ORIGINAL ARTICLE

Check for updates

Which estimator to measure local governments' cost efficiency? The case of Spanish municipalities

Isabel Narbón-Perpiñá[1] · Maria Teresa Balaguer-Coll[2] · Marko Petrović[3] · Emili Tortosa-Ausina[3]

Abstract

We analyse overall cost efficiency in Spanish local governments during the crisis period (2008–2015). To this end, we first consider some of the most popular nonparametric methods to evaluate local government efficiency, data envelopment analysis and free disposal hull, as well as recent proposals, namely the order-m partial frontier and the nonparametric estimator proposed by Kneip et al. (Econom Theory 24(6):1663–1697, 2008). Second, to compare the four methods and choose the most appropriate one for our particular context and dataset (local government cost efficiency in Spain), we carry out an experiment via Monte Carlo simulations and discuss the relative performance of the efficiency scores under various scenarios. Our results suggest that there is no one approach suitable for all efficiency analysis. We find that for our sample of 1846 Spanish local governments, the average cost efficiency would have been between 0.5417 and 0.7543 during the period 2008–2015, suggesting that Spanish local governments could have achieved the same level of local outputs with about 25% and 46% fewer resources.

✉ Emili Tortosa-Ausina
tortosa@uji.es

Extended author information available on the last page of the article

1 Introduction

Managing available resources efficiently at all levels of government (central, regional, and municipal) is essential, particularly in times of crisis, such as the one that until recently had serious effects on several European countries. Given that increasing taxes and deficit is politically costly (Doumpos and Cohen 2014), a reasonable way to operate in difficult circumstances (not only during crises) is to improve economic efficiency (De Witte and Geys 2011). Since local regulators must provide the best possible services at the lowest cost, developing a system for evaluating local government performance that allows benchmarks to be set over time could have relevant implications (Ferreira Da Cruz and Cunha Marques 2014). As a consequence, local government efficiency has attracted much scholarly interest in the field of public administration and the literature now is extensive (see, for comprehensive reviews, Narbón-Perpiñá and De Witte 2018a, b). However, despite the high number of empirical contributions, a major challenge is the lack of clear, standard methodology to evaluate municipalities' efficiency.

Although this problem is well known in the efficiency measurement literature, in local government most studies focus on one approach only, few have attempted to use two or more methods for comparative purposes. For instance, De Borger and Kerstens (1996a) analysed local governments in Belgium using five different reference technologies, two nonparametric (data envelopment analysis or DEA, and free disposal hull or 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, suggested using different methods to control for the robustness of results whenever the problem of choosing the "best" reference technology is unsolved. Other studies drew similar conclusions after comparing the efficiency estimates of DEA and the stochastic frontier approach (SFA),[1] or DEA and FDH or other nonparametric variants.[2] Some contributions have also attempted to compare parametric and nonparametric efficiency methods (mainly SFA with DEA or other DEA variants) by using Montecarlo simulations.[3,4] However, none of these studies concluded that there is a superior method (Andor and Hesse 2014).

Since there is no obvious way to choose an efficiency estimator, the method selected may affect the efficiency analysis (Geys and Moesen 2009b) and could lead to biased results. Therefore, if local government decision makers set a benchmark based on an "incorrect" efficiency score, the economic impact might not be neutral, mislabelling some inefficient municipalities as efficient and vice versa. Hence, although we note that the measurement technique might be not entirely neutral in the case of local

---

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

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

[3] Some important studies in this area are those of Banker et al. (1993), Banker et al. (1996), Ruggiero (1999), Resti (2000), Ruggiero (2007), Krüger (2012), Badunenko et al. (2012) and Andor and Hesse (2014).

[4] Essentially, these methods use generated artificial data, to investigate situations that influence the performance of the efficiency estimators under different assumptions by comparing the "true" efficiency levels to those estimated through one or more of these techniques (Resti 2000).

governments' efficiency, one should ideally report efficiency scores that are more reliable, or closer to the "truth" (Badunenko et al. 2012).[5]

The present investigation addresses these issues by comparing four nonparametric methodologies and uncovering which measures might be more appropriate to assess local government cost efficiency in Spain. In other words, we attempt to ascertain which method leads to the most reliable results when evaluating cost efficiency in our particular dataset. The study contributes to the literature in three specific aspects. First, we seek to compare four nonparametric methodologies that cover traditional and more recently developed nonparametric proposals, namely DEA, FDH, the order-m partial frontier (Cazals et al. 2002) and the bias-corrected DEA estimator proposed by Kneip et al. (2008). These techniques have been widely used in the previous literature, but little is known about their performance in comparison with each other.

Second, we attempt to determine which of these methods should be applied to measure cost efficiency in a given situation. In contrast to previous local government efficiency literature, which has regularly compared techniques and made alternative proposals, we carry out an experiment via Monte Carlo simulations, identifying those methods that *perform better* in different settings. In doing so, we adapt the simulated technology in order to adequately describe the characteristics of the local government sector by employing a cost function setting with multiple outputs. Finally, based on the simulation results, we discuss the relative performance of the efficiency estimators under various scenarios and seek to determine which method should be used in each one.

Our final contribution is to identify which methodologies perform better with our particular dataset. From the simulation results, we determine in which scenario our data lies in, and follow the suggestions related to the performance of the estimators for this scenario. Therefore, we use a consistent method to choose an efficiency estimator, which provides a significant contribution to previous literature in local government efficiency. We use a sample of 1846 Spanish local governments of municipalities between 1000 and 50,000 inhabitants for the period 2008–2015. 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.

The sample is also relevant in terms of the period analysed. The economic and financial crisis that started in 2007 had a huge impact on most Spanish local government revenues and finances in general. In addition, the budget constraints became stricter with the law on budgetary stability,[6] which introduced greater control over public debt and public spending. Under these circumstances, issues related to Spanish local government efficiency have gained relevance and momentum. Evaluation techniques give the opportunity to identify policy programmes that work, to analyse aspects of a programme that can be improved, and to identify other public programmes that do not meet the stated objectives. In fact, gaining more insights into the amount of local government inefficiency might help to further support effective policy measures

---

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

[6] *Ley General Estabilidad Presupuestaria* (2007, 2012), or Law on Budgetary Stability.

to correct and/or control it. Therefore, it is obvious that obtaining here a reliable efficiency score would have relevant economic and political implications.[7]

Our results suggest that there is no one approach suitable for all efficiency analysis. When using these results for policy decisions, local regulators must be aware of which part of the distribution is of particular interest and if the interest lies in the efficiency scores or the rankings estimates. We find that for our sample of Spanish local governments, all methods showed some room for improvement in terms of possible cost efficiency gains, although they present large differences in inefficiency levels. Both DEA and FDH methodologies showed the most reliable efficiency results, according to the findings of our simulations. Therefore, our results indicate that the average cost efficiency would have been between 0.5417 and 0.7543 during the period 2008–2015, suggesting that Spanish local governments could have achieved the same level of local outputs with about 25% and 46% fewer resources. From a technical point of view, the analytical tools introduced in this study make an interesting contribution that examines the possibility of using a consistent method to choose an efficiency estimator, and the results provide evidence on how efficiency could certainly be assessed to provide some additional guidance for practitioners, scholars and policy makers.

The paper is organised as follows: Section 2 gives an overview of the methodologies applied to determine the cost efficiency. Section 4 describes the data used. Section 3 shows the methodological comparison experiment and the results for the different scenarios. Section 5 suggests which methodology performs better with our dataset and presents and comments on the most relevant efficiency results. Finally, Sect. 6 summarises the main conclusions.

2 Methodologies

Frontier or *best practice* methods aim to model the frontier of the technology, rather than modelling the average use of the technological possibilities. As indicated by Bogetoft and Otto (2010), this has certain advantages (from a practical point of view), since it is better to "learn from the best" than to "imitate mediocre performance". In this benchmarking literature, however, and as indicated in Introduction, there is no consensus as to the most appropriate method to measure efficiency.[8]

We distinguish between nonparametric (such as DEA) and parametric methods (such as SFA), and several subcategories are found within each family of methods. The main difference between the two measurement techniques is that whereas the former

---

[7] In this respect, Law 27/2013, of Rationalisation and Sustainability of the Local Administration (LRSAL, *Ley de Racionalización y Sostenibilidad de la Administración Local*) is the most significant reform since Law 7/1985, on Local Government (LBRL, *Ley Reguladora de las Bases del Régimen Local*). The three objectives of the Law are: (i) to guarantee the financial sustainability of all public administrations; (ii) to strengthen confidence in the stability of the Spanish economy; and (iii) to strengthen Spain's commitment to the European Union in terms of budget stability. Thus, the four principles of previous legislation are maintained: budgetary stability, multi-annuality, transparency and effectiveness, and efficiency in the allocation of public resources. This contributes to ultimately reinforcing some of the elements of the previous law, introducing three new principles: financial sustainability, responsibility, and institutional loyalty.

