Delivery offices cost frontier : a robust non parametric approach with exogenous variables.

C. Cazals*, P. Dudley** ,
J.-P. Florens*, S. Patel**, F. Rodriguez**

* IDEI, University of Toulouse - France
** Royal Mail Group

1. Introduction:

This paper deals with an analysis of the efficiency of delivery post offices from an estimated cost frontier.

Various methodologies are available to estimate the efficiency frontier, including the stochastic or deterministic parametric approach, where we have to specify a particular functional form for the frontier function, and the non parametric approach, where no particular form is assumed for this frontier.

Our focus in this paper is to apply a robust non parametric approach for the cost frontier estimation, called the order-m frontier (Cazals, Florens and Simar (2002)), to extend it to take into account some environmental variables. We illustrate our approach using a cross-section data set on delivery offices.

We give first in the section 2 an overview of existing standard methodologies to estimate efficiency frontier. In the section 3, we present the order-m frontier approach, with an extension to include explanatory variables. Section 4 presents the results estimated using our data set on delivery offices. Section 5 concludes.

2. Estimation of efficiency frontier: a review

The objective of the estimation of a frontier function is to compare performance of different production units, in terms of production or cost. We can analyse a production frontier, where we search for a given level of input the unit which produces the maximum output, or cost frontier, where we search for a given level of output the unit which produces at a minimal cost. To be consistent with the empirical part of this paper, we will focus here on the cost frontier.

We consider cross-sectional data for a sample of N production units. For each unit we observe the production cost, denoted by C, outputs, denoted by Y, and some environmental variables, denoted by Z. These environmental variables represent exogenous factors with a possible influence on the production process. For the postal delivery activity which is studied in this paper, outputs may be different types of delivered mail and environmental variables may be the delivery area of the delivery offices, the number of delivery points, or the density of the delivery zone (number of delivery points per square kilometre).

For the empirical efficiency analysis of production units the task is to examine the relation between cost and outputs and exogenous factors at the frontier for the data cloud (that is at the bottom of the data cloud in the case of a cost frontier) in order to obtain a “best practice” function to be used to evaluate the distance of each producer to this function.

Usually the cost model is written as follows:

$$C = f(Y,Z)e^{u} ,$$

or more often after a logarithmic transformation of the variables (that *is* $c=Ln\ C$, $y=Ln Y$ and $z=Ln\ Z$):

$$c = \varphi(y,z) + u$$

where $\varphi(y,z)$ is the unknown cost frontier and u represents the deviation of the observed cost to the cost frontier and is a random variable.

There are two categories of methodologies to analyze this frontier: the parametric approach which assumes a functional form is chosen for the frontier *φ(.)*, and a non parametric approach where we do not specify a particular form for *φ(.)*.

2.1 Parametric frontier:

In a parametric frontier model we have to choose a specific form for the frontier function (as Cobb-Douglas or translog for example), and we estimate then a given number of parameters *θ*. The frontier function is written : *φ(y,z; θ)*. According to the assumption made on the random deviation term *u,* we consider deterministic or stochastic frontier models.

- *Deterministic parametric frontier:*

With this approach any deviation of the observations from the cost frontier is attributed only to inefficiency - then in the previous specification u represents the inefficiency term. Mainly two methods can be considered to estimate the parameters: mathematical programming and regression type models.

(i) Mathematical programming:

Aigner and Chu (1968) are the first to have developed this technology. For a sample of N decision units, the parameters of the cost frontier are obtained as the solution of either a linear programming model, expressed as:

$$\min_{\theta} \sum_{i=1}^{N} \left| c_i - \varphi(y_i, z_i; \theta) \right|$$

$$\text{subject to } c_i - \varphi(y_i, z_i; \theta) \geq 0, i = 1, \cdots, N,$$

or a *quadratic* programming model, expressed as:

$$\min_{\theta} \sum_{i=1}^{N} \left(c_i - \varphi(y_i, z_i; \theta) \right)^2$$

$$\text{subject to } c_i - \varphi(y_i, z_i; \theta) \geq 0, i = 1, \cdots, N.$$

In these models we have to calculate the parameter vector θ that minimizes the sum of deviations (for linear model) or the sum of squared deviations (for quadratic model) above the frontier.
The main drawback of this method is that the parameters are "calculated" rather than "estimated", and then the statistical properties are not well established.

(ii) Regression type models (Greene (1993), Lovell (1993)) :

The frontier model is written :

$$c = \varphi(y, z; \theta) + u,$$

where we assume that $E(u \mid Y = y, Z = z) = \mu$, where μ is a constant.

