(x, y) = sup\,\{\theta | (x\,|\theta y) \in \Psi_m^z\,(x)\} = E\left[\max_{i=1,\dots,m}\{\min_{j=1,\dots,q}(\frac{Y_i^j}{y^j})\}\,|X \le x, Z = z\right] \qquad (4)$$

Where $\Psi_m^z\,(x) = \left\{(x', y) \in \mathbb{R}_+^{p+q} \mid x' \le x, Y_i \le y, Z = z, i = 1, \dots, m\right\}$

2. Redo step 1 for $b = 1, \dots, B$, where B is large.
3. Finally, $\hat{\theta}_{m,n}\,(x, y \mid z) \approx \frac{1}{B} \sum_{b=1}^{B} \tilde{\theta}_m^{b,z}\,(x, y)$. (5)

The efficient frontier corresponds to those points where $\hat{\theta}_{m,n}\,(x, y \mid z) = 1$. In this case the score can be lower than one. This would mean that the school is labeled as super-efficient, since the order-m frontier exhibits lower levels of outputs than the school under assessment.

[9] See De Witte and Kortelainen (2013) for a detailed explanation of the advantages of their method.

3.2. Spatial Study

The next step is to introduce the effect of space into the analysis, for which it is necessary to work with spatial data. Specifically, we use UTM coordinates to validate the geographic location of each school. In our sample, we expect a high spatial interdependence among schools. For instance, student results can be affected by the geographic location of the school or by the environment of the area where the school is operating. As we explained above, the main technique we use to conduct this second part of the analysis is Spatial Econometrics (SE) (Anselin, 1988a).

Firstly, we perform an Ordinary Least Squares (OLS) regression by taking the school demand index as the dependent variable and, as the explanatory variable, the conditional efficiency score obtained in the first stage.

$$Demand = \gamma + \beta * \tilde{\theta}_m^{b,z} + \varepsilon \qquad (6)$$

Where $Demand$ is the school demand ratio and $\tilde{\theta}_m^{b,z}$ denotes the conditional order-*m* efficiency scores.

Secondly, we study the distribution of the data through exploratory spatial data analysis (ESDA) and then, apply statistical tests to detect the existence of spatial dependence. Finally, we fix the previous model (6) considering the spatial problems detected, thus obtaining a better approximation of school demand.

ESDA methodology is used to study patterns and associations of spatial data. It is equivalent to a descriptive analysis of the spatial distribution of the variable under study. To carry out this analysis maps and specific techniques are commonly used to describe spatial distributions, identify spatial outliers and spatial clusters (Moreno and Vayá, 2000). Anselin (1988a) presents a classification using different techniques for ESDA. Table 1 summarizes a set of indicators that allow us to test the presence of a spatial autocorrelation scheme at the univariate level. In this case, H_0 would be a non-spatial autocorrelation (i.e., a variable is randomly distributed in space) against the alternative hypothesis H_a: there is a significant association of similar or dissimilar values between neighboring regions.

< Table 1 around here >

ESDA also includes other techniques that enable, through maps, to complement the results obtained from previous tests. Some of the most valuable are the *box map* (useful to identify outliers), the *Moran's scatterplot* (the x-axis shows the observations of the standard variable under study and the y-axis represents the normalized spatial lag of the same variable) and its associated *scatter map* (represents the map of the territory).

Once obtained an idea about the spatial distribution of the data and confirmed the existence of spatial dependence, the next step is to design suitable model that allows us to correct it[10]. Spatial dependence can appear in a regression model as a consequence of the existent correlation in the dependent variable (substantive spatial autocorrelation), in one or more independent variables, or because of the existence of a spatial dependency scheme in the error term (residual spatial autocorrelation). This can be translated into different ways of incorporating spatial dependence in regression models through the spatial weight matrix, or contacts matrix, W and the spatial lag operator. Firstly, let us define W as:

$$W = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1N} \\ w_{21} & 0 & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & 0 \end{bmatrix} \quad (7)$$

W is a non-square stochastic matrix whose elements (w_{ij}) reflect the intensity of the relationship between each pair of regions i and j. There is no single way to define the weights, but those weights must always be non-negative and finite (Anselin, 1988a). The matrix W has to be standardized by dividing each element w_{ij} by the sum of the elements of each row. To carry out our analysis we use a contacts matrix based on distance. Thus, the intensity of the interdependence between two regions decreases with the distance between them. We consider it to be the best option to classify neighboring schools. In normal circumstances it is difficult to find two schools that are physically adjacent or share a boundary.

Secondly, it is important to introduce the spatial lag. This operator is a weighted average of random variables at neighboring locations (Anselin, 2000; Moreno and Vayá, 2000, 27). The spatial lag operator is obtained as the product of the matrix W by the observations vector of a random variable y, i.e., Wy. Thus, each element of a spatial lagged variable is equal to:

$$\sum_{j=1}^{n} w_{ij} \, y_j \quad (8)$$

Where w_{ij} refers to the weights of W and y is an Nx1 vector of space observations of the random variable.

---

[10] For space reasons we focus on overall modeling of these variants. For further detail on specific aspects and implications of these techniques on standard econometric techniques or variants of SE, consult Anselin (1988a), Moreno and Vayá (2000) and Elhorst (2012).

Defining the spatial regression model to be used requires starting from a general lineal regression model like:

$$y = X\beta + u \tag{9}$$

Where $u \sim N(0, \sigma^2 I)$, y is an $N \times 1$ vector, X is a matrix of K exogenous variables, u is the white noise perturbation term and N is the number of observations. Variants of the regression model that incorporate the spatial dependence are, first, lag models when the dependence is substantive. In this case the model could be:

$$y = \rho W y + X\beta + u \tag{10}$$

Where Wy is the spatial lag of y and ρ is the autoregressive parameter which contains the intensity of the interdependence among units.

Similarly, spatial correlation could be present in the perturbation error:

$$y = X\beta + \varepsilon \tag{11}$$

Where $\varepsilon = \lambda W \varepsilon + u$, λ is the autoregressive parameter which contains the intensity of the interdependences.

Mixed structures are also available, in which both substantive and residual spatial autocorrelation exist, as well as spatially correlated explanatory variables.

$$y = \rho W_1 y + X\beta_1 + W_2 R \beta_2 + \varepsilon \tag{12}$$

Where $\varepsilon = \lambda W_3 \varepsilon + u$, X is a $N \times K_1$ matrix of exogenous variables, R is a $N \times K_2$ matrix of exogenous variables which are spatially lagged.

As in the case of the detection process, there are a number of spatial statistics to contrast the above structures. In all cases, the null hypothesis is that spatial autocorrelation does not exist. Table 2 lists the most commonly statistical tests used. The type of spatial correlation depends on the values of these statistics. Those which take a higher value will indicate the kind of spatial dependence detected in the data.

< Table 2 around here >

Finally, we estimate a valid model to explain the school demand. In this case, the Maximum Likelihood approach (ML) is among the most widely used[11] (see Anselin (1988a) for a detailed explanation of the estimation process) [12].

[11] Other alternative estimation methods proposed in the literature include instrumental variables, generalized method of moments, bootstrapping techniques or Bayesian estimation. See Anselin (1988a) for further review.

[12] R software (version 3.0.1) was used in all these operations.

3.3. Data and variables

Based on a previous study on school efficiency (López-Torres and Prior, 2013), we use a specific database from the Catalan Evaluation Council of the Education System (*Consell d'Avaluació del Sistema Educatiu de la Generalitat de Catalunya*). The sample includes 1,695 primary schools for the academic year 2009-2010, covering almost all the public schools in Catalonia. The relevant unit of observation is the school, as we do not have access to students' data. We aware about the importance of having students level data and the problems that aggregation could cause (this topic has been treated in the literature, e.g. Hanushek et al. 1996, among others). Table 3 collects the variables used for the efficiency analysis.

