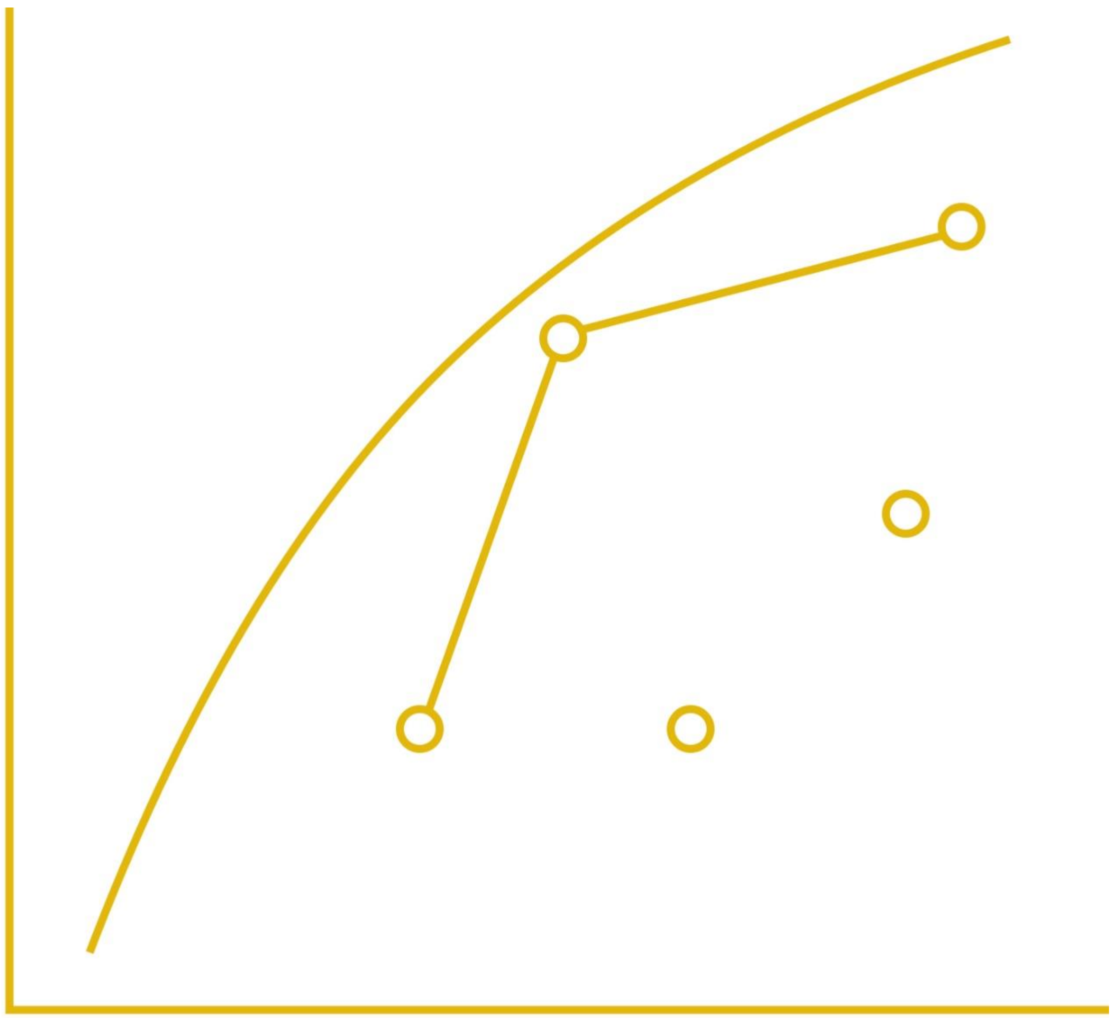
Iteratively Weighted Least Squares as an Alternative Frontier Methodology | Applied to the Local Administrative Public Services Industry

IPSE Studies | Jos L.T. Blank



Iteratively Weighted Least Squares as an Alternative Frontier Methodology | Applied to the Local Administrative Public Services Industry

Jos L.T. Blank

Institute for Public Sector Efficiency Studies

Delft, August 2018

Colofon

Productie en lay-out: Stichting IPSE Studies

Druk: Stichting IPSE Studies

Den Haag/Delft, 1 november 2018

E-mail: info@ipsestudies.nl

Internet: www.ipsestudies.nl

ISBN: 978-90-827258-2-7

Institute for Public Sector Efficiency Studies
IPSE Studies aims to promote research of public sector efficiency and effectiveness in the Netherlands as to contribute to policy discussions on the efficient organization of the public sector. The foundation coordinates scientific research and consultancy, maintains intensive (international) contacts with both policy and scientific communities and communicates its research through publications and the organization of seminars and conferences.

Content

Abstract

The academic literature provides excellent methodologies to identify best practices and to calculate inefficiencies by stochastic frontier analysis. However, these methodologies are regarded as a black box by policy makers and managers and therefore results are hard to accept. This paper proposes an alternative class of stochastic frontier estimators, based on the notion that some observations contain more information than others about the true frontier. If an observation is likely to contain much information, it is assigned a large weight in the regression analysis. In order to establish the weights, we propose an iterative procedure. The advantages of this more intuitive approach are its transparency and its easy application. The method is applied to Dutch local administrative services (LAS) in municipalities. The method converges quickly and produces reliable estimates. About 25% of the LAS are designated as efficient. The average efficiency score is 93%. For the average sized LAS no economies of scale exist.

JEL codes: C31; D24; I12; O39

Keywords: weighted least squares, frontier analysis, efficiency, local public services

1 Introduction

The recent financial and economic crises are forcing many administrations to cut budgets in various areas of public services. Particularly the European countries that are, or were, under direct budgetary supervision by the Euro Group and/or the IMF, such as Greece and Spain, are experiencing a tremendous impact on the service levels in education, healthcare, and infrastructure industries. Since these services are of great importance in the structural improvement of their economies – or in a broader sense, in the maintenance of their social welfare – the pressure on these services is great. The only way to balance shrinking budgets and the need for structural improvement is to enhance performance in these sectors. This implies that more effort must be put into finding ways to improve performance in the public sectors. This concerns good and reinforcing policies at both a government level and the level of the management of the executing public institutions. Academics can play an important role in identifying best practices in order to disseminate knowledge about which types of internal and external governance, incentive structures, market regulations, capacity planning, etc. might improve performance.