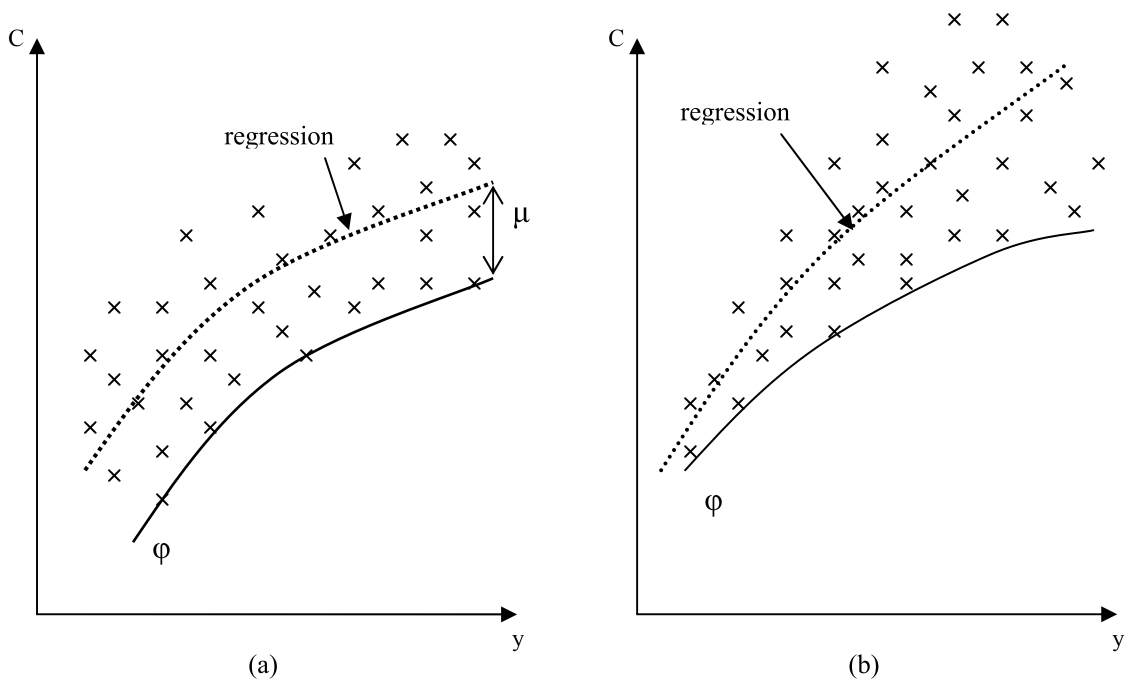
Then we obtain a regression model expressed as:

$$c = \varphi(y, z; \theta) + \mu + \varepsilon,$$

where ε is such that $E(\varepsilon \mid y, z) = 0$.

Estimation methods of this model are "corrected ordinary least squares" (COLS) or "modified ordinary least squares" (MOLS). With the COLS, in a first stage we apply standard OLS on data, and in a second stage we shift down the intercept to obtain a frontier that bounds data from below. The MOLS method is very similar to the COLS method. The difference comes from the assumption of an explicit one-sided distribution for inefficiency component ε (half-normal or exponential). The intercept is shifted down by the mean for the assumed one-sided distribution. However with this method, there is no guarantee that the frontier envelop all data from below.

This methodology relies on an extremely strong hypothesis: the frontier is parallel to the standard regression function, that is the frontier function is assumed to have the same shape as the central tendency. The graph 1 shows two examples of data samples, where for one of them this method can be reasonable (case (a)) and for the other (case (b)) it is not.



Graph. 1

➢ *Stochastic parametric models:*

These models were first introduced by Aigner and al. (1977) and Meeusen and Broeck (1977). The stochastic frontier is based on a composed error model. It decomposes the deviation of observations from the frontier between inefficiency and usual "noise" component which captures other stochastic effects and unobserved heterogeneity. A stochastic frontier model is written :

$$c = \varphi(y, z; \theta) + \upsilon + \varepsilon$$

where υ is the usual random term (stochastic noise) and has a two-sided distribution, and ε represents the inefficiency and has a one-sided distribution.

Maximum likelihood method can be used to estimate this model after selection of a distribution for *ε* and υ. Most common choices for these distributions are a Normal distribution for υ, and a half or truncated Normal, an Exponential or a Gamma distribution for *ε*.

