I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/54

Unobserved heterogeneity and endogeneity in nonparametric frontier estimation

SIMAR, L., VANHEMS, A. and I. VAN KEILEGOM

Unobserved heterogeneity and endogeneity in nonparametric frontier estimation

Léopold Simar[*]
leopold.simar@uclouvain.be

Anne Vanhems[§]
a.vanhems@tbs-education.fr

Ingrid Van Keilegom [*,**]
ingrid.vankeilegom@uclouvain.be.

December 18, 2013

Abstract

In production theory and efficiency analysis, firm efficiencies are measured by their distances to a production frontier, which is the geometrical locus of optimal combinations of inputs and outputs. It is today recognized that in the presence of heterogenous conditions (like environmental factors) that influence the shape and the position of the frontier, traditional measures of efficiency obtained in the space of inputs/outputs have no sensible economic meaning. This is because the benchmark frontier may not be attainable for a firm facing different heterogenous conditions. Using a nonparametric approach, this can be corrected by using conditional frontiers and conditional efficiency scores developed in the literature. In this paper we extend these concepts in the case where the heterogeneity is not observed. We propose and analyze a model where the heterogeneity variable is linked to a particular input (or output). It is defined as the part of the input (or the output), independent from some instrumental variable through a nonseparable nonparametric model. We discuss endogeneity issues involved in this model. Under certain regularity assumptions, we show that the model is identified, we propose nonparametric estimators of the conditional frontier and the conditional efficiency score, and analyze their asymptotic properties. When using FDH estimators we prove the asymptotic convergence to a Weibull distribution, whereas when using the robust order-m estimators we obtain the asymptotic normality of the estimators. The method is illustrated with some simulated and real data examples. A Monte-Carlo experiment shows how the procedure works for finite samples.

Key Words: Unobserved Heterogeneity, Endogeneity, Nonparametric Frontiers, Robust Estimation of Frontiers, Conditional Efficiency.

JEL Classification: Primary C13; secondary C14; C49

[*]Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium. Research supported by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy).

[§]University of Toulouse, Toulouse Business School and Toulouse School of Economics, France.

[**]Research also supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain").

1 Introduction

In production theory and efficiency analysis, the technical efficiency of a production unit (a firm) is measured by an appropriate distance of this unit to the production frontier, which is the geometrical locus of optimal combinations of the inputs and outputs in the set of attainable production plans (the production set). The economic theory underlying this analysis dates back to the works of Koopmans (1951), Debreu (1951) and Farrell (1957). The first empirical analysis where the production set, its boundary and the resulting efficiency are estimated from a sample of observed units is due to Farrell (1957).

Parametric models have been used in the econometric literature starting from the works of Aigner and Chu (1968) or Greene (1980) for parametric deterministic frontier models and of Aigner et al. (1977), Meeusen and van den Broeck (1977) using stochastic frontier models. Nonparametric approaches have been developed after the pioneering work of Farrell (1957) with the DEA estimator, popularized by Charnes et al. (1978) and the FDH estimator of Deprins et al. (1984). These nonparametric approaches are based on envelopment techniques in the space of inputs and outputs and so are sensitive to outliers and extreme points in the cloud of observed points. Robust alternatives have been proposed in Cazals et al. (2002), Aragon et al. (2005) and Daouia and Simar (2007) using partial frontiers (order-m and order-α quantile frontiers). Today, these nonparametric estimators have been analyzed from a statistical point of view and inference on efficiency estimates is available, mainly by using bootstrap techniques (see the recent survey Simar and Wilson, 2013 for details, and the references therein).

It is now recognized that in the presence of heterogeneous conditions (environmental factors, ...) that may affect the frontier level, the measures based only on inputs and outputs may have no economic meaning, because the units are benchmarked against a frontier level that may not be attainable under their environmental conditions (see Simar and Wilson, 2007, 2011 for a detailed discussion). A solution to this problem is to use attainable sets and frontier levels that may depend on these heterogeneous conditions: this is the idea of defining conditional efficiency scores, initiated by Cazals et al. (2002) and extended in Daraio and Simar (2005). Here too the statistical properties of these estimators and their robust versions, have been established (in Cazals et al., 2002, Daouia and Simar, 2007 and Jeong et al., 2010).

So far, these conditional approaches are based on the assumption that these heterogeneous conditions are known and observed. However, this may not be the case. We may infer that some latent factors influence the production process and in particular the set of attainable

combinations of inputs and outputs. The objective of our paper is to propose a model where we may have unobserved heterogeneity. In this paper we propose one approach that allows to identify and estimate this latent variable. This will be achieved through a model where the heterogeneity variable is linked to a particular input (or an output). It is defined as the part of the input (or the output), independent from some instrumental variable through a non separable nonparametric model. This model involves endogeneity issues that will be discussed. Under usual regularity assumptions, we show that the model is identified, we propose a nonparametric estimator and analyze its asymptotic properties. To the best of our knowledge we are not aware of existing results for handling latent or unobserved heterogeneity in nonparametric frontier models.

The paper is organized as follows. Section 2 summarizes the basic notations and concepts of conditional efficiency measures, and gives the basic model for introducing unobserved heterogeneity in the production process. It also gives the natural nonparametric estimators of the elements of the model. Then, Section 3 establishes the asymptotic properties of our estimators. Section 4 gives a few simple illustrative examples with simulated and real data, and Section 5 indicates how the procedure works for finite samples through a limited Monte-Carlo experiment. Finally, Section 6 contains some conclusions and ideas for future research, Appendix A gives an original way to derive optimal bandwidths for estimating conditional efficiencies sharing monotonicity properties, and Appendix B contains the proofs of the asymptotic results.

2 The Model

2.1 Conditional efficiency scores

We first summarize the existing tools for handling the presence of observable heterogeneous (or environmental) factors $Z \in \mathcal{Z} \subseteq \mathbb{R}^d$ in a production process where inputs $X \in \mathbb{R}_+^p$ are used to produce the output $Y \in \mathbb{R}_+$.[1] The production set is the set of technically possible combinations of inputs and output. In the presence of environmental factors, when $Z = z$ this is defined as

$$\Psi(z) = \{(x, y) \in \mathbb{R}_+^p \times \mathbb{R}_+ \mid x \text{ can produce } y, \text{ when } Z = z\}. \tag{2.1}$$

[1]We will do the presentation in an output orientation where firms try to reach the maximal possible output for a given level of inputs. The same could be done in an input orientation where the firms try to reduce their input (cost) $X \in \mathbb{R}_+$ for a given level of their outputs $Y \in \mathbb{R}_+^q$.

This set is the support of the joint random variable (X, Y), conditionally on $Z = z$. The marginal support of the variables (X, Y) is given by

$$\Psi = \{(x, y) \in \mathbb{R}^p_+ \times \mathbb{R}_+ \mid x \text{ can produce } y\} = \bigcup_{z \in \mathcal{Z}} \Psi(z). \tag{2.2}$$

It is the support of the joint marginal distribution of (X, Y). The "separability condition" described in Simar and Wilson (2007, 2011) is the assumption that $\Psi(z) = \Psi$ for all $z \in \mathcal{Z}$. The traditional Farrell-Debreu efficiency score for a firm operating at level (x, y) is defined by the distance in the output direction to the upper boundary of Ψ, and it is given by

$$\lambda(x, y) = \sup\{\lambda \mid (x, \lambda y) \in \Psi\}. \tag{2.3}$$

The separability assumption is a strong assumption, and if it is not fulfilled, the efficiency score $\lambda(x, y)$ is meaningless since the frontier may not be reachable for the firm facing the environmental conditions z. The same is obviously true for their nonparametric DEA or FDH estimators.