< Table 3 around here >

Regarding the selection of variables, different methodological approaches can be taken, but the output used in most of them is the academic results from aptitude tests that are homogeneous for all students. Following the literature (e.g. Smith and Mayston, 1987; Johnes, 2006; De Witte and Kortelainen, 2013; Grosskopf et al. 2013) we consider as output variables the sum of the arithmetic means of the students' marks in the sixth grade general test conducted in Catalonia and the number of students who pass the exams.

In terms of inputs, students usually spend resources in order to study (Ray, 1991). Most of the studies in the literature distinguish between quality of teachers and the physical conditions of the school as the main resources[13] (e.g. Opdenakker and Van Damme, 2001; Johnson and Ruggiero, 2011; Silva-Portela et al. 2013). In this category, we include the number of teachers employed and students enrolled.

Finally, several empirical studies have estimated the impact of non-discretionary factors on school outcomes. They can have different origins (environmental factors external to the school or complexity factors belonging at school (Harrison et al. 2012)). The majority of empirical papers reveal that students' educational and socioeconomic environment explain the differences in their achievement (Ruggiero, 1998; Muñiz, 2002; Muñiz et al. 2006; Rubenstein et al. 2007; Mancebón and Muñiz, 2008; Cordero et al. 2008, 2010; De Witte and Kortelainen, 2013; Thieme et al. 2013). Therefore, according to the previous literature the environment of the school is captured by 14

[13] See Hanushek (1986, 2003) that deals with the importance (or not) of including teacher quality in the efficiency assessment.

variables found to be significant in the separability tests explained above. We include two ordered variables (X_{nd1} and X_{nd2}) referring to the home environment. We also take into account one unordered variable (X_{nd3}) to capture teachers' commitment inside the school, and 11 continuous variables (X_{nd4} - X_{nd14}) related to the complexity inside the school.

Summary statistics for efficiency variables are provided in Table 4. As can be seen, there are some very small schools with only 4 students and at the other extreme, larger schools with 730 students. This information demonstrates the breadth of our sample, which includes schools operating in municipalities of different sizes. Later we test for any significant differences in the role of location by controlling for typology of municipality. Non-discretionary factors reveal some interesting aspects. First, although a high number of parents are unemployed, they usually have professional qualifications and those who are working have administrative positions. Second, the combined effect of availability of innovation projects and school stability shows the school's commitment to educational quality.

< Table 4 around here >

In addition, Table 5 presents the correlation matrix among efficiency variables. Given the large number of non-discretionary factors defining the internal and external environment of the school, we decided to conduct a multicollinearity study to detect possible significant relationship and collinearity problems. Both the Tolerance and VIF tests show values that are not disturbing. In all the cases Tolerance is higher than 0.3 and VIF is lower than 3 (see Belsley et al. 1980 for thresholds).

< Table 5 around here >

We also present the variables we apply in the spatial study in Table 6 and some descriptive statistics about them in Table 7.

< Table 6 around here >

The main variable we want to explain is school demand. This is a ratio between the number of enrollment applications from families and the places offered by the school. As Table 7 shows, on average schools do not cover the total places available, indicating that they have the capacity to take more students, which translates in improvement possibilities to attract new students and therefore greater demand from parents. The last variable, population, enables us to divide the sample into two groups in order to fulfill the second specific objective, and identify whether the location is more important in rural or in urban municipalities. We divided the sample following the Eurostat criterion:

municipalities with fewer than 5,000 inhabitants were classified as rural and those with 5,000 inhabitants or more as urban.

< Table 7 around here >

4. EMPIRICAL RESULTS

In order to facilitate the explanation of the results, we divide this section into the same stages as explained above in Section 3.

4.1. Robust Non-Parametric Efficiency Estimations

In the first part of the efficiency analysis we consider schools' outcomes without controlling for non-discretionary factors. We estimate the unconditional and robust order-*m* model. Summary statistics on the unconditional efficiency scores are presented in Table 8.

< Table 8 around here >

As can be seen, school performance amounts to 1.12, on average (θ_{uncond} in Table 8). This means that in our sample schools could perform better if they imitated the best practice schools. The number of students who pass the course and grades could increase, on average, by 12%. It is important to note that our sample has some super-efficient schools which are performing better than the average *m* schools they were benchmarked with.

We next control by environmental variables $Z \in \mathbb{R}^r$. To do so, we estimate the conditional and robust order-*m* model for each *Z*. In this part of the analysis we want to know whether each $z_i \in Z$ is a separable or non-separable factor in order to include it in the final efficiency estimation. To conduct this separability analysis we run the test explained in note 6 by applying the related samples non-parametric Wilcoxon test (Wilcoxon, 1945). The results are shown in Table 9.

< Table 9 around here >

Testing for the inclusion of each Z in the conditional model shows that some of them are irrelevant and do not have a significant influence on the production process. Specifically, we find variables such as unidentified parents, school age, number of changes in the school principal and teachers' absenteeism do not influence the school's outcomes. Given this insignificant relationship we decided to exclude them from the conditional order-*m* final estimation.

Finally, we control for heterogeneity among schools by running the final conditional order-*m* efficiency model with the non-separable non-discretionary factors. As previously mentioned, we follow De Witte and Kortelainen's (2013) proposal. Thus, taking into account school environment, the average conditional efficiency score rises to 1.2 (θ_{cond} in Table 8). This means that when we control for the environment schools performance worsens (in other words they have more opportunities to improve when they are benchmarked with schools that have a similar environment). As a result, the efficiency score is lower in the unconditional order-*m* than in the conditional model, on average.

This is a surprising result as the literature usually negatively classifies the impact of school environment on school outcomes (e.g., Muñiz, 2002, Corman, 2003; Cordero et al. 2010, among others). In order to better explain this controversial result, we perform a non-parametric regression with the ratio of the conditional and unconditional efficiency scores as a dependent variable and the exogenous variables as explanatory variables, as in De Witte and Kortelainen (2013). The significance test is presented in Table 10.

< Table 10 around here >

As can be seen, some of the variables significantly impact on the efficiency ratio and the schools' outputs. Firstly, the average effect on efficiency is positive and significant for the two ordered variables (socio-economic and educational level). That means the larger the *z*, the greater the outcomes the school can achieve. In practical terms, when we compare inefficient and efficient schools with similar socio-economic and educational levels, the potential output increases as the environment plays a favorable role in the targets to be achieved. Secondly, some of the continuous variables reduce the inefficiency (the efficiency score is lower as we are in an output orientation model) due to the way they are defined. This is the case of the number of students with special educational needs and the dropout rate. These findings are in line with the literature (e.g., De Witte and Kortelainen, 20013; Feng and Sass, 2013).

Finally, the most surprising results come from the continuous variables unemployed and grants. *Ceteris paribus,* these two variables positively affect the potential school outputs. For instance, we can confirm that the larger the number of unemployed parents, the better for students' potential outcomes. Although this result can initially appear controversial, we think it has a logical interpretation which corresponds to the reality in many households. To better explain this astonishing result, we turn to the theory of social promotion posed by Ouchi (2003). This author demonstrates that students attending the worst public school in the US (the Goudy Elementary School) achieved exceptional results in their general tests thanks to the perseverance and commitment of the parents and the school principal[14]. Specifically, *"this school is located in an immigrant neighborhood on the far north end of Chicago where 26 languages are spoken every day. The teachers, students, and families were devastated by the negative publicity. 98% of the students are from low-income homes and thus qualify for free or reduced-price lunches under a federal program. However, on the Iowa Test of Basic Skills used in Chicago schools, reading scores rose from the 14.9th percentile to an astounding 56th (above the state and national averages). Math scores have also skyrocketed, from the 24.7th percentile to the 63rd"* (Ouchi, 2003, 3-7).

