WP-EC 2010-02

Output complexity, environmental conditions, and the efficiency of municipalities: a robust approach*

M. Teresa Balaguer-Coll, Diego Prior and Emili Tortosa-Ausina**

Abstract

Over the last few years, many studies have analyzed the productive efficiency of local governments in different countries. An accurate definition of their output bundles -i.e., the services and facilities they provide to their constituencies- is essential to this research. However, several difficulties emerge in this task. First, since in most cases the law only establishes the minimum amount of services and facilities to provide, it may well be the case that some municipalities go beyond the legal minimum and, consequently, might be labeled as inefficient when compared to other municipalities which stick to the legal minimum. Second, municipalities face very different environmental conditions, which raises some doubts about the plausibility of an unconditional analysis. This study tackles these problems by proposing a metafrontier analysis in which the efficiency of municipalities is evaluated after splitting them into clusters according to various criteria (output mix, environmental conditions, size). We perform our estimations using order-m frontiers, given their robustness to outliers and immunity to the curse of dimensionality. We provide an application to Spanish municipalities, and results show that both output mix and, more especially, environmental conditions, should be controlled for, since efficiency differences between municipalities in different groups are notable.

Keywords: efficiency, environmental conditions, local government, metafrontier, order-*m*
JEL Classification: D24, D60, H71, H72

Resumen

Durante los últimos años muchos trabajos han venido analizando la eficiencia productiva de las corporaciones locales de una gran variedad de países. Para este tipo de estudios resulta crucial una definición precisa de los servicios e infraestructuras que los municipios proporcionan a sus ciudadanos. Sin embargo, esta tarea presenta varias dificultades. En primer lugar, dado que en muchas circunstancias la ley únicamente establece los servicios mínimos que debe proporcionar un municipio, puede darse el caso de que algunos municipios vayan más allá de este mínimo legal y, consecuentemente, sean clasificados como ineficientes al compararlos con otros municipios que se ciñen al mínimo. En segundo lugar, las corporaciones locales operan en condiciones ambientales muy dispares, lo cual genera dudas acerca de la factibilidad de un análisis incondicional. Este trabajo aborda estas cuestiones proponiendo un análisis metafrontera en el que la eficiencia de las corporaciones locales se evalúa tras clasificarlas en distintos grupos de acuerdo con criterios múltiples (output mix, condiciones ambientales, tamaño). Las estimaciones son llevadas a cabo utilizando fronteras orden-*m*, debido a la robustez que presentan frente a observaciones atípicas y la inmunidad a la "maldición de la dimensionalidad" (*curse of dimensionality*). Llevamos a cabo una aplicación a los municipios españoles, y los resultados indican que tanto el *output mix* como, sobre todo, las condiciones ambientales, deberían ser tenidas en cuenta al evaluar la eficiencia, pues las diferencias en la eficiencia de los municipios en los distintos grupos son notables.

Palabras clave: eficiencia, condiciones ambientales, gobierno local, metafrontera, orden-*m*

---

* M.T. Balaguer-Coll and D. Prior acknowledge the financial support of the Ministerio de Educación y Ciencia (SEC2003-04770). E. Tortosa-Ausina acknowledges the financial support of Fundació Caixa Castelló-Bancaixa (P1.1B2008-46), and the Ministerio de Ciencia e Innovación (ECO2008-03813/ECON and ECO2008-05908-C02-01/ECON).

** M.T. Balaguer-Coll: Universitat Jaume I. D. Prior: Universitat Autònoma de Barcelona and IESEG School of Management. E. Tortosa-Ausina: Universitat Jaume I and Ivie. Contact author: tortosa@uji.es

1. Introduction

Over the last few years, a wide range of studies has analyzed the productive efficiency of municipalities from multiple perspectives. The empirical evidence now available is increasing, and can be divided into two groups. On the one hand, some studies analyze efficiency in the provision of a specific service such as refusal collection (Brueckner, 1981; Ruggiero, 1995; Bosch et al., 2000). However, in some countries this type of study presents certain disadvantages related to the difficulties in assigning the amount of input usage by each specific service. On the other hand, many studies have considered a global perspective, taking into account that local governments provide their constituencies with a wide variety of services and facilities from the municipal budget. The literature is fairly extensive, yet scattered in time—studies have not emerged abruptly but rather have appeared sporadically over time.[1]

Therefore, this is a relevant topic whose importance has further increased in the last few years in some countries. From a viewpoint of European integration and the Stability and Growth Pact (SGP), it is of utmost importance that all layers of government manage their resources *efficiently*, in order to facilitate and maintain European Economic and Monetary Union. The interest is even higher in certain European countries such as Spain, where municipalities face tighter budget constraints since the passing of the law on budget stability ("Ley General de Estabilidad Presupuestaria", 2001), which establishes mechanisms to control public debt and public spending in pursuit of a balanced budget.

A shared problem faced by the second category of studies referred to in the first paragraph is the difficulty to accurately define, and measure, what it is that local governments produce. In most cases the problems arise due to the impossibility of directly quantifying the supply of public services. The Spanish case is not free from that criticism, although its magnitude is lessened, largely due to the availability of data on most of the public services that municipalities are bound to provide, which also includes information on output quality. However, the law only establishes the *minimum* services and facilities each municipality must provide, which varies according to their population. Nothing prevents a particular municipality from going *beyond* this legal minimum and providing not only more of each compulsory service (such as discretionally increasing the area of public parks), but also providing additional services and facilities whose input usage may be substantial. Such a municipality would be mislabeled as inefficient when compared to other municipalities which stick to the legal minimum. Therefore, we can consider that modelling municipalities' output is difficult; we may even talk about *output complexity* (Haynes, 2003).

However, what is difficult to measure is the amount of services and facilities provided by each municipality *beyond* the legal minimum. Some reports acknowledge this reality (see Vilalta and Mas, 2006), but their results are generally based on surveys carried out on a limited number of municipalities. Previous studies dealing with this reality are, for instance, Bennett and DiLorenzo (1982), Marlow and Joulfaian (1989) or Merrifield (1994). According to some of these papers, the additional costs incurred by many municipalities are quite large, although they also vary a great deal across observations. These studies

[1] See, for instance, Brueckner and Wingler (1984), Deller (1992), Taylor (1995), De Borger and Kerstens (1996), Grossman et al. (1999), Hughes and Edwards (2000) and, more recently, Sampaio de Sousa and Stošić (2005) Balaguer-Coll et al. (2007), or Barankay and Lockwood (2007).

constitute evidence supporting the hypothesis that some municipalities provide their constituencies with a larger amount of services and facilities, not only because of different needs in different constituencies, but also because of the different environmental conditions each municipality faces.

In addition to this, there is a large body of literature on the differing environmental conditions that DMUs (Decision Making Units) face in different contexts—not only in local government. Since the pioneering contributions by Banker and Morey (1986a,b), many studies have analyzed the issue. Most of the ensuing literature has been of an applied nature (see, for instance Bos and Kool, 2006), although several theoretical refinements to the initial methodologies have also been proposed (see, for instance Ruggiero, 2004). This body of literature postulates that environmental conditions may have a strong impact on DMUs' performance. In the case of municipalities, this would imply that different local governments face different constituencies in terms of economic and social conditions, to whose needs municipalities may be more or less responsive irrespective of the amount of services they are obligated to provide. Municipalities also have different characteristics in terms of geography (including, for instance, rugged terrain, or urban sprawl); different sectoral production characteristics (in terms of tourism, etc.); and other characteristics.

We consider municipalities may provide more services and facilities than the legally established minimum because of the different environmental conditions they face. Tourist municipalities may face budget strains due to the highly increased personnel needs they have during their high season. Other municipalities may face higher costs because of urban sprawl. According to Solé-Ollé and Hortas Rico (2008), the urban spatial structure of many Spanish cities has not only an environmental impact, but also a major impact on municipal finances. In other cases, reasons might be more involved, such as wealthy constituencies asking for additional services, or rising expenditures on security because of the rapid population increases experienced by some Spanish cities. As documented by Vilalta and Mas (2006), in a study applied to a sample of the province of Barcelona, more than 30% of municipal expenditures were discretionary. In these circumstances, one may reasonably expect that some municipalities will be mislabeled as inefficient (or, at least, more inefficient than other municipalities) simply because of this wide variety of scenarios. Therefore, it would be more appropriate to compare only local governments facing similar environmental conditions.

This study approaches these problems through a two-stage analysis that firstly analyzes the productive efficiency of municipalities, and secondly divides them into groups according to different classifications. In the second stage we consider three criteria to classify municipalities into different groups. The first one considers clusters according to local governments' output mixes, in order to compare the efficiency of municipalities with similar output bundles; in this way we can control for the fact that some of them might provide services and facilities *beyond* the legal minimum—hence, more complex. The second criterion constructs groups of municipalities for which we include information on environmental conditions. The third criterion forms groups according to size, since the levels of services each municipality must provide hinges on the level of population. In our application to Spanish municipalities, results show that both output mix and the different environments that municipalities face are issues to control for, since the differences between municipalities affiliated to different groups turned out to be statistically significant.

