IDEI Working Paper Series n° 625

May, 2010

"Nonparametric Frontier Estimation from Noisy Data"

Jean-Pierre FLORENS,
Sébastien VAN BELLEGEM and
Maik SCHWARZ

Institut d'Economie Industrielle (IDEI) – Manufacture des Tabacs
Aile Jean-Jacques Laffont – 21, allée de Brienne – 31000 TOULOUSE – FRANCE
Tél. + 33(0)5 61 12 85 89 – Fax + 33(0)5 61 12 86 37 – **www.idei.fr** – contact@cict.fr

INSTITUT D'ECONOMIE INDUSTRIELLE

Nonparametric frontier estimation from noisy data*

Maik Schwarz[1] Sébastien Van Bellegem[2] Jean-Pierre Florens[2]

May 11, 2010

Abstract

A new nonparametric estimator of production a frontier is defined and studied when the data set of production units is contaminated by measurement error. The measurement error is assumed to be an additive normal random variable on the input variable, but its variance is unknown. The estimator is a modification of the m-frontier, which necessitates the computation of a consistent estimator of the conditional survival function of the input variable given the output variable. In this paper, the identification and the consistency of a new estimator of the survival function is proved in the presence of additive noise with unknown variance. The performance of the estimator is also studied through simulated data.

Keywords: Production frontier, deconvolution, measurement error, efficiency analysis

*This work was supported by the "Agence National de la Recherche" under contract ANR-09-JCJC-0124-01 and by the IAP research network nr P6/03 of the Belgian Government (Belgian Science Policy). Comments from Ingrid Van Keilegom and an anonymous referee were most helpful to improve the final version of the manuscript. The usual disclaimer applies.

[1]Institut de statistique, biostatistique et sciences actuarielles (Université catholique de Louvain)

[2]Toulouse School of Economics (GREMAQ)

1 Introduction

The modelling and estimation of production functions have been the topic of many research papers on economic activity. A classical formulation of this problem is to consider production units characterized by a vector of inputs $x \in \mathbb{R}_+^p$ producing a vector of outputs $y \in \mathbb{R}_+^q$. The set of production possibilities is denoted by Φ and is a subset of $\mathbb{R}_+^{p+q}$ on which the inputs x can produce the outputs y. Following Shephard (1970), several assumptions are usually imposed on Φ: convexity, free disposability and strong disposability. Free disposability means that if (x, y) belongs to Φ and if x', y' are such that $x' \geqslant x$ and $y' \leqslant y$ then $(x', y') \in \Phi$. Strong disposability requires that one can always produce a smaller amount of outputs using the same inputs.

The boundary of the production set is of particular interest in the efficiency analysis of production units. The efficient frontier in the input space is defined as follows. For all $y \in \mathbb{R}_+^q$, consider the set $\rho(y) = \{x \in \mathbb{R}_+^p | (x, y) \in \Phi\}$. The radial efficiency boundary is then given by

$$\varphi(y) = \{x \in \mathbb{R}_+^p \;:\; x \in \rho(y), \theta x \notin \rho(y) \; \forall 0 < \theta < 1\}$$

for all y. Similarly, an efficient frontier in the output space may be defined (e.g. Färe, Grosskopf, & Knox Lovell, 1985).

In empirical studies, the attainable set Φ is unknown and has to be estimated from data. Suppose a random sample of production units $\mathcal{X}_n = \{(X_i, Y_i) \in \mathbb{R}_+^{p+q} \;:\; i = 1, \ldots, n\}$ is observed. We assume that each unit (X_i, Y_i) is an independent replication of (X, Y). The joint probability measure (X, Y) on $\mathbb{R}_+^{p+q}$ describes the production process. The support of this probability measure is the attainable set Φ, and looking for an estimator of the efficiency boundary is related to the estimation of the support of (X, Y).

Out of the large literature on the estimation of the attainable set, nonparametric models appeared to be appealing since they do not require restrictive assumptions on the data generating process of $\mathcal{X}_n$. Deprins, Simar, and Tulkens (1984) have introduced the Free Disposal Hull (FDH) estimator that is defined as

$$\hat{\Phi}_{fdh} = \{(x, y) \in \mathbb{R}_+^{p+q} \;:\; y \leqslant Y_i, x \geqslant X_i, i = 1, \ldots, n\}$$

and became a popular estimation method (e.g. De Borger, Kerstens, Moesen, & Vanneste, 1994; Leleu, 2006). The convex hull of $\hat{\Phi}_{fdh}$, called the Data Envelopment Analysis (DEA), is the smallest free disposal convex set covering the data (e.g. Seiford & Thrall, 1990). Among the significant results on this subject, we like to mention the asymptotic results proved in Kneip, Park, and Simar (1998) for the DEA and Park, Simar, and Weiner (2000) for the FDH.

The consistency of the FDH estimator and other data envelopment techniques is only achieved when the production units are observed without noise, that is when $\mathbb{P}((X_i, Y_i) \in \Phi) = 1$. However, FDH in particular is very sensitive to the contamination of the data by measurement errors or by outliers (e.g. Cazals, Florens, & Simar, 2002; Daouia, Florens, & Simar, 2009). Measurement errors are frequently encountered in economic data bases, and therefore there is a need for developing more robust estimation procedures of the production frontier.

In Cazals et al. (2002), a new nonparametric estimator has been proposed to overcome the nonparametric frontier estimation from contaminated samples. When $p = 1$ and under the free disposability assumption, they show that the frontier function $\varphi(y)$ can be written as

$$\varphi(y) = \inf\{x \in \mathbb{R}_+ \text{ such that } S_{X|Y \geqslant y}(x) < 1\}, \tag{1.1}$$

where $S_{X|Y\geqslant y}(x) = \mathbb{P}(X \geqslant x | Y \geqslant y)$ denotes the conditional survival function. If $X^1, \ldots, X^m$ are m independent replications of $(X|Y \geqslant y)$ for an integer $m > 1$, then a key observation in Cazals et al. (2002) is that the expected minimum input functions

$$\varphi_m(y) := \mathbb{E}\left(\min\{X^1, \ldots, X^m\}|Y \geqslant y\right) \qquad m = 1, 2, 3, \ldots \tag{1.2}$$

are such that

$$\varphi_m(y) := \int_0^\infty \left\{S_{X|Y\geqslant y}(u)\right\}^m \mathrm{d}u \tag{1.3}$$

and $\varphi_m(y)$ converges pointwise to the frontier $\varphi(y)$ as m tends to infinity (assuming the existence of $\varphi_m(y)$ for all m). The functions $\varphi_m(y)$ are nonparametrically estimated in Cazals et al. (2002) from nonparametric estimators of the conditional survival function $S_{X|Y\geqslant y}$. The empirical survival function is defined by $\hat{S}_{X,Y}(x,y) = n^{-1}\sum_i \mathbb{1}(X_i \geqslant x, Y_i \geqslant y)$ and the empirical version of $S_{X|Y\geqslant y}$ is thus given by

$$\hat{S}_{X|Y\geqslant y} = \frac{\hat{S}_{X,Y}(x,y)}{\hat{S}_Y(y)} \tag{1.4}$$

where $\hat{S}_Y(y) = n^{-1}\sum_i \mathbb{1}(Y_i \geqslant y)$. Cazals et al. (2002) have studied the asymptotic properties of the frontier estimator

$$\hat{\varphi}_{m,n}(y) := \int_0^\infty \left\{\hat{S}_{X|Y\geqslant y}(u)\right\}^m \mathrm{d}u \tag{1.5}$$

that is called the m-frontier estimator. They argue that this estimator is less sensitive to extreme values or noise in the sample of production units than FDH or DEA-type estimators.