[8] Some excellent monographs in the field are those by Fried et al. (1993), Färe et al. (1994), Färe et al. (1985), Charnes et al. (1994) and, more recently, Daraio and Simar (2007a), among many others.

is quite flexible and not subject to the "parametric straitjacket", it cannot differentiate between noise and inefficiency. In contrast, parametric methods can make such a distinction, but must select functional forms—also for the distribution of efficiency. Therefore, selecting a particular methodology used to boil down to choosing between "the lesser of two evils" (Berger 1993). However, both parametric and nonparametric methods have evolved, and new proposals are now available (some of which have been reviewed in Fried et al. 2008).

In this section, we present our four different nonparametric techniques to measure cost efficiency,[9] namely, DEA, FDH, order-*m* and Kneip et al.'s (2008) bias-corrected DEA estimator, which we will refer to as KSW.

2.1 Data envelopment analysis (DEA) and free disposal hull (FDH)

Data envelopment analysis (DEA, Charnes et al. 1978; Banker et al. 1984) and free disposal hull (FDH, Deprins et al. 1984) are efficiency measurement techniques which share the underpinnings of adopting a frontier approach, in which efficient units (municipalities) lie on the empirical frontier, whereas the rest are defined as inefficient. Both DEA and FDH share their nonparametric nature, as well as the fact of being based on linear programming techniques. We consider an input-oriented DEA model because public sector outputs are established externally (the minimum services that local governments must provide), and it is therefore more appropriate to evaluate efficiency in terms of the minimisation of inputs (Balaguer-Coll and Prior 2009).[10]

The mathematical formulation for the cost efficiency measurement (Färe et al. 1994) corresponding to DEA aims to minimise costs, for given levels of outputs, by solving the following programme for each local government and each sample year[11]:

---

[9] Different types of efficiency can be distinguished, depending on the data available for inputs and outputs: *technical efficiency* (TE) requires data on quantities of inputs and outputs, while *allocative efficiency* (AE) requires additional information on input prices. When these two measures are combined, we obtain the *economic efficiency*, also called *cost efficiency* ($CE = TE \cdot AE$). In the public sector, there are often no prices for public goods due to the sector non-market nature (Kalb et al. 2012). In this paper, we measure local government cost efficiency since we have information relative to specific costs, although it is not possible to decompose it into physical inputs and input prices.

[10] Due to the multi-input and/or multi-output nature of local governments, considering frontier methods (particularly using nonparametric methods such as DEA) has been a very popular choice. However, as suggested by an anonymous referee, estimating the impact of a chapter of local expenditure on the production of good or service using a panel of municipalities that are homogeneous after controlling for observable and unobservable characteristics is an option that, although not usual in this literature, might also be pondered. Unfortunately, due to the way local governments report their accounting information, it is not possible to allocate the resources (as shares of the budget) to provide a particular service or infrastructure. Even if this information were currently available, it would only be possible to evaluate how efficiently the specific service analysed were provided.

[11] Alternatively to this input (or cost minimisation) orientation, an output orientation might also be adopted. However, in public sector studies, in general, and local government, in particular, it is generally adopted an input orientation for a variety of reasons such as their (in several cases) non-controllable nature.

$$\min_{(\theta_k, \lambda_1 \ldots \lambda_n)} \theta_k$$

$$\text{s.t.} \qquad \theta_k x_k - \sum_{i=1}^{n} \lambda_i x_i \geq 0$$

$$-y_{k,p} + \sum_{i=1}^{n} \lambda_i y_{i,p} \geq 0 \qquad\qquad (1)$$

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad\qquad p = 1, \ldots, P$$

$$\lambda_i \geq 0 \qquad\qquad i = 1, \ldots, n$$

where x_k and x_i represent the observed inputs (i.e., the total costs) corresponding to municipalities k and i, respectively. Similarly, $y_{k,p}$ and $y_{i,p}$ denote the observed outputs for unit k and i with respect to output p; λ_i are the relative weights which describe the relative importance of the unit considered to determine the virtual reference used as a comparison in order to evaluate unit k; n are the total number of observations; and θ_k represents the cost efficiency coefficient for each k municipality. The 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 a similar size.

The free disposal hull (FDH) estimator proposed by Deprins et al. (1984) is an extension of DEA whose main difference is that the former drops the convexity assumption. Therefore, FDH cost efficiency is defined as follows:

$$\min_{(\theta_k, \lambda_1 \ldots \lambda_n)} \theta_k$$

$$\text{s.t.} \qquad \theta_k x_k - \sum_{i=1}^{n} \lambda_i x_i \geq 0$$

$$-y_{k,p} + \sum_{i=1}^{n} \lambda_i y_{i,p} \geq 0 \qquad\qquad (2)$$

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad\qquad p = 1, \ldots, P$$

$$\lambda_i \in \{0, 1\} \qquad\qquad i = 1, \ldots, n$$

Finally, the solution to the mathematical linear programming problems (1) and (2) yields optimal values for θ_k. 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, it is well known that they present several drawbacks, such as the influence of extreme values and outliers, the "curse of dimensionality"[12] or the difficulty of drawing classical statistical inference. Hence, we also consider two alternatives to DEA and FDH estimators that are able to overcome most of these drawbacks. The first is order-m (Cazals et al. 2002), a partial frontier approach that mitigates the influence of outliers and the curse of dimensionality, and the second is

---

[12] An increase in the number of inputs or outputs, or a decrease in the number of units for comparison, implies higher efficiencies Daraio and Simar (2007a).

Kneip et al.'s (2008) bias-corrected DEA estimator (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 that involves the concept of partial frontier. The order-*m* estimator, for finite *m* units, does not envelope all data points and is consequently less extreme. In the input orientation case, this method uses as a benchmark the expected minimum level of input achieved among a fixed number of *m* local governments producing at least output level *y* (Daraio and Simar 2007a). The value *m* represents the number of potential units against which we benchmark the analysed unit. Following (Daraio and Simar 2007a, p. 72), the order-m input efficiency scores are estimated as follows:

1. First, for a given output level *y*, draw a sample *m* with replacement among those $y_{i,m}$, such that $y_{i,m} \geq y$.
2. The FDH estimator is applied by using the sub-sample drawn in step 1, estimating the efficiency coefficient $\hat{\theta}_i$.
3. Steps 1 and 2 are repeated *B* times, obtaining *B* efficiency coefficients, $\hat{\theta}_i^b$ ($b = 1, \ldots, B$).
4. Finally, we obtain $\hat{\theta}_{i,m}$ which is the mean of the estimated *B* efficiency coefficients: $\theta_{i,m} = \frac{1}{B} \sum \hat{\theta}_i^b$

If *m* goes to infinity, the order-*m* estimator converges to FDH. The most reasonable value of *m* is determined as the value for which the super-efficient observations becomes constant (Daraio and Simar 2005). Note that order-*m* scores are not bounded by 1 as DEA or FDH. 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 2007a).

2.2.2 Kneip et al.'s (2008) bias-corrected DEA estimator (KSW)

The KSW (Kneip et al. 2008) is a bias-corrected DEA estimator which derives the asymptotic distribution of DEA via bootstrapping techniques. Simar and Wilson (2008) noted that DEA and FDH estimators are biased by construction, implying that the true frontier would lie under the DEA estimated frontier. Badunenko et al. (2012) explained that, 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". This procedure provides consistent statistical inference of efficiency estimates (i.e., bias and confidence intervals for the estimated efficiency scores).

In order to implement the bootstrap procedure (based on sub-sampling), first let $s = n^d$ for some $d \in (0, 1)$, where *n* is the sample size and *s* is the sub-sample size. Then, the bootstrap is outlined as follows:

1. First, a bootstrap sub-sample $S_s^* = (X_i^*, Y_i^*)_{i=1}^s$ is generated by randomly drawing (independently, uniformly, and with replacement) s observations from the original sample, S_n.
2. Apply the DEA estimator, 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, $\hat{\theta}_b^*(b = 1, \ldots, B)$, 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 us to approximate the unknown distribution of $n^{2/(p+q+1)}(\frac{\theta^*(x,y)}{\theta(x,y)} - 1)$. The values p and q are the output and input, respectively. Moreover, θ^* and θ represents the true score coming from a known and an unknown data-generating process. The bias-corrected DEA efficiency score, which is adjusted by the s sub-sample size, is given by:

$$\theta_{bc} = \theta^* - Bias^* \tag{3}$$

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

$$Bias^* = \left(\frac{s}{n}\right)^{2/(p+q+1)} \left[\frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^* - \theta^*\right] \tag{4}$$

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 $\theta(x, y)$

$$\left[\frac{\hat{\theta}(x, y)}{1 + n^{-2/(p+q+1)}\delta_{1-\alpha/2,s}}, \frac{\hat{\theta}(x, y)}{1 + n^{-2/(p+q+1)}\delta_{\alpha/2,s}}\right] \tag{5}$$

3 Methodological comparison

In contrast to the previous literature, in this section, we compare DEA, FDH, order-m, and KSW approaches following the method proposed by Badunenko et al. (2012).[13] Our aim is to uncover which measures perform best with our particular dataset, that is, which ones are the most appropriate to measure local government efficiency in Spain in order to provide useful information for local governments' performance decisions.