Therefore, in the first stage municipalities are classified into groups according to different criteria given that, according to the hypotheses formulated, local governments in each group might be facing different production opportunities. They respond by making choices from different sets of feasible input-output combinations. These so-called technology sets differ because of differences in available stocks of physical, human and financial capital (e.g., type of machinery, size and quality of the labor force, access to foreign exchange), economic infrastructure (e.g., number of ports, access to markets), resource endowments (e.g., quality of soils, climate, energy resources) and any other characteristics of the physical, social and economic environment in which production takes place. As indicated by O'Donnell et al. (2008), such differences have led efficiency researchers to estimate separate production frontiers for different groups of DMUs.[2]

However, the literature on efficiency analysis has faced severe problems when dealing with the evaluation of DMUs in different groups, which are assumed to have different technologies. Authors such as Battese and Rao (2002) or Battese et al. (2004) argue that the efficiencies of DMUs that operate under a given production technology are not comparable with those of DMUs operating under different technologies. In our setting, this would imply that it is not possible to compare the efficiencies of municipalities in the different groups formed according to our three criteria. Battese and Rao (2002) propose a solution based on the concept of metafrontier[3] in the context of efficiency measurement using stochastic frontier analysis (SFA), which refined Battese et al. (2004). In their paper, Battese and Rao (2002) assume there are two different data-generation mechanisms for the data—one with respect to the stochastic frontier, the other with respect to the metafrontier model. In contrast, Battese et al. (2004) assume that the metafrontier function is an overarching function that encompasses the deterministic components of the stochastic frontier production functions for those DMUs operating under different technologies.

Yet in public sector applications such as the measurement of local government efficiency, most studies have been using nonparametric techniques such as DEA (Data Envelopment Analysis), or its nonconvex version (Free Disposable Hull, FDH), for a variety of reasons (Fox, 2001). O'Donnell et al. (2008) have recently filled this gap in the literature, by extending the metafrontier to DEA and alternative SFA approaches for estimating both metafrontiers and group frontiers. However, their solutions are not entirely satisfactory, mainly because of the *curse of dimensionality* that generally affects efficiency scores obtained using DEA. As indicated by Daraio and Simar (2007), increasing the number of inputs or outputs, or decreasing the number of units being compared, leads to higher efficiencies, simply as a result of a statistical artifact. Multiple applications in disparate fields are affected by this issue (see, among many others Maudos et al., 2002), which is especially severe in fields such as mutual fund evaluation, where difficulties arise in defining the number of inputs and outputs (Joro and Na, 2002). Our approach to deal with this problem is based in the order-m frontier initially proposed by Cazals et al. (2002). We modify their algorithm

[2]For example, separate frontiers have been estimated for universities in Canada (McMillan and Chan, 2006), Australia (Worthington and Lee, 2008) and the United Kingdom (Glass et al., 1995), and for bank branches in South Africa (O'Donnell and Westhuizen, 2002) and Spain (Lovell and Pastor, 1997).

[3]The metafrontier function was first introduced by Hayami (1969) and Hayami and Ruttan (1970, 1971). As indicated by these authors (Hayami and Ruttan, 1971, p. 82), "the metaproduction function can be regarded as the envelope of commonly conceived neoclassical production functions."

to control for the existence of municipalities facing different technologies (i.e., municipalities in different groups), in such a way that the efficiencies found are now comparable because the curse of dimensionality problem is strongly alleviated. Therefore, we contribute to this growing field of research in which, as pointed out by Battese et al. (2004) in their conclusions, "further theoretical and applied studies with other models for technical inefficiency effects are clearly desirable."

The plan of the paper is as follows. Section 2 and Section 3 provide details on the techniques employed to both measure efficiency and assess the statistical differences between the efficiencies obtained. Section 4 specifies the particularities of the data employed. Section 5 presents and comments on the most relevant results, and finally Section 6 summarizes with some concluding remarks.

2. Methodology: efficiency measurement

In the first stage we measure efficiency for all municipalities in our sample, regardless of their different characteristics. Therefore, we consider a common non-convex Free Disposable Hull (FDH) frontier for all observations as follows:

$$\begin{aligned} &\min_{\{\alpha_k^{FDH}, Z_1, Z_2, \ldots, Z_S\}} \quad \alpha_k^{FDH}, \\ &\text{s.t.} \quad \alpha_k^{FDH} TC^k - \textstyle\sum_{s=1}^{S} Z_s TC_s \geq 0, \\ &\qquad\qquad \textstyle\sum_{s=1}^{S} Z_s y_{s,i} - y_{k,i} \geq 0, \quad i = 1, \ldots, I \\ &\qquad\qquad\qquad Z_s \in \{0, 1\}, \quad s = 1, \ldots, S \end{aligned} \tag{1}$$

where TC_s is total cost for municipality s, $s = 1, \ldots, S$, and $y_{s,i}$ represents the value of its i^{th} output, $i = 1, \ldots, I$, and Z_s denotes the intensity level at which the s observation is conducted.

The FDH methodology is particularly suited to detect the most obvious cases of inefficiency as this technique is very demanding with regard to inefficiency measurement. For each municipality labeled as FDH-inefficient, at least one other municipality with superior performance can be found in the sample. Under some other technological assumptions (e.g., for the convex Data Envelopment Analysis, DEA, models) it may well be the case that the inefficiency coefficient depends entirely on the assumption of convexity.

At this point, two aspects of the FDH methodology deserve special attention: *efficiency by default* and *outliers*. In the absence of a sufficient number of similar municipalities for a comparison, a municipality is labeled as efficient by default. This ranking of efficiency does not result from any effective superiority, but rather is due to the lack of information that would allow pertinent comparisons. In addition, by construction, the FDH concept of efficiency applies both to the municipality that presents the lowest level of spending and to those with the highest values for at least one output indicator. This extreme form of the sparsity bias that characterizes the FDH technique leads to lack of discrimination among production units and constitutes a shortcoming of the FDH approach.

As for outliers, by definition nonparametric frontiers are defined by the extreme values of the dimensional space of inputs and outputs. Thus, the appearance of outliers (atypical observations that differ significantly from the rest of the data) may considerably influence efficiency computations. It is therefore necessary to verify that the divergence does not result from evaluation errors. However, once the reliability of the data set has been confirmed this kind of information may provide valuable information.

Recent work has established the statistical properties of the FDH estimator (Kneip et al., 1998; Simar and Wilson, 2000) so that inference is now possible either by using asymptotic results or by means of bootstrap. Simar and Wilson (2000) present a survey on this issue as well as a detailed examination of the statistical properties of the nonparametric estimators in a multivariate context. Like other nonparametric measures, FDH estimators suffer from the curse of dimensionality due to their slow convergence rate.

Taken together, the above mentioned problems may be serious enough to jeopardize the FDH estimates. To solve these problems some additional procedures are required in order to make FDH estimates more robust. Several approaches have already been proposed in the literature. Wilson (1993, 1995) introduced descriptive methods to detect influential observations in nonparametric efficiency calculations. More recent developments include the order-m frontiers (Simar, 2003). The order-m approach, based on the concept of expected maximal (minimal) output (input or cost) function, yields frontiers of varying degrees of robustness. The order-m frontiers allow for statistical inference while keeping their nonparametric nature. We briefly describe this approach below.

Consider a positive fixed integer m. For a given level of input (x_0) and output (y_0), the estimation defines the expected value of maximum of m random variables (Y_1,...,Y_m), drawn from the conditional distribution of the output matrix Y observing the condition $Y_m \geq y_0$. Formally, the proposed algorithm (algorithm I) to compute the order-m estimator has the following steps:

1. For a given level of y_0, draw a random sample of size m with replacement among those y_{sm}, such that $y_{sm} \geq y_0$.

2. Compute Program (1) and estimate $\widetilde{\alpha}_s$.

3. Repeat steps 1 and 2 B times and obtain B efficiency coefficients $\widetilde{\alpha}_s^b (b = 1, 2, \ldots, B)$. The quality of the approximation can be tuned by increasing B, but in most applications $B = 200$ seems to be a reasonable choice.

4. Compute the empirical mean of B samples as:

$$\alpha_s^m = \frac{1}{B} \sum_{b=1}^{B} \widetilde{\alpha}_s^b \tag{2}$$

As m increases, the number of observations considered in the estimation approaches the observed units that meet the condition $y_{sm} \geq y_0$ and the expected order-m estimator in each one of the b iterations ($\widetilde{\alpha}_s^b$) tends toward the FDH (α_s^{FDH}). Thus, m is an arbitrary positive integer value, but it is always convenient to observe the fluctuations of the $\widetilde{\alpha}_s^b$ coefficients depending on the level of m. For acceptable values of

m, α_s^m will normally present values smaller than the unity (this indicates that these units are inefficient, as total costs can be reduced without modifying the production plan). When $\alpha_s^m > 1$, the s unit can be labeled as superefficient, as the order-m frontier exhibits a higher total cost.

As mentioned above, the order-m estimation is an excellent tool to mitigate the problems of dimensionality and the presence of extreme observations and outliers. However, this evaluation will be of little use if part of the inefficiency found hinges on the output complexity or the different environmental conditions that local governments face, which could lead to biased estimates of the frontier, and hence misleading policy implications.

