INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

STOCHASTIC FDH/DEA ESTIMATORS FOR FRONTIER ANALYSIS

SIMAR, L. and V. ZELENYUK

Stochastic FDH/DEA estimators for Frontier Analysis

Léopold Simar*

Institut de Statistique - Université Catholique de Louvain

IDEI, GREMAQ - Toulouse School of Economics

Valentin Zelenyuk

KEI - Kyiv Economics Institute

KSE - Kyiv School of Economics

National University, Kyiv-Mohyla Academy

May 12, 2008

Abstract

In this paper we extend the work of Simar (2007) introducing noise in nonparametric frontier models. We develop an approach that synthesizes the best features of the two main methods in the estimation of production efficiency. Specifically, our approach first allows for statistical noise, similar to Stochastic Frontier Analysis (even in a more flexible way), and second, it allows modelling multiple-inputs-multiple-outputs technologies without imposing parametric assumptions on production relationship, similar to what is done in non-parametric methods (DEA, FDH, etc. . .). The methodology is based on the theory of local maximum likelihood estimation and extends recent works of Park, Kumbhakar, Simar and Tsionas (2007) and Park, Simar and Zelenyuk (2006). Our method is suitable for modelling and estimation of the marginal effects onto inefficiency level jointly with estimation of marginal effects of input. The approach is robust to heteroskedastic cases and to various (unknown) distributions of statistical noise and inefficiency, despite assuming simple anchorage models. The method also improves DEA/FDH estimators, by allowing them to be quite robust to statistical noise and especially to outliers, which were the main problems of the original DEA/FDH. The procedure shows great performance for various simulated cases and is also illustrated for some real data sets.

Key words : Stochastic Frontier, Nonparametric Frontier, Local Maximum Likelihood
JEL Classification: C13, C14, C2

---

1 Frontier Models, Basic Notations

1.1 Introduction

Nonparametric frontier models are very appealing because they do not impose restrictive assumptions on the shape of the frontier and on the stochastics that governs the inefficiencies. Nonparametric estimators (DEA/FDH) are very popular and are based on envelopment techniques. Statistical inference is now available and robust versions (order-m and order-α quantile frontiers) have been developed to reduce the sensitivity of these estimators to outliers and extreme values (see Daraio and Simar, 2007 and Simar and Wilson, 2008 for recent surveys). But the basic drawback of these models is that they do not allow for noise. Recently, Simar (2007) describes a Data Generating Process allowing to handle nonparametric stochastic frontier in a full multivariate setup and suggests a new estimator which improves the performance of FDH/DEA estimators in the presence of noise. The limitation of the approach is that the size of the noise (in term of noise to signal ratios) has to be small to moderate.

In this paper, we extend the idea and propose an approach which allows to estimate nonparametric stochastic frontier in a more general setup, with no restriction on the size of the noise. The methodology is based on the theory of local maximum likelihood estimation and extends recent works of Park, Kumbhakar, Simar and Tsionas (2007) and Park, Simar and Zelenyuk (2006), by adapting the tool to the full multivariate model described in Simar (2007). In a sense, the method allows to clean the data from the noise and allows in a second step to derive *stochastic versions* of DEA/FDH estimators. Our method is also suitable for modeling and estimation of the marginal effects onto inefficiency level jointly with estimation of marginal effects of input. The approach is robust to heteroskedastic case and to various (unknown) distributions of statistical noise and of inefficiency. As expected, in the case of the presence of noise in the DGP, the proposed stochastic DEA/FDH estimators outperform the standard DEA/FDH estimators, but interestingly, if the DGP does not contain any noise, the stochastic DEA/FDH performs as well, if not better, than the classical DEA/FDH.

Our paper is organized as follows. In the end of this section we briefly remind the basic concepts and notations for defining the efficient frontier and the efficiency measures, with their classical deterministic DEA/FDH estimators. In Section 2, we define our stochastic model and explain how we can estimate it by using local maximum likelihood methods. We also derive our stochastic version of the DEA/FDH estimators. Section 3 illustrates the performance of our estimator with some simulated data sets as well as with real data sets from the literature. Section 4 concludes. The mathematical details for deriving the local likelihood function, and its customized version to our model, are presented in Appendix A

1.2 Efficient frontier and efficiency scores

We begin by introducing some basic concepts of nonparametric efficiency measurement. The presentation is in the spirit of Simar and Wilson (2008) and it is limited to the output oriented case; the same could be done in the input orientation case. Suppose producers use an input vector $x \in \mathbb{R}_+^p$ to produce an output vector $y \in \mathbb{R}_+^q$. The production set of the feasible input-output combinations can be defined as:

$$\Psi = \left\{ (x,y) \in \mathbb{R}_+^{p+q} \,|\, x \text{ can produce } y \right\}. \tag{1.1}$$

Often, various assumptions are made on the attainable set, such as free disposability, convexity, returns to scale, etc. Free disposability of inputs and of outputs means, if $(x,y) \in \Psi$, then any (x',y') such that $x' \geq x$ and $y' \leq y$ is also in Ψ. Often, but not always, convexity of Ψ is also assumed. We will consider the variable returns to scale case, but of course, the presentation could be adapted to any other returns to scale assumption.

The Farrell-Debreu output (technical-) efficiency score (Debreu, 1951 and Farrell, 1957) of a firm operating at the level (x,y) is defined as:

$$\lambda(x,y) = \sup \left\{ \lambda \,|\, (x, \lambda y) \in \Psi \right\}. \tag{1.2}$$

A value $\lambda(x,y) = 1$ means that the producer operating at the level (x,y) is output efficient, while a value $\lambda(x,y) > 1$ indicates inefficiency, in the sense that it is possible to have the radial increase in all the outputs $\lambda(x,y)$-times in order to reach the efficient frontier. For a given level of input and an output mix, the efficient level of output is then given by:

$$y^\partial(x,y) = \lambda(x,y)y. \tag{1.3}$$

Note that under free disposability, the Farrell-Debreu output efficiency can also be defined as:

$$\lambda(x,y) = \sup\{\lambda \,|\, H_{X,Y}(x, \lambda y) > 0\}, \tag{1.4}$$

where $H_{X,Y}(x,y) = \text{Prob}(X \leq x, Y \geq y)$ is the probability of the firm (x,y) to be dominated (see Cazals, Florens and Simar, 2002, for details)

Since the attainable set Ψ is unknown, so is its frontier $y^\partial(x,y)$ and the input efficiency score $\lambda(x,y)$. The best we can do is to estimate these quantities from a sample of i.i.d. observations $\mathcal{X} = \{(X_i, Y_i) \,|\, i = 1, \ldots, n\}$, generated according to the joint probability measure of (X,Y).

In order to make inference, we need to define the Data Generating Process (DGP), *i.e.*, the statistical model providing the assumptions under which the random sample (X_i, Y_i)

has been generated. Nonparametric models avoid restrictive parametric assumptions on the various components of the DGP, in order to provide inference in a very flexible and general framework.

1.3 Deterministic frontier models and DEA/FDH estimators

Most of the existing literature on nonparametric frontier has been first developped in the framework of *deterministic frontier models*, where it is assumed that $\text{Prob}((X_i, Y_i) \in \Psi) = 1$. In this framework the Free Disposal Hull (FDH) estimator has been proposed by Deprins *et al.* (1984) when only free disposabilty of inputs and outputs are assumed and Data Envelopment Analysis (DEA) estimator, initiated by Farrell (1957), assumes in addition the convexity of the attainable set.

The FDH is the most natural nonparametric estimator obtained by replacing the unknwon $H_{X,Y}(x, y)$ in (1.4) by its empirical version, the proportion of points in the sample dominating (x, y) *i.e.* $\widehat{H}_{X,Y,n}(x, y) = (1/n) \sum_{i=1}^{n} \mathbb{1}(X_i \leq x, Y_i \geq y)$ where $\mathbb{1}(\cdot)$ is the usual indicator function. It turns out (Cazals *at al.*, 2002) that the corresponding estimator of the attainable set is given by the free disposal hull of the sample points:

$$\widehat{\Psi}_{FDH,n} = \left\{ (x, y) \in \mathbb{R}_+^{p+q} \mid X_i \leq x, Y_i \geq y, \text{ for all } (X_i, Y_i) \in \mathcal{X} \right\}. \tag{1.5}$$

The DEA estimator is the convex hull of the FDH estimator, it can be written as:

$$\widehat{\Psi}_{DEA,n} = \left\{ (x, y) \in \mathbb{R}_+^{p+q} \mid \sum_{i=1}^{n} \gamma_i X_i \leq x, \sum_{i=1}^{n} \gamma_i Y_i \geq y, \right.$$
$$\left. \gamma_i \geq 0, \ i = 1, \ldots, n \text{ such that } \sum_{i=1}^{n} \gamma_i = 1 \right\}. \tag{1.6}$$

The resulting estimators of the efficiency scores can be obtained by plugging the estimators of Ψ in (1.2). This results in sorting algorithms for computing $\widehat{\lambda}_{FDH,n}(x, y)$ and solution of a linear program for obtaining $\widehat{\lambda}_{DEA,n}(x, y)$.

Deterministic frontier models have the drawbacks of (i) not allowing random noise in the DGP and, as a result, (ii) being very sensitive to extreme data points and outliers. The latter issue is addressed by extensions of the DEA/FDH estimators to partial orders frontiers, as the order-m frontiers (Cazals *et al.*, 2002) and the order-α quantile frontiers (Daouia and Simar, 2007), but in essence, they remain deterministic models since they do not allow for the presence of noise in the DGP. Simar and Wilson (2008) propose a survey on the properties of these nonparametric estimators and their variations and they provide details on how to make inference in these models. For the purpose of the paper here it is sufficient to remind that the rate of convergence of the FDH estimator is $n^{1/(p+q)}$ whereas for the DEA (with the additional assumption of convexity of Ψ) the achieved rate is $n^{2/(p+q+1)}$.

2 A Stochastic Frontier Model and its Estimation

2.1 Introducing noise

When introducing noise in the model, we are not sure that all the observations are really in the attainable set and envelopment estimators are no more appropriate since they would also envelop noisy data points. This complicates the analysis and the estimation of the frontier because we have to identify inefficiency from noise when analyzing the distance from a data point to the frontier. This family of models allowing noise are known as *stochastic frontier models* and, due to the complexity of the models, the literature was first developped in the simpler parametric setup.

The pionneering work in this framework is Aigner, Lovell and Schmidt (1977), and Meeusen and van den Broek (1977). In their approaches (and all the existing variants), specific parametric analytical forms are needed for the shape of the boundary of Ψ, and for the probability structure of the noise and of the efficiency distributions. Typically, the models take the (log-)linear form

$$Y_i = \beta_0 + \beta' X_i + v_i - u_i, \ i = 1, \ldots, n \tag{2.1}$$

where, the inputs/outputs are usually in the log-scale and the random inefficiency term is, for instance, $u_i \overset{iid}{\sim} Exp(\lambda)$ or $u_i \overset{iid}{\sim} \mathcal{N}^+(0, \sigma_u^2)$ and the noise term is $v_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_v^2)$. Generally, it is supposed that v is independent of (u, X) and that also u is independent of X. These approaches work well and have well established properties but they are limited by all these restrictive parametric hypotheses.

Recently new attemps have been developped to try to handle noise in multivariate nonparametric forntier models. Simar (2007) offers a recent survey on the topic explaining the limitations of the different approaches proposed so far (this includes, *e.g.*, Atkinson and Primont, 2002, Gstach, 1998, Land, Lovell and Thore, 1993, or Petersen and Olesen, 1995) and suggests, as a first attempt, a DGP allowing to introduce noise in a fully multivariate nonparametric setup. It is noticed there, using results from Hall and Simar (2002), that a full nonparametric specification of the DGP leads to non-identifiable models. However, if the noise is not too big in terms of noise to signal ratio, Simar (2007) proposes stochastic versions of DEA/FDH estimators that perform better than the envelopment estimators in the presence of noise and perform as well when in fact there is no noise in the DGP. It appears also that these estimators are much more robust to extreme values and outliers than the original DEA/FDH estimators. Still the approach is limited to situations where the noise has a small size.

Recently, Kumbhakar, Park, Simar and Tsionas (2007) have proposed a local maximum

likelihood approach for nonparametric stochastic frontier models. They analyze local polynomial models which encompass a particular anchoring parametric model like, *e.g.*, model (2.1). The local model is a parametric one and so is well identified, but by localizing the polynomial approximations, the model shares the flexibilty of nonparametric models. This is a very promising approach, but so far, it was limited to one-dimensional response variable (input or output).

The idea of this paper is to extend the Kumbhakar *et al.* (2007) ideas to a full multivariate setup by using the full multivariate DGP suggested by Simar (2007). This allows to handle noise in a quite general multivariate nonparametric stochastic frontier setup. In addition, the method provides attractive stochastic versions of the DEA/FDH estimators that circumvents the main weakness of these methods - presence of noise and especially outliers.

2.2 The Data Generating Process

We first summarize the basic idea of Simar (2007). By analogy with the parametric stochastic frontier models, the noise is introduced in the output space (when output-oriented). In parametric models, like (2.1), we have, in the log-scale, a parametric frontier model with additive inefficiency $(-u_i)$ and noise (v_i). In mutlivariate setups, since the efficiency measures are radial measures, the noise can be introduced, as it is done for the inefficiency, in the appropriate radial direction (here output), by using polar coordinates (see *e.g.*, Kneip, Park and Simar, 1998).

