Which estimator to measure local governments' cost efficiency? Evidence from Spanish municipalities

Isabel Narbón-Perpiñá
Universitat Jaume I

Maria Teresa Balaguer-Coll
Universitat Jaume I

Marko Petrović
Universitat Jaume I

Emili Tortosa-Ausina
Universitat Jaume I and Ivie

Abstract

We analyse the overall cost efficiency in local governments in Spain during the crisis period (2008–2013). For this, we first consider not only the most popular methods to evaluate local governments' efficiency, DEA (Data Envelopment Analysis) and FDH (Free Disposal Hull), but also more recent proposals such as the order-m partial frontier as well as the nonparametric estimator proposed by Kneip, Simar and Wilson (2008), which share their nonparametric flavour. Second, we compare the methodologies employed to measure efficiency. Contrary to previous literature, where there has been a regular comparison between techniques and several proposals of alternative techniques, we follow the method employed in the study of Badunenko et al. (2012), with the aim to compare the different methods used and choose the one which performs better with our particular dataset, i.e., the one which is more appropriate to measure local government cost efficiency in Spain. We carry out the experiment via Monte Carlo simulations and discuss the relative performance of the efficiency scores under various scenarios. We find that the results of the experiment depend on the value of Λ parameter, i.e., the relative sizes of the inefficiency and the error terms. Results suggest that our particular sample of 1,574 Spanish local governments lies in scenario 6, where DEA and FDH methodologies did the best job at estimating the efficiency scores given what we found in our simulations. Accordingly, our results indicate that the average cost efficiency would have been between 0.54 and 0.77 during the period 2008–2013

Keywords: efficiency, local government, nonparametric frontiers

JEL Classification: C14, C15, H70, R15

1. Introduction

Managing the available resources efficiently at all levels of government (central, regional, and municipal) is essential, even more if we consider the current international economic crisis scenario (Balaguer-Coll et al., 2013). Given that increasing taxes as well as deficit is politically costly (Doumpos and Cohen, 2014), a reasonable way to deal with this context is to improve economic efficiency (De Witte and Geys, 2011), which in cost terms means that an entity should produce a particular level of output in the cheapest way. In this setting, since local regulators must provide the best possible local services at the lowest possible cost, developing a system for evaluating local government performance that would allow to set benchmarks over time could have practical relevance (Da Cruz and Marques, 2014). However, measuring the performance of local governments is usually highly complex.

The study of local government efficiency has been a topic of high interest in the field of public administration. In fact, we find a large body of literature that covers several countries (such as Balaguer-Coll et al. (2007) in Spain, Geys et al. (2013) in Germany or Štastná and Gregor (2015) in Czech Republic).[1] However, despite the high number of empirical contributions, an important problem shared by the studies which analyse local government performance is the lack of a clear and standard methodology to perform efficiency analysis. This is not a trivial question as previous literature has widely proposed different frontier techniques, both parametric and nonparametric, to analyse technical, cost or other forms of efficiency in local governments.

Despite this problem is a well-known issue on the efficiency measurement literature, only a small number of studies has attempted to use two or more alternative approaches in a comparative way. For instance, De Borger and Kerstens (1996a) analysed local governments in Belgium using five different reference technologies, two nonparametric (DEA and FDH) and three parametric frontiers (one deterministic and two stochastic). They found large differences in the efficiency scores for identical samples and, as a consequence, they suggested using different methods to control for the robustness of the results as long as the problem of choosing the "best" reference technology is not solved. Other studies compared the efficiency estimates

[1] For a comprehensive literature review on efficiency measurement in local governments see Narbón-Perpiñá and De Witte (2016a,b).

between DEA and SFA,[2] or between DEA and FDH or other nonparametric variants,[3] leading to similar conclusions.

Since there is no obvious way to choose an efficiency estimator, the method chosen may affect the efficiency analysis (Geys and Moesen, 2009b) and could provide biased results. Therefore, if the decision makers at the local governments' level (based on an incorrect efficiency score) set a benchmark, this could result in a non-negligible economic impact. Accordingly, as indicated by Badunenko et al. (2012), if the selected method overestimates the efficiency scores, some local governments may not be penalised and, as a result, their inefficiencies will persist. In contrast, if the efficiency scores are underestimated some local governments would be regarded as "low performers" and could be unnecessarily penalised. Hence, although we note that each particular methodology leads to different cost efficiency results for each local government, one should ideally report efficiency scores that will be more reliable, or closer, to the *truth* (Badunenko et al., 2012).[4]

The present investigation pays attention to these issues by comparing different nonparametric methodologies and uncovering which measures might be more appropriate to assess local government cost efficiency in Spain. More specifically, the study contributes to the literature in three aspects. First, we seek to compare four nonparametric methodologies that cover traditional and recently developed nonparametric framework, namely Data Envelopment Analysis (DEA), Free Disposal Hull (FDH), order-*m* frontier and the bias corrected DEA estimator proposed by Kneip et al. (2008), being the first two the most popular towards the nonparametric field, and the last ones two more recent proposals.

Second, we attempt to determine which of these methods should be applied for cost efficiency measurement in a given situation. In contrast to previous literature, where there has been a regular comparison between techniques and several proposals of alternative ones, we follow the method employed in the study of Badunenko et al. (2012), with the aim to compare the different methods used and choose the ones which perform better in different settings. We carry out the experiment via Monte Carlo simulations and we discuss the relative performance of the efficiency estimators under various scenarios.

[2]Athanassopoulos and Triantis (1998); Worthington (2000); Geys and Moesen (2009b); Boetti et al. (2012); Nikolov and Hrovatin (2013); Pevcin (2014)

[3]Balaguer-Coll et al. (2007); Fogarty and Mugera (2013); El Mehdi and Hafner (2014)

[4]We will elaborate further on this *a priori* ambitious expression.

As a final contribution, we uncover which methodologies perform better with our particular dataset. From the simulation results, we determine in which scenario our data lies in, and we follow the suggestions related to the performance of the estimators for this scenario. We focus on a sample of 1,574 Spanish local governments between 1,000 and 50,000 inhabitants for the period 2008–2013. While other studies based on Spanish data (as well as data from other countries) focus on a specific region or year, our study examines a much larger sample of Spanish municipalities comprising various regions for several years. Moreover, the relevance of the sample is also related to the period under analysis. The economic and financial crisis which started in 2007/2008 has had a huge impact on most of Spanish local governments' revenues and finances in general. In addition, the budget constraints became stricter with the law on budgetary stability[5], which set up more control on public debt and public spending. Under these circumstances, issues related to the efficiency of Spanish local governments gains some relevance and momentum.

Our results suggest that our particular sample of 1,574 Spanish local governments lies in scenario 6, where DEA and FDH methodologies did the best job at estimating the efficiency scores given what we found in our simulations (DEA slightly underestimate efficiency while FDH slightly overestimate it). Accordingly, our results indicate that the average cost efficiency would have been between 0.54 and 0.77 during the period 2008–2013, suggesting that Spanish local governments could have achieved the same level of local outputs with about 0.23 to 0.36 fewer resources. In addition, DEA and KSW methodologies did the best job at identifying the ranks of the efficiency scores.

The paper is organised as follows: Section 2 gives an overview of the methodologies used to determine the cost efficiency. Section 3 specifies the particularities of the data employed. Section 4 shows the methodological comparison experiment and the results for the different scenarios. Section 5 gives a suggestion of which methodology performs better with our particular dataset and presents and comments the most relevant efficiency results. Finally, section 6 summarizes the main conclusions.

[5] *Ley General Presupuestaria* (2007,2012), or General Law on the Budget.

2. Methodologies

In the present study, we focus on the *efficiency*[6] of the provision of public goods and services. It is possible to distinguish different types of efficiency depending on the data available for inputs and outputs. In this way, *technical efficiency* requires data on inputs and outputs quantities, while *allocative efficiency* requires additional information on input prices. When these two measures are combined, we obtain the *economic efficiency*, also called *cost efficiency* when the economic objective is based on cost minimization. However, if data on costs is available, but data on prices and physical units is not, cost efficiency can be measured but not decomposed (Balaguer-Coll et al., 2007). We must notice that public sector goods and services are frequently unpriced, due to their non-market nature (Kalb et al., 2012). Within this context, since there is no data available about input prices, in the present study we measure local government cost efficiency using data on municipal budgets as input costs.

In addition, we consider four different nonparametric techniques to measure cost efficiency, namely, Data Envelopment Analysis (DEA) (Charnes et al., 1978; Banker et al., 1984), Free Disposal Hull (FDH) (Deprins et al., 1984), order-m (Cazals et al., 2002), and the bias corrected DEA estimator of Kneip et al. (2008), which we will refer to as KSW, being the first two the most popular within the nonparametric field and the others two relatively recent proposals. We focus on nonparametric methodologies as opposed to the parametric ones, due to their less restrictive assumptions and greater flexibility. For a detailed review of the main differences between parametric and nonparametric frontier techniques, see Murillo-Zamorano (2004) and Bogetoft and Otto (2010). In addition, the evolution of parametric and nonparametric methodologies has not been parallel, and several proposals have leaned towards the nonparametric field, overcoming most of their limits (Daraio and Simar, 2007; Bădin et al., 2014).

2.1. Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH)

DEA (Charnes et al., 1978; Banker et al., 1984) is a nonparametric methodology based on linear programming to estimate and compare the relative efficiency of different units (DMUs). We consider an input-oriented DEA model (Sampaio de Sousa and Stošić, 2005; Balaguer-Coll

[6]See Coelli et al. (2005) and Fried et al. (2008) for an introduction to efficiency measurement.

et al., 2007) because in public sector outputs are established externally (the minimum services that local governments must provide) and, consequently, it is more appropriate to evaluate efficiency in terms of the minimization of inputs (Balaguer-Coll and Prior, 2009).