More precisely if we assume for example $\upsilon \sim N(0,\sigma_\upsilon^2)$ and $\varepsilon \sim N^+(0,\sigma_\varepsilon^2)$, and *ε* and *v* are distributed independently of each other and of y and z, the log-likelihood function can be written as :

$$l(\theta,\sigma,\lambda) = \text{constant} - NLn\sigma + \sum_{i=1}^{N} Ln\Phi\left(\frac{u_i\lambda}{\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N} u_i^2$$

where $u_i = \upsilon_i + \varepsilon_i$, $\sigma^2 = \sigma_\upsilon^2 + \sigma_\varepsilon^2$, $\lambda = \frac{\sigma_\varepsilon}{\sigma_\upsilon}$, and *Φ(.)* is the standard normal cumulative distribution function.

After estimation we obtain the residuals $\hat{u}_i$, i= 1, …, N, which contain information on the inefficiency $\hat{\varepsilon}_i$. We use the conditional distribution of *ε* given *u* to estimate the decision unit specific inefficiency (Jondrow, Lovell, Materov and Schmidt (1982). We have:

$$E(\varepsilon_i \mid u_i) = \sigma_* \left(\frac{\phi(\frac{u_i\lambda}{\sigma})}{1-\Phi(-\frac{u_i\lambda}{\sigma})} + \frac{u_i\lambda}{\sigma} \right)$$

where $\sigma_* = \frac{\sigma_\varepsilon^2\sigma_\upsilon^2}{\sigma^2}$

and the inefficiency for the decision unit *i* is $E_i = \exp(-E(\varepsilon_i \mid u_i))$.

An alternative estimator of the inefficiency term, given by Battese and Coelli (1988), is:

$$E(e^{\varepsilon_i} \mid u_i) = \left(\frac{1-\Phi(\sigma_* - \frac{\mu_{*i}}{\sigma_*})}{1-\Phi(-\frac{\mu_{*i}}{\sigma_*})} \right) \exp(-\mu_{*i} + 0.5\sigma_*^2)$$

where $\mu_{*i} = \frac{u_i\sigma_\varepsilon^2}{\sigma^2}$. These two measures can be different since the exponential is a non linear function ($\exp(E(\varepsilon_i \mid u_i)) \neq E(\exp(\varepsilon_i) \mid u_i)$). (See Kumbhakar and Lovell (2000) for more technical details about stochastic frontier analysis).

The main drawback of these parametric models is that they are based on some strong assumptions. First we have to select a particular functional form for the frontier function : we then add a risk of specification error. Second, we have to assume specific distributions for the random "noise' and the inefficiency term. Different distributions may give different results for the inefficiency values. And finally the assumption of independence for inefficiency term and noise with respect to outputs and environmental variables is a strong assumption; it is possible that there is some relation between inefficiency and the levels of outputs and exogenous variables.

2.2 Non parametric frontier:

This approach does not rely on a particular functional form for the cost frontier. The most popular approaches for the estimation of a non parametric frontier are the Free Disposal Hull (FDH) approach and the Data Envelopment Analysis (DEA) approach.

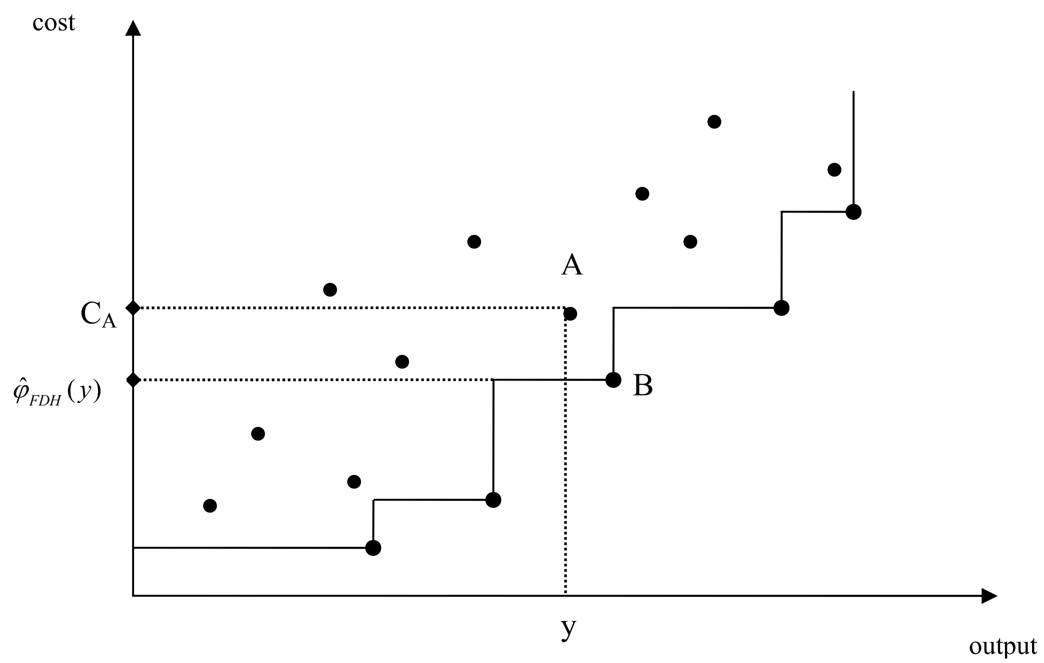
Usually the method used to introduce the influence of exogenous variables on efficiency is to estimate, first, the frontier without exogenous variables and, in a second stage, to explain the derived estimated inefficiencies by exogenous variables using a regression function. We briefly set out below the methodologies for FDH and DEA approaches.