The separability condition may be interpreted as an exogeneity condition on Z when defining the support of (X, Y). On the other hand, when the separability condition is not fulfilled, then the support of (X, Y) is linked to Z and we can say that Z is an endogenous variable with respect to the support of (X, Y). Endogeneity when estimating the support of a random variable by usual nonparametric estimators is not of the same nature as the usual endogeneity in regression models. Here, the distribution of the distance of a point to the frontier can depend on Z (and even on X) without affecting the estimation of the support of Y given $X \leq x$.[2] By contrast, when Z is "endogenous" in the sense that it affects this support, it matters a lot. We will come back to this later.

To overcome these difficulties, Cazals et al. (2002) and Daraio and Simar (2005) introduce the concept of conditional frontier and of conditional efficiency score for a unit operating at the level (x, y) and facing conditions z. It is defined as

$$\lambda(x, y|z) = \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi(z)\}, \tag{2.4}$$

where the production process is described by the conditional distribution of (X, Y) given $Z = z$. It is convenient to characterize this distribution by the probability of being dominated[3]:

$$H_{X,Y|Z}(x, y|Z = z) = P(X \leq x, Y \geq y|Z = z) = S_{Y|X,Z}(y|X \leq x, Z = z)F_{X|Z}(x|Z = z), \tag{2.5}$$

[2]This fact was already acknowledged in Florens and Simar (2005).

[3]Here and in the sequel, inequalities on vectors should be understood component by component.

where $S_{Y|X,Z}$ denotes a survival function and $F_{X|Z}$ a cumulative distribution function; we note also the nonstandard conditioning on X ($X \leq x$) and the usual conditioning $Z = z$ for the environmental variables. Under the free disposability assumption,[4] the output conditional score (see e.g. Daraio and Simar, 2005), can also be defined for all x such that $F_{X|Z}(x|Z = z) > 0$ as

$$\begin{aligned} \lambda(x, y|z) &= \sup\{\lambda > 0 \mid H_{XY|Z}(x, \lambda y|Z = z) > 0\} \quad (2.6) \\ &= \sup\{\lambda > 0 \mid S_{Y|X,Z}(\lambda y|X \leq x, Z = z) > 0\}. \quad (2.7) \end{aligned}$$

Nonparametric estimators are obtained by plugging-in empirical versions of the probabilities appearing on the right hand side of these equations. The conditioning on $Z = z$ requires the use of smoothing techniques and the derivation of the optimal bandwidth for Z.

In our particular setup where Y is univariate, the frontier can be described by a conditional production function:

$$\phi(x, z) = \sup\{y \mid F_{Y|X,Z}(y|X \leq x, Z = z) < 1\}, \quad (2.8)$$

since for univariate y, $F_{Y|X,Z}(y|X \leq x, Z = z) = 1 - S_{Y|X,Z}(y|X \leq x, Z = z)$. Similarly, the output conditional efficiency score can be defined as

$$\lambda(x, y|z) = \sup\{\lambda > 0 \mid F_{Y|X,Z}(\lambda y|X \leq x, Z = z) < 1\}, \quad (2.9)$$

where for univariate Y, it is obvious that $\phi(x, z) = \lambda(x, y|z)\, y$.

Given a sample of observations $\mathcal{X} = \{(X_i, Y_i, Z_i)|i = 1, ..., n\}$, we can compute a nonparametric kernel estimator of $F_{Y|X,Z}$ by using

$$\widehat{F}_{Y|X,Z}(y|x, z) = \frac{\sum_{i=1}^{n} \mathbb{1}(X_i \leq x, Y_i \leq y) K_{h_z}(Z_i - z)}{\sum_{i=1}^{n} \mathbb{1}(X_i \leq x) K_{h_z}(Z_i - z)}, \quad (2.10)$$

where $\mathbb{1}(\cdot)$ is the indicator function, h_z is a (univariate) bandwidth converging to zero when n tends to infinity, $K(u_1, ..., u_s) = \prod_{j=1}^{s} k(u_j)$, with k a univariate kernel function and s an arbitrary dimension, and where for any bandwidth h we let $k_h(\cdot) = k(\cdot/h)/h$ and $K_h(u_1, ...u_s) = \prod_{j=1}^{s} k_h(u_j)$.

Nonparametric estimators of ϕ and λ are obtained by replacing $F_{Y|X,Z}$ by $\widehat{F}_{Y|X,Z}$ in equations (2.8) and (2.9). The practical computation of $\widehat{\lambda}$ can be easily derived from equation (2.10):

$$\widehat{\lambda}(x, y|z) = \max_{\{i \ | X_i \leq x, ||Z_i - z|| \leq h_z\}} \left(\frac{Y_i}{y} \right), \quad (2.11)$$

[4]The free disposability assumption means that if (x, y) is achievable, then $(\tilde{x}, \tilde{y})$ is also achievable once $(\tilde{x} - x, y - \tilde{y}) \geq 0$. It is technically possible to waste resources.

where $\|Z-z\| = \max_{j=1,...,d} |Z_j - z_j|$ and provided the support of k is $[-1,1]$.[5] Note that this is a localized version of the traditional FDH estimator. The statistical properties of nonparametric FDH estimators of these conditional measures are now established (Jeong et al. 2010), and tests of the "separability" or of the exogeneity condition of Z on the support of (X,Y) have been suggested in Daraio et al. (2010).[6]

As any envelopment estimators, the FDH is very sensitive to outliers and extreme data points. Therefore Cazals et al. (2002) introduced the order-m frontier as a less extreme frontier more robust to outliers. This is done for both marginal and conditional frontiers. We limit the presentation to the conditional case, which is mostly used in what follows. Consider an integer $m \geq 1$ and let $(Y_1, ..., Y_m)$ be m independent and identically distributed random variables generated by the distribution of Y given $X \leq x$ and $Z = z$. The expected maximum conditional production function of order m, denoted by $\phi_m(x,z)$, is the real function defined for any (x,z) by

$$\begin{aligned} \phi_m(x,z) &= E(\max(Y_1,...,Y_m)|X \leq x, Z=z) \\ &= \int_0^\infty \left[1 - F_{Y|X,Z}(y|X \leq x, Z=z)^m\right] dy. \end{aligned} \tag{2.12}$$

The associated order-m conditional efficiency score is defined by

$$\lambda_m(x,y|z) = \int_0^\infty \left[1 - F_{Y|X,Z}(uy|X \leq x, Z=z)^m\right] du, \tag{2.13}$$

where again, for univariate Y, we have $\phi_m(x,z) = \lambda_m(x,y|z)\, y$. A nonparametric estimator of $\lambda_m(x,y|z)$ is obtained by replacing the conditional distribution $F_{Y|X,Z}$ by its nonparametric estimator

$$\widehat{\lambda}_m(x,y|z) = \int_0^\infty \left[1 - \widehat{F}_{Y|X,Z}(uy|X \leq x, Z=z)^m\right] du. \tag{2.14}$$

The statistical properties of nonparametric estimators of conditional and unconditional order-m efficiency scores are provided in Cazals et al. (2002). For the marginal (unconditional) case, we only mention that they reach the parametric $\sqrt{n}$ rate of convergence with a normal asymptotic distribution. We have similar results for the conditional case, where

[5] Bădin et al. (2010) indicate how to adapt the procedure of Hall et al. (2004) to the particular setup here. We describe in Appendix A an original way to derive an optimal bandwidth giving estimates of $\lambda(x,y|z)$ sharing desirable monotonicity properties.