In this article, we slightly amend this claim, and show that, when the noise level on the data does not vanish as the sample size n grows, then the m-estimator is no longer asymptotically consistent. When the noise level is too high, we show that consistency may be recovered when a robust estimate of the conditional survival function is plugged-in in the integral in (1.5). By “robust estimate”, we think here of an estimator of $S_{X|Y\geqslant y}$ that is consistent even in the presence of a non-vanishing noise in the sample.

In this article, a new robust estimator of the survival function is studied when the inputs X are contaminated by an additive error. We show the consistency of the estimator under the assumption that we only have a partial information on the distribution of the error. More precisely, we assume that the additive noise is a zero-mean Gaussian random variable with an unknown variance.

The paper is organized as follows. In Section 2, we give an overview of existing methods to nonparametrically estimate a density from noisy observations when the distribution of the noise is partially unknown. In Section 3, we define a new estimator of the survival function in the univariate case, when the data are contaminated by an additive Gaussian random noise with an unknown variance. We prove the asymptotic consistency of our estimator. Finite sample properties are also considered through Monte Carlo simulations. In Section 4, we define and illustrate on data a new robust m-frontier estimator that is defined similarly to the estimator in (1.5), except that our robust estimator of the conditional survival function is plugged in the integral. The consistency of the robust m-frontier estimator is also established theoretically in this section. The last section summarizes the results of this paper and suggests future directions of research.

2 Density estimation from noisy observations

Estimating the distribution of a real random variable X from a noisy sample is a standard problem in nonparametric statistics. The usual setting is to assume independent and identically distributed (iid) observations from a random variable Z such that $Z = X + \varepsilon$, where ε represents an additive error independent of X. Many research papers focus on the accurate estimation of the cumulative distribution function (cdf) of X under the assumption that the cdf of ε is known. The additive measurement error implies that the density of Z, if it exists, is the convolution between the density of ε and the one of X:

$$f^Z(z) = f^\varepsilon \star f^X(z) := \int_{-\infty}^{\infty} f^\varepsilon(t) f^X(z-t)\mathrm{d}t \, .$$

Based on this result, most estimators of f^X studied in the literature use the Fourier transform of the densities since the Fourier coefficients of the convolution are the product of the coefficients:

$$\psi^Z(\ell) = \psi^\varepsilon(\ell)\psi^X(\ell), \qquad \ell \in \mathbb{Z}$$

where $\psi^U(\ell) := \mathbb{E}\{\exp(i\ell U)\}$ denotes the ℓ-th Fourier coefficient of a density f^U. A usual estimator of $\psi^Z(\ell)$ is (a functional of) the empirical characteristic function of the sample $(Z_1, \ldots, Z_n)$, i.e.

$$\hat{\psi}^Z(\ell) := \frac{1}{n}\sum_{i=1}^{n} \exp(i\ell Z_i), \qquad \ell \in \mathbb{Z} \, .$$

From this estimator and under the condition that f_ℓ is known and nonzero, the standard estimators are based on the inverse Fourier transform of $\hat{\psi}^Z(\ell)/\psi^\varepsilon(\ell)$ (e.g. Carroll & Hall, 1988; Fan, 1991). Alternative estimators have also been studied in the literature, for instance in the wavelet domain (Pensky & Vidakovic, 1999; Johnstone, Kerkyacharian, Picard, & Raimondo, 2004; Bigot & Van Bellegem, 2009).

The exact knowledge of the cdf of the error is however not realistic in many empirical studies. If we want to relax the condition that the cdf of the error is known, one major obstacle is that the cdf of X is no longer identifiable. To circumvent this problem, at least three research directions may be found in the literature.

A first approach assumes that an independent sample from the measurement error ε is available in addition to the sample of Z. From the independent observation of ε, the density f^{ε} is identified and so is the target density f^X. A nonparametric estimator from the sample of ε's can be constructed, and then used in the construction of the estimator of f^X (Neumann, 2007; Johannes & Schwarz, 2009; Johannes, Van Bellegem, & Vanhems, 2010). If this approach may be realistic for a set of practical situations (e.g. in some problems in biostatistics and astrophysics), it is hardly applicable in production frontier estimation.

A second approach is to assume various sampling processes. Li and Vuong (1998) suppose that repeated measurements for one single value of X are available, such as $Z_j = X + \varepsilon_j$ for $j = 1, \ldots, m$. Assuming further that X, ε_1, and ε_2 are mutually independent, $\mathbb{E}(\varepsilon_j) = 0$, and that the characteristic functions of X and ε are non-zero everywhere, they show how the latter characteristic functions can be expressed as functions of the joint characteristic function of (Z_1, Z_2). From this representation it follows that the cumulative distribution function (cdf) of both X and ε can be identified from the observation of the pair (Z_1, Z_2). The joint characteristic function of (Z_1, Z_2) can be estimated from a sample of (Z_1, Z_2) and then used to derive an estimator of f^X. The characteristic functions of X and ε, denoted by ψ^X and ψ^{ε}, can then be computed using the above-mentioned representation. Delaigle, Hall, and Meister (2008) have considered this setting and present modified kernel estimators which, if the number of repeated measurements is large enough, can perform as well as they would under known error distribution.

A related situation is when there are repeated measurements of X in a multilevel model. In Neumann (2007) it is assumed that $Z_{ij} = X_i + \varepsilon_{ij}$ for $j = 1, \ldots, N$ and $i = 1, \ldots, n$ are observed (see also Meister, Stadtmüller, & Wagner, 2010). In this sampling process, the identification of the cdf of X is ensured by a condition on the zero-sets of the characteristic functions of X and ε. Let $\mathcal{Z} = (Z_{i1}, \ldots, Z_{iN})'$, $\psi^{\mathcal{Z}}$ its characteristic function, and $\hat{\psi}^{\mathcal{Z}}$ the empirical characteristic function of $\mathcal{Z}$. A consistent estimator of the density of X is obtained by minimizing the discrepancy

$$\int_{\mathcal{R}^n} \left|\psi^X(t_1 + \ldots + t_n)\psi^{\varepsilon}(t_1) \cdots \psi^{\varepsilon}(t_n) - \hat{\psi}^{\mathcal{Z}}_n(t_1, \ldots, t_n)\right| h(t_1, \ldots, t_n) \mathrm{d}t_1 \ldots \mathrm{d}t_n$$

over certain classes of possible characteristic functions ψ^X and ψ^{ε} of X and ε respectively.

