

NONPARAMETRIC INSTRUMENTAL VARIABLES ESTIMATION FOR EFFICIENCY FRONTIER

Catherine CAZALS*

catherine.cazals@tse-fr.eu

Frédérique FÈVE *

frederique.feve@tse-fr.eu

Jean-Pierre FLORENS *

jean-pierre.florens@tse-fr.eu

Léopold SIMAR *,§

leopold.simar@uclouvain.be.

January 9, 2015

Abstract

The paper investigates endogeneity issues in nonparametric frontier models. It considers a nonseparable model for a cost function $C = \varphi(Y, U)$ where C and Y are the cost and the output, U is uniform in $[0, 1]$ and φ is increasing with respect to U . The cost frontier corresponds to $U = 0$ and U can be interpreted as a normalized level of inefficiency. The endogeneity issue arises when Y is dependent of U . For identification and estimation, we use a nonparametric instrumental variables estimator of the model for fixed value $U = \alpha$, and obtain an estimate of the α -quantile cost frontier $\varphi(Y, \alpha)$. This involves the solution of a non linear integral equation. If the true frontier $\varphi(Y, 0)$ is wanted, it is then estimated by estimating the bias correction $\varphi(Y, 0) - \varphi(Y, \alpha)$ under additional regularity conditions. The procedure is illustrated through a simulated sample and with an empirical application to the efficiency of post offices.

Key Words: Endogeneity in frontier models, Instrumental variable quantile, Non linear integral equation, Landweber iteration, Tail index estimation.

JEL Classification: C14, C26, D24

*Toulouse School of Economics, Toulouse, France. The authors acknowledge Alain Roset and the support from The Group La Poste, Paris, France.

§Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Louvain-la-Neuve, Belgium. Research supported by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy).

1 Introduction

The important literature on production efficiency and frontier analysis has been essentially concentrated on the analysis of conditional distributions. Production models analyze the conditional distribution of the production (the output) given the inputs levels and cost frontier models consider the conditional distribution of the cost given the outputs. In these two cases, environmental variables may also be introduced but there are always treated as additional conditioning variables. This attention to conditional models is verified for parametric and nonparametric (DEA, FDH) models and for both deterministic and stochastic versions of the frontier. Some models consider conditional model to a given value (cost distribution given the output) and others are conditional to an inequality (the distribution of cost given the fact that the outputs are larger than some values). Both are conditional models.

Conditioning to some information is equivalent in the econometric literature to consider this information as exogenous information. Implicitly it is assumed that the process generating the conditioning elements does not contain any relevant information on the parameter of interest and that the conditional model identifies this parameter of interest.

In frontier analysis the exogeneity assumption has a particular meaning. Consider the case of a cost function, the exogeneity assumption means that the level of the outputs of a given firm is generated independently (or mean independently or non correlated) of the level of inefficiency of this firm. This property may be unrealistic in many situations. For instance, consider the case where a manager has to assign the quantity of outputs to produce to different production units. If the manager has some information on the level of inefficiency of each unit this may influence his choice (see e.g. Marschak and Andrews, 1944). Even if the manager is replaced by a population of consumers the demand may go to firms considered as being more efficient by the consumers. We will see below that treatment models allow to understand the mechanism driving endogeneity: in a cost model, the outputs are assigned by randomization but may contain bias selection. The same argument applies in production models where the quantity of inputs may suffer from bias selection. Our results will show that even in models where only inefficiency is present (no noise), DEA and FDH estimators can suffer from endogeneity.

The price to pay to relax the exogeneity assumption is that we need to extend the model. Our strategy belongs to the class of instrumental variables models. We will replace the conditioning of the cost to the outputs by conditioning to observed instruments. We also assume that these instruments explain sufficiently the endogenous variables to guarantee identifiability of the model.

Alternative strategies may be used to address the endogeneity question: we may adopt a control function approach where the introduction of a supplementary variable eliminates the endogeneity or we may adopt a more structural approach by the introduction of an explicit link between endogenous variables and inefficiency (see Simar et al., 2014). However the approach of the latter is quite different, Simar et al. analyze particular models where the endogeneity is introduced by some missing (unobserved) variables characterizing heterogeneity. In separable models the endogeneity problem may be addressed in two ways; by control functions (see Newey et al., 1999) or by instrumental variables (see Darolles et al., 2011). But to the best of our knowledge, the theory for the control function approach is not yet available for estimating quantile functions. In nonseparable models, which is our setup here, identification and estimation may be achieved by using nonparametric instrumental variables (IV) models.

So in this paper we will indeed concentrate our attention to nonseparable models satisfying some instrumental variables conditions. Basically if C is the cost and Y the outputs we analyze models defined by an equation $C = \varphi(Y, U)$ where U has a uniform in $[0, 1]$ distribution independent from some instruments W and where φ is an increasing function of U . So U may be interpreted as a normalized inefficiency and the frontier is then equal to $\varphi(\cdot, 0)$; more generally the α -quantile of C is $\varphi(\cdot, \alpha)$. As explained below, for identification reasons, a direct estimation of $\varphi(\cdot, 0)$ is impossible. Then our strategy will be to estimate $\varphi(\cdot, \alpha)$ for some fixed values of α (this will require the solution of non linear integral equations), and then in a final step, to correct the bias between this quantile and the frontier. To the best of our knowledge, our paper is the first tentative to estimate a nonparametric frontier in the presence of endogeneity by applying nonseparable instrumental models.

Traditionally, parametric cost frontier models have often used models for the expectation of the cost and then by additional assumption on the error term, try to get estimates of the frontier (e.g., COLS, MOLS, etc.). It seems much more natural to concentrate the model on the full distribution of the cost described by its quantile function, small quantiles approaching the cost frontier. In this perspective, nonseparable models are quite natural, because they are based on the fact that a quantile function can be represented by a monotone transformation of a uniform variable U on $[0, 1]$. In addition, in the frontier setup, U is directly interpretable as the inefficiency. Separable models can be considered as a special cases of nonseparable models.

The interest for our model is justified in Section 2 where the endogeneity in frontier models is presented in terms of treatment model. This clarifies what are exactly the issues

of endogeneity in this particular setup. Section 3 is devoted to the nonparametric estimation of the α -quantile using the iterative resolution of a non linear integral equation. The bias correction for the estimation of the true function is analyzed in Section 4. In Section 5 we give some numerical illustrations of our method in a simulated example and in an empirical application. This allows to understand how to implement the estimator and its various components in practice. Section 6 concludes. Some technical details for the asymptotic properties of our estimator are displayed in Appendix A.

2 A treatment model for frontier analysis

One of the difficulties for the econometric analysis of endogeneity and for the understanding of its consequences is related to a correct definition of the parameter of interest. An important progress in this formalization has been realized in the context of treatment models which is based on the concept of counterfactual models, see e.g. Heckman and Vytlačil, (2006). This concept allows the distinction between “fixing” the level of a variable and “conditioning” to the observation of this variable. For example a demand equation considers the reaction of the demand to any possible fixed level of price but we only observe a price generated by a market equilibrium and conditioning the demand by the observed price does not characterize the demand equation. The observation mechanism of the price creates endogeneity. Essentially we will say that we have an endogeneity problem if the parameter or the function of interest is not characterized by the assumed conditional distribution.

To simplify our presentation we consider a univariate cost model explained by a vector of outputs of dimension p . Also for simplification the model does not introduce additional environmental variables. The extensions to production functions is straightforward at least for the case of a single output and multiple inputs. As illustrated below, the concepts of exogeneity or endogeneity are easy to define if we adopt a presentation based on treatment models where the distinction between a counterfactual model and an observed model is introduced. This will be made in three steps: the counterfactual specification, an assignment mechanism and a process generating the observations.

2.1 The counterfactual specification

The first element of this model is the counterfactual specification. Let $\eta \in \mathbb{R}^p$ be a latent vector of the levels of the outputs. This multivariate index is non random and takes its value in all the possible values of the outputs which play here the role of the (continuous and

multivariate) treatment. For each possible value of η there exists a cost, which is a random variable $C_\eta \geq 0$ and the (counterfactual) cost frontier is denoted by φ_η . It is defined as the minimum possible value of C_η . If the distribution of the random C_η is characterized by its cumulative distribution function F_η , or by its survivor function $S_\eta = 1 - F_\eta$, both assumed continuous, we have:

$$\begin{aligned}\varphi_\eta &= \inf\{c | F_\eta(c) > 0\} \\ &= \inf\{c | S_\eta(c) < 1\}.\end{aligned}\tag{2.1}$$

Let us underline that we have defined C_η as the cost corresponding to a value of the outputs exactly equal to η and not larger or equal to η , as is often done in the frontier literature (see Cazals et al., 2002). In the case of monotone frontier these two definitions are equivalent but the first one is necessary for our construction below.

Many models are possible for the definition of the cost distribution. In general the family $\{C_\eta\}_\eta$ may be viewed as a stochastic process indexed by η . We restrict the class of models by considering a single noise model based on the following nonseparable specification

$$C_\eta = \varphi_\eta(U)\tag{2.2}$$

where, for any η , $\varphi_\eta(\cdot)$ is a strictly increasing function of U and where, without loss of generality, U has a uniform distribution between 0 and 1.

We would like to point that this specification can be seen as being restrictive because the distribution of U is identical for any η , even if two units of production receiving two levels of η will have two different drawings of U , but from the same uniform distribution. We may imagine, as often the case in treatment models, more complex models with multivariate source of heterogeneity, like $C_\eta = \varphi_\eta(U_0, U_1, \dots, U_r)$ for some $r > 1$, or models where U is replaced by a process U_η indexed by η . For example if η only takes a value in some finite set $\{1, \dots, K\}$ we may imagine a vector (U_1, \dots, U_K) of noises and a relation $C_\eta = \varphi_\eta(U_\eta)$. We will not pursue these extensions here and we will develop our approach on the simple model (2.2), which stays in the usual framework of nonparametric frontiers, where the heterogeneity has only one random component U .

The specification (2.2) implies obviously that the frontier is given by

$$\varphi_\eta = \varphi_\eta(0).\tag{2.3}$$

We call this function the true frontier of the cost model. More generally we have

$$F_\eta(c) = \varphi_\eta^{-1}(c),\tag{2.4}$$

and the function $\varphi_\eta(\alpha)$ is the α -quantile of the distribution of C_η . The frontier may be viewed as an extreme quantile of order 0 and the random element U may be interpreted as a normalized level of inefficiency. Indeed the value $\varphi_\eta^{-1}(c)$ is the percentage of firms more efficient than the firm with a cost c among the firms producing the outputs η .