To this end, we carry out the experiment via Monte Carlo simulations. We first define the data-generating process, the parameters and the distributional assumptions on data, adapted to the local government framework in order to make our simulation more realistic. Second, we consider the different methodologies and take several standard measures to compare their behaviour. Next, after running the simulations, we

---

[13] Badunenko et al. (2012) compared parametric methodologies, represented by the nonparametric kernel SFA estimator of Fan et al. (1996), with nonparametric, represented by the bias-corrected DEA estimator of Kneip et al. (2008). They assess the performance of these estimators via Monte Carlo simulations and discuss which estimator should be employed in various scenarios. Finally, they consider how these estimators work in practice by determining which scenario corresponds to three different data sets.

discuss the relative performance of the efficiency estimators under the various scenarios. Finally, we decide which methods are the most appropriate to measure local government efficiency in Spain.

3.1 Simulations

Several previous studies analysing local government cost efficiency with parametric techniques used the SFA estimator developed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977) as a model to estimate cost frontiers.[14] These studies 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 researchers 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).[15] We consider a Cobb-Douglas cost function (CD). For the baseline case, we assume constant returns to scale (CRS) ($\gamma = 1$).[16] We establish $\alpha = 1/3$ and $\beta = \gamma - \alpha$.[17]

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 (υ) is normally distributed $N(0, \sigma_\upsilon^2)$, and the true cost efficiency is $TCE = exp(-u)$, where u is half-normally distributed $N^+(0, \sigma_u^2)$ and independent from υ. We introduce the true error and inefficiency terms in the frontier formulation, which takes the following expression:

$$c = y_1^\alpha \cdot y_2^\beta \cdot exp(\upsilon + u), \qquad (6)$$

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

We simulate six different combinations for the error and inefficiency terms, in order to model various real scenarios. Table 1 contains the matrix of the different scenarios. It shows the combinations when σ_υ 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 (σ_υ), while the columns represent the variation of the inefficiency term (σ_u). The first

---

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

[15] As we will see in Sect. 4, local governments are considered as multiproduct organisations in which the joint use of their resources (input or costs) give rise to several services and facilities (outputs). In the experiment, for simplicity, we use a multi-output model with two outputs.

[16] In Sect. 3.4, we consider robustness checks with increasing and decreasing returns to scale to make sure that our simulations accurately represent the performance of our methods.

[17] We use $\alpha = 1/3$ given that it is a common value for calibration of the Cobb-Douglas function in the related literature (see for instance Badunenko et al. 2012). Robustness checks concerning different values of α show similar qualitative results to the baseline experiment and are available upon request.

Table 1 Combinations of error and inefficiency terms in Monte Carlo simulations to model scenarios

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

We simulate six different combinations for the error (σ_υ) 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. The Λ parameter is the ratio between σ_u and σ_υ, which sets each scenario

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 variation in inefficiency is relatively larger. The Λ parameter, which sets each scenario, is the ratio between of σ_u and σ_υ.

Within this context, scenario 1 is the case when the error and the inefficiency terms are relatively small ($\sigma_u = 0.01$, $\sigma_\upsilon = 0.01$, $\Lambda = 1.0$), 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 ($\sigma_u = 0.1$, $\sigma_\upsilon = 0.05$, $\Lambda = 2.0$), which means that the data is relatively noisy and the units are relatively inefficient.

For all simulations we consider 2000 Monte Carlo trials, and we analyse two different sample sizes, $n = 100$ and 200.[18,19] We note that nonparametric estimators do not take into account the presence of noise; however, we want to check how it affects the performance of our estimators since all data tend to have noise.[20]

3.2 Measures to compare the estimators' performance

In order to compare the relative performance of our four nonparametric methodologies, we consider the following median measures over the 2000 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

---

[18] Krüger (2012) notes that the low number of replications, ranging from 5 to 100, is a weakness in most previous studies on Monte Carlo investigation of efficiency measurement methods.

[19] To ease the computational process, we use samples of $n = 100$ and 200 to conduct simulations. In Sect. 3.4, we consider a robustness check with a bigger sample size ($n = 500$) to ensure that our simulations accurately represent the performance of our data.

[20] In Sect. 3.4, we consider a robustness check with no noise to ensure that our simulations accurately represent the performance of our data.

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 efficiencies. 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, where n_c and n_d are the number of concordant and discordant pairs in the data set, respectively. This test identifies the differences in the ranking distributions of the true and the estimated ranks.

We also compare the densities of cost efficiency across all Monte Carlo simulations (i.e., densities for the TCE and the estimated efficiencies of the four estimators) in order to report a more comprehensive description of the results, not only restrict them to a single summary statistic—the median. We use violin plots to compare the true distribution for different percentiles of our sample.[21] For each draw, we sort the data by the relative value of true efficiency and we plot densities at the 5%, 50% and 95% percentiles. Accordingly, we can analyse the performance of each estimator for a specific part of our sample. So, for example, if we were interested in estimating the poorer performers, we would focus on which estimator perform best only at the 5% percentile the efficiency distribution.

3.3 Relative performance of the estimators

Table 2 provides baseline results for the performance measures of the cost efficiency with the CD cost function. First we observe that the median bias of the cost efficiency scores is negative in DEA and KSW in all cases. This implies that the 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, implying a tendency to overestimate the true efficiency. Bias for all methodologies tends 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. The RMSE is smaller when σ_v is small, except for FDH in scenario 5 and order-m in scenarios 3 and 5. Moreover, the RMSE of the cost efficiency estimates increases 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. This 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 indicate underestimation (overestimation) of cost efficiencies. In this setting, DEA and KSW systematically underestimate the true efficiency. Moreover, as the sample size increases, so does the percentage of underestimated results. In contrast, FDH and order-m tend to overestimate the true efficiency, but as the sample size increases overestimated results decrease. Finally, we analyse Kendall's τ for the efficiency ranks between true and estimated efficiency

---

[21] The violin plot combines the density trace (or smoothed histogram) and the box plot (initially conceived by Tukey) into a single figure that reveals the structure found within the data. The name *violin plot* originated due to the early studies using these procedures, which resulted in graphics with the appearance of a violin. For details, see Hintze and Nelson (1998).

Table 2 Baseline results with Cobb-Douglas cost function

	Bias[a]				RMSE[b]				Upward bias[c]				Kendall's τ[d]			
	DEA	FDH	Order-m	KSW	DEA	FDH	Order-m	KSW	DEA	FDH	Order-m	KSW	DEA	FDH	Order-m	KSW
s1: $\sigma_\upsilon = 0.01$, $\sigma_u = 0.01$																
$n = 100$	−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: $\sigma_\upsilon = 0.05$, $\sigma_u = 0.01$																
$n = 100$	−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: $\sigma_\upsilon = 0.01$, $\sigma_u = 0.05$																
$n = 100$	−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: $\sigma_\upsilon = 0.05$, $\sigma_u = 0.05$																
$n = 100$	−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: $\sigma_\upsilon = 0.01$, $\sigma_u = 0.1$																
$n = 100$	−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

Table 2 continued

	Bias[a]				RMSE[b]				Upward bias[c]				Kendall's τ[d]			
	DEA	FDH	Order-m	KSW	DEA	FDH	Order-m	KSW	DEA	FDH	Order-m	KSW	DEA	FDH	Order-m	KSW
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

This table provides the baseline results for the performance of the methodologies in the Monte Carlo experiment. We simulate six scenarios, which represent different combinations for the error (σ_v) and inefficiency (σ_u) terms

[a] 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

[b] The $RMSE$ (root-mean-squared error) measures the standard deviation or error from the true efficiency

[c] The upward bias is the proportion of estimated efficiencies larger than the true efficiencies (returning a value of 1). The desired value is 0.5. The values less (greater) than 0.5 indicate underestimation (overestimation) of cost efficiencies

[d] Kendall's τ shows the correlation coefficient for the efficiency ranks between true and estimated efficiency scores

scores. In each scenario and sample size, DEA and KSW have a larger Kendall's τ; they therefore perform best at identifying the ranks of the efficiency scores.

We also analyse other percentiles of the efficiency distribution, since it is difficult to conclude from the table which methods perform better. Figures 1, 2 and 3 show results for the 5th, 50th and 95th percentiles of true and estimated cost efficiencies. We compare the distribution of each method with the TCE.[22] For visual simplicity, we show only the case when $n = 100$. Figures with sample size $n = 200$ do not vary greatly and are available upon request.

The figures show that results depend on the value of the Λ parameter. As expected, when the variance of the error term increases our results are less accurate. Note that nonparametric methodologies assume the absence of noise. In contrast, when the variance of the inefficiency term relative to the variance of the error term increases, our results are more precise.

Under **scenario 1** (see Fig. 1a, c, e), 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 5th and 50th percentiles. Moreover, we observe that FDH performs well in estimating the efficiency units in the 95th percentile.

Although **scenario 4** (see Fig. 2b, d, f) 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. On the other hand, we see from the 5th percentile that both FDH and order-m tend to overestimate efficiency. However, at the 50th and 95th percentiles, both methods perform better at estimating the efficiency units since their median values and density modes are closer to the TCE distribution.