We introduce the mathematical formulation for the cost efficiency measurement (Färe et al., 1994). The minimal cost efficiency can be calculated by solving the following program for each local government and each sample year:

$$
\begin{aligned}
& min_{\theta,\lambda}\theta \\
& \quad s.t. \quad && y_{ri} \leq \textstyle\sum_{i=1}^{n} \lambda_i y_{ri}, && r = 1, \ldots, p \\
& && \theta x_{ji} \geq \textstyle\sum_{i=1}^{n} \lambda_i x_{ji}, && j = 1, \ldots, q \\
& && \lambda_i \geq 0, && i = 1, \ldots, n \\
& && \textstyle\sum_{i=1}^{n} \lambda_i = 1
\end{aligned}
\tag{1}
$$

where for n observations there are q inputs producing p outputs. The $n \times p$ output matrix, r, and the $n \times q$ input matrix, j, represent the data for all n local governments. Specifically, for each unit (municipality) under evaluation i we consider an input vector x_{ji} to produce outputs y_{ri}. The last constraint ($\sum_{i=1}^{n} \lambda_i = 1$) implies variable returns to scale (VRS), which assures that each DMU is compared only with others of similar sizes.

A further extension of DEA model with variable returns to scale was proposed by Deprins et al. (1984), called Free Disposal Hull (FDH). The main difference with DEA is that it drops the convexity assumption. The FDH cost efficiency is defined as follows:

$$
\begin{aligned}
& min_{\theta,\lambda}\theta \\
& \quad s.t. \quad && y_{ri} \leq \textstyle\sum_{i=1}^{n} \lambda_i y_{ri}, && r = 1, \ldots, p \\
& && \theta x_{ji} \geq \textstyle\sum_{i=1}^{n} \lambda_i x_{ji}, && j = 1, \ldots, q \\
& && \lambda_i \in \{0, 1\}, && i = 1, \ldots, n
\end{aligned}
\tag{2}
$$

Finally, by solving linear programming problems (1) and (2) we obtain the cost efficiency coefficient θ, i.e., the optimal (minimal) input quantities of producing y_r. Local governments with efficiency scores of $\theta < 1$ are inefficient, while efficient units receive efficiency scores of $\theta = 1$.

2.2. Robust variants of DEA and FDH

The traditional nonparametric techniques DEA and FDH have been widely applied in efficiency analysis; however, they present several drawbacks. One limitation of both DEA and FDH is that they are sensitive to outliers and extreme values. Since these techniques envelope all data, the efficient frontier is determined by the observations which are extreme points (Simar and Wilson, 2008) and, as a consequence, the presence of outliers strongly influence the estimated frontier as well as the efficiency scores of all observations. This problem can be addressed by using "partial" frontiers for being more robust to extremes or outliers in data. Moreover, these "partial" estimators do not suffer from the "curse of dimensionality[7]", an important problem that generally affects efficiency scores obtained using DEA and FDH (Daraio and Simar, 2007). Finally, another important drawback of traditional nonparametric approaches is the difficulty of making statistical inference. Nevertheless, bootstrap methods such as those proposed by Simar and Wilson (1998, 2000) made possible statistical inferences (consistent analysis, bias correction, confidence intervals, test of hypothesis and so on) about efficiency.

Hence, in this study we consider some variants of DEA and FDH estimators that are able to overcome some crucial drawbacks of the *traditional* nonparametric methods. On the one hand, we will use order-m (Cazals et al., 2002), which is a partial frontier approach that mitigates the influence of outliers, extreme values and the curse of dimensionality. On the other hand, the bias corrected DEA estimator of Kneip et al. (2008) (KSW), which allows for consistent statistical inference by applying bootstrap techniques.

2.2.1. Order-m

Order-m frontier (Cazals et al., 2002) is a robust alternative to DEA and FDH estimators which involves the concept of partial frontier, opposed to the traditional full frontier. The order-m estimator, for finite m units, does not envelope all data points and consequently, is less extreme. In the input orientation case, this method uses as benchmark the expected minimum input achievable among a fixed number of m units producing at least output level y. Hence,

[7] As indicated by Daraio and Simar (2007), the "curse of dimensionality" implies that an increase in the number of inputs or outputs, or a decrease in the sample under analysis (i.e., the number of units for comparison), implies higher efficiencies.

the order-m input efficiency score (Daraio and Simar, 2007) is given by:

$$\hat{\theta}_m(x,y) = E[(\hat{\theta}_m(x,y)|Y \geqslant y)] \tag{3}$$

The value m represents the number of potential units against we benchmark the analysed unit (i.e., how efficient is a local government compared with m local governments.). If m goes to infinity, the order-m estimator converges to FDH. As suggested by Daraio and Simar (2005), the most reasonable value of m is determined as the value for which the super-efficient observations becomes constant.

Note that order-m scores are not bounded by 1. A value greater than 1 indicates super-efficiency, showing that the unit operating at the level (x, y) is more efficient than the average of m peers randomly drawn from the population of units producing more output than y (Daraio and Simar, 2007).

2.2.2. Bias corrected DEA estimator of Kneip et al. (2008) (KSW)

The KSW (Kneip et al., 2008) is a bias corrected DEA estimator which derives the asymptotic distribution of DEA via bootstrapping techniques. As indicated by Simar and Wilson (2008), DEA and FDH estimators are biased by construction, implying that the *true* frontier would be located under the DEA-estimated frontier. As a consequence, DEA scores (i.e., relative to the estimated frontier) are too "optimistic". The bootstrap procedure to correct this bias, based on sub-sampling, uses the idea that the known distribution of the difference between estimated and bootstrapped efficiency scores mimics the unknown distribution of the difference between the true and the estimated efficiency scores (Badunenko et al., 2012). In addition, the KSW procedure allows for consistent statistical inference of efficiency estimates (i.e., bias and confidence intervals for the estimated efficiency scores.).

Therefore, in order to implement the bootstrap procedure (based on sub-sampling), first let $s = n^d$ for some $d \in (0,1)$, where n and s are the sample and sub-sample size, respectively. The optimal d depends on the dimensionality of the problem. Following, the bootstrap considers the following scheme:

1. First, a bootstrap sample $S_s^* = (X_i^*, Y_i^*)_{i=1}^s$ is generated by drawing (independently, uniformly and with replacement) s observations from the original sample, S_n.

2. DEA estimator is applied, where the technology set is constructed with the sub-sample drawn in step (1), to construct the bootstrap estimates $\hat{\theta}^*(x,y)$.

3. Steps (1) and (2) are repeated B times, using the resulting bootstrap values to approximate the conditional distribution of $s^{2/(p+q+1)}(\frac{\hat{\theta}^*(x,y)}{\theta^*(x,y)} - 1)$, which allows to approximate the unknown distribution of $n^{2/(p+q+1)}(\frac{\hat{\theta}^*(x,y)}{\theta^*(x,y)} - 1)$. The values p and q are the output and input quantities, respectively. The bias-corrected DEA efficiency score, which is adjusted by the s subsample size, is given by:

$$\theta_{bc}(x,y) = \theta^*(x,y) - Bias^* \tag{4}$$

where the bias is adjusted by employing the s sub-sample size.

$$Bias^* = \left(\frac{s}{n}\right)^{2/(p+q+1)} \left[\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*_b(x,y) - \theta^*(x,y)\right] \tag{5}$$

4. Finally, for a given $\alpha \in (0,1)$, the bootstrap values are used to find the quantiles $\delta_{\alpha/2,s}$, $\delta_{1-\alpha/2,s}$ in order to compute a symmetric $1-\alpha$ confidence interval for $\theta(x,y)$

$$\left[\frac{\hat{\theta}(x,y)}{1+n^{-2/(p+q+1)}\delta_{1-\alpha/2,s}}, \frac{\hat{\theta}(x,y)}{1+n^{-2/(p+q+1)}\delta_{\alpha/2,s}}\right] \tag{6}$$

3. Sample, data and variables

We consider a sample of Spanish local governments between 1,000 and 50,000 inhabitants for the 2008–2013 period. The information on inputs and outputs was obtained from the Spanish Ministry of the Treasury and Public Administrations (*Ministerio de Hacienda y Administraciones Públicas*). In particular, outputs were obtained from a survey on local infrastructures and facilities (*Encuesta de Infraestructuras y Equipamientos Locales*). Information on inputs was obtained from local governments' budget expenditures. The final sample contains 1,574 Spanish municipalities for every year, after removing all the observations for which information on inputs or outputs was not available for the sample period (2008–2013). Specifically, there was no information for the Basque Country, Navarre,[8] the regions of Catalonia and Madrid, as well as the provinces of Burgos, Huesca, Guadalajara and Huelva.

Inputs are derived from the local governments' budget expenditures, and are representative of the cost of the municipal services provided. Using budget expenditures as inputs is consistent with the literature (e.g., Balaguer-Coll et al., 2007, 2010; Zafra-Gómez and Muñiz-Pérez, 2010; Fogarty and Mugera, 2013; Da Cruz and Marques, 2014). We construct an input measure, which represents the total local governments' costs (X_1), by including different municipal expenditures as personnel expenses, expenditures on goods and services, current transfers, capital investments and capital transfers.

Outputs are related to the minimum specific services and facilities provided by each municipality. Our selection is based on the article 26 of the Spanish law which regulates the local system (*Ley reguladora de Bases de Régimen Local*), which establishes the minimum services and facilities that each municipality must provide compulsorily—depending on their size. Specifically, all governments must provide public street lighting, cemeteries, waste collection and street cleaning services, drinking water to households, sewage system, access to population centres, paving of public roads, and regulation of food and drink. The selection of outputs is consistent with the literature (e.g., Balaguer-Coll et al., 2007; Balaguer-Coll and Prior, 2009; Zafra-Gómez and Muñiz-Pérez, 2010; Bosch-Roca et al., 2012). Note that differently from previous studies in other European countries, Spanish local governments do not have any responsibility in the area of education, care for elderly and health services.