The parameter of interest is thus the frontier $\varphi_\eta(0)$ but the α -quantile $\varphi_\eta(\alpha)$ has been shown to be a robust approximation of the frontier, in particular in presence of outliers. In many applications, one may wish to eliminate a percentage of extreme observations in order to get “reasonable” results. This elimination corresponds to a specific choice of α . See e.g. Aragon et al. (2005) and Daouia and Simar (2007) for similar concepts of α -quantile frontiers and their properties from a robustness theory perspective. We will see below that $\varphi_\eta(0)$ is difficult to estimate in presence of endogeneity but that it may be recovered from the $\varphi_\eta(\alpha)$ function by using an approach suggested in Daouia et al. (2010).

As a remark, in the particular case where the model is separable, we may decompose (2.2) as

$$C = \varphi_\eta(0) + g(U), \tag{2.5}$$

with $g(\cdot) \geq 0$. So, in this case C is decomposed into the true frontier plus a positive random element which may be written as a transformation of the uniform distribution. The function g is strictly increasing and such that $g(0) = 0$. Then the α -quantile is equal to $\varphi_\eta(0) + g(\alpha)$ and the difference between this quantile and the true frontier is $g(\alpha)$.

2.2 The assignment mechanism

The second component of the model is an assignment mechanism. In the set up here of cost efficiency, it is the allocation rule of the outputs to the units of production. We denote by $Y \in \mathbb{R}^p$ the assigned level of outputs.

If the random element Y is independent of U , the outputs are exogenously assigned. We will see below that in this case, the model reduces to the conditional distribution of C given Y . If this property is no longer true we say that Y is endogenous. Since we address more general nonseparable models, we will follow the IV approach. The IV model considers a random vector $W \in \mathbb{R}^q$ called the instruments that is assumed to be fully independent of U and we assume that Y is generated conditionally to W and U . We will see later that Y should be sufficiently dependent of W given U to warrant the identification of $\varphi_\eta(\alpha)$. Note that the full independence between W and U is indeed required in the case of nonseparable models.

As a remark we notice that in the particular case of separable models (2.5), the condition $E(g(U)|W) = \text{constant}$ is sufficient to identify $\varphi_\eta(0)$ (up to a constant identified by $\varphi_\eta(0) = 0$ for example) but if g is unknown and if we need to assume that this property holds for any g , we actually assume the full independence.

2.3 The process generating the data

The last element of the construction of the model defines the observations. We assume that we observe an iid sample of W, Y and C_Y (denoted C). This means that we cannot observe C_η for any η but only for the assigned value Y . In the class of models we consider the observations are related by the formulae:

$$C = C_Y = \varphi_Y(U), \quad (2.6)$$

For ease of notations and when necessary, we will also denote the latter as $\varphi(Y, U)$, where as above, $\varphi(\cdot, U)$ is monotone in U . In general, we assume that U and Y are dependent, so Y is endogenous. Note that if Y is independent of U (the exogeneity case), $\varphi(Y, U)$ is the conditional quantile of C given Y at the level U and $\varphi_Y^{-1}(C) = F(C|Y)$. The model reduces then to a usual nonparametric nonseparable model as it has been studied by Chernozukov and Hansen (2005).

Partial exogeneity or Boundary independence

We would like to mention that our concept of exogeneity (full independence between Y and U) may be viewed as being too strong. Indeed let us consider a case where Y and U are not independent but where the minimum value of the support of U given $Y = \eta$ is equal to 0 for any value of η (this may define a “boundary independence”). Then, due to the increasing property of φ with respect to U we have

$$\inf\{C|Y = \eta\} = \inf\{\varphi(\eta, U)|Y = \eta\} = \varphi(\eta, \inf\{U|Y = \eta\}) = \varphi(\eta, 0). \quad (2.7)$$

So, the conditional frontier of C given $Y = \eta$ is equal to $\varphi(\eta, 0)$ even if the conditional quantiles of C given $Y = \eta$ at level $\alpha > 0$ is not equal to $\varphi(\eta, \alpha)$. But this is enough to identify the cost frontier and empirical versions of the $\inf\{C|Y = \eta\}$ (including e.g., FDH or DEA estimators) will still provide consistent estimators of the cost frontier. But of course, as soon as we want to estimate objects defined in the interior of the support of $C|Y = \eta$ (like the robust partial frontiers or the α -quantile frontier defined above), this is no more the case. So, even if we have the boundary independence and (2.7), we need instrumental techniques like the one derived in the next section if we want to estimate α -frontiers.

3 Instrumental variables estimation of α -frontiers

As we have seen in the previous section the estimation problem of α -frontier under endogeneity may be treated by the estimation of a model $C = \varphi(Y, U)$ where U is independent of the instruments W . The function $\varphi(\eta, \alpha)$ then gives this α -frontier that we may also call the α -instrumental quantile. If the true frontier $\varphi(\eta, 0)$ is the parameter of interest a bias correction will be necessary and this will be considered in the next section.

Identification and estimation of the quantile function under endogeneity has been recently treated in several papers, e.g. see Chernozukov et al. (2007), Horowitz and Lee (2007), Gagliardini and Scaillet (2012) and Dunker et al. (2014) among others. We first recall a local identification condition and we present an iterative regularization scheme for the estimation.

3.1 Identification issues

We first need to show how the characterization of the function φ leads to the resolution a non linear integral equation and then we will derive a local identification condition. We first need to introduce some notations. Let us define

$$F(c, \eta|w) = P(C \leq c|W = w, Y = \eta)f_{Y|W}(\eta|W = w), \quad (3.1)$$

where $f_{Y|W}(\cdot|W = w)$ is the conditional density of Y given $W = w$. With some abuse of notation we can interpret this function as $F(c, \eta|w) = P(C \leq c, Y = \eta|W = w)$ where we understand P as a cumulative distribution for C but as a density for Y . Then we can establish the following preliminary result.¹

Lemma 3.1. *For all $\alpha \in [0, 1]$ we have the integral equation*

$$\int F(\varphi(\eta, \alpha), \eta|w)d\eta - \alpha = 0 \quad (3.2)$$

Proof. Since U and W are independent and U is uniform on $[0, 1]$ we have $P(U \leq u|W = w) = u$. The left hand term is equal to $\int P(U \leq u, Y = \eta|W = w)d\eta$. Since for all η , $\varphi(\eta, U)$ is strictly increasing in U we have

$$\begin{aligned} \int P(U \leq u, Y = \eta|W = w)d\eta &= \int P(\varphi(\eta, U) \leq \varphi(\eta, u), Y = \eta|W = w)d\eta \\ &= \int P(C \leq c, Y = \eta|W = w)d\eta, \end{aligned}$$

¹This is rather standard in the above literature but we prefer to be explicit for the unfamiliar reader.

where the least equality is obtained since if $Y = \eta$, $C = \varphi(\eta, U)$. This completes the proof. \square

First note that the equation (3.2) restricted to $\alpha = 0$ cannot identify $\varphi(y, 0)$. Indeed as $F(c, \eta|w) = 0$ for any $c \leq \varphi(\eta, 0)$ all the functions smaller or equal to the true $\varphi(\eta, 0)$ satisfy the condition. This is the reason why we consider (3.2) for any α even if it would be sufficient to consider this equation for $\alpha \in [0, \alpha_0]$, with $0 < \alpha_0 \leq 1$.

In order to elaborate an identification condition, we have to define the functional spaces involved in this relation. For each α we consider the space $L_Y^2(\alpha)$ of square integrable functions $\varphi(\eta, \alpha)$ relatively to the true conditional distribution of Y given $U = \alpha$, i.e.

$$L_Y^2(\alpha) = \{\varphi \mid \int \varphi(z)^2 f_{Y|U}(z|U = \alpha) dz < \infty\}. \quad (3.3)$$

Let us define the following operator $T(\varphi)$

$$T(\varphi)(\alpha, w) = \int F(\varphi(\eta, \alpha), \eta|w) d\eta \quad (3.4)$$

So the integral equation (3.2) can be written for a given α as

$$T(\varphi)(\alpha, w) = \alpha. \quad (3.5)$$

The operator $T(\varphi)$ transforms a function $\varphi \in L_Y^2(\alpha)$ into a function $\psi(w, \alpha) \in L_W^2(\alpha)$, where $L_W^2(\alpha)$ is the space of functions of W and α square integrable with respect to the conditional distribution of W given $U = \alpha$ (actually the latter is equal to the marginal distribution on W). As T is not a conditional expectation operator this property is actually an assumption on F , the joint conditional “distribution” of C, Y given W . We also assume that this distribution has a continuous derivative with respect to C equal to $f_{C,Y|W}(c, \eta|w)$, which is the joint density of (C, Y) given $W = w$. In that case T has a Frechet derivative for any $\tilde{\varphi} \in L_Y^2(\alpha)$ equal to

$$T'_\varphi(\tilde{\varphi}) = \int \tilde{\varphi}(\eta, \alpha) f_{C,Y|W}(\varphi(\eta, \alpha), \eta|w) d\eta \quad (3.6)$$

computed at the function φ . We see that this is a linear operator as a function of $\tilde{\varphi}$ from $L_Y^2(\alpha)$ to $L_W^2(\alpha)$.

The function φ is locally identified if $T'_\varphi(\tilde{\varphi}) = 0$ implies $\tilde{\varphi} = 0$ (see Florens and Sbaï, 2010 and Chen et al., 2014). This means that T'_φ is one to one. Equivalently assuming that

for the true value φ , $\frac{\partial}{\partial \alpha} \varphi(\eta, \alpha) \neq 0$ the Frechet derivative may be rewritten as:

$$\begin{aligned} T'_\varphi(\tilde{\varphi}) &= \int \frac{\tilde{\varphi}(\eta, \alpha)}{\frac{\partial \varphi}{\partial \alpha}(\eta, \alpha)} f_{Y,U|W}(\eta, \alpha|w) d\eta \\ &= \int \frac{\tilde{\varphi}(\eta, \alpha)}{\frac{\partial \varphi}{\partial \alpha}(\eta, \alpha)} f_{Y|W,U}(\eta|w, \alpha) d\eta, \end{aligned} \quad (3.7)$$

where $f_{Y,U|W}(\eta, \alpha|w) = \frac{\partial \varphi}{\partial \alpha}(\eta, \alpha) f_{C,Y|W}(\varphi(\eta, \alpha), \eta|w)$ is indeed the density of Y, U given $W = w$ obtained from the density of C, Y given $W = w$ by simple change of variable. It turns out that this density is nothing else than $f_{Y|W,U}(\eta|w, \alpha)$, the conditional density of Y given $W = w$ and $U = \alpha$, since U and W are independent and U is uniform on $[0, 1]$. So we have the following proposition.