Similarly, in **scenario 2** (see Fig. 1b, d, f), when the error term is relatively large but 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 the poorest results as the dispersion of TCE is much more squeezed than the estimators' distributions. Therefore, when Λ is small, all four methodologies perform less well in predicting efficiency scores.

Scenario 3 (see Fig. 2a, c, e), the error term is relatively small but the inefficiency term is relatively large. Because the Λ value has increased, all methodologies do better at predicting the efficiency scores. At the 5th and 50th 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 95th percentile FDH does better at estimating the efficient units, while DEA and KSW slightly underestimate efficiency and order-m slightly overestimates it.

In **scenario 5** (see Fig. 2a, c, e), the error variation is relatively small, but the inefficiency variation is very large. This scenario shows the most favourable results
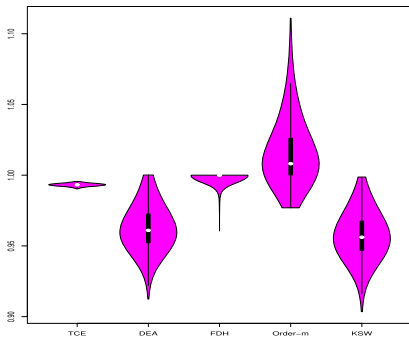
---

[22] We consider that a particular methodology has a better or worse performance depending on the similarities found between its efficiency distribution and the true efficiency distribution.
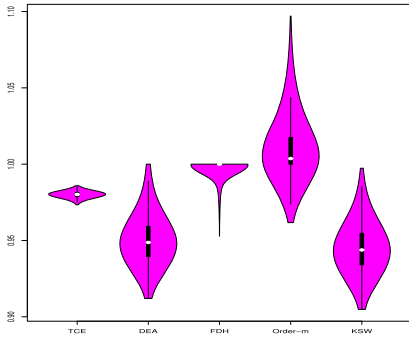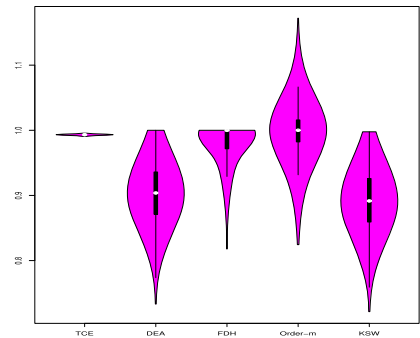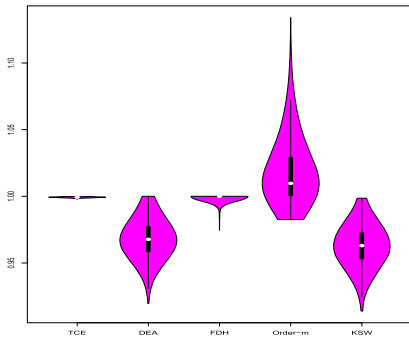
(a) S1, 5th percentile, n=100
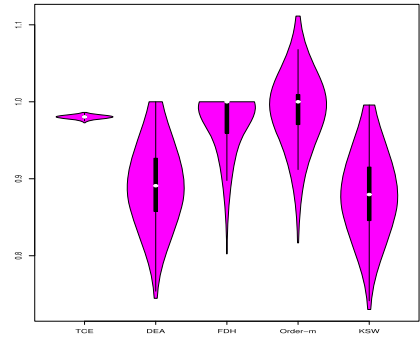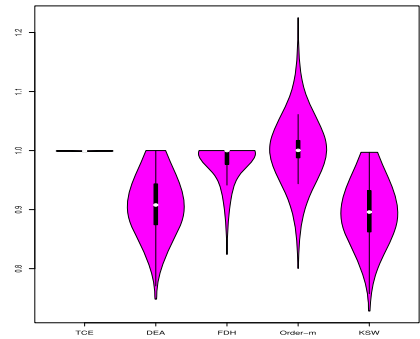
(b) S2, 5th percentile, n=100

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

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

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

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

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

because the TCE distribution is highly dispersed and therefore better represents the estimators' performance. At the 5th and 50th percentiles, DEA and KSW densities are very close to the true distribution of efficiency, while FDH and order-*m* overestimate it. In contrast, at the 95th percentile FDH seems to be closer to the TCE although it slightly overestimates it.

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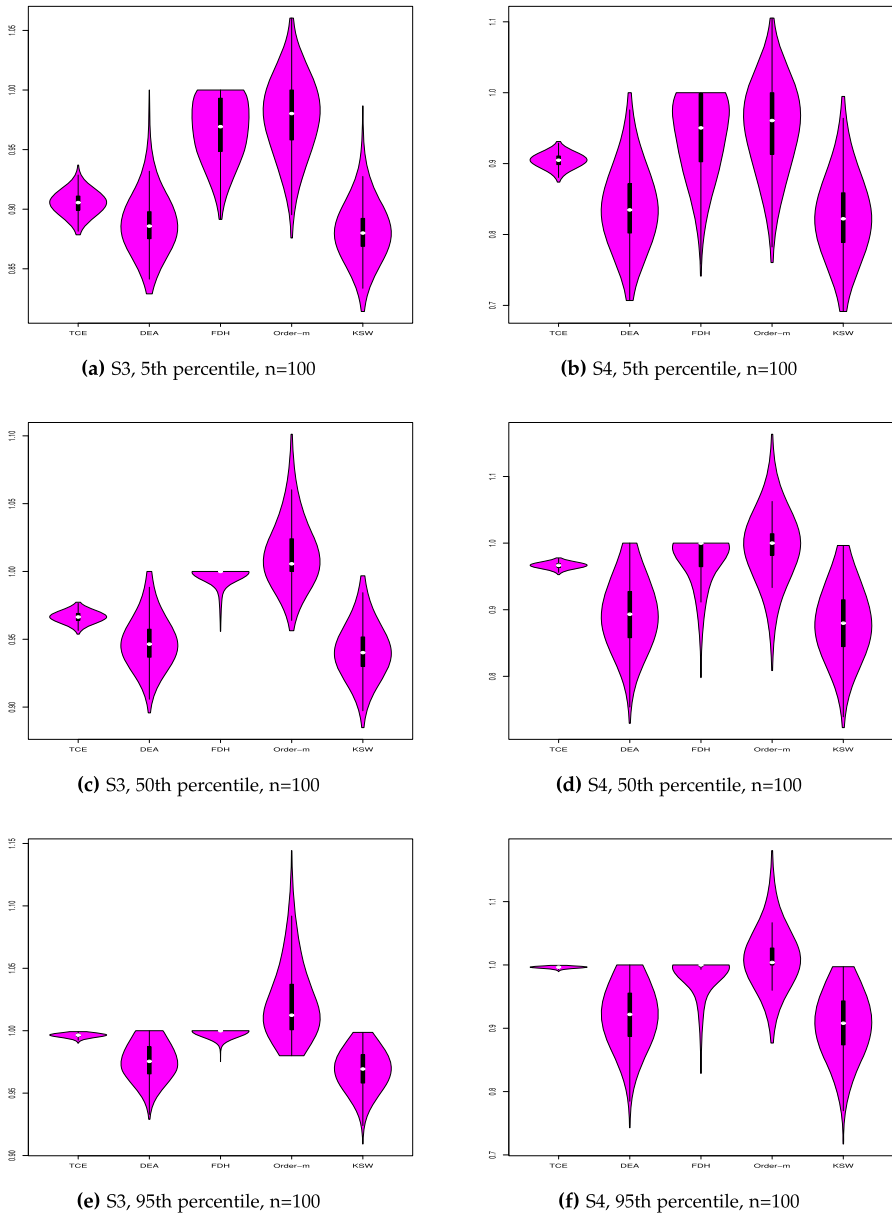
(f) S4, 95th percentile, n=100

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

Finally, in **scenario 6** (see Fig. 3b, d, f) 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 scenarios 2 and 4), all the estimators perform better. At the 5th and 50th percentiles, DEA and KSW slightly underestimate efficiency and FDH and order-*m* slightly overestimate it (in terms of median values and density