Parents encouraged their children to obtain the best marks they could in order to escape that negative environment and find a good job in the future. The school principal exactly matched the needs of his unique population of students. He delegated most decisions to his teachers, who solved the problems by providing their students with a good education. *"They focused everyone on student achievement, not complaining about the poor children who were in the neighborhood"* (Ouchi, 2003, 4).

This true story demonstrates that a school's good results are not only a question of environment; parents and teachers also have an important role to play. If the worst school in the US could become one of the best, then every school can be successful. For those who believe that a neighboring school made up of families with economic needs from homes in poverty or with a high level of unemployment cannot achieve high academic levels, the Goudy school proves otherwise.

4.2. Spatial Study

After the efficiency study, the next step is to introduce the effect of space into the analysis. Our main purpose is to analyze whether school location has an impact on the relationship between the level of technical quality of public schools and the school demand index. As previously mentioned, we apply SE techniques in order to detect the possible effect of space on school demand. Thus, we start with the exploratory spatial data analysis (ESDA) to study patterns and associations of spatial data. Then we conduct two spatial regression models, one with the entire sample and the other

[14] See the work by Ouchi (2003) for a comprehensive explanation about the situation of this school.

distinguishing between rural and urban municipalities in order to give answer to our specific objective.

To this end, we first performed an ESDA that enables us to identify different patterns of spatial association and regional clusters or atypical locations, which is particularly important to characterize the Catalonian landscape of school demand. Our empirical analysis begins with an initial picture of the distribution of school demand, presented in Figure 1.

< Figure 1 around here >

The figure reveals relatively significant disparities in the proportion of school demand across Catalonia. Specifically, we can draw two different conclusions. First, while the most remote municipalities have low school demand, the highest values are concentrated in cities or central regions. The first group includes the more distant towns of Lérida and Tarragona, which present lower values (in blue) compared to the city centers of Barcelona and Gerona where a higher demand for schools is seen (in red). Secondly, school demand does not seem to be randomly distributed across space. We can observe a positive spatial association between adjacent areas because they show similar school demand values. Figure 2 represents the associated box map. This figure reinforces the previous idea, appearing again a positive spatial dependence in the distribution of school demand. Thus, the areas grouped in the same quartile also form clusters in space.

< Figure 2 around here >

Some caution is recommended when interpreting the data shown in Figures 1 and 2, since the conclusions that might be drawn are highly sensitive to the number and width of the different intervals used to represent the variable of interest. Additional analyses should be performed to determine the degree of spatial interdependence between the values of the study variable at different geographic locations. For this reason, we supplemented the preliminary evidence provided by these Figures with a formal analysis of the possible presence of spatial autocorrelation in our sample. To this end, we calculated Moran’s I and Getis and Ord’s G global tests of spatial autocorrelation (Table 11). As we noted previously, we use a standardized *W* matrix defined by the distance among schools calculated from the UTM coordinates.

< Table 11 around here >

The result of the global tests provides us with standardized values of 0.3403 and 0.5974, respectively, which are significant at the 0.1 percent level. This is evidence of a pattern of positive spatial association in this context, which is consistent with the initial impression drawn from Figures 1 and 2. We can conclude that in Catalonia, schools located in spatially adjacent zones tend on the whole to exhibit a similar degree of demand. To further confirm this finding, we also constructed the corresponding Moran's *scatterplot* (Figure 3) and the *scatter* map associated to the Moran's local test (Figure 4) for the school demand distribution. As can be seen from Figure 3, the majority of the schools considered are located in quadrants I and III. This confirms that Catalonia is characterized by the presence of spatial clusters of areas with similar levels of school demand while there are relatively few cases in which a zone registers a value of the analyzed variable that is markedly different from the average of its neighbors.

< Figure 3 around here >

< Figure 4 around here >

Figure 4 shows how the concentrations of high values of the analyzed variable are situated in the city centers of Barcelona and Gerona. On the other hand, the groupings of zones characterized by a low proportion of school demand are located in Lérida and Tarragona.

The analysis performed so far is useful to describe the spatial distribution of demand in public schools in Catalonia, but it is not suitable to quantify the magnitude of regional differences in the variable of interest. To do so, and following common practice in SE (Florax and Folmer, 1992) we start by estimating the model proposed in equation (6) by OLS and performing various spatial dependence tests based on the residuals provided by the OLS estimations. Specifically, we calculated the Lagrange multiplier tests for the spatial error (LM-ERR) and the spatial lag models (LM-LAG) proposed, respectively, by Burridge (1980) and Anselin (1988b) as well as their robust versions (RLM ERR and RLM LAG, respectively). Table 12 reveals that the results of these tests lead in all cases to the rejection of the null hypothesis of absence of residual spatial dependence.

< Table 12 around here >

Indeed, according to the decision rule proposed by Anselin and Rey (1997), the values of the various Lagrange multiplier tests calculated suggest that in this context the spatial lag model is preferable to the spatial error model. However, we see no great difference in their values. For this reason we conduct the SARMA test to contrast both types of spatial dependence by combining the basic statistics LM ERR and LM LAG. The null hypothesis in this case is $H_0\colon \rho = 0;\ \lambda = 0$. As can be seen in Table 12, we can reject the null hypothesis and confirm that our data have substantive and residual dependence. Therefore, we can conclude that, first, the demand index of a school in an area i is affected systematically by the demand index of schools in neighboring areas (substantive spatial dependence). Second, there are interdependencies in the school demand index among schools located in neighboring areas due to, among other factors, spillover effects between neighboring areas (residual spatial dependence).

To correctly introduce the effect of spatial dependence detected, the next step is to estimate the spatial model using the ML approach. Table 13 shows the results. As can be seen, a spatial autoregressive structure should be included first, in the dependent variable (model 1 LAG). Then, we contrast the model with a spatial autoregressive structure included in the error term (model 2 ERR). In model 3 (SARAR) we include both substantive and residual effects together. Finally, in model 4 (SARARMIX) we add a spatial lag in the independent variable. As can be seen, SARARMIX is the more complete model. Thus, we compare the base model (OLS) with SARARMIX (model 4) in order to explain the spatial dependence.

< Table 13 around here >

In the OLS model the conditional efficiency score has a negative and significant relationship with the school demand index. As we are in an output orientation model, this finding fits with our intuition: the more inefficient a school (higher $\tilde{\theta}_m^{b,z}$), the less demand it will have from parents. The estimated coefficient (β = -3.11**) reveals this to be a strong relationship. However, when we include the school location effect we find this relationship remains negative, although the coefficient and the significance are lower (β= -2.85*). This means that when we control by space we find that school location exerts a moderating effect in the relationship between efficiency and demand. Space reduces or smoothes the impact of efficiency on demand. In other words, the demand index of one school depends not only on its efficiency, but also on the efficiency of its neighboring schools.

In addition, the two main spatial autoregressive parameters (ρ and λ) are statistically significant, thus confirming the previous conclusions from the tests. Specifically, we find a strong and positive relationship between the demand index of neighboring schools and the demand from the unit under assessment (ρ = 0.88***). That means the area or zone in which the school is operating is important in the parents' decision. For instance, if neighboring schools have a high level of demand, the demand for my school can also be higher as a consequence of the spatial spillover effect. If we turn to the second autoregressive parameter (λ = -1.8***) we can conclude that there are omitted variables that vary systematically over space which are relevant and could negatively affect the school demand index. These types of non-crucial variables may be tangible or intangible assets that neighboring schools might have and can negatively impact on the demand index of the school under assessment (such as new furniture, a computer room or extracurricular classes). Finally, the autoregressive parameter of the independent variable (α = 2.29) is not significant. Summarizing, the school's demand index depends on the efficiency of the school, the efficiency of neighboring schools, the area where it is operating and some other non-crucial variables that can be systematic in other schools. This model is consistent as the common factors test is positive and significant (LR-COMFAC = 53.09***).