We consider first the DGP, generating points inside Ψ and then we will introduce the noise. The data (X_i, Y_i) are iid random variables generated according to a density $f_{X,Y}(x, y)$ having support Ψ. We can formulate the joint density in terms of the (partial-)polar coordinates (ω, η, x), where we use polar coordinates only for the output, so that $(x, y) \Leftrightarrow (\omega, \eta, x)$, where $\omega \in \mathbb{R}_+$ is the modulus and $\eta \in [0, \pi/2]^{q-1}$ is the amplitude (angle) of the vector y. To be explicit, $\omega = \sqrt{y^T y}$ and $\eta = (\eta_1, \dots, \eta_{q-1})$ where for $j = 1, \dots, q-1$, $\eta_j = \arctan(\frac{y^{j+1}}{y^1})$ if $y^1 > 0$ or $\eta_j = \frac{\pi}{2}$ if $y^1 = 0$. Often, η is referred as the "output mix", since it determines the position of ray y.

The joint density $f_{X,Y}(x, y)$ induces a density $f_{\omega,\eta,X}(\omega, \eta, x)$ on the polar coordinates and we decompose this joint density as follows:

$$f_{\omega,\eta,X}(\omega, \eta, x) = f_\omega(\omega \mid \eta, x) \, f_{\eta,X}(\eta, x),$$

where we assume all the densities exist. For a given (η, x) the frontier point $y^\partial(x, y)$ has a modulus which can be described through the upper boundary of the support of the density $f_\omega(\omega \mid \eta, x)$:

$$\omega(y^\partial(x, y)) = \sup\{\omega \in \mathbb{R}^+ \mid f_\omega(\omega \mid \eta, x) > 0\} \tag{2.2}$$

5

Note that the Farrell-Debreu output efficiency score can be redefined as

$$\lambda(x,y) = \frac{\omega(y^{\partial}(x,y))}{\omega(y)} \geq 1. \tag{2.3}$$

In a deterministic frontier framework, in order to achieve consistency of DEA/FDH estimators, we need the free disposability of Ψ assumption when FDH estimators are used and in addition, the convexity of Ψ when DEA estimators are used. Kneip *et al.* (1998) and Park, Simar and Weiner (2000) also introduce two additional regularity conditions:

A1. The function $\lambda(x,y)$ is differentiable in both arguments (smoothness of the frontier);

A2. For all x and η, $f_{\omega}(\omega(y^{\partial}(x,y)) \mid \eta, x) > 0$ (positive density on the efficient frontier).

All what precedes defines a DGP that generates data points inside the "deterministic" frontier of Ψ.

The radial-noise can now be introduced through the univariate modulus ω, conditionally on the input level x and on the output mix η (this is the multivariate analog of the parametric stochastic model (2.1)). We suppose here that the observations are made on noisy data in the output radial-direction: so we observe n i.i.d. random variables $\{(X_i, Y_i),\ i = 1, \ldots, n\}$, with polar coordinates (ω_i, η_i, X_i) where

$$\omega_i = \omega(y^{\partial}(X_i, Y_i))\, e^{-u_i}\, e^{v_i}, \tag{2.4}$$

where $\omega(y^{\partial}(X_i, Y_i))$, as defined in (2.2) is the frontier level. It is only a function of (η_i, X_i). Finally, in the log-scale, the model could be written as

$$\log \omega_i = r(\eta_i, X_i) - u_i + v_i, \tag{2.5}$$

with $u_i > 0$ and $E(v_i | \eta_i, X_i) = 0$. Here $r(\eta, x)$ is an unknown function determining the (log of the) frontier level for a given input level x and an output mix η. Conditionally on (η, x), the random elements u and v are assumed to be independent of each other, but in our approach, we do not need the usual homoskedastic assumption: u and v can be stochastically dependent of (η, x). Assumption $[A2]$ above implies that we consider conditional densities for u such that $f_u(0 \mid \eta, x) > 0$.

2.3 Local maximum likelihood estimation

For a given point of interest (x, y) with polar coordinates (ω, η, x), the problem is back to an univariate nonparametric stochastic frontier problem: given (η, x), we search for the estimation of $r(\eta, x) = \log \omega(y^{\partial}(x,y))$, from a random sample (ω_i, η_i, X_i), $i = 1, \ldots, n$.

For the estimation technique, we may adapt the nonparametric approach developed in Kumbhakar *et al.* (2007) to our particular setup. It is based on local polynomial maximum likelihood estimation techniques. The idea is to choose a local polynomial approximation for the unknown function $r(\eta, x)$ and to choose a localized parametric model for the shape parameters of the distribution of the random convolution error $\varepsilon = -u + v$, where other local polynomial approximations will be used. It is known from the litterature on local maximum likelihood (see *e.g.* Fan and Gijbels, 1993) that the local parametric specification can be viewed as an anchorage model but its localisation provides models sharing the flexibility of nonparametric models.

In summary, the nonparametric localized model can be written[1] as

$$\log \omega = r(\eta, x) - u + v, \tag{2.6}$$

where $(u|\eta, x) \sim |\mathcal{N}(0, \sigma_u^2(\eta, x))|$ and $(v|\eta, x) \sim \mathcal{N}(0, \sigma_v^2(\eta, x))$, u and v being independent conditionally on η, x. All functional parameters r, σ_u^2 and σ_v^2 are unknown functions and will be approximated by local polynomials. The procedure provides for any value of (η, x) estimates $\widehat{r}(\eta, x)$, $\widehat{\sigma}_u^2(\eta, x)$ and $\widehat{\sigma}_v^2(\eta, x)$ and if needed estimates of their partial derivatives with respect to the elements of η and of x.

The mathematical details for estimating this nonparametric model by local maximum likelihood techniques, and their customized version to our setup, can be found in the Appendix A. In our applications below we will use local linear approximations for the frontier and variance functions, because we know from Park, Simar and Zelenyuk (2006) that a lower order polynomials for the variance terms could jeopardize the properties of the estimator of the frontier function. Since we are also interested in the estimation of the marginal effects of the inputs on the efficiency level, we will even use a local quadratic polynomials for the functional σ_u^2. In principle, higher order approximations could be used but computational burden increases exponentially especially when the dimension of x and y is high.

2.4 Estimation of the individual efficiencies and of the frontier

Once the model (2.6) has been estimated, the estimation of the individual efficiency score for a particular fixed point $(x, y) \Leftrightarrow (\omega, \eta, x)$ is easy: we have indeed under our parametrization

$$E(u \mid \eta, x) = \sqrt{\frac{2}{\pi}} \sigma_u(\eta, x), \tag{2.7}$$

---

[1]Any other parametric model could be chosen as anchorage for the stochastic part of the model: in our simulations below the chosen parametrization turned out to be very flexible, providing good results even if the anchorage model was not the true one.

where $\sigma_u(\eta, x)$ can be estimated by $\sqrt{\widehat{\sigma}_u^2(\eta, x)}$. This can be done for any value of (η, x) and a sensitivity analysis could be performed to stress the influence of the output mix and/or the input level on the efficiencies.

As explained in the Appendix, if a quadratic local polynomial is used to approximate the function $\sigma_u^2(\eta, x)$, estimates of the partial derivatives of $\sigma_u^2(\eta, x)$ with respect to any of its component can be directly obtained. This allows to recover estimates of the partial derivatives of $E(u \mid \eta, x)$. Indeed, denoting (η, x) by z, we have

$$\frac{\partial E(u \mid z)}{\partial z_j} = \sqrt{\frac{1}{2\pi \sigma_u^2(z)}} \frac{\partial \sigma_u^2(z)}{\partial z_j},$$

where z_j is the jth component of z, $j = 1, \ldots, p + q - 1$.

When individual efficiencies are wanted for data points in the original sample, the problem is more complicated since these data points are perturbated by noise. This is a well-known identification problem in stochastic frontier models but we can use, as suggested by Kumbhakar et $al.$ (2007) the Jondrow et $al.$ (1982) procedure to predict the individual efficiency score of a particular data point from the predicted convoluted error $\varepsilon = \log \omega - r(\eta, x)$. The procedure is as follows: again denoting (η, x) by z, it can be shown that

$$\left(u \mid \varepsilon, z \right) \sim \mathcal{N}^+(\mu^*(z), \sigma^{2*}(z))$$

$i.e.,$ a truncated (positive values) normal, where

$$\begin{aligned}
\mu^*(z) &= \frac{-\varepsilon \, \sigma_u^2(z)}{\sigma_u^2(z) + \sigma_v^2(z)} \\
\sigma^{2*}(z) &= \frac{\sigma_u^2(z) \, \sigma_v^2(z)}{\sigma_u^2(z) + \sigma_v^2(z)}.
\end{aligned}$$

In particular, we can compute

$$E(u \mid \varepsilon, z) = \frac{\sigma_u(z) \, \sigma_v(z)}{\sqrt{\sigma_u^2(z) + \sigma_v^2(z)}} \left[\frac{\varphi(-\nu(\varepsilon, z))}{\Phi(-\nu(\varepsilon, z))} - \nu(\varepsilon, z) \right], \qquad (2.8)$$

where $\nu(\varepsilon, z) = \dfrac{\varepsilon \sigma_u(z)}{\sigma_v(z) \sqrt{\sigma_u^2(z) + \sigma_v^2(z)}}$, and φ and Φ denote the standard normal pdf and cdf.

Now, as in Jondrow et $al.$ (1982), a point predictor of the individual efficiency score for a data point (ω_i, η_i, X_i) generated by the DGP, can be obtained by replacing in (2.8) the unknown functions σ_u^2 and σ_v^2 by their local maximum likelihood estimates $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_v^2$ and ε by $\widehat{\varepsilon}_i = \log \omega_i - \widehat{r}(\eta_i, X_i)$. It is defined as

$$\widehat{u}(\eta_i, X_i) = \frac{\widehat{\sigma}_u(\eta_i, X_i) \, \widehat{\sigma}_v(\eta_i, X_i)}{\sqrt{\widehat{\sigma}_u^2(\eta_i, X_i) + \widehat{\sigma}_v^2(\eta_i, X_i)}} \left[\frac{\varphi(-\widehat{\nu}(\widehat{\varepsilon}_i, \eta_i, X_i))}{\Phi(-\widehat{\nu}(\widehat{\varepsilon}_i, \eta_i, X_i))} - \widehat{\nu}(\widehat{\varepsilon}_i, \eta_i, X_i) \right]. \qquad (2.9)$$

We know of course (see Simar and Wilson, 2007), as in the full parametric setup, that this procedure provides point predictor with uncertain statistical properties since they are based on the observation of one residual $\widehat{\varepsilon}_i$. It turns out in our simulated example below that the method performs reasonably well in our setup.

The predicted Farrell-Debreu efficiency score for this point could be given by

$$\widehat{\lambda}(X_i, Y_i) = e^{\widehat{u}(\eta_i, X_i)} \geq 1. \tag{2.10}$$

The reference (efficient) point for the unit (X_i, Y_i) in rectangular coordinates will be estimated as $(X_i, \widehat{\lambda}(X_i, Y_i) Y_i)$.

2.5 Stochastic version of DEA/FDH estimators

Thus, the above method provides, for any (x, y), an estimate of the frontier in the output direction. Of course this estimator will not show the usual properties of DEA/FDH estimators (free disposability and/or convexity). Our estimator $\widehat{r}(\eta, x)$ is consitent even under these additional assumptions, but if such properties are desired, we suggest to modify our estimator by using the appropriate FDH or DEA program on the projections of a selected grid of points (x_k, y_k), $k = 1, \ldots K$, on the estimated frontier where K could be large enough to avoid introducing additional noise in this second step estimation (FDH or DEA)[2]. The resulting estimated frontier is the free disposal hull (if FDH is used) or the convex free disposal hull (if DEA is used) of our nonparametric estimator derived above.

The whole procedure provides a stochastic version of the DEA/FDH frontier. It may be summarized as follows.

[1] Transform all the grid points (x_k, y_k) into polar coordinates (ω_k, η_k, x_k).

[2] Compute for each point (x_k, y_k), the estimates $\widehat{r}(\eta_k, x_k)$ by the method described in Section 2.3, using the original sample $\mathcal{X} = \{(X_i, Y_i) | i = 1, \ldots, n\}$.

[3] Project the grid points on the estimated frontier to obtain (x_k, y_k^*), where

$$y_k^* = \frac{\exp(\widehat{r}(\eta_k, x_k))}{\omega_k} y_k \tag{2.11}$$

---

[2] A natural choice for the grid of points could be the original data ($K = n$). Indeed, the noise in the radial direction does not influence the procedure, since the interest is to have points on the estimated frontier. We could alternatively select a grid of K values in polar coordinates (ω_k, η_k, x_k), where the (η_k, x_k) would be randomnly chosen in the range of the observed (η_i, X_i). The values of ω_k are rather arbitrary and do not play any role in the procedure, so they could also be randomly chosen in the range of the observed (noisy) ω_i.

[4] For any given fixed value of interest (x, y), run a FDH or DEA program (output-oriented) with reference set $\mathcal{X}^* = (x_k, y_k^*), k = 1, \ldots, K$ to compute an efficiency estimator $\widetilde{\lambda}(x, y)$ using (1.5) for FDH and (1.6) for DEA. The stochastic DEA/FDH estimate of the efficient frontier in the output direction for any fixed point (x, y) is finally given by:

$$\widetilde{y}^{\partial}(x, y) = \widetilde{\lambda}(x, y)\, y \tag{2.12}$$

We could also compute $\widetilde{\lambda}(X_i, Y_i)$ for $i = 1, \ldots, n$, by running n DEA or FDH programs for each of the data points generated by the DGP, with the reference set $\mathcal{X}^*$. Of course, due to the presence of noise in the data some of the resulting values might be smaller than 1. Due to the lack of information on the noise structure, we are unable to identify in $\widetilde{\lambda}(X_i, Y_i) = \exp(\widetilde{u_i - v_i})$, the part which is due to noise from the part due to inefficiency. We use the $\widetilde{}$ notation to stress the fact that we are looking here to residuals with respect to the (convex) free disposal hull of our estimated nonparametric frontier. The only part we are able to predict for an individual observation is the efficiency score (2.10) defined from our local likelihood frontier estimate, namely, $\widehat{\lambda}(X_i, Y_i) = \exp(\widehat{u}_i)$.