Repeated measurements of multilevel sampling appear in some economic situations, for instance when production units are observed over time (a case considered e.g. in Park, Sickles, & Simar, 2003; Daskovska, Simar, & Van Bellegem, 2010).

A third approach to recover the identification of X in spite of the noise ε is to assume that the cdf of ε is only *partially unknown*. A realistic case for practical purposes is to assume that ε is normally distributed, but the variance of ε is unknown. Of course the cdf of X is not identified in this setting, and it is necessary to restrict the class of cdfs of X in order to recover identification.

Several recent research papers have proposed identification restrictions on the class of X given a partial knowledge about the cdf of the noise. Butucea and Matias (2005) assume that the error density, is "s-exponential" meaning that its Fourier transform, ψ^{ε}, satisfies

$$b \exp(-|u|^s) \leqslant |\psi^{\varepsilon}(u)| \leqslant B \exp(-|u|^s)$$

for some constants b, B, s and $|u|$ large enough. In their approach the error density is supposed to be known up to its scale σ (called "noise level"). As for the density f^X, both polynomial and exponential decay of its Fourier transform are shown to lead to a fully identified model. To define an estimator, let ψ_σ^ε be the Fourier transform of (σf^ε). The key to the estimation of σ is the observation that the function $|F(\tau, u)| = |\psi^Z(u)|/|\psi_\tau^\varepsilon(u)|$ diverges as $u \to \infty$ when $\tau > \sigma$ and that it converges to 0 otherwise. Let $\hat{F}(\tau, u_n) = |\hat{\psi}^Z(u_n)|/|\psi_\tau^\varepsilon(u_n)|$. Then Butucea and Matias (2005) show that

$$\hat{\sigma}_n = \inf\{\tau > 0 : |\hat{F}(\tau, u_n)| \geqslant 1\}$$

yields a consistent estimator of σ for some well balanced sequence u_n. This estimator is then used to deconvolve the empirical density of Z and to get an estimator of the density of X. Some extensions are proposed in Butucea, Matias, and Pouet (2008), where the error density is assumed to have a stable symmetric distribution with $\psi^\varepsilon(u) = \exp(-|\gamma u|^s)$ in which γ represents some known scale parameter and s is an unknown index, called the self-similarity index.

A similar setting is considered in Meister (2006). In this paper, the error is supposed to be normally distributed with an unknown variance parameter. Identification is recovered by assuming that ψ^X lies in $\{\psi : c_1|u|^{-\beta} \leqslant |\psi(u)| \leqslant c_2|u|^{-\beta}$ for all $u \gg 0\}$ for some strictly positive constants c_1, c_2.

In Meister (2007), it is assumed that ψ^ε is known on some arbitrarily small interval $[-\nu, \nu]$ and that it belongs to some class

$$\mathcal{G}_{\mu,\nu} = \{f \text{ is a density such that } \|f\|_\infty \leqslant C, |\psi^f(t)| \geqslant \mu \ \forall |t| \geqslant \nu\}.$$

The target density f^X is assumed to belong to

$$\mathcal{F}_{S,C,\beta} = \{f \text{ is a density such that } \int_{-S}^{S} f(u)\mathrm{d}u = 1 \text{ and } \int |\psi^f(t)|^2(1+t^2)^\beta \mathrm{d}t \leqslant C\},$$

that is in the class of densities with compact support that are uniformly bounded in the Sobolev norm. Empirically the direct access to ψ^X via Fourier deconvolution is only restricted to the interval $[-\nu, \nu]$. However, it is shown using a Taylor expansion that ψ^X is uniquely determined by its restriction to $[-\nu, \nu]$, and therefore is everywhere identified.

Because the deconvolution of the density of Z is solved via the Fourier transform, most of the assumptions on X or ε recalled above are expressed in terms of their characteristic functions. They appear to be *ad hoc* assumptions, although they could be difficult to interpret econometrically. In Schwarz and Van Bellegem (2010), an identification theorem is proved on the target density under assumptions that are not expressed in the Fourier domain. It is instead assumed that the measurement error ε is normally distributed with an unknown variance parameter, and that f^X lies in the class of densities that vanish on a set of positive Lebesgue measure. This restriction on the class of target densities is reasonable for our purpose of frontier estimation, in which it is structurally assumed that the density of X (or the conditional density of $(X|Y \geqslant y)$) is zero beyond the frontier. Since this is a natural assumption in the setting of frontier estimation, we use this framework in the next section in order to estimate a survival function from noisy data.

3 A new estimator of the survival function from noisy observations

3.1 Identification of the survival function

Suppose we observe a sample $\{Z_1, \dots, Z_n\}$ of n independent replications of Z from the model

$$Z = X + \varepsilon\ , \tag{3.1}$$

where ε is a $N(0, \sigma^2)$ random variable, independent from X, and with an unknown variance σ^2. As explained in the previous section, the probability density of Z is the convolution $\phi_\sigma \star f^X$, where f^X is the probability density of X and ϕ_σ denotes the Normal density with standard error σ. The following theorem, quoted from Schwarz and Van Bellegem (2010), defines a set of identified probability distributions f^X from model (3.1). The survival function S^X of X will hence be identified on that set from the observation of Z.

Theorem 3.1. *Define the following set of probability distributions:*

$$\mathcal{P}_0 := \{P \text{ distribution } : \exists \text{ Borel set } A \text{ such that } |A| > 0 \text{ and } P(A) = 0\},$$

where $|A|$ denotes the Lebesgue measure of A. The model defined by (3.1) is identifiable for the parameter space $\mathcal{P}_0 \times (0, \infty)$. In other words, for any two probability measures $P^1, P^2 \in \mathcal{P}_0$ and $\sigma_1, \sigma_2 > 0$, we have that $\phi_{\sigma_1} \star P^1 = \phi_{\sigma_2} \star P^2$ implies $P^1 = P^2$ and $\sigma_1 = \sigma_2$.

3.2 A consistent estimator

From model (3.1), we also observe after a straightforward calculation that the survival function of Z, denoted by S^Z, follows a convolution formula:

$$S^Z(z) = \phi_\sigma \star S^X(z)$$

where S^X is the survival function of the variable X and ϕ_σ denotes the density function of a $N(0, \sigma^2)$ random variable.