The last part of this article focuses on whether differences exist between school demand indexes by type of municipality. As seen above in the descriptive statistics (Table 7) Catalonia is composed of very diverse municipalities. Rural municipalities are likely to be further away from city centers and this can have an effect on the school demand and the options available to families. In a previous study (López-Torres and Prior, 2013) we found a significant negative relationship between the concentration index (measured by Herfindahl index) and parents' demand in large municipalities. However, this relationship was not significant in rural towns. We now want to test whether space exerts the same moderating effect in neighboring schools as we found before. To do so, we carry out a sample division following the Eurostat criterion (limit of 5,000 inhabitants). Table 14 lists the results of the spatial contrasting tests.

< Table 14 around here >

The results are consistent with our intuition and add robustness to those obtained in our previous study. We find no spatial dependence in rural municipalities, while we detect residual and substantive autocorrelation structures in urban municipalities. As we can see in Table 14, none of the tests is significant in the case of rural municipalities. This leads us to conclude that the demand for a school in a rural area depends solely on how that school is managed, finding no spatial effect of neighboring schools. This is due to the remoteness of these municipalities on the map (Figure 5).

< Figure 5 around here >

5. CONCLUSIONS AND POLICY IMPLICATIONS

The main goal of this paper was to analyze the relationship between the level of technical quality of public schools (measured by the efficiency score) and the school demand index, paying particular attention to the role played in this context by spatial effects. Our sample consists of 1,695 primary public schools in Catalonia (Spain) which is a considerably wide geographic setting including almost all available public schools in Catalonia (81% of all Catalonian schools). We excluded schools offering special education only, and those for which there were no available data on the students' results. This paper endeavors to respond to our main purpose by presenting a specific approach that distinguishes itself from the previous literature in two major aspects. First, this is the first study to offer an analysis of the role played by geographic location in explaining the spatial distribution of the school demand index in the Spanish context. Second, unlike previous analyses in the school literature, from a methodological perspective our paper applies spatial econometric techniques (Anselin 1988a) that allow us to capture the spatial characteristics of the data and the influence of geographic proximity in shaping the school demand index within Catalonia. This approach is particularly useful in the regional context as SE has become such a prominent topic in the recent related literature (Anselin, 2009).

Our findings reveal important differences in school demand across Catalonia. In addition, the empirical evidence reveals the presence of positive spatial autocorrelation. This implies that school demand is not randomly distributed across space. In contrast, physically adjacent zones tend, on the whole, to exhibit a similar demand index. Indeed, several clusters of regions with similar values to the study variable were detected, but distinct from the neighboring zones. The groupings of regions with a significantly high school demand are situated in big cities (for instance, the city centers of Barcelona and Gerona). On the other hand, the clusters characterized by a low demand index are located in the most remote municipalities (the more distant towns in the provinces of Lérida and Tarragona). The analysis performed in this paper highlights the importance of spatial effects in explaining the spatial distribution of the school demand across Catalonian public schools.

In order to strengthen these findings, we carried out a causal analysis of the observed regional differences. Bearing in mind the consequences of ignoring the presence of spatial dependence, we estimated a model incorporating a spatial autoregressive structure in the dependent variable (spatial lag model) and in the error term (spatial error model). It is important to note here that, as far as we are aware, this is the first time a spatial model has been used to explain school demand in any geographic setting. We have found a few articles in the literature that analyze education-related issues using SE techniques, but none of them takes into account the school demand index (e.g., Zanzig, 1997; Marlow, 2000; Hoxby; 2000; Millimet and Rangaprasad, 2006; Gu, 2012a, b).

The estimated model indicates that the more inefficient a school is, the less demand it receives from parents. We find that school location exerts a moderating effect in the relationship between efficiency and demand, especially in urban municipalities. Space reduces or smoothes the impact of efficiency on demand. In other words, the demand index of one school depends not only on its efficiency, but also on the efficiency of neighboring schools. In addition, it should be pointed out that the results obtained clearly show the importance of spatial effects in explaining the regional distribution of school demand. The empirical evidence also indicates that the transmission of spatial spillover effects across schools belonging to different neighboring areas is relevant. That means the zone in which the school is operating is important to the parents' decision and this affects the demand index.

These results might have significant implications. First, the paper contributes to the current literature as it uses a robust methodological approach, scarcely applied in the literature to date, to analyze school efficiency and school demand focusing on location. Second, it also provides valuable information for public authority decision makers facilitating the implementation of improvement programs in less demanded schools. Thus, it can contribute to higher levels of school quality, motivation, and competition within the system. In this context, the magnitude of territorial imbalances in school demand should encourage Spanish policy makers to introduce additional efforts to reduce the existing differences among the regions by considering the following scenario: taking into account the relevance of spatial effects in this setting, a selective policy to encourage the adoption of innovative teaching plans should be developed at regional level. Thus, an active school quality policy put into practice in a specific neighborhood might not only affect the number of schools in that area, but might also influence the school demand in adjacent zones.

Despite these implications the paper has some limitations that should be noted. In particular, the spatial autocorrelation observed in our study may be partially caused by other geographically correlated factors not included in the analysis. The spatial model reveals the existence of other non-crucial omitted variables that vary systematically over space, and which are relevant and could negatively affect the school demand index. These variables can refer to tangible or intangible assets that neighboring schools might have and can negatively impact the demand index of the school under assessment. Further research is required on this point. Another limitation is the lack of student level data, which prevented us from measuring the first part of the analysis, the efficiency score, in greater depth. Likewise, the availability of information for several years would have allowed us to study the evolution over time of neighboring disparities in school demand across Catalonian regions. If we can obtain these data, it will be possible to model students' and schools' behaviors in space and time, and use the results of such models to gain information about trends and spatial spillover effects at an individual, school and regional level.

Acknowledgements

The authors are grateful to the *Consell Superior d'Avaluació del Sistema Educatiu de Catalunya* for providing the data, with a special mention to Dr. Paquita Grané-Terradas. The authors also acknowledge the financial support of the Spanish Ministerio de Ciencia e Innovación (ECO2010-18967/ECON) and the FPU grant (number 12/01341). We also thank Mika Kortelainen for allowing us to use the specific order-*m* code in R software, and Abaghan Ghahraman (PhD Student) for helping us with the computational part.

6. REFERENCES

Anselin, L. (1988a). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers. The Netherlands.

Anselin, L. (1988b). "Lagrange multiplier test diagnostic for spatial dependence and spatial heterogeneity". *Geographical Analysis*, 20(1), 1-17.

Anselin, L. (2000). "Spatial Econometrics" in B. Baltagi (Ed.). *Companion to Econometrics*. Basil Blackwell, Oxford, UK.

Anselin, L. (2007). "Spatial econometrics in RSUE: retrospect and prospect". *Regional Science and Urban Economics*, 37(4), 450-456.

Anselin, L. (2009). "Thirty years of spatial econometrics". *Papers in Regional Science*, 89(1), 3-25.

Anselin, L. and Rey, S.J. (1997). "Introduction to the special issue on spatial econometrics". *International Regional Science Review*, 20(1, 2), 1-8.

Arbia, G. (2001). "Modeling the geography of economic activities on a continuous space". *Papers in Regional Science*, 80, 11-424.

Arbia, G. (2011). "A lustrum of SEA: Recent research trends following the creation of the Spatial Econometrics Association". *Spatial Economic Analysis*, 6(4), 377-395.