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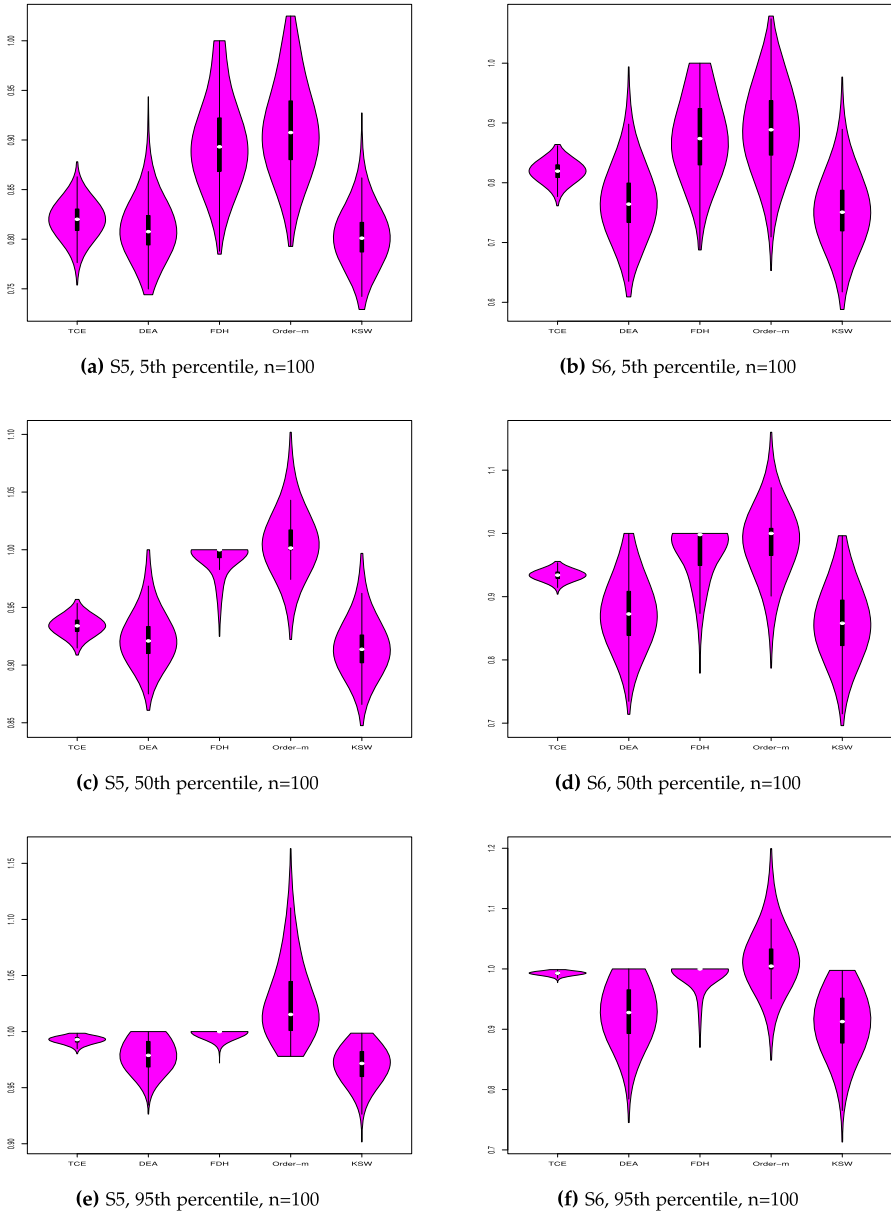
(f) S6, 95th percentile, n=100

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

mode). However, despite all methods being quite close to the TCE distribution, DEA underestimates less than KSW, and FDH overestimates less than order-*m*. Finally, at the 95th 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.

To sum up, in this subsection we have provided the baseline results for the relative performance of our four nonparametric methodologies. We have considered four median measures as well as other percentiles of the efficiency distribution. We found that the performance of the estimators vary greatly according to each particular scenario. However, we observe that both DEA and KSW consistently underestimate efficiency in nearly all cases, while FDH and order-*m* tend to overestimate it. Moreover, we note that DEA and KSW perform best at identifying the ranks of the efficiency scores. In Sect. 3.5, we will explain in greater detail which estimator to use in the various scenarios.

3.4 Robustness checks

We consider a number of robustness checks to verify that our baseline experiment represents the performance of our estimators. Results for each robustness test are given in the extra Annex.

- 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* are slightly worse than the baseline experiment. All methods perform better at estimating the true ranks, except order-*m* in scenario 1. In short, we find that when noise is absent, DEA and KSW have a greater performance.
- Greater variation in the inefficiency term: In the baseline experiment we set different values for the inefficiency term ($\sigma_u = 0.01, 0.05$ and 0.1). We also consider the case in which bigger efficiency shocks exist ($\sigma_u = 0.2$ and $\sigma_u = 0.3$). As expected, we observe some small improvement in the performance of the median measures for DEA and KSW, while the performance for FDH and order-*m* slightly decrease. All methods perform better when estimating the true efficiency ranks. In general, despite the small quantitative variations, the results of the baseline experiment seem to hold.
- 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, $n = 500$. There is a slight deterioration in the performance of DEA and KSW, while FDH and order-*m* vary depending on the scenario. However, the results only differed slightly. We find no 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 a slight deterioration in the performance of DEA and KSW estimators. Performance for order-*m* improves with decreasing returns to scale and deteriorates 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*: Following Daraio and Simar's (2007a) suggestion, in order to choose the most reasonable value of *m* we considered different *m* sizes

($m = 20,\ 30$ and 40). In our application, the baseline experiment sets $m = 30$. In general, compared with the other m values, there are some quantitative changes (i.e., performance with $m = 20$ worsens, while with $m = 40$ it improves slightly); however, the qualitative results from the baseline case seem to hold.

In sort, we find that after considering several robustness checks, we do not see any major differences from the baseline experiment. Therefore, despite the initial assumptions done, our simulations accurately depict the performance of our estimators.

3.5 Which estimator in each scenario

Based on the above comparative analysis of the four methodologies' performance, inspired by our results as well as Badunenko et al.'s (2012) proposal, we summarise which ones should be used in the various scenarios, assuming that the simulations remain true for different data-generating processes. Table 3 suggests which estimators to use for each scenario when taking into account the efficiency scores. The first row in each scenario shows the relative magnitudes of the estimators compared with the True Cost Efficiency (TCE), while the rest of the rows suggest which estimators to use for each percentile (5th, 50th or 95th). In some cases, the methodologies vary little in terms of identifying the efficiency scores.

Badunenko et al. (2012) conclude that if the Λ value is small, as in scenario 2 ($\Lambda = 0.2$), the efficiency scores and ranks will be poorly estimated.[23] This scenario yields the worst results, since the estimators are far from the "truth". Although Table 3 suggests scenario 2, we do not recommend efficiency analysis for this particular scenario, since it would be inaccurate.

Although scenarios 1 and 4 present better results than scenario 2 (when $\Lambda = 1$), estimators also perform poorly at predicting the true efficiency scores. In scenario 1, FDH seems to be the best method to estimate efficiency in all percentiles; however, DEA should also be considered at the 5th percentile (the TCE remains between DEA and FDH at this percentile). Similarly, in scenario 4 FDH predominates at the 5th percentile, although DEA should also be considered. On the other hand, both FDH and order-m perform better at the 50th and 95th percentiles. For efficiency rankings, DEA and KSW methodologies show a fairly good performance when ranking the observations in both scenarios.

Similarly, scenario 6 performs better than scenarios 1 and 4, since the variation of the inefficiency term increases and, as a consequence, the value of Λ also increases ($\Lambda = 2$). In this scenario the best methodologies for estimating the true efficiency scores seem to be DEA and FDH at the 5th and 50th percentiles, and FDH (followed by order-m) at the 95th percentile. In contrast, DEA and KSW methodologies are better at ranking the observations.

In scenario 3, the Λ value increases again ($\Lambda = 5$), and all the methodologies predict the efficiency scores more accurately. For the 5th and 50th percentiles, the closest estimator to the true efficiency seems to be DEA (followed by KSW). At the 95th percentile FDH is the best method. For the rankings, however, DEA and KSW provide more accurate estimations of the efficiency rankings.

---

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

Table 3 Relative performance of the efficiency scores at the 5th, 50th and 95th percentiles

	$\sigma_u = 0.01$	$\sigma_u = 0.05$	$\sigma_u = 0.1$
$\sigma_v = 0.01$	**Scenario 1**	**Scenario 3**	**Scenario 5**
	KSW<DEA<TCE<FDH<order-m	KSW<DEA<TCE<FDH<order-m	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
$\sigma_v = 0.05$	**Scenario 2**	**Scenario 4**	**Scenario 6**
	KSW<DEA<FDH<TCE<order-m	KSW<DEA<TCE<FDH<order-m	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

This table summarises which estimators to use for each scenario. Each scenario represent different combinations for the error (σ_v) and inefficiency (σ_u) terms. The first row of each scenario represents the relative magnitudes of the estimators compared with the True Cost Efficiency (TCE), while the rest of the rows suggest which estimators to use for each percentile (5th, 50th or 95th)

Finally, scenario 5 has the largest Λ value ($\Lambda = 10$). Here, the estimators perform best at estimating efficiency and ranks. DEA (followed by KSW) performs better at the 5th and 50th percentiles and FDH at the 95th percentile. DEA and KSW excel at estimating the efficiency rankings.