[6] To simplify the notation we use in our presentation, and in all the theoretical developments throughout the paper, a univariate bandwidth h_z. This is without any loss of generality and in practice we can use a vector of bandwidths h_z each having the same optimal size. Then $||Z-z|| \leq h_z$ has to be understood componentwise.

the rate of convergence is deteriorated by the smoothing in (2.10), i.e. we have the rate $\sqrt{nh_z^d}$ when using the same bandwidth for all elements of Z. With the optimal size for the bandwidths this leads to the rate $\sqrt{n^{4/(4+d)}}$, indicating the curse of dimensionality when using multivariate Z.

2.2 Unobserved heterogeneity

Consider now the case where the environmental variable characterizing the heterogeneity in the production process is not observed and we denote it by $V \in \mathbb{R}$. At this stage we limit our presentation to the case of a univariate unobserved factor. We will see below how to add other multivariate observable factors Z. In order to identify the unobserved variable V, we assume that it is linked to one input, say X^1, where we decompose X as $X = (X^1, X^{(-1)}) \in \mathbb{R} \times \mathbb{R}^{p-1}$. We model this link with the help of an instrumental variable W as follows:

$$X^1 = \psi(W, V), \tag{2.15}$$

where W is an observed variable correlated with X^1 and independent from V. The model defined in (2.15) is nonseparable in V and has been studied in several papers in econometrics (see for example, Matzkin, 2003). We also impose a classical assumption of monotonicity of ψ with respect to the second argument V, and assume, without loss of generality, that V is uniformly distributed on $[0, 1]$. It is known that under these assumptions, V can be identified by the conditional distribution of X^1 given W:

$$V = F_{X^1|W}. \tag{2.16}$$

Looking to the link equation (2.15) we can view this as defining the latent variable V as the part of the input X^1 that is independent of the instrument W. For instance the variable W could be a measure of the size of the firms, which certainly affects the level of the input (say labor), but another latent factor may influence the input variable X^1. So, in a sense, X^1 is endogenous in the production process, because it is determined by the variable W and this latent factor V (that might be a factor linked to environmental conditions, socio-economic conditions, ...).

Another illustration is the measure of the return to schooling by ranking the earnings of individuals with respect to education. Here the variable Y measures the earnings of an individual and X measures the level of education (by the number of years spent in school). Of course, there may also exist unobserved characteristics of the individuals that may impact both the level of education and the earnings, and to take into account this

unobserved heterogeneity, Card (1995) suggests to work with the proximity to college as an instrumental variable W. In the same vein, Angrist and Krueger (2001) emphasize that people make schooling choices by comparing the costs and benefits of alternatives. Then, one possible source of instrument W could be the differences in costs due to for example loan policies or institutional constraints (see also Angrist and Krueger 1991 for more details).

We have seen above that if we neglect in the production process an heterogeneous factor when estimating the production frontier and the efficiency scores, we end up with a meaningless measure, unless the heterogeneous factor is "separable" or exogenous when defining the support of the production variables (X, Y). So it is safer to estimate the performances of the firms by conditioning on this latent factor and estimating the frontier function $\phi(x, v)$ or equivalently the conditional efficiency $\lambda(x, y|v) = \phi(x, v)/y$. This may be achieved by "estimating" or predicting the value of the heterogeneity term for each observation in the following way. Due to (2.16), a natural estimator of V_i is given by

$$\begin{aligned} \widehat{V}_i &= \widehat{F}_{X^1|W}(X_i^1|W_i) \\ &= \frac{\sum_{k=1}^n \mathbb{1}(X_k^1 \leq X_i^1) K_{h_w}(W_i - W_k)}{\sum_{k=1}^n K_{h_w}(W_i - W_k)}. \end{aligned} \tag{2.17}$$

The resulting nonparametric estimators of the efficiency measure λ, that take into account this unobserved heterogeneity as well as other observed heterogeneity factors $Z \in \mathbb{R}^d$, are then defined for any point (x, y) facing the heterogeneity conditions v and z as

$$\begin{aligned} \widehat{\lambda}(x, y|v, z) &= \sup\{\lambda \,|\widehat{F}_{Y|X,\widehat{V},Z}(\lambda y|X \leq x, (\widehat{V}, Z) = (v, z)) < 1\} \\ \widehat{\lambda}_m(x, y|v, z) &= \int_0^\infty \left[1 - \widehat{F}_{Y|X,\widehat{V},Z}(uy|X \leq x, (\widehat{V}, Z) = (v, z))^m\right] du, \end{aligned}$$

where by analogy with the case of only observed heterogeneity Z, $\widehat{F}_{Y|X,\widehat{V},Z}(\cdot)$ will be computed from the sample of observed and estimated values $(X_i, Y_i, \widehat{V}_i, Z_i)$, $i = 1, \ldots, n$ as follows:

$$\widehat{F}_{Y|X,\widehat{V},Z}(y|x, v, z) = \frac{\sum_{i=1}^n \mathbb{1}(X_i \leq x, Y_i \leq y) K_{h_v}(\widehat{V}_i - v) K_{h_z}(Z_i - z)}{\sum_{i=1}^n \mathbb{1}(X_i \leq x) K_{h_v}(\widehat{V}_i - v) K_{h_z}(Z_i - z)}. \tag{2.18}$$

This allows us to estimate the conditional frontier $\phi(x, v, z)$ for any value of (x, v, z). Of course, in practice, the efficiency scores will be evaluated at the observed data points, leading to $\widehat{\lambda}(X_i, Y_i, |\widehat{V}_i, Z_i)$ and $\widehat{\lambda}_m(X_i, Y_i, |\widehat{V}_i, Z_i)$ respectively.

To investigate the potential effect of V and Z on the frontier, we can adapt the procedure suggested in Bădin et al. (2012) and compare the conditional measures to their unconditional counterpart. In practice we compute the ratios $\widehat{\lambda}(X_i, Y_i, |\widehat{V}_i, Z_i)/\widehat{\lambda}(X_i, Y_i)$, where $\widehat{\lambda}(X_i, Y_i)$ is

the FDH estimator computed by neglecting the potential effect of V and Z. The same could be done by using the ratios of the robust order-m version of the estimators. As explained in detail in Daraio and Simar (2005, 2007), the nonparametric regression of these ratios on the values $\widehat{V}_i$ and Z_i is helpful to see if this latent variable has no effect or a favorable/unfavorable effect on the frontier of possibilities. The analysis of the order-m ratios for small values of m (e.g. $m = 1$) gives more information on the effect of V and Z on the middle of the distribution of the inefficiencies (see Bădin et al., 2012 for more details). In Section 3 we analyze the asymptotic properties of our estimators, whereas in Section 4 we illustrate the method with some simple examples.

2.3 Endogeneity issues

We have already seen that in a sense, in our setup, X^1 can be viewed as endogenous since it is determined by the two independent components W and V. There is another way to view the endogeneity issues that our model can handle. Again, to simplify the presentation and the notation, we avoid to explicitly introduce other observable environmental variables Z, but we know that this can be done.