As a result, in an attempt to generate a balanced set of outputs, we have chosen 6 output variables to measure services and facilities that municipalities provide. Due to the difficulties in measuring public sector outputs, in some cases it is necessary to use proxy variables, an assumption which has been widely applied in the literature. Based on the studies by De Borger and Kerstens (1996a,b), many of these output variables should be considered as crude proxies for the services delivered by municipalities, due to the unavailability of more direct outputs.

Table 1 reports the minimum services that all local government must provide for the 2008–2013 period, as well as the different output indicators used to evaluate the services, whereas table 2 reports descriptive statistics for inputs and outputs for the same period. We include

[8]The Basque Country and Navarre do not have to present this information to the Spanish Ministry of the Treasury and Public Administrations because they have its own autonomous system and, consequently, they are not included in the State Economic Cooperation.

the median instead of the mean with the intention of avoiding the outliers' distortion.

4. Methodological comparison

Differently to previous literature, in this section we compare DEA, FDH, order-m and KSW approaches following the method proposed by Badunenko et al. (2012). Our aim is to uncover which measures perform better with our particular dataset, i.e., the ones which are more appropriate to measure local governments' efficiency in Spain.

Therefore, we carry out the experiment via Monte Carlo simulations. We first define the data generating process, the parameters and the distributional assumptions on data. Second, we consider the different methodologies and we take several standard measures to compare their behaviour. Next, after running the simulations, we discuss the relative performance of the efficiency estimators under the various scenarios. Finally, we decide which methods are more appropriate to measure local governments' efficiency in Spain.

4.1. Simulations

Most previous studies which analysed local governments' cost efficiency with parametric techniques used the Stochastic Frontier Approach (SFA) developed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977) as a model to estimate cost frontiers.[9] They considered the input-oriented efficiency where the dependent variable is the level of spending or cost, and the independent variables are output levels. As a parametric approach, SFA establishes the best practice frontier on the basis of a specific functional form, most commonly Cobb-Douglas or Translog. Moreover, it allows to distinguish between measurement error and inefficiency term.

Following this scheme, we conduct simulations for a production process with one input or cost (c) and two outputs (y_1 and y_2).[10] We consider a Cobb-Douglas cost function (CD). For the baseline case, we assume constant returns to scale (CRS) ($\gamma = 1$).[11] We establish $\alpha = 1/3$ and $\beta = \gamma - \alpha$.

[9]See, for instance, the studies of Worthington (2000), De Borger and Kerstens (1996a), Geys (2006), Ibrahim and Salleh (2006), Geys and Moesen (2009b,a), Kalb (2010), Geys et al. (2010), Kalb et al. (2012) or Šťastná and Gregor (2015), Lampe et al. (2015), among others.

[10]For simplicity, we use a multi-output model with two outputs instead of six.

[11]In subsection 4.4, we consider robustness checks with increasing and decreasing returns to scale to make sure that our simulations accurately represent the performance of our methods.

We simulate observations for outputs y_1 and y_2, which are distributed uniformly on the $[1, 2]$ interval. Moreover, we assume that the true error term (v) is normally distributed $N(0, \sigma_v^2)$ and the true cost efficiency is $TCE = exp(-u)$, where u is half-normally distributed $N(0, \sigma_u^2)$ and independent from v. We introduce the true error and inefficiency terms in the frontier formulation, which takes the following expression:

$$c = y_1^{\alpha} \cdot y_2^{\beta} \cdot exp(v + u), \tag{7}$$

where c are total costs and y_1 and y_2 are output indicators. For reasons previously explained in section 2, there are no observable variation on input prices, so that input prices are ignored (see, for instance, the studies of Kalb (2012), Pacheco et al. (2014)).

We simulate six different combinations for the error and inefficiency terms, in order to model various real scenarios. Table 3 contains the matrix of the different scenarios. It shows the combinations when σ_v takes values 0.01 and 0.05 and σ_u takes values 0.01, 0.05 and 0.1. The rows in the table represent the variation of the error term, while the columns represent the variation of the inefficiency term. The first row is the case where the variation of the error term is relatively small, while the second row shows a large variation. The first column is the case where the inefficiency term is relatively small, while the second and third columns represent the cases where the inefficiency variation are relatively larger. Λ parameter, which sets each scenario, is the ratio between of σ_u and σ_v.

Within this context, scenario 1 is the case when the error and the inefficiency terms are relatively small ($\sigma_u = 0.01$, $\sigma_v = 0.01$, $\Lambda = 1.0$), which means that the data has been measured with little noise and the units are relatively efficient, while scenario 6 is the case when the error and the inefficiency terms are relatively large ($\sigma_u = 0.1$, $\sigma_v = 0.05$, $\Lambda = 2.0$), which means that the data is relatively noisy and the units are relatively inefficient. For all simulations we consider 2,000 Monte Carlo trials, and we analyse two different sample sizes, n= 100 and 200.[12]

[12]To ease the computational process, we use n= 100 and 200 sample size to conduct simulations. In subsection 4.4, we consider a robustness check with a bigger sample size ($n = 500$) to make sure that our simulations accurately represent the performance of our data.

4.2. Measures to compare the estimators' performance

In order to compare the relative performance of the methodologies, we consider the following median measures, over the 2,000 simulations. We use median values instead of the average, since it is more robust to skewed distributions.

- $Bias(TCE) = \frac{1}{n}\sum_{i=1}^{n}(\widehat{TCE_i} - TCE_i)$
- $RMSE(TCE) = [\frac{1}{n}\sum_{i=1}^{n}(\widehat{TCE_i} - TCE_i)^2]^{1/2}$
- $UpwardBias(TCE) = \frac{1}{n}\sum_{i=1}^{n} 1 \cdot (\widehat{TCE_i} > TCE_i)$
- Kendall's τ (TCE)= $\frac{n_c - n_d}{0.5n(n-1)}$

where $\widehat{TCE_i}$ is the estimated cost efficiency of municipality i in a given Monte Carlo replication (by a given method) and TCE_i is the true efficiency score. The Bias reports the difference between the estimated and true efficiency scores. When it is negative (positive), the estimators are underestimating (overestimating) the true efficiency. The $RMSE$ (Root Mean Squared Error) measures the standard deviation or error from the true efficiency. The Upward Bias is the proportion of $\widehat{TCE}$ larger than the true ones. It measures the percentage of overestimated or underestimated cost efficiencies. Finally, the Kendall's τ test represents the correlation between the predicted and true cost efficiencies, being n_c and n_d the number of concordant and discordant pairs in the data set, respectively. It identifies the differences in the ranking distributions of the true and the estimated ranks.

Additionally, we also compare the densities of cost efficiency across all Monte Carlo simulations, in order to report a more comprehensive description of the results, and not only restrict them to a single summary statistic—the median. For each draw, we sort the data by the relative value of true efficiency. Since we are interested in comparing the true distribution for different percentiles of our sample, we show violin plots for 5%, 50% and 95% percentiles.

4.3. Relative performance of the estimators

Tables 4 provides baseline results for the performance measures of the cost efficiency with the CD cost function. First, looking at the bias of the cost efficiency scores, we observe that the median bias is always negative in DEA and KSW. This implies that DEA and KSW estimators tend to underestimate the true cost efficiency in all scenarios. FDH and order-m present positive median bias except for scenario 2 in FDH. This implies that FDH and order-m tend to overestimate the true efficiency. Bias for all methodologies tend to increase with the sample size when the bias is negative and decrease when the bias is positive, except for order-m in scenarios 1, 3 and 5. Focusing on RMSE, it is smaller when σ_v is small, except for FDH in scenario 5 and order-m in scenarios 3 and 5. Moreover, we can see that the RMSE of the cost efficiency estimates increase with the sample size for all cases except for FDH in scenarios 1, 3, 5 and 6 and order-m in scenarios 5 and 6.

We also consider the Upward Bias. It shows the percentage of observations for which cost efficiency is larger than the true value (returning a value of 1). The desired value is 0.5. The values less (greater) than 0.5 indicates underestimation (overestimation) of cost efficiencies. In this setting, DEA and KSW systematically underestimate the true efficiency. Moreover, as the sample size goes up the percentage of underestimated results increases. On the contrary, FDH and order-m tend to overestimate the true efficiency but, as the sample size goes up overestimated results go down. Finally, we analyse Kendall's τ for the efficiency ranks between true and estimated efficiency scores. In each scenario and sample size, DEA and KSW have a larger Kendall's τ, so they do a better job at identifying the ranks.

We also analyse other percentiles of the efficiency distribution, since it is difficult to conclude from the table which methods perform better. Figures 1 to 3 show results for the 5^{th}, 50^{th} and 95^{th} percentiles of true and the estimated cost efficiencies. We compare the distribution of each method with the TCE. For visual simplicity, we show only the case when $n = 100$. Figures with sample size $n = 200$ are not much different and are available upon request.

Figures indicate that results depend on the value of Λ parameter. As expected, when the variance of the error term increase our results are less accurate (note that nonparametric methodologies assume the absence of noise). On the contrary, when the variance of the inefficiency term increase, our results are more precise.

Under **scenario 1** (see Figures 1a, 1c and 1e), when both error and inefficiency terms are relatively small, DEA and KSW methodologies consistently underestimate efficiency (their distributions are below the true efficiency in all percentiles). If we consider median values and density modes, order-m tends to overestimate efficiency in all percentiles, while FDH also tends to overestimate efficiency at the 5^{th} and 50^{th} percentiles. Moreover, we observe that FDH does a good job at estimating the efficiency units when looking at the 95^{th} percentile.

Although **scenario 4** (see Figures 2b, 2d and 2f) is the opposite case to scenario 1, when both error and inefficiency terms are relatively large, they have the same value of Λ. As in scenario 1, DEA and KSW methodologies consistently underestimate efficiency. Otherwise, when looking at the 5^{th} percentile, both FDH and order-m tend to overestimate efficiency. However, at the 50^{th} and 95^{th} percentiles both methods do a better job at estimating the efficiency units since their median values and density modes are closer to the TCE distribution.