2.6 Statistical open issues

As explained in the Appendix A, Section A.3, for all the examples below, we use local linear approximation for the functional r and $\log \sigma_v^2$ and local quadratic approximation for the functional $\log \sigma_u^2$. In the nonparametric estimation of the efficient frontier by local maximum likelihood methods, we need to localize the likelihood by using a kernel function and a bandwidth. We have chosen a global bandwidth with a multivariate gaussian kernel.

If we denote by $d = p + q - 1$ the dimension of the regressors $z = (\eta, x)$ in the non-parametric stochastic frontier model (2.6), we use the standard d-variate normal density $K(u) = 1/(2\pi)^{d/2} \exp\{-u'u/2\}$ for the kernel, and as bandwidth matrix we may choose the scaled matrix $hS^{1/2}$, where S is the empirical covariance matrix of the d components of $Z_i = (\eta_i, X_i)$. By doing so, the selection of optimal bandwidths is in terms of a univariate baseline bandwidth h. So we obtain for instance in (A.6):

$$K_h(Z_i - z) = \frac{1}{(2\pi)^{d/2}h^d|S|^{1/2}} \exp\{-\frac{1}{2h^2}(Z_i - z)'S^{-1}(Z_i - z)\}, \tag{2.13}$$

The selection of an optimal bandwidth in practice, is discussed below.

By direct application of Theorem 2.2 of Kumbhakar *et al.* (2007), we know that for local linear approximations and under regularity conditions, the optimal bandwidth that balances bias and variance of the resulting estimator of the functionals $(r, \log \sigma_v^2, \log \sigma_u^2)$ has

to be of the order $n^{-1/(p+q+3)}$. This corresponds to a rate of convergence of $n^{2/(p+q+3)}$, for the Root Mean Squared Error (RMSE) of the estimators. The Theorem also establishes the asymptotic normality of the estimators (see Kumbhakar *et al.*, 2007 for details).

We have not yet investigated the asymptotic properties of the stochastic DEA/FDH version of our estimator. We know by experience (see *e.g.* Kneip *et al.*, 2003) that this is a difficult task; this is left for future work. But by an appropriate choice of K, the DEA step (if Ψ is convex and free disposal) and the FDH step (if Ψ is only free disposal) should not deteriorate the consistency property of the local polynomial MLE of the frontier function.

In fact our stochastic version of the DEA/FDH estimators can be viewed as a two-step procedure where the first step tries to pre-whiten the cloud of points from noise by a flexible nonparametric method (this idea was already suggested in a panel data context by Kneip and Simar, 1996). From the simulated examples below, it appears that the approach seems to be very powerful. In addition, it seems that transforming the local polynomial MLE obtained in the first step to obtain a free disposal attainable set (and eventually a free disposal convex set) by the appropriate FDH program (or the DEA program, resp.) often results in substantive improvements of the resulting estimator. Still, simulated samples are not a proof and theoretical work is needed to assess definitely the properties of these attractive two-step estimators.

3 Examples

3.1 Simulated examples

We used in this section some simulated data set to illustrate the flexibility of our approach. In order to evaluate the relative performances of our estimator we compute the Mean Integrated Squared Error (MISE) between the true frontier level and the nonparametricly estimated one. This MISE is defined as follows (here again the presentation is for the output orientation):

$$MISE = \int_{\Psi} ||y^{\partial}(x,y) - \hat{y}^{\partial}(x,y)||^2 f(x,y)\, dx\, dy,$$

where $y^{\partial}(x,y)$ is the true frontier (as defined in (1.3)) and $\hat{y}^{\partial}(x,y)$ is the selected estimator of the frontier level, $f(x,y)$ is the density of (x,y) defined over Ψ and $||\cdot||$ stands for the Euclidean norm. In practice, we will estimate this quantity for a given simulated example by its empirical counterpart:

$$\widehat{MISE} = \frac{1}{n} \sum_{i=1}^{n} ||y^{\partial}(X_i, Y_i) - \hat{y}^{\partial}(X_i, Y_i)||^2. \tag{3.14}$$

In what follows we will compute this $\widehat{MISE}$ for the DEA/FDH estimator and for their stochastic versions.

For a simulated sample we compute the optimal bandwidth h_{opt} for our Stochastic-DEA/FDH estimator by chosing the value of h which minimizes $\widehat{MISE}$. For real data set, this will be done by using Leave-One-Out Least-Squares crossvalidation techniques (see practical details below)[3].

3.1.1 Bivariate Cobb-Douglas frontier

We simulate first a sample of size n according the Cobb-Douglas model:

$$Y_i = X_i^{0.5} \exp(-u_i) \exp(v_i),$$

where $X_i \sim U(0,1)$, $u_i \sim Exp\,(\mu = 1/6)$, where μ is the mean (and standard deviation) of u_i, and the noise is $v_i \sim N(0, \sigma_v^2)$ where $\sigma_v = \rho_{\text{nts}} \times \mu$. The different scenarios below correspond to various selection of n and ρ_{nts}. Table 1 summarizes the scenarii and the results, whereas Figure 1 displays the resulting pictures.

---

[3]In all the simulated examples, we provided also the plots of the various nonparametric estimators of the efficiency against the true efficiency scores. They are not reproduced to save space. As expected, they confirm the qualitative comments coming from the pictures displayed below for the estimation of the frontier levels.

Case	Description	$\widehat{MISE}_{DEA}$	$\widehat{MISE}_{Stoch-DEA}$	h_{opt}
case (a): $n = 100$	$\rho_{\text{nts}} = 0$	0.000127	0.000053	0.1
case (b): $n = 103$	3 outliers	0.0993	0.0010	0.34
case (c): $n = 100$	$\rho_{\text{nts}} = 1$	0.0501	0.0057	0.83
case (d): $n = 200$	$\rho_{\text{nts}} = 1$	0.0507	0.0028	0.42
case (e): $n = 500$	$\rho_{\text{nts}} = 1$	0.0998	0.0054	0.42
case (e): $n = 500$	$\rho_{\text{nts}} = 2$	0.7053	0.0184	1.20

Table 1: *Summary of the results for Example 1.*

It should be noted that in this first example, the anchorage model is wrong for the stochastic part of the model. The results show that the performance of the stochastic-DEA are quite good. As seen from Table 1 even if there is no noise in the DGP, our estimators behaves better than the DEA estimator. This is probably the case because the nonparametric likelihood smoothing allows to build the second stage DEA with more efficient support points (this will be confirmed in most of our simulated examples and also for the stochastic FDH). The second scenario (no noise but 3 outliers) is quite enlightening on the power of our estimator to resist to a few, but important, outliers (here, as expected, the standard DEA is very sensitive to these 3 outliers). Once the size of the noise increases, the DEA estimator behaves quite badly (in particular with big sample size and $\rho_{\text{nts}} = 2$) compared to our stochastic-DEA. Of course, since the noise (and the inefficiency) is multiplicative in the original units of X and Y, the performance of the nonparametric fit (and of the stochastic DEA) is less good on the right part of the panels. Remember that our first step nonparametric estimator of the stochastic frontier does not impose free disposability of the resulting attainable set, and so at some points has a poor fit, but the DEA estimator at the second step corrects it well.
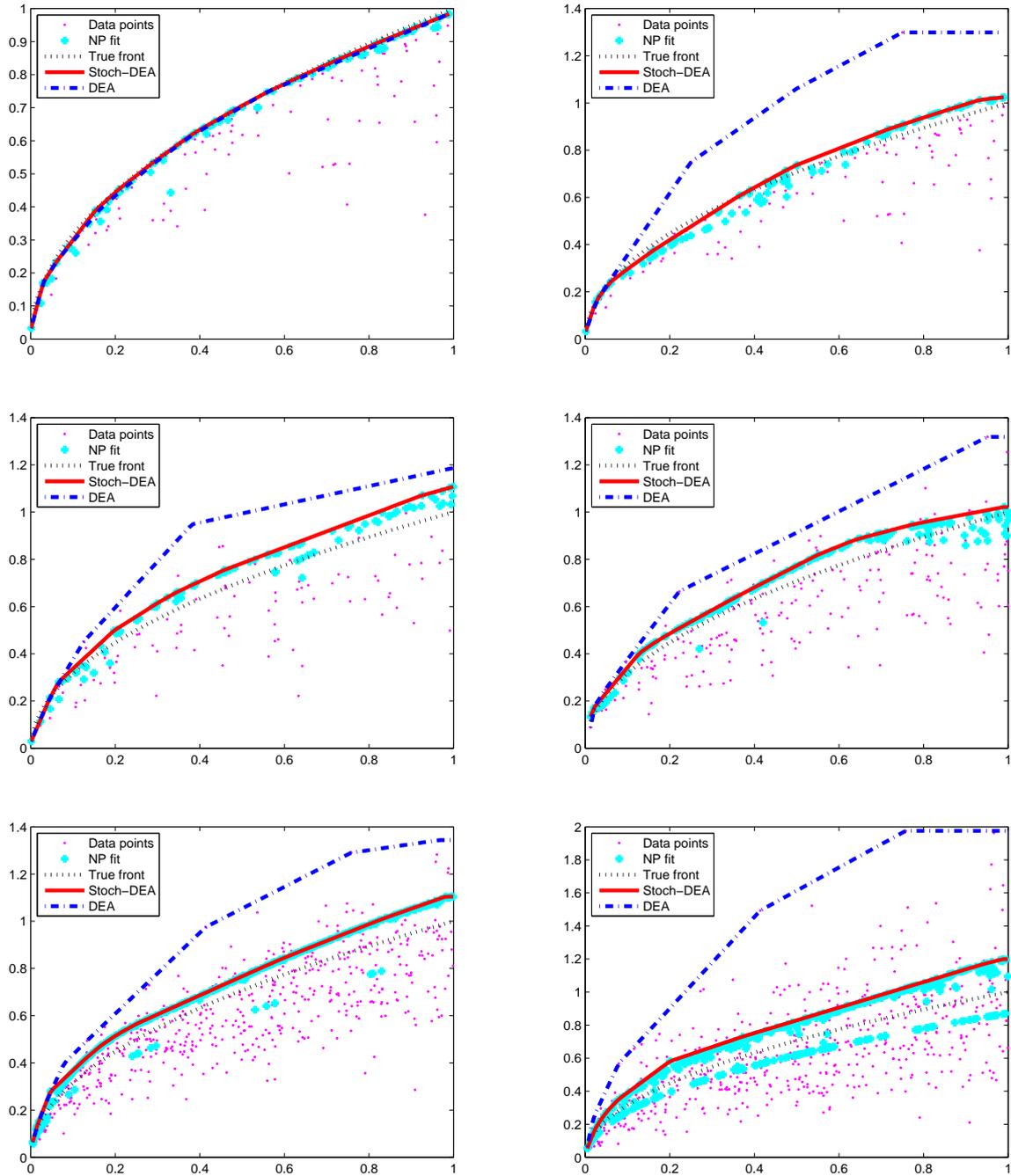
Figure 1: *Figures for Example 1, from left to right and top to bottom, we have the 6 cases described in Table 1: (a)* $n = 100, \rho_{\mathrm{nts}} = 0$, *(b)* $n = 103, \rho_{\mathrm{nts}} = 0 + 3$ *outliers, (c)* $n = 100, \rho_{\mathrm{nts}} = 1$, *(d)* $n = 200, \rho_{\mathrm{nts}} = 1$, *(e)* $n = 500, \rho_{\mathrm{nts}} = 1$, *(f)* $n = 500, \rho_{\mathrm{nts}} = 2$.

3.1.2 Bivariate Logit frontier

Here we have a nonconvex attainable set and we will compare the performance of the stochastic-FDH with the standard FDH. We generate the data sets according to the following model

$$Y_i = \frac{1 + \exp(7)}{\exp(7)} \frac{\exp(-3 + 10X_i)}{(1 + \exp(-3 + 10X_i))} \exp(-u_i) \exp(v_i),$$

where, as for the first example, $X_i \sim U(0,1)$, $u_i \sim Exp\,(\mu = 1/6)$, where μ is the mean (and standard deviation) of u_i, and the noise is $v_i \sim N(0, \sigma_v^2)$ where $\sigma_v = \rho_{\text{nts}} \times \mu$. The frontier function has been scaled so that $0 \le Y_i \le 1$ when $0 \le X_i \le 1$.

Note that here the anchorage model is wrong for both the frontier function and the stochastic part of the model. The results are presented in Table 2 and Figure 2. We could expect here less encouraging results since the anchorage model is entirely wrong, however, we observe that the results are qualitatively very similar than those of the first example above, showing very good performance of our estimator even when there is no noise, and greater ability than the FDH to resist to outliers and to the size of the noise. Most of the comments made for example 1 could be repeated here. In particular, we see that the first step local linear MLE of the frontier function does not provide a free disposal attainable set (monotone frontier), but the FDH correction of the second step does a good job.