Our estimator of S^X is approximated in a sieve as follows. For any integers $k, D > 0$, define $\Delta^{(k,D)} := \{\delta \in \mathbb{R}^k\ :\ 0 \leqslant \delta_1 \leqslant \ldots \leqslant \delta_k \leqslant D\}$ and for $\delta \in \Delta^{(k,D)}$ define

$$S_\delta(t) := \frac{1}{k} \sum_{j=1}^{k} \mathbb{1}(\delta_j > t)\ . \tag{3.2}$$

For any $\delta \in \Delta^{(k,D)}$, denote by P_δ the probability distribution corresponding to the survival function S_δ. The choice of the approximating function is performed minimizing the contrast function

$$\gamma(S, \zeta; T) := \int_{-\infty}^{\infty} \big|(\phi_\zeta \star S)(t) - T(t)\big| h(t) \mathrm{d}t,$$

where h is some strictly positive probability density ensuring the existence of the integral.

We are now in position to define our estimator of the survival function. Let $(k_n)_{n\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$ be two positive, divergent sequence of integers. The estimator $(S_{\hat{\delta}(n)}, \hat{\sigma}_n)$ is defined by

$$(\hat{\delta}(n), \hat{\sigma}_n) := \underset{\substack{\delta\in\Delta^{(k_n,D_n)} \\ \sigma\in[0,D_n]}}{\arg\min}\ \gamma(S_\delta, \sigma; \hat{S}_n^Z)\ , \tag{3.3}$$

where $\hat{S}_n^Z := n^{-1}\sum_{k=1}^n \mathbb{1}(Z_k > t)$ is the empirical survival function of Z. Note that the argmin is attained because it is taken over a compact set of parameters. Though, it is not necessary unique. If it is not, an arbitrary value among the possible solutions may be chosen.

Theorem 3.2. *The estimator $(S_{\hat{\delta}(n)}, \hat{\sigma}_n)$ is consistent in the sense that*

$$P^X_{\hat{\delta}_n} \xrightarrow{\mathcal{L}} P^X \qquad \text{and} \qquad \hat{\sigma}_n \to \sigma$$

almost surely as $n \to \infty$, where $\xrightarrow{\mathcal{L}}$ denotes weak convergence of probability measures.

The proof of this result is based on some technical lemmas and can be found in the appendix below.

To illustrate the estimator, we now present the result of a Monte Carlo experiment. The estimator of the standard deviation σ of the noise is of particular interest. In the following experiment, we consider two designs for the input X. One is uniformly distributed over $[1, 2]$, and the other is a mixture $U[1, 2] + Exp(1)$. In both cases the density of X is zero below 1, and in the second case the support of X is not bounded to the right. For various true values of σ, we calculate the estimators $(\hat{\delta}(n), \hat{\sigma}_n)$ for sample sizes $n = 100$, 200 and 500. No particular optimization over the value of k (appearing in (3.2)) is provided, except that we increase k as the sample size increases. For the considered sample sizes, we set $k = 10\, n^{1/2}$. The minimization of the contrast function is calculated using the algorithm `optim` in the R software. For this algorithm, we have chosen the initial values of δ_j to be equispaced values over the interval $[0, 3]$ and the initial value of σ is the empirical standard deviation of the sample $Z_1, \ldots, Z_n$.

Tables 1 and 2 show the result of the Monte Carlo simulation using $B = 2000$ replications of each design. The mean and standard deviation of $\sigma - \hat{\sigma}_n$ over the B replications are displayed. Some results are not reported for very small sizes, because a stability problem has been observed, especially in the mixture case. In these cases, the `optim` algorithm did not often converge (a similar phenomenon has been observed using the `nlm` algorithm). It also has to be mentioned that the stability is very sensitive to the choice of k and to the choice of initial values for δ and σ. For larger sample sizes, or larger values of the noise, the results overall improve with the sample size.

4 Robust m-frontier estimation in the presence of noise

4.1 Inconsistency of the m-frontier estimator

Let us now consider our initial problem of consistently estimating the production frontier $\varphi(y)$ from a sample of production units (X_i, Y_i), where X_i is the input and Y_i is the output.

	True σ		
n	1	2	5
100		1.30	-1.08
		(1.05)	(0.51)
200	0.91	0.07	-0.38
	(3.84	(0.45)	(0.45)
500	0.37	0.06	0.14
	(0.30)	(0.44)	(0.49)

Table 1: The inputs simulated in this experiment are uniformly distributed over $[1, 2]$. For each sample size and noise level, we compute the mean of $\sigma - \hat{\sigma}_n$ from $B = 2000$ replications (the standard deviation is given between parentheses)

	True σ		
n	1	2	5
100		2.84	-0.92
		(7.80)	(7.15)
200		-0.49	-0.49
		(6.32)	(5.92)
500	1.78	0.029	0.014
	(5.90)	(4.88)	(6.69)

Table 2: The inputs simulated in this experiment are a mixture $U[1, 2] + Exp(1)$. For each sample size and noise level, we compute the mean of $\sigma - \hat{\sigma}_n$ from $B = 2000$ replications (the standard deviation is given between parentheses)

To simplify the discussion, we assume that the dimension of the input and the output are $p = q = 1$.

In the introduction we have recalled the definition of the m-frontier estimator in equation (1.5). Compared to the FDH or DEA estimator, this nonparametric frontier estimator provides a more robust estimator of the frontier in the presence of noise. In Cazals et al. (2002, Theorem 3.1) it is also proved that for any interior point y in the support of the distribution Y and for any $m \geqslant 1$, it holds that

$$\hat{\varphi}_{m,n}(y) \to \varphi_m(y) \qquad \text{almost surely as } n \to \infty \tag{4.1}$$

where $\varphi_m(y)$ is the expected minimum input function of order m given in equation (1.2).

When the input of the production units is contaminated by an additive error, the actually observed inputs are

$$Z_i = X_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

instead of X_i, for some positive, unknown variance parameter σ^2. If σ^2 does not vanish asymptotically, the limit appearing in (4.1) is no longer given by the expected minimum input function (1.2). Instead we get

$$\hat{\varphi}_{m,n}(y) \to \mathbb{E}\left(\min\{Z_1, \ldots, Z_m\} | Y \geqslant y\right) \qquad \text{almost surely as } n \to \infty\ .$$

The expectation appearing on the right hand side is not (1.2) because the support of the variable Z is the whole real line. Therefore, the m-frontier estimator does not converge to the desired target function, due to the non-vanishing error variance. Note that this is in contrast with the approach of Hall and Simar (2002) or Simar (2007). In the two latter references, the noise level is assumed to be asymptotically negligible.