Similarly, in **scenario 2** (see Figures 1b, 1d and 1f), when the error term is relatively large while the inefficiency term is relatively small, DEA and KSW tend to underestimate the true efficiency scores, while FDH and order-m appear to be close to the TCE distribution (in terms of median values and mode). This scenario yields to the worst results since the dispersion of TCE is much squeezed compared to the estimators' distributions. Therefore, when Λ is small, all methodologies do a poorer job at predicting the efficiency scores.

Scenario 3 (see Figures 2a, 2c and 2e) is the case when the error term is relatively small while the inefficiency term is relatively large. Since Λ value increase, all methodologies do a better job at predicting the efficiency scores. At the 5^{th} and 50^{th} percentiles, we observe that DEA and KSW underestimate efficiency, while order-m and FDH tend to overestimate it. However, if we consider the median and density modes, DEA (followed by KSW) is closer to the TCE distribution in both percentiles. At the 95^{th} percentile FDH does a better job at estimating the efficient units, while DEA and KSW slightly underestimate efficiency and order-m slightly overestimate it.

Otherwise, **scenario 5** (see Figures 2a, 2c and 2e) is the case when the error variation is relatively small while the inefficiency variation is very large. This scenario shows the most favourable results because the dispersion of the TCE distribution is much scattered and, as a consequence, it represents better the performance of the estimators. At the 5^{th} and 50^{th} percentiles DEA and KSW densities are very close to the true distribution of efficiency, while FDH and order-m overestimate it. Otherwise, at the 95^{th} percentile FDH seems to be closer to the TCE although it slightly overestimates the true efficiency.

Finally, **scenario 6** (see Figures 3b, 3d and 3f) is the case when the error term is relatively large and the inefficiency term is even larger. Again, we observe that when the variation of the inefficiency term increases (compared with scenario 2 and 4), all the estimators perform better. At the 5^{th} and 50^{th} percentiles, DEA and KSW slightly underestimate efficiency and FDH and order-m slightly overestimate it. However, despite all methods are quite close to the TCE distribution, DEA underestimates less than KSW and FDH overestimates less than order-m. Finally, at the 95^{th} percentile FDH (followed by order-m) is the best method to determine a higher number of efficient units because its mode and median values are closer to the true efficiency.

4.4. Robustness checks

We consider a number of robustness checks to verify that our baseline experiment represent the performance of our estimators. Results for each robustness are given in Appendices A-D.

- No noise: All our nonparametric estimators assume the absence of noise. However, in the baseline experiment we include noise in each scenario. In this situation, we consider the case where there is no noise in the data generating process. Results show that DEA, and KSW perform better at predicting the efficiency scores, while FDH and order-m remain slightly worse to the baseline experiment. Moreover, all methods do a better job at estimating the true ranks, except order-m in scenario 1. In short, we find that when the absence of noise exists, DEA and KSW perform better.

- Changes in sample size: The baseline experiment analyses two different sample sizes, n= 100 and 200. We also consider the case where the sample size is very large, that is, the case where n= 500. There exist a little deterioration in DEA and KSW performance, while FDH and order-m vary depending on the scenario. However, the results only differed slightly. We do not see any qualitative changes from the baseline results.

- Returns to scale: The baseline experiment assumes CRS technology. We also consider the case where the technology assumes decreasing and increasing returns to scale ($\gamma = 0.8$ and $\gamma = 1.2$). We find the performance of DEA and KSW estimators slightly deteriorated. Performance for order-m is improved with decreasing returns to scale and deteriorated with increasing returns to scale, while FDH varies depending on the scenario. However, despite these minor quantitative differences, the qualitative results do not change.

- Different m values for order-m: As suggested by Daraio and Simar (2007), in order to choose the most reasonable value of m, we have considered different m sizes ($m = 20, 30$ and 40). In our application, the baseline experiment sets $m = 30$. In general, when comparing with the other m values we observe some quantitative changes (i.e., performance with $m = 20$ gets worse, while with $m = 40$ slightly improves), however the qualitative results from the baseline case seem to hold.

4.5. Which estimator in each scenario

Based on the prior analysis on the relative performance of the different methodologies, we sum up which and when they must be used, assuming that the simulations remain true for different data generating processes. Table 5 suggests which estimators to use for each scenario when taking into account the efficiency scores. The first row in each scenario shows the relative size of the median estimates, while the rest of the rows suggest which estimators use depending on the percentile (5^{th}, 50^{th} or 95^{th}). In some cases different methodologies are relatively similar in terms of identifying the efficiency scores.

As the study of Badunenko et al. (2012) concludes, if Λ value is small, as in scenario 2 ($\Lambda = 0.2$), the efficiency scores and ranks will be poorly estimated[13]. This scenario yields to the worst results, since the estimators are far from the "truth". Despite in table 5 we have suggestions for scenario 2, we do not recommend efficiency analysis for this particular scenario, since it would be inaccurate.

Although scenario 1 and 4 present better results than scenario 2 (when $\Lambda = 1$), estimators also do a poor job at predicting the true efficiency scores. In scenario 1, FDH seems to be the best method to estimate efficiency in all percentiles, however, also DEA should be considered at the 5^{th} percentile (the TCE remains between DEA and FDH at this percentile). Similarly, in scenario 4 FDH is the dominant method at the 5^{th} percentile, however, DEA also should be considered. Otherwise, both FDH and order-m perform better at the 50^{th} and 95^{th} percentiles. When looking at the efficiency rankings, DEA and KSW methodologies do a fair job at ranking the observations in both scenarios.

Similarly, scenario 6 performs better than scenario 1 and 4, since the variation of the inefficiency term increases and, as a consequence, the value of Λ also increases ($\Lambda = 2$). In

[13]It is difficult to obtain the inefficiency from a relatively large noise component.

this scenario seems that the methodologies which estimate better the true efficiency scores are DEA and FDH at the 5^{th} and 50^{th} percentiles, and FDH (followed by order-m) at the 95^{th} percentile. Otherwise, when focusing on the rankings, DEA and KSW methodology do a better job at ranking the observations.

In scenario 3, the value of Λ increases again ($\Lambda = 5$), and all methodologies do a better job at predicting the efficiency scores. When looking at the 5^{th} and 50^{th} percentiles, DEA (followed by KSW) seems to be the estimator closer to the true efficiency. At the 95^{th} percentile FDH is the method which does the best job. Otherwise, when considering the rankings, DEA and KSW are the methods which estimate the efficiency rankings better.

Finally, scenario 5 is the case where the value of Λ is larger ($\Lambda = 10$). Here, the estimators do a good job at estimating efficiency and ranks. DEA (followed by KSW) performs better when looking at the 5^{th} and 50^{th} percentiles and FDH at the 95^{th} percentile. In addition, DEA and KSW do a really good job at estimating the efficiency rankings.

5. Which estimator performs better with Spanish local governments'

Finally, in this section we define which methodologies are more appropriate to measure local governments' efficiency in Spain. First, we estimate Λ values for our particular dataset via the nonparametric kernel estimator of Fan et al. (1996), hereafter FLW[14]. The estimated Λ value helps to determine the scenario in which our data lies in. Second, we have to refer to table 5 and choose the appropriate estimators for our particular needs. Therefore, table 6 reports results of Λ parameters for our sample of 1,574 Spanish local governments between 1,000 and 50,000 inhabitants for the 2008–2013 period.

As we can observe, results of Λ estimates are closer to 2, which corresponds to scenario 6. We see that the goodness-of-fit measure (R^2) of our empirical data remains around 0.8. The summary statistics for the overall cost-efficiency results averaged over all municipalities for each year can be found in table 7. Moreover, figure 4 provide the violin plots of the estimated cost efficiencies to show further interpretation of results[15].

In scenario 6, DEA and FDH were the methods which performed better at the 5^{th} and 50^{th}

[14]In the appendix we describe how to obtain Λ measures via FLW derived from a cost function.

[15]For visual simplicity, we plot together years 2008–2013, however they do not differ much and individual plots are available upon request.

percentiles of the distribution (the former slightly underestimates efficiency while the latter slightly overestimates it), and FDH was the method which performed better at the the 95^{th} percentile. Therefore, the true efficiency would remain between DEA and FDH results, given what we found in our simulations. In this context, DEA results indicate that the average cost efficiency during the period 2008–2013 is 0.54, while the average in FDH is 0.77, so we expect that the true cost efficiency scores are found between 0.54 and 0.77. Moreover, looking at the lower quartile (Q1), average scores are 0.42 in DEA and 0.61 in FDH, so we expect that the true efficiency scores at the lower end of the distribution are between 0.42 and 0.61. Similarly, looking at the upper quartile (Q3), FDH average scores are 0.99, and we expect that these estimated efficiencies are similar to the true ones.

Looking at the efficiency scores shown by KSW, they are smaller than those reported by DEA and FDH (the average efficiency scores in KSW for the period 2008–2013 is 0.48). Our belief, based on our Monte Carlo simulations, is that KSW methodology consistently underestimates the true efficiency scores. On the contrary, all the statistics estimated by order-*m* methodology are bigger than those shown in DEA and FDH (the average efficiency scores in order-*m* for the period 2008–2013 is 0.83). Therefore, given what we have seen in the experiment, we believe that order-*m* method overestimates the true efficiency scores.

Otherwise, when looking at the rank estimates we must note that in scenario 6, DEA and KSW methodologies did the best job at identifying the ranks of the efficiency scores. Table 8 shows the rank correlation between the cost efficiency estimates of the different methodologies. As we observed in our Monte Carlo experiment, DEA and KSW have a high correlation between their rank estimates since they have a similar distribution of the rankings. Moreover, despite the fact that there exists a relatively high correlation between order-*m* and FDH rank estimates with DEA and KSW, these last two outperform order-*m* and FDH.

6. Conclusion

Over the last years, there have been many empirical research studies that have focused on the evaluation of efficiency in local governments. However, despite the high number of empirical contributions, there is still a lack of a clear and standard methodology to perform efficiency analysis. Since there is no obvious way to choose an estimator, the method chosen may affect the efficiency results, and could provide "unfair" or biased results. Therefore, we note that each particular methodology leads to different cost efficiency results for each local government, but one must provide efficiency scores that will be more reliable or closer to the *truth* (Badunenko et al., 2012).