- *FDH frontier:*

This method was first proposed by Deprins, Simar and Tulkens (1984). The frontier for a given value of output, *y*, is computed as follows:

$$\hat{\varphi}_{FDH}(y) = \min_{i:Y_i \geq y} C_i$$

That is, for a unit producing a level *y* of output, $\hat{\varphi}_{FDH}$ is the minimum cost of all the observed all observation units producing at least a level *y* of output. The graph 2 shows an example of this frontier.



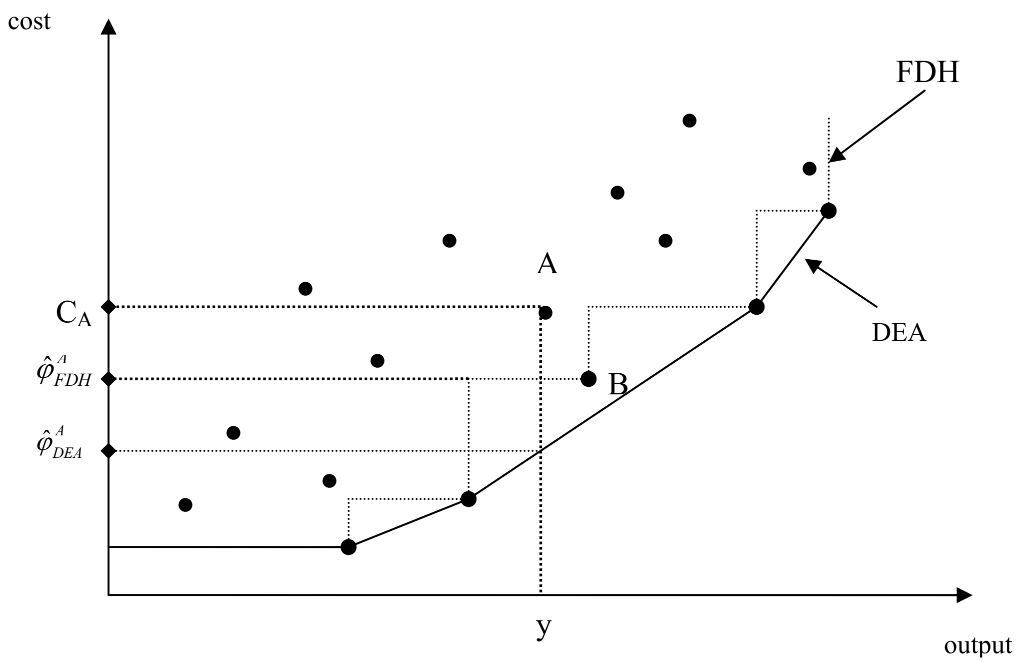
Graph 2

On this graph the firm A produces *y* units of output. Among all observations that produce at least *y*, the minimum cost is realised by the firm B. Then the value of the frontier for a level y of output is $\hat{\varphi}_{FDH}(y)$, corresponding here to the observed cost for firm B.

The cost efficiency is represented by the distance between observed costs and the frontier, for example, for the firm A on the graph, the cost efficiency can be measured by the ratio $\hat{\varphi}_{FDH}(y)/C_A$. In this case, a firm is efficient when this ratio is equal to 1 (firm B is efficient), and inefficient if this ratio is lower than one.

- *DEA frontier:*

The DEA method is similar to the FDH method except that a condition of convexity is added to the determination of the frontier. A typical DEA cost frontier is shown in graph 3 (where we report also the FDH frontier in dot line).