Therefore, our objective is to define a process that can estimate the impact on efficiency of output complexity, environmental conditions and, in general, other options that consider classifying municipalities in different groups. This estimation is possible following the proposals of Battese et al. (2004) and O'Donnell et al. (2008) for estimating a metafrontier production function. This process (algorithm II) contains the following steps:

1. Use cluster analysis to classify the S units in $S_1, S_2, \ldots, S_C$ groups.

2. Following the algorithm to estimate the order-m efficiency coefficients, complete steps 1 to 4 of algorithm I to estimate the efficiency coefficients ($\alpha_s^{m,S_1}, \alpha_s^{m,S_2}, \ldots, \alpha_s^{m,S_C}$) for the municipalities classified in each one of the clusters $S_1, S_2, \ldots, S_C$. In order to facilitate the cross-comparison of the results, irrespective of the number of units classified in each cluster, the same value for m will be assigned in all the estimations. By doing this, the problems of dimensionality and the potential impact of the outliers will be neutralized.

3. After completing the conditional frontiers in step 2 of algorithm II, apply steps 1 to 4 of the order-m estimation to the complete sample and estimate the efficiency coefficients with respect to the metafrontier α_s^m.

4. Estimate the technology gap ratio (TGR) separating the conditional and the metafrontier as the ratio $(\alpha_s^m/\alpha_s^{m,S_1}), (\alpha_s^m/\alpha_s^{m,S_2}), \ldots, (\alpha_s^m/\alpha_s^{m,S_C})$.

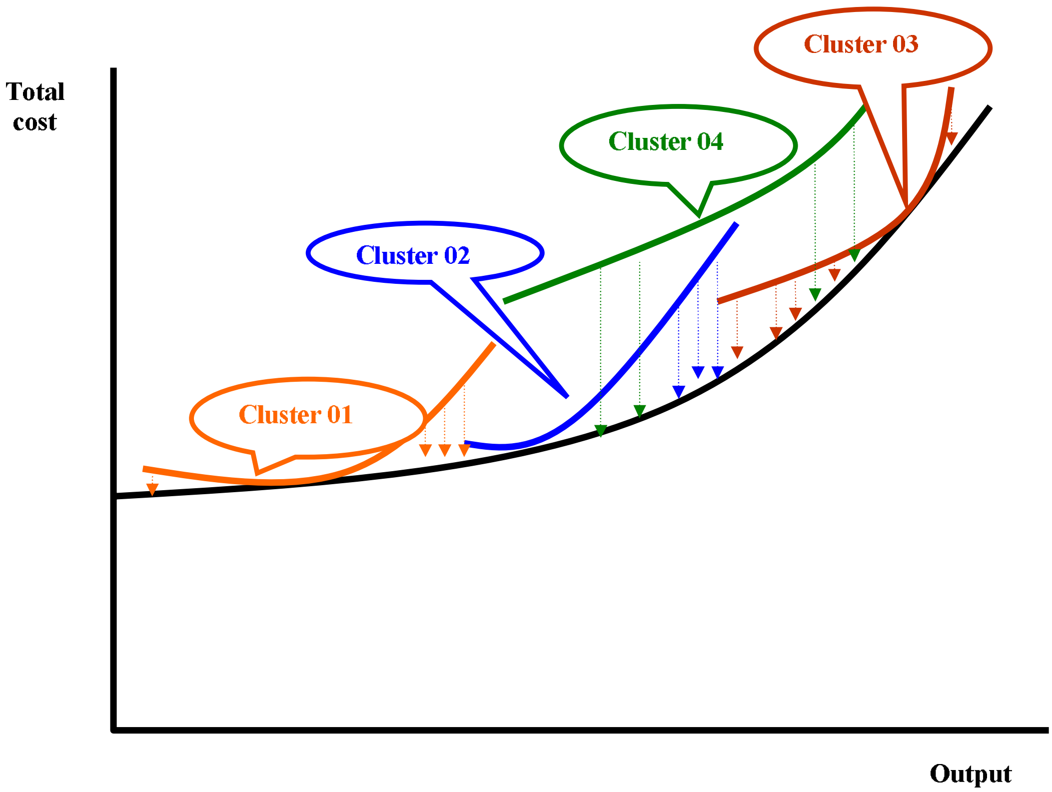
Figure 1 presents an illustration of a simple case with one output and the total cost. At a given output level, the technology gap ratio (TGR) is defined as the lowest possible cost within the metafrontier divided by the lowest total cost at the conditional specific cluster.[4]

3. Testing the closeness between efficiency distributions

Once computation of efficiency scores have been computed there are multiple ways to display and compare results (El-Mahgary and Lahdelma, 1995). We consider some methods which provide us with more

[4]We have borrowed the concept of technology gap ratio from Battese and Rao (2002), Battese et al. (2004) and O'Donnell et al. (2008). They use the term *technology* because they consider DMUs operating under different technologies. Although we also talk about different technologies, in order to facilitate comparisons with their papers, we must acknowledge we do not explicitly test whether the municipalities in the different groups have different technologies. Therefore, when referring to *technology* we are only suggesting whether the effect of each hypothesis (output mix or environmental conditions) matters or not.

Figure 1: Metafrontier function model



accurate information (see, for instance Li, 1996; Li et al., 2009). If we based our interpretations on a number of summary statistics only, we would miss a considerable amount of relevant information. Most of these methods are based on kernel smoothing to nonparametrically estimate the density functions corresponding to both α_s^m and α_s^{m,S_C} indices.[5] The kernel density estimator $\hat{f}$ for a univariate density f based on a sample of S efficiency indices (either α_s^m or α_s^{m,S_C}) is $\hat{f}(x) = (Sh)^{-1}\sum_{s=1}^{S} K((x-\alpha_s^m)/h)$, where s is the municipality index, α_s^m represents its efficiency index, x is the evaluation point, h is the bandwidth, and K is a kernel function satisfying certain properties. Additional decisions concerned the kernel and bandwidth choice. For choice of kernel, we considered the Gaussian method for its ease of computation. Regarding bandwidth selection, the plug-in method suggested by Wand and Jones (1994) is increasingly adopted in the literature.[6]

We must control for the fact that efficiency indices are bounded between $(0, 1]$. The Silverman (1986)-Schuster (1985) reflection method is based on the idea of "reflecting" the probability mass lying beyond unity—where, theoretically, no probability mass should exist. The kernel estimate disregarding the boundary condition can be shown to be biased and inconsistent (Simar and Wilson, 1998).[7] After estimating densities, the Li (1996) test shows whether the observed *visual* differences are statistically significant or not. The test is based on measuring the distance between densities $f(x)$ and $g(x)$ for which we consider the mean integrated square error, i.e.:

$$\begin{aligned} L = L\big(f(x), g(x)\big) = \int_x \big(f(x)-g(x)\big)^2 dx &= \int_x \big(f^2(x)+g^2(x)-2f(x)g(x)\big)dx \\ &= \int_x \big(f(x)dF(x)+g(x)dG(x)-2g(x)dF(x)\big) \end{aligned} \quad (3)$$

where F and G are candidates for the distribution of X, with density functions $f(x)$ and $g(x)$ and $\hat{f}$ is the nonparametric kernel estimator referred to above. Given that $\hat{f} = (Sh)^{-1}\sum_{s=1}^{S} K\big((x-\alpha_s^m)/h\big)$, and $\hat{g} = (Sh)^{-1}\sum_{s=1}^{S} K\big((y-\alpha_s^{m,S_C})/h\big)$ then, in practice, an estimator for L is:

$$\begin{aligned} \tilde{L} &= \int_x \big(\hat{f}(x)-\hat{g}(x)\big)^2 dx \\ &= \frac{1}{S^2 h}\sum_{s=1}^{S}\sum_{\substack{t=1\\ t\neq s}}^{S}\Big[K\Big(\frac{\alpha_s^m-\alpha_t^m}{h}\Big)+K\Big(\frac{\alpha_s^{m,S_C}-\alpha_t^{m,S_C}}{h}\Big)-K\Big(\frac{\alpha_s^{m,S_C}-\alpha_t^m}{h}\Big)-K\Big(\frac{\alpha_s^m-\alpha_t^{m,S_C}}{h}\Big)\Big] \end{aligned} \quad (4)$$

The integrated square error is also required for estimating the statistic on which the test is based, the

[5]Several monographs cover this topic in depth. See, for instance, Silverman (1986), Pagan and Ullah (1999) or, more recently, Li and Racine (2007).

[6]It can also be easily accessed, as it is implemented in many software packages such as `R`.

[7]Although we used a variety of statistical procedures to perform all computations, the `FEAR` package by Paul W. Wilson for the statistical software `R` provides codes for computing both efficiencies and densities, considering the reflection method. See `http://business.clemson.edu/Economic/faculty/wilson/`.

expression for which is as follows:

$$T = \frac{Sh^{1/2}\tilde{L}}{\hat{\sigma}} \tag{5}$$

where

$$\hat{\sigma} = \frac{1}{S^2 h \pi^{1/2}} \sum_{t=1}^{S} \sum_{s=1}^{S} \Big[K\Big(\frac{\alpha_s^m - \alpha_t^m}{h}\Big) + K\Big(\frac{\alpha_s^{m,S_C} - \alpha_t^{m,S_C}}{h}\Big) + 2K\Big(\frac{\alpha_s^m - \alpha_t^{m,S_C}}{h}\Big)\Big] \tag{6}$$

and h is the bandwidth.