However, one might conclude (although not based on solid empirical evidence) that in many cases governments and those managing public institutions are operating in the dark. Academics fail both to bridge the gap between practice and theory, and to provide policymakers and management with evidence-based policy and management measures to strive for optimal strategies and business conduct. The academic literature provides excellent methodologies to identify best practices (see e.g. Fried et al. 2008) in stochastic frontier analysis (SFA) and data envelopment analysis (DEA). There are numerous examples of applications in public service industries, such as in health industry (see e.g. Blank & Valdmanis, 2008) or water and power utilities (see e.g. Bottasso et al., 2011; Murillo-Zamorano & Vega-Cervera, 2001). The technique is also being applied to compare the performance of countries or industries in different countries (see e.g. Shao & Lin, 2016; Chen & Lin, 2009).

Unfortunately, most of the researchers in this field have "lost" themselves in the techniques, instead of paying attention to practical and policy-relevant issues, and no connection is made with public administration research. Moreover, in the public administration literature, few references relate public administration concepts to empirical research based on frontier techniques. Gill and Meier (2000) complained more than a decade ago that frontier or best practices techniques were not being applied: "Public administration research has fallen notably behind research in related fields in terms of methodological sophistication. This hinders the development of empirical investigations into substantive questions of interest to practitioners and academics."

One may wonder why the frontier techniques have not become common practice in public administration research. A possible explanation is that these techniques are based on sophisticated mathematical economics, econometrics, and statistics. Aside from the technical problems researchers might face in applying these techniques, a strong reason also lies in the fact that policymakers and managers do not have faith in the results derived from these complex and rather non-transparent methodologies. It is not the mathematics that are involved causing acceptance problems but more the conceptual issues behind these techniques. Except for the seminal work by Meier and Gill (2000) in their *What Works: A New Approach to Program and Policy Analysis*, few serious attempts have been made to introduce more accessible and transparent methodologies that produce the same results as the existing state-of-the-art frontier techniques. In this paper, we present a more appealing technique that is based on the original ideas of Gill and Meier, and that provides results that are similar to SFA while presenting fewer computational problems. Meier and Gill take a rather modest position in stating that there technique is "..not a technique for estimating parameters but rather a qualitative exploratory and prescriptive procedure for isolating performance" (Meier & Gill, 2000: p10). In this paper we are more explicit about how weights should be derived and how the weighting scheme should be chosen in order to establish SFA look-alike results, which in turn leads to a less modest claim than Meier and Gill. As an illustration, the proposed method is applied to Dutch local public administrative services, namely the department of municipalities that provide drivers' licenses, passports, and other such documents.

This paper is organized as follows. In Section 2, we discuss the state of the art in efficiency measurement, and various types of frontier analysis techniques. In Section 3, we look at a few conceptual issues concerning the alternative method. In Section 4, we introduce a formal description of our model and the estimation procedure. In Section 5, we apply the method to a set of Dutch local government services data. We conclude the paper in Section 6.

2 Methodology

Best practices in public sector service delivery can be identified by various techniques. One of the most popular is the stochastic frontier analysis (SFA) methodology suggested by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broek (1977). This technique has become a standard in the parametric estimation of production and cost (or any other value) function. It is based on the idea that production (or cost) can be empirically described as a function of a number of inputs (or outputs and input prices), a stochastic term reflecting errors, and a stochastic term reflecting efficiency. In this approach, a stochastic term is added to an ordinary least squares (OLS) equation, where it is assumed that it follows a distribution with a non-negative support. This stochastic term is supposed to pick up the inefficiency for each firm. Maximum likelihood techniques can be used to estimate the parameters of the function and the parameters of the distribution of the stochastic components. For extensive discussions on this technique see, for example, Kumbakhar and Lovell (2000) and Fried, Lovell, and Schmidt (2008).

SFA is now very popular, and it has been applied in numerous empirical work (for extensive literature reviews, see also Fried, Lovell, and Schmidt (2008) and Blank (2000)). Nevertheless, the approach has also been widely criticized. The criticisms focus on two major points, namely the a priori specification of the production (or cost) function, and the assumptions about the distribution of the stochastic terms (see e.g. Ondrich & Ruggiero, 2001). Although both criticisms can be overcome to a certain extent, by using flexible forms and different assumptions about the distribution of the stochastic variables in the analysis, the rigidity might be seen as a problem. A third type of criticism, which is not expressed very often, is of a conceptual nature: The methodology suggests observing an unobservable (the efficiency), which can be derived from another unobservable (the measurement and specification error), within a rather complex econometric framework. Those who try to explain this approach to the non-initiated, such as managers and policymakers, are met with skepticism and disbelief. A technique like data envelopment analysis (DEA), which actually seeks observations that form the envelope, is far more appealing and more transparent. This is why in real-life problems, DEA has become a very popular tool in applied work. Another conceptual framing of SFA may tackle the problem and make the technique more accessible to non-experts. In applied research both methods are often combined (see e.g. Murillo-Zamorano & Vega-Cervera, 2001).

The original work by Aigner, Lovell, and Schmidt (1977) derives the stochastic frontier approach in the case of a single equation model, in which we can estimate only the technical or the cost efficiency. If we are interested not only in technical or cost efficiencies but also in allocative efficiencies, we need a multiple equations approach that will allow the underutilization or overutilization of inputs to be derived. However, the estimation of a multiple equations model – with a far-reaching decomposition of the underlying stochastic variables for measurement errors, and technical and allocative efficiency – is very troublesome. In particular, the theoretical linkage between the cost function and the input demand equations is extremely difficult to handle (this is the so-called Greene problem; Greene (1980)). The theoretical linkage follows from Shephard's Lemma and should also be reflected in the error structure of the model. Although some interesting solutions have been proposed – for example, applying shadow cost models (see Blank, 2009; Kumbhakar, 1997) or using Bayesian estimation techniques (see e.g. Tabak & Langsch Tecles, 2010) – new estimation problems occur. These approaches obviously suffer from an even greater lack of transparency. Since the suggested approach here actually measures the frontier, the Greene problem will be mitigated. In our example in Section 4 we only focus on a single equation approach. The relevance of aforementioned remark therefore is for further research.