The inconsistency of the m-frontier estimator is illustrated in Figures 1 and 2. The true production frontier in this simulation is given by $\varphi(y) = \sqrt{y}$ and is displayed by the dashed line. We have simulated 200 production inputs from the model $X_i = Y_i^2 + E_i$, where $E_i \sim Exp(1)$. The production inputs are then contaminated by an additive noise, so that the observed inputs are $Z_i = X_i + \varepsilon_i$ instead of X_i, where ε_i are independently generated from a zero mean normal variable with standard error $\sigma = 2$.

The FDH estimator computed in Figure 1 is known to be inconsistent in this situation, because it is constructed under the assumption that all production units are in the production set Φ with probability one. Figure 2 shows the m-frontier of Cazals et al. (2002) for $m = 1$ and 50 respectively (cf. (1.5)). As discussed in Cazals et al. (2002), an appropriate choice of m is delicate and, as far as we know, there is no automatic procedure to select it from the data. If m is too low, the m-frontier is not a good estimator of the production function. In the theory of Cazals et al. (2002), m is an increasing parameter with respect to the sample size. For large values of y, the estimator is above the true frontier.

For larger values of m, as shown in Figure 2, the estimator is close to the FDH estimator. Because the value of m increases with n in theory, the two estimators will be asymptotically close. This illustrates the inconsistency of the m-frontier in the case where the noise on the data is not vanishing with increasing sample size.

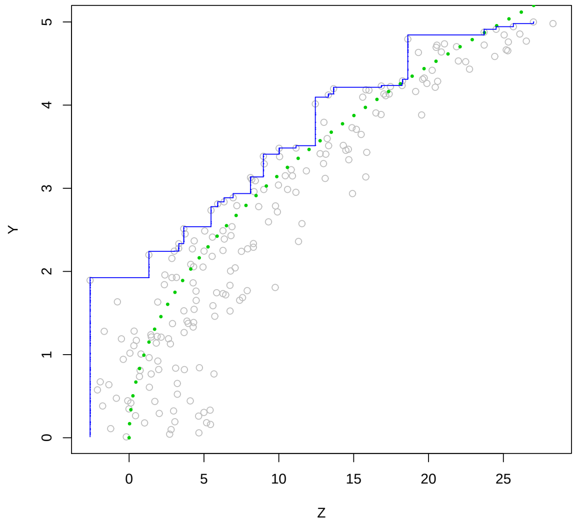

Figure 1: The gray points are the simulated production units and the thick line is the true production frontier. The solid line is the Free Disposal Hull (FDH) estimator of the frontier.

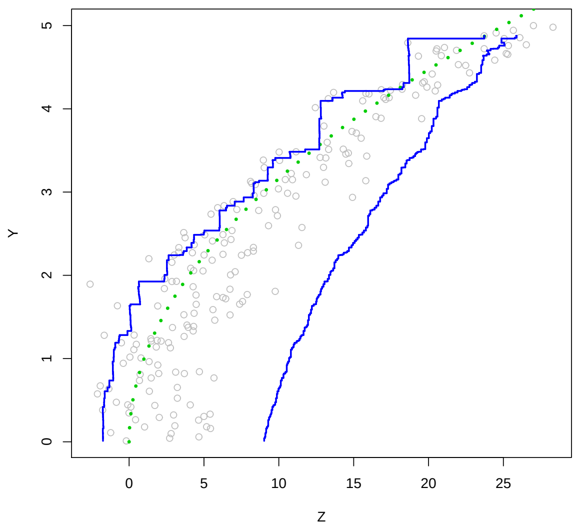

Figure 2: Using the same data as in Figure 1, the two solid lines are the m-frontier estimator with $m = 1$ and $m = 50$ respectively.

4.2 Robust m-frontier estimation

In order to recover the consistency of the m-frontier, we need to plug-in a consistent estimator of the conditional survival function in (1.3). The construction of the estimator is easy from the above results if we assume that the additive noise to the inputs is independent from the input X and the output Y. Let y be a point in the output domain where the support of Y is strictly positive. Restricting the data set to $(Z_i|Y_i \geqslant y)$, we can construct the empirical conditional survival function $\hat{S}_{Z|Y\geqslant y}$ using the usual nonparametric estimator (1.4). Note that this estimator does not require any regularization parameter such as a bandwidth. In analogy to (3.3), we also define

$$(\hat{\delta}(n), \hat{\sigma}_n) := \underset{\substack{\delta\in\Delta^{(k_n,D_n)} \\ \sigma\in[0,D_n]}}{\arg\min}\ \gamma(S_\delta, \sigma; \hat{S}_{Z|Y\geqslant y})\ . \tag{4.2}$$

The final robust m-frontier estimator is then given by

$$\hat{\varphi}^{rob}_{m,n}(y) = \int_0^\infty \left\{S_{\hat{\delta}(n)}(u)\right\}^m \mathrm{d}u\ . \tag{4.3}$$

Note that this integral is easy to compute since $S_{\hat{\delta}(n)}$ is a step function. The following result establishes the consistency of this new estimator under a condition on the parameter m.

Proposition 4.1. *Suppose we observe production units $\{(Z_i, Y_i); i = 1, \ldots, n\}$ in which the univariate inputs are such that $Z_i = X_i + \varepsilon_i$, where ε_i models a measurement error that is independent from X_i and Y_i, normally distributed with zero mean and unknown variance σ^2. Consider the robust m-frontier estimator given by equations (4.2) and (4.3) and let m_n be a strictly divergent sequence of positive integers such that*

$$\{S_{\hat{\delta}(n)}(\varphi(y))\}^{m_n} \to 1 \tag{4.4}$$

almost surely as $n \to \infty$. Then $\hat{\varphi}^{rob}_{m_n,n}(y) \to \varphi(y)$ almost surely as $n \to \infty$.

This result illustrates well the role of the parameter $m = m_n$, which has to tend to infinity at an appropriate rate as $n \to \infty$ in order to achieve consistency of the robust frontier estimator. Indeed, if m_n is bounded by some $M > 0$, Fatou's Lemma implies that almost surely

$$\lim_{n\to\infty} \hat{\varphi}^{rob}_{m_n,n}(y) \geqslant \int_0^\infty \left\{S_{X|Y\geqslant y}(u)\right\}^M \mathrm{d}u = \varphi(y) + \int_{\varphi(y)}^\infty \left\{S_{X|Y\geqslant y}(u)\right\}^M \mathrm{d}u.$$

Except for the trivial case where the true conditional survival function is the indicator function of the interval $(-\infty, \varphi(y))$, the last integral on the right hand side is strictly positive. This shows that the robust estimator asymptotically overestimates the true frontier $\varphi(y)$ if m_n does not diverge to infinity.