The Li (1996) test requires some assumptions to be met such as independently distributed efficiency scores in each sub-group (and identically within each sub-group). However, order-m efficiency estimates are dependent in the statistical sense, since perturbations of observations lying on the estimated frontier will in most cases cause changes in efficiencies estimated for other observations. As indicated by Simar and Zelenyuk (2006), under these circumstances the Li (1996) test has to be modified in several ways. First, we must control for the issue of bounded support in order-m density estimation. However, as demonstrated by Simar and Zelenyuk (2006), although controlling for the boundary effect is important in density estimation, the statistic based on the reflection method is "essentially the same as the original Li (1996) test, with the difference being a factor of $\sqrt{2}$ and the fact that the bandwidth used in estimation of the statistic is obtained from the data with reflection rather than the original data". Therefore, since the independency issue is not negligible, we should then follow Simar and Zelenyuk (2006), who provide a way of adapting the Li (1996) test to the order-m context via bootstrapping techniques to improve its performance. These authors provide consistent bootstrap estimates of the p-values of the Li (1996) test as follow:

$$\hat{p} = \frac{1}{B} \sum_{b=1}^{B} I\{\hat{L}^b > \hat{L}\}, \tag{7}$$

where $b = 1, \ldots, B$ is the number of bootstrap replicates. The p-values are then adapted to our context, where the true efficiency scores are replaced by order-m estimates. Simar and Zelenyuk (2006) consider somewhat *ad hoc* methods to solve the discontinuity problem generated by the spurious probability mass at the unity—recall that, by construction, at least one observation will always be on the frontier, and in most circumstances the number will be quite large. We adopt one of their proposed methods (Algorithm II; see Simar and Zelenyuk, 2006), based on computing and bootstrapping the Li (1996) statistic using the sample of order-m estimates where those equal to unity are "smoothed" away from the boundary. We add a small noise, within, say, 5% of the empirical distribution of $\hat{\alpha}_s^m$, disregarding those equal to the unity, but with an order of magnitude smaller than the noise of the estimation. The smoothing procedure is performed via:

$$\hat{\alpha}_s^m = \begin{cases} \hat{\alpha}_s^m + \varepsilon_s, & \text{if } \hat{\alpha}_s^m = 1; \\ \hat{\alpha}_s^m, & \text{otherwise.} \end{cases} \tag{8}$$

where $\varepsilon_i = \text{Uniform}(0, \min\{S^{-2/(I+1+1)}, a\})$, a is the α-quantile of the empirical distribution of $\hat{\alpha}_s^m$ ignoring those equal to the unity.

4. Data, inputs, and outputs

We perform the analysis for a sample of Spanish municipalities with a population over 1,000 inhabitants for year 2000. Both input and output data are provided by the Spanish Ministry for Public Administration. Information on outputs is gathered through the survey on local infrastructures and facilities (*Encuesta de Infraestructuras y Equipamientos Locales*), which is performed with 5-year frequency and, consequently, constrains our sample period. Information on inputs basically consists of different types of costs, and is taken from local government budgetary data. This data is available for every year. The regions that meet our criteria (data for year 2000, and data for both inputs and outputs) were Andalusia, Aragon, Asturias, the Canary Islands, Cantabria, Castile-Leon, Castile-La Mancha, Extremadura, Murcia, La Rioja, and the Valencian Community. The final sample was made up of 1,198 municipalities. There was no information for the remaining regions for several reasons. At the time of the study, Madrid had not yet presented information on its outputs. Catalonia, the Basque Country and Navarra do not have to provide the Spanish Ministry for Public Administration with this information.

Measuring the production process at municipal level is usually more difficult than in other sectors/industries. We can distinguish three stages in this process of transforming inputs into outputs (Bradford et al., 1969). In the first stage primary inputs (labor, equipment and external services) are transformed into intermediate outputs (e.g., hours of traffic control or the extension of police services). In the second stage, intermediate outputs are transformed into direct outputs. This is what Bradford et al. (1969) call *D-outputs*, which are ready for *consumption* by the population. In the third stage, the direct outputs ultimately have welfare effects on consumers (e.g., increasing perceptions and feelings of safety and welfare). The third stage of the process can be directly captured by outcome indicators (labeled *C-outputs* by Bradford et al., 1969), which reflect the degree to which direct outputs translate into welfare improvements as perceived by consumers.

The efficiency of municipalities can be measured at each stage of this production process. However, under normal circumstances this will be difficult because data might be either unavailable or simply poor, making it difficult to distinguish between primary inputs, intermediate outputs, direct outputs, and final welfare effects. For this reason the analysis is usually confined to analyzing the first and second phases of this process, i.e., the links between primary inputs and direct outputs. We base our selection of outputs on the services and facilities provided by each municipality. All local authorities must provide public street lighting, cemeteries, waste collection and street cleaning services, drinking water to households, access to population centers, surfacing of public roads, and regulation of food and drink. In some cases we must select proxies for these services and facilities. As pointed out by De Borger and Kerstens (1996), population is assumed to proxy for the various administrative tasks undertaken by municipalities, but it is clearly not a direct output of local production. Other relevant outputs, such as provision of primary and secondary education, are not the responsibility of Spanish municipalities.

Spanish law requires municipalities to provide minimum services depending on their size. Some of the minimum services and facilities must be provided by all municipalities, while others are only binding for larger municipalities (with populations of over 5,000, 20,000, and 50,000, the boundaries that define the different categories). The second column in Table 1 reports information on the minimum services that each category of municipalities must provide. The third column indicates the selected output indicators to measure the different services and facilities. Our output choice was driven by the minimum services and facilities. The list of outputs for year 2000, along with summary statistics, are reported in Table 2. The choice was also driven by previous studies on local government efficiency in other European countries, since for the most part they are endowed with the same competencies.[8]

Selecting inputs is much easier, as it is based on budgetary variables reflecting municipalities' costs. Our definition reflects the economic structure of Spanish local government expenditures, details of which are reported by Spanish legislation,[9] that considers three basic categories: current or ordinary expenditures, capital expenditures, and financial expenditures. Within these, current expenditures are further divided into four chapters, or categories, which account for: i) personnel expenditure; ii) current goods and services expenditures; iii) financial expenditures; iv) current transfers. Capital expenditures are also broken down into either real investments, or capital transfers. The former is what Table 2 refers to as capital expenditures (X_4), i.e., all expenditures local governments implement: i) to produce or acquire capital goods; ii) to acquire necessary goods to provide local services in the right conditions; or iii) financial expenditures that are suitable for amortization. On the other hand, capital transfers (X_5) refer to the payments to institutions to finance certain investments. Descriptive statistics for year 2000 are provided in Table 2. Since our analysis is entirely confined to overall cost efficiency, the fact that some local government departments may be actually sharing some costs does not raise any particular issue.

We must also select those variables to include when carrying out the cluster analysis for classifying municipalities into groups according to the different hypotheses. This task is not easy given that we face certain relevant constraints. First, there is no well-established theory as to which variables constitute the "environmental conditions" that might impact on each municipality's cost structure. Second, the available information is also limited. In a number of contexts the relevance of considering groups is unquestionable. As indicated by O'Donnell et al. (2008), in most practical settings DMUs can be grouped *a priori* on the basis of geographical, economic and/or political boundaries, to name a few.[10] However, it is not entirely clear how groups must be formed. In the absence of "natural" boundaries for the different groups, multivariate statistical techniques such as cluster analysis are available for determining both the number of groups and group membership. In our particular setting, the *a priori* classification or "natural" boundaries would be those based on size, since the limits of each group are determined by the extent of their powers. In contrast, we need the cluster analysis technique to classify municipalities according to either output mix or environmental conditions.

Regarding the classification based on output complexity, the variables selected to construct the groups

[8]Differences are basically confined to the area of education which, in Spain, corresponds to regional and central governments.

[9]See Ministerial Order (*Orden Ministerial*), September 20$^{\text{th}}$, 1989.

[10]As also indicated by O'Donnell et al. (2008), if the analysis were conducted in an SFA framework, it would also be possible to conduct statistical tests concerning the number of groups. El-Gamal and Inanoglu (2005) propose other methods to circumvent the use of multivariate analysis techniques.

Table 1: Output indicators based on minimum services provided

	Minimum services provided	Output indicators
All local governments	Public street lighting.	Number of lighting points
	Cemetery	Total population
	Waste collection	Waste collected
	Street cleaning	Street infrastructure surface area
	Supply of drinking water to households	Population, street infrastructure surface area
	Access to population centres	Street infrastructure surface area
	Surfacing of public roads	Street infrastructure surface area
	Regulation of food and drink	Total population
In local governments with populations of over 5,000, in addition	Public parks	Surface of public parks
	Public library	Total population, public buildings surface area
	Market	Market surface area
	Treatment of collected waste	Waste collected
In local governments with population of over 20,000, in addition	Civil protection	Total population
	Provision of social services	Total population, public buildings surface area, assistance centers surface area
	Fire prevention and extinction	Street infrastructure surface area
	Public sports facilities	Total population, public buildings surface area
	Abattoir	Total population
In local governments with populations of over 50,000, in addition	Urban passenger transport service	Total population, total surface area
	Protection of the environment	Total surface area

Table 2: Summary statistics for inputs and outputs, year 2000

Inputs[a]	Mean	Std.dev.
Wages and salaries (X_1)	6,304.18	9,044.30
Expenditure on goods and services (X_2)	4,953.61	7,862.17
Current transfers (X_3)	1,072.51	1,748.62
Capital expenditure (X_4)	5,905.78	8,513.16
Capital transfers (X_5)	276.52	1,031.29
Outputs		
Population (Y_1)	6,118.54	6,991.48
Number of lighting points (Y_2)	1,100.07	1,022.89
Tons of waste collected (Y_3)	2,984.87	4,787.84
Street infrastructure surface area (Y_4)	256,097.13	261,653.68
Public buildings surface area[b](Y_5)	599.84	1,198.27
Market surface area[b](Y_6)	748.21	1,867.12
Registered surface area of public parks[b](Y_7)	202.44	1,193.32
Assistance centers surface area (Y_8)	1,858.22	4,015.22
# of observations	1,198	

[a] In thousands of euros, converted from 1995 pesetas (1 euro=166.386 pesetas).