Estimating a production, cost, or profit frontier (hereinafter "frontier") would be pointless were all firms to operate at full efficiency. Although one could use OLS to estimate the parameters of the model, in reality some firms are inefficient, which makes the estimation of the frontier a challenging task. This problem could be solved by neglecting the inefficient firms and only taking account of efficient firms. However, this method implies a priori knowledge of whether a firm is efficient, and knowledge about the efficient firms is generally not available prior to the estimation of a production frontier. Therefore, other methods for addressing this problem have been proposed.

An alternative to the original SFA approach is the thick frontier analysis (TFA) developed by Berger and Humphrey (1991). This approach, which is based on the idea of selecting efficient firms, allows the estimation of a single or a multiple equation. The technique uses a selection of firms in the top 10% (or any other percentage) and the bottom 10%. The production (or cost) function for both subsamples is estimated separately. Cost efficiencies are subsequently derived by taking the ratio of the average cost of the worst practice firms and the best practice firms. TFA has a number of advantages. Seemingly unrelated regression allows for a straightforward estimate of a system of a cost function and the corresponding share equations. TFA does not require any rigid assumptions about the distributions of the error components, nor does it suffer from the Greene problem. It is a conceptually very transparent and appealing approach, although it does have some serious drawbacks. It does not provide firm-specific cost efficiencies, but only rather general cost efficiency scores. From an econometric point of view, there is a loss of information, due to the discard of a large subset of observations, and it is questionable whether the researcher has the luxury of losing so many degrees of freedom.

Another approach to estimating a frontier – one that can be regarded as a successor to TFA – is provided by Wagenvoort and Schure (2006), who show how efficient firms can be identified if panel data are available. They use a recursive thick frontier approach (RTFA), dropping the most inefficient firm at each iteration. In each step, the firm-specific efficiency is calculated by averaging the residuals of a firm over the whole time period. Their final step consists of using the fully efficient firms to estimate the frontier. Although it is intuitively appealing, RTFA also has some serious drawbacks. It can be applied only to panel data. Further, it is assumed that inefficiency is time-invariant. This implies that a firm cannot change its efficiency in time – which is a rather rigid assumption, particularly in the case of a long time span. Another drawback is that it still depends on the assumption of a 0–1 probability of being efficient.

Our approach has some similarities to another alternative, namely quantile regression (see e.g. Koenker & Hallock, 2001). Whereas the method of least squares provides an estimate of the conditional mean of the dependent variable, quantile regression provides an estimate of the conditional median or any other quantile. In the case of the conditional median, the objective is to minimize the absolute residuals. For other quantiles, the absolute deviations are assigned an asymmetric weight (Koenker & Hallock, 2001).

In frontier analysis, for instance, one may choose the 75% quantile or the 90% quantile. The interesting aspect of this method is that it actually assigns more weight to observations that are close (conditionally on the explanatory variables) to the desired quantile. Thus, in contrast to TFA, it does not drop or ignore a number of observations. Although promising results have been achieved with this method, it also lacks transparency, perhaps even more than SFA does. The concept is very hard to understand, calculations are based on linear programming techniques, and no straightforward statistical inferences can be made. Further, it cannot be applied to systems of equations.

Our method also has a strong resemblance to earlier work by Meier and Gill (2000), who focused on investigating subgroups in a given sample by applying a method called substantively weighted least squares (SWLS). In an iterative procedure, SWLS selects the outliers from standard least squares (e.g., observations with residuals above 3 times the standard deviation of the residuals), and re-estimates the model by assigning weights equal to 1 to observations in the selection, and weights smaller than 1 to observations outside the selection. In an iterative procedure, the weights corresponding to the observations outside the selection are successively decreased. Although this method is quite appealing, it has no direct link to the standard productivity and efficiency literature, and the weights are handled in the iterations in a rather ad hoc way.

Our approach combines the best of many worlds. We argue that whether a firm is fully efficient does not concern a 0–1 probability, but is probabilistic. We therefore introduce weights to the observations and show the way in which a weighting scheme can be implemented in order to determine which firms are likely to be efficient and which are likely to be inefficient. At the same time, we are able to preserve the transparency of the RTFA and the SWLS method by applying standard least squares techniques and without losing any degrees of freedom, as occurs in RTFA (by creating a subsample of selected observations). With respect to the SWLS method, our approach does not assign common and rather arbitrary weights to the observations outside the selection. Instead, we use weights that reflect the probability of being efficient or nearly efficient, which implies a minimum loss of information, and therefore leads to more efficient estimates of the model parameters.

Our concept also translates to a cross-section setting so as to avoid the need for panel data. This also implies that we do not need to assume that inefficiency is time-invariant, which can also be regarded as a rather restrictive assumption in many efficiency models that are based on panel data.

Thus, our approach is related to the concept of stochastic frontier analysis, but is far more conceptually appealing. As in TFA, it can also be applied to multiple equation systems, while avoiding the Greene problem. Our alternative incorporates information derived from all the available data. It is based on an iterative weighted least squares (IWLS) method and can easily be programmed in standard econometric software.

3 Economic model and estimation

We start with the cost function, although the method may be applied to any other model (see e.g. Färe & Primont, 1995). We assume that the firm is cost-minimizing and that the total cost can be represented by a cost function *c(y,w)*, where *y* and *w* are a vector of outputs and input prices, respectively, that meets all the requirements it entails. For convenience, we rewrite the cost equations in terms of logarithms and add an error term.

$$\ln(C) = c(\ln(y)\,, \ln(w)) + \epsilon \qquad (1)$$

With:

C = total costs;

y = vector of outputs;

w = vector of input prices;

ε = error term.