Lemma 3.2. *A sufficient condition for local identification is given by the conditional strong identification or conditional completeness condition*

$$\forall \lambda \in L^2_Y(\alpha) \quad E(\lambda(Y, U)|W = w, U = \alpha) = 0 \Rightarrow \lambda = 0 \quad a.s. \quad (3.8)$$

Proof. By (3.7) we translated the condition that $T'_\varphi(\tilde{\varphi}) = 0$ implies $\tilde{\varphi} = 0$ in terms of the conditional density of $Y|W, U$. A sufficient condition to get this property is indeed the conditional strong identification (or completeness) (3.8) as described in Chapter 5 of Florens et al. (1990). \square

Hereafter, we assume that condition (3.8) is satisfied. This assumption is actually a dependence condition between Y and W (given U). There exists no function of Y orthogonal to functions of W (for fixed α) except the zero function. The previous lemma shows in particular that $T'_\varphi(\tilde{\varphi})$ is the conditional expectation of $\frac{\tilde{\varphi}}{\frac{\partial \varphi}{\partial \alpha}}$ given w and α . This remark proves that T'_φ is a bounded operator and then the problem is locally ill-posed.

3.2 Solving the integral equation

Since solving the integral equation (3.5) in φ is a locally ill-posed inverse problem, we need a regularization method. We will use a regularized recursive method which involves the adjoint operator of T'_φ . In practice we will use estimators of T and of the adjoint operator to provide an estimator of φ . The adjoint operator of T'_φ is denoted T'^*_φ and is the linear operator from $L^2_W(\alpha)$ to $L^2_Y(\alpha)$ characterized by the following equation: for any $\tilde{\varphi} \in L^2_Y(\alpha)$ and any $\tilde{\psi}(w, \alpha) \in L^2_W(\alpha)$ we have

$$\int [T'_\varphi(\tilde{\varphi})(w, \alpha)] \tilde{\psi}(w, \alpha) f_{W|U}(w|U = \alpha) dw = \int \tilde{\varphi}(\eta, x) [T'^*_\varphi(\tilde{\psi})(\eta, \alpha)] f_{Y|U}(\eta|U = \alpha) d\eta,$$

where in our case $f_{W|U}(w|U = \alpha) = f_W(w)$. Then plugging (3.6) in the left hand side term allows to identify the adjoint operator as

$$T'_{\varphi^*}(\tilde{\psi})(\eta, \alpha) = \int \tilde{\psi}(w, \alpha) \frac{f_{C,Y,W}(\varphi(\eta, \alpha), \eta, w)}{f_{Y|U}(\eta|U = \alpha)} dw \quad (3.9)$$

The Landweber algorithm (see Engl et al., 1996 or Kaltenbacher et al., 2008) to solve equation (3.5) is as follows: we start with some starting point $\varphi_0(\eta, \alpha)$ not too far from φ and we form the recurrence relation

$$\varphi_{k+1}(\eta, \alpha) = \varphi_k(\eta, \alpha) + T'_{\varphi_k^*}(\alpha - T(\varphi_k)). \quad (3.10)$$

As the norm of T'_{φ} may be not smaller or equal to 1, (3.10) should introduce a γ factor multiplying T'_{φ^*} . This factor is fixed and we dropout this factor to simplify the formulae. In the simulation and in the estimations presented below, we keep $\gamma = 1$. The algorithm is known to converge to a fixed point of (3.10) which is a solution of (3.5). Since the operators T and T'_{φ^*} are unknown, they will be estimated, this introduce additional noise and so the algorithm has to be stopped at some step k . These practical issues are discussed in details in Fève and Florens (2014). We particularize now their presentation to our setup here.

The estimation of T and T'_{φ^*} are done by usual kernel smoothing method. Let K a standard kernel univariate density and denote \bar{K} the corresponding cumulative distribution function and h_C, h_Y and h_W three bandwidths corresponding to the variables C, Y and W . In order to simplify the presentation and the notations, we assume that W is uniformly distributed in $[0, 1]$ and that C, Y and W are univariate variables.² Then:

$$\begin{aligned} \hat{T}(\varphi) &= \int \hat{F}(\varphi(\eta, \alpha), \eta|w) d\eta \\ &= \int \frac{1}{n} \sum_{i=1}^n \bar{K} \left(\frac{c_i - \varphi(\eta, \alpha)}{h_C} \right) \frac{1}{h_Y} K \left(\frac{\eta - y_i}{h_Y} \right) \frac{1}{h_W} K \left(\frac{w - w_i}{h_W} \right) d\eta. \end{aligned} \quad (3.11)$$

Using standard kernels and bandwidths properties,³ the integral in η may be approximated by

$$\hat{T}(\varphi) = \frac{1}{n} \sum_{i=1}^n \bar{K} \left(\frac{c_i - \varphi(y_i, \alpha)}{h_C} \right) \frac{1}{h_W} K \left(\frac{w - w_i}{h_W} \right). \quad (3.12)$$

²For multivariate Y and W , product kernels could be used with a vector of bandwidths for h_Y and h_W . If W is not uniform, when conditioning on W , we have to divide all the expressions by the kernel density estimate of W .

³One standard argument in kernel estimation is that under regularity conditions of the function g , $\int h^{-1} K((z - z_i)/h) g(z) dz = g(z_i) + O(h)$.

A similar estimator of the adjoint of the derivative is obtained by

$$(\widehat{T}'_{\varphi}(\tilde{\psi}))(\eta, \alpha) = \frac{\int \tilde{\psi}(w, \alpha) \sum_{i=1}^n \frac{1}{h_C} K\left(\frac{c_i - \varphi(\eta, \alpha)}{h_C}\right) \frac{1}{h_Y} K\left(\frac{\eta - y_i}{h_Y}\right) \frac{1}{h_W} K\left(\frac{w - w_i}{h_W}\right) dw}{\sum_{i=1}^n \frac{1}{h_U} K\left(\frac{u_i - \alpha}{h_U}\right) \frac{1}{h_Y} K\left(\frac{\eta - y_i}{h_Y}\right)},$$

where h_U is some appropriate bandwidth for U . We can use the same type of approximation as above for the integral at the numerator. For the denominator we cannot use the kernel in U since the u_i are not observed. Here we use the following approximation as $h_U \rightarrow 0$,

$$\frac{1}{h_U} K\left(\frac{u_i - \alpha}{h_U}\right) = \frac{1}{h_C} K\left(\frac{c_i - \varphi(y_i, \alpha)}{h_C}\right) + o(h_U),$$

with $h_C = \varphi'_y(u)h_U$. This results can be viewed as a version of the delta method for kernels where $h_U \rightarrow 0$, plays the usual role of $\sigma_n^2 \rightarrow 0$ in the delta method (see e.g. Serfling, 1980, Theorem A, Section 3.1). This leads to our final estimator of the adjoint operator

$$(\widehat{T}'_{\varphi}(\tilde{\psi}))(\eta, \alpha) = \frac{\sum_{i=1}^n \tilde{\psi}(w_i, \alpha) K\left(\frac{c_i - \varphi(\eta, \alpha)}{h_C}\right) K\left(\frac{\eta - y_i}{h_Y}\right)}{\sum_{i=1}^n K\left(\frac{c_i - \varphi(y_i, \alpha)}{h_C}\right) K\left(\frac{\eta - y_i}{h_Y}\right)} \quad (3.13)$$

Now for the estimation of φ , we can use the Landweber recursive scheme described above. With our estimators, the algorithm can be written as

$$\hat{\varphi}_{k+1} = \hat{\varphi}_k + \widehat{T}'_{\varphi_k}(\alpha - \widehat{T}(\hat{\varphi}_k)). \quad (3.14)$$

This algorithm needs a starting point φ_0 and a stopping rule. As usual in nonparametric estimation the stopping rule k has contradictory effects on the variance and on the bias: increasing k increases the variance but decreases the bias. The regularization in that case is coming from the stopping rule. The value of k should increase to ∞ with n but at a “slow” rate in order to control the illposedness. The asymptotic properties of the resulting estimator are described in the next section. They imply selection of bandwidths of optimal order, which is the one obtained by using the usual rule of thumb (see below).

Practical implementation

The practical implementation of the algorithm is described in details in Fève and Florens (2014). It is shown there that by discretizing the algorithm and computing the function

$\varphi(y_i, \alpha)$ at the observations $y = (y_1, \dots, y_n)^t$ given by the sample, the iteration scheme can be written very simply in matrix form. Denote $\bar{\alpha} = \alpha i_n$, where i_n is a n -vector of ones, then we have

$$\hat{\varphi}_{k+1}(y, \alpha) = \hat{\varphi}_k(y, \alpha) + B(\varphi_k)(\bar{\alpha} - b(\varphi_k)), \quad (3.15)$$

where $b(\varphi_k) = \hat{T}(\hat{\varphi}_k)(\alpha, w)$ is a $(n \times 1)$ column vector computed for each element w_i , $i = 1, \dots, n$ and $B(\varphi_k)$ is a $(n \times n)$ matrix with element (i, j) given by

$$b_{i,j} = \frac{K\left(\frac{c_j - \hat{\varphi}_k(y_i, \alpha)}{h_C}\right) K\left(\frac{y_i - y_j}{h_Y}\right)}{\sum_{r=1}^n K\left(\frac{c_r - \hat{\varphi}_k(y_r, \alpha)}{h_C}\right) K\left(\frac{y_i - y_r}{h_Y}\right)}.$$

As starting value, we suggest to start with the conditional quantile of order α of C given Y characterized by $F^{-1}(\alpha|Y = \eta)$ where $F(c|\eta)$ is the conditional cumulative distribution function of C given $Y = \eta$. Many estimators of the conditional quantile have been proposed in the literature (see e.g. Van der Vaart, 1998, for the non conditional case). This choice for the starting point is based on the hypothesis that the endogeneity bias is not “too big”. We will see below in the numerical illustration that the algorithm is rather robust to the choice of its starting value. As discussed in Section 5.1 the choice of this initial function seems not essential in our simulation or estimation. As usual in optimization problem this sensitivity to starting value may be checked.