Graph 3

The cost efficiency measurement for firm A in this case, $\hat{\varphi}^A_{DEA}/C_A$, is smaller than the one we obtained with FDH. We can also notice that the firm B is ranked as inefficient when we apply this DEA methodology because it is above the DEA frontier. Indeed, typically efficiency measures based on DEA are less than or equal to efficiency measures based on FDH, and the number of efficient observations is larger with FDH method than with DEA method.

The main advantages of DEA and FDH methods are that no hypothesis is required for the form of the frontier and they are easy to compute.

But these two methods, as they envelop all observation points, are very sensitive to outliers. Moreover as they are deterministic models they fail to account for the influence of statistical noise.

A typical approach in the literature to take into account the influence of environmental variables and statistical noise is to use the cost efficiency measure as the dependant variable in a parametric regression on the environmental variables[1]. However the statistical properties of the obtained estimators with this regression are not available (Simar (1992)).

An original non parametric estimation method of frontier has been developed by Cazals, Florens and Simar (2002) with the advantage of being more robust to the outliers. This frontier is called the "*order-m frontier*" (or "*m-expected frontier*").

3. The order-m frontier:

This method is a regularization method which "eliminates" some points, such that the frontier does not envelop all the observation points. It estimates an expected minimal cost frontier rather than the full frontier as with FDH or DEA method. Without environmental variables and for the case of a single output the order-m cost frontier is defined as follows. For a given level of output y, we consider *m* random variables $C_1, \cdots, C_m$ drawn from the conditional distribution of the cost *C*, given $Y \geq y$, where *m* is an integer greater or equal to 1. Then we may define the expected value of the minimum of these *m* random variables as[2]:

$$\varphi_m(y) = E\left[\min\left(C_1, \cdots C_m\right) | Y \geq y\right] = \int_0^\infty \left[S_c(c \mid y)\right]^m dc$$

where $S_c(c \mid y) = \Pr(C \geq c \mid Y \geq y)$ is the conditional survivor function of the cost C, given the output $Y \geq y$. This order-m frontier does not realize the full efficiency cost frontier (it does not envelop the data cloud) but rather gives the expected minimum cost among *m* production units drawn from all production units producing at least a level *y* of output.

An estimator of this function is:

$$\hat{\varphi}_{m,n}(y) = \hat{E}\left[\min\left(C_1, \cdots C_m\right) | Y \geq y\right] = \int_0^\infty \left[\hat{S}_{c,n}(c \mid y)\right]^m dc$$

The integral can be computed in practice, but a simple Monte-Carlo procedure can be applied to easily approximate the empirical expectation. For a given level of output y, we draw a sample of size *m* with replacement among those c_i where $y_i \geq y$. We denote this sample $(c_1^b, \cdots c_m^b)$ and we compute $\tilde{\varphi}_m^b(y) = \min_{j=1,\cdots,m} (c_j^b)$. We redo the two previous stages for b= 1, …, B, with B large; and finally we approximate $\hat{\varphi}_{m,n}(y)$ by $\frac{1}{B}\sum_{b=1}^{B} \tilde{\varphi}_m^b(y)$.

It is interesting to note that $\hat{\varphi}_m$ is a decreasing function of *m*; the limiting case when $m \rightarrow \infty$ achieves the full efficient frontier (FDH frontier), that is: $\lim_{m \to \infty} \hat{\varphi}_{m,n}(y) \to \hat{\varphi}_{FDH}(y)$; and for m fixed we have:

$$\sqrt{n}\left(\hat{\varphi}_{m,n}(y) - \varphi_m(y)\right) \rightarrow N(0, \sigma^2(y))$$