Equation (1) can be estimated by a certain minimum distance estimator or, if one wants to check for heterogeneity, with fixed or random effects, which will result in consistent estimates of the parameters if $E[\varepsilon|y,w] = 0$. However, if some firms are inefficient – that is, they have a cost that is higher than what can be explained – the cost function or random noise with $E[\varepsilon] > 0$, causing biases in the parameters of equation (1).

We can reduce these biases by estimating equation (1) with weighted least squares, and assigning the "ill-behaving" observations a small weight and the "well-behaving" observations a large weight. Weighted least squares (WLS), which is also referred to as generalized least squares (GLS), is a widely used econometric technique; however, since the weights are generally not observable, they have to be estimated (see e.g. Verbeek, 2012). Our proposed weighting scheme is based on the residuals obtained after equation (1) has been estimated with least squares (LS),[1] as we know that firms that are highly inefficient, and thus likely to bias the results, will have a large residual $\hat{\varepsilon}$, where $\hat{\varepsilon}$ is the estimate of ε. The transformation of residuals into weights can be reflected by a weighting function $\omega(\hat{\varepsilon})$, which satisfies the requirements that it is monotonously non-decreasing in $\hat{\varepsilon}$ and always

[1] If equation (1) is estimated with fixed effects, the weights can also be based on the fixed effects, which would make our estimator into a generalized version of the estimator, suggested by Wagenvoort and Schure (2006).

non-negative. Simple examples of functions that satisfy these requirements are rank($-\hat{\varepsilon}$) or exp(-$\hat{\varepsilon}$) (in the case of the cost function). Although not strictly necessary for estimation, we should also like to impose a direct correspondence between the weights and the probability of firms being efficient. In the case that actual cost are below estimated cost (i.e. $\hat{\varepsilon} < 0$), the firm is assumed to be efficient and the corresponding weight is set at 1. Formally, ω($\hat{\varepsilon}$)=1 if $\hat{\varepsilon} < 0$.

Since the weighting scheme depends on $\hat{\varepsilon}$, an iterative reweighted least squares procedure can be implemented. This procedure is used for some robust regression estimators, such as the Huber W estimator (Guitton, 2000). This similarity is not a coincidence, since our proposed estimator can also be considered a robust type of regression. This implies that, after each WLS estimation, new $\hat{\varepsilon}$s are calculated, which are then used to generate new weights, which in turn are used in a next stage WLS estimation, until the convergence criterion is met. The convergence criterion we use requires that the parameter estimates do not differ by more than 1% from the previous stage. Note that if the parameter estimates are stable or almost stable, the residuals and the corresponding weights are also stable, implying that there is no more information available in the data to identify a firm that is probably more efficient than another.

Implementing the weights in the estimation procedure is straightforward. Instead of minimizing the sum of the squared residuals, the sum of the squared weighted residuals is minimized. Observations that show large deviations from the frontier will therefore contribute less to establishing the parameters of the cost function.

The way in which $\hat{\varepsilon}$ should be transformed into weights is obviously as debatable as the distributional assumption for the efficiency component in SFA. The weighting scheme should reflect the trade-off between noise and inefficiency. If one expects all firms to be efficient, the deviation from the frontier, captured by $\hat{\varepsilon}$, is mostly determined by noise. If a weighting scheme is used that rapidly reduces the weight if $\hat{\varepsilon}$ is increasing, the assessment of the level of the frontier will be overly optimistic, since firms that perform very well due to luck will be assigned a larger weight. On the other hand, if a weighting scheme that is virtually flat for all $\hat{\varepsilon}$ is used and many firms are inefficient, the estimation of the frontier will be too low, since firms that are effectively very inefficient will still be considered quite efficient. One way to determine the amount of noise versus inefficiency is to examine the skewness of the LS residuals. It is easy to implement other weighting schemes and see whether the results differ. This is another advantage of our approach over the SFA approach, which requires one to calculate the convolution of two random variables and derive the maximum likelihood, if one wants to use another distribution.

We also want to know the levels of inefficiency, rather than only the parameters of the cost function. In our proposed model, it is not obvious how these levels can be calculated, since for the calculation of $\mathrm{E}[\mu|\hat{\varepsilon}]$, where μ is the level of efficiency, we need at least the probability density distributions of μ and ε. However, our estimator does not require one to make an assumption about the distribution of μ. Nevertheless, this does not mean that we are unable to say something about the efficiency of each firm. Ondrich and Ruggiero (2001), for instance, showed that if a normal distribution is assumed to be noise, the ranking of $\hat{\varepsilon}$ is equal to the ranking of μ. Therefore, our model enables us to specify the efficiency ranking for each firm.

Although the distributional assumptions about the efficiency term are not necessary for the estimation, we might still use them to derive the efficiency scores. Therefore, we introduce the two usual unobservables *u* and *v*, representing the efficiency/inefficiency and the error term, respectively, so that ε = *u* +*v*. We simplify the original problem of Aigner et al. (1977) by estimating the distribution of only the error term, instead of both components simultaneously. Since we have identified the cost frontier, we are able to select a subsample of observations that satisfy *u* = 0, that is, all observations with an observed cost lower than or equal to frontier cost (*v* ≤ 0) and thus a weight of one. This sample can be seen as the fully efficient sample, which is in accordance with Kumbhakar et al. (2013), who developed a model that allows for fully efficient firms. Note that we are not able to identify observations that satisfy *u* = 0 and *v* ≥ 0, namely efficient firms with an observed cost greater than the frontier cost. We therefore assume that |*v*| in the subsample is distributed as $N^+(0, \sigma_v^2)$. The variance σ_v^2 can now be estimated by the sum of squared residuals, divided by the number of observations in the subsample (denoted as $\hat{\sigma}_v^2$). Furthermore, in the full sample, we assume that the subsample is representative of the variance of the random errors, and that random errors are distributed as $N(0, \hat{\sigma}_v^2)$. Since we now have an estimate of the variance of the random errors, we are also able to conditionally derive the expected efficiency from the residuals by applying, for instance, Materov's formula (Kumbhakar & Lovell, 2000: p.78):

$$M(\hat{u}_i|\hat{\varepsilon}_i) = \hat{\varepsilon}_i \left(\frac{\hat{\sigma}_u^2}{\hat{\sigma}_\varepsilon^2}\right) \text{ if } \hat{\varepsilon}_i \geq 0; = 0 \text{ otherwise} \quad (2)$$

with:

$$\hat{\sigma}_u^2 = \hat{\sigma}_\varepsilon^2 - \hat{\sigma}_v^2$$

The efficiency score then equals:

$$Eff_i = \exp(-M(\hat{u}_i|\hat{\varepsilon}_i) \quad (3)$$

There are, of course, other alternatives (see e.g., Kumbakhar & Lovell, 2000). Note that in comparison with the original paper by Jondrow et al. (1982), in our model we have swapped the roles of the random error and efficiency components. It is important to stress that we do not apply the distributional assumptions a priori to the errors and efficiency components in the estimation procedure as Jondrow et al. (1982) do , but do so only in the derivation of the efficiency scores.