[b] In square meters.

are similar to those chosen as outputs, dividing them by population. This helps to control for the fact that, as pointed out in the introduction, some municipalities might go beyond the legal minimum and provide an amount of services which does not correspond to their size. Therefore, we would only compare municipalities with similar output mixes, i.e., only those in the same group. We have selected as many variables to construct the clusters as outputs. The details are reported in Table 3.

Regarding the classification based on environmental conditions, we have chosen some of the variables provided by the *Anuario Estadístico de La Caixa*.[11] Although they can be partly judged as *ad hoc*, and although factors influencing the amount, allocation and distribution of local public spending, we consider these provide a rough idea of the environmental conditions facing each municipality that might have an impact on municipal finances. Summary statistics are reported in Table 4, and the particular definition of each variable considered for performing the cluster analysis follows:

Total surface area: (divided by population). This indicator is similar to density, which is usually defined as urbanized land per person. Although some authors (Solé-Ollé and Hortas Rico, 2008) consider it is more appropriate to use urbanized land per person, this variable was not available for all municipalities in our sample. We consider this is an appropriate proxy for urban sprawl, although other indicators can be considered (for instance, street surface area divided by total surface ($ENV1$).

Tourist index: this index may contribute to rising municipal expenditures, at least during some months of the year. Comparing these municipalities only with their peers might be more appropriate ($ENV2$).

Economic status: municipalities with wealthier populations might be facing higher types of requirements. These constituencies may be willing to pay more taxes but, in return, they will demand

[11] An annual report provided by the largest Spanish savings bank, *La Caixa*.

Table 3: Clusters based on output mix, medians, year 2000

Group	# observations	Output mix 1	Output mix 2	Output mix 3	Output mix 4	Output mix 5	Output mix 6	Output mix 7
Cluster 1	300	0.0357	0.3755	56.5099	0.0370	0.0000	2.2358	0.0040
Cluster 2	407	0.0204	0.3825	30.1844	0.0315	0.0638	2.4404	0.0083
Cluster 3	135	0.0273	0.3601	43.5196	0.0095	0.0000	7.5103	0.0438
Cluster 4	116	0.0547	0.4442	108.7945	0.1267	0.0000	2.8484	0.0000
Cluster 5	100	0.0315	0.3670	65.8584	0.5475	0.0000	1.6693	0.0000
Cluster 6	72	0.0234	0.8920	36.4134	0.0410	0.0406	2.3416	0.0102
Cluster 7	25	0.0309	0.3643	48.4396	0.0723	0.0000	45.9444	0.0111
Cluster 8	14	0.0303	2.1255	46.8649	0.0324	0.0075	3.9791	0.0093
Cluster 9	20	0.0256	0.3535	60.3041	0.0680	1.0086	3.9054	0.0163

[a] The MANOVA analysis indicated that differences between the different group means were statistically significant, as shown by Wilks-$\Lambda = 0.565$, corresponding to p-value=0.000. The Pillai-Bartlett statistic yielded analogous results.

Output mix 1:	lighting points/population	Output mix 5:	public markets area/population
Output mix 2:	waste collected/population	Output mix 6:	public parks area/population
Output mix 3:	street surface area/population	Output mix 7:	assistance centers/population
Output mix 4:	public buildings surface area/population		

more services and facilities ($ENV3$).

Industrial activities: (divided by population). Municipalities in which industrial activity is high may be affected by a different, probably higher, cost structure, since this type of activity requires higher investments in infrastructures, security, anti-pollution policies, etc., which may—although not necessarily—be offset by higher tax revenues ($ENV4$).

Total number of cars: (divided by population). This might be related to $ENV3$. Some wealthy constituencies have higher levels of education also (in relative terms) and might have a preference for non-polluting means of transport ($ENV5$).

Unemployment: higher unemployment might entail higher crime and, therefore, increased demands for security ($ENV6$).

Total population growth, 1991–98 (percentage): municipalities facing higher population growth might have had to increase the services and facilities at their own expense, because of the speed at which the population's demands have increased. If central and regional governments have not reacted promptly, they might face a sudden imbalance between their revenues and their costs ($ENV7$).

Construction: (divided by population). Municipalities where construction was higher have also raised high tax revenues, which might have led to inefficient management of these increased revenues. In addition, the population levels in these municipalities might have increased sharply,[12] which might have driven some municipalities to increase their social expenditures, as well as other expenses such as civil protection, security, etc., some of which are not included in the list of minimum services municipalities must provide ($ENV8$).

Agricultural vehicles/Total number of vehicles (percentage): this information would indicate whether it is a rural municipality whose needs might differ from others with different sectoral specializations. It might also be a proxy for urban sprawl ($ENV9$).

Number of bank branches: (divided by total population). This would indicate another type of specialization and, in addition, it might proxy for the economic level of the municipality ($ENV10$).

We admit the selection of these variables is somewhat *ad hoc*. However, we consider they represent a realistic summary of the different socioeconomic conditions affecting each municipality.

Some decisions involved in performing cluster analysis are the measure of similarity as well as the clustering method. Regarding the former, one of the most popular choices is the Euclidean square distance. Regarding the latter, although there are several alternatives, the Ward hierarchial clustering method has the advantage of maximizing intra-group homogeneity and inter-group heterogeneity. In addition, the technique is robust to outliers and groups are not too dissimilar in size. However, the criterium that must

[12]Indeed, the strong increase of population in Spain over the last 10 years has been unevenly distributed across regions, and it has been especially higher in areas where construction grew more in relative terms.

Table 4: Clusters based on environmental variables, medians, year 2000[a]

Group	# observations	ENV1	ENV2	ENV3	ENV4	ENV5	ENV6	ENV7	ENV8	ENV9	ENV10
Cluster 1	444	33.3422	0.0000	3.0000	12.0445	0.4057	3.0000	–3.6000	5.9265	3.3090	0.8857
Cluster 2	54	218.3355	2.5000	5.0000	14.4055	0.5359	3.6500	25.1000	8.3535	4.4866	0.7296
Cluster 3	96	310.2710	2.0000	5.0000	13.6200	0.5077	5.0000	1.8500	6.0925	4.1590	0.7703
Cluster 4	201	69.8400	1.0000	4.0000	20.7760	0.5063	3.3000	–1.1000	11.7070	6.2500	1.0717
Cluster 5	130	99.3975	2.0000	3.0000	11.7470	0.4087	5.7000	1.8500	5.6840	3.8620	0.7624
Cluster 6	147	20.7421	1.0000	5.0000	24.5230	0.4988	2.4000	–5.3000	13.8750	6.9830	2.1816
Cluster 7	71	75.1765	1.0000	4.0000	16.4570	0.5374	3.2000	2.6000	8.1300	19.5360	0.9643
Cluster 8	44	108.6135	2.0000	5.0000	39.9995	0.6188	2.7500	16.1500	19.9020	5.1865	1.2016

[a] The MANOVA analysis indicated that differences between the different group means were statistically significant, as shown by Wilks-$\Lambda = 0.502$, corresponding to p-value=0.000. The Pillai-Bartlett statistic yielded analogous results. Results for some clusters are not reported due to their very small number of observations, basically outliers which could not be classified in other groups. These totalled 11 observations which, compared with the 1,198 municipalities in our sample, is not noteworthy.

ENV 1:	surface area/population	ENV 6:	unemployment rate
ENV 2:	tourist index	ENV 7:	population growth, 1991–98
ENV 3	economic status	ENV 8:	construction/population
ENV 4	industrial activities/population	ENV 9:	agricultural vehicles/number of vehicles
ENV 5	number of cars/population	ENV 10:	number of bank branches/population

Table 5: Order-m efficiencies (metafrontier), summary statistics, year 2000

Group	Mean	Median	Max.	Min.	Std.dev.	% of eff. obs.
All municipalities, metafrontier	0.9118	1.0000	1.4207	0.2344	0.1818	0.3088
Group 1, metafrontier	0.8904	1.0000	1.4207	0.2344	0.2048	0.1068
Group 2, metafrontier	0.9417	1.0000	1.1364	0.3772	0.1343	0.5979
Group 3, metafrontier	0.9787	1.0000	1.0000	0.6376	0.0730	0.9014

ultimately guide the decision on both the methodology and the optimal number of groups is whether the final groups are sensible in any way, and whether statistical differences exist among group centroids.

Battese et al. (2004) point out the importance of analyzing whether all municipalities share the same technology. If all municipality-level data were generated from a single production function and the same underlying technology, there would be no good reason for estimating the efficiency levels of municipalities relative to a metafrontier. We assume that if statistically significant differences existed among the different groups, it would constitute evidence in favor of comparing municipalities with those in their group only.