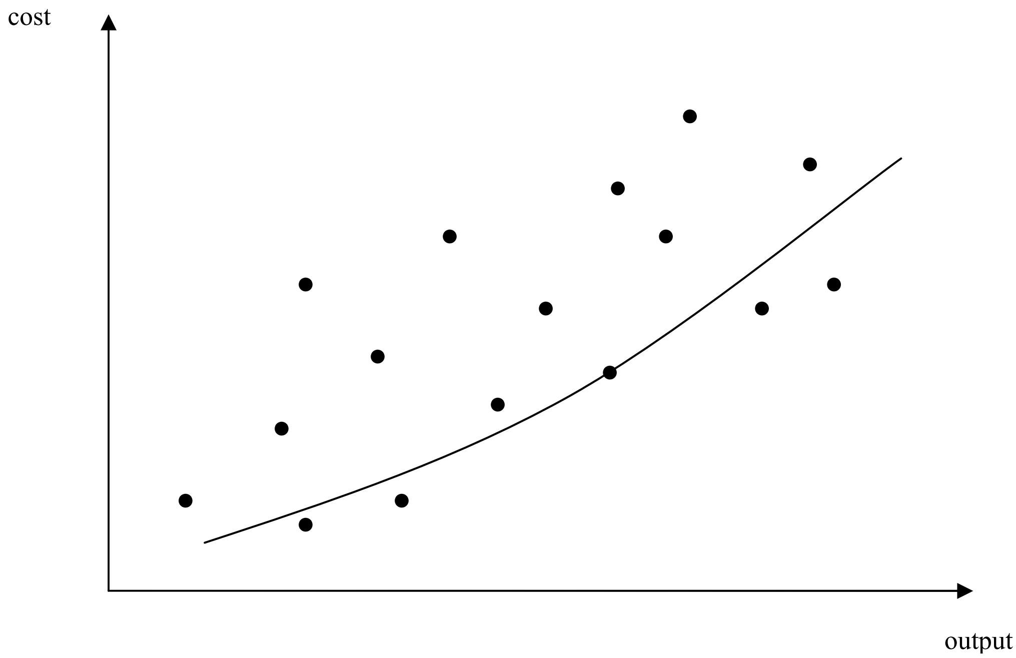
as $n \rightarrow \infty$. The expression for $\sigma^2(y)$ is given in Cazals, Florens and Simar (2002).

---

[1] Some models extend this two-stage method to a three-stage method, in which the third stage uses a DEA or FDH model to evaluate again inefficiency after adjustment of cost or outputs (Fried, Lovell, Schmidt, Yaisawarng (2002)).

[2] See Cazals, Florens and Simar (2002) for more technical details.

The graph 4 illustrates an example of such an order-m cost frontier. The observations above the frontier are inefficient, whereas the observations near or below the frontier are efficient or "super-efficient". A firm is super efficient if the efficiency ratio (frontier cost on observed cost) is greater than one.



Graph 4

This approach can be extended to take into account environmental variables, Z, which may have an influence on the cost frontier (see Cazals, Florens and Simar (2002), Daraio and Simar (2005)).

The idea is to be able to compare costs for decision units with similar values for the environmental characteristics. An appropriate method is to condition the production process to a given value z of Z. We replace the conditional survivor function $S_c(c \mid y)$ in the definition of the order-m frontier given previously, by the conditional survivor function $S_c(c \mid y, z) = \Pr(C \geq c \mid Y \geq y, Z = z)$. A smoothing technique is required to estimate this conditional survivor function, like a kernel method for example. In this case, the estimator of this conditional order-m frontier is:

$$\hat{\varphi}_{m,n}(y \mid z) = \int_0^\infty \left[\frac{\sum_{i=1}^{n} \mathbf{I}(c_i \geq c, y_i \geq y) K(\frac{z_i - z}{h_n})}{\sum_{i=1}^{n} \mathbf{I}(y_i \geq y) K(\frac{z_i - z}{h_n})} \right]^m dc$$

where **I**(.) is the indicator function, *K(.)* is a kernel and h_n is the smoothing bandwidth. We observe that we are faced with the problem of the curse of dimensionality in the dimension of Z, arising from the non parametric method and the use of the smoothing technique which impose limits on the introduction of more exogenous variables.

Again this estimator can be evaluated with a Monte-Carlo procedure similar to the previous one, except in the first stage where we draw here a sample of size m with replacement with a probability $\frac{K(\frac{z_i-z}{h_n})}{\sum_{j=1}^{n} K(\frac{z_j-z}{h_n})}$, among observations C_i such that $Y_i \geq y$.

In this paper we suggest the use of another smoothing technique: a "k-nearest-neighbor" (k-NN) method. The idea is that when estimating the efficient (order-m) cost, $\varphi_m(y_0 \mid z_0)$, for a decision unit with values (c_0, y_0, z_0) for cost, output and environmental variable respectively, we consider only the *k* observations nearest to z_0. An estimator of this conditional order-m frontier can be written:

$$\hat{\varphi}_{m,n}(y_0 \mid z_0) = \hat{E}\left[\min\left(C_1, \cdots C_m\right) \mid Y \geq y_0, \left|Z - z_0\right| \leq h_0\right],$$

where h_0 is a varying bandwidth, that is its value is such that exactly k observations verify $\left|Z - z_0\right| \leq h_0$. This estimator can be computed again with a Monte-Carlo procedure where in the first stage we draw the sample of size m with replacement within the sub-sample of k observations, among those C_i such that $Y_i \leq y$.