On the other hand, if m_n increases too fast in the sense that the condition in (4.4) does not hold, then $\hat{\varphi}^{rob}_{m_n,n}(y)$ may asymptotically underestimate the true frontier $\varphi(y)$ as one can see decomposing the integral from (4.3) into

$$\int_0^\infty \left\{S_{\hat{\delta}(n)}(u)\right\}^{m_n} \mathrm{d}u = \int_0^{\varphi(y)} \left\{S_{\hat{\delta}(n)}(u)\right\}^{m_n} \mathrm{d}u + \int_{\varphi(y)}^\infty \left\{S_{\hat{\delta}(n)}(u)\right\}^{m_n} \mathrm{d}u.$$

The second integral on the right hand side tends to 0 almost surely for $n \to \infty$ as we explain in the proof of Proposition 4.1. As for the first one, the integrand converges to a non-negative monotone function S with $S(\varphi(y)) < 1$, and hence the integral may tend to a limit that is smaller than the true frontier $\varphi(y)$. However, this need not be the case, and thus the condition in (4.4) is sufficient but not necessary.

Summarizing the above discussion, the sufficient condition in (4.4) implicitly defines an appropriate rate at which m_n may diverge to infinity such that the new robust frontier estimator is consistent. This rate depends on characteristics of the true conditional survival function, and we do not know at present how to choose it in an adaptive way. Nevertheless, the simulations show that even for finite samples, large choices of m do not deteriorate the performance of the robust estimator.

The estimator is computed for each possible value of y. In practice, it is not necessary to estimate the standard deviation of the noise for each y. We can first estimate the noise level using the marginal data set of inputs only, and use the techniques developed in Section 4. We then use this estimated value in (4.2) even as an initial parameter of the `optim` algorithm, or as a fixed, known parameter of the noise standard deviation.

Figure 3 shows the estimator on the simulated data of Figure 1. As for the standard m-frontier, the robust m-frontier with $m = 1$ is not a satisfactory estimator. The interesting fact about the robust m-frontier is that it does not deteriorate the frontier estimation for large values of m. For the sake of comparison with Figure 2, Figure 3 also displays the robust m-frontier estimator with $m = 50$. This estimator does not cross the true production frontier and does not converge to the FDH estimator.

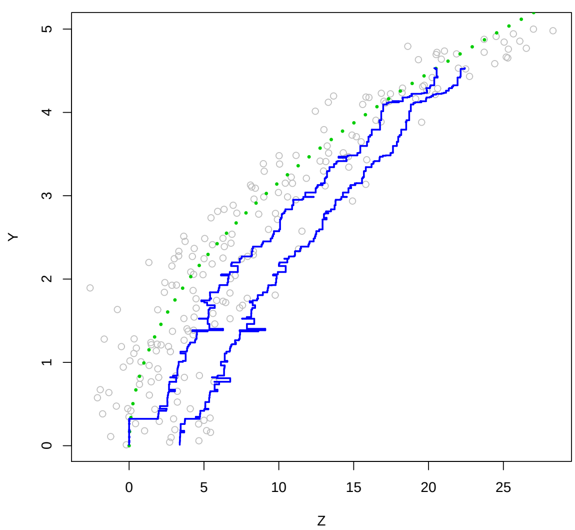

Figure 3: Using the same data as in Figure 1, the two solid lines are the robust m-frontier estimator with $m = 1$ and $m = 50$ respectively.

5 Conclusion and further research

One original idea in this paper is to consider stochastic frontier estimation when the data generating process has an additive noise on the inputs. The noise is not assumed to vanish asymptotically. In this situation, the m-frontier estimator introduced by Cazals, Florens, and Simar (2002) is still a valuable tool in robust frontier estimation, but it requires to plug-in a consistent estimator of the conditional survival function in order to be consistent itself.

Constructing this consistent estimator is a deconvolution problem. We have solved this problem in this paper. An important feature of our results is that the noise level is not known, and therefore needs to be estimated from a cross section of production units.

Measurement errors are frequently encountered in empirical economic data, and the new robust estimator is designed to be consistent in this setting. The rate of convergence of the estimator is however unknown. This study might be of interest for future research in efficiency analysis.

As it was suggested by a referee, one might also be interested in the case where the measurement error is in the output rather than in the input variable. We would like to end this paper by explaining how the above methods can be transferred to this problem and where the limitations are. In this setting, in contrast to Section 4, the inputs X_i are directly observed, but only a contaminated version

$$W_i = Y_i + \eta_i, \qquad \eta_i \sim N(0, \sigma^2) \tag{5.1}$$

of the true output variables Y_i is observed, with η_i independent from X_i and Y_i. Let us briefly discuss the case where both the input and the output spaces are one-dimensional, i.e. $p = q = 1$. As the frontier function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ given in (1.1) is strictly increasing, its inverse function $\varphi^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ exists. The efficiency boundary can be described by either of the functions φ and φ^{-1}. Estimating φ^{-1} is thus equivalent to estimating φ itself. The inverse frontier function can be written as

$$\varphi^{-1}(x) = \inf\{y \in \mathbb{R}_+ \ : \ F_{Y|X \leqslant x}(y) = 1\},$$

where $F_{Y|X\leqslant x}$ denotes the conditional distribution function of Y given $X \leqslant x$. To apply the robust m-frontier methodology we therefore need to estimate the conditional distribution function $F_{Y|X\leqslant x}$. From the model (5.1), one can easily show that the estimation of $F_{Y|X\leqslant x}$ is again a deconvolution problem, and recalling that $F_{Y|X\leqslant x} = 1 - S_{Y|X\leqslant x}$, we can define

$$(\hat{\delta}(n), \hat{\sigma}_n) := \mathop{\arg\min}_{\substack{\delta \in \Delta^{(k_n, D_n)} \\ \sigma \in [0, D_n]}} \gamma(S_\delta, \sigma; \hat{S}_{W|X\leqslant x}) \ \text{ and } \ \hat{F}_n := 1 - S_{\hat{\delta}(n)}$$

in analogy to Section 4.2. $\hat{F}_n$ is the deconvolving estimator of the conditional distribution function $F_{Y|X\leqslant x}$. We proceed by defining the robust m-frontier estimator of φ^{-1} as

$$\hat{\varphi}^{-1}_{m,n}(x) := A - \int_0^A \left\{\hat{F}_n(u)\right\}^m \mathrm{d}u,$$

where $A > 0$ is some constant fixed in advance. Let m_n be a strictly divergent sequence such that $\{\hat{F}_n(\varphi(x))\}^{m_n} \to 1$ almost surely as $n \to \infty$. In analogy to Proposition 4.1, it can be shown that for such a sequence, $\hat{\varphi}^{-1}_{m_n,n}(x)$ is consistent if $A > \varphi^{-1}(x)$. Otherwise, $\hat{\varphi}^{-1}_{m_n,n}(x)$ tends to A almost surely. This suggests the following adaptive choice of A. First, one computes the estimator with some arbitrary initial value of A. If the result is close to A, recompute it repeatedly for increasing values of A until a value smaller than A is obtained.