For the estimation, in order to get consistent estimators of the operators, we select the bandwidths to have the appropriate order. For sake of simplicity we fix all the bandwidths to the one given by the usual rule of thumb ($1.06 \hat{\sigma} n^{-1/5}$ where $\hat{\sigma}$ is the empirical standard deviation of the corresponding variable). Here too, we have seen in our examples that the results are robust to the particular choice of the bandwidths.

A more important tuning element to fix is the stopping rule, i.e. the number of iterations. A data driven choice of the stopping rule has been suggested and motivated by Florens and Racine (2013) and by Fève and Florens (2010, 2014). It is based on the analysis of the curve, as a function of k , defined as

$$a(k) = k \|\alpha - \hat{T}(\hat{\varphi}_k)\|^2. \quad (3.16)$$

This function first increases from 0 to a local maximum and then decreases to a minimum. The argument is to define the selected value of k at this minimum. This argument is justified by some intuitive argument explained in details in Fève and Florens (2014). It is also justified by our simulations and we give in Appendix A theoretical justification of this rule.

3.3 Asymptotic properties

The asymptotic properties of our estimator are difficult to establish. The literature on numerical analysis has extensively analyzed estimation by iterated methods for nonlinear inverse problems (see e.g. Kaltenbacher et al., 2008). However this has been done in a slightly different setting where the basic equation to solve is of the type $r = T(\varphi)$, but the operator T is known and the error of estimation is only driven by r .

The econometric literature has been developed for estimating nonparametric instrumental quantiles (see Horowitz and Lee, 2007, Gagliardini and Scaillet, 2012, and Dunker et al., 2014). The type of estimator and the regularization technique are however different. To the best of our knowledge, there are no existing results directly applicable to our setup and a detailed proof for this case would be out of the scope of our paper. However we can give a natural intuitive argument to obtain the desired results.

First we prove the consistency property of our estimator. As in all nonlinear model, the asymptotic analysis is obtained through the approximated linear model, using appropriate Taylor expansions. In our case, the derivations go as follows.⁴ Our model can be written as in (3.5), $\alpha = T(\varphi)$ and can be linearized around the true value of φ we denote by φ_{\dagger} :

$$\alpha - T(\varphi_{\dagger}) = T'_{\varphi_{\dagger}}(\psi), \text{ where } \psi = \varphi - \varphi_{\dagger}. \quad (3.17)$$

This approximation can be viewed a linear model $r = K\psi$ where r and K have to be estimated. We know (see e.g. Carrasco et al., 2006 or Carrasco, 2012) that under regularity conditions, the Landweber estimation of the latter equation converges to ψ (or that for an appropriate choice of k (function of n), $\|\hat{\varphi}_k - \varphi_{\dagger}\| \xrightarrow{P} 0$). Moreover we have

$$\|\hat{\varphi}_k - \varphi_{\dagger}\|^2 = O\left(k n^{-\frac{2s}{2s+q}} + \left(\frac{1}{k}\right)^{\beta}\right), \quad (3.18)$$

where k is the number of iterations, s is the regularity of $F(\varphi(\eta), \alpha), \eta|w$ (in practice, it is the minimum of the order of the kernel used and the order of differentiability of F). Then $n^{\frac{2s}{2s+q}}$ is the rate of convergence achieved for the estimation of $\int F(\varphi(\eta), \alpha), \eta|w d\eta$. This rate assume the usual optimal choice of the bandwidths in the kernel estimates. The parameter $\beta > 0$ accounts for the regularity of φ which is formalized by the condition $\varphi - \varphi_{\dagger} \in \mathcal{R}\left(T'_{\varphi_{\dagger}}^* T'_{\varphi_{\dagger}}\right)^{\beta/2}$ (for details and discussion on this condition see Carrasco et al., 2006 or Darolles et al., 2011, typically β is related to the smoothness of φ).

Furthermore, it is implicitly assumed that T'_{φ}^* and T'_{φ} have been optimally estimated (see Darolles et al., 2011) and that the hypothesis allowing to neglect the estimation of

⁴A sketch of the proof with more details is given in Appendix A.

these operators are satisfied. Then we know that the optimal speed for k is of the order $\left(n^{\frac{2s}{2s+q}}\right)^{1/(\beta+1)}$. This leads to $\|\hat{\varphi}_k - \varphi_{\dagger}\|^2 = O\left(n^{-\frac{2s}{2s+q}}\right)^{\beta/(\beta+1)}$. This rate of convergence can be interpreted as the rate achieved for the nonparametric estimation of $\alpha - T(\varphi_{\dagger})$ reduced by the factor $\beta/(\beta + 1)$ due to the inversion (i.e. solving the integral equation).

4 Estimation of the Full frontier: Bias Correction

So far we have a consistent procedure for estimating the α -quantile frontier $\varphi_{\eta}(\alpha)$, in a nonseparable model with endogeneity. In most application, the researcher will be happy with this because he could prefer a robust estimation of an object less extreme than the full frontier, robust to outliers and extreme data points (see e.g. Aragon et al., 2005, Daouia and Simar, 2007, Daouia et al., 2010, 2012, 2014 for a motivation in the study of quantile frontiers).

However, the researcher may also estimate the full frontier $\varphi_{\eta}(0)$ which, as explained above, is not identified by the integral equation (3.2). The relation between $\varphi_{\eta}(\alpha)$ and $\varphi_{\eta}(0)$ is complex in general, but this can be simplified by assuming an additional assumption.

Assumption 4.1. *We assume that the relation between the α -quantile frontier and the full frontier has the form*

$$\varphi_{\eta}(\alpha) = \varphi_{\eta}(0) + a_{\eta}\alpha^{b_{\eta}}, \quad (4.1)$$

where a_{η} and b_{η} are function of the output levels η .

In fact we will see below that we only need this to be true for values of α in some neighborhood of zero, but the assumption (4.1) will allow to derive consistent estimator of the a_{η} and b_{η} for all η .

Note that the assumption (4.1) is compatible with nonseparable models. In the case of separable models, like in (2.5), we have the additional assumptions that a_{η} and b_{η} are constant with respect to η , which facilitates their estimation.

Remark 4.1. *It should be pointed that the assumption (4.1) is not very restrictive, since it can be viewed as the first element of a Taylor expansion of $\varphi_{\eta}(\alpha)$ around zero. But even more, under the general assumption that, as $c \downarrow \varphi_{\eta}(0)$*

$$F_{\eta}(c) = \ell_{\eta}(c - \varphi_{\eta}(0))^{\rho_{\eta}} + o(c - \varphi_{\eta}(0))^{\rho_{\eta}}, \quad (4.2)$$

with $\ell_{\eta} > 0$ and $\rho_{\eta} > 0$. It is easy to show that this implies

$$\varphi_{\eta}(\alpha) = \varphi_{\eta}(0) + a_{\eta}\alpha^{b_{\eta}} + o(\alpha^{b_{\eta}}), \quad (4.3)$$

where $a_\eta = (\ell_\eta)^{-1/\rho_\eta}$ and $b_\eta = 1/\rho_\eta$. The regularity condition (4.2) is standard in Extreme Value Theory (EVT) and is motivated in details in Daouia et al. (2010, 2012). To summarize, in EVT, a necessary and sufficient condition under which a minimum of an iid sample of a random variable C minus the lower boundary of the support of C , converges in distribution to a non-degenerate distribution, is an extreme value regularity condition involving a slowly varying function and a tail index (here ρ_η). A convenient way to characterize this condition is to consider slowly varying function that can be approximated by a constant (here ℓ_η). This leads to the condition (4.2), where the tail index measures the rate of decline to zero of the cdf $F_\eta(\cdot)$ when it reaches its lower boundary (the frontier point $\varphi_\eta(0)$). In particular if $\rho_\eta = 1$ the density f_η has a jump at the frontier point, as is often assumed in most of the nonparametric frontier models, while ℓ_η, ρ_η are assumed to be known in all the parametric approaches.

The above remark motivates the assumption (4.1) and the estimation of $\varphi_\eta(\alpha)$ for several small values of α may be used to correct the bias $\varphi_\eta(\alpha) - \varphi_\eta(0)$. We may adopt a Pickand's style estimation similar to what has been proposed in Daouia et al.(2010) and (2012).

From (4.1) it is easy to show that for any fixed value $\lambda > 1$

$$\frac{\varphi_\eta(\lambda\alpha) - \varphi_\eta(\alpha)}{\varphi_\eta(\lambda^2\alpha) - \varphi_\eta(\lambda\alpha)} = \lambda^{-b_\eta},$$

which leads to

$$b_\eta = (\log \lambda)^{-1} \log \left(\frac{\varphi_\eta(\lambda^2\alpha) - \varphi_\eta(\lambda\alpha)}{\varphi_\eta(\lambda\alpha) - \varphi_\eta(\alpha)} \right). \quad (4.4)$$

Then it is easy to check that

$$a_\eta = \frac{\varphi_\eta(c\alpha) - \varphi_\eta(\alpha)}{(c\alpha)^{b_\eta} - \alpha^{b_\eta}}. \quad (4.5)$$

Under the assumption (4.1), we obtain consistent estimator of a_η and b_η by plugging in the above formulae the consistent estimators of the corresponding quantile frontier $\hat{\varphi}_\eta(\cdot)$ (see the argument in Daouia et al., 2012). So the resulting estimator of the full frontier is given by

$$\hat{\varphi}_\eta(0) = \hat{\varphi}_\eta(\alpha) - \hat{a}_\eta \alpha^{\hat{b}_\eta}. \quad (4.6)$$

As a final remark, it is well known from EVT, that the estimation of the tail index ρ_η (or of $b_\eta = \rho_\eta^{-1}$) is not easy and requires a large number of data. In our setup here of frontier estimation we could make the assumption that the density of C given η has a jump on the frontier, as in most nonparametric frontier models (implying that $b_\eta = 1$). But a much less restrictive assumption is that the behavior of this density near the frontier is

homogeneous with respect to η , so b_η would be a constant, say b . This implicitly assumes that the α -quantile frontier obeys the equation

$$\varphi_\eta(\alpha) = \varphi_\eta(0) + a_\eta \alpha^b, \quad (4.7)$$

at least for small values of α . We come back to these issues in the numerical illustrations.