In this setting, the current paper has attempted to compare four different nonparametric estimators, which are DEA, FDH, order-m and KSW. All these approaches are well studied in previous literature, but little is known about their performance related to each other. In contrast to previous literature, where there has been a regular comparison between techniques and several proposals of alternative ones, we have followed the method employed in the study of Badunenko et al. (2012), with the aim to compare the different methods used and choose the ones which performed better with our particular dataset, i.e., the ones which are more appropriate to measure local government cost efficiency in Spain.

We have found that the results of the experiment depend on the value of Λ parameter, i.e., the relative sizes of σ_v and σ_u. Our results are more precise when the ratios of the inefficiency term to the error term are larger, while they are less accurate when the ratio is smaller.

We estimate Λ values for our sample of Spanish local governments for the 2008–2013 period. The results suggest that our particular dataset lies in scenario 6. In this scenario, DEA and FDH methodologies did the best job at estimating the efficiency scores given what we found in our simulations. Accordingly, our results indicate that the average true cost efficiency would be between 0.54 and 0.77 during the period 2008–2013, suggesting that Spanish local governments could achieve the same level of local outputs with about 0.23 to 0.36 fewer resources. In addition, when looking at the rank estimates in scenario 6, DEA and KSW methodologies did the best job at identifying the ranks of the efficiency scores.

References

Aigner, D., Lovell, C. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Athanassopoulos, A. D. and Triantis, K. P. (1998). Assessing aggregate cost efficiency and the related policy implications for Greek local municipalities. *INFOR*, 36(3):66–83.

Bădin, L., Daraio, C., and Simar, L. (2014). Explaining inefficiency in nonparametric production models: the state of the art. *Annals of Operations Research*, 214(1):5–30.

Badunenko, O., Henderson, D. J., and Kumbhakar, S. C. (2012). When, where and how to perform efficiency estimation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(4):863–892.

Balaguer-Coll, M. T. and Prior, D. (2009). Short-and long-term evaluation of efficiency and quality. An application to Spanish municipalities. *Applied Economics*, 41(23):2991–3002.

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.

Balaguer-Coll, M. T., Prior, D., and Tortosa-Ausina, E. (2010). Decentralization and efficiency of local government. *The Annals of Regional Science*, 45(3):571–601.

Balaguer-Coll, M. T., Prior, D., and Tortosa-Ausina, E. (2013). Output complexity, environmental conditions, and the efficiency of municipalities. *Journal of Productivity Analysis*, 39(3):303–324.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9):1078–1092.

Boetti, L., Piacenza, M., and Turati, G. (2012). Decentralization and local governments' performance: how does fiscal autonomy affect spending efficiency? *FinanzArchiv: Public Finance Analysis*, 68(3):269–302.

Bogetoft, P. and Otto, L. (2010). *Benchmarking with DEA, SFA, and R*, volume 157. Springer Science & Business Media, New York.

Bosch-Roca, N., Mora-Corral, A. J., and Espasa-Queralt, M. (2012). Citizen control and the efficiency of local public services. *Environment and Planning C: Government and Policy*, 30(2):248.

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

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., and Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. Springer Science & Business Media, Heidelberg.

Da Cruz, N. F. and Marques, R. C. (2014). Revisiting the determinants of local government performance. *Omega*, 44:91–103.

Daraio, C. and Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis*, 24(1):93–121.

Daraio, C. and Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, volume 4. Springer Science & Business Media, New York.

De Borger, B. and Kerstens, K. (1996a). Cost efficiency of Belgian local governments: a comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics*, 26(2):145–170.

De Borger, B. and Kerstens, K. (1996b). Radial and nonradial measures of technical efficiency: an empirical illustration for Belgian local governments using an FDH reference technology. *Journal of Productivity Analysis*, 7(1):41–62.

De Witte, K. and Geys, B. (2011). Evaluating efficient public good provision: theory and evidence from a generalised conditional efficiency model for public libraries. *Journal of Urban Economics*, 69(3):319–327.

Deprins, D., Simar, L., and Tulkens, H. (1984). *The Performance of Public Enterprises: Concepts and Measurements*, chapter Measuring labor inefficiency in post offices, pages 243–267. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, NorthHolland.

Doumpos, M. and Cohen, S. (2014). Applying Data Envelopment Analysis on accounting data to assess and optimize the efficiency of Greek local governments. *Omega*, 46:74–85.

El Mehdi, R. and Hafner, C. M. (2014). Local government efficiency: the case of Moroccan municipalities. *African Development Review*, 26(1):88–101.

Fan, Y., Li, Q., and Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business & Economic Statistics*, 14(4):460–468.

Färe, R., Grosskopf, S., and Lovell, C. K. (1994). *Production Frontiers*. Cambridge University Press, Cambridge.

Fogarty, J. and Mugera, A. (2013). Local government efficiency: evidence from Western Australia. *Australian Economic Review*, 46(3):300–311.

Fried, H. O., Lovell, C. K., and Schmidt, S. S. (2008). *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press, New York.

Geys, B. (2006). Looking across borders: a test of spatial policy interdependence using local government efficiency ratings. *Journal of Urban Economics*, 60(3):443–462.

Geys, B., Heinemann, F., and Kalb, A. (2010). Voter involvement, fiscal autonomy and public sector efficiency: evidence from German municipalities. *European Journal of Political Economy*, 26(2):265–278.

Geys, B., Heinemann, F., and Kalb, A. (2013). Local government efficiency in German municipalities. *Raumforschung und Raumordnung*, 71(4):283–293.

Geys, B. and Moesen, W. (2009a). Exploring sources of local government technical inefficiency: evidence from Flemish municipalities. *Public Finance and Management*, 9(1):1–29.

Geys, B. and Moesen, W. (2009b). Measuring local government technical (in)efficiency: an application and comparison of FDH, DEA and econometric approaches. *Public Performance and Management Review*, 32(4):499–513.

Ibrahim, F. W. and Salleh, M. F. M. (2006). Stochastic frontier estimation: an application to local governments in Malaysia. *Malaysian Journal of Economic Studies*, 43(1/2):85.

Kalb, A. (2010). The impact of intergovernmental grants on cost efficiency: theory and evidence from German municipalities. *Economic Analysis and Policy*, 40(1):23–48.

Kalb, A. (2012). What determines local governments' cost-efficiency? The case of road maintenance. *Regional Studies*, 48(9):1–16.

Kalb, A., Geys, B., and Heinemann, F. (2012). Value for money? German local government efficiency in a comparative perspective. *Applied Economics*, 44(2):201–218.

Kneip, A., Simar, L., and Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24(06):1663–1697.

Lampe, H., Hilgers, D., and Ihl, C. (2015). Does accrual accounting improve municipalities' efficiency? Evidence from Germany. *Applied Economics*, 47(41):4349–4363.

Meeusen, W. and Van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–444.

Murillo-Zamorano, L. R. (2004). Economic efficiency and frontier techniques. *Journal of Economic Surveys*, 18(1):33–77.

Narbón-Perpiñá, I. and De Witte, K. (2016a). Local governments' efficiency: a systematic literature review – Part I. Technical Report No. 20/2016, Economics Department, Universitat Jaume I, Spain.

Narbón-Perpiñá, I. and De Witte, K. (2016b). Local governments' efficiency: a systematic literature review – Part II. Technical Report No. 21/2016, Economics Department, Universitat Jaume I, Spain.

Nikolov, M. and Hrovatin, N. (2013). Cost efficiency of Macedonian municipalities in service delivery: does ethnic fragmentation matter? *Lex Localis*, 11(3):743.

Pacheco, F., Sanchez, R., and Villena, M. (2014). A longitudinal parametric approach to estimate local government efficiency. Technical Report No. 54918, Munich University Library, Germany.

Pevcin, P. (2014). Efficiency levels of sub-national governments: a comparison of SFA and DEA estimations. *The TQM Journal*, 26(3):275–283.

Sampaio de Sousa, M. C. and Stošić, B. (2005). Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers. *Journal of Productivity Analysis*, 24(2):157–181.

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. *The Measurement of Productive Efficiency (H. Fried, CAK Lovell and SS Schmidt Eds), Oxford University Press, Oxford*, pages 421–521.

Štastná, L. and Gregor, M. (2015). Public sector efficiency in transition and beyond: evidence from Czech local governments. *Applied Economics*, 47(7):680–699.

Worthington, A. C. (2000). Cost efficiency in Australian local government: a comparative analysis of mathematical programming anf econometrical approaches. *Financial Accountability & Management*, 16(3):201–223.

Zafra-Gómez, J. L. and Muñiz-Pérez, Antonio, M. (2010). Overcoming cost-inefficiencies within small municipalities: improve financial condition or reduce the quality of public services? *Environment and Planning C: Government and Policy*, 28(4):609–629.

Estimation of Λ

We use the following semi-parametric stochastic cost frontier model:

$$C_i = g(y_i) + \varepsilon_i, \qquad i = 1, \ldots.., n, \tag{8}$$

where y_i is a $p \times 1$ vector of random regressors (outputs), $g(.)$ is the unknown smooth function and ε_i is a composed error term which has two components: (1) v_i, the two-sided random error term which is assumed to be normally distributed N(0, σ_v^2), and (2) u_i, the cost efficiency term which is half-normally distributed ($u_i \geq 0$). These two error components are assumed to be independent.

We use available data on cost (municipal budgets) due to the difficulty to use market prices to measure public services. Hence the assumption allows us to omit the factor prices from the model.