With our nonseparable link model (2.15) we can always write

$$Y = \phi(X, V) - U, \tag{2.19}$$

where $U \geq 0$ is a measure of the inefficiency. This way of writing is familiar to parametric approaches. With our assumptions above (W and V are independent and ψ is strictly monotone) and with the minimal assumption that U is exogenous in defining the support of (X, Y), we can appreciate the endogeneity issue raised if we neglect our latent variable V.

If we define the unconditional (to V) frontier as usual by $\varphi(x) = \sup\{y \mid F_{Y|X}(y|X \leq x) < 1\}$ and write the analog of (2.19) as

$$Y = \varphi(X) - \widetilde{U}, \tag{2.20}$$

we see, by comparing with (2.19) that the support of $\widetilde{U}$ depends on V and so depends on X^1, which introduces the endogeneity issue already discussed above. The endogeneity is an issue not because the distribution of $\widetilde{U}$ depends on X^1 (as in usual regression models) but because its support depends on X^1.

2.4 Related literature

In this subsection, we would like to highlight the links between our model and the existing literature. As far as we know, there are no existing results on handling unobserved heterogeneity in nonparametric frontier models. However, in econometrics, several nonparametric models have already been studied to take into account unobserved heterogeneity among individuals. In Matzkin (2003) the generic model considered is the nonseparable model $X^1 = \psi(W, V)$, where V is the error term of the nonparametric regression that now cannot be neglected anymore. In Imbens and Newey (2009), the model studied takes into account unobserved heterogeneity and also endogeneity of some explanatory variables since they consider two equations: $Y = m(X, \epsilon)$ with X^1 linked to ϵ (endogeneity issue), and $X^1 = \psi(W, V)$. In this case, the error term V acts as a control function to identify the model.

Another aspect of the econometrics literature linked to our problem is nonparametric estimation with generated covariates. Indeed, the production frontier we estimate is now defined through a generated covariate $V = F_{X^1|W}$ that is unknown and needs to be estimated. Although this framework has already been studied in econometrics (Imbens and Newey 2009, Mammen, Rothe and Schienle 2012, Vanhems and Van Keilegom 2013, among others), it has never been analyzed in a production frontier setting. In particular, one important issue is to study the impact of the generated covariate on the asymptotic properties of the estimator. One very nice feature of our model is that the effect of estimating our generated covariate disappears asymptotically and the estimated frontier behaves asymptotically as if the heterogeneity V would be observed. This property is quite surprising since the estimation of V is nonparametric and the associated estimator $\widehat{V}$ may converge slowly. Note that the impact of the estimation of a generated covariate on the asymptotic variance has been investigated in Mammen, Rothe and Schienle (2012) for example, and they prove that in some cases the effect of estimating the generated covariate may disappear in the asymptotic variance.

3 Asymptotic Properties

We start by showing that the conditional full production frontier $\widehat{\phi}(x, v, z)$ and the conditional efficiency score $\widehat{\lambda}(x, y|v, z)$ follow asymptotically a Weibull distribution. Recall that for any $a, b > 0$ and any random variable T, we have that $T \sim \text{Weib}(a, b)$ if $P(T \leq t) = 1 - \exp(-at^b)$ for $t > 0$. The regularity conditions under which the results below are valid,

as well as the proofs of these results, can be found in Appendix B.

Theorem 3.1. *Assume (A1)(a),(b), (A2)–(A6) and (C1)–(C4). Then, for any x, y, z and v, there exist constants $\mu_{xvz} > 0$ and $\mu_{xy|vz} > 0$, such that*

$$(nh_v h_z^d)^{1/(p+1)}\big(\phi(x,v,z) - \widehat{\phi}(x,v,z)\big) \xrightarrow{d} Weib\big(\mu_{xvz}, p+1\big)$$

and

$$(nh_v h_z^d)^{1/(p+1)}\big(\lambda(x,y|v,z) - \widehat{\lambda}(x,y|v,z)\big) \xrightarrow{d} Weib\big(\mu_{xy|vz}, p+1\big).$$

The formulas of μ_{xvz} and $\mu_{xy|vz}$ have a complicated structure. In the unconditional case they can be found in e.g. Park, Simar and Weiner (2000) (Definition A.2, page 873). The conditional case was considered in Jeong, Park and Simar (2010) (Theorem 1, page 114).

It is interesting to note that the asymptotic distribution is the same as in the case where V would be observed. This is remarkable, since the estimator of V has a nonparametric (and so a slow) rate of convergence.

We now turn to the asymptotic properties of the estimator $\widehat{\lambda}_m(x,y|v,z)$ of the order m-efficiency score. As in Cazals, Florens, Simar (2002), we will show that the estimator is asymptotically normally distributed and converges to the true order m-efficiency score $\lambda_m(x,y|v,z)$. Again, we note that the asymptotic variance is the same as in the case where V would be observed.

Theorem 3.2. *Assume (A1)(a),(c), (A2)–(A6).*

(i) Then, for any x, y, z and v,

$$\widehat{\lambda}_m(x,y|v,z) - \lambda_m(x,y|v,z) = \widetilde{\lambda}_m(x,y|v,z) - \lambda_m(x,y|v,z) + o_P((nh_v h_z^d)^{-1/2}),$$

where

$$\widetilde{\lambda}_m(x,y|v,z) = \int_0^\infty \Big[1 - \widehat{F}_{Y|X,V,Z}\left(uy|X \le x, (V,Z) = (v,z)\right)^m\Big]\, du$$

is the estimator of order m based on the true (unobservable) variable V.

(ii) Moreover, for any x, y, z and v,

$$(nh_v h_z^d)^{1/2}(\widehat{\lambda}_m(x,y|v,z) - \lambda_m(x,y|v,z)) \to N\big(0, \sigma^2(x,y|v,z)\big),$$

with

$$\sigma^2(x,y|v,z) = \mathrm{Var}\Big(A(X,Y;x,y,v,z)|(V,Z) = (v,z)\Big) f_{V,Z}(v,z) \int K^2(u)\, du,$$

where

$$
\begin{aligned}
&A(X,Y;x,y,v,z)\\
&= \frac{m}{F_{X|V,Z}(x|(V,Z)=(v,z))\ f_{V,Z}(v,z)}\,\mathbb{1}(X\le x)\\
&\times\left\{\int_0^\infty \Big(1-F_{Y|X,V,Z}(uy|X\le x,(V,Z)=(v,z))^{m-1}\,\mathbb{1}(Y\le uy)\Big)du-\lambda_m(x,y|v,z)\right\}.
\end{aligned}
$$

Remark 3.1. If we define $\widehat{\phi}_m(x,v,z)=\widehat{\lambda}_m(x,y|v,z)\,y$, then we immediately recover the asymptotic properties of $\widehat{\phi}_m(x,v,z)$ from the theorem above.

4 Illustrative Examples

4.1 Simulated data

We will first illustrate how the estimation procedure works in practice in two simulated examples. In the first example, we suppose that the effect of the unobserved variable V on the frontier is monotone and in the second one we will simulate data with an inverted U-shaped effect. In Example 1 the frontier function is defined by

$$\phi(x,v)=\sqrt{x}\exp(-v). \tag{4.1}$$

Then the inefficiency U generates observations Y_i below the frontier ϕ. We have chosen U to be independent of (X,V) with $U\sim|N(0,\sigma_u^2)|$ and $\sigma_u^2=0.05$. Therefore, for a given (x,v) we have

$$Y=\phi(x,v)\exp(-U). \tag{4.2}$$

The link between X and V is modeled through an instrument W according to the following model:

$$X=3W\exp(V), \tag{4.3}$$

where $V\sim\text{Unif}(0,1)$ and $W\sim\text{Unif}(0,1)$ independently of each other.