This estimator is thus robust with respect to noise in the output variable, but note that it is not obvious how to generalize this procedure to a multi-dimensional setting. Moreover, it is not clear how one could cope with a situation with error in both variables. These questions could be subject to further investigation.

A Proofs

A.1 Proof of Theorem 3.2

In order to show the consistency of the robust frontier estimator, we first need to prove two lemmas.

Lemma A.1. *The estimator $(S_{\hat{\delta}(n)}, \hat{\sigma}_n)$ satisfies*

$$\gamma(S_{\hat{\delta}(n)}, \hat{\sigma}_n; \hat{S}^Z_n) \to 0 \quad \text{as } n \to \infty.$$

Proof. By the triangle inequality, we have, for any $(S', \sigma') \in \mathbb{C} \times \mathbb{R}^+$,

$$\begin{aligned} \gamma(S_{\hat{\delta}(n)}, \hat{\sigma}_n; \hat{S}^Z_n) &= \min_{\substack{\delta \in \Delta^{(k_n,D_n)} \\ \tilde{\sigma} \in [0,D_n]}} \gamma(S_\delta, \tilde{\sigma}; \hat{S}^Z_n) \\ &\leqslant \min_{\substack{\delta \in \Delta^{(k_n,D_n)} \\ \sigma \in [0,D_n]}} \gamma(S_\delta, \sigma; S^X \star \phi_\sigma) + \gamma(S^X, \phi_\sigma; \hat{S}^Z_n). \end{aligned} \tag{A.1}$$

Let $\eta > 0$ and $T > 0$ be such that $\int_T^\infty S^X(x)\,dx \leqslant \eta/2$. For n sufficiently large, we have $\sigma \leqslant D_n$ and there is $\delta \in \Delta^{(k_n,D_n)}$ with $\int_0^T |(S_\delta - S^X)(x)|\,dx \leqslant \eta/2$, such that $\int_{\mathbb{R}} |(S_\delta - S^X)(x)|\,dx \leqslant \eta$. It follows that the first term on the right hand side of (A.1) is a null sequence, because

$$\gamma(S_\delta, \sigma; S^X \star \phi_\sigma) \leqslant \|(S_\delta - S^X) \star \phi_\sigma\|_{L^1} \leqslant \|S_\delta - S^X\|_{L^1} \|\phi_\sigma\|_{L^1} \leqslant \eta.$$

The second term is also a null sequence by virtue of Glivenko-Cantelli's and Lebesgue's Theorem. □

Lemma A.2. *The estimator $S_{\hat{\delta}(n)}$ defined by (3.3) satisfies*

$$(P_{\hat{\delta}(n)} \star \phi_{\hat{\sigma}_n}) \xrightarrow{\mathcal{L}} P^Z$$

almost surely as $n \to \infty$.

Proof. The survival function S^Z is continuous everywhere as it can be written as a convolution with some normal density. Therefore, the convergence

$$\hat{S}_n^Z(x) \xrightarrow{n\to\infty} S^Z(x) \qquad \text{a.s.}$$

holds for every $x \in \mathbb{R}$. Hence, by Lebesgue's theorem,

$$\gamma(S^X, \sigma; \hat{S}_n^Z) \xrightarrow{n\to\infty} 0 \qquad \text{a.s.}$$

The triangle inequality, together with Lemma A.1, implies

$$\gamma(S_{\hat{\delta}(n)}, \hat{\sigma}_n; S^Z) \leqslant \gamma(S_{\hat{\delta}(n)}, \hat{\sigma}_n; \hat{S}_n^Z) + \gamma(S^X, \sigma; \hat{S}_n^Z) \xrightarrow{n\to\infty} 0 \qquad \text{a.s.}$$

A continuity argument implies

$$(S_{\hat{\delta}(n)} \star \phi_{\hat{\sigma}_n})(x) \xrightarrow{n\to\infty} S^Z(x) \qquad \text{a.s.}$$

for every $x \in \mathbb{R}$, which is in fact weak convergence and hence concludes the proof. □

Our proof of consistency also needs the following two lemmas. The first one is quoted from Schwarz and Van Bellegem (2010), the second one is an immediate consequence of Lemma 3.4 from the same article.

Lemma A.3. *Let Q_n be a sequence of probability distributions and σ_n a sequence of positive real numbers. Suppose further that $(Q_n \star N(0, \sigma_n))_{n\in\mathbb{N}}$ converges weakly to some probability distribution. Then, there exist an increasing sequence $(n_k)_{k\in\mathbb{N}}$, a probability distribution Q_∞, and a constant $\sigma_\infty \geqslant 0$ such that*

$$Q_{n_k} \xrightarrow{\mathcal{L}} Q_\infty \quad \textit{and} \quad \sigma_{n_k} \to \sigma_\infty$$

as $n \to \infty$.

Lemma A.4. *A weakly convergent sequence of probability distributions that have all their mass on the positive axis has its limit in $\mathcal{P}_0$.*

We are now in position the prove the consistency theorem.

Proof of Theorem 3.2. For probability distributions P, P' and positive real numbers σ, σ', define the distance $\Delta(P, \sigma; P', \sigma') := d(P, P') + |\sigma - \sigma'|$, where $d(\cdot,\cdot)$ denotes the Lévy distance, which metrizes weak convergence. The theorem is hence equivalent to

$$\Delta(P_{\hat{\delta}(n_k)}, \hat{\sigma}_{n_k}; P^X, \sigma) \xrightarrow{n\to\infty} 0$$

almost surely. The proof is obtained by contradiction. Suppose that there is some $d > 0$ and an increasing sequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\Delta(P_{\hat{\delta}_{n_k}}, \hat{\sigma}(n_k); P^X, \sigma) > d$$

for all $k \in \mathbb{N}$.