Badin, L., Daraio, C., and Simar, L. (2010). “Optimal bandwidth selection for conditional efficiency measures: A data-driven approach”. *European Journal of Operational Research,* 201(2), 633-640.
Barrow, L. (2002). “School choice through relocation: evidence from the Washington, D.C. area”. *Journal of Public Economics,* 86, 155-189.
Belsley, D., Kuhn, E. and Welsh, R. (1980). *Regression diagnostic identifying influential data and source of collinearity*. New York: John Wiley.
Bera, A.K. and Yoon, M.J. (1992). “Simple diagnostic tests for spatial dependence”. University of Illinois. Department of Economics.
Burridge, P. (1980). “On the Cliff-Ord test for spatial autocorrelation”. *Journal of the Royal Statistical Society B,* 42, 107-108.
Cazals, C., Florens, J. P. and Simar, L. (2002). “Nonparametric frontier estimation: A robust approach”. *Journal of Econometrics*, 106, 1-25.
Cliff, A. and Ord, J. (1972). “Testing for spatial autocorrelation among regression residuals”. *Geographical Analysis,* 4, 267-284.
Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation*. London, Pion.
Cordero, J.M., Pedraja, F. and Salinas, J. (2008). “Measuring Efficiency in Education: An Analysis of Different Approaches for Incorporating Non-Discretionary Inputs”. *Applied Economics,* 40(10), 1323-1339.
Cordero, J.M., Pedraja, F. and Santín, D. (2010). “Enhancing the Inclusion of Non-Discretionary Inputs in DEA”. *Journal of the Operational Research Society,* 61, 574-584.
Corman, H. (2003). “The effects of state policies, individual characteristics, family characteristics, and neighborhood characteristics on grade repetition in United States”. *Economics of Education Review,* 22, 409-420.
Daraio, C. and Simar, L. (2005). “Introducing environmental variables in nonparametric frontier models: A probabilistic approach”. *Journal of Productivity Analysis,* 24, 93-121.
Daraio, C., Simar, L. and Wilson, P. (2010). “Testing whether two-stage estimation is meaningful in non-parametric models of production”. *working paper.*
De Witte, K. and Kortelainen M. (2013). “What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables”. *Applied Economics*, 45, 2401-2412.
Elhorst, J.P. (2012). “Applied Spatial Econometrics: Raising the Bar”. *Spatial Economic Analysis,* 51(1), 9-28.
Feng, L., and Sass, T.R. (2013). “What makes special-education teachers special? Teacher training and achievement of students with disabilities”. *Economics of Education Review,* 36, 122-134.
Florax, R. and Folmer, H. (1992). “Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators”. *Regional Science and Urban Economics,* 22, 404-432.
Getis, A. and Ord, J. (1992). “The analysis of spatial association by use of distance statistics”. *Geographical Analysis,* 24, 189-206.
Grosskopf, S., Hayes, K.J., Taylor, L.L. and Weber, W.L. (2013). “Centralized or Decentralized Control of Resources? A Network Model”. Paper presented at EWEPA’13 conference, Helsinki (Finland).
Gu, J. (2012a). “Spatial dynamics and determinants of county-level expenditure in China”. *Asia Pacific Education Review,* 13(4), 617-634.
Gu, J. (2012b). “Spatial recruiting competition in Chinese higher education system”. *Higher Education,* 63(2), 165-185.

Hanushek, E.A. (1986). "The economics of schooling: Production and efficiency in public schools". *Journal of Economic Literature*, 24, 1141-1177.

Hanushek, E.A. (2003). "The failure of input-based schooling policies". *The Economic Journal*, 113, 64-98.

Hanushek, E.A., Rivkin, S.G. and Taylor, L.L. (1996). "Aggregation and the Estimated Effects of School Resources". *The Review of Economics and Statistics,* 78(4), 611-627.

Hanushek, E.A. and Kimko, D.D. (2000). "Schooling, labor-force quality, and the growth of nations". *American Economic Review,* 90, 1184–1208.

Harrison, J., Rouse, P. and Armstrong, J. (2012). "Categorical and continuous non-discretionary variables in data envelopment analysis: A comparison of two-stage models". *Journal of Productivity Analysis,* 37(3), 261-276.

Hoxby, C.M. (2000). "Does competition among public schools benefit students and taxpayers?" *American Economic Review*, 90, 1209–1238.

Johnes, J. (2006). "Data Envelopment Analysis and its Application to the Measurement of Efficiency in Higher Education". *Economics of Education Review*, 25, 273-288.

Johnson, A.L. and Ruggiero, J. (2011). "Nonparametric Measurement of Productivity and Efficiency in Education". *Annals of Operations Research*, forthcoming. DOI 10.1007/s10479-011-0880-9.

Krugman, P. (1991a). "Increasing Return and Economic Geography". *Journal of Political Economy,* 99, 438-499.

Krugman, P. (1991b). *Geography and Trade*. MIT Press, Cambridge MA.

Krugman, P. (1998). "What's New about the New Economic Geography?". Oxford *Review of Economic Policy*, 14(2), 7-17.

López-Torres, L. and Prior, D. (2013). "Do parents perceive the technical quality of public schools? An activity analysis approach". *Regional and Sectoral Economic Studies*, 13(3), 39-60.

Mancebón, M.J. and Muñiz, M. (2008). "Private versus Public High Schools in Spain: Disentangling Managerial and Programme Efficiencies". *Journal of the Operational Research Society,* 59(7), 892-901.

Marlow, M.L. (2000). "Spending, school structure, and public education quality: evidence from California". *Economics of Education Review,* 19, 89–106.

Millimet, D.L. and Rangaprasad, V. (2006). "Strategic competition amongst public schools". *Regional Science and Urban Economics,* 37, 199-219.

Moran, P. (1948). "The interpretation of statistical maps". *Journal of the Royal Statistical Society B*, 10, 243-251.

Moreno, R. and Vayá, E. (2000). *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*. UB 44 manuals, Edicions Universitat de Barcelona.

Muñiz, M. (2002). "Separating Managerial Inefficiency and External Conditions in Data". *European Journal of Operational Research,* 143(3), 625-643.

Muñiz, M., Paradi, J., Ruggiero, J. and Yang, Z. (2006). "Evaluating Alternative DEA Models Used to Control for Non-Discretionary Inputs". *Computers and Operations Research,* 33(5), 1173-1183.

Ngware, M. W., Oketch, M. and Ezeh, A. C. (2011). "Quality of Primary Education Inputs in Urban Schools: Evidence from Nairobi". *Education and Urban Society*, 43(1), pages 91-116.

Opdenakker, M. C. and Van Damme, J. (2001). "Relationship between School Composition and Characteristics of School Process and their Effect on Mathematics Achievement". *British Educational Research Journal*, 27(4), 407-432.