We generate n i.i.d. observations (V_i,W_i,U_i) and we derive the resulting observations (X_i,Y_i), $i=1,\dots,n$ according to model (4.3) for X_i and then according to (4.2) for Y_i. Here the effect of V on the true frontier is monotone decreasing. In this example, ignoring the unobserved V_i would lead to estimating the wrong marginal frontier $\varphi(x)=\sqrt{x}$ using only the sample (X_i,Y_i), $i=1,\dots,n$. In the Monte-Carlo study given in Section 5, we will investigate the statistical behavior of our estimator based on 500 simulated samples of

size $n = 100, 200, 400$ and 1000. In this illustrative section we only consider one particular sample of size $n = 100$.

In the second example, we follow the same scheme as above, but the true frontier is now given by

$$\phi(x, v) = \sqrt{x} \exp(-(2(v - 0.5))^2), \tag{4.4}$$

and the link between X, V and the instrument W is given by

$$X = 3W + V. \tag{4.5}$$

Except for the two latter equations, all elements of the simulation are the same as in the first example. So here the effect of V on the frontier of possibilities has an inverted U-shaped form.

Figure 1 displays for the two examples the shape of the true frontier and of the wrong marginal frontier in the full space (X, V). Figures 2 and 3 indicate how the procedure is able to detect the effect of the unobserved heterogeneity on the frontier, and this even with a sample of size as small as $n = 100$. As explained above and suggested in the literature (see e.g., Daraio and Simar, 2005, 2007 or Bădin et al., 2012) we provide plots of the ratios of the conditional efficiency scores over the marginal ones, as a function of the heterogeneous variable. So we compute for each data point the ratio $\widehat{\lambda}(X_i, Y_i|V = \widehat{V}_i)/\widehat{\lambda}(X_i, Y_i)$ for the full frontier and for its robust alternative $\widehat{\lambda}_m(X_i, Y_i|V = \widehat{V}_i)/\widehat{\lambda}_m(X_i, Y_i)$ and plot them against $\widehat{V}_i$. To help the interpretation of the plot we also display the nonparametric (local-linear) fit of the cloud of points. We clearly see in the two examples that the regression curve is able to recover the expected shape in the two scenarios. We also note the similar shape of the nonparametric regressions, although we use a rather small value of m, namely $m = 20$. We know that if m increases, the order-m efficiencies converge to the full efficiency measures. But for smaller values of m we get an idea of the effect of V on the distribution of the inefficiencies.

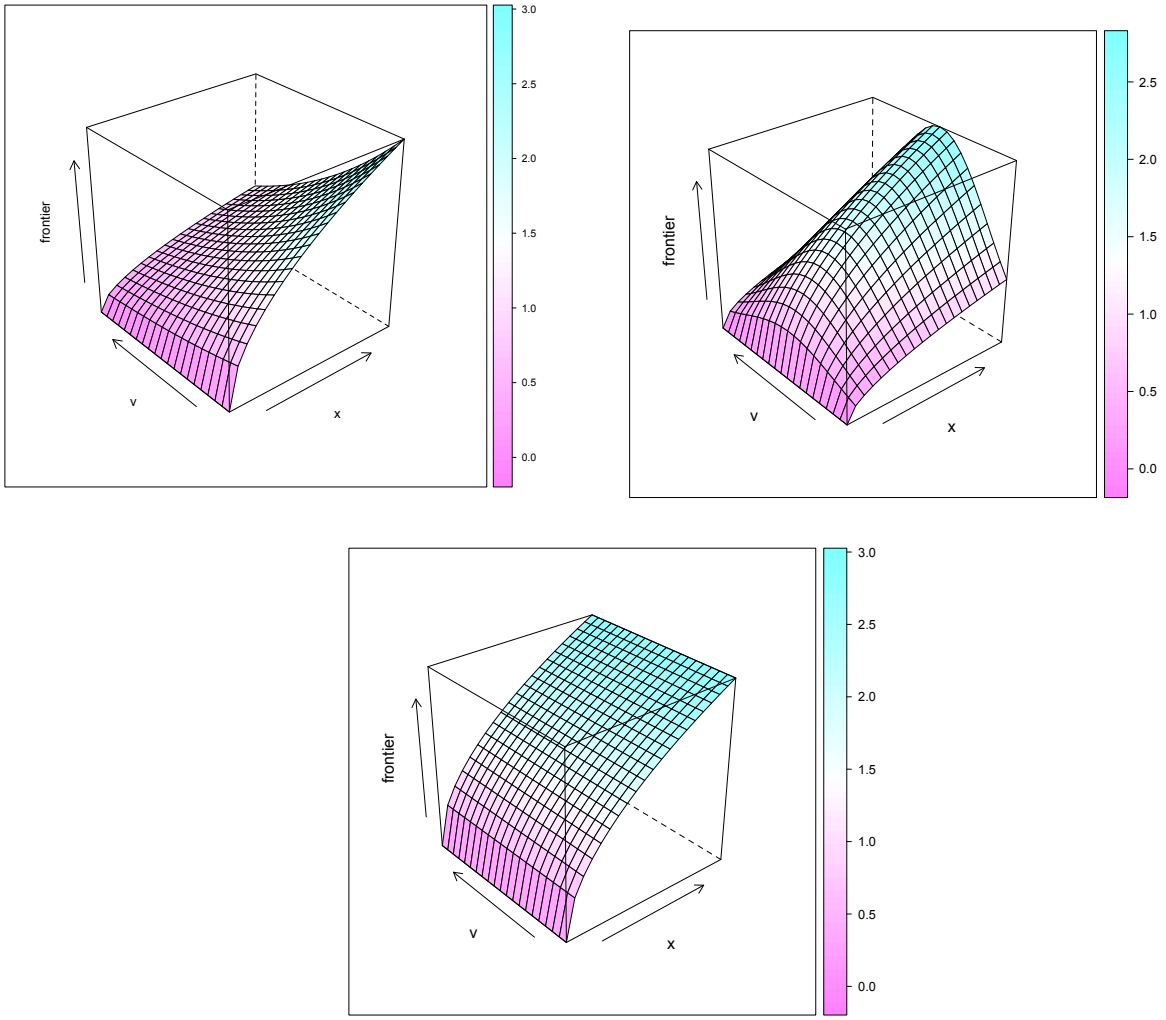

Figure 1: *Top panels: true frontier levels for Example 1 (left panel) and Example 2 (right panel), as a function of X and V. The bottom panel is the marginal (wrong) frontier, ignoring the variable V, for both examples.*