4 Application to Dutch local administrative services

4.1 Model specification

We apply the well-known translog cost function model (Christensen et al., 1973; Christensen & Greene, 1976). In general, the model includes first- and second-order terms, as well as cross-terms between outputs and input prices on the one hand, and a time trend on the other hand. These cross-terms with a time trend represent the possible different natures of technical change. Cross-terms with outputs refer to output-biased technical change, while cross-terms with input prices refer to input-biased technical change.

$$
\begin{aligned}
\ln(C) = {} & a_0 + \sum_{m=1}^{M} b_m \ln(Y_m) + \sum_{n=1}^{N} c_n \ln(W_n) + \sum_{o=1}^{O} d_o \ln(Z_o) + \tfrac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M} b_{mm'} \ln(Y_m)\ln(Y_{m'}) + \\
& \tfrac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} c_{nn'} \ln(W_n)\ln(W_{n'}) + \tfrac{1}{2}\sum_{'0=1}^{O}\sum_{o'=1}^{O} d_{oo'} \ln(Z_o)\ln(Z_{o'}) + \sum_{m=1}^{M}\sum_{n=1}^{N} e_{mn} \ln(Y_m)\ln(W_n) + \\
& \sum_{o=1}^{O}\sum_{n=1}^{N} f_{on} \ln(Z_o)\ln(W_n) + \tfrac{1}{2}\sum_{o=1}^{O}\sum_{m=1}^{M} g_{ij} \ln(Z_o)\ln(Y_m) + h_0\, T + \\
& \sum_{m=1}^{M} i_{1i}\, T \ln(Y_m) + \sum_{n=1}^{N} j_{1i}\, T \ln(W_n) + \varepsilon
\end{aligned}
\qquad (4)
$$

With:

C = total costs;

Y_m = output m (m = 1,.., M);

T = year of observation;

W_n = price of input n (n = 1,.., N);

Z_o = fixed input o (o = 1,.., O).

a_o, b_m, c_n, d_o, b_{mn}, $c_{nn'}$, $d_{oo'}$, e_{mn}, f_{on}, g_{om}, h_o, i_{1m} *en* j_{in} parameters to be estimated.

Homogeneity of degree 1 in prices and symmetry is imposed by applying constraints to some of the parameters to be estimated. In formula:

$$b_{mm'} = b_{m'm} \quad ; \quad c_{nn'} = c_{n'n} \quad ; \quad d_{oo'} = d_{o'o}$$

$$\sum_{n=1}^{N} c_n = 1; \sum_{n=1}^{N} c_{nn'} = 0\ (\forall n'); \sum_{n=1}^{N} e_{mn} = 0\ (\forall m); \sum_{o=1}^{O} f_{on} = 0\ (\forall k); \sum_{n=1}^{N} j_{1n} = 0 \tag{5}$$

Equation (1) can be estimated by OLS or, if one wants to control for heterogeneity, with fixed or random effects, which will result in consistent estimates of the parameters if $E[\varepsilon|Y,W] = 0$. However, if some firms are inefficient – that is, they have a cost that is higher than can be explained – the cost function or random noise $E[\varepsilon] > 0$, causing a bias in α_0. Moreover, if the input mix of a firm is partly determined by the firm's expectations of its efficiency level, we even have $E[\varepsilon|Y,W] \neq \mu$, where μ is a constant. This also causes biases in the other parameters of equation (1).

4.2 Data

The data for this study cover the period 2005–10. They were obtained from municipal annual accounts by Statistics Netherlands (CBS). Annual financial and services data were collected by means of surveys covering all the local administrative services (LASs) in the Netherlands. For the purpose of this study, the data were checked for missing or unreliable data. Various consistency checks were performed on the data, in order to ensure that changes in average values and in the distribution of values across time were not excessive. After eliminating observations whose dataset contained inaccurate or missing values, we had an unbalanced panel dataset of 2,683 observations over the 6 years of the study. There are approximately 400 observations for each year.

The main service of the LASs is the provision of passports, driving licenses, and national identity cards, as well as birth, death, and marriage certificates that are retrieved from the local registry upon the request of citizens. We define three specific outputs: the (unweighted) sum of passports, identity cards, and driving licenses; the (unweighted) number of excerpts from municipal databases (such as death and birth certificates); and the number of marriages (which is included as arranging civil marriage ceremonies is an important activity of this part of local government).

Resources include all types of staff, material supplies, and capital input. Unfortunately, the data do not allow a distinction to be made between these different resources; therefore, the total input of resources is expressed by total costs only. Since we are dealing with data on a number of years, costs are deflated by the GDP price index. (for more details see Hulst van & de Groot, 2011). We do not distinct any environmental factors in our analysis. Table 1 provides the statistical descriptives of the data.

Table 1 Descriptives

	Mean	Std Dev.	Minimum	Maximum
Documents	11,044.3	16,657.3	280	223,050
Excerpts	4,180.0	9,219.0	72	116,995
Marriages	150.7	228.6	10	3397
Total cost (x 1000 euro)	1,322.7	3,540.7	15.2	67,206.5

Our pooled dataset for 2005–10 contains 2,683 cases. We refer to the extensive research report for more details (van Hulst & de Groot, 2011).