Case	Description	$\widehat{MISE}_{FDH}$	$\widehat{MISE}_{Stoch-FDH}$	h_{opt}
case (a): $n = 100$	$\rho_{\text{nts}} = 0$	0.0018	0.0007	0.09
case (b): $n = 103$	3 outliers	0.0820	0.0003	0.21
case (c): $n = 100$	$\rho_{\text{nts}} = 0.5$	0.0089	0.0019	0.30
case (d): $n = 100$	$\rho_{\text{nts}} = 1$	0.0770	0.0011	0.47
case (e): $n = 200$	$\rho_{\text{nts}} = 1$	0.1018	0.0061	0.23
case (f): $n = 500$	$\rho_{\text{nts}} = 1$	0.1021	0.0023	0.20

Table 2: *Summary of the results for Example 2.*
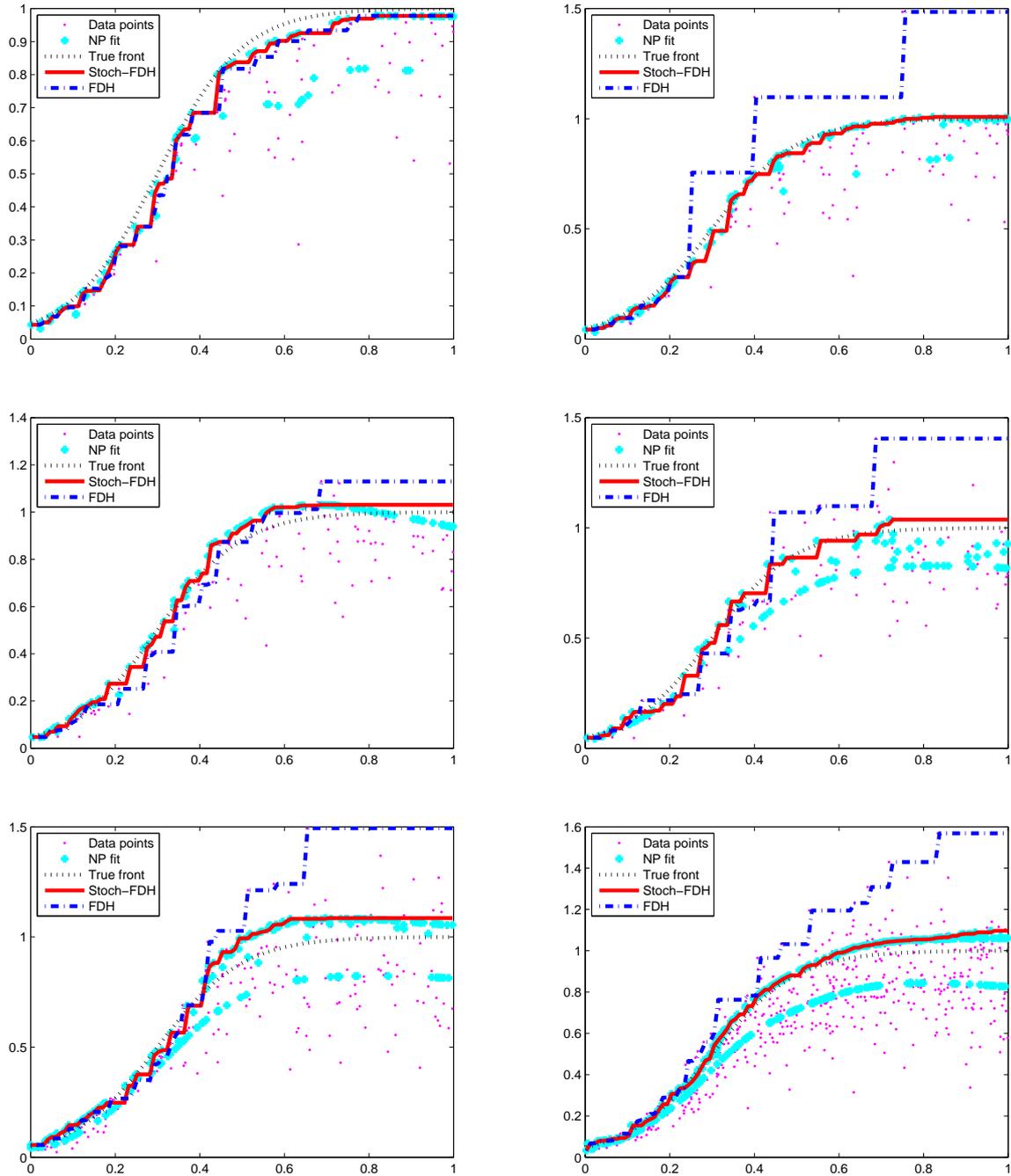
Figure 2: *Figures for Example 2, from left to right and top to bottom, we have the 6 cases described in Table 2: (a) $n = 100, \rho_{\text{nts}} = 0$, (b) $n = 103, \rho_{\text{nts}} = 0 + 3$ outliers, (c) $n = 100, \rho_{\text{nts}} = 0.5$, (d) $n = 100, \rho_{\text{nts}} = 1$, (e) $n = 200, \rho_{\text{nts}} = 1$, (f) $n = 500, \rho_{\text{nts}} = 1$.*

3.1.3 Multivariate Log-Linear frontier

For the multivariate case we have $p = 2$ inputs and $q = 2$ outputs, we adapt the scenario of Kneip, Simar and Wilson (2003) to an output oriented case. We generate input levels X_1, X_2 from the uniform distribution on $[10, 20]$. Next, we computed efficient output levels by generating ζ uniform on $\left[\frac{1}{9}\frac{\pi}{2}, \frac{8}{9}\frac{\pi}{2}\right]$ and setting $Y_1^{\partial} = X_1^{0.4} X_2^{0.4} \times \cos(\zeta)$ and $Y_2^{\partial} = X_1^{0.4} X_2^{0.4} \times \sin(\zeta)$. We then generated the observed output levels by setting $Y_1 = Y_1^{\partial} e^{-u} e^{v}$ and $Y_2 = Y_2^{\partial} e^{-u} e^{v}$ and where $u \sim 0.2 |N(0, \sigma_u^2)|$, with $\sigma_u = 0.5$ and $v \sim N(0, \sigma_v^2)$ where $\sigma_v = \rho_{\text{nts}} \times \sqrt{(\pi - 2)/\pi}\, \sigma_u$. We note that here the anchorage model is not the true model (remember we use the local likelihood method in the polar coordinates, which makes a difference when $p + q > 2$). The results are displayed in Table 3. We observe again that our stochastic DEA behaves very well even if there is no noise and of course, as expected, as the size of the noise increases our estimator resists much better than the DEA. This is particularly true when the sample size increases (see the cases with $n = 500$). Of course, with $p = q = 2$ and $n = 100$, it is not a big sample especially when the noise has the same size as the inefficiency term.

Case	Description	$\widehat{MISE}_{DEA}$	$\widehat{MISE}_{Stoch-DEA}$	h_{opt}
case (a): $n = 100$	$\rho_{\text{nts}} = 0.00$	0.7431	0.0958	1.00
case (b): $n = 100$	$\rho_{\text{nts}} = 0.50$	0.7033	0.1306	3.20
case (c): $n = 100$	$\rho_{\text{nts}} = 1.00$	5.4602	0.2574	3.25
case (d): $n = 200$	$\rho_{\text{nts}} = 1.00$	6.9569	0.2366	15
case (e): $n = 500$	$\rho_{\text{nts}} = 1.00$	9.5637	0.0546	2.40
case (f): $n = 500$	$\rho_{\text{nts}} = 2.00$	160.0901	0.3464	11

Table 3: *Summary of the results for Example 3.*

It is of course difficult to see a cloud a point in 4 dimensions, but to give an idea on the quality of the fit we display, in Figure 3 for each of the 6 cases, the plots of the modulus of the true and of the estimated frontier against the modulus of the inputs values for all the observations in the sample. It provides only a partial picture but gives an idea of the difference between the quality of the results for the DEA and the stochastic DEA in each case. These figures confirm the qualitative comments given above.
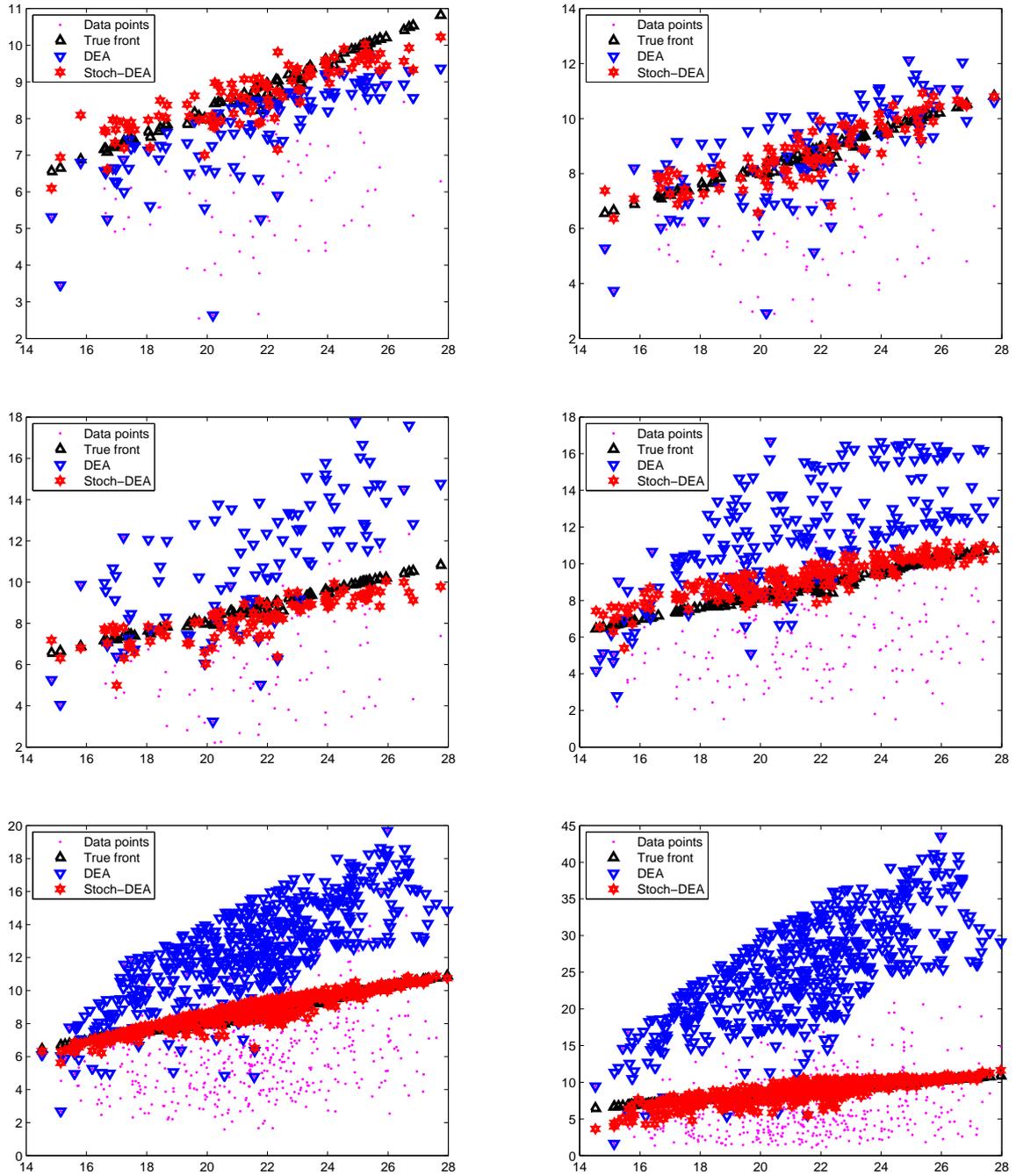
Figure 3: *Pictures for Example 3, from top to bottom, we have the 6 cases described in Table 3: (a) $n = 100, \rho_{\mathrm{nts}} = 0$, (b) $n = 100, \rho_{\mathrm{nts}} = 0.5$, (c) $n = 100, \rho_{\mathrm{nts}} = 1$, (d) $n = 200, \rho_{\mathrm{nts}} = 1$. (e) $n = 500, \rho_{\mathrm{nts}} = 1$, (d) $n = 500, \rho_{\mathrm{nts}} = 2$.*

3.1.4 Multivariate Quadratic frontier with heteroskedastic inefficiencies

In this scenario, the anchorage model (log-linear) is wrong for both the frontier model and the variance terms. We select the folowing model for the frontier levels: $Y_1^{\partial} = 10 + 20X_1 + 20X_2 - (X_1^2 + X_2^2 + 2X_1X_2) \times \cos(\zeta)$ and $Y_2^{\partial} = 10 + 20X_1 + 20X_2 - (X_1^2 + X_2^2 + 2X_1X_2) \times \sin(\zeta)$, where X_1, X_2 are generated from the uniform distribution on $[2, 4]$ and ζ is as in the preceding example. We then generated the observed output levels by setting $Y_1 = Y_1^{\partial} e^{-u(x)} e^v$ and $Y_2 = Y_2^{\partial} e^{-u(x)} e^v$ and where $u|X = x \sim |N(0, (\sigma_u \times (x_1 + x_2)/6)^2)|$, with $\sigma_u = 0.3$. This corresponds to a variation of the standard deviation of U by a factor 2 over the range of X. Finally we select an homoskedastic noise with $v \sim N(0, \sigma_v^2)$ where $\sigma_v = \rho_{\text{nts}} \times \sqrt{(\pi - 2)/\pi}\, \sigma_u$, so that ρ_{nts} measure an average noise to signal ratio. We obtain the following results displayed in Table 4 and Figure 4.