We derive the concentrated log-likelihood function $\ln l(\Lambda)$ and maximize it over the single parameter Λ:

$$\max_{\Lambda} \ln l(\Lambda) = \max_{\Lambda} \left\{ -n \ln \hat{\sigma} + \sum_{i=1}^{n} \ln \left[1 + \Phi\left(\frac{\hat{\epsilon}_i}{\hat{\sigma}} \Lambda \right) \right] - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} \hat{\epsilon}_i{}^2 \right\}, \tag{9}$$

with $\hat{\epsilon}_i = C_i - \hat{E}(C_i|y_i) + \mu(\hat{\sigma}^2, \Lambda)$ and

$$\sigma^2 = \left\{ \frac{1}{n} \sum_{i=1}^{n} [C_i - \hat{E}(C_i|y_i)]^2 \bigg/ \left[1 - \frac{2\Lambda^2}{\pi(1+\Lambda^2)} \right] \right\}^{1/2}, \tag{10}$$

where $\hat{E}(C_i|y_i)$ is the kernel estimator of the conditional expectation $E(C_i|y_i)$ and it is given as:

$$\hat{E}(C_i|y_i) = \sum_{j=1}^{n} C_j \cdot K\left(\frac{y_i - y_j}{h} \right) \bigg/ \sum_{j=1}^{n} K\left(\frac{y_i - y_j}{h} \right), \tag{11}$$

where $K(.)$ is the kernel function and $h = h_n$ is the smoothing parameter. For further details about the estimation procedure see Fan et al. (1996).

Table 1: Minimum services provided by all local governments and output variables

	Minimum services	Output indicators
In all municipalities:	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	Length water distribution networks (m)
	Sewage system	Length sewerage networks (m)
	Access to population centres	Street infrastructure surface area
	Paving of public roads	Street infrastructure surface area
	Regulation of food and drink	Total population

Table 2: Descriptive statistics for data in inputs and outputs, period 2008-2013

	Mean	**S.d.**
INPUTS[1]		
Total costs (X_1)	6,856,864.55	7,990,865.20
OUTPUTS		
Total population (Y_1)	7,555.36	8,460.33
Street infrastructure surface area[2](Y_2)	336,673.55	325,808.07
Number of lighting points (Y_3)	1,519.78	1,567.02
Tons of waste collected (Y_4)	4,216.73	19,720.07
Length water distribution networks[2] (Y_5)	50,503.12	93,877.89
Length sewerage networks[2] (Y_6)	29,650.29	32,424.83

[1]: In thousands of €.

[2]: In square meters.

Table 3: Combinations of error and inefficiency in Monte Carlo simulations to model scenarios

We simulate six different combinations for the error (σ_v) and inefficiency (σ_u) terms, in order to model various real scenarios. The rows represent the variation of the error term, while the columns represent the variation of the inefficiency term. Λ parameter is the ratio between of σ_u and σ_v, which sets each scenario.

	$\sigma_u = 0.01$	$\sigma_u = 0.05$	$\sigma_u = 0.1$
$\sigma_v = 0.01$	s1: $\Lambda = 1.0$	s3: $\Lambda = 5.0$	s5: $\Lambda = 10.0$
$\sigma_v = 0.05$	s2: $\Lambda = 0.2$	s4: $\Lambda = 1.0$	s6: $\Lambda = 2.0$

Table 4: Baseline results with Cobb-Douglas cost function

	Bias				RMSE				Upward Bias				Kendall's τ			
	DEA	**FDH**	**Order-m**	**KSW**	**DEA**	**FDH**	**Order-m**	**KSW**	**DEA**	**FDH**	**Order-m**	**KSW**	**DEA**	**FDH**	**Order-m**	**KSW**
s1: $\sigma_v = 0.01$, $\sigma_u = 0.01$																
n=100	-0.0298	0.0074	0.0231	-0.0350	0.0330	0.0096	0.0307	0.0375	0.0500	0.9800	0.9900	0.0100	0.2491	0.1227	0.0615	0.2549
n=200	-0.0348	0.0070	0.0287	-0.0391	0.0376	0.0094	0.0370	0.0415	0.0250	0.9600	0.9850	0.0050	0.2573	0.1590	0.0790	0.2588
s2: $\sigma_v = 0.05$, $\sigma_u = 0.01$																
n=100	-0.0892	-0.0111	0.0049	-0.1006	0.1013	0.0338	0.0399	0.1111	0.0500	0.6400	0.6900	0.0100	0.0681	0.0551	0.0511	0.0687
n=200	-0.1028	-0.0205	0.0043	-0.1130	0.1134	0.0428	0.0466	0.1225	0.0250	0.5000	0.6200	0.0050	0.0707	0.0597	0.0542	0.0705
s3: $\sigma_v = 0.01$, $\sigma_u = 0.05$																
n=100	-0.0182	0.0322	0.0477	-0.0246	0.0238	0.0392	0.0548	0.0285	0.1200	1.0000	1.0000	0.0600	0.6753	0.4443	0.3169	0.6877
n=200	-0.0239	0.0289	0.0512	-0.0293	0.0281	0.0351	0.0581	0.0325	0.0700	0.9900	1.0000	0.0350	0.6843	0.5192	0.3812	0.6911
s4: $\sigma_v = 0.05$, $\sigma_u = 0.05$																
n=100	-0.0707	0.0133	0.0303	-0.0832	0.0857	0.0410	0.0528	0.0960	0.0900	0.7500	0.8100	0.0400	0.3060	0.2547	0.2415	0.3059
n=200	-0.0849	0.0024	0.0279	-0.0963	0.0972	0.0421	0.0565	0.1072	0.0500	0.6300	0.7550	0.0250	0.3132	0.2710	0.2564	0.3123
s5: $\sigma_v = 0.01$, $\sigma_u = 0.1$																
n=100	-0.0101	0.0525	0.0684	-0.0177	0.0203	0.0624	0.0768	0.0238	0.2000	1.0000	1.0000	0.1000	0.8057	0.5928	0.5146	0.8182
n=200	-0.0170	0.0453	0.0689	-0.0232	0.0230	0.0537	0.0763	0.0275	0.1150	0.9950	1.0000	0.0600	0.8174	0.6586	0.5738	0.8254
s6: $\sigma_v = 0.05$, $\sigma_u = 0.1$																
n=100	-0.0580	0.0347	0.0519	-0.0724	0.0755	0.0591	0.0722	0.0869	0.1300	0.8200	0.8800	0.0700	0.4996	0.4170	0.4030	0.4990
n=200	-0.0726	0.0207	0.0485	-0.0854	0.0867	0.0530	0.0717	0.0974	0.0750	0.7200	0.8400	0.0400	0.5065	0.4403	0.4332	0.5065

Table 5: Relative performance of the efficiency scores at the 5^{th}, 50^{th} and 95^{th} percentiles

This table summarises which estimators to use for each scenario. The first row of each scenario represents the relative median values of the estimators, while the rest of the rows suggest which estimators use depending on the percentile (5^{th}, 50^{th} or 95^{th}).

	σ_u**=0.01**	σ_u**=0.05**	σ_u**=0.1**
σ_v**=0.01**	**s1**: **KSW<DEA<TCE<FDH<order-*m***	**s3**: **KSW<DEA<TCE<FDH<order-*m***	**s5**: **KSW<DEA<TCE<FDH<order-*m***
	5: DEA or FDH	5: DEA	5: DEA
	50: FDH	50: DEA	50: DEA
	95: FDH	95: FDH	95: FDH
σ_v**=0.05**	**s2**: **KSW<DEA<FDH<TCE<order-*m***	**s4**: **KSW<DEA<TCE<FDH<order-*m***	**s6**: **KSW<DEA<TCE<FDH<order-*m***
	5: order-*m* or FDH	5: DEA or FDH	5: DEA or FDH
	50: order-*m* or FDH	50: order-*m* or FDH	50: DEA or FDH
	95: order-*m* or FDH	95: order-*m* or FDH	95: order-*m* or FDH

Table 6: Estimates for Λ value for Spanish local governments

This table contains the results of the Λ parameters of our sample of 1,574 Spanish local governments for the period 2008–2013. Λ values help to determine the scenario in which our data lies in. We also report the goodness-of-fit measure (R^2) of our empirical data.

	2008	**2009**	**2010**	**2011**	**2012**	**2013**
Λ	2.0596	2.2143	1.7256	1.5953	1.8283	1.8371
R^2	0.7980	0.8331	0.8250	0.8244	0.8209	0.8478

Table 7: Summary statistics for efficiency results in Spanish local governments

	DEA						
	Mean	**Median**	**Min**	**Max**	**S.d.**	**Q1**	**Q3**
2008	0.4943	0.4689	0.0437	1.0000	0.1876	0.3611	0.6038
2009	0.5843	0.5740	0.1257	1.0000	0.1677	0.4633	0.6830
2010	0.5212	0.4953	0.1312	1.0000	0.1718	0.4017	0.6135
2011	0.5314	0.5092	0.1359	1.0000	0.1728	0.4104	0.6237
2012	0.5316	0.5128	0.1079	1.0000	0.1749	0.4077	0.6429
2013	0.5712	0.5591	0.1138	1.0000	0.1817	0.4458	0.6823
2008–2013	0.5390	0.5199	0.1097	1.0000	0.1761	0.4150	0.6415

	FDH						
	Mean	**Median**	**Min**	**Max**	**S.d.**	**Q1**	**Q3**
2008	0.7444	0.7678	0.0808	1.0000	0.2276	0.5644	1.0000
2009	0.8186	0.8563	0.2045	1.0000	0.1841	0.6821	1.0000
2010	0.7761	0.7848	0.1559	1.0000	0.1961	0.6251	1.0000
2011	0.7453	0.7434	0.2037	1.0000	0.2108	0.5808	0.9892
2012	0.7630	0.7737	0.1497	1.0000	0.2076	0.6104	1.0000
2013	0.7619	0.7721	0.1497	1.0000	0.2055	0.6104	0.9999
2008–2013	0.7682	0.7830	0.1574	1.0000	0.2053	0.6122	0.9982