5. Results

Table 5 provides summary statistics on *unconditioned* efficiency for order-m efficiency scores. Results are reported for all municipalities, and also for the different size categories, given the differences in their powers. Results have been obtained by specifying a common frontier—the metafrontier—for all 1,198 observations. Thus, although results are split into different municipality size categories, they correspond to the same common frontier. The results corresponding to *all* municipalities, regardless of their powers, are displayed in the first row. Average efficiency is 91.18%, which is a high value given that municipalities would become fully efficient if they were able to decrease their total costs by 8.82% only. However, this is an *average* effect which varies across municipalities. The values at both tails of the distribution suggest that a remarkable variety of behaviors exist, since the minimum is 23.44%, whereas the maximum is 142.07%. In the former case, cost inefficiency is high, whereas the latter refers to cases of super-efficiency—units which lie beyond the frontier and can be regarded as outliers. This finding is important, since it constitutes a clear advantage of the order-m frontiers over DEA or FDH, which are strongly affected by the existence of outliers. In the case of order-m frontiers these extreme observations are labeled as super-efficient and do not affect the efficiencies found for other observations.

Table 5 also reports order-m efficiencies for the different categories of municipalities split by population and, consequently, levels of powers. The smallest municipalities in the sample (those with populations between 1,000 and 5,000) are, on average, the most inefficient. Mean efficiency is 89.04%, close to the global mean value of 91.18%. These results are partly similar because this is, by and large, the category with more observations. In contrast, medium sized municipalities (with populations between 5,000 and 20,000) and large municipalities (with populations over 20,000) show higher efficiencies. Not only is average efficiency higher (94.17% and 97.87%, respectively), but also the number of municipalities lying *on* the frontier (i.e., fully efficient municipalities) is much higher (59.79% and 90.14%, respectively).[13]

However, the most interesting results are those obtained for the different clusters, constructed either using output mix or environmental variables. A description of these clusters is provided in Table 3 (clusters based on output mix) and Table 4 (clusters based on environmental variables). In order to facilitate interpretations, we provide a lower panel below each table that reports a summary of the variables included to form the clusters. Regarding the clusters based on output mixes, as reported in Table 3, differences between the municipalities in each group (in terms of the selected variables) are noteworthy, even though the size of some of these clusters is remarkable. For instance, group 1 is made up of 300 municipalities, roughly 1/4 of the total sample. Ideally, it would be desirable to have clusters containing fewer observations to facilitate comparisons. However, some clusters were difficult to split into further groups, despite considering the Ward method to cluster observations (which tends to form equally-sized clusters). This is an interesting finding, which would corroborate the fact that many municipalities indeed do different things, making comparisons misleading.

In some cases, the medians for some clusters and variables differ substantially; this is the case for cluster 2 in $OUTMIX5$, cluster 3 in $OUTMIX6$, cluster 4 in $OUTMIX3$, cluster 5 in $OUTMIX4$, cluster 6 in $OUTMIX2$ and $OUTMIX5$, cluster 7 in $OUTMIX6$, cluster 8 in $OUTMIX2$, and cluster 9 in $OUTMIX5$. Therefore, the clusters excel in some particular variables, even taking into account that some of them contain many observations—compared with the total sample size. In addition, although there is a wide consensus that the multivariate technique of cluster analysis is flawed, especially because of the multiple decisions it involves, the MANOVA analysis indicated that the differences between the identified groups and variables were indeed significant.[14]

With respect to the clusters based on environmental variables, as indicated in Table 4, the differences found among them (with respect to the variables included in the analysis) are also noteworthy, and significant at the 1% level. In this case, differences are more difficult to distinguish, because of the narrow range of variation for some of these variables (for instance, $ENV2$ or $ENV3$). However, certain groups excel in some variables. See, for instance, cluster 3 in $ENV1$, cluster 8 in $ENV4$, cluster 2 and 6 in $ENV7$, cluster 8 in $ENV8$, cluster 7 in $ENV9$ or cluster 6 in $ENV10$.

It may be claimed that the municipalities in the groups, regardless of the criteria followed to create them, might be considered to possess different technologies. As indicated in the introduction, Battese et al. (2004) propose a method for comparing the efficiencies of DMUs in different groups in the context of Stochastic Frontier Analysis, which has been extended to the DEA context by O'Donnell et al. (2008). Our methodology, based on order-m indicators, allows us to control for group membership in the context of efficiency measurement via nonparametric techniques and at the same time, take into account the severity of the *curse of dimensionality*. We report the results obtained following our approach in Table 6. On average—and this result holds for both categories of clusters—municipalities are much closer to their

[13]For explanations for the different levels of inefficiency found see, for instance, Balaguer-Coll et al. (2007).

[14]Although, technically, MANOVA compares the means of the groups, not the medians, as reported in Table 3 and Table 4.

Table 6: Order-m efficiency scores and metatechnology ratios, all municipalities, summary statistics, year 2000

Group	Index	Mean	Median	Max.	Min.	Std.dev.
Clusters based on output mix variables	Cluster-specific efficiencies	0.9910	1.0000	1.9583	0.2425	0.1832
	Technology gap	0.9249	0.9996	1.1949	0.3659	0.1343
Clusters based on environmental variables	Cluster-specific efficiencies	0.9892	1.0000	1.5218	0.2493	0.1447
	Technology gap	0.9203	0.9895	1.1682	0.2766	0.1264
Clusters based on size	Cluster-specific efficiencies	0.9264	1.0000	1.6983	0.2428	0.1807
	Technology gap	0.9849	1.0000	1.2117	0.4559	0.0598

frontiers, as documented by average efficiencies closer to unity. This result holds for both clusters based on output mix variables (99.10%) and clusters based on environmental variables (98.92%). In the case of clusters based on size, as one might *a priori* expect, results are quite similar to those of the unconditioned case (Table 5).

However, Table 6 also reveals that classifying municipalities into different groups does not *per se* explain away the remaining efficiencies. The maximum values for the three clustering criteria are well above the unity, suggesting that a non-negligible number of outliers exist. These are what Andersen and Petersen (1993) call super-efficient units.[15]

More specific results are reported in Table 7, Table 8, and Table 9. They report basic summary statistics of the technology gap ratio, the efficiencies obtained from the group frontiers (CE^g), and the metafrontier (CE^*). In the case of clusters based on output mix, the widest gap between group efficiencies and metafrontier efficiencies corresponds to municipalities in cluster 2, for which average $CE_2^g = 1.0508$ and average $CE_2^* = 0.9071$. As a result, the technology gap ratio is the lowest ($TGR_2 = 0.8776$). In contrast, for cluster 8 the technology gap ratio is the highest ($TGR_8 = 1.0009$) which should be interpreted inversely, i.e., the efficiencies for municipalities in this group are very similar for the group frontier ($CE^g = 1.0006$) and metafrontier ($CE^* = 1.0016$). In general, although some groups show remarkable discrepancies between group efficiencies and metafrontier efficiencies (clusters 1 and 2), for many others the gap is narrower (well above 0.90). Although this result might constitute evidence against our initial hypothesis, we should bear in mind that clusters 1 and 2 are indeed the largest ones, with 300 and 407 observations included in each of them. Therefore, for more than half of the municipalities in our sample, it is more reasonable to compare them only with the municipalities in their output mix group.

Table 8 reports analogous information as in Table 7 for clusters constructed using environmental variables. In this case, supporting evidence for our initial hypothesis is stronger as, on average, the technology gap ratios are much lower than in Table 7. The widest gap is found for cluster 8, whose average $TGR_8 = 0.8718$, whereas the lowest gap is found for cluster 3 ($TGR_3 = 0.9421$). However, in this case the clusters are, on average, much closer to their respective group frontiers than in the case of output mix clusters—on average, most of them show CE^g values in the vicinity of 1. Therefore, regardless of the

[15]The existence of outliers also partly underlies the remarkably high average values for the efficiency scores.

Table 7: Order-m efficiencies and metatechnology ratios, clusters based on output mix variables

Group		Mean	Median	Minimum	Maximum	Std.Dev.	% eff. obs.
Cluster 1	Group efficiency, CE^g	0.9651	1.0000	0.3222	1.6823	0.1669	0.1900
	Technology gap ratio	0.9083	0.9678	0.4094	1.0466	0.1266	
	Metafrontier, CE^*	0.8809	1.0000	0.2672	1.4207	0.2047	
Cluster 2	Group efficiency, CE^g	1.0508	1.0000	0.4205	1.9583	0.2278	0.3808
	Technology gap ratio	0.8776	0.9529	0.4089	1.0000	0.1489	
	Metafrontier, CE^*	0.9071	1.0000	0.3772	1.2286	0.1731	
Cluster 3	Group efficiency, CE^g	0.9300	1.0000	0.2944	1.1426	0.1582	0.3037
	Technology gap ratio	0.9751	1.0000	0.4835	1.0384	0.0870	
	Metafrontier, CE^*	0.9090	1.0000	0.3049	1.1412	0.1795	
Cluster 4	Group efficiency, CE^g	0.9796	1.0000	0.3445	1.2002	0.1043	0.1638
	Technology gap ratio	0.9785	1.0051	0.3659	1.1949	0.1272	
	Metafrontier, CE^*	0.9611	1.0040	0.2710	1.3073	0.1699	
Cluster 5	Group efficiency, CE^g	0.9225	1.0000	0.2425	1.0874	0.1623	0.2300
	Technology gap ratio	0.9922	1.0006	0.4667	1.1652	0.0908	
	Metafrontier, CE^*	0.9165	1.0000	0.2536	1.1682	0.1854	
Cluster 6	Group efficiency, CE^g	0.9972	1.0000	0.6334	1.1546	0.0749	0.4861
	Technology gap ratio	0.9418	1.0000	0.3846	1.0730	0.1252	
	Metafrontier, CE^*	0.9412	1.0000	0.3923	1.2117	0.1527	
Cluster 7	Group efficiency, CE^g	0.9942	1.0000	0.9207	1.0191	0.0229	0.6000
	Technology gap ratio	0.9883	1.0000	0.7597	1.0145	0.0494	
	Metafrontier, CE^*	0.9832	1.0000	0.7076	1.0339	0.0619	
Cluster 8	Group efficiency, CE^g	1.0006	1.0000	1.0000	1.0088	0.0023	0.5714
	Technology gap ratio	1.0009	1.0000	0.8551	1.1033	0.0502	
	Metafrontier, CE^*	1.0016	1.0000	0.8551	1.1129	0.0517	
Cluster 9	Group efficiency, CE^g	0.9686	1.0000	0.3371	1.0187	0.1487	0.6500
	Technology gap ratio	0.9625	1.0000	0.6663	1.1369	0.1215	
	Metafrontier, CE^*	0.9411	1.0000	0.2344	1.1581	0.1963	