Ouchi, W. G. (2003). “*Making schools work: A revolutionary plan to get your children the education they need”.* Simon and Schuster. New York.
Paelinck, J. and Klaassen, L. (1979). *Spatial Econometrics*, Saxon House, Farnborough.
Pinkse, J. and Slade, M.E. (2010). “The future of spatial econometrics”. *Journal of Regional Science*, 50(1), 103-117.
Ray, S.C. (1991). “Resource Use Efficiency in Public Schools: A Study of Connecticut Data”. *Management Science,* 37(12), 1620-1628.
Rubenstein, R., Schwartz, A.E., Stiefel, L. and Bel Hadj Amor, H. (2007). “From districts to schools: The distribution of resources across schools in big city school districts”. *Economics of Education Review,* 26(5), 532-545.
Ruggiero, J. (1998). “Non-Discretionary Inputs in Data Envelopment Analysis”. *European Journal of Operational Research*, 111, 461-469.
Silva-Portela, M.C.A., Camacho, A.S. and Keshvari, A. (2013). “Assessing the Evolution of School Performance and Value-Added: Trends over Four Years”. *Journal of Productivity Analysis,* 39, 1-14.
Simar, L. (2003). “Detecting outliers in frontiers models: A simple approach”. *Journal of Productivity Analysis,* 20, 391-423.
Simar, L. and Wilson, P. (2007). “Estimation and Inference in Two-Stage, SemiParametric Models of Production Processes”. *Journal of Econometrics,* 136(1), 31-64.
Smith, P. and Mayston, D. (1987). “Measuring Efficiency in the Public Sector”. *OMEGA International Journal of Management Science,* 15(3), 181-189.
Thieme, C., Prior, D. and Tortosa-Ausina, E. (2013). “A Multilevel Decomposition of School Performance Using Robust Nonparametric Frontier”. *Economics of Education Review*, 32, 104-121.
Wilcoxon, F. (1945). “Individual Comparisons by Ranking Methods”. *Biometrics,* 1, 80-83.
Zanzig, B.R. (1997). “Measuring the impact of competition in local government education markets on the cognitive achievement of students”. *Economics of Education Review,* 16, 431–444.

Table 1. Global and local statistics of spatial association

	Statistics	***Features***	***Meaning***
Global	**Moran's I (1948)** $I = \frac{N}{S_0} * \frac{\sum_{ij}^{n} w_{ij} \, (x_i - \bar{x})\left(x_j - \bar{x}\right)}{S_{i1}^{N} \, (x_i - \bar{x})^2}$ $i \neq j$	x_i= value of the variable x in region i. $\bar{x}$= sample mean of the variable x. w_{ij}= weights of the matrix W. N= sample size. $S_0 = \sum_i \sum_j w_{ij}$.	After standardization: Z (I) > 0 and significant: positive autocorrelation. Z (I) < 0 and significant: negative autocorrelation.
Global	**Getis and Ord's G(d) (1992)** $G(d) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \, (d) x_i x_j}{\sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j}$ $i \neq j$	Two pairs of regions *i* and *j* are neighbors if they are within a predetermined distance *d* ($w_{ij}(d)$ = 1 or 0 otherwise).	Z (G (d)) > 0 and significant: higher concentration values. Z (G (d)) < 0 and significant: lower concentration values.
Local	**Moran's Local I** $I_i = \frac{z_i}{\sum_i z_i^2 / N} \sum_{j \in J_i} w_{ij} \, z_j$ $i \neq j$	z_i= value of the normalized variable corresponding to the region *i*. J_i= set of neighboring regions to *i*.	After standardization: Z(I_i) > 0 and significant: cluster or similar values around *i*. Z(I_i) < 0 and significant: cluster or dissimilar values around *i*.
Local	**Getis and Ord's Local G(d)** $New - G_i^* = \frac{\sum_{j=1}^{N} w_{ij} x_j - W_i^* \bar{x}}{s \left\{ \frac{[N S_{1i}^* - W_i^{*2}]}{N-1} \right\}^{1/2}}$ $i \neq j$	$W_i^* = W_i + w_{ii}$ $S_{1i}^* = \sum_j w_{ij}^2$ $s^2 = \frac{1}{N-1} \sum_j \left(x_j - \bar{x}\right)^2$	$New - G_i^*$> 0 and significant: cluster or similar and higher values around *i*. $New - G_i^*$< 0 and significant: cluster or similar and lower values around *i*.

Source: Compiled from Moreno and Vayá (2000, 33-44).

Table 2: Some spatial autocorrelation statistics in the regression model

Spatial dependence type	***Test type***	***Statistics***	***Features***
Residual	Ad-hoc	**Moran's I (Cliff and Ord, 1972)** $I = \frac{N}{S}\frac{e'We}{e'e}$	e = OLS residues. N = sample size. S = sum of all w_{ij} W matrix.
	ML	**LM-ERR (Burridge, 1980)** $LM - ERR = \frac{\left[e'We/_{s^2}\right]^2}{T_1}$	s^2= estimation of residual variance. T_1= $tr\ (W'W + W^2)$.
		LM-EL (Bera and Yoon, 1992) $LM - EL = \frac{\left[\frac{e'We}{s^2 - T_1(RJ_{\rho-\beta})^{-1}e'Wy/s^2}\right]^2}{\left[T_{1-}T_1^2(RJ_{\rho-\beta})^{-1}\right]}$	$RJ_{\rho-\beta} = [T_1 + (WX\beta)'M(WX\beta)/s^2$ $M = I - X(X'X)^{-1}X'$
Substantive	ML	**LM-LAG (Anselin, 1988b)** $LM - LAG = \frac{[e'Wy/s^2]^2}{RJ_{\rho-\beta}}$	All the terms are known.
		LM-LE (Bera and Yoon, 1992) $LM - LE = \frac{\left[\frac{e'Wy}{s^2} - e'We/s^2\right]^2}{RJ_{\rho-\beta} - T_1}$	All the terms are known.
Both	ML	**SARMA Test (Anselin, 1988b)** $SARMA = \frac{\left[\frac{e'Wy}{s^2} - e'We/s^2\right]^2}{RJ_{\rho-\beta} - T_1} + \frac{(\frac{e'We}{s^2})^2}{T_1}$	All the terms are known.

Source: compiled from Moreno and Vayá (2000, 38-44).

Table 3. Description of variables for efficiency study

Type	*Variable*		*Description*
Discretionary input	X_1	Students	Total number of regular students.
	X_2	Teachers	Total number of teachers at the school
Non-discretionary factor	X_{nd1}	Socio-economic level	Employment status of families (mean). 0. Unclassifiable (housewives, unemployed). 1. Other workers (commercial, administrative). 2. Middle managers. 3. Technicians, professionals. 4. General managers. 5. Entrepreneurs with/without employees.
	X_{nd2}	Educational level	Parents' education (mean). 0. No education 1. Primary education. 2. Secondary Education. 3. Intermediate Professional Training. 4. Baccalaureate (post-compulsory school). 5. Higher Professional Training, 6. Graduate.7. Post-Graduate. 8. PhD.
	X_{nd3}	Innovation	Availability of Innovation Projects (0. No 1. Yes)
	X_{nd4}	Unemployed	Number of parents unemployed.
	X_{nd5}	Grants	Percentage of applied grants.
	X_{nd6}	Economic needs	Percentage of students with some economic need due to the employment situation at home.
	X_{nd7}	Immigrants	Percentage of non-Spanish students.
	X_{nd8}	Late incorporations	Percentage of newly incorporated students (halfway through the year).
	X_{nd9}	New students	Percentage of newly incorporated students (at the beginning of an academic year).
	X_{nd10}	Students' mobility	Percentage of newly incorporated students plus drop-out students (New enrollments + Exits / Total enrollment).
	X_{nd11}	Educational needs	Percentage of students with special educational needs (additional supporting classes).
	X_{nd12}	New teachers	Percentage of newly incorporated teachers (at the beginning of an academic year).
	X_{nd13}	Dropout rate	Percentage of student absences during the academic year (students absent more than 75% of all days).
	X_{nd14}	Stability	Average number of years a principal holds his/her position.
Output	Y_1	Grades	Average test mark obtained by the school's students in a general sixth grade test.
	Y_2	Pass rate	Total enrolled – repeaters – absentee students (with more than 75% absences each quarter).

Source: Self devised.