4.3 Estimation results

Since we are dealing with a relatively large number of cross-sectional units (> 400) and a limited number of periods (6 years), we ignore the fact that we are dealing with panel data (with respect to intra-firm correlations): The between variance is obviously far more important than the within variance. We estimate the cost frontier for 2005–10, with year fixed effects to allow for an annual shift of the frontier due to technological progress or other relevant changes to the production structure.

In our analysis, we use the following weighting scheme:

$$w = \frac{1}{\left(1+\frac{\hat{\varepsilon}}{\sigma_{\hat{\varepsilon}}}\right)} \text{ if } \hat{\varepsilon} > 0, \text{ else } w = 1 \qquad (6)$$

With:

$\sigma_{\hat{\varepsilon}}$ = the standard deviation of the least squares residuals.

As explained in the theoretical section, the weighting scheme is such that the weights are directly related to the efficiency scores. Efficient firms have weights equal to 1, while inefficient firms have efficiency scores equaling the weights multiplied by a constant (equal to the ratio of variances).

However, it is easy to implement other weighting schemes and see whether the results differ. This is another advantage of our approach over the SFA approach, which requires calculation of the convolution of two random variables and the derivation of the maximum likelihood, if one wants to use another distribution. As it turns out, our results were quite robust for another weighting scheme, based on rank numbers. In the case of IWLS estimation, we assume convergence if the maximum change in the parameters is less than 1% and the procedure stops. For convergence we needed 12 iterations in our application. So far we have found any problems with convergence whatsoever, which is a persistent problem in numerous SFA applications.

In order to get some insight between possible differences between SFA and IWLS we also estimated the cost function model with SFA, assuming that the efficiency component follows a half normal distribution. Both frontier methods are estimated using standard maximum likelihood and least squares methods with TSP software. Table 2 shows the estimates according to both estimation procedures.

Table 2 Estimates of frontier cost function by SFA and IWLS

		SFA		IWLS	
Variable		Est.	St.Err.	Est.	St.Err.
2006	a_2	0.034	0.021	0.037	0.016
2007	a_3	-0.097	0.025	-0.119	0.019
2008	a_4	-0.021	0.022	-0.056	0.017
2009	a_5	0.022	0.024	-0.014	0.019
2010	a_6	0.098	0.023	0.060	0.018
constant	a_0	-0.412	0.028	-0.362	0.015
Documents	b_1	0.598	0.103	0.638	0.086
Excerpts	b_2	0.238	0.091	0.227	0.071
Marriages	b_3	0.122	0.035	0.128	0.024
Doc. x Doc.	b_{11}	0.311	0.317	0.161	0.262
Doc. x Exc.	b_{12}	-0.096	0.268	-0.095	0.211
Doc. x Mar.	b_{13}	-0.120	0.085	-0.063	0.058
Exc. x Exc.	b_{22}	0.102	0.242	0.240	0.180
Exc. x Mar.	b_{23}	0.002	0.080	-0.130	0.052
Mar. x Mar.	b_{33}	0.192	0.056	0.347	0.033
$\boldsymbol{\sigma}$	σ_ε	0.368	0.014	0.292	
$\boldsymbol{\sigma_u/\sigma_v}$	λ	1.211	0.156	0.624	

A comparison of the outcomes of the SFA estimates and the IWLS, shows that a number of the estimated parameters are quite similar, in particular the parameters corresponding to the production terms in the equation (b_1, b_2 and b_3). Consequently, the calculated cost flexibilities for the average firm are almost identical ($\sum b_m = 0.96 \text{ versus } 0.99$). The parameters corresponding to the cross terms may show some differences, but none of them are significantly different (b_{11}, b_{12}, b_{22}, b_{23} *and* b_{33}). The same holds for the trend parameters (a_2-a_7), representing the frontier shift from year to year. As expected, all the parameter estimates according to the IWLS estimation are more efficient.

In order to underline the plausibility of the estimates, we derived a few other economically relevant outcomes. The first concerns the cost efficiency scores. Figure 1 shows the distribution of the efficiency scores in 2010, based on the IWLS estimation.

Figure 1 Distribution of cost efficiency scores, 2010

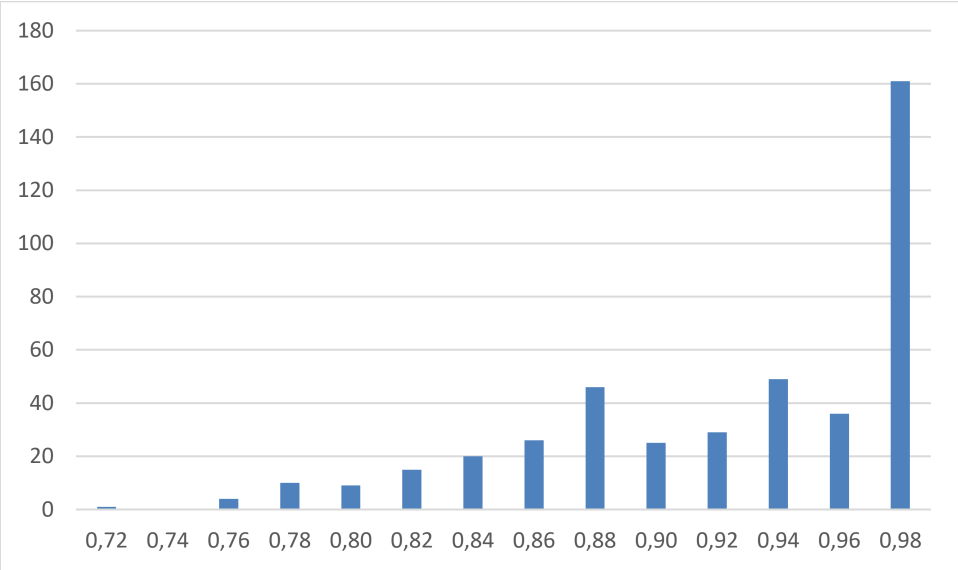


Figure 1 shows that in 2010, approximately one quarter of the LASs were efficient or almost efficient. Furthermore, the inefficient LASs show a plausible pattern of inefficiencies. The average efficiency is 94%, with a standard deviation of 6%. The minimum efficiency score is 69%. The efficiency scores between the years are very robust (not presented in the figure): The average efficiency scores over the years vary between 0.94 and 0.95. Comparing the IWLS efficiency scores to the SFA scores it appears that the IWLS scores are higher. The average difference is 7 percentage points. However, this difference refers only to the absolute level of the efficiency scores. The correlation between both types of efficiency scores equals almost 100% and the rank correlation equals 98%. Further, it shows that all the SFA identified efficient firms are also IWLS efficient, and that 81% of the IWLS efficient firms are also SFA efficient.