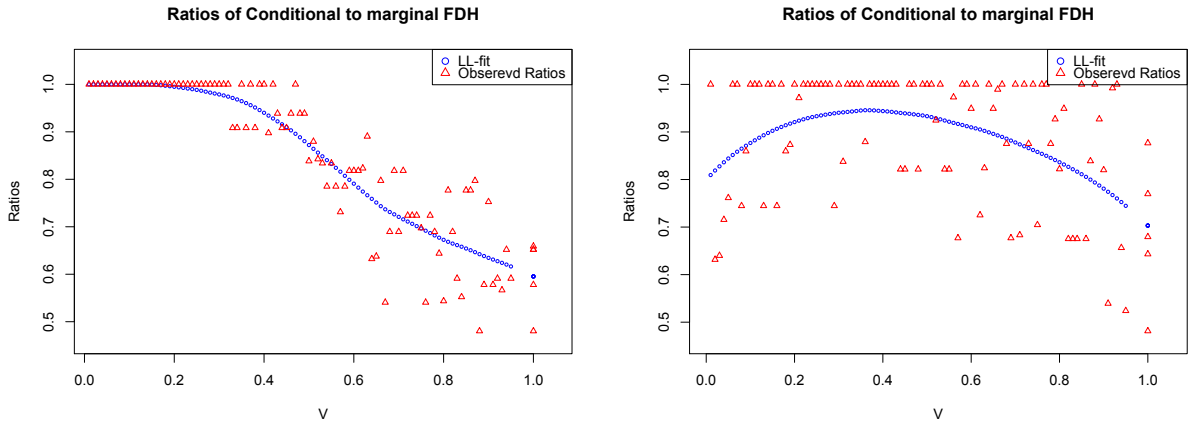

Figure 2: *Estimated ratios* $\widehat{\lambda}(X_i, Y_i|V = \widehat{V}_i)/\widehat{\lambda}(X_i, Y_i)$ *of conditional full efficiency scores over marginal ones, as a function of* $\widehat{V}_i$*, for Example 1 (left panel) and Example 2 (right panel). The sample size is* $n = 100$*. The local linear fit of the cloud of ratios is also displayed.*

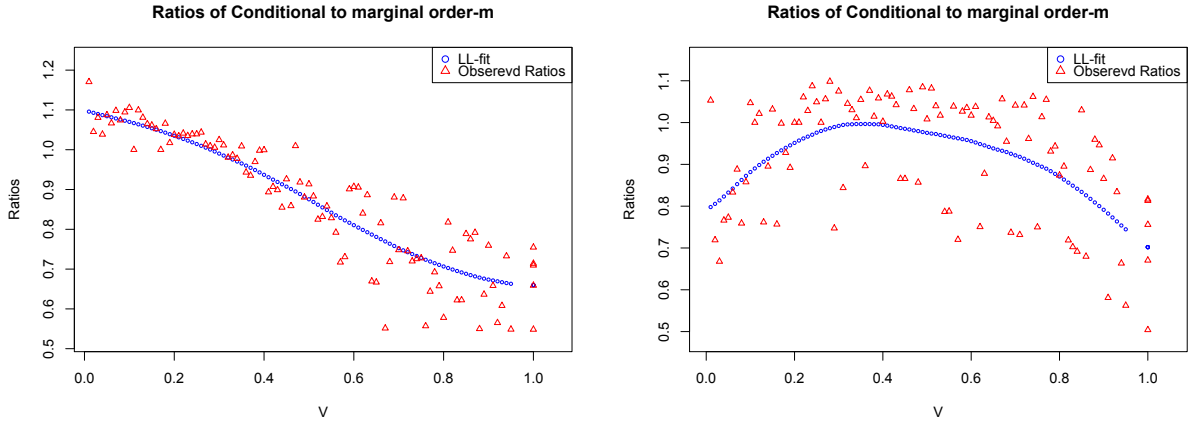

Figure 3: *Estimated ratios* $\widehat{\lambda}_m(X_i, Y_i|V = \widehat{V}_i)/\widehat{\lambda}_m(X_i, Y_i)$ *of conditional order-m efficiency scores over marginal ones, as a function of* $\widehat{V}_i$*, for Example 1 (left panel) and Example 2 (right panel). Here,* $m = 20$ *and the sample size is* $n = 100$*. The local linear fit of the cloud of ratios is also displayed.*

Finally, we would like to know whether the procedure does not introduce spurious effects of V on the frontier. Therefore, we simulate a sample of size $n = 100$ under the same scenario as in Example 1, except that now, the support of the frontier is not depending on V. The picture of the true frontier is the one in the bottom panel of Figure 1. To make the story

clear, in this example, we replace equation (4.1) by the following one:

$$\phi(x, v) = \sqrt{x}, \tag{4.6}$$

and we keep all other elements of the scenario identical as in Example 1, with X determined by V as in (4.3). The ratios of the conditional over the marginal efficiency scores are depicted in Figure 4: we clearly see that the ratios of the full frontiers indicate that there is no dependence on V, and for the ratio of the order-m frontiers there is no significant difference for $m = 20$.[7] As we know, the two panels would be more similar by choosing larger values of m.

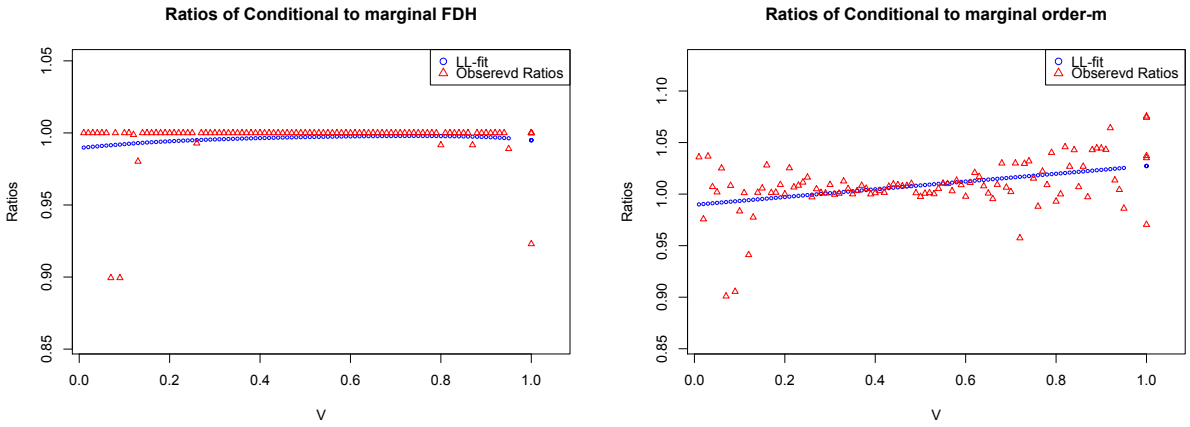

Figure 4: *Estimated ratios of conditional efficiency scores over marginal ones for the full frontier (left panel) and the order-m frontier (right panel), with no effect of V on the support of the frontier. Here, $m = 20$ and the sample size is $n = 100$. The local linear fit of the cloud of ratios is also displayed.*

So we can conclude from these three simple simulated data sets, that the procedure is able to recover the shape of the effect of the latent factor V on the production process, including the case of no effect.

[7]Other simulated trials gave very similar results. As expected, many of them, corresponding to large values of h_v for determining $\widehat{F}_{Y|X,V}(y|x, v)$ (say values of h_v much larger than the range of V), gave ratios equal to 1 for all data points.

4.2 Real data illustrations

Banks example

We first illustrate our procedure with a real dataset coming from the banking sector.[8] Simar and Wilson (2007) analyzed these data based on Aly et al. (1990) using data on 6.955 US commercial banks observed at the end of the fourth quarter of 2002.

The original dataset contains 3 inputs (purchased funds, core deposits and labor) and 4 outputs (consumer loans, business loans, real estate loans, and securities held) for banks. Bădin et al. (2012) explained that the inputs can be aggregated in a one dimensional input measure, without loosing much information and the same is true for the outputs. The final output Y is highly correlated (more than 0.93) with all the original outputs (a measure of the level of the loans) and the same is true for the input X (correlation with the original inputs of more than 0.97). This facilitates the presentation of our empirical illustration.