Table 4. Summary statistics: Catalonian Public Schools, 2009/2010

Variable	*N*	*Min*	Q_{25}	*Mean*	*S.D.*	*Median*	Q_{75}	*Max*
X_1	1,695	4.00	113.00	259.03	165.74	228.00	442.00	730.00
X_2	1,695	1.00	12.00	21.16	11.39	20.00	32.00	52.00
Y_1	1,695	30.81	67.04	71.29	8.49	71.00	76.56	95.89
Y_2	1,695	4.00	111.00	255.55	163.80	226.00	416.00	727.00
X_{nd1}	1,695	0.00	2.00	1.78	0.48	2.00	2.00	5.00
X_{nd2}	1,695	2.00	5.00	5.17	0.83	5.00	6.00	8.00
X_{nd3}	1,695	0.00	0.00	0.51	0.50	1.00	1.00	1.00
X_{nd4}	1,695	1.00	29.00	68.26	48.15	68.00	97.00	265.00
X_{nd5}	1,695	0.00	9.00	19.61	15.09	16.00	27.00	100.00
X_{nd6}	1,695	0.00	0.00	3.64	7.99	0.00	4.00	88.00
X_{nd7}	1,695	0.00	5.00	15.66	15.11	11.00	22.00	85.00
X_{nd8}	1,695	0.00	0.00	1.95	3.86	0.00	2.00	44.00
X_{nd9}	1,695	0.00	0.00	1.62	3.73	0.00	2.00	48.00
X_{nd10}	1,695	0.00	2.00	38.76	7.89	5.00	9.00	89.00
X_{nd11}	1,695	0.00	0.00	2.28	3.37	1.00	3.00	33.00
X_{nd12}	1,695	0.00	1.00	2.91	2.81	2.00	3.00	33.00
X_{nd13}	1,695	0.00	0.00	0.83	2.29	0.00	1.00	35.00
X_{nd14}	1,695	1.00	4.00	7.29	4.66	6.00	9.00	33.00

Source: Self devised.

Table 5. Correlation Matrix

	X_1	X_2	Y_1	Y_2	X_{nd1}	X_{nd2}	X_{nd3}	X_{nd4}	X_{nd5}	X_{nd6}	X_{nd7}	X_{nd8}	X_{nd9}	X_{nd10}	X_{nd11}	X_{nd12}	X_{nd13}	X_{nd14}
X_1	1																	
X_2	.97**	1																
Y_1	.98**	.94**	1															
Y_2	.96**	.97**	.99**	1														
X_{nd1}	-.14**	-.19**	-.06*	-.14**	1													
X_{nd2}	0.02	-.06*	.12**	0.03	.53**	1												
X_{nd3}	.09**	.09**	.09**	.09**	0.02	0.02	1											
X_{nd4}	.81**	.80**	.74**	.80**	-.36**	-.30**	.05*	1										
X_{nd5}	-.14**	-.06**	-.20**	-.14**	-.32**	-.53**	0.02	.09**	1									
X_{nd6}	0.05	.1**	0.00	0.04	-.27**	-.37**	-0.01	.18**	.37**	1								
X_{nd7}	.09**	.18**	0.02	.09**	-.47**	-.48**	0.02	.24**	.43**	.39**	1							
X_{nd8}	0.03	.06**	0.00	0.03	-.21**	-.22**	0.01	.13**	.15**	.18**	.37**	1						
X_{nd9}	.08**	.12**	.05*	.08**	-.19**	-.18**	0.03	.13**	.15**	.11**	.34**	.24**	1					
X_{nd10}	-.13**	-.07**	-.18**	-.13**	-.32**	-.32**	0.02	0.03	.25**	.19**	.44**	.18**	.15**	1				
X_{nd11}	-0.02	0.00	-0.03	-0.02	-0.03	-.10**	0.03	0.00	.12**	.19**	.08**	.11**	0.02	-0.01	1			
X_{nd12}	-.39**	-.39**	-.40**	-.39**	0.01	-.07**	-.057*	-.27**	.07**	-0.04	-.06*	-0.03	-.07**	.19**	0.04	1		
X_{nd13}	.22**	.24**	.19**	.20**	-.15**	-.17**	0.01	.23**	.11**	.13**	.20**	.12**	.09**	.09**	.07**	-.06*	1	
X_{nd14}	.08**	.11**	.09**	.08**	0.01	-.07**	0.04	0.02	.08**	.05*	.08**	0.01	0.03	0.00	0.04	-.15**	0	1

Source: Self devised.

Table 6. Description of spatial variables

Type	*Variable*		*Description*
Independent	θ_{cond}	Conditional Efficiency score	Efficiency index; reflects the school's performance controlling for environmental variables
Dependent	***Demand***	School demand	Enrollment applications / places offered
Necessary to separate the sample	***Popul***	Population	Number of inhabitants

Source: Self devised.

Table 7. Summary statistics for spatial variables

Variable	N	Min	Q_{25}	Mean	S.D.	Median	Q_{75}	Max
θ_{uncond}	1,695	0.97	0.98	1.12	0.04	1.09	1.14	1.20
θ_{cond}	1,695	0.98	1.01	1.20	0.01	1.1	1.21	1.25
Demand	1,695	0	0.64	0.84	0.47	0.84	1.24	8.00
Population	1,695	102	2,235	193,393.87	466,638.61	16,341	253,782	1,615,908
Rural	567	102	508	1,487.22	1,219.58	1,029	3,479	4,970
Urban	1,128	5,016	16,341	289,858.59	547,230.24	51,912	1,615,908	1,615,908

Source: Self devised.

Table 8. Efficiency estimations

Variable	N	Min	Q_{25}	Mean	S.D.	Median	Q_{75}	Max
θ_{uncond}	1,695	0.97	0.98	1.12	0.04	1.09	1.14	1.20
θ_{cond}	1,695	0.98	1.01	1.20	0.01	1.1	1.21	1.25

Source: Self devised.

Table 9. Wilcoxon test results[15]

	Wilcoxon Signed-Ranks Test (1)		
	Statistic	**0.1% significance level**	**Decision**
Socio-economic level	-19.52***	H_0 rejected	Non-separable
Educational level	-10.42***	H_0 rejected	Non-separable
Innovation	-17.47***	H_0 rejected	Non-separable
Unemployed	-26.59***	H_0 rejected	Non-separable
Grants	-27.13***	H_0 rejected	Non-separable
Economic needs	-13.25***	H_0 rejected	Non-separable
Immigrants	-27.70***	H_0 rejected	Non-separable
Late incorporations	-16.79***	H_0 rejected	Non-separable
New students	-8.91***	H_0 rejected	Non-separable
Students' mobility	-23.53***	H_0 rejected	Non-separable
Educational needs	-15.87***	H_0 rejected	Non-separable
New teachers	-8.83***	H_0 rejected	Non-separable
Dropout rate	-24.04***	H_0 rejected	Non-separable
Stability	-24.47***	H_0 rejected	Non-separable
Unidentified	-0.97	H_0 not rejected	Separable
Unemployed/ID found	-0.75	H_0 not rejected	Separable
Age	-0.92	H_0 not rejected	Separable
Changes	-0.88	H_0 not rejected	Separable
Teachers' absenteeism	-0.84	H_0 not rejected	Separable
NOTE	(1) The hypothesis evaluated with the nonparametric Wilcoxon matched-pairs signed-ranks test (1945) is whether or not the median of the difference scores equals zero in the underlying populations represented by the sampled experimental conditions. If a significant difference is obtained, it indicates a high likelihood that the two sampled conditions represent two different populations. The Wilcoxon matched-pairs signed-ranks test is based on the assumption that the distribution of the difference scores in the populations represented by the two samples is symmetric about the median of the population of difference scores.		

Source: Self devised

Table 10. Significance test

Variable	***Statistic***	***Impact on efficiency***	***Impact on potential school outcomes***
Socio-economic level	0.07***	Increases	Favorable
Educational level	0.08***	Increases	Favorable
Unemployed	0.09***	Increases	Favorable
Grants	0.01***	Increases	Favorable
Economic needs	0.01		
Immigrants	0.01		
Late incorporations	0.02		
New students	0.01		
Students' mobility	0.03		
Educational needs	-0.01*	Decreases	Unfavorable
New teachers	0.01		
Dropout rate	-0.05*	Decreases	Unfavorable
Stability	0.01		

Source: Self devised.