4 Sample, data, and variables

We consider a sample of Spanish local governments of municipalities between 1000 and 50,000 inhabitants for the 2008–2015 period.[24, 25] The information on inputs and outputs was obtained from the Spanish Ministry of the Treasury and Public Administrations (*Ministerio de Hacienda y Administraciones Públicas*). Specific data on 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 1846 Spanish municipalities for every year (representing 22.74%), after removing all the observations for which information on inputs or outputs was not available for some of the years of the sample period (2008–2015). Specifically, there was no information for the Basque Country, Navarre,[26] the regions of Catalonia and Madrid, nor for the provinces of Burgos and Huesca.[27]

Inputs are representative of the cost of the municipal services provided. Using budget expenditures as inputs is consistent with previous literature (e.g., Balaguer-Coll et al. 2007, 2010; Zafra-Gómez and Muñiz-Pérez 2010; Fogarty and Mugera 2013; Ferreira Da Cruz and Cunha Marques 2014; Narbón-Perpiñá et al. 2019) since data on the costs incurred by each local government in the provision of each municipal service and facility (i.e., in physical units and their corresponding input prices) is not available. Accordingly, we construct an input measure, representing total local government costs (X_1), that includes various municipal expenditures taken from the implemented (or executed) municipal budgets: 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 article 26 of the Spanish law which regulates the local system (*Ley reguladora de Bases de Régimen Local*). It establishes the minimum services and facilities that each municipality is legally obliged to provide, depending on their size. Specifically, all governments must provide public street lighting, cemeteries,

---

[24] The restriction of the sample to municipalities between 1000 and 50,000 inhabitants is due to the limited availability of financial data for municipalities with population below 1000 and the lack of data on local services and facilities coming from the survey on local infrastructures and facilities for municipalities over 50,000 inhabitants.

[25] Spanish local governments are characterised by their very diverse populations and territorial distributions. For instance, in 2013 almost 60.35% of municipalities had populations below 1000, and accounted for only 3.14% of the total population.

[26] 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 their own autonomous system, and consequently, they are not included in the State Economic Cooperation.

[27] Data missing from the survey on local infrastructures and facilities: Madrid (2008–2015), Burgos (2009), Huesca (2011–2015) and Catalonia (2012–2015).

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 in contrast to previous studies in other European countries, we do not include outputs such as the provision of primary and secondary education, care for the elderly or health services, since they do not fall within the responsibilities of Spanish municipalities.[28]

As a result, we chose six output variables to measure the services and facilities municipalities provide. Due to the difficulties in measuring public sector outputs, in some cases, it is necessary to use proxy variables for the services delivered by municipalities given the unavailability of more direct outputs (De Borger and Kerstens 1996a, b), an assumption which has been widely applied in the literature. Table 4 reports the minimum services that all local government were obliged to provide for the 2008–2015 period, as well as the output indicators used to evaluate the services. Table 5 reports descriptive statistics for inputs and outputs for the same period.[29]

The period under analysis was turbulent and the effects of the crisis might have affected municipalities differently. This is a very interesting issue but, in our opinion, lies beyond the aims of our paper. Specifically, the level of involvement of Spanish municipalities in feeding the housing bubble varied markedly across them—it was high, on average, but not generalised. It would therefore be worth investigating the links between efficiency and urban development during these years, including also data for presence of bank branches and their geographical distribution. This would also be related with second-stage issues, i.e., with analysing the determinants of municipal efficiency, a topic that we do not explicitly address in our study but on which at least two survey studies exist (Narbón-Perpiñá and De Witte 2018b; Aiello and Bonanno 2019), and very recently.[30]

---

[28] Defining and measuring the services and infrastructures is particularly challenging for additional reasons. Among them, we find that some municipalities might go beyond the legal minimum and, in addition, to provide services and facilities for which information is not always available. These are the so-called "gastos impropios", and some (few) reports have made attempts to measure them. This is the case of Vilalta and Mas (2006), who estimated the spending of municipalities on services and infrastructures they were not bounded to provide for the municipalities in the province of Barcelona. With this motivation, Balaguer-Coll et al. (2007) proposed a methodology whose aims was to compare municipalities only with those facing similar environmental conditions and choosing similar output mixes. For other contexts different to the Spanish one, see, among others, Bennett and DiLorenzo (1982), Marlow and Joulfaian (1989) and Merrifield (1994).

[29] The literature on efficiency analysis has considered, from its very beginnings, how environmental variables might effect efficiency scores. One of the seminal contributions was that by Banker and Morey (1986), but more recently conditional efficiency models such as Daraio and Simar (2007b) have also acknowledge this reality. In the case of the Spanish context, Balaguer-Coll et al. (2013) have proposed a methodology to address this issue, and in a geographically close context (Portugal), Cordero et al. (2017) have followed a conditional efficiency approach to deal with similar issues. The survey study by Narbón-Perpiñá and De Witte (2018b) also reviews the contributions dealing explicitly with the determinants of municipal efficiency (focusing not only on the Spanish case). In the case we are dealing with here, we consider that combining the aims of these studies with ours would strain the space limits to unreasonable levels.

[30] Other shocks such as the *Plan E*, although interesting to model, would not only require a specific investigation but also information difficult to obtain, in addition to several modifications to our approach (probably having to consider a conditional efficiency model). Some contributions (Bellod Redondo 2015)

Table 4 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 of water distribution networks (m)
Sewage system	Length of sewer 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 5 Descriptive statistics for data in inputs and outputs, average values for the period 2008–2015

	Median	S.d.
INPUTS[a]		
Total costs (X_1)	6,554,849.98	7,833,578.91
OUTPUTS		
Total population (Y_1)	7017.36	8141.78
Street infrastructure surface area[b](Y_2)	331,864.63	320,926.78
Number of lighting points (Y_3)	1471.80	1432.94
Tons of waste collected (Y_4)	3726.89	10,994.20
Length of water distribution networks[b](Y_5)	47,395.48	71,743.54
Length of sewer networks[b](Y_6)	29,383.79	31,287.88

[a] In thousands of euros
[b] In square metres

5 Which estimator performs better with Spanish local governments'

Finally, in this section we identify the most appropriate methodologies to measure local government efficiency in Spain. First, we estimate Λ values for our particular dataset via Fan et al.'s (1996) nonparametric kernel estimator, hereafter FLW.[31] The estimated Λ value helps to determine in which scenario our data lies (see Table 1). Second, we refer to Table 3, check the recommendations for our scenario, and choose the appropriate estimators for our particular needs.

Table 6 reports results of the Λ parameters for our sample of 1846 Spanish local governments for municipalities between 1000 and 50,000 inhabitants for the 2008–

---

that analysed aspects of the *Plan* concluded that although its size was not negligible, its total magnitude was very difficult to estimate (ranging between 1.43% and 3.1% of total GDP), as well as its overall effect. In addition, it is also complex to calculate the exact quantities that each municipality spent that corresponded to the *Plan E.*

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

Table 6 Estimates to determine the scenario for Spanish local governments dataset

	2008	2009	2010	2011	2012	2013	2014	2015
Λ	2.1535	2.2640	1.8618	1.7480	2.0107	1.9501	1.7736	1.9797
R^2	0.7847	0.8207	0.8132	0.8103	0.8093	0.8236	0.8261	0.8226

This table contains the results of the Λ parameters of our sample of 1846 Spanish local governments for the period 2008–2015. Λ values help to determine in which scenario our data lies. We also report the goodness-of-fit measure (R^2) of our empirical data

Table 7 Summary statistics for efficiency results in Spanish local governments

	Mean	Median	Min	Max	Q1	Q3
DEA						
2008	0.4785	0.4537	0.0440	1.0000	0.3492	0.5828
2009	0.5740	0.5625	0.0981	1.0000	0.4588	0.6718
2010	0.5220	0.5033	0.1141	1.0000	0.4114	0.6028
2011	0.5232	0.5019	0.1329	1.0000	0.4093	0.6095
2012	0.5140	0.4981	0.1036	1.0000	0.3953	0.6090
2013	0.5672	0.5584	0.1463	1.0000	0.4594	0.6628
2014	0.5887	0.5709	0.1356	1.0000	0.4698	0.6837
2015	0.5661	0.5471	0.1093	1.0000	0.4484	0.6610
2008–2015	0.5417	0.5245	0.1105	1.0000	0.4252	0.6354
FDH						
2008	0.7254	0.7362	0.0785	1.0000	0.5431	0.9857
2009	0.7970	0.8261	0.1195	1.0000	0.6509	1.0000
2010	0.7628	0.7674	0.1833	1.0000	0.6060	1.0000
2011	0.7408	0.7457	0.1792	1.0000	0.5757	0.9841
2012	0.7406	0.7476	0.1190	1.0000	0.5789	0.9732
2013	0.7563	0.7613	0.1946	1.0000	0.5958	0.9866
2014	0.7523	0.7489	0.1443	1.0000	0.6022	0.9734
2015	0.7593	0.7597	0.1447	1.0000	0.6028	0.9850
2008–2015	0.7543	0.7616	0.1454	1.0000	0.5944	0.9860
Order-m						
2008	0.7847	0.7919	0.0817	1.9386	0.6039	1.0000
2009	0.8472	0.8659	0.1358	1.6568	0.7043	1.0000
2010	0.8187	0.8315	0.1926	1.5159	0.6630	1.0000
2011	0.7975	0.7984	0.1938	1.6070	0.6339	1.0000
2012	0.7958	0.7999	0.1451	1.8822	0.6364	1.0000
2013	0.8068	0.8055	0.2053	1.8952	0.6497	1.0000
2014	0.8086	0.8100	0.1834	2.0638	0.6589	1.0000
2015	0.8151	0.8159	0.1550	1.6661	0.6643	1.0000
2008–2015	0.8093	0.8149	0.1616	1.7782	0.6518	1.0000