5 Numerical illustrations: a simulated example and a real data application

We illustrate how our approach can be implemented through some examples. The first one is a nonseparable simulated data set and the second uses data from post offices of an european operator. Through all the applications in this paper we have chosen the bandwidths by the usual rule of thumb since they give the appropriate orders for getting the desired asymptotic properties. We have done for the simulated example some sensitivity analysis and compare the results when all the chosen bandwidths are divided by a factor 2 and also when all the bandwidths are multiplied by the same factor 2. To save place, we do not reproduce the results, but the fact of using these different bandwidths had almost no consequences on the resulting estimates of the α -quantile frontiers. So this choice seems to be less crucial in our setup here than in the classical problem of density estimation. Of course, more elaborated techniques could be used, e.g. based on least-squares cross validation, but in our numerical examples, it was not necessary.

5.1 Simulated example

The data generating process to produce outputs and costs is constructed in a nonseparable model of the type described in the paper. We generate first an i.i.d. sample of size n for U and W from two independent uniform distributions in $[0, 1]$. The endogenous output Y is thus generated for given $U = u$ and $W = w$ is drawn by a uniform distributions in $[u + w, 0.75u + w + 1]$. Finally we define C by the nonseparable model

$$C = Y^{a_1} + a_2 (\sqrt{Y} + a_3)U,$$

where for the illustration we have chosen $a_j = 2$, for $j = 1, 2, 3$. This DGP provides n values of (C_i, Y_i, W_i) and we fix $n = 500$.

Figure 1 displays the cloud of the n points generated according our model, the true frontier, our frontier estimator for three different (bias corrected) quantiles and the FDH

estimate. The figure gives a clear picture of what happens and we can clearly see the effect of the endogeneity of Y if we use, e.g., traditional estimators, like the FDH. The FDH estimate envelops the cloud of points whereas the true frontier is much more below the cloud, mainly for larger values of the output Y . The endogeneity bias is illustrated by the difference between the FDH and the true frontiers.

For the bias correction we used the simplified model (4.7) (the estimation of individual values b_η would require much more data). For the common \hat{b} we used the least-squares estimator of the regression of the numerator on the denominator appearing in (4.4). With $c = 2$ and $\alpha = 0.05$ we obtain $\hat{b} = 0.8427$ (whereas the true value is 1). Then with this value we computed the values \hat{a}_η at each of the data points by using (4.5). This leads to the bias-corrected frontier estimates given by (4.6) and illustrated in Figure 1 for three particular quantiles $\alpha = 0.05, 0.1, 0.2$.

We will now discuss the sensitivity of the estimate of some α -quantile frontiers to various elements needed for their practical implementation. First we could ask which quantile to use, we expect having better estimates for large values of α , but then the bias correction will be more important. For smaller values of α the bias correction will be easier, but estimating extreme quantiles is always more difficult (by definition, less data are available near the extrema). Figure 2 illustrates the quality of the estimation of the quantile functions for the three particular quantiles $\alpha = 0.05, 0.1, 0.2$, along with their corresponding true values. We see indeed that we have better estimation with bigger α s.

Another important issue is the stopping rule in the Landweber algorithm (3.15). It has been described in (3.16). Figure 3 illustrates the case $\alpha = 0.20$ which leads to the optimal value $k = 110$. To investigate the sensitivity of the final estimator of the α -frontier to the number of iterations, Figure 4 displays (for $\alpha = 0.20$) the starting value $\hat{\varphi}_0(\alpha)$ (the naive conditional quantile) and the resulting estimators after $k = 10, 110, 2500$ iterations. The true 0.20-quantile is also displayed for ease of comparison. We see that the optimal value gives indeed the best choice.

It is also interesting to see if the choice of the starting value in the Landweber algorithm has some importance in our problem. It seems that the final estimator is rather stable with respect to this choice. We analyze this sensitivity for the α -frontier with $\alpha = 0.2$ and Figure 5 provides some interesting pictures. We display there the final value of the resulting estimator along with the value of the starting estimate. We suggest three different strategies, (i) the QC corresponds to the choice suggested in the paper above, i.e. the empirical α -quantile of the conditional of C , given Y (which would be the final solution in case Y would be exogenous)

(ii) the HO choose as starting value the nonparametric α instrumental regression estimate of C on Y and (iii), the strategy “0” is the choice a function equal to the zero, for all the y 's. Interestingly, we can indeed see on Figure 5 that even if the three starting values are quite different, the final estimator are rather similar, except for some minor divergences on the right part of the display, due to some edge effect (no so much observations there). This is rather good news.

A final point is somewhat different, it address the natural question if the instrumental variable estimator introduce some spurious noise in the estimation process when in fact there is no endogeneity. To investigate this issue we compare the results obtained for the α -frontier by using the naive quantile estimator (which is biased in case of endogeneity) with our instrumental variable estimator in two cases: where we have endogeneity (in the simulated case described above) and in the case where in fact Y is exogenous (independent of U). Here, for $W = w$, we simulate the values of Y as uniform on the interval $[w, w + 1]$. The results are displayed in the two panels of Figure 6 and are very encouraging. First, we see in the left panel (Y is exogenous), that our instrumental estimator behaves very well and is similar to the consistent naive quantile estimator. Second, the right panel, illustrates again clearly the bias introduced by endogeneity: the naive quantile is inconsistent and the instrumental estimator is as described above, quite similar to the true quantile function.

As a conclusion of this numerical illustration, we see that the implementation of the procedure is not so complicated, the iterations are given in matrix notation in (3.15), that the determination of an optimal value of k works rather well, and that the starting value of the algorithm is not so crucial. We also have seen that the procedure does not introduce spurious bias, in case the outputs Y are in fact exogenous. The more delicate part of the process is when the full frontier $\varphi(\eta, 0)$ is wanted. In this case, unless the sample size is huge, we need some additional regularity condition on the model, like (4.7) to get sensible estimator of the tail index.

5.2 Cost efficiency of Post Offices

We illustrate now our approach with the analysis of data on post offices (Delivery Offices, DO) of a French operator. To facilitate the illustration we consider one output Y which is an aggregate of the different activities of post offices. The cost C is measured by the number of hours of work. We take as single instrument W , the population of the sector of post offices. We discuss the endogeneity issue and the choice of the instruments below.

As in the previous simulated case, we have estimated the α -quantile for $\alpha = 0.05, 0.1$

and 0.2 by using our instrumental variable approach for nonseparable models. We used for the illustration a random sample of $n = 300$ offices coming from a more complete data set. Figure 7 displays the main results. To obtain the bias correction, we assume, as in the simulated case, the hypothesis (4.7) getting estimates of b ($\hat{b} = 0.3658$) and of a_η for all η . The determination of the number of iterations in the Landweber algorithm provided a similar picture as the one displayed for the simulated case leading for the three selected quantiles to the optimal values $k = 2854, 3180$ and 3199 respectively.

Looking to Figure 7 we can compare our bias corrected frontier estimates with the traditional FDH estimate of the frontier, we see that there is a serious bias of the latter due to the endogeneity. This endogeneity may be explained by a negative dependence between the inefficiency level and the quantity of outputs. A delivery office which deliver an important amount of mail may organize its production process in a more efficient way. The estimated frontier corrected by instrumental variables shows the importance of fixed costs for small units. The population of the corresponding sector is a natural instrument: it is related to the level of the output and it is natural to assume that the inefficiency component is independent of the population.

Table 1 gives the cost, output and the efficiency scores of 20 randomly chosen offices in the sample. As can be seen from the table, the bias of endogeneity of the FDH estimator gives too optimistic measures of efficiency defined for unit i as the ratio c_i^∂/c_i , where c_i^∂ is the corresponding frontier estimate for the unit i . This is the case here because most of FDH frontier estimates (which envelops the cloud of points), are much higher than our instrumental estimator of the frontier, as seen in Figure 7.

As a final remark, we also estimated a separable model with the same data, i.e. assuming the model (2.5). This model has the advantage of simplifying the estimation of the bias correction, because the assumption (4.7) is still true but with $a_\eta = a$, being a constant. The corresponding results are displayed in Figure 8. We see that the results are reasonably similar but still different. Are these differences significant? This advocates the need for deriving some testing procedure for testing the separability assumption which facilitates the estimation of the full frontier. This is kept for future researches.

6 Conclusion

The paper investigates endogeneity issues in nonparametric frontier models. We have chosen an input orientation where we want to estimate a cost frontier. We show in a treatment effect

model how endogeneity can be introduced, linking the output level Y and the inefficiency term U .

We show that nonseparable models are natural in this framework, so we have the cost function $C = \varphi(Y, U)$ where C and Y are the cost and the output, U is uniform in $[0, 1]$ and φ is increasing with respect to U . The cost frontier corresponds to $U = 0$. The endogeneity issue arises when Y is dependent of U and our results have shown that even in models where only inefficiency is present (no noise), DEA and FDH estimators can suffer from endogeneity.

For identification and estimation, we use a nonparametric instrumental variables estimator of the model for fixed value $U = \alpha$, with an instrument W independent of U but dependent of Y . We obtain an estimate of the α -quantile cost frontier $\varphi(Y, \alpha)$. This involves the solution of a non linear integral equation, and we use the Landweber iteration method. We provide the rates of convergence and practical guidelines for the implementation of the algorithm.

If the true frontier $\varphi(Y, 0)$ is wanted, we have to estimate the bias correction $\varphi(Y, 0) - \varphi(Y, \alpha)$ and this can be achieved under an additional regularity condition. The procedure is illustrated through a simulated sample and with an empirical application to the efficiency of post offices.

Many complementary questions may be addressed. Among others we may quote the asymptotic distribution of the quantile estimators and the properties of the bootstrap in this case, deriving a test for separability and a test of the exogeneity assumption.

In particular, one may also ask if the full independence between U and W is really needed, as we did in this paper. We wonder under which assumption, a boundary independence between U and W , like

$$\min\{U|W = w\} = 0, \text{ for any } w,$$

would be sufficient for the estimation of the frontier. Some preliminary experiments in a separable cost function example were encouraging and the idea would be an appealing extension of our results here. The theory of the identification and of the estimation under this weaker hypothesis is certainly on our research agenda.

A Appendix: Asymptotics properties

In this appendix, we consider the problem

$$T(\varphi) = \alpha, \text{ where } T(\varphi) = \int F(\varphi(\eta, \alpha), \eta|w) d\eta, \quad (\text{A.1})$$

and we have the two sequences of approximations defined as

$$\varphi_k = \varphi_{k-1} + T'_{\varphi_{k-1}}(\alpha - T(\varphi_k)) \quad (\text{A.2})$$

$$\hat{\varphi}_k = \hat{\varphi}_{k-1} + \hat{T}'_{\hat{\varphi}_{k-1}}(\alpha - \hat{T}(\hat{\varphi}_k)) \quad (\text{A.3})$$

The elements of (A.3) are all related to a sample of size n but they are not indexed by n to simplify the notations. We can prove the following convergence result.