Table 8: Order-m efficiencies and metatechnology ratios, clusters based on environmental variables

Group		Mean	Median	Minimum	Maximum	Std.Dev.	% eff. obs.
Cluster 1	Group efficiency, CE^g	1.0008	1.0000	0.3251	1.5098	0.1603	0.2297
	Technology gap ratio	0.9155	0.9573	0.4106	1.0131	0.1019	
	Metafrontier, CE^*	0.9183	1.0000	0.2672	1.3073	0.1775	
Cluster 2	Group efficiency, CE^g	0.9923	1.0000	0.5929	1.0813	0.0615	0.4630
	Technology gap ratio	0.9200	1.0000	0.5185	1.0181	0.1381	
	Metafrontier, CE^*	0.9144	1.0000	0.4611	1.0503	0.1536	
Cluster 3	Group efficiency, CE^g	0.9895	1.0000	0.4945	1.4171	0.1124	0.5833
	Technology gap ratio	0.9421	1.0000	0.2766	1.0003	0.1242	
	Metafrontier, CE^*	0.9326	1.0000	0.2768	1.2117	0.1588	
Cluster 4	Group efficiency, CE^g	0.9811	1.0000	0.2493	1.5218	0.1863	0.3284
	Technology gap ratio	0.9358	0.9788	0.4822	1.0070	0.0949	
	Metafrontier, CE^*	0.9176	1.0000	0.2344	1.4207	0.1882	
Cluster 5	Group efficiency, CE^g	1.0022	1.0000	0.4365	1.2965	0.1010	0.4231
	Technology gap ratio	0.9187	0.9985	0.3577	1.0045	0.1449	
	Metafrontier, CE^*	0.9209	1.0000	0.3049	1.1232	0.1658	
Cluster 6	Group efficiency, CE^g	0.9557	1.0000	0.3818	1.2284	0.1382	0.1361
	Technology gap ratio	0.9190	1.0000	0.4005	1.1682	0.1547	
	Metafrontier, CE^*	0.8805	1.0000	0.3115	1.2286	0.2037	
Cluster 7	Group efficiency, CE^g	0.9987	1.0000	0.8312	1.1401	0.0404	0.3944
	Technology gap ratio	0.9090	1.0000	0.3203	1.0829	0.1680	
	Metafrontier, CE^*	0.9095	1.0000	0.3203	1.1193	0.1756	
Cluster 8	Group efficiency, CE^g	0.9655	1.0000	0.3860	1.0712	0.1303	0.3182
	Technology gap ratio	0.8718	0.9760	0.3092	1.1522	0.1856	
	Metafrontier, CE^*	0.8476	0.9791	0.3092	1.1568	0.2254	

Table 9: Order-m efficiencies and metatechnology ratios, clusters based on size

Group		Mean	Median	Minimum	Maximum	Std.Dev.	% eff. obs.
Cluster 1	Group efficiency, CE^g	0.9011	1.0000	0.2428	1.6983	0.2008	0.1068
	Technology gap ratio	0.8972	0.9773	0.3659	1.1949	0.1541	
	Metafrontier, CE^*	0.8904	1.0000	0.2344	1.4207	0.2048	
Cluster 2	Group efficiency, CE^g	0.9665	1.0000	0.3955	1.6893	0.1388	0.5979
	Technology gap ratio	0.9683	1.0000	0.5640	1.1364	0.0749	
	Metafrontier, CE^*	0.9417	1.0000	0.3772	1.1364	0.1343	
Cluster 3	Group efficiency, CE^g	0.9788	1.0000	0.6376	1.0020	0.073	0.9014
	Technology gap ratio	0.9866	1.0000	0.7564	1.0000	0.0491	
	Metafrontier, CE^*	0.9787	1.0000	0.6376	1.0000	0.0730	

Table 10: Distribution hypothesis tests (Li, 1996; Simar and Zelenyuk, 2006)

Efficiency distributions compared	t-statistic	p-value
Metafrontier vs. output mix clusters	3.4814	0.0002
Metafrontier vs. environmental clusters	7.8215	0.0000
Metafrontier vs. size clusters	2.7929	0.0026

cluster considered, the environmental conditions faced by the different municipalities play a remarkable role, leading us to mislabel them as inefficient. Note also that the differences found between CE^g and CE^* are irrespective of the number of observations in each cluster.

Figure 2 shows densities estimated using kernel smoothing of unconditioned and cluster-specific frontiers. The tighter probability mass at unity shows that observations are indeed much closer to those in their groups than to observations in other groups. These differences are also significant, as indicated by the p-values in Table 10, which were obtained by applying the Simar and Zelenyuk (2006) test. The substantial amount of probability mass found at the upper tail of the output mix distribution (Figure 2.a) indicates that controlling for group membership contributes in a more modest way of explaining efficiency differentials than in the case of clusters formed using environmental variables.

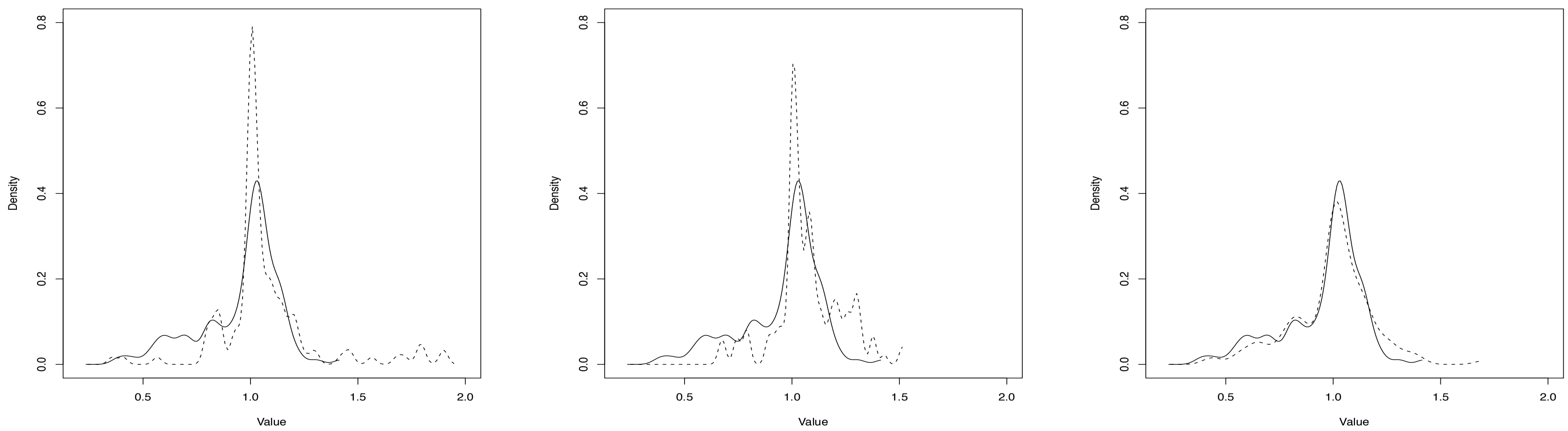
6. Concluding remarks

Over the last few years, a relevant area of research in the field of public economics and regional science and urban economics has been the analysis of the efficiency of lower layers of government such as regional governments or, as it is the case in this study, local governments. The topic is not only relevant *per se*, but also because of recent events such as the economic and financial crisis. In some countries such as Spain, economic activity has stalled, resulting in a sharp reduction of tax revenues and a simultaneous rapid increase of public spending. Under these circumstances, the efficient management of public resources at all levels of government becomes even more important.

Among the different levels of government, the literature devoted to the analysis of municipality efficiency is now relatively large. However, although the number of theoretical and applied contributions is high, there are still some critical issues which remain unsolved. One refers to the definition of municipality output. This is a thorny question, especially if we take into account that, on the one hand, municipalities face a budget constraint and, on the other, the law requires them to provide a minimum amount of services and facilities. A related issue is the remarkably varied environmental conditions—in our case defined as socioeconomic variables—that each municipality faces, which may have a marked effect on their performance.