The parameter *k* plays the role of a smoothing parameter, similar to the bandwidth for the kernel smoother. We observe a selection problem for this parameter. Various methods of selection exist in the literature for non parametric econometrics (see Härdle (1990), Pagan and Ullah (1999)). We apply here the likelihood cross-validation criterion (as in Daraio and Simar (2005)). The underlying idea is to optimize the estimation of the density of Z.

We obtain varying bandwidths h_i, i=1, .., n, such that we have the same number k of observations in the neighbourhood of z when estimating the frontier.

We have to find the value of *k* which maximises the score function[3]:

$$CV(k) = \frac{1}{n}\sum_{i=1}^{n} \log(\hat{f}_k^{(-i)}(z_i))$$

where

$$\hat{f}_k^{(-i)}(z_i) = \frac{1}{(n-1)h_i} \sum_{j=1, j\neq i}^{n} K\left(\frac{z_j - z_i}{h_i}\right)$$

is the leave-one-out kernel density estimate of *Z*, h_i is the local bandwidth such that there exist *k* observations z_j ,verifying $|z_j - z_i| \leq h_i$ and K(.) is a kernel.

4. Cost frontier estimation for delivery offices:

We study the delivery cost efficiency by means of the order-m frontier using a sample provided by Royal Mail of 1108 delivery offices observed in 2003/04. For each delivery office, the cost variable, C, is the total labour hours and the output, Q, is the "weighted" mail volume provided by Royal Mail.

[3] For more details see Silverman (1986).

The analysis includes consideration of the environmental variable : traffic per delivery point denoted QDP. Indeed it has been shown from other analysis to be a significant variable in explaining the cost function for delivery offices.

As a first step, the order-m frontier is estimated without the environmental variable, for different values of *m*. As mentioned in the presentation of the methodology, this cost frontier does not envelop all the data points, but it converges toward the FDH efficiency frontier as *m* increases. In what follows, all delivery offices on or below the frontier will be qualified as "efficient" delivery offices.

Table 1 shows the results for different values of *m* if the order -*m* frontier is estimated without the environmental variable. It gives the percentage of efficient delivery offices and the percentage of potential cost saving under this assumption. This latter value is the cost (number of labour hours) that can be saved if all inefficient delivery offices become efficient (shift on the frontier) in percentage of observed total cost (total number of labour hours).

For values of *m* between 80 and 120, the percentage of super-efficient delivery offices lies between 10% and 15% and the potential cost saving is around 20-21% where the order –*m* frontier is estimated without the explanatory variable. The FDH efficient frontier, which takes no account of heterogeneity and explanatory variables, gives a potential saving of 27.8% with 3.9% efficient delivery offices. The table indicates that this value is being reached for the order-m frontier for large values of *m* as it converges toward the FDH frontier.

M	% of efficient DO	% of potential cost saving
20	38.3%	12.6 %
30	29.6%	14.9%
50	20.4%	17.7%
80	14.8%	19.9%
100	12.0%	20.9%
120	10.4%	21.6%
Infinity (=FDH)	3.9%	27.8%

Table 1: Efficiency results with no explanatory variables

The main drawback of this particular application of the order-m frontier methodology is that it considers that all deviation from frontier to be due to inefficiency (like with standard deterministic non parametric approaches such FDH or DEA). A more satisfactory application is to take into account some environmental variables (observed heterogeneity) into the estimation process that may be able to explain some of the apparent inefficiency within FDH arising from comparison of each office with offices with dissimilar environmental characteristics.

The order-m frontier was extended conditional to the environmental variable QDP (traffic per delivery point), following the definition given in the previous section. After application of the likelihood cross-validation criterion with the Gaussian kernel we find the optimal k value, k_{opt}, equal to 242. Table 2 shows the results for the order-m frontier conditional to the environmental variable QDP, for different values of *m* and for values of *k* fixed to its optimal value and also k_{opt} ±50% (that is 121 and 363), in order to examine the sensitivity of the results to the value of k.

	k=121		k=242		k=363	
m	**% of efficient DO**	**% of potential cost saving**	**% of efficient DO**	**% of potential cost saving**	**% of efficient DO**	**% of potential cost saving**
20	49.28%	7.18%	46.66%	8.11%	45.85%	8.49%
80	31.41%	10.12%	24.46%	12.24%	21.23%	13.2%
120	30.6%	10.54%	22.02%	12.99%	18.23%	14.15%

Table 2: Efficiency results of different m and k values using volume per delivery point as an explanatory variable

We must first notice that the results are relatively stable over large range of k. The introduction of this environmental variable reduces inefficiency: potential cost saving lies between 8% and 13% for k_{opt}, according to the values of m.