Case	Description	$\widehat{MISE}_{DEA}$	$\widehat{MISE}_{Stoch-DEA}$	h_{opt}
case (a): $n = 100$	$\rho_{\text{nts}} = 0.00$	26.0628	6.5757	1.05
case (b): $n = 100$	$\rho_{\text{nts}} = 0.50$	21.6157	7.8178	1.35
case (c): $n = 100$	$\rho_{\text{nts}} = 1.00$	138.4493	27.1243	2.00
case (d): $n = 200$	$\rho_{\text{nts}} = 0.50$	22.8037	3.3958	1.90
case (e): $n = 200$	$\rho_{\text{nts}} = 1.00$	163.3461	13.9084	3.10
case (f): $n = 500$	$\rho_{\text{nts}} = 1.00$	392.4182	8.0690	4.00

Table 4: *Summary of the results for Example 4.*

Globally we have the same comments than in the preceding multivariate case, which is rather positive since in the scenario here, the anchorage model is far from the true model and we have in addition heteroskedasticity of the inefficiency term. Note that the higher values of the MISE is largely due to the different scaling of Y^{∂} in this example compared with the preceding one. The behavior of the MISE for both estimators as a function of the sample size and the noise is exactly the one we expected. Clearly in this 4-dimensional setup, $n = 100$ is a small sample size when $\rho_{\text{nts}} = 1$.
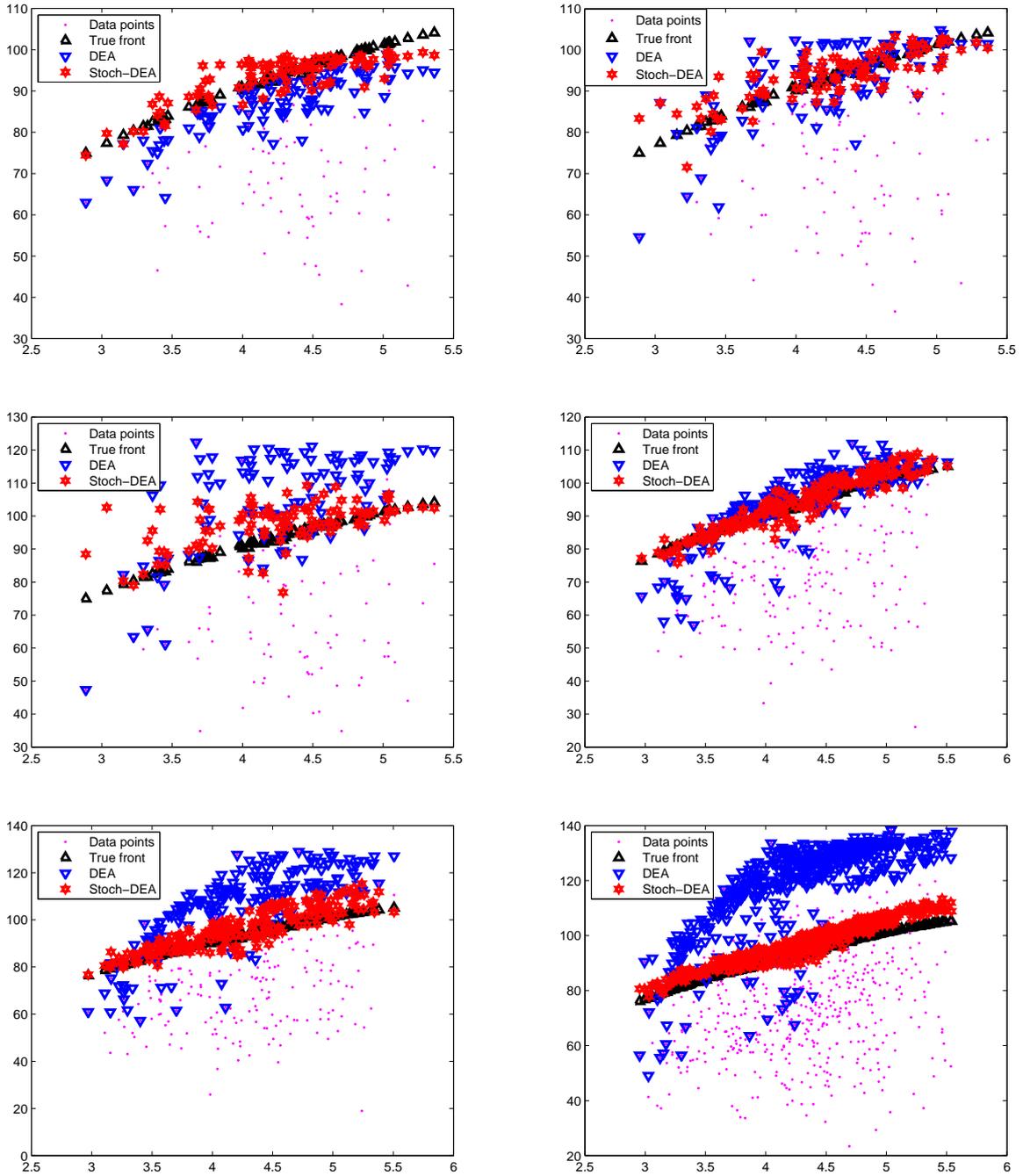
Figure 4: *Pictures for Example 4, from top to bottom, we have the 6 cases described in Table 4: (a)* $n = 100, \rho_{\mathrm{nts}} = 0$, *(b)* $n = 100, \rho_{\mathrm{nts}} = 0.5$ *, (c)* $n = 100, \rho_{\mathrm{nts}} = 1$, *(d)* $n = 200, \rho_{\mathrm{nts}} = 0.5$. *(e)* $n = 200, \rho_{\mathrm{nts}} = 1$, *(d)* $n = 500, \rho_{\mathrm{nts}} = 1$.

3.2 Real data examples

3.2.1 Bandwidth selection

With real data, we select the optimal badwith for the nonparametric frontier estimation by local maximum likelihood method by using, as in Kumbhakar *et al.* (2007), a leave-one-out least squares crossvalidation procedure. We compute for a grid of values of h

$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} [\log \omega_i - (\widehat{r}_0^{(i)}(\eta_i, X_i) - \widehat{u}_i^{(i)})]^2 \tag{3.15}$$

where $\widehat{r}_0^{(i)}$ and $\widehat{u}^{(i)}$ are the leave-one-out version of the local linear estimators derived above, by using the bandwidth h.

When the sample size is large, to avoid the computational burden, we limit the estimation of $CV(h)$ by taking the mean on the rhs of (3.15) over a random subset of m points from the original sample, where $m \ll n$.

3.2.2 Schools example

We can go back to the classical example of the 70 observations in the Program Follow Through (PFT) data used by Charnes, Cooper and Rhodes (1981) to analyze an experimental education program administered in US schools. Observations 1–49 correspond to the PFT schools, while observations 50–70 correspond to the remaining 21 Non-PFT schools.

Each observation consists of five input variables ($p = 5$) describing characteristics of the school and three output variables ($q = 3$) obtained from results of standardized tests in this school. It is clear that we cannot obtain sensible results with so few observations in a 8-dimensional input-output space. It turns out that doing simple data analysis, the 5 inputs are higly correlated and the same is true for the 3 outputs. So, by using factorial methods described for this setup in Daraio and Simar (2007), we can reduce the dimension of the problem to two dimensions: one input factor (FI) being the best linear combination of the scaled inputs and one output factor (FO) being the best linear combination of the scaled outputs (these factors are obtained by the eigenvectors of the second moment matrices of the two groups of variables, see Daraio and Simar, 2007, for details).

It turns out that the input factor FI (that summarizes 95 % of the total inertia of the 5 inputs) has correlations (0.89, 0.98, 0.96, 0.97, 0.74) with the 5 original inputs. It can be interpreted (looking to the eigenvector) as an average of the standardized inputs. The output factor FO (that summarizes 99 % of the total inertia of the 3 outputs) has correlations (0.99, 0.99, 0.98) with the 3 original outputs. Here again this factor is interpreted as an average of the standardized outputs. So in this example, we do not lose too much information

by looking to the pictures in two dimensions: one input and one output. The results are shown in Table 5: we have chosen an output orientation and the less restrictive variable returns to scale (VRS) case. The column headed "NonPar Eff" gives the values of the individual efficiency scores computed by (2.10) from local likelihood estimate of the frontier. The last row of the Table gives the averages over the 70 scores.

The Stochastic-DEA and the DEA (and their FDH versions, not imposing convexity) are displayed in Figure 5. Looking to these results, it seems that the Stochastic-FDH provides the best fit (see the right panel of Figure 5). Imposing convexity in the second step seems to be inappropriate in this example. This is confirmed when we remark that the average of the stochastic FDH efficiency scores is very similar to the average of the nonparametric effciencies.

The approach provides also very informative complementary tools. Figure 6 displays the boxplots of the 70 local estimates of the variance functions coming from the local MLE procedure. It turns out that there is not so much variation among the 70 units and that the noise is of the same order (or smaller) than the inefficiency variances. Figure 7 allows to appreciate the effect of the input factor on the output and on the inefficiency σ_u^2. This confirms a positive slope (rather stable across units) of the frontier and very few influence of the input on the efficiency (most of the local derivatives are equal or not far from zero, see the right panel of Figure 7).
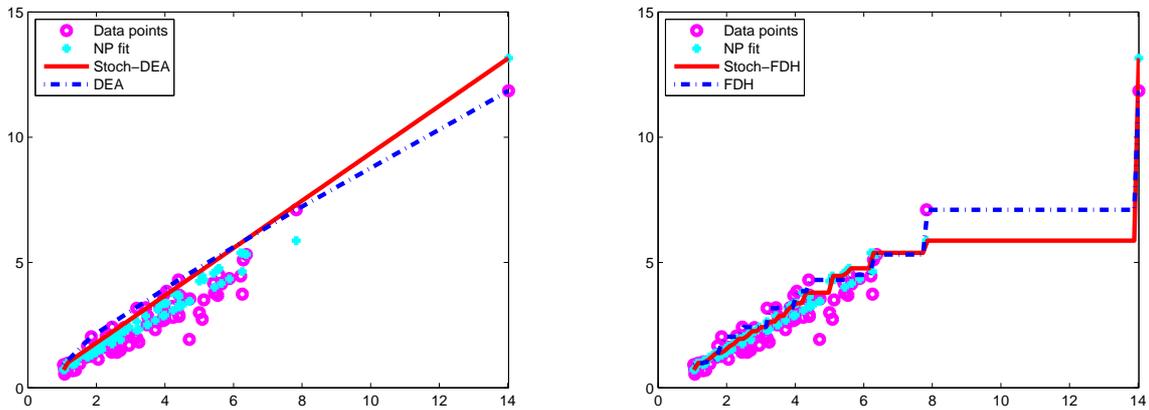
Figure 5: *Nonparametric estimates of the frontier for the School data, output factor FO vs input factor FI: DEA on the left panel and FDH on the right panel.*
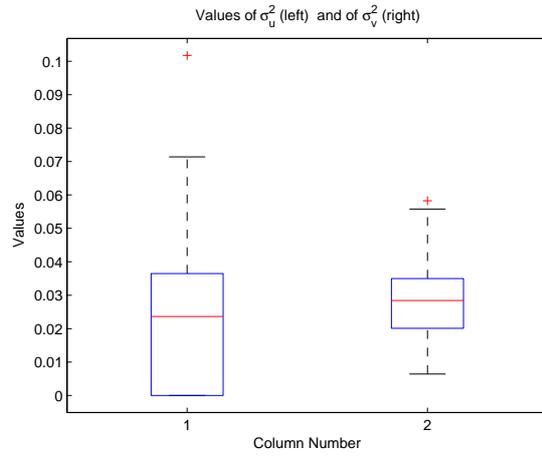
Figure 6: *Boxplot of the 70 estimated variance functions σ_u^2 (left) and σ_v^2 (right).*
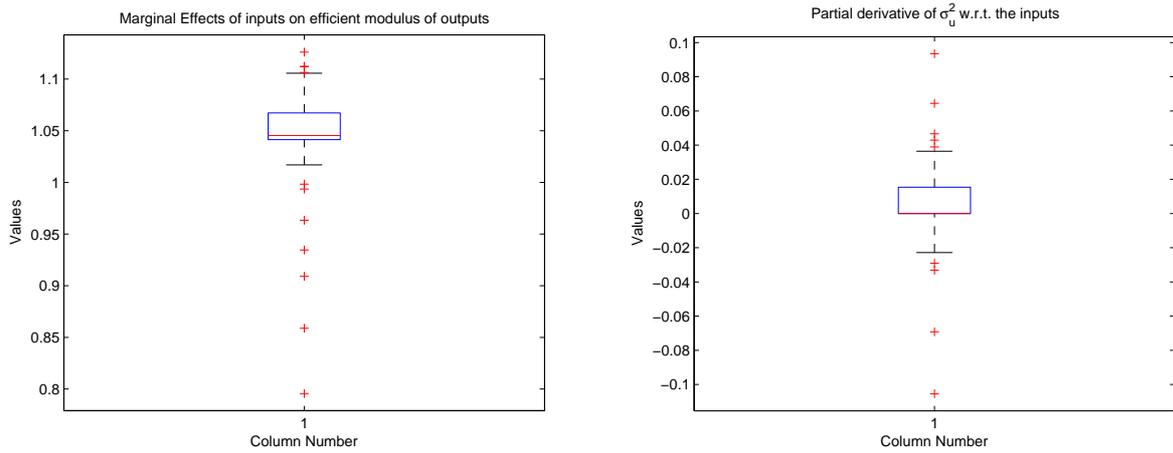
Figure 7: *Local effects of the input factor FI on the frontier level of the output factor FO (left panel) and on the inefficiency level (right panel).*