One of the major drawbacks of the TFA is that it requires sampling from a stratified sample. Since in this procedure we do not stratify the sample at all, it is questionable whether, regardless of certain characteristics, each LAS has an equal probability of being identified as an efficient LAS. Obvious characteristics that may affect the probability of being efficient/inefficient are the size and the year. We therefore inspected the distribution of the efficiency scores in relation to year and size. Figure 2 shows the number of efficient LASs in each year of the sample.

Figure 2 Number of efficient local administrative services by year

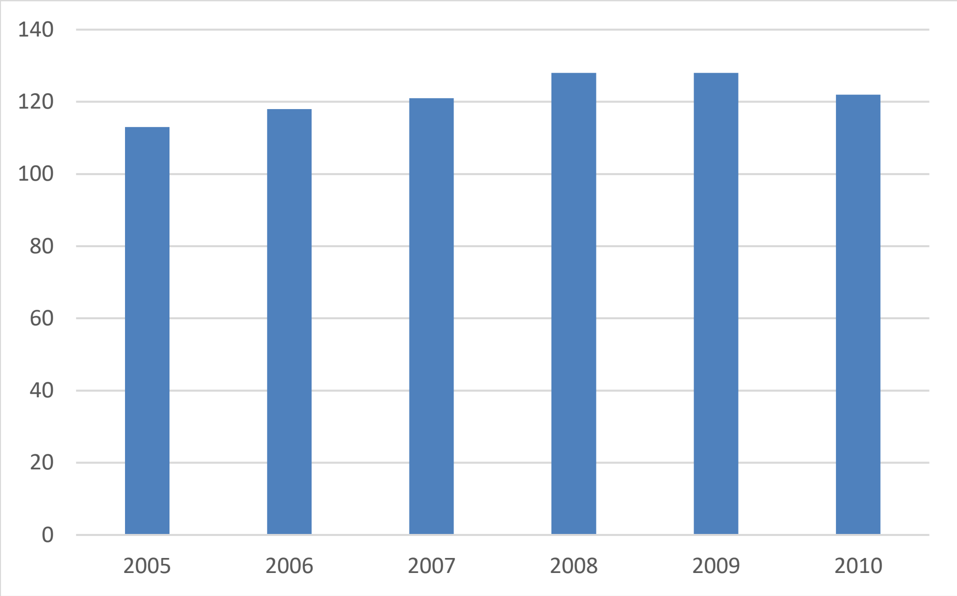


Figure 2 shows that the final selection of efficient LASs is quite uniformly distributed over the years, varying between 116 and 124, indicating that there is an equal probability of a municipality in a certain year to belong to the frontier. This shows that the procedure does not tend to favor a particular year.

Figure 3 shows the frequency distribution with respect to the size of the LASs (divided into four quartiles with respect to total cost).

Figure 3 Number of efficient local administrative services by size, 2010

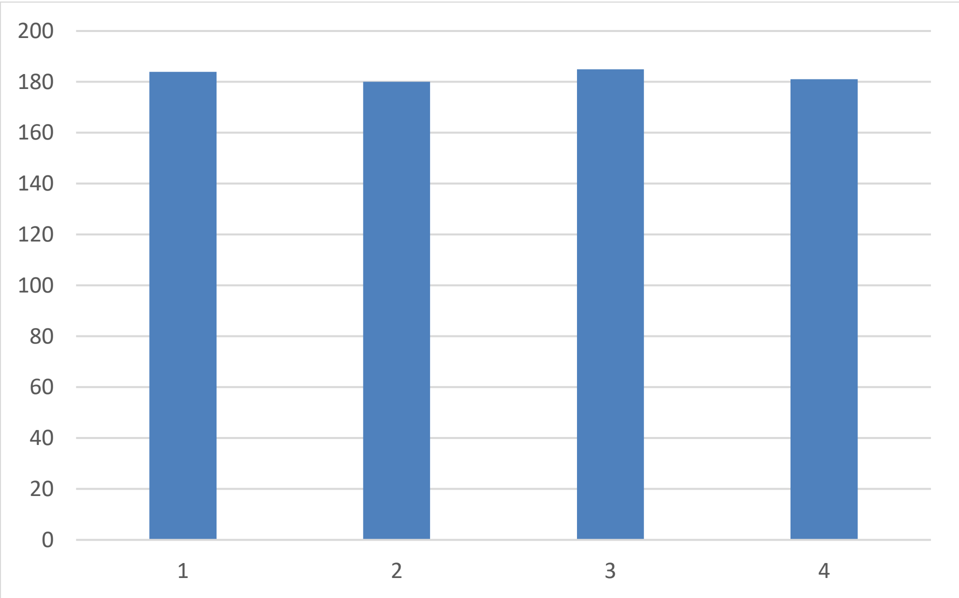


Figure 3 also shows that all the size categories are well represented by a substantial number of efficient LASs.

One of the restrictive assumptions in RTFA concerns the firm-specific efficiency through time. Since in our approach we allow for time varying efficiency, we are able to check this assumption. Based on the calculated total variance (0.0028), between variance (0.0021), and within variance (0.0007) of the residuals, it shows that one quarter of total variance can be attributed to the within variance and three quarters to the between variance. From this we can conclude that there is some consistency in the hospital efficiency through time, but that the assumption of constant firm-specific efficiency does not hold.