The set of results chosen within this range depends on the degree of conservatism taken over the value of *m*; lower values of *m* reflect a more conservative comparison of delivery offices and estimation of potential savings. The value of *m* affects the number of offices against which any individual office is being compared and the lower the *m* the smaller is the size of the sample being used for this comparison thereby reducing the inefficiency estimate. While the individual ranking of the inefficiency results show a relatively high correlation of more than 0.9 over the range of m (more precisely, 0.9365 between m=20 and m=80, and 0.994 between m=80 and m=120), they also show that the value of *m* does affect the scale of the inefficiency estimate at some individual offices (as might be expected from applying different samples for comparison implicit in the different m values). Therefore, in practice, there remains an element of judgement in forming a view of the potential overall level of efficiency improvement from this analysis, though the range is significantly reduced by the inclusion of the environmental variable.

5. Conclusion:

This paper focuses on the cost efficiency analysis of delivery offices where the cost is measured by the number of labour hours. It reviews the existing methodologies that could be applied and some of their drawbacks; for example, parametric models impose some strong assumptions where a specific functional form has to be chosen for the relation between cost and output and also generally a distribution for the efficiency component has to be chosen.

A non parametric methodology is preferred in this paper. The most frequently used applications of this methodology are non parametric deterministic frontiers such FDH or DEA frontier. These frontiers 'envelop' the data, as all the observations are above the frontier; then these techniques are very sensitive to extreme values and to outliers.

We use a non parametric estimator of the efficient frontier, called the order-m frontier, based on the concept of 'expected minimum cost', which is more robust to outliers as it does not envelop all the data. Moreover, in the case of delivery offices, as the sample size is reasonably large and thereby very susceptible to heterogeneity, we introduce the environmental factor “traffic per delivery point” in the model in order to capture existing heterogeneity that might otherwise erroneously be ascribed as being inefficiency within the cost of delivery offices.

The main result for delivery offices is an estimated potential cost saving (in terms of labour hours) which is between 8% and 13% of the total cost of Royal Mail when we estimate the frontier with an environmental variable (volume per delivery point). The inclusion of the single environmental variable reduces the estimated cost saving from a range of 15% to 20% without this variable. The difference in estimates arising from the inclusion of environmental variables emphasises the importance of taking into account observed heterogeneity of delivery offices through such variables in the frontier estimation. However as the estimation of the conditional frontier requires a smoothing technique we are faced with the curse of dimensionality, which imposes a limit on the introduction more exogenous variables with smaller sample sets.

References:

Aigner, D. and S. Chu (1968): "On Estimating the Industry Production Function", *American Economic Review*, 58: 826-839.

Aigner, D, C.A.K Lovell. and S. Schmidt (1977): "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics, 6, 21-37.*

Cazals C., J.-P. Florens and L. Simar (2002): "Nonparametric Frontier Estimation: a Robust Approach", *Journal of Econometrics*, 106(1), 1-15.

Daraio, C. and L. Simar (2005): "Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach", *Journal of Productivity Analysis*, 24, 93-121.

Deprins, D., L. Simar and H. Tulkens (1984): Measuring Labor-Efficiency in Post Offices, in: M. Marchand, P. Pestieau, H. Tulkens (Eds), *The Performance of Public Enterprises: Concepts and Measurement* (Morth-Holland, Amsterdam), 243-267.

Fried, H., C.A.K. Lovell, S. Schmidt and S. Yasawarng (2002): Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis", *Journal of Productivity Analysis*, 17, 157-174.

Greene, W. (1980): "Maximum likelihood estimation of Econometric Frontier Functions", *Journal of Econometrics*, 13(1): 27-56.

Härdle; W. (1990): *Applied Nonparametric Regression.* Econometric Society Monographs 19. Cambridge University Press.

Kumbhakar, S.C. and C.A.K. Lovell (2000): *Stochastic Frontier Analysis.* Cambridge University Press.

Lovell, C.A.K (1993): Production Frontiers and Productive Efficiency", in: H. Fried, C.A.K. Lovell & S. Schmidt (Eds), *The Measurement Productive Efficiency: Techniques and Applications* (Oxford University press), 3-67.

Meeusen, W., and J. Van Den Broeck (1977): "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Errors", *International Economic Review,* 18, 435-444.

Pagan, A. and A. Ullah (1999): *Nonparameric Econometrics*, Cambrige University Press.

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

Simar, L. (1992): "Estimating Efficiencies Frontier Models with Panel Data: a Comparison of Parametric, Non-parametric and Semi-parametric Methods with Bootstrapping", *Journal of Productivity Analysis*, 3, 167-203.