[15] As can be seen in Table 9, we initially had more than 14 non-discretionary factors, i.e., those which were separable. We decided not to include them in section 3.3 to avoid confusions.

Table 11. Moran's I and Getis and Ord's G global tests

	Moran's I		Getis and Ord's G	
Variable	***Statistic***	***S.D.***	***Statistic***	***S.D.***
Demand	0.3403***	8.9141	0.5974***	7.3894
NOTE	*W*=Distance Matrix			

*** = 0.1% significance level
Source: Self devised

Table 12. OLS estimation and Lagrange multiplier tests for spatial dependence

Variable	***OLS***	
	Estimated coefficient	**p-value**
Constant	3.96***	0.00
β	-3.11**	0.01
LM ERR	71.35***	0
LM LAG	72.29***	0
RLM ERR	8.36**	0.004
RLM LAG	14.69***	0
SARMA	73.07***	0

Notes: the dependent variable is the school demand index.
The independent variable is the conditional efficiency score.
** = 1% significance level
*** = 0.1% significance level
Source: Self devised

Table 13. Regression analysis: Matrix type: distance matrix

Independent variable = conditional efficiency					
	BASE MODEL	MODEL 1	MODEL 2	MODEL 3	MODEL 4
Variable/Model	**OLS**	**LAG**	**ERR**	**SARAR**	**SARARMIX**
	Estimated coefficients (p-value)				
Constant	3.96***	3.43**	3.79**	5.20***	0.74
	0.00	0.00	0.00	0.00	0.85
β	-3.11**	-2.98*	-2.94*	-2.93*	-2.85*
	0.01	0.01	0.01	0.01	0.01
ρ		0.48***		-1.80***	0.88***
		0.00		0.00	0.00
λ			0.48***	0.89***	-1.80***
			0.00	0.00	0.00
α					2.29
					0.57
LM test for residual autocorrelation		27.77***			
		0.00			
LN L		-1121.59	-1121.68	-1109.57	-1109.53
TESTS					
LR-COMFAC				53.02***	53.09***
				0.00	0.00
AIC		2251.20	2251.40	2221.10	2209.10

Notes: the dependent variable is the school demand index.
*= 5% significance level
** = 1% significance level
*** = 0.1% significance level

Source: self devised

Table 14. Tests to detect spatial dependence by areas

	Rural		Urban	
Tests	**Statistic**	**p-value**	**Statistic**	**p-value**
Moran's I	0.02	0.11	0.06*	0.04
Getis and Ord's G	0.02	0.05	0.09	0.26
LM ERR	2.07	0.15	2.80	0.09
LM LAG	2.08	0.15	3.34	0.07
RLM ERR	0.02	0.90	5.64*	0.02
RLM LAG	0.04	0.85	6.18*	0.01
SARMA	2.10	0.35	8.98*	0.01

* = 5% significance level

Source: self devised

Figure 1. Spatial distribution of school demand

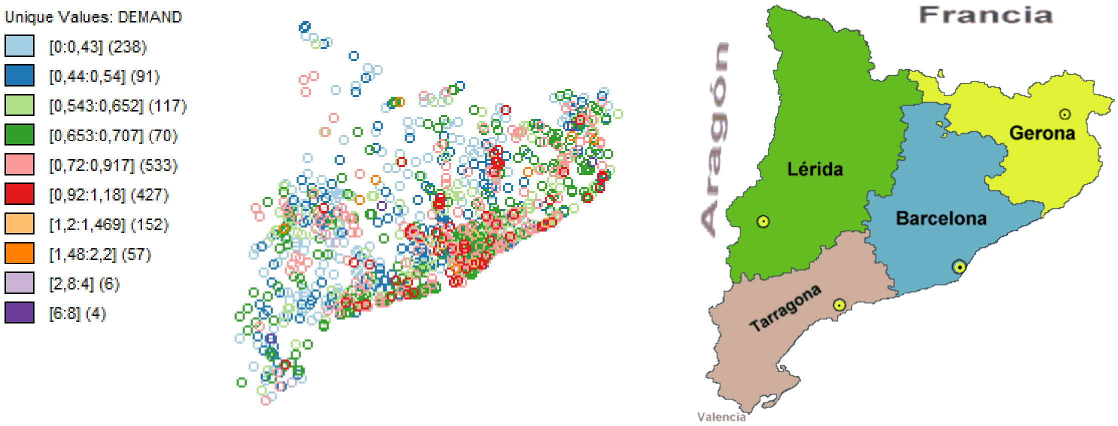

Figure 2. School demand box map

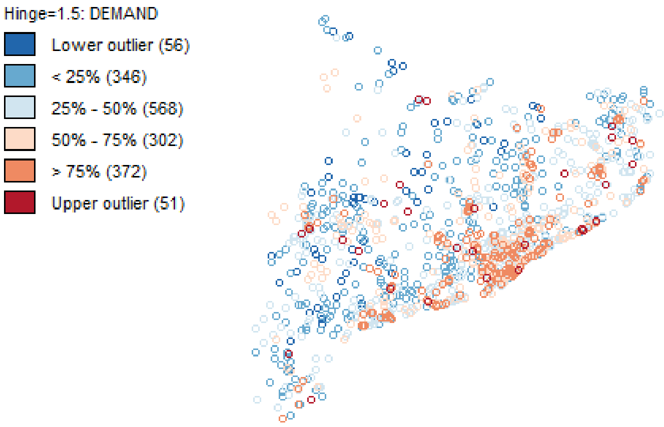

Figure 3. Moran's *Scatterplot*

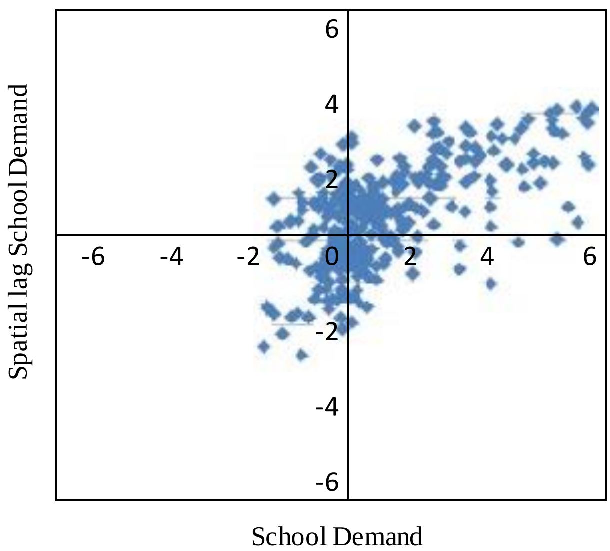

Figure 4. School demand *scatter* map

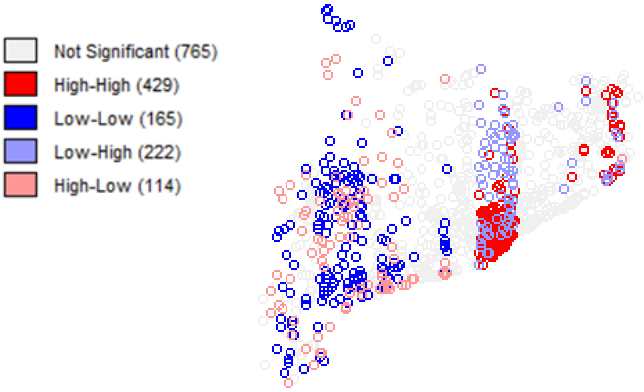

Figure 5. Spatial distribution of school demand in rural areas