We consider as observed environmental condition the variable Z defined by a measure of the diversity of the services proposed by the banks (see Aly. et al., 1990, for details). Previous studies used the same data ($n = 303$ units) and the same variables, and all were using an input oriented framework (Simar and Wilson, 2007, Bădin et al., 2012 and Florens et al., 2014). They also considered the size of the banks (measured by the log of the total assets) as another observed environmental factor that might influence the production process. The analysis was able to detect a slight negative effect of the diversity and almost no effect of the size variable on the shift of the frontier (see e.g. Figure 14 in Florens at al., 2014).

In our approach we rather use a model where latent heterogeneity can play a role in fixing the level of the output (loans). This can happen e.g. when the level of the output is linked to the size of the banks, but that there exists a non-observable factor that may also influence the level of the loans. This could be related to some socio-economic or profile of the customers of each particular bank. So, here we could use the size as an instrument W to identify this latent factor.

So, here we estimate the conditional input-oriented efficiency, conditional on (V, Z) where Z is the diversity index and is observed. To save space we do not reproduce all the formulae for the conditional input measures that can be found in the references above. To investigate the effect of (V, Z) on the production process we produce, as for the univariate case, a picture of the ratios between the conditional and the unconditional (marginal) efficiency scores. In this bivariate example, we only provide the nonparametric regression surface of these ratios

[8] We would like to thank Paul W. Wilson who provided us this dataset.

as a function of (V, Z). This is displayed in Figure 5 and 6. We clearly see a positive monotone effect of the latent factor (here we are in input orientation, for larger values of V, the two frontiers, marginal and conditional are more similar, the shift is more important for smaller values of V). The diversity does not play an important role when V is large. However, we see a more important inverted U-shaped effect for small values of V. Again, a more deep analysis, using information of the environment of each bank would help to give a sensible interpretation of this effective unobserved heterogeneity factor, which in our model is defined as the part of the output (loans) independent from the instrument W (the size of the banks). Table 1 gives the results for 10 randomly chosen banks in our sample. We note how the conditional measures may differ for some points from the unconditional ones, for example, the rankings of the units would change.

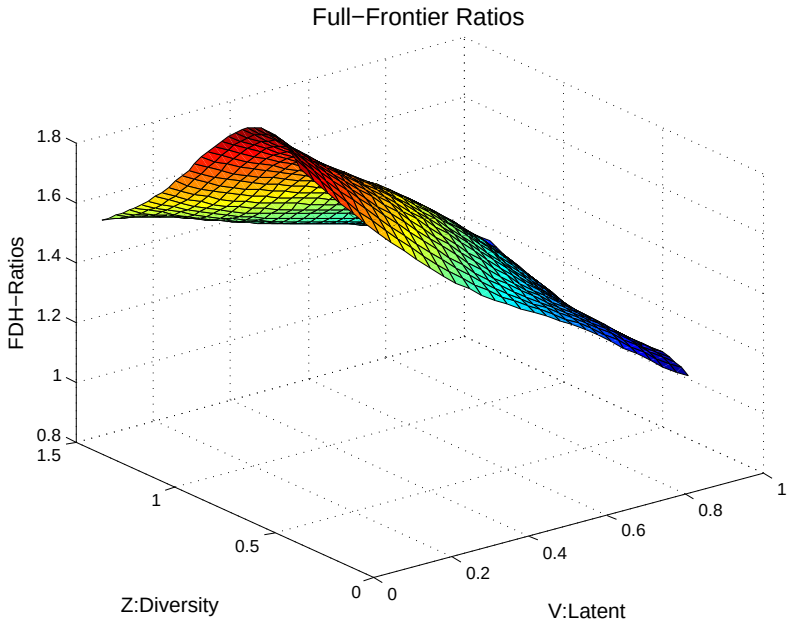

Figure 5: *Estimated ratios of conditional over marginal efficiencies for the full frontier, for the BANK data. Here, the sample size is* $n = 303$.

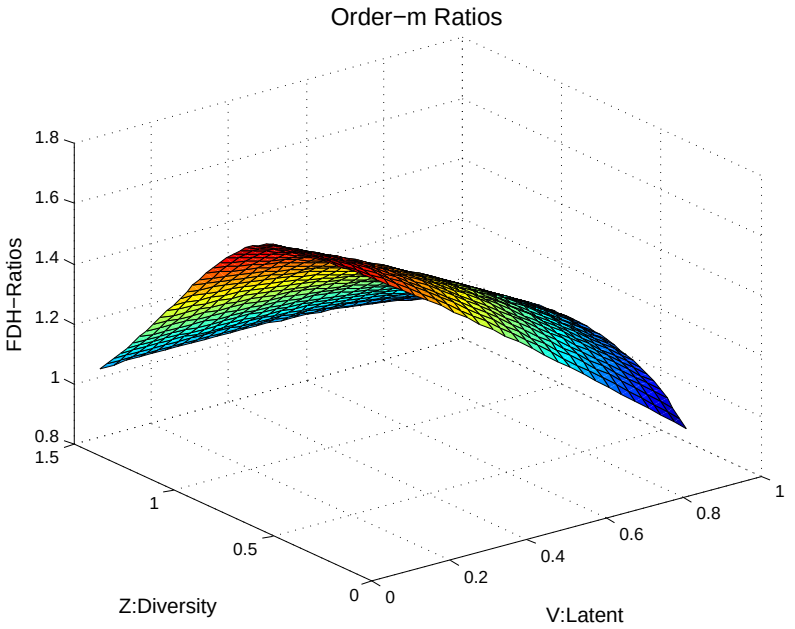

Figure 6: *Estimated ratios of conditional over marginal efficiencies for the order-m frontier, for the BANK data. Here,* $m = 20$ *and the sample size is* $n = 303$.

Units	$\widehat{\theta}^{in}(x,y)$	$\widehat{\theta}^{in}(x,y\|v,z)$	$\widehat{\theta}^{in}_m(x,y)$	$\widehat{\theta}^{in}_m(x,y\|v,z)$
3	0.7228	1.2765	0.9899	1.2779
251	0.6258	1.4112	0.6853	1.4112
99	0.8921	1.2295	1.0749	1.2369
150	1.0000	1.0000	1.1118	1.0662
197	0.6862	1.0647	0.7547	1.0648
220	1.0000	1.0000	1.2862	1.2139
237	1.0000	1.0000	1.2720	1.3496
107	0.6577	1.4212	0.7208	1.4212
39	0.7324	1.2499	0.9940	1.2666
166	0.8192	1.4248	0.9496	1.4250
mean	0.8004	1.1785	0.9889	1.2343

Table 1: *Input efficiency scores for 10 randomly selected banks. The mean is taken over the* $n = 303$ *observations.*

Schools example

For this example we chose the popular data set from Charnes et al. (1981) (referred to as the CCR data), where the performance of 70 schools is analyzed, 49 of them having benefited

from Program Follow Through (PFT) and 21 called Non-Follow Through (NFT). The paper gives the data for 3 outputs (achievements on students on some tests) and 5 inputs, describing 4 characteristics of the family and the number of teachers. In our illustration and due to the limited number of data points, we will limit our model to the simplest model of production with one input (the number of teachers) and one output (an average of the 3 achievement tests which are in fact highly correlated). The latter paper also gives in the appendix the number of students.