Table 7 continued

	Mean	Median	Min	Max	Q1	Q3
KSW						
2008	0.4274	0.4067	0.0397	1.0000	0.3048	0.5229
2009	0.5344	0.5250	0.0887	1.0000	0.4214	0.6337
2010	0.4703	0.4516	0.0968	1.0000	0.3719	0.5444
2011	0.4713	0.4526	0.1222	1.0000	0.3694	0.5521
2012	0.4600	0.4469	0.0786	1.0000	0.3517	0.5447
2013	0.5243	0.5160	0.1182	1.0000	0.4276	0.6140
2014	0.5495	0.5377	0.1265	1.0000	0.4438	0.6382
2015	0.5213	0.5074	0.1060	1.0000	0.4138	0.6175
2008–2015	0.4948	0.4805	0.0971	1.0000	0.3881	0.5835

Fig. 4 Violin plots of cost efficiency estimates in Spanish local governments, average values for the period 2008–2015
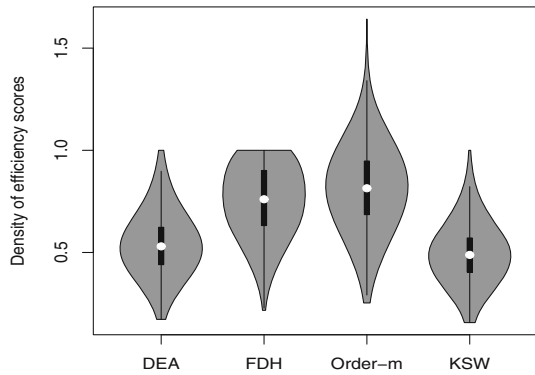
2015 period. The results of the Λ estimates range from 1.74 to 2.26, which are closer to 2 and correspond to scenario 6. Moreover, the goodness-of-fit measure (R^2) of our empirical data lies at around 0.8. The summary statistics for the overall cost-efficiency results averaged over all municipalities for each year are reported in Table 7. Figure 4 shows the violin plots of the estimated cost efficiencies for further interpretation of results.[32]

In scenario 6, the DEA and FDH methods performed better than the others at the 5th and 50th percentiles of the distribution (the former slightly underestimates efficiency, while the latter slightly overestimates it), and FDH (followed by order-m) performed better at the 95th percentile. Therefore, the true efficiency would lie between the results of DEA and FDH both at the median and the lower percentiles, while FDH perform best at estimating the benchmark units. When using these results for performance decisions, local managers must be aware of which part of the observations are of particular interest and whether interest lies in the efficiency score or the ranking. In this context, DEA results indicate that the average cost efficiency during the period

---

[32] For visual simplicity, we plot together years 2008–2015, however they do not differ greatly and individual plots are available upon request.

Table 8 Rank correlation Kendall coefficients between the average cost efficiency estimates of the four methodologies for the period 2008–2015

	DEA	FDH	Order-*m*	KSW
DEA	1.0000	0.6928	0.6853	0.8998
FDH	0.6928	1.0000	0.8379	0.6484
Order-*m*	0.6853	0.8379	1.0000	0.6342
KSW	0.8998	0.6484	0.6342	1.0000

2008–2015 at the central part of the distribution is 0.5417, while the average in FDH is 0.7543, so we expect the true cost efficiency scores to lie between 0.5417 and 0.7543. Moreover, average scores at the lowest quartile (Q1) are 0.4252 in DEA and 0.5944 in FDH, so we expect the true efficiency scores at the lower end of the distribution to lie between 0.4252 and 0.5944. Similarly, the average FDH scores at the upper quartile (Q3) are 0.9860, so we expect these estimated efficiencies will be similar to the true ones.

The efficiency scores shown by KSW are smaller than those reported by DEA and FDH (the average efficiency scores in KSW for the period 2008–2015 are 0.3881 for the lowest quantile (Q1), 0.4948 for the mean and 0.5835 for the upper quartile (Q3)). Based on our Monte Carlo simulations, we believe that KSW methodology consistently underestimates the true efficiency scores. In contrast, all the statistics estimated by order-*m* methodology are larger than those shown in DEA and FDH (the average efficiency scores in order-*m* for the period 2008–2015 are 0.6518 for the lowest quantile (Q1), 0.8093 for the mean and 1.0000 for the upper quartile (Q3)). Therefore, the experiment leads us to understand that the order-*m* method overestimates the true efficiency scores.

As regards the rank estimates, note that in scenario 6, DEA and KSW methodologies performed best at identifying the ranks of the efficiency scores. Table 8 shows the rank correlation between the average cost efficiency estimates of the four methodologies for the period 2008–2015. As our Monte Carlo experiment showed, DEA and KSW have a high correlation between their rank estimates because of their similar distribution of the rankings. Accordingly, our results show a relatively high correlation between the rank estimates of these two estimators (0.8998). Moreover, although there is a relatively high correlation between order-*m* and FDH rank estimates with DEA and KSW, the latter two outperform order-*m* and FDH. As a consequence, DEA and KSW estimators would be preferred to identify the efficiency rankings, but order-*m* and FDH will not necessarily produce poor efficiency rankings.

6 Conclusion

Over the last years, many empirical research studies have set out to evaluate efficiency in local governments. However, despite this high academic interest there is still a lack of a clear, 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. In this context, if local regulators use efficiency

analysis models in order to set a benchmark and they take a decision based on an incorrect efficiency score, it could have relevant economic and political implications. Therefore, we note that each methodology leads to different cost efficiency results for each local government, but one method 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: DEA, FDH, order-*m*, and the estimator proposed by Kneip et al. (2008) (KSW). All these approaches have been widely studied in the previous literature, but little is known about their performance in comparison with each other. Indeed, no study has compared these efficiency estimators. In contrast to previous literature, which has regularly compared techniques and made several proposals for alternative ones, we compare the different methods used via Montecarlo simulations and choose the ones which performed better with our particular dataset, in other words, the most appropriate methods to measure local government cost efficiency in Spain.

Our data included 1846 Spanish local governments between 1000 and 50,000 inhabitants for the period 2008–2015. Note that the period considered is also important, since the economic and financial crisis that started in 2007 has had a huge impact on most Spanish local government revenues and finances in general. Under these circumstances, identifying a method for evaluating local governments' performance to obtain reliable efficiency scores and set benchmarks over time is even more important, if possible.

In general, we have observed that there is no approach suitable for all efficiency analysis. Beyond the obvious academic interest, if using efficiency results for policy decisions, local regulators must be aware of which part of the efficiency distribution is of particular interest (for example, identifying benchmark local governments might be important to decide penalty decisions to poor performers) and if the interest lies in the efficiency scores or the rankings, i.e., it should be considered where and when to use a particular estimator. It is obvious that obtaining reliable efficiency scores might have some implications for local management decisions. Therefore, gaining deeper insights into the issue of local government inefficiency might help to further support effective policy measures, both those that might be appropriate as well as those that are not achieving their objectives.

We learn that, for our sample of Spanish local governments, all methods showed some room for improvement in terms of possible cost efficiency gains, although some differences in the inefficiency levels obtained were also present. The methodologies which perform better with our sample of Spanish local governments are the DEA and FDH methods at the median and lower tail of the efficiency distribution (the former slightly underestimates efficiency while the latter slightly overestimates it), and FDH (followed by order-*m*) for local governments with higher performance, according to the findings in our simulations. Specifically, the results suggested that the average true cost efficiency would range between 0.5417 and 0.7543 during the period 2008–2015, suggesting that Spanish local governments could achieve the same level of local outputs with between 25% and 46% fewer resources. Similarly, the true efficiency scores at the lowest quantile would lie between 0.4252 and 0.5944, and at the upper quartile would be around 0.9860. Further, DEA and KSW methodologies performed best at identifying the ranks of the efficiency scores.

The obtained results provide evidence as to how efficiency could certainly be assessed as close as possible in order to provide some additional guidance for policy makers. In addition, these results are particularly important given the overall financial constraints faced by Spanish local governments during the period under analysis, which have come under increasing pressure to meet strict budgetary and fiscal constraints without reducing their provision of local public services. Therefore, identifying accurately efficiency gains might help to limit the adverse impact of spending cuts on local governments' service provision.

We also note that the effects on the methodological choice identified in this paper might be valid only for our sample dataset. However, the analytical tools introduced in this study could have significant implications for researchers and policy makers who analyse efficiency using data from different countries. From a technical point of view, our results are obtained using a consistent method, which provides a significant contribution to previous literature in local governments efficiency. We emphasise that few studies from this literature have attempted to use two or more alternative approaches in a comparative way (Narbón-Perpiñá and De Witte 2018a). Therefore, from a policy perspective one should take care when interpreting results and drawing conclusions from these research studies that have used only one particular methodology, since their results might be affected by the approach taken. We think that the implementation of our proposed method to compare different efficiency estimators would represent an interesting contribution that provides the opportunity for further research in this particular issue, given the lack of a clear and standard methodology to perform efficiency analysis.