Units i	NonPar Eff	$\hat{\lambda}_{FDH}$	$\tilde{\lambda}_{FDH}$	$\hat{\lambda}_{DEA}$	$\tilde{\lambda}_{DEA}$
1	1.1248	1.0000	1.0551	1.1448	1.1447
2	1.1452	1.2679	1.1577	1.3560	1.2634
3	1.0000	1.2025	1.0594	1.2286	1.1638
4	1.2458	1.4285	1.4289	1.7795	1.5882
5	1.3337	1.3426	1.4414	1.5571	1.4414
6	1.0000	1.4054	1.0336	1.4845	1.2243
7	1.0000	1.7417	1.4475	1.7957	1.6262
8	1.2835	1.4374	1.4206	1.6035	1.5477
9	1.1889	1.5102	1.3306	1.5148	1.4289
10	1.1042	1.0000	1.0575	1.1682	1.1475
11	1.0000	1.1663	1.0276	1.1725	1.1067
12	1.0663	1.0000	0.9117	1.0634	0.9899
13	1.0088	1.4764	1.3007	1.4789	1.3947
14	1.1442	1.0663	1.1229	1.3711	1.2046
15	1.1107	1.2615	1.0309	1.3698	1.1386
16	1.0000	1.1390	1.2082	1.3659	1.3368
17	1.0000	1.0000	0.8317	1.1268	0.9257
18	1.0000	1.1163	1.0041	1.2235	1.1204
19	1.1143	1.0293	1.0918	1.2286	1.2015
20	1.0000	1.0000	0.8311	1.0146	0.9162
21	1.0613	1.0000	0.8764	1.0376	0.9687
22	1.0953	1.0000	0.9471	1.1745	0.9998
23	1.0001	1.0338	1.0932	1.2651	1.2529
24	1.0000	1.0000	0.8066	1.0636	0.9187
25	1.0000	1.2720	1.1441	1.3888	1.2709
26	1.0000	1.2130	1.0687	1.2884	1.2310
27	1.0746	1.0000	0.8995	1.0803	0.9869
28	1.1251	1.1596	1.0820	1.3394	1.1778
29	1.0303	1.4440	1.4503	1.8926	1.6738
30	1.1902	1.6123	1.3518	1.7553	1.5242
31	1.0000	1.6220	1.3918	1.8683	1.6420
32	1.0000	1.6957	1.3575	1.7731	1.4612
33	1.1814	1.0102	1.2075	1.2966	1.2943
34	1.1679	1.2273	1.2738	1.3991	1.3562
35	1.0000	1.0888	1.1514	1.2862	1.2660
36	1.2027	1.7115	1.4687	1.9105	1.6680
37	1.1239	1.3977	1.1496	1.4997	1.2979
38	1.1387	1.0000	1.0043	1.1155	1.0253
39	1.1623	1.1845	1.1667	1.2748	1.1959
40	1.1054	1.1023	1.0630	1.3485	1.1994
41	1.0000	1.1945	1.2224	1.4584	1.3573
42	1.0000	1.1208	0.9618	1.2890	1.1324
43	1.0000	1.3690	1.3485	1.4796	1.3888
44	1.0000	1.0000	0.8269	1.0000	1.0282
45	1.0000	1.0000	0.9567	1.2648	1.1153
46	1.0000	1.2063	1.4419	1.5583	1.5572
47	1.0977	1.0000	1.0234	1.1933	1.1069
48	1.1733	1.0775	1.2809	1.6362	1.3892
49	1.1911	1.5659	1.3129	1.6846	1.4589

Units i	NonPar Eff	$\hat{\lambda}_{FDH}$	$\tilde{\lambda}_{FDH}$	$\hat{\lambda}_{DEA}$	$\tilde{\lambda}_{DEA}$
50	1.2133	1.1681	1.2784	1.4212	1.3947
51	1.2148	1.7918	1.4642	1.9483	1.6200
52	1.0631	1.0000	0.8810	1.0000	0.9427
53	1.2080	1.2195	1.3276	1.6203	1.4631
54	1.0880	1.0000	1.0132	1.1130	1.1151
55	1.0783	1.0087	0.9412	1.2026	1.0643
56	1.1150	1.2640	1.0914	1.4555	1.2288
57	1.1612	1.2211	1.1959	1.4393	1.3325
58	1.0000	1.0000	0.7354	1.0000	0.8115
59	1.1266	1.0000	1.1099	1.0000	1.1099
60	1.0000	1.2745	1.0414	1.4453	1.2153
61	1.0000	1.4643	1.5420	1.9820	1.7223
62	1.0000	1.0000	0.7769	1.0000	0.7769
63	1.1493	1.3347	1.1947	1.6232	1.3174
64	1.0000	1.3775	1.2390	1.5959	1.4738
65	1.1616	1.0000	1.1591	1.4883	1.2339
66	1.0000	2.2310	1.9656	2.3636	2.2569
67	1.5580	1.5758	1.6355	1.7817	1.7243
68	1.0000	1.3487	1.2316	1.4476	1.3494
69	1.0000	1.2538	1.0000	1.3037	1.0669
70	1.1693	1.4548	1.3779	1.8076	1.5595
mean	1.0900	1.2413	1.1618	1.4101	1.2805

Table 5: *Results for the Charnes et al., (1981) School data.*

3.2.3 Banks example

We illustrate finally our tool with the Bank data set[4] used by Simar and Wilson (2007). They are data on 322 commercial banks operating in the U.S., from the fourth-quarter FDIC Reports of Income and Condition (Call Reports) for 2002. The choice of the inputs and outputs is motivated in Simar and Wilson, with references herein: we have here 3 inputs (purchased funds, core deposits, and labor) and 4 outputs (consumer loans, business loans, real estate loans, and securities held) for banks. We estimate output efficiencies with VRS assumption.

A part of the results are given for illustration in Table 6 and in Figure 8, where the 322 banks and the DEA and the stochastic DEA and the nonparametric fit (the local MLE fit) are displayed on the scale of the modulus of the outputs versus the modulus of the inputs. The right panel of the figure is a zoom of the smallest units. Of course, we will not go into the details of the analysis here but the following remarks can be given. First, we do not display the FDH pictures to save place because they are very similar. Second, it should be noticed that most of the FDH scores are equal to one, but when we introduce noise, the

---

[4]We would like to thank Paul W. Wilson who provided us this data set.

FDH estimates of the distances to the frontier becomes different form one. Some of them, when much lower than one, could warn for potential outliers (like units #5, #316 or #322 in Table 6).
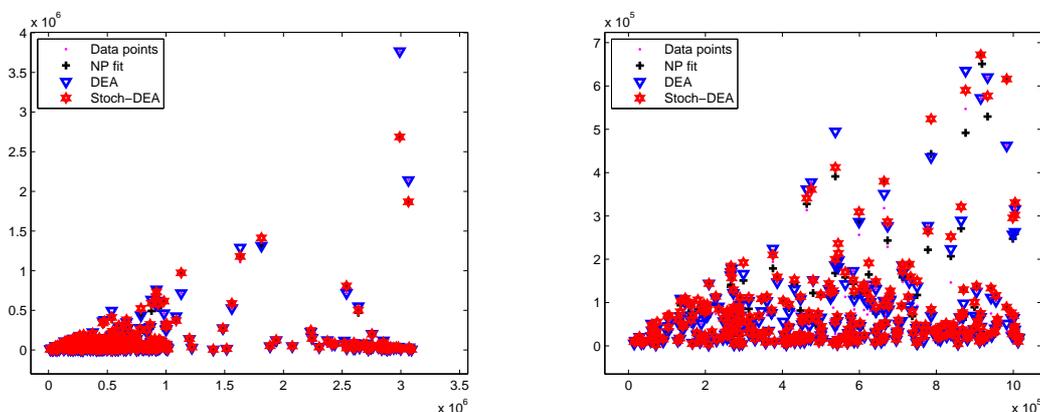
Figure 8: *DEA and Stochatic DEA in units of modulus of outputs versus modulus of the inputs. The right panel is a zoom on the smallest units.*

More interesting is the picture of the variations of the variance functions over the 322 local estimates. This appears in Figure 9. It seems that the size of the noise is rather small with respect to the variance of the inefficiency term, but also that there is a lot of heteroskedasticity in $\sigma_u^2(x)$.

By looking to Figure 10 (right panel) we see that the first and the third inputs (purchase funds and labor) have some positive influence on the efficiency level (negative partial derivatives of $\sigma_u^2(x)$) but that on the average, the second input (core deposit) is less influential on the efficiency levels. As expected, the three inputs have a positive effect on the frontier level (measured by its modulus).

Of course these rough comments should be complemented by a careful analysis from experts of the banking activity. The idea here was to illustrate the kind of tools that can come out from our approach and how they can be useful to enlighten the efficiency analysis of a sector of activity.
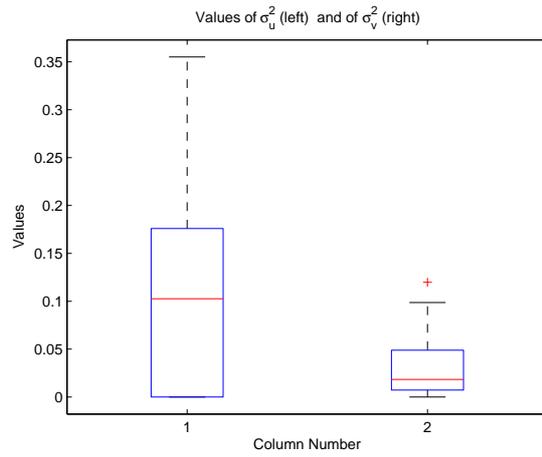
Figure 9: *Boxplot of the 322 estimated variance functions σ_u^2 (left) and σ_v^2 (right).*
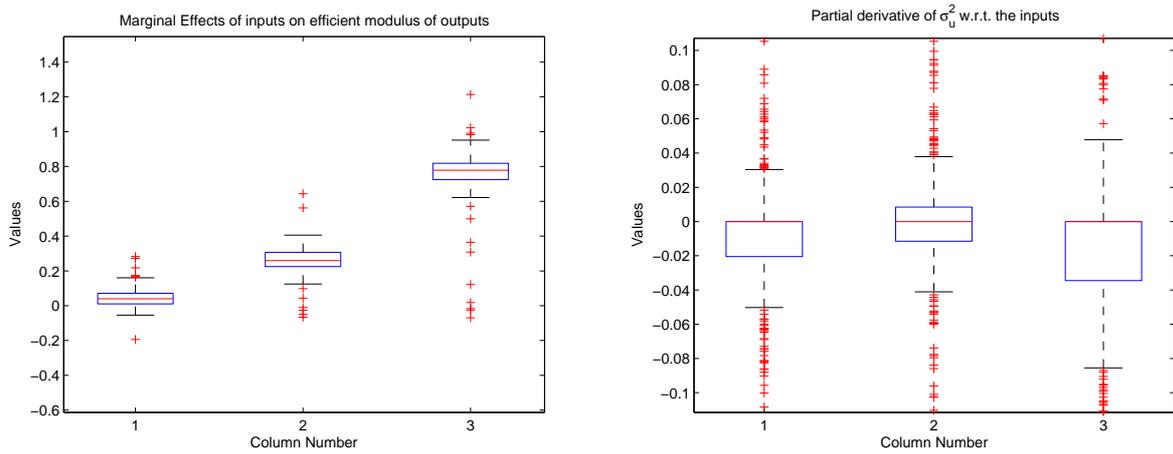
Figure 10: *Local effects of the 3 inputs on the frontier level (left panel) and on the inefficiency level (right panel): input 1 is purchase funds, input 2 is core deposit and input 3 is labor.*

27

Units i	NonPar Eff	$\tilde{\lambda}_{FDH}$	$\tilde{\lambda}_{FDH}$	$\tilde{\lambda}_{DEA}$	$\tilde{\lambda}_{DEA}$
1	1.2572	1.0000	1.2603	1.2039	1.3565
2	1.4241	1.0000	1.4791	1.3292	1.4791
3	1.1261	1.0000	1.0713	1.0145	1.1310
4	1.7023	1.0000	1.7670	1.1005	1.7670
5	1.0000	1.0000	0.8038	1.0000	0.9992
6	1.0000	1.0000	1.8138	1.4178	1.8830
7	1.7997	1.0000	1.8172	1.5304	1.8180
8	1.5359	1.0000	1.5790	1.3558	1.5790
9	1.6474	1.0000	1.7151	1.6116	1.7186
10	2.7640	1.0000	2.7894	1.6459	2.7894
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
312	1.0972	1.0000	1.0646	1.0093	1.1731
313	1.2129	1.0000	1.2224	1.0218	1.2224
314	1.3232	1.0000	1.4084	1.1334	1.4084
315	1.5488	1.0000	1.5633	1.4061	1.5633
316	1.0000	1.0000	0.8626	1.1392	1.0097
317	1.2004	1.0000	1.1886	1.1614	1.3275
318	1.0000	1.0000	1.0398	1.0462	1.0398
319	1.2228	1.0000	1.2291	1.1530	1.2972
320	1.0000	1.0000	1.0000	1.2651	1.0528
321	1.0000	1.0000	1.0022	1.0000	1.1926
322	1.0000	1.0000	0.7149	1.0000	1.0359
mean	1.2435	1.0069	1.2629	1.1906	1.3689

Table 6: *Results for some units from the Banks data set.*

4 Conclusions

This paper provides one promising way to handle noise in a multivariate nonparametric stochastic frontier setup. The examples above with simulated data indicate it is very flexible and robust. The results are reasonable even when the chosen anchorage model is wrong. The appraoch also allows the researcher to shed lights on the effect of the inputs on the frontier level but also on the level of the efficiency.

The procedure can handle multi-inputs and multi-outputs situations and can be easely adapted to both input and output orientations.