	Order-m						
	Mean	**Median**	**Min**	**Max**	**S.d.**	**Q1**	**Q3**
2008	0.8089	0.8255	0.0834	1.9813	0.2353	0.6312	1.0000
2009	0.8691	0.8926	0.2122	1.7369	0.2005	0.7318	1.0013
2010	0.8385	0.8515	0.2172	1.8080	0.2032	0.6938	1.0000
2011	0.8088	0.8100	0.2368	2.0281	0.2197	0.6497	1.0000
2012	0.8222	0.8358	0.1797	1.8914	0.2169	0.6644	1.0000
2013	0.8209	0.8328	0.1785	1.9204	0.2175	0.6609	1.0000
2008–2013	0.8281	0.8414	0.1846	1.8944	0.2155	0.6720	1.0002

	KSW						
	Mean	**Median**	**Min**	**Max**	**S.d.**	**Q1**	**Q3**
2008	0.4421	0.4239	0.0400	1.0000	0.1720	0.3183	0.5454
2009	0.5384	0.5297	0.1179	1.0000	0.1575	0.4250	0.6370
2010	0.4541	0.4294	0.0563	1.0000	0.1603	0.3420	0.5399
2011	0.4752	0.4558	0.1178	1.0000	0.1572	0.3697	0.5558
2012	0.4677	0.4477	0.0134	1.0000	0.1650	0.3503	0.5687
2013	0.4846	0.4711	0.0118	1.0000	0.1619	0.3678	0.5848
2008–2013	0.4770	0.4596	0.0595	1.0000	0.1623	0.3622	0.5719

Table 8: Rank correlation Kendall coefficients between the cost efficiency estimates of the different methodologies

2008	DEA	FDH	Order-m	KSW
DEA	1.0000	0.6035	0.5976	0.9264
FDH	0.6035	1.0000	0.8114	0.5958
Order-m	0.5976	0.8114	1.0000	0.5661
KSW	0.9264	0.5958	0.5661	1.0000

2009	DEA	FDH	Order-m	KSW
DEA	1.0000	0.4736	0.5589	0.9300
FDH	0.4736	1.0000	0.7305	0.4506
Order-m	0.5589	0.7305	1.0000	0.5208
KSW	0.9300	0.4506	0.5208	1.0000

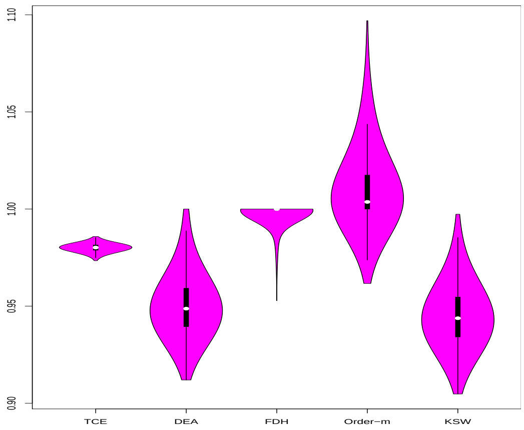
2010	DEA	FDH	Order-m	KSW
DEA	1.0000	0.4959	0.5817	0.9030
FDH	0.4959	1.0000	0.6964	0.4636
Order-m	0.5817	0.6964	1.0000	0.5314
KSW	0.9030	0.4636	0.5314	1.0000

2011	DEA	FDH	Order-m	KSW
DEA	1.0000	0.5931	0.5925	0.9098
FDH	0.5931	1.0000	0.8014	0.5456
Order-m	0.5925	0.8014	1.0000	0.5343
KSW	0.9098	0.5456	0.5343	1.0000

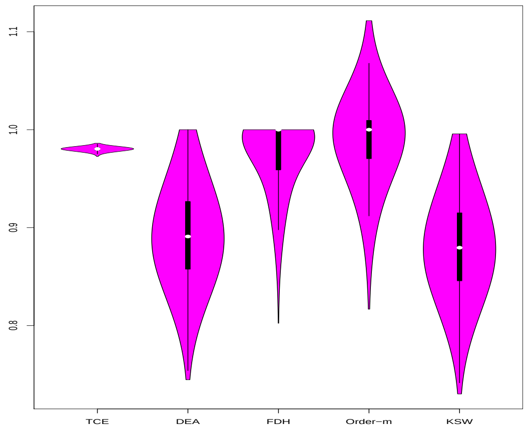
2012	DEA	FDH	Order-m	KSW
DEA	1.0000	0.6140	0.5652	0.9136
FDH	0.6140	1.0000	0.8093	0.5813
Order-m	0.5652	0.8093	1.0000	0.5121
KSW	0.9136	0.5813	0.5121	1.0000

2013	DEA	FDH	Order-m	KSW
DEA	1.0000	0.5237	0.5555	0.9128
FDH	0.5237	1.0000	0.7105	0.4817
Order-m	0.5555	0.7105	1.0000	0.4996
KSW	0.9128	0.4817	0.4996	1.0000

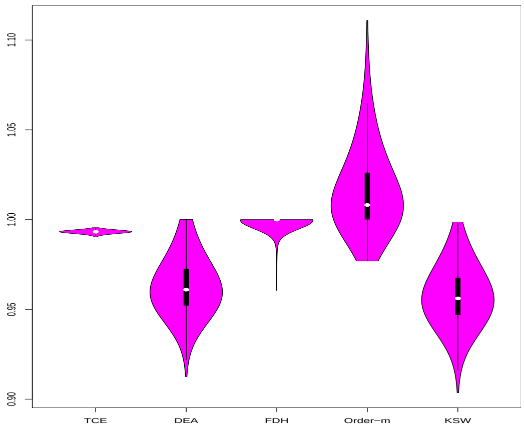
Figure 1: Violin plots in scenario 1 and 2 for 5^{th}, for 5^{th}, 50^{th} and 95^{th} percentiles of cost efficiency estimates under a Cobb-Douglas cost function.



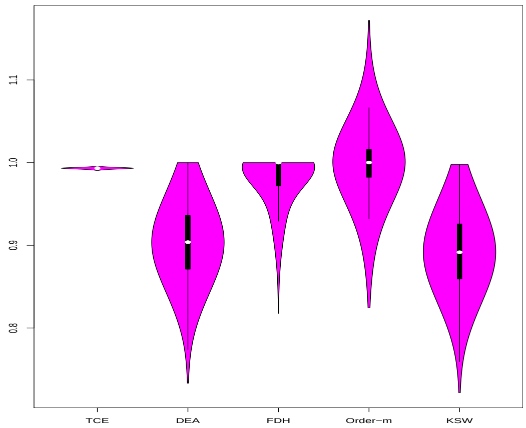
(a) S1, 5th percentile, n=100



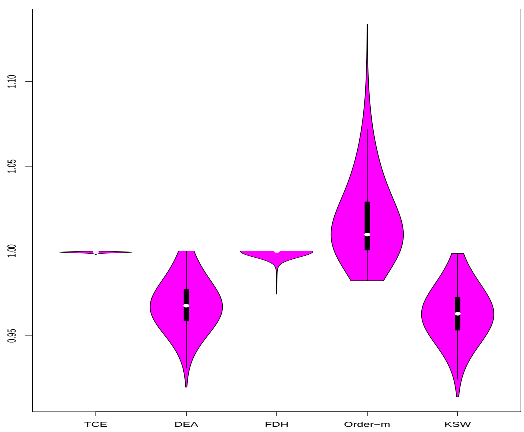
(b) S2, 5th percentile, n=100



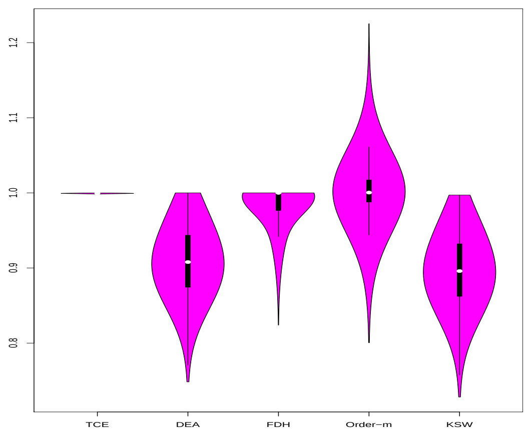
(c) S1, 50th percentile, n=100



(d) S2, 50th percentile, n=100

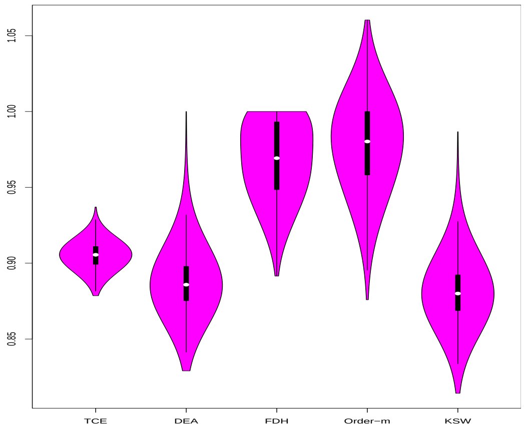


(e) S1, 95th percentile, n=100

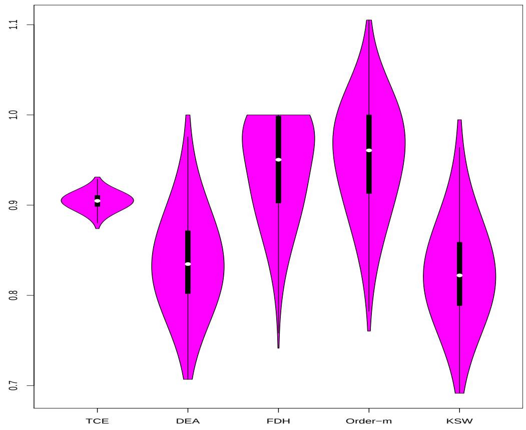


(f) S2, 95th percentile, n=100

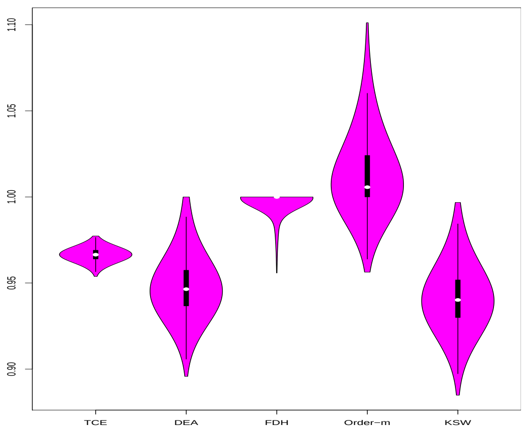
Figure 2: Violin plots in scenario 3 and 4 for 5th, for 5th, 50th and 95th percentiles of cost efficiency estimates under a Cobb-Douglas cost function.