Proposition A.1. *Under the assumptions*

(i) *There exists a unique solution to equation (A.1) and the algorithm (A.2) converge in norm to φ , i.e., $\|\varphi_k - \varphi\| \rightarrow 0$ if $k \rightarrow \infty$.*

(ii) *\hat{F} and \hat{f} converge in probability in L^2 -norm to F and f respectively.*

We have

$$\|\hat{\varphi}_k - \varphi\| \rightarrow 0, \text{ in probability.} \quad (\text{A.4})$$

Proof. Let $\varepsilon > 0$. We have

$$\|\hat{\varphi}_k - \varphi\| \leq \|\hat{\varphi}_k - \varphi_k\| + \|\varphi_k - \varphi\|.$$

Using (i), there exists $k(\varepsilon)$ such that for $k \geq k(\varepsilon)$ we have $\|\varphi_k - \varphi\| < \varepsilon$.

For any k larger than $k(\varepsilon)$ and for all ζ , there exists $n(\varepsilon, k, \zeta)$ such that for all $n \geq n(\varepsilon, k, \zeta)$ we have $\text{Prob}(\|\hat{\varphi}_k - \varphi_k\| > \varepsilon/2) < \zeta$. This property is satisfied because for fixed k , $\hat{\varphi}_k$ is a continuous transformation of \hat{F} and \hat{f} assumed to be consistent.

Then we have $\forall \varepsilon, \exists k(\varepsilon)$ such that $\forall k > k(\varepsilon)$ and $\forall \zeta, \exists n(\varepsilon, k, \zeta), \text{Prob}(\|\hat{\varphi}_k - \varphi\| > \varepsilon) < \zeta$, which completes the proof. \square

We can discuss the assumption (i). It has been shown (see Kaltenbacher et al., 2008, Theorem 2.4) that this property is true if for all $\varphi, \bar{\varphi}$ in a neighborhood of the starting point φ_0 we have

$$\|T(\varphi) - T(\bar{\varphi}) - T'_{\varphi}(\varphi - \bar{\varphi})\| \leq \theta \|T(\varphi) - T(\bar{\varphi})\|,$$

where $\theta < 1/2$ and if $\|T'_\varphi\| \leq 1$. As remarked in Section 3.2, this second property may be obtained through a multiplication of T'_φ by a suitable γ factor. So the assumption is quite similar to the usual assumptions made for justifying Newton-Raphson iterations techniques.

We now give a sketch of the proof for the rate of convergence for the linearized version of our model. Using the notations introduced in Section 3, the model to solve may be written as $\alpha = T(\varphi)$. It can be linearized around the true value φ_\dagger as

$$\alpha - T(\varphi_\dagger) = T'_{\varphi_\dagger}(\psi), \text{ where } \psi = \varphi - \varphi_\dagger.$$

This approximation can be viewed as a linear model $r = K \psi$ where r and K have to be estimated. Denote \hat{r} and \hat{K} these estimators. The Landweber solution of the estimated equation can be written as

$$\hat{\psi}_k = \sum_{j=0}^{k-1} (I - \hat{K}^* \hat{K})^j \hat{K}^* \hat{r}, \quad (\text{A.5})$$

where k is the number of iterations. We have the following result

Proposition A.2. *Under regularity conditions given in the proof, we have*

$$\|\hat{\psi}_k - \psi\|^2 = O\left(\left(n^{-\frac{2s}{2s+q}}\right)^{\beta/(\beta+1)}\right), \quad (\text{A.6})$$

where q is the dimension of W , s characterizes the regularity of $F(c, y|w)$ and β , the regularity of the function φ (more details on s and β are given in the proof).

Proof. Adding and subtracting terms from (A.5) we obtain

$$\begin{aligned} \hat{\psi}_k - \psi &= \sum_{j=0}^{k-1} (I - \hat{K}^* \hat{K})^j \hat{K}^* (\hat{r} - \hat{K} \psi) \\ &\quad + \left[\sum_{j=0}^{k-1} (I - \hat{K}^* \hat{K})^j \hat{K}^* \hat{K} - \sum_{j=0}^{k-1} (I - K^* K)^j K^* K \right] \psi \\ &\quad + \sum_{j=0}^{k-1} (I - K^* K)^j K^* K - \psi, \\ &= \mathcal{J}_{1n} + \mathcal{J}_{2n} + \mathcal{J}_{3n}. \end{aligned}$$

Clearly, for the first term we have

$$\|\mathcal{J}_{1n}\| \leq \left\| \sum_{j=0}^{k-1} (I - \hat{K}^* \hat{K})^j \hat{K}^* \right\| \times \|\hat{r} - \hat{K} \psi\|,$$

where we know (see Engl et al., 1996) that $\left\| \sum_{j=0}^{k-1} (I - \hat{K}^* \hat{K})^j \hat{K}^* \right\| = O(\sqrt{k})$. Since $\|\hat{r} - \hat{K} \psi\| = \|\hat{T}(\varphi) - T(\varphi)\|$ and the last difference is given by $\int (\hat{F}(\varphi(\eta, \alpha), \eta|w) -$

$F(\varphi(\eta, \alpha), \eta|w))d\eta$, under suitable choice of the order of bandwidths (like the one we use in practice), we have $\|\hat{r} - \hat{K}\psi\| = O(n^{-s/(2s+q)})$, where s is the minimum between the order of the kernel (generally, 2) and the regularity of F (order of differentiability). So $\|\mathcal{J}_{1n}\| = O(\sqrt{k}n^{-s/(2s+q)})$.

From Carrasco et al. (2006), Proposition 3.14, it is seen that $\|\mathcal{J}_{2n}\| = O(k)\|\hat{K}^*\hat{K} - K^*K\|(1/k)^{\beta/2}$, where $\beta > 0$ characterize the regularity of the function φ . More precisely, it is formalized by the ‘‘source condition’’ $\varphi - \varphi_{\dagger} \in \mathcal{R}\left(T'_{\varphi_{\dagger}}{}^*T'_{\varphi_{\dagger}}\right)^{\beta/2}$ (for details and discussion on this condition see Carrasco et al., 2006 or Darolles et al., 2011). Since $\|\hat{K}^*\hat{K} - K^*K\|$ has a rate given by the rate of convergence of the density estimate of $f_{C,Y,W}(\varphi(\eta, \alpha), \eta, w)$, we obtain, again under suitable choice of the bandwidths, an order $O(n^{-s/(2s+p+q+1)})$. So we end up with $\|\mathcal{J}_{2n}\| = O((1/k)^{\beta/2-1}n^{-s/(2s+p+q+1)})$.

Finally, the third term has an order given by $\|\mathcal{J}_{3n}\| = O((1/\sqrt{k})^{\beta})$ (see Engl et al., 1996 or Proposition 3.11 in Carrasco et al., 2006).

Putting all these results together and considering the squares of the norms we have

$$\|\hat{\psi} - \psi\|^2 = O\left(kn^{-2s/(2s+q)} + (1/k)^{\beta-2}n^{-2s/(2s+p+q+1)} + (1/k)^{\beta}\right). \quad (\text{A.7})$$

Now, the approach is rather standard, we first define k_0 , the value of k that balances the first and the third term of (A.7). After simple analytical manipulations, we find that $k_0 = O\left((n^{2s/(2s+q)})^{1/(\beta+1)}\right)$. Now we have to check that with this value k_0 the central term of (A.7) is negligible. This will provide an additional condition on β : we must have $\beta + 1 \geq 2 \times \left(\frac{2s+p+q+1}{2s+q}\right)$, which assumes that φ is smooth enough to control the term \mathcal{J}_{2n} and is rather standard in this literature (see e.g. Darolles et al., 2011). With this k_0 we obtain the final order $O\left((n^{-2s/2s+q})^{\beta/(\beta+1)}\right)$ which completes the proof. \square

Finally, we can prove the properties of our stopping rule for the linearized version of our model. We minimize $k\|\alpha - \hat{T}(\hat{\varphi}_k)\|^2$ but we consider the linear approximation of our criterion, given by $k\|(\alpha - \hat{T}(\varphi_{\dagger})) - \hat{T}'^*_{\varphi_{\dagger}}(\psi_k)\|^2$ where as above $\psi = \varphi - \varphi_{\dagger}$. This can be denoted as in the preceding proof as $k\|\hat{r} - \hat{K}\psi_k\|^2$, or equivalently

$$\begin{aligned} & k\|\hat{r} - \hat{K}\sum_{j=0}^{k-1}(I - \hat{K}^*\hat{K})^j\hat{K}^*\hat{r}\|^2 = k\|(\hat{r} - \hat{K}\psi) + \hat{K}(I - \sum_{j=0}^{k-1}(I - \hat{K}^*\hat{K})^j\hat{K}^*)\hat{K}\psi\|^2 \\ & = O\left(k\|\hat{r} - \hat{K}\psi\|^2 + k\left\|\left(\hat{K}(I - \sum_{j=0}^{k-1}(I - \hat{K}^*\hat{K})^j\hat{K}^*\hat{K}) - K(I - \sum_{j=0}^{k-1}(I - K^*K)^jK^*)\right)\psi\right\|^2\right. \\ & \quad \left.+ k\left\|K(I - \sum_{j=0}^{k-1}(I - K^*K)^jK^*)K\psi\right\|^2\right). \end{aligned}$$

Under the assumptions of Proposition A.2 and under the conditions given after (A.7) this term is

$$O\left(k\left(n^{\frac{-2s}{2s+q}} + (1/k)^{\beta+1}\right)\right).$$

The last term is the regularization bias of $K\psi$ which belongs to $(T'_{\varphi_{\dagger}} T'_{\varphi_{\dagger}})^{(\beta+1)/2}$. The minimum of this function gives k proportional to the optimal rate.