Appendix: Estimation of Λ

We use the following semiparametric stochastic cost frontier model:

$$C_i = g(y_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{7}$$

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) υ_i, the two-sided random error term which is assumed to be normally distributed N$(0, \sigma_\upsilon^2)$, 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 of using 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 maximise it over the single parameter Λ:

$$\max_{\Lambda} \ln l(\Lambda) = \max_{\Lambda} \left\{ -n \ln \hat{\sigma} + \sum_{i=1}^{n} \ln \left[1 + \Phi \left(\frac{\hat{\epsilon}_i}{\hat{\sigma}} \Lambda \right) \right] - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \right\}, \quad (8)$$

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

$$\sigma^2 = \left\{ \frac{1}{n} \sum_{i=1}^{n} [C_i - \hat{E}(C_i|y_i)]^2 \middle/ \left[1 - \frac{2\Lambda^2}{\pi(1+\Lambda^2)} \right] \right\}^{1/2}, \quad (9)$$

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:

$$\hat{E}(C_i|y_i) = \sum_{j=1}^{n} C_j \cdot K\left(\frac{y_i - y_j}{h} \right) \middle/ \sum_{j=1}^{n} K\left(\frac{y_i - y_j}{h} \right), \quad (10)$$

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).

References

Aiello F, Bonanno G (2019) Explaining differences in efficiency: a meta-study on local government literature. J Econ Surv 33(3):999–1027

Aigner D, Lovell CK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econom 6(1):21–37

Andor M, Hesse F (2014) The StoNED age: the departure into a new era of efficiency analysis? A monte carlo comparison of StoNED and the "oldies" (SFA and DEA). J Prod Anal 41(1):85–109

Athanassopoulos AD, Triantis KP (1998) Assessing aggregate cost efficiency and the related policy implications for Greek local municipalities. INFOR 36(3):66–83

Badunenko O, Henderson DJ, Kumbhakar SC (2012) When, where and how to perform efficiency estimation. J R Stat Soc Ser A (Stat Soc) 175(4):863–892

Balaguer-Coll MT, Prior D (2009) Short-and long-term evaluation of efficiency and quality. An application to Spanish municipalities. Appl Econ 41(23):2991–3002

Balaguer-Coll MT, Prior D, Tortosa-Ausina E (2007) On the determinants of local government performance: a two-stage nonparametric approach. Eur Econ Rev 51(2):425–451

Balaguer-Coll MT, Prior D, Tortosa-Ausina E (2010) Decentralization and efficiency of local government. Ann Reg Sci 45(3):571–601

Balaguer-Coll MT, Prior D, Tortosa-Ausina E (2013) Output complexity, environmental conditions, and the efficiency of municipalities. J Prod Anal 39(3):303–324

Banker RD, Chang H, Cooper WW (1996) Simulation studies of efficiency, returns to scale and misspecification with nonlinear functions in DEA. Ann Oper Res 66(4):231–253

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30(9):1078–1092

Banker RD, Gadh VM, Gorr WL (1993) A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis. Eur J Oper Res 67(3):332–343

Banker RD, Morey RC (1986) The use of categorical variables in data envelopment analysis. Manag Sci 32:1613–1627

Bellod Redondo JF (2015) Plan E: la estrategia keynesiana frente a la crisis en España. Rev Econ Crít 20:4–22

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

Berger AN (1993) "Distribution-free" estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions. J Prod Anal 4:261–292

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

Bogetoft P, Otto L (2010) Benchmarking with DEA, SFA, and R, vol 157. Springer, New York

Bosch-Roca N, Mora-Corral AJ, Espasa-Queralt M (2012) Citizen control and the efficiency of local public services. Environ Plan C Gov Policy 30(2):248

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

Charnes A, Cooper WW, Lewin AY, Seiford LM (1994) Data envelopment analysis: theory, methodology and applications. Kluwer, Boston

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Cordero JM, Pedraja-Chaparro F, Pisaflores EC, Polo C (2017) Efficiency assessment of Portuguese municipalities using a conditional nonparametric approach. J Prod Anal 48(1):1–24

Daraio C, Simar L (2005) Introducing environmental variables in nonparametric frontier models: a probabilistic approach. J Prod Anal 24(1):93–121

Daraio C, Simar L (2007a) Advanced robust and nonparametric methods in efficiency analysis: methodology and applications, vol 4. Springer, New York

Daraio C, Simar L (2007b) Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. J Prod Anal 28(1–2):13–32

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

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

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

Deprins D, Simar L, Tulkens H (1984) The performance of public enterprises: concepts and measurements. In: Marchand M, Pestieau P, Tulkens H (eds) Chapter Measuring labor inefficiency in post offices. North Holland, Amsterdam, pp 243–267

Doumpos M, 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, Hafner CM (2014) Local government efficiency: the case of Moroccan municipalities. Afr Dev Rev 26(1):88–101

Fan Y, Li Q, Weersink A (1996) Semiparametric estimation of stochastic production frontier models. J Bus Econ Stat 14(4):460–468

Ferreira Da Cruz N, Cunha Marques R (2014) Revisiting the determinants of local government performance. Omega 44:91–103

Fogarty J, Mugera A (2013) Local government efficiency: evidence from Western Australia. Austral Econ Rev 46(3):300–311

Färe R, Grosskopf S, Lovell CAK (1985) The measurement of efficiency of production. Studies in productivity analysis. Kluwer-Nijhoff Publishing, Dordrecht

Färe R, Grosskopf S, Lovell CK (1994) Production frontiers. Cambridge University Press, Cambridge

Fried HO, Lovell CAK, Schmidt SS (eds) (1993) The measurement of productive efficiency: techniques and applications. Oxford University Press, Oxford

Fried HO, Lovell CK, Schmidt SS (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. J Urban Econ 60(3):443–462

Geys B, Heinemann F, Kalb A (2010) Voter involvement, fiscal autonomy and public sector efficiency: evidence from German municipalities. Eur J Polit Econ 26(2):265–278

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

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

Hintze JL, Nelson RD (1998) Violin plots: a box plot-density trace synergism. Am Stat 52(2):181–184

Ibrahim FW, Salleh MFM (2006) Stochastic frontier estimation: an application to local governments in Malaysia. Malays J Econ Stud 43(1/2):85

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

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

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

Kneip A, Simar L, Wilson PW (2008) Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. Econom Theory 24(6):1663–1697

Krüger JJ (2012) A Monte Carlo study of old and new frontier methods for efficiency measurement. Eur J Oper Res 222(1):137–148

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

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

Meeusen W, Van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18(2):435–444

Merrifield J (1994) Factors that influence the level of underground government. Public Finance Rev 22(4):462–482

Narbón-Perpiñá I, Balaguer-Coll M, Tortosa-Ausina E (2019) Evaluating local government performance in times of crisis. Local Gov Stud 45(1):64–100

Narbón-Perpiñá I, De Witte K (2018a) Local governments' efficiency: a systematic literature review–part I. Int Trans Oper Res 25(2):431–468

Narbón-Perpiñá I, De Witte K (2018b) Local governments' efficiency: a systematic literature review–part II. Int. Trans. Oper. Res. 25(4):1107–1136

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

Pacheco F, Sanchez R, 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. TQM J 26(3):275–283

Resti A (2000) Efficiency measurement for multi-product industries: a comparison of classic and recent techniques based on simulated data. Eur J Oper Res 121(3):559–578

Ruggiero J (1999) Efficiency estimation and error decomposition in the stochastic frontier model: a Monte Carlo analysis. Eur J Oper Res 115(3):555–563

Ruggiero J (2007) A comparison of DEA and the stochastic frontier model using panel data. Int Trans Oper Res 14(3):259–266

Simar L, Wilson PW (2008) The measurement of productive efficiency and productivity growth. Chapter Statistical inference in nonparametric frontier models: recent developments and perspectives. Oxford University Press, Oxford, pp 421–521

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

Vilalta M, 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

Worthington AC (2000) Cost efficiency in Australian local government: a comparative analysis of mathematical programming and econometrical approaches. Financ Account Manag 16(3):201–223

Zafra-Gómez JL, Muñiz-Pérez AM (2010) Overcoming cost-inefficiencies within small municipalities: improve financial condition or reduce the quality of public services? Environ Plan C Gov Policy 28(4):609–629

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Isabel Narbón-Perpiñá[1] · Maria Teresa Balaguer-Coll[2] · Marko Petrović[3] · Emili Tortosa-Ausina[3] (ORCID)

   Isabel Narbón-Perpiñá
   isabel.narbon@uab.es

   Maria Teresa Balaguer-Coll
   bcoll@uji.es

   Marko Petrović
   petrovic@uji.es

[1]   Department of Business, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

[2]   Department of Finance and Accounting, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló de la Plana, Spain

[3]   Department of Economics, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló de la Plana, Spain