By Lemma A.2, we know that the distributions given by $(S_{\hat{\delta}(n)} \star \phi_{\hat{\sigma}_n})$ converge almost surely weakly to P^Z. Lemma A.3 implies that there is a distribution P_∞, some $\sigma_\infty \geqslant 0$, and a sub-sequence $(n'_k)_{k\in\mathbb{N}}$ such that almost surely

$$P_{\hat{\delta}(n'_k)} \xrightarrow{\mathcal{L}} P_\infty \qquad \text{and} \qquad \hat{\sigma}_{n'_k} \to \sigma_\infty,$$

which implies the almost sure pointwise convergence of $S_{\hat{\delta}_{n'_k}}$ to S_∞. Fatou's lemma then implies

$$\gamma(S_\infty, \sigma_\infty; S^Z) \leqslant \liminf_{k\to\infty} \gamma(S_{\hat{\delta}(n'_k)}, \hat{\sigma}_{n'_k}; S^Z) = 0 \qquad \text{a.s.},$$

where the last equality holds because of Lemma A.2. Hence, $\gamma(S_\infty, \sigma_\infty; S^Z) = 0$, and using continuity again, we conclude that $S_\infty \star \phi_{\sigma_\infty} = S^X \star \phi_\sigma$. Or equivalently, in terms of distributions, $P_\infty \star \phi_{\sigma_\infty} = P^X \star \mathcal{N}_\sigma$. As all the distributions $P_{\hat{\delta}(n'_k)}$ have their mass on the positive axis, Lemma A.4 implies that $P_\infty \in \mathcal{P}_0$, and hence that $P_\infty = P^X$ and $\sigma_\infty = \sigma$, which is in contradiction to the assumption and concludes the proof. □

A.2 Proof of Proposition 4.1

We begin the proof by plugging-in the sequence m_n into the robust estimator and by splitting up the integral occurring in (4.3) into

$$\int_0^\infty \left\{S_{\hat{\delta}(n)}(u)\right\}^{m_n} \mathrm{d}u = \int_0^{\varphi(y)} \left\{S_{\hat{\delta}(n)}(u)\right\}^{m_n} \mathrm{d}u + \int_{\varphi(y)}^\infty \left\{S_{\hat{\delta}(n)}(u)\right\}^{m_n} \mathrm{d}u =: A_n + B_n$$

with obvious definitions for A_n and B_n. We have that $B_n \to 0$ almost surely as n tends to infinity. To see this, let $t_n := \varphi(y) \vee \sup\{t \in \mathbb{R} \,:\, S_{\hat{\delta}(n)}(t) = 1\}$ and decompose B_n further into

$$\int_{\varphi(y)}^\infty \{S_{\hat{\delta}(n)}(u)\}^{m_n} \,\mathrm{d}u = \int_{\varphi(y)}^{t_n} 1 \,\mathrm{d}u + \int_{t_n}^\infty \{S_{\hat{\delta}(n)}(u)\}^{m_n} \,\mathrm{d}u. \tag{A.2}$$

Firstly, $t_n \to \varphi(y)$ as $n \to \infty$ because of the consistency of $S_{\hat{\delta}(n)}$. Therefore, the first integral on the right hand side of (A.2) tends to 0 as $n \to \infty$. Secondly, $S_{\hat{\delta}(n)}$ is non-increasing and strictly smaller than 1 on (t_n, ∞) for every $n \in \mathbb{N}$. As the sequence $S_{\hat{\delta}(n)}$ is further surely point-wise convergent on $\mathbb{R}$, the other integral of the decomposition in (A.2) also tends to 0.

It remains to show that $A_n \to \varphi(y)$ almost surely as $n \to \infty$. Since $S_{\hat{\delta}(n)}$ is non-increasing and $S_{\hat{\delta}(n)}(0) = 1$, we have that $s_n \leqslant \varphi(y)$. On the other hand, $s_n \geqslant \varphi(y)\,\{S_{\hat{\delta}(n)}(\varphi(y))\}^{m_n}$, which proves the result by virtue of the assumption. □

References

Bigot, J., & Van Bellegem, S. (2009). Log-density deconvolution by wavelet thresholding. *Scandinavian Journal of Statistics*, *36*, 749–763.

Butucea, C., & Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric deconvolution model. *Bernoulli*, *11*, 309–340.

Butucea, C., Matias, C., & Pouet, C. (2008). Adaptivity in convolution models with partially known noise distribution. *Electronic Journal of Statistics*, *2*, 897–915.

Carroll, R., & Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, *83*, 1184–1186.

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

Daouia, A., Florens, J., & Simar, L. (2009). *Regularization in nonparametric frontier estimators* (Discussion Paper No. 0922). Université catholique de Louvain, Belgium: Institut de statistique, biostatistique et sciences actuarielles.

Daskovska, A., Simar, L., & Van Bellegem, S. (2010). Forecasting the Malmquist productivity index. *Journal of Productivity Analysis*, *33*, 97–107.

De Borger, B., Kerstens, K., Moesen, W., & Vanneste, J. (1994). A non-parametric free disposal hull (FDH) approach to technical efficiency: an illustration of radial and graph efficiency measures and some sensitivity results. *Swiss Journal of Economics and Statistics*, *130*, 647–667.

Delaigle, A., Hall, P., & Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, *36*, 665–685.

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

Fan, J. (1991). On the optimal rate of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, *19*, 1257–1272.

Färe, R., Grosskopf, S., & Knox Lovell, C. (1985). *The measurements of efficiency of production* (Vol. 6). New York: Springer.

Hall, P., & Simar, L. (2002). Estimating a changepoint, boundary, or frontier in the presence of observation error. *Journal of the American Statistical Association*, *97*, 523-534.

Johannes, J., & Schwarz, M. (2009). *Adaptive circular deconvolution by model selection under unknown error distribution* (Discussion Paper No. 0931). Université catholique de Louvain, Belgium: Institut de statistique, biostatistique et sciences actuarielles.

Johannes, J., Van Bellegem, S., & Vanhems, A. (2010). Convergence rates for ill-posed inverse problems with an unknown operator. *Econometric Theory*. (forthcoming)

Johnstone, I., Kerkyacharian, G., Picard, D., & Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society Series B*, *66*, 547–573.

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

Leleu, H. (2006). A linear programming framework for free disposal hull technologies and cost functions: Primal and dual models. *European Journal of Operational Research*, *168*, 340–344.

Li, T., & Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, *65*, 139–165.

Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, *16*, 195–211.

Meister, A. (2007). Deconvolving compactly supported densities. *Mathematical Methods in Statistics*, *16*, 195–211.

Meister, A., Stadtmüller, U., & Wagner, C. (2010). Density deconvolution in a two-level heteroscedastic model with unknown error density. *Electronic Journal of Statistics*, *4*, 36–57.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, *98*, 1955–1968.

Park, B. U., Sickles, R. C., & Simar, L. (2003). Semiparametric efficient estimation of AR(1) panel data models. *Journal of Econometrics*, *117*, 279-311.

Park, B. U., Simar, L., & Weiner, C. (2000). The FDH estimator for productivity efficiency scores: asymptotic properties. *Econometric Theory*, *16*, 855–877.

Pensky, M., & Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, *27*, 2033–2053.

Schwarz, M., & Van Bellegem, S. (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, *80*, 236–241.

Seiford, L., & Thrall, R. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometrics*, *46*, 7–38.

Shephard, R. W. (1970). *Theory of cost and production functions*. Princeton, NJ: Princeton University Press.

Simar, L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise? *Journal of Productivity Analysis*, *28*, 183–201.