5 Conclusions

This paper proposes an alternative class of stochastic frontier estimators. Instead of making distributional assumptions about the error and efficiency component in the econometric specification of a cost function model (or any other model), this class is based on the notion that some observations contain more information than others about the true frontier. If an observation is likely to contain much information, it is assigned a large weight in the regression analysis. In order to establish the weights, we propose an iterative procedure. Since no a priori information is available, the first step consists of running a standard least squares (LS) method. Weights can subsequently be determined by the residuals obtained and a user-specified weighting function. The weights obtained allow for weighted least squares (WLS) to be applied. Since the WLS residuals will differ from the LS residuals, new weights are determined by means of an iterative procedure. In each step, the weights are updated and a new WLS regression is estimated. Since the negative residuals, by definition, represent the error component, the variance of these errors can easily be calculated and used as an estimator of the variance of the normal distribution of the noise. Similar to SFA, expected inefficiency and noise can be derived for all the other observations. The iterative procedure stops as soon as the change in the parameters between two iterations is less than a given threshold value.

The advantages of this approach include that it has high transparency, it is easy to apply to a fully specified model, and it is flexible. It allows the direct ascertainment of which observations largely determine the frontier. Its easy application to a fully specified model refers to a model that includes a cost function and its corresponding share equations. Its flexibility pertains to the use of several alternative weighting functions and the ease of testing for the sensitivity of the outcomes.

The model was applied to a set of Dutch local administrative services data that comprised 2,683 observations. The outcomes are promising. The model converges rather quickly and presents reliable estimates of the parameters, the cost efficiencies, and the error components. About 25% of the local administrative services are designated as efficient. The average efficiency score is approximately 93%. For the average sized LAS no economies of scale exist.

6 References

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

Berger, A.N., & Humphrey, D.B. (1991). The Dominance of Inefficiencies over Scale and Product Mix Economies in Banking. *Journal of Monetary Economics, 28*(1), 117-148.

Blank, J.L.T. (2000). *Public provision and performance: contributions from efficiency and productivity measurement*. Amsterdam: Elsevier.

Blank, J.L.T. (2009). Non-maximizing output behavior for firms with a cost-constrained technology. *Journal of Productivity Analysis, 31*(1), 27-32.

Blank, J.L.T., & Valdmanis, V.G. (2008). *Evaluating hospital policy and performance: contributions from hospital policy and productivity research*. Oxford: Elsevier JAI.

Bottasso, A., Conti, M., Piacenz, M., & Vannoni, D. (2011). The appropriateness of the poolability assumption for multiproduct technologies: Evidence from the English water and sewerage utilities. *International Journal of Production Economics, 130*(1), 112-117.

Chen, Y.H., & Lin, W.T. (2009). Analyzing the relationships between information technology, inputs substitution and national characteristics based on CES stochastic frontier production models. *International Journal of Production Economics, 120*(2), 552-569.

Christensen, L.R., & Greene, W.H. (1976). Economies of Scale in U.S. Electric Power Generation. *Journal of Political Economy, 84*(4), 655-676.

Christensen, L.R., Jorgenson, D.W., & Lau, L.J. (1973). Transcendental Logarithmic Production Frontiers. *The Review of Economics and Statistics, 55*(1), 28-45.

Färe, R., & Primont, D. (1995). *Multi-Output Production and Duality: Theory and applications*. Dordrecht: Kluwer Academic Publishers.

Fried, H.O., Lovell, C.A.K., & Schmidt, S.S. (2008). *The measurement of productive efficiency and productivity growth*. New York: Oxford University Press.

Gill, J., & Meier, K.J. (2000). Public Administration Research and Practice: A Methodological Manifesto. *Journal of Public Administration Research and Theory, 10*(1), 157-199.

Greene, W.H. (1980). On the estimation of a flexible frontier production model. *Journal of Econometrics, 13*(1), 101-115.

Guitton, A. (2000). Stanford Lecture Notes on the IRLS algorithm. Retrieved from http://sepwww.stanford.edu/public/docs/sep103/antoine2/paper_html/index.html

Hulst van, B.L., & de Groot, H. (2011). Benchmark Burgerzaken, Een empirisch onderzoek naar de kostendoelmatigheid van burgerzaken *IPSE-Studies reserch reeks*. Delft: IPSE-Studies (English summary).

Jondrow, J., Lovell, C.A.K., Materov, I.S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics, 19*(2-3), 233-238.

Koenker, R., & Hallock, K.F. (2001). Quantile Regression. *The Journal of Economic Perspectives*, *15*(4), 143-156.

Kumbhakar, S.C. (1997). Modelling Allocative Inefficiency in a Translog Cost Function and Cost Share Equations. *Journal of Econometrics, 76*, 351-356.

Kumbhakar, S.C., & Lovell, C.A.K. (2000). *Stochastic frontier analysis*. New York: Cambridge University Press.

Kumbhakar, S.C., Parmeter, C.F., & Tsionas, E.G. (2013). A zero inefficiency stochastic frontier model. *Journal of Econometrics*, *172*(1), 66-76.

Meeusen, W., & Van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*(8), 435–444.

Meier, K.J., & Gill, J. (2000). *What works: a now approach to program amd policy analysis*. Boulder: Westview Press.

Murillo-Zamorano, L.R., & Vega-Cervera, J.A. (2001). The use of parametric and non-parametric frontier methods to measure the productive efficiency in the industrial sector: A comparative study. *International Journal of Production Economics*, *69*(3), 265-275.

Ondrich, J., & Ruggiero, J. (2001). Efficiency measurement in the stochastic frontier model. *European Journal of Operational Research*, *129*(2), 434-442.

Shao, B.B.M., & Lin, W.T. (2016). Assessing output performance of information technology service industries: Productivity, innovation and catch-up. *International Journal of Production Economics*, *172*, 43-53.

Tabak, B.M., & Langsch Tecles, P. (2010). Estimating a Bayesian stochastic frontier for the Indian banking system. *International Journal of Production Economics, 125*(1), 96-110.

Verbeek, M. (2012). *A guide to modern econometrics* (4 ed.). Chichester: John Wiley & sons, Ltd.

Wagenvoort, R.J.L.M., & Schure, P.H. (2006). A Recursive Thick Frontier Approach to Estimating Production Efficiency*. *Oxford Bulletin of Economics and Statistics*, *68*(2), 183-201.