References

- [1] Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005), Nonparametric frontier estimation: A conditional quantile-based approach, *Econometric Theory* 21, 358–389.
- [2] Bakunshinskii, A. B. and Kokurin, M.Y. (2004), *Iterative Methods for Approximate Solution of Inverse Problems*, Springer, Dordrecht.
- [3] Carrasco, M. (2012), A regularization approach to the many instruments problem, *Journal of Econometrics*, 170, 383–398.
- [4] Carrasco, M., Florens, J.-P. and Renault E. (2006), Linear Inverse Problems in Structural Econometrics: Estimation based on Spectral Decomposition and Regularization, in *Handbook of Econometrics*, Heckman, J.J. and E.E. Leamer, eds., volume 6B, North Holland.
- [5] Cazals, C. Florens, J.-P. and Simar L. (2002), Nonparametric Frontier Estimation: A robust Approach, *Journal of Econometrics*, 106, 1-25.
- [6] Chen, X., Chernozhukov, V., Lee, S. and Newey, W. K. (2014), Local Identification of Nonparametric and Semiparametric Models, *Econometrica*, 82(2), 785–809.
- [7] Chernozhukov, V. and Hansen, C. (2005), An IV Model of Quantile Treatment Effects, *Econometrica*, 73, 245–261.
- [8] Chernozhukov, V., Imbens, G.W., and Newey, W. K. (2007), Instrumental Variable Estimation of Nonseparable Models, *Journal of Econometrics*, 139, 4–14.
- [9] Daouia, A., Florens, J.-P. and Simar L. (2010), Frontier Estimation and Extreme Value Theory, *Bernoulli*, 16, 1039–1063.
- [10] Daouia, A., Florens, J.-P. and Simar L. (2012), Regularization of Nonparametric Frontier Estimators”, *Journal of Econometrics*, 168, 342–354.
- [11] Daouia, A. and L. Simar (2007), Nonparametric efficiency analysis: a multivariate conditional quantile approach, *Journal of Econometrics*, 140, 375–400.
- [12] Daouia, A., Simar, L. and P.W. Wilson (2014), Measuring Firm Performance Using Nonparametric Quantile-type Distances. Discussion paper 2013/14, ISBA, UCL, to appear *Econometric Review*.

- [13] Darolles, S., Fan, Y., Florens, J.-P. and Renault E. (2011), Nonparametric Instrumental Regression, *Econometrica*, 79, 1541–1565.
- [14] Dunker, F., Florens, J.-P., Hohage, T., Johannes, J. and Mammen, E. (2014), Iterative Estimation of Solutions to Noisy Nonlinear Operator Equations in Nonparametric Instrumental Regression, forthcoming in *Journal of Econometrics*, 178(3), 444–455.
- [15] Engl, H. W., Hanke, M., and Neubauer, A. (1996), *Regularization of Inverse Problems*, Kluwer Academic Publisher, Dordrecht, Boston, London.
- [16] Fève, F., Florens J.-P. (2010), The practice of nonparametric estimation by solving inverse problems. The example of transformation models, *Econometric Journal*, 13, 1–27.
- [17] Féve, F., Florens J.-P. (2014), Iterative Algorithm for Nonparametric Estimation of the Instrumental Variables Quantile, *Economics Letters*, 123, 300–304.
- [18] Florens, J.-P., Mouchart, M. and J.-M. Rolin (1990), *Elements of Bayesian Analysis*, Marcel Dekker, New-York.
- [19] Florens, J.-P. and Racine, J.,(2013), Nonparametric Instrumental Derivatives, mimeo TSE.
- [20] Florens, J.-P. and Sbaï E. (2010), Local Identification in Empirical Games of Incomplete Information, *Econometric Theory*, 26, 1638–1662.
- [21] Gagliardini, P., Scaillet, O. (2012), Nonparametric Instrumental Variables Estimators of Structural Quantile Effects, *Econometrica*, 80(4), 1533–1562.
- [22] Heckman, J.J. and E. Vytlacil (2006), Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation, in *Handbook of Econometrics*, Heckman, J.J. and E.E. Leamer, eds., volume 6B, North Holland.
- [23] Horowitz, J. L. and Lee, S. (2007), Nonparametric Instrumental Variables Estimation of a Quantile Regression Model, *Econometrica*, 75, 1513–1518.
- [24] Imbens, G. and Newey, W. (2009), Identification and Estimation of Triangular Simultaneous Equations Models without Additivity, *Econometrica*, 77, 1481–1512.
- [25] Kaltenbacher, B., Neubauer, A. and Scherzer, O. (2008), Iterative Regularization Methods for Nonlinear Ill-posed Problems, *Radon Series on Computational and Applied Mathematics* de Gruyter, Berlin.

- [26] Marschak, J. and Andrews, W. (1944), Random simultaneous equations and the theory of production, *Econometrica*, 12, 143–205.
- [27] Newey, W. Powell, J. and Vella F. (1999), Nonparametric Estimation of Triangular Simultaneous Equations Models, *Econometrica*, 67, 565–604.
- [28] Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.
- [29] Simar, L., Vanhems, A. and Van Keilegom, I. (2014), Unobserved Heterogeneity and Endogeneity in Nonparametric Frontier Estimation, *Discussion paper Institut de Statistiques*, Université de Louvain.
- [30] Torgovitsky, A. (2012), Identification of Nonseparable Models with General Instruments. *Preprint*.
- [31] Van der Vaart, A.W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Office	Cost	Output	IV-eff	FDH
1	1.00	1.00	0.0998	0.6091
2	0.20	0.38	0.4560	0.8299
3	0.86	2.36	0.2443	0.6304
4	0.72	1.36	0.1599	0.7356
5	0.46	0.62	0.2037	0.7267
6	0.42	0.58	0.2209	0.6830
7	1.48	2.46	0.1522	0.6806
8	1.05	1.65	0.1273	0.5819
9	0.40	0.74	0.2380	0.7232
10	1.22	3.05	0.3045	0.5107
11	0.30	0.55	0.3039	0.6757
12	0.92	2.04	0.1825	0.5982
13	3.66	8.81	0.7132	0.7486
14	1.23	1.54	0.1022	0.5067
15	0.37	0.62	0.2485	0.7656
16	0.15	0.13	0.6288	1.0000
17	1.15	3.14	0.3525	0.5411
18	2.06	3.92	0.3523	0.6562
19	0.97	1.91	0.1597	0.6262
20	0.71	1.48	0.1699	0.7385

Table 1: *Cost efficiency scores for 20 randomly selected offices. A value “1” indicates a unit on the cost frontier. In order to keep the confidentiality of the data, cost and output are normalized to 1 for the first office.*

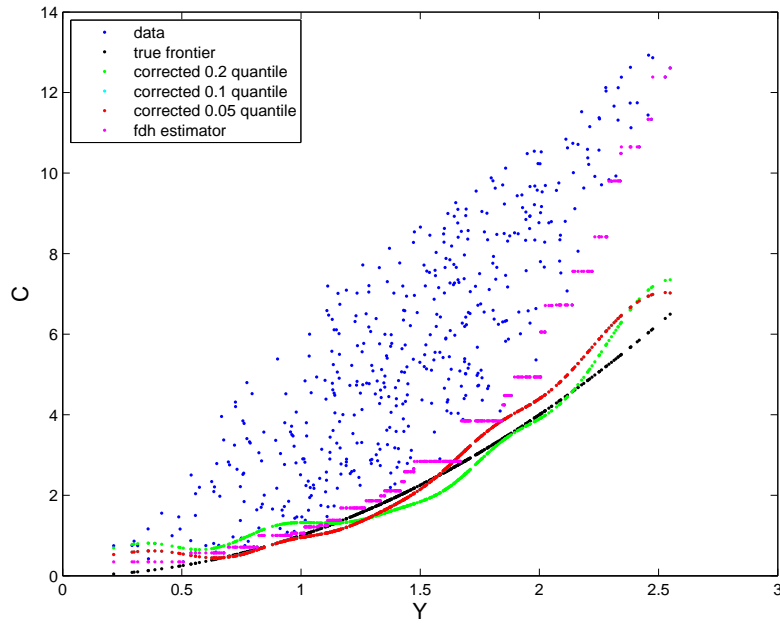


Figure 1: *Simulated example. Final results and FDH estimate. The true is the convex curve in black dots. The FDH is the stair case function and the bias-corrected quantiles are the wiggled curves. The results for $\alpha = 0.1$ are hidden by the curve for $\alpha = 0.05$. The estimator for $\alpha = 0.05$ is in red dots staying below the one corresponding to $\alpha = 0.2$ (green dots) for the first left part of the picture, then passing above at the right part.*

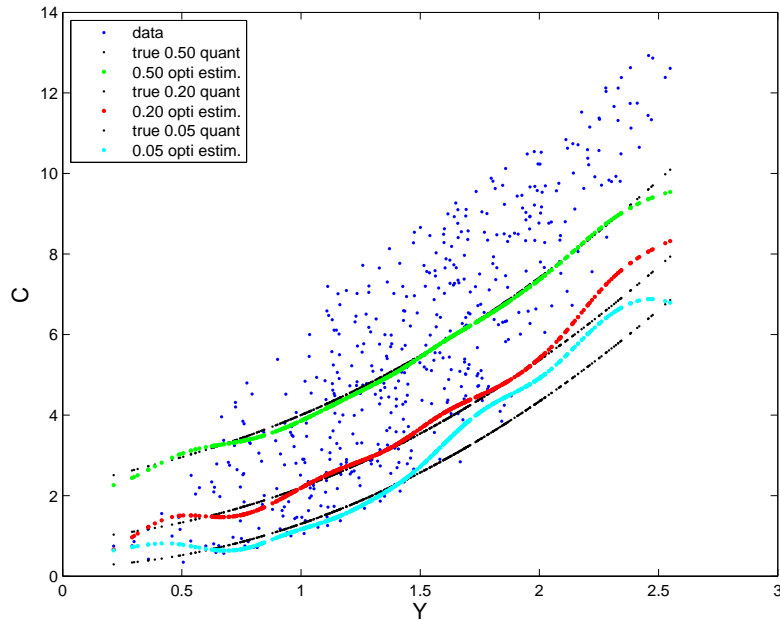


Figure 2: *Simulated example. Estimation of α -frontiers with the true ones.*

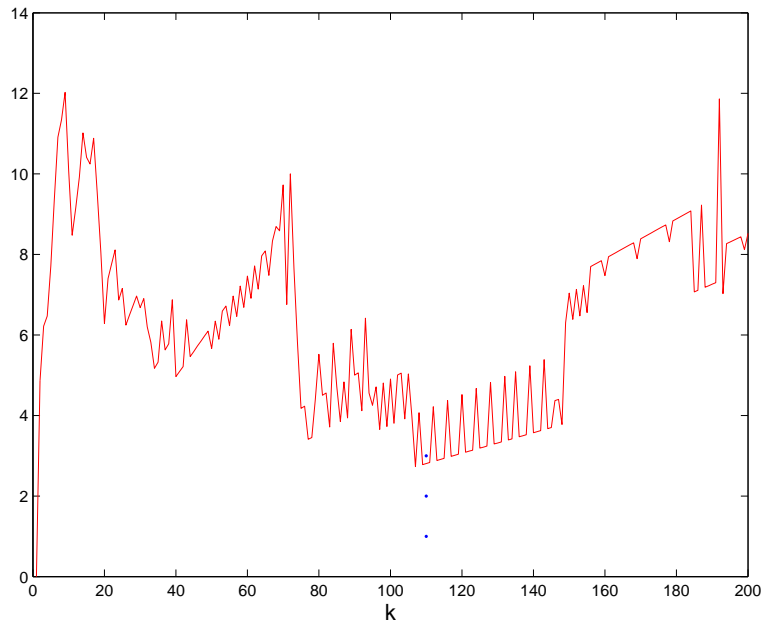


Figure 3: *Simulated example. Criterion for the stopping rule for $\alpha = 0.2$. The optimal value is selected as $k = 110$.*