The numerical burden is linked to the bandwidth selection (leave-one-out least-squarres cross validation). For example for the bank data set (322 units and an 8-dimensional input/output space), the optimal bandwidth was selected on a grid search base, evaluating the criterion on only 100 randomnly chosen data points. As far as the sample size is large (as in this example) the procedure is rather stable, but it took a couple of hours on a 3Gherz machine, using a Matlab code. The numerical optimization (maximization of the local likelihood) can be unstable when using second order methods (gradient, etc.). Finally we preferred to make all our computations with the slower, but much more robust, Nelder-simplex method.

Some of the statistical issues are still open. The statistical theory of the local polynomial

MLE in this frontier setup has been done (see Kumbhakar *et al.*, 2007). We hope to able in future work to derive the full properties of our stochastic versions of the DEA/FDH estimators, but we know this is a serious challenge for further research.

Appendix

A   Local Polynomial Estimation

This appendix summarizes the main aspects of local maximum likelihood estimation using local polynomial approximation for the functional parameters of a model. The presentation of these techniques in a multiple regression framework with high order polynomials can be notationally very complex. In order to simplify the notation, in this appendix, we will use a general notation, independent of the preceding one, but we will give the correspondance when necessary.

We consider indeed a general nonparametric stochastic frontier model which can be viewed as a nonparametric regression of a univariate dependent variable Y (in Section 2, $Y \equiv \log \omega$) on a d-dimensional variable Z (in Section 2, we had $Z \equiv (\eta, X)$ with $d = p+q-1$), where locally (for $Z = z$), the error term is the convolution of a positive random variable u (inefficiency) with a known distribution and a real random variable v having also a known distribution with zero mean, the parameters of these distributions being unknown function of z. In its most general version the model could be presented as follows.

We observe a set of i.i.d. random variables (Z_i, Y_i) for $i = 1, \ldots, n$ with $Z_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$, where

$$Y_i = f(Z_i) - u_i + v_i \tag{A.1}$$

for some unknown function f, where, conditionally on $Z = z$, u and v are independent and $\varepsilon = -u + v$ has a known conditional continuous distribution $G(\cdot, \tau(z))$, where τ is some k-valued unknown function (in our setup τ wil be defined as the pair of functions $(\sigma_u^2, \sigma_v^2)^T$, but more general structure could be analyzed, see below). The unknown functions f and τ are called the functional parameters of the model.

For the ease of the notation, we first introduce the simple case of constant parameters for the convolution parameters and then generalize for more general models allowing for heteroskedasticity. Then we will customize the likelihood function for the model used in our simulations and in our real data examples.

A.1 Convolution with Constant Parameters

In this model, conditionally on $Z = z$, $\varepsilon = v - u$ has a continuous distribution $G(\cdot, \tau)$, where τ represents the unknown parameters of the convolution, that are assumed to be constant. In other words, the conditional density of Y given $Z = z$ equals

$$\phi(y, f(z), \tau) \equiv g_\varepsilon(y - f(z), \tau),$$

where $g_\varepsilon(\varepsilon, \tau) = \partial G(\varepsilon, \tau)/\partial \varepsilon$. Our main interest here is estimation of the function f and eventually its derivatives. We now describe the local polynomial estimation of f in a general setting of multivariate Z.

Define $\ell = \log \phi$. Then, the conditional log-likelihood equals $\sum_{i=1}^{n} \ell(Y_i, f(Z_i), \tau)$. Let z be a point at which one wants to estimate the values of the function f and its derivatives. A local conditional log-likelihood is obtained by replacing f in the conditional log-likelihood by its mth order polynomial approximation in a neighborhood of z and putting the weight $K_h(Z_i - z)$ for each observation (Z_i, Y_i), where $K_h(u) = h^{-d}K(h^{-1}u)$, K is a d-variate kernel function, typically a symmetric density function defined on $\mathbb{R}^d$, and h is a positive scalar, called the bandwidth. Precisely, it is given by

$$
\begin{aligned}
L_n(\theta_0, &\theta_1, \ldots, \theta_{r(m)-1}, \tau; z) \\
&= \sum_{i=1}^{n} \ell\left(Y_i, \theta_0 + \theta_1(Z_{i1} - z_1) + \cdots + \theta_{r(m)-1}(Z_{id} - z_d)^m, \tau\right) K_h(Z_i - z), \quad \text{(A.2)}
\end{aligned}
$$

where $r(m) - 1$ is the total number of partial derivatives up to order m, i.e., $r(m) = \sum_{j=0}^{m} \binom{j+d-1}{d-1}$. Here and below, $Z_i \equiv (Z_{i1}, \ldots, Z_{id})^T$ and $z \equiv (z_1, \ldots, z_d)^T$.

The mth order local polynomial estimators of f and its derivatives at z are obtained by maximizing $L_n(\theta_0, \theta_1, \ldots, \theta_{r(m)-1}, \tau; z)$. For example, $\hat{f}(z) = \hat{\theta}_0(z)$ and the estimator of $f'(z) \equiv [\partial f(z)/\partial z_1, \ldots, \partial f(z)/\partial z_d]^T$ is given by $\hat{f}'(z) = [\hat{\theta}_1(z), \ldots, \hat{\theta}_d(z)]^T$, where

$$\left(\hat{\theta}_0(z), \hat{\theta}_1(z), \ldots, \hat{\theta}_{r(m)-1}(z), \tilde{\tau}(z)\right) = \arg\max_{\theta_0, \ldots, \theta_{r(m)-1}, \tau} L_n(\theta_0, \theta_1, \ldots, \theta_{r(m)-1}, \tau; z). \quad \text{(A.3)}$$

The estimator $\tilde{\tau}$ is obtained locally in the above local polynomial estimation procedure, and thus depends on z. This can be improved by maximizing the full likelihood with f being replaced by its estimator $\hat{f}$, i.e, a better estimator is given by

$$\hat{\tau} = \arg\max_\tau \sum_{i=1}^{n} \ell(Y_i, \hat{f}(Z_i), \tau). \quad \text{(A.4)}$$

One may further update the estimators $\hat{\theta}_j(z)$ by maximizing with respect to $\theta_0, \theta_1, \ldots, \theta_{r(m)-1}$ only the likelihood $L_n(\theta_0, \theta_1, \ldots, \theta_{r(m)-1}, \hat{\tau}; z)$ where τ on the right hand side of (A.3) is replaced by $\hat{\tau}$.

In the particular case where quadratic local approximation is used to estimate f, the likelihood (A.2) can be written as

$$L_n(\theta_0, \theta_1, \ldots, \theta_{r(2)-1}, \tau; z)$$
$$= \sum_{i=1}^n \ell\left(Y_i, \theta_0 + \boldsymbol{\theta}_1^T(Z_i - z) + \frac{1}{2}(Z_i - z)^T\boldsymbol{\Theta}(Z_i - z), \tau\right) K_h(Z_i - z), \quad \text{(A.5)}$$

where $\boldsymbol{\theta}_1 = (\theta_1, \ldots, \theta_d)^T$ and $\boldsymbol{\Theta} = \boldsymbol{\Theta}^T$ represents the $d(d+1)/2$ last elements of $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{r(2)-1})$ associated with the quadratic terms but written here and below in a quadratic form, for the sake of notational simplicity. Here, after maximizing this likelihood relative to the full set of parameters $(\theta_0, \theta_1, \ldots, \theta_{r(2)-1}, \tau)$, we have directly $\hat{f}(z) = \hat{\theta}_0(z)$ and $\hat{f}'(z) = \boldsymbol{\theta}_1(z)$.

Remark A.1. *For practical computation of the maximum in (A.3), in order to avoid maximization under the constraint $\boldsymbol{\Theta} = \boldsymbol{\Theta}^T$, we should rather use the explicit full expression*

$$\sum_{k=1}^d \sum_{j=1}^k \theta_{k,j}(Z_{ik} - z_k)(Z_{ij} - z_j),$$

in the place of the quadratic form $\frac{1}{2}(Z_i - z)^T\boldsymbol{\Theta}(Z_i - z)$ in (A.5).

A.2 Convolution with Functional Parameters

In this subsection we discuss the case where the shape parameter τ is also a smooth function of z. We know indeed from previous experience in a similar context (see Park *et al.*, 2006) that fiting a local constant shape parameters whereas the true shape parameters are not constant, provides poor estimates of f and f'. It is also shown that choosing a lower order polynomial for the variance functional τ than for f could jeopardize the accuracy of the estimation of the latter.

In addition, if partial derivatives of the functional parameter $\tau(z)$ are of interest, quadratic local polynomial approximation for $\tau(z)$ should be useful. As pointed in Kumbhakar *et al.* (2007), the order of the polynomials for the functionals f and for τ do not have to be the same. They can be adapted to the object of interest for the researcher (derivative or not), reminding that higher orders involve notational and computational complexities. So parcimony is also advisable in this framework. We suggest the following startegy, if partial derivatives of some functional parameter are of interest, we will choose a quadratic local polynomial for this parameter, otherwise, by the principle of parcimony we will choose local linear approximations.

The presentation below is for the general case of a m_1^{th} order local polynomial for the function f and a m_2^{th} order local polynomial for the (vector-valued) function τ. When the

k-valued shape parameter is also a functional parameter, the conditional log-likelihood is now given by $\sum_{i=1}^{n} \ell(Y_i, f(Z_i), \tau(Z_i))$, where $\ell(y, \nu, \omega) = \log g_\varepsilon(y - \nu, \omega)$. Hence we obtain the following local conditional log-likelihood:

$$
\begin{aligned}
&L_n(\theta_0, \ldots, \theta_{r(m_1)-1}, \tau_0, \ldots, \tau_{r(m_2)-1}; z) \\
&= \sum_{i=1}^{n} \ell\left(Y_i, \theta_0 + \theta_1(Z_{i1} - z_1) + \cdots + \theta_{r(m_1)-1}(Z_{id} - z_d)^{m_1},\right. \\
&\qquad\qquad \left. \tau_0 + \tau_1(Z_{i1} - z_1) + \cdots + \tau_{r(m_2)-1}(Z_{id} - z_d)^{m_2}\right) K_h(Z_i - z), \qquad \text{(A.6)}
\end{aligned}
$$

where each $\tau_j \in \mathbb{R}^k$.

The local polynomial estimators of f, τ and their derivatives at z are obtained by maximizing $L_n(\theta_0, \theta_1, \ldots, \theta_{r(m_1)-1}, \tau_0, \ldots, \tau_{r(m_2)-1}; z)$, i.e.,

$$
\begin{aligned}
&\left(\hat{\theta}_0(x), \ldots, \hat{\theta}_{r(m_1)-1}(z), \hat{\tau}_0(z), \ldots, \hat{\tau}_{r(m_2)-1}(z)\right) \qquad\qquad\qquad\qquad \text{(A.7)} \\
&= \arg \max_{\theta_0, \ldots, \theta_{r(m_1)-1}, \tau_0, \ldots, \tau_{r(m_2)-1}} L_n(\theta_0, \theta_1, \ldots, \theta_{r(m_1)-1}, \tau_0, \ldots, \tau_{r(m_2)-1}; z).
\end{aligned}
$$

As above $\hat{f}(z) = \hat{\theta}_0(z)$, $\hat{f}'(z) = [\hat{\theta}_1(z), \ldots, \hat{\theta}_d(z)]^T$ and for the vector τ we have the k-dimensional vector $\hat{\tau}(z) = \hat{\tau}_0(z)$ and the $(d \times k)$ matrix $\hat{\tau}'(z) = \widehat{\frac{\partial \tau^T}{\partial z}} = [\hat{\tau}_1(z), \ldots, \hat{\tau}_d(z)]^T$.

We know that under regularity conditions and appropriate bandwitdhs selection, these are consistent estimators, asymptotically normally distributed with a bias converging to zero when $n \to \infty$ (see Kumbhakar *et al.*, 2007 for the theoretical details). To summarize, if we choose, as we did, a bandwidth matrix $H = hS^{1/2}$, it can be shown by direct application of Theorem 2.2 in Kumbhakar *et al.* (2007), that for linear approximation of all the functionals, the optimal bandwidth that balances bias and variance is given by $h \equiv n^{-1/(d+4)}$: this corresponds to a rate of convergence of the nonparametric estimators of the functionals of the order $n^{2/(d+4)}$. In our setup here, this means $n^{2/(p+q+3)}$, that can be compared to the DEA and FDH rates of convergences (see footnote 2).

A.3 The Customized Model used in the Applications

We first have to select the distribution for the convolution term. As is well known in local MLE procedure, this can be viewed as selecting an anchorage parametric model for the convolution term; the localized version will encompass this anchorage model, ensuring much more flexibility, as shown *e.g.* in Kumbhakar *et al.*, 2007, in Park *et al.*, 2006 and in some of our examples in Section 3.

We will focus here on the two-parameters ($k = 2$) distribution $G(\cdot, \tau)$ obtained from the Normal/Half-Normal convolution case as proposed in Aigner *et al.* (1977). To be explicit, we choose a local normal $\mathcal{N}(0, \sigma_v^2(z))$ for the noise v and a local half-normal $\mathcal{N}^+(0, \sigma_u^2(z))$

for the inefficiency u. An alternative could be the Normal/Exponential convolution also analyzed by Aigner *et al.* (1977), but to save place we will not present this case. Although theoretically valid, we also avoid one sided distributions for the inefficiency term u that would be characterized by two free parameters (scale and shape) as the Gamma (Greene, 1990) or the truncated normal (Stevenson, 1980) due to the identification issues raised by Ritter and Simar (1997). The local version of our model showed to be flexible enough to handle many different situations.