We deal with these issues using a two-stage procedure. In the first stage, municipalities are classified into groups using cluster analysis taking into account variables based on their output mixes and environmental conditions. We identify groups of municipalities that share some important features in these fields, and the differences among them were statistically significant. In the second stage we assess how

Figure 2: Densities for order-m efficiency scores, metafrontier vs. clusters, year 2000



a) **Metafrontier vs. output mix** b) **Metafrontier vs. environmental** c) **Metafrontier vs. size** results vary when considering that each municipality should be compared with those in its group rather than with all municipalities.

The efficiency literature has dealt with the existence of groups and varying environmental conditions following different criteria. Some authors (Battese and Rao, 2002; Battese et al., 2004; O'Donnell et al., 2008) have proposed a methodology to compare the decision making units in different groups under the assumption that technology differs across groups and, therefore, one should estimate both group frontiers and a metafrontier. We deal with these concepts in the context of the order-m frontiers proposed by Cazals et al. (2002), in order to tackle relevant problems such as the existence of outliers and the curse of dimensionality. The severity of these problems, especially the latter, has not been fully acknowledged by the efficiency literature (Simar and Wilson, 2008, p.441).

Our results indicate that both hypotheses are relevant, especially that referring to the relevance of different environmental conditions—i.e., it is essential to control for the environment surrounding each municipality. Although the literature on efficiency had proposed ways to control for them, our methods are more robust to outliers and alleviate the curse of dimensionality. However, we must admit the groups constructed, both using output mix and environmental variables, were partly subjective in terms of number of groups and composition because of the technique employed in their formation—cluster analysis. Ideally, the research agenda should address how to define groups more objectively when natural boundaries between them do not exist.

References

Andersen, P. and Petersen, N. C. (1993). A procedure for ranking efficient units in Data Envelopment Analysis. Management Science, 39(10):1261–1264.

Balaguer-Coll, M. T., Prior, D., and Tortosa-Ausina, E. (2007). On the determinants of local government performance: A two-stage nonparametric approach. European Economic Review, 51(2):425–451.

Banker, R. D. and Morey, R. C. (1986a). Efficiency analysis for exogenously fixed inputs and outputs. Operations Research, 34:513–521.

Banker, R. D. and Morey, R. C. (1986b). The use of categorical variables in Data Envelopment Analysis. Management Science, 32:1613–1627.

Barankay, I. and Lockwood, B. (2007). Decentralization and the productive efficiency of government: Evidence from Swiss cantons. Journal of Public Economics, 91(5-6):1197–1218.

Battese, G. E. and Rao, D. (2002). Technology gap, efficiency, and a stochastic metafrontier function. International Journal of Business, 1(2):87–93.

Battese, G. E., Rao, D., and O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. Journal of Productivity Analysis, 21(1):91–103.

Bennett, J. T. and DiLorenzo, T. J. (1982). Off-budget activities of local government: The bane of the tax revolt. Public Choice, 39(3):333–342.

Bos, J. W. B. and Kool, C. J. M. (2006). Bank efficiency: The role of bank strategy and local market conditions. Journal of Banking and Finance, 30(7):1953–1974.

Bosch, N., Pedraja, F., and Suárez-Pandiello, J. (2000). Measuring the efficiency of Spanish municipal refuse collection services. Local Government Studies, 26(3):71–90.

Bradford, D., Malt, R., and Oates, W. (1969). The rising cost of local public services: Some evidence and reflections. National Tax Journal, 22:185–202.

Brueckner, J. K. (1981). Congested public goods: The case of fire protection. Journal of Public Economics, 15:45–58.

Brueckner, J. K. and Wingler, T. L. (1984). Public intermediate inputs, property values, and allocative efficiency. Economics Letters, 14:245–250.

Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. Journal of Econometrics, 106:1–25.

Daraio, C. and Simar, L. (2007). Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications. Studies in Productivity and Efficiency. Springer, New York.

De Borger, B. and Kerstens, K. (1996). Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. Regional Science and Urban Economics, 26:145–170.

Deller, S. C. (1992). Production efficiency in local government: A parametric approach. Public Finance/Finances Publiques, 47:32–44.

El-Gamal, M. and Inanoglu, H. (2005). Inefficiency and heterogeneity in Turkish banking: 1990–2000. Journal of Applied Econometrics, 20:641–664.

El-Mahgary, S. and Lahdelma, R. (1995). Data Envelopment Analysis: Visualizing the results. European Journal of Operational Research, 83(3):700–710.

Fox, K. J. (2001). Efficiency in the Public Sector, volume 1 of Studies in Productivity and Efficiency. Kluwer Academic Publishers, Boston.

Glass, J. C., McKillop, D. G., and Hyndman, N. (1995). Efficiency in the provision of university teaching and research: An empirical analysis of U.K. universities. Journal of Applied Econometrics, pages 61–72.

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

Hayami, Y. (1969). Sources of agricultural productivity gap among selected countries. American Journal of Agricultural Economics, 51(3):564–575.

Hayami, Y. and Ruttan, V. W. (1970). Agricultural productivity differences among countries. The American Economic Review, 60(5):895–911.

Hayami, Y. and Ruttan, V. W. (1971). Agricultural Development: An International Perspective. Johns Hopkins University Press Baltimore.

Haynes, P. (2003). Managing Complexity in the Public Services. Open University Press.

Hughes, P. A. N. and Edwards, M. E. (2000). Leviathan vs. Lilliputian: A Data Envelopment Analysis of government efficiency. Journal of Regional Science, 40(4):649–669.

Joro, T. and Na, P. (2002). Data Envelopment Analysis in mutual fund evaluation: a critical review. Research Report 02-2, Department of Finance and Management Science, School of Business, University of Alberta, Edmonton, Alberta.

Kneip, A., Park, B. U., and Simar, L. (1998). A note on the convergence of nonparametric dea estimators for production efficiency scores. Econometric Theory, 14:783–793.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. Econometric Reviews, 15:261–274.

Li, Q., Maasoumi, E., and Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. Journal of Econometrics, 148(2):186–200.

Li, Q. and Racine, J. S. (2007). Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton and Oxford.

Lovell, C. A. K. and Pastor, J. T. (1997). Target setting: An application to a bank branch network. European Journal of Operational Research, 98(2):290–299.

Marlow, M. L. and Joulfaian, D. (1989). The determinants of off-budget activity of state and local governments. Public Choice, 63(2):113–123.

Maudos, J., Pastor, J. M., and Pérez, F. (2002). Competition and efficiency in Spanish banking sector: The importance of specialisation. Applied Financial Economics, 12(7):505–516.

McMillan, M. L. and Chan, W. H. (2006). University efficiency: A comparison and consolidation of results from stochastic and non-stochastic methods. Education Economics, 14(1):1–30.

Merrifield, J. (1994). Factors that influence the level of underground government. Public Finance Review, 22(4):462.

O'Donnell, C. and Westhuizen, G. (2002). Regional comparisons of banking performance in South Africa. South African Journal of Economics, 70(3):224–240.

O'Donnell, C. J., Prasada Rao, D. S., and Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empirical Economics, 34:231–255.

Pagan, A. and Ullah, A. (1999). Nonparametric Econometrics. Themes in modern econometrics. Cambridge University Press, Cambridge.

Ruggiero, J. (1995). On the measurement and causes of technical inefficiency in local public services: With an application to public education. Journal of Public Administration Research and Theory, 5(4):403–428.

Ruggiero, J. (2004). Performance evaluation when non-discretionary factors correlate with technical efficiency. European Journal of Operational Research, 159(1):250–257.

Sampaio de Sousa, M. and Stošić, B. (2005). Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers. Journal of Productivity Analysis, 24:157–181.

Schuster, E. (1985). Incorporating support constraints into nonparametric estimators of densities. Communications in Statistics, 14:1123–1136.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. Journal of Productivity Analysis, 20(3):391–424.

Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management Science, 44(1):49–61.

Simar, L. and Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. Journal of Productivity Analysis, 13(1):49–78.

Simar, L. and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In Fried, H., Lovell, C. A. K., and Schmidt, S. S., editors, The Measurement of Productive Efficiency, chapter 4, pages 421–521. Oxford University Press, Oxford, 2nd edition.

Simar, L. and Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. Econometric Reviews, 25(4):497–522.

Solé-Ollé, A. and Hortas Rico, M. (2008). Does urban sprawl increase the costs of providing local public services? Document de treball 6, Institut d'Economia de Barcelona, Barcelona.

Taylor, L. L. (1995). Allocative efficiency and local government. Journal of Urban Economics, 37:201–211.

Vilalta, M. and Mas, D. (2006). El gasto de carácter discrecional de los ayuntamientos y su financiación. ejercicios 2002 y 2003. Elementos de debate territorial 23, Diputació de Barcelona (Xarxa de Municipis), Barcelona.

Wand, M. P. and Jones, M. C. (1994). Multivariate plug-in bandwidth selection. Computational Statistics, 9:97–116.

Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. Journal of Business and Economic Statistics, 11(3):319–23.

Wilson, P. W. (1995). Detecting influential observations in Data Envelopment Analysis. Journal of Productivity Analysis, 6(1):27–45.

Worthington, A. and Lee, B. (2008). Efficiency, technology and productivity change in Australian universities, 1998–2003. Economics of Education Review, 27(3):285–298.