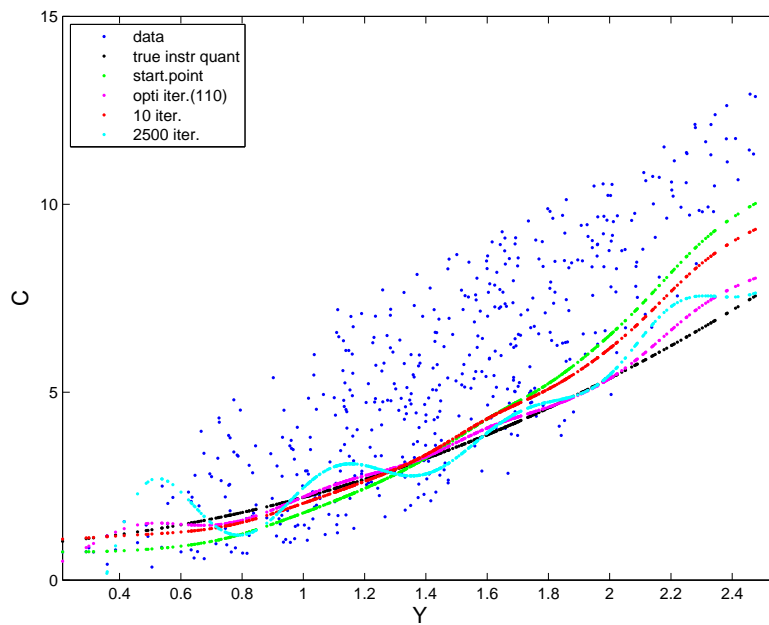


Figure 4: *Simulated example. Sensitivity of the results to k for $\alpha = 0.2$, the number of iterations in the Landweber algorithm. The optimal value is $k = 110$.*

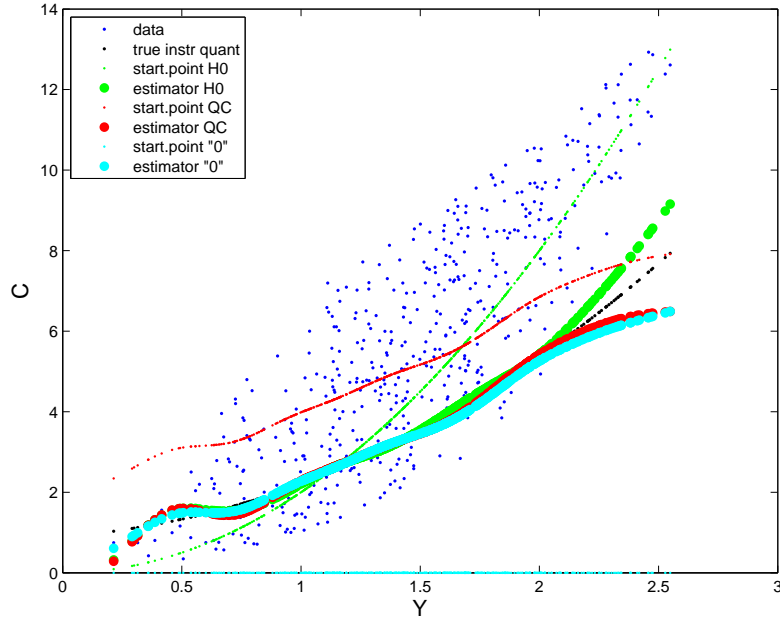


Figure 5: *Simulated example. Sensitivity of the results to the starting value for $\alpha = 0.2$. Starting values for HO is the nonparametric instrumental regression estimate, for QC is the conditional quantile of $C|Y$ and "0" is the value zero.*

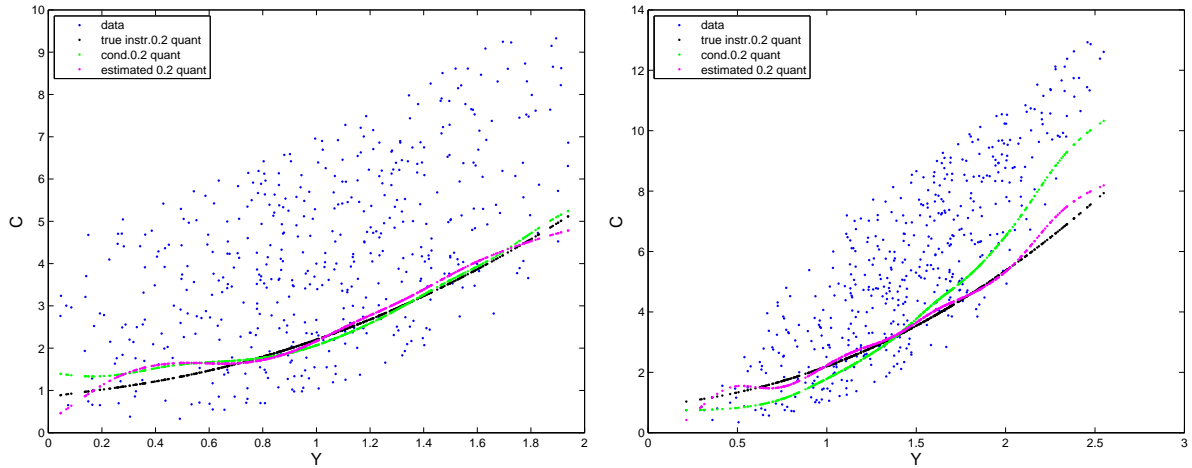


Figure 6: *Estimation of the α -frontier for $\alpha = 0.2$ by using the instrumental variable estimator and by using the empirical quantile of $C|Y$. Left panel, Y exogenous, right panel Y endogenous.*

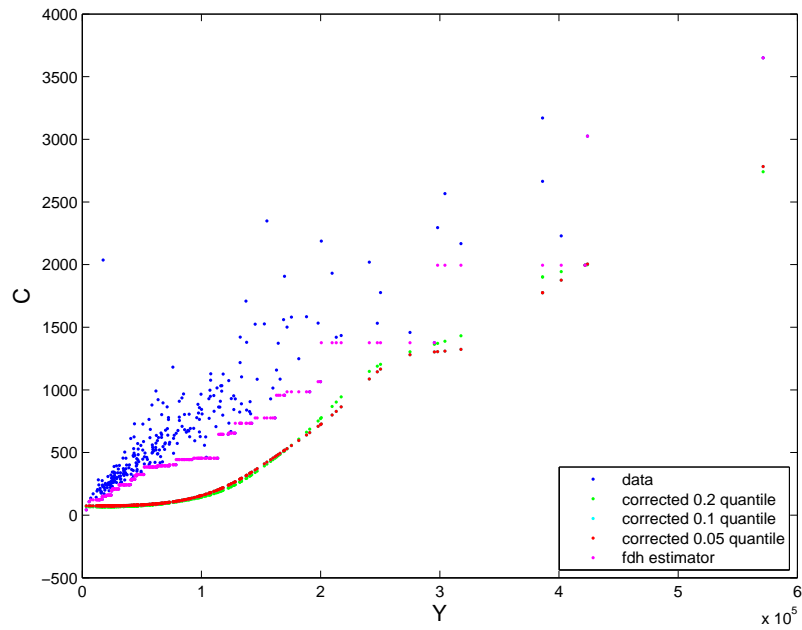


Figure 7: *Data on post offices. Final results for the nonseparable model and the FDH estimate. The bias-corrected quantiles are for $\alpha = 0.05, 0.1$ and 0.2 .*

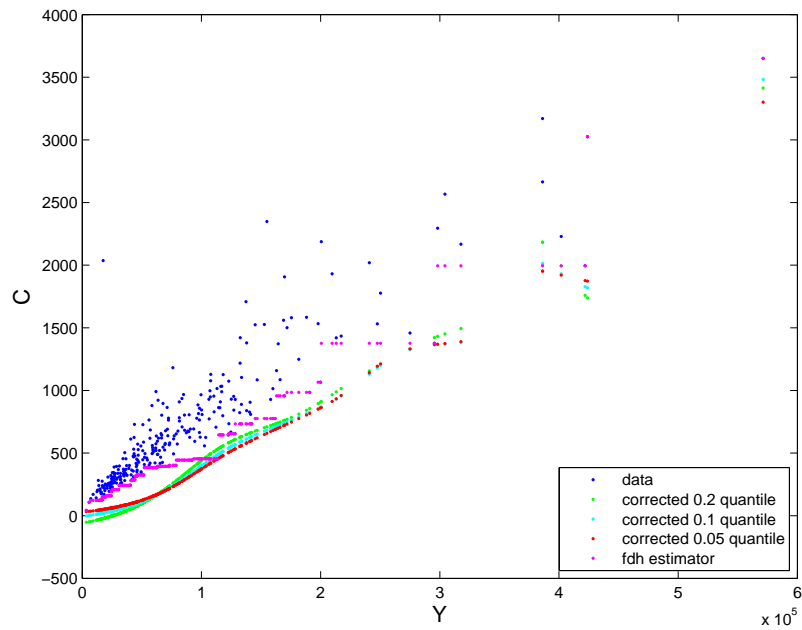


Figure 8: *Data on post offices. Final results for the separable model and the FDH estimate. The bias-corrected quantiles are for $\alpha = 0.05, 0.1$ and 0.2 .*