So our nonparametric localized model can be written, as in (A.1)

$$Y = f(Z) - u + v,$$

where $(u|Z = z) \sim |\mathcal{N}(0, \sigma_u^2(z))|$ and $(v|Z = z) \sim \mathcal{N}(0, \sigma_v^2(z))$, u and v being independent conditionally on Z. The conditional probability density function of $\varepsilon = v - u$ is given by[5]

$$g_\varepsilon(\varepsilon|Z = z) = \frac{2}{\sqrt{\sigma_u^2(z) + \sigma_v^2(z)}} \varphi\left(\frac{\varepsilon}{\sqrt{\sigma_u^2(z) + \sigma_v^2(z)}}\right) \Phi\left(-\varepsilon\sqrt{\frac{\sigma_u^2(z)}{\sigma_v^2(z)(\sigma_u^2(z) + \sigma_v^2(z))}}\right),$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of a standard normal variable. Note that we keep the parametrization in terms of the functionals σ_u^2 and σ_v^2 because we could be interested in estimating their derivatives, but any other standard parametrization could be used (*e.g.* in terms of $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$).

As suggested in Kumbhakar *et al.* (2007), in order to avoid nonnegativity restrictions on the variance functions in the local polynomial approximations, we rather choose the following coordinate system for the shape parameters: $\tau(z) = (\widetilde{\sigma}_u^2(z), \widetilde{\sigma}_v^2(z))^T$ where $\widetilde{\sigma}_u^2(z) = \log(\sigma_u^2(z))$ and $\widetilde{\sigma}_v^2(z) = \log(\sigma_v^2(z))$.

The conditional pdf of Y given Z is thus given by:

$$\phi(y, f(z), \tau(z)) = \frac{2}{\sqrt{\exp(\widetilde{\sigma}_u^2(z)) + \exp(\widetilde{\sigma}_v^2(z))}} \varphi\left(\frac{(y - f(z))}{\sqrt{\exp(\widetilde{\sigma}_u^2(z)) + \exp(\widetilde{\sigma}_v^2(z))}}\right)$$
$$\Phi\left(-(y - f(z))\frac{\exp(\widetilde{\sigma}_u^2(z)/2 - \widetilde{\sigma}_v^2(z)/2)}{\sqrt{\exp(\widetilde{\sigma}_u^2(z)) + \exp(\widetilde{\sigma}_v^2(z))}}\right)$$

---

[5]In the case of the estimation of a cost function where $\varepsilon = v + u$, we would obtain:

$$g_\varepsilon(\varepsilon|Z = z) = \frac{2}{\sigma(z)} \varphi\left(\frac{\varepsilon}{\sigma(z)}\right) \Phi\left(\varepsilon\frac{\nu(z)}{\sigma(z)}\right)$$

After some analytical manipulation, it is found that

$$
\begin{aligned}
\ell((y, f(z), \tau(z))) \;=\; & \log(\phi(y, f(z), \tau(z))) \propto -\frac{1}{2} \log \left[\exp(\widetilde{\sigma}_u^2(z)) + \exp(\widetilde{\sigma}_v^2(z)) \right] \\
& - \frac{1}{2} \frac{(y - f(z))^2}{\exp(\widetilde{\sigma}_u^2(z)) + \exp(\widetilde{\sigma}_v^2(z))} \\
& + \log \Phi \left(-(y - f(z)) \frac{\exp(\widetilde{\sigma}_u^2(z)/2 - \widetilde{\sigma}_v^2(z)/2)}{\sqrt{\exp(\widetilde{\sigma}_u^2(z)) + \exp(\widetilde{\sigma}_v^2(z))}} \right),
\end{aligned}
$$

where the constants have been eliminated. The conditional local log-likelihood is

$$
\begin{aligned}
L(f, \tau) \;\propto\; & \sum_{i=1}^{n} \left[-\frac{1}{2} \log \left[\exp(\widetilde{\sigma}_u^2(Z_i)) + \exp(\widetilde{\sigma}_v^2(Z_i)) \right] - \frac{1}{2} \frac{(Y_i - f(Z_i))^2}{\exp(\widetilde{\sigma}_u^2(Z_i)) + \exp(\widetilde{\sigma}_v^2(Z_i))} \right. \\
& \left. + \log \Phi \left(-(Y_i - f(Z_i)) \frac{\exp[\widetilde{\sigma}_u^2(Z_i)/2 - \widetilde{\sigma}_v^2(Z_i)/2]}{\sqrt{\exp(\widetilde{\sigma}_u^2(Z_i)) + \exp(\widetilde{\sigma}_v^2(Z_i))}} \right) \right],
\end{aligned}
$$

Introducing the local polynomial approximations for the functional parameters f, $\widetilde{\sigma}_u^2$ and $\widetilde{\sigma}_v^2$, would follow the general presentation of Section A.2 above. But since we are interested in the partial derivative of σ_u^2 only, we will choose the parcimonious model with local linear approximations for f and $\widetilde{\sigma}_v^2$ and a quadratic approximation for $\widetilde{\sigma}_u^2$, at the point z. The conditional local log-likelihhod is finally given by

$$
\begin{aligned}
L_n(\theta_0, \boldsymbol{\theta}_1, \sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}, \sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2; z) \propto & \\
\sum_{i=1}^{n} \left[-\frac{1}{2} \log \left[\psi_u(\sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}; Z_i - z) + \psi_v(\sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2; Z_i - z) \right] \right. & \\
- \frac{1}{2} \frac{(Y_i - \psi_f(\theta_0, \boldsymbol{\theta}_1; Z_i - z)))^2}{\psi_u(\sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}; Z_i - z) + \psi_v(\sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2; Z_i - z)} & \\
+ \log \Phi \left(\frac{-(Y_i - \psi_f(\theta_0, \boldsymbol{\theta}_1; Z_i - z))\sqrt{\psi_u(\sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}; Z_i - z)}}{\sqrt{\psi_v(\sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2; Z_i - z)}\sqrt{\psi_u(\sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}; Z_i - z) + \psi_v(\sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2; Z_i - z)}} \right) \right] & \\
\times K_H(Z_i - z), & \quad (A.8)
\end{aligned}
$$

where for the ease of notation we used the shortcuts ψ for the following expressions:

$$
\begin{aligned}
\psi_f(\theta_0, \boldsymbol{\theta}_1; Z_i - z) &= \theta_0 + \boldsymbol{\theta}_1^T (Z_i - z) \\
\psi_v(\sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2; Z_i - z) &= \exp \left(\sigma_{v0}^2 + \boldsymbol{\sigma}_{v1}^{2\,T} (Z_i - z) \right) \\
\psi_u(\sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}; Z_i - z) &= \exp \left(\sigma_{u0}^2 + \boldsymbol{\sigma}_{u1}^{2\,T} (Z_i - z) + (1/2)(Z_i - z)^T \boldsymbol{\Sigma}_{u2}(Z_i - z) \right)
\end{aligned}
$$

Here the local parameters are the three scalars $(\theta_0, \sigma_{u0}^2, \sigma_{v0}^2)$ plus the three $(d \times 1)$ vectors $\boldsymbol{\theta}_1, \boldsymbol{\sigma}_{u1}^2$ and $\boldsymbol{\sigma}_{v1}^2$, and the $(d \times d)$ symmetric matrix $\boldsymbol{\Sigma}_{u2}$ containing $d(d+1)/2$ free parameters. At the total, we have $3(d+1) + d(d+1)/2$ free local parameters.

The local polynomial estimator of the model is given by $(\widehat{\theta}_0, \widehat{\sigma}_{u0}^2, \widehat{\sigma}_{v0}^2)$ where

$$(\widehat{\theta}_0, \widehat{\boldsymbol{\theta}}_1, \widehat{\sigma}_{u0}^2, \widehat{\boldsymbol{\sigma}}_{u1}^2, \widehat{\boldsymbol{\Sigma}}_{u2}, \widehat{\sigma}_{v0}^2, \widehat{\boldsymbol{\sigma}}_{v1}^2) =$$

$$\arg \max_{\theta_0, \boldsymbol{\theta}_1, \sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}, \sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2} L_n(\theta_0, \boldsymbol{\theta}_1, \sigma_{u0}^2, \boldsymbol{\sigma}_{u1}^2, \boldsymbol{\Sigma}_{u2}, \sigma_{v0}^2, \boldsymbol{\sigma}_{v1}^2; z). \qquad (A.9)$$

These local estimators are indeed local since the solution of this maximization problem depends on the value of z. In particular $\widehat{f}(z) = \widehat{\theta}_0(z)$ and $\widehat{\widetilde{\sigma}}^2_u(z) = \widehat{\sigma}_{u0}^2(z)$ and $\widehat{\widetilde{\sigma}}^2_v(z) = \widehat{\sigma}_{v0}^2(z)$, where the "tildes" are used here to remind that we are estimating the log of the variance functions. In addition, $\widehat{\frac{\partial}{\partial z} \widetilde{\sigma}}^2_u(z) = \widehat{\boldsymbol{\sigma}}_{u1}^2(z)$. For practical computations, see Remark A.1.

If a quadratic approximation would be needed for f and $\widetilde{\sigma}_v^2$ too, we should add in (A.8) a quadratic term for these two elements. This would involve two additional symmetric matrices $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}_v$ that would appear in quadratic forms as we did it above in (A.8) with $\boldsymbol{\Sigma}_u$. This would add $2d(d+1)$ free parameters in the optimization (A.9). Again, practical organization of the optimization should follow Remark A.1 for the 3 involved symmetric matrices.

Estimates of $\sigma_u^2(z)$, $\sigma_v^2(z)$ and $\frac{\partial}{\partial z}\sigma_u^2(z)$ in the original units can be directly obtained from their logarithmic versions, $\widehat{\sigma}^2_u(z) = \exp\left(\widehat{\widetilde{\sigma}}^2_u(z)\right)$ and $\widehat{\sigma}^2_v(z) = \exp\left(\widehat{\widetilde{\sigma}}^2_v(z)\right)$. For the derivatives we have $\widehat{\frac{\partial}{\partial z}\sigma^2}_u(z) = \widehat{\sigma}^2_u(z) \widehat{\frac{\partial}{\partial z}\widetilde{\sigma}^2}_u(z)$.

References

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

[2] Atkinson, S.E. and D. Primont (2002). Stochastic estimation of firm technology, and productivity growth using shadow cost and distance function, *Journal of Econometrics*, 108, 203–225.

[3] Cazals, C. Florens, J.P. and L. Simar (2002), Nonparametric Frontier Estimation: a Robust Approach , *Journal of Econometrics*, 106, 1–25.

[4] Charnes, A., Cooper W.W. and E. Rhodes (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2 (6), 429-444.

[5] Daouia, A. and L. Simar (2007), Nonparametric efficiency analysis: a multivariate conditional quantile approach, *Journal of Econometrics*, 140, 375–400.

[6] Daraio, C. and L. Simar (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, Springer, New-York.

[7] Deprins, D., Simar, L. and H. Tulkens (1984). Measuring labor inefficiency in post offices. In *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243–267.

[8] Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

[9] Farrell, M.J. (1957), The Measurement of Productive Efficiency, *Journal of Royal Statistical Society*, Series A, 120:3, 253–281.

[10] Greene, W.H. (1990), A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46, 141-163.

[11] Gstach, D. (1998), Another approach to Data Envelopment Analysis in noisy environments: DEA+, *Journal of Productivity Analysis*, 9, 161–176.

[12] Hall, P. and L. Simar (2002), Estimating a Changepoint, Boundary or Frontier in the Presence of Observation Error, *Journal of the American Statistical Association*, 97, 523–534.

[13] Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982), On the estimation of technical inefficiency in stochastic frontier production models, *Journal of Econometrics*, 19, 233–238.

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

[15] Kneip, A. and L. Simar (1996), A General framework for Frontier Estimation with Panel data, *Journal of Productivity Analysis*, 7, 187–212.

[16] Kumbhakar, S.C. , Park, B.U., Simar, L. and E.G. Tsionas (2007), Nonparametric stochastic frontiers: a local likelihood approach, *Journal of Econometrics*, 137, 1, 1–27.

[17] Land, K.C., Lovell, C.A.K. and S. Thore (1993), Chance-constrained data envelopment analysis, *Managerial and Decison Economics*, 14(6), 541–554.

[18] Meeusen, W. and J. van den Broek (1977), Efficiency estimation from Cobb-Douglas production function with composed error, *International Economic Review*, 8, 435-444.

[19] Park, B. Simar, L. and Ch. Weiner (2000), The FDH Estimator for Productivity Efficiency Scores : Asymptotic Properties, *Econometric Theory*, Vol 16, 855–877.

[20] Park, B., Simar, L. and V. Zelenyuk (2006), Local Likelihood Estimation of Truncated Regression and its Partial Derivatives: Theory and Application, Discussion paper 0606, Institut de Statistique, UCL.

[21] Petersen, N.C. and O.B. Olesen (1995), Chance-constrained efficiency evaluation, *Management Science*, 41, 442–457.

[22] Ritter, C. and L. Simar (1997), Pitfall of Normal-Gamma stochastic frontier models, *Journal of Productivity Analysis*, 8, 167–182.

[23] Simar, L. (2007), How to Improve the Performances of DEA/FDH Estimators in the Presence of Noise, *Journal of Productivity Analysis*, 28, 183–201.

[24] Simar, L. and P.W. Wilson (2007), Estimation and Inference in Cross-Sectional Stochastic Frontier Models. Discussion paper 0524, revised version 2007, Institut de Statistique, UCL.

[25] Simar, L. and P.W. Wilson (2008), Statistical Inference in Nonparametric Frontier Models: recent Developments and Perspectives, in *The Measurement of Productive Efficiency*, 2nd Edition, Harold Fried, C.A.Knox Lovell and Shelton Schmidt, editors, Oxford University Press, 2008.

[26] Stevenson, R.E. (1980), Likelihood functions for generalized stochastic frontier estimation, *Journal of Econometrics*, 13, 57–66