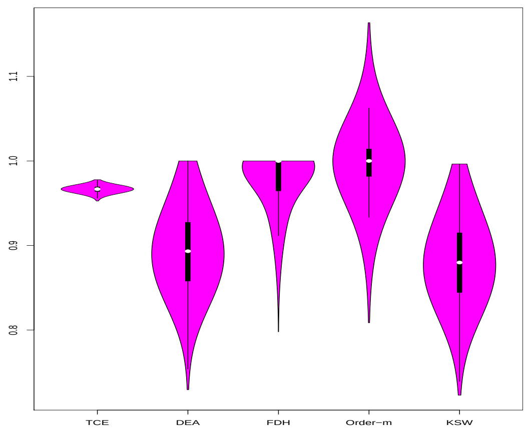
(a) S3, 5th percentile, n=100



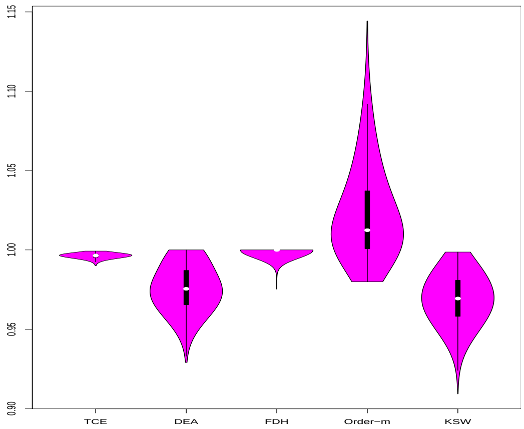
(b) S4, 5th percentile, n=100



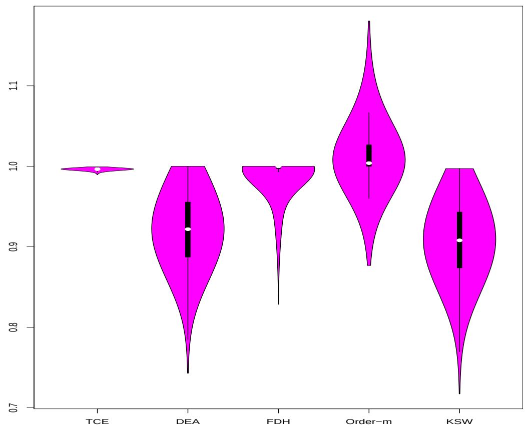
(c) S3, 50th percentile, n=100



(d) S4, 50th percentile, n=100

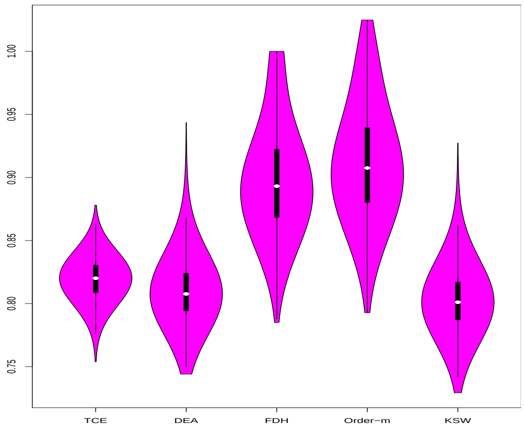


(e) S3, 95th percentile, n=100

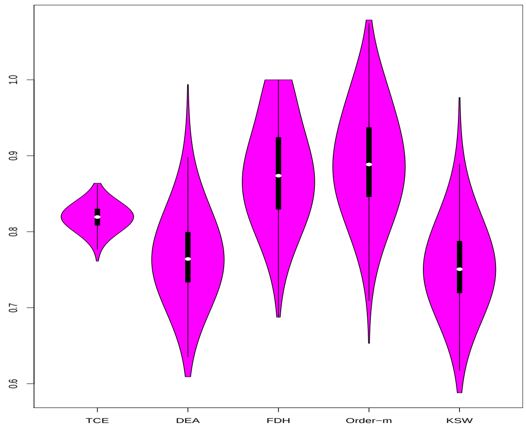


(f) S4, 95th percentile, n=100

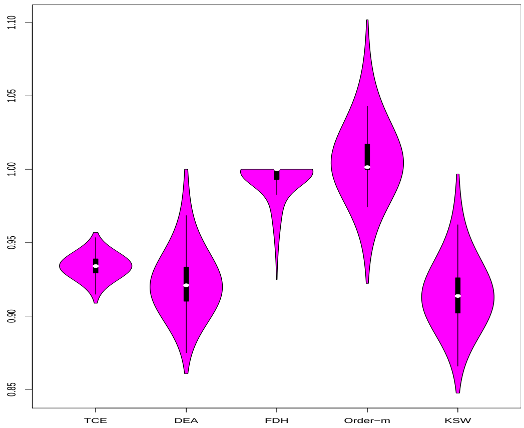
Figure 3: Violin plots in scenario 5 and 6 for 5th, for 5th, 50th and 95th percentiles of cost efficiency estimates under a Cobb-Douglas cost function.



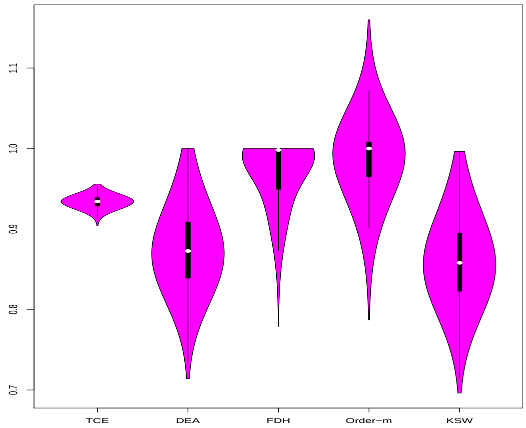
(a) S5, 5th percentile, n=100



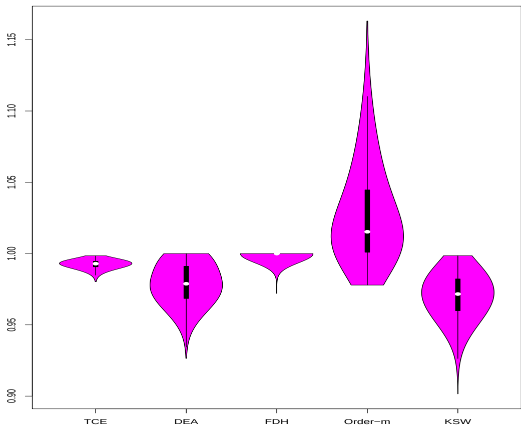
(b) S6, 5th percentile, n=100



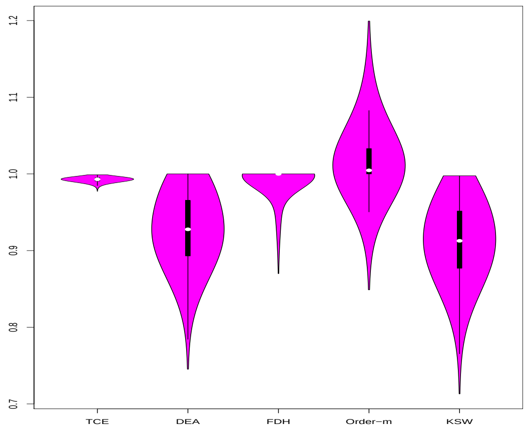
(c) S5, 50th percentile, n=100



(d) S6, 50th percentile, n=100



(e) S5, 95th percentile, n=100



(f) S6, 95th percentile, n=100

Figure 4: Violin plots of cost efficiency estimates in Spanish local governments

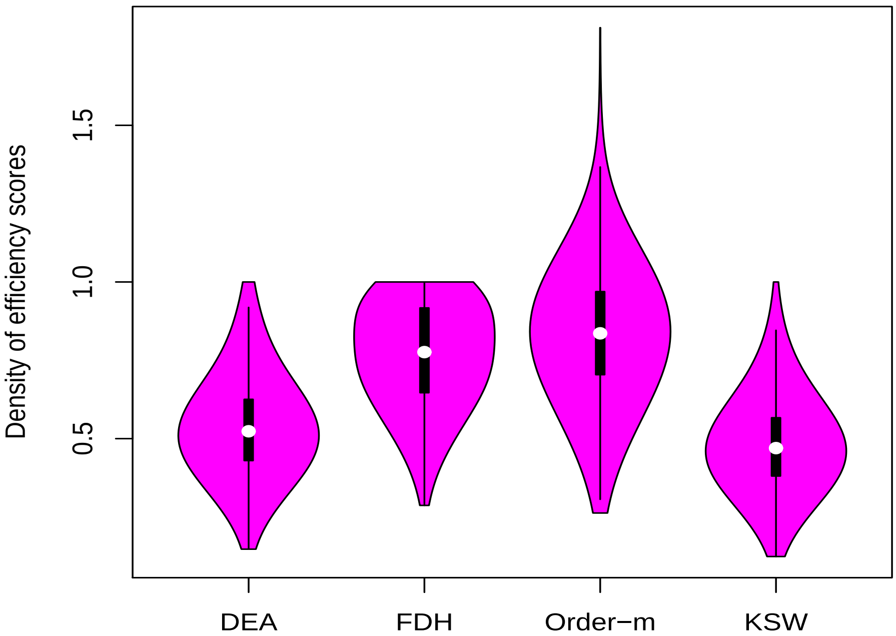


Figure 5: 2008–2013