We argue that if the number of teachers is certainly linked to the size of the school (measured as proxy by the number of students), the choice of the number of teachers is not completely exogenous but is also determined by some latent factor, measuring facts like the socio-economical environment of the school or other similar characteristics. We will use as an instrument W the number of students. We delete observation 59 from our sample because the precise information on W was not available. So we end up with 69 schools.

The results on the potential effect of the latent variable V defined as "the part of the chosen number of teachers which is not related to the number of pupils" are shown in Figure 7.

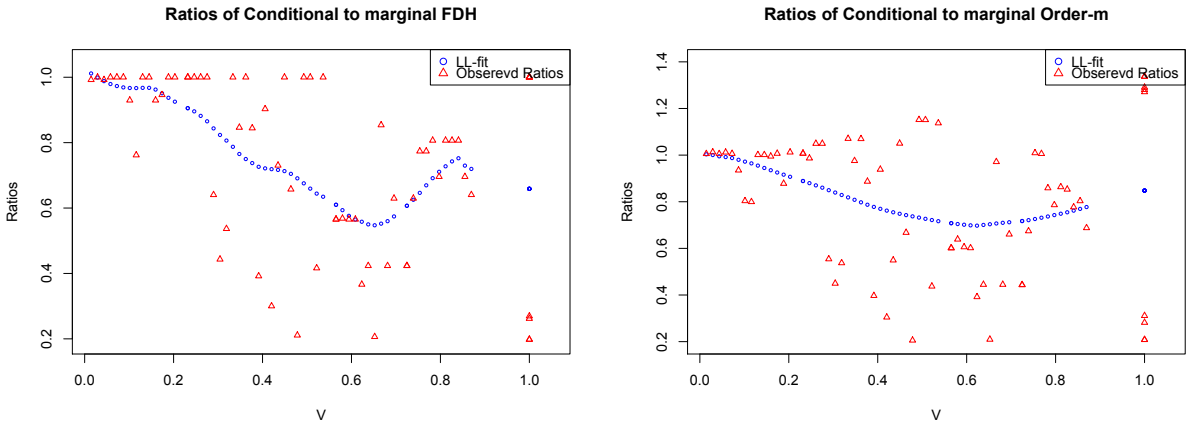

Figure 7: Estimated ratios of conditional over marginal efficiencies for the full frontier (left panel) and the order-m frontier (right panel), for the CCR data. Here, $m = 20$ and the sample size is $n = 69$.

We observe a clear U-shaped effect and it would be interesting to be able to link the values of $\widehat{V}_i$ to pertinent socio-economic variables describing the environment of the school to get insights in the interpretation of V and its consequences on the frontier levels. In this exercise, we could for instance argue that V is linked to the inverse of some socio-economical level of the schools. For small values of V, when V increases the ratio of teachers over

pupils may increase, but with no effect for correcting the increasingly "bad" environment V. Therefore, the frontier is declining. However at a certain level, when V gets larger this ratio of teacher per pupils is large enough to start to have a positive effect and so to correct the "bad" environment and improve the level of score that is reachable. Of course this explanation is just an exercise, and more data would be necessary to obtain a more clear picture for the interpretation of this latent factor.

Table 2 gives for each school participating to the PFT program the resulting efficiency measures and Table 3 does the same for the NFT schools not participating to the program.

5 Monte-Carlo Simulations

For analyzing the statistical behavior of our estimators for finite samples, we simulated 500 Monte-Carlo (MC) samples according to the scenario of Example 1 described above by equations (4.3) and (4.2). We did the experiment for $n = 100, 200, 400$ and 1000. The numerical burden for computing the optimal bandwidth by cross validation becomes enormous when n increases. Therefore, for the case $n = 1000$ we only realized 200 MC replications.

We focus the analysis here on one particular fixed point in the middle of the picture (top left panel in Figure 1) defined as follows: $v_0 = w_0 = 0.5$ and $u_0 = E(U) = \sqrt{2\pi}\sigma_u = 0.1784$. Then, by the model we have $x_0 = 3w_0 \exp(v_0)$ and $y_0 = \sqrt{x_0}\exp(-v_0)\exp(-u_0)$. So the true (Farrell) efficiency score we are estimating is $\lambda(x_0, y_0|v_0) = \exp(-u_0) = 1.1953$. We fix $m = 20$ as above, and compute (by an independent Monte-Carlo simulation) the true value of the order-m efficiency score and obtain $\lambda_m(x_0, y_0|v_0) = 1.0752$. Table 5 gives for each case the MC estimates of the bias and of the MSE (computed relative to the true value of the full and order-m efficiencies). The numbers between parentheses are the MC standard deviations of the MC estimates: it allows to appreciate the significance of the bias and MSE estimates.

The results are exactly as we expected from the asymptotic theory: our estimators behave better as the sample size increases and we see that, as it should be, the "naive" marginal estimators (FDH and order-m, computed relative to the wrong marginal frontier) have a bad behavior. We see that the improvement when going from $n = 100$ to $n = 200$ is small but significant, whereas when going to larger sample sizes the difference becomes much clearer. Also, for large n we observe, as expected by the theory, a better behavior of the estimator of the conditional order-m efficiency relative to the conditional FDH estimator.

	$\widehat{\lambda}_{m,i}$	$\widehat{\lambda}_i$	$\widehat{\lambda}_{m,i\|\widehat{V}_i}$	$\widehat{\lambda}_{i\|\widehat{V}_i}$
1	0.865404	1	0.996333	1
2	1.2494	1.31253	0.999052	1
3	1.24351	1.43709	0.999072	1
4	2.47564	2.63519	1.92354	2.12605
5	4.67432	4.74155	1.42345	1.42346
6	2.22439	2.25639	1.23293	1.44513
7	2.32482	2.47465	1.39819	1.39927
8	1.42408	1.52646	1.14444	1.41896
9	1.45647	1.56118	1.24238	1.25955
10	0.985212	1.13858	1.03461	1.13858
11	0.980471	1.02986	1.02899	1.02986
12	0.952039	1	0.999252	1
13	1.42911	1.52121	1.34107	1.37379
14	2.51915	2.55133	1	1
15	2.24833	2.36155	0.999997	1
16	1.22471	1.41536	1.19502	1.19784
17	0.998713	1	0.999999	1
18	1.52307	1.75992	1.73227	1.75992
19	0.992931	1.05681	0.999641	1
20	0.986545	1	0.998676	1
21	1.03345	1.10732	1.10561	1.10732
22	1.53924	1.56137	1.55029	1.56137
23	1.20928	1.63309	1.6169	1.63309
24	0.987388	1	0.999998	1
25	1.66146	1.76853	0.999859	1
26	1.1644	1.23947	1	1
27	0.994257	1.00782	0.99971	1
28	1.19057	1.20578	1.11321	1.20578
29	4.7764	4.8451	1	1
30	2.8131	3.01456	1.54515	2.20221
31	2.79212	2.99207	1.09435	1.09604
32	5.97505	6.06099	1.22853	1.27827
33	1.02282	1.1712	0.993431	1
34	1.28703	1.48725	1.48217	1.48725
35	1.004	1.07576	0.998747	1
36	2.31815	2.35149	2.28718	2.35149
37	1.86343	1.89023	1.87681	1.89023
38	1.70083	1.70302	1.49338	1.70302
39	1.12738	1.18417	0.999972	1
40	1.84214	1.97416	1.21661	1.24285
41	1.55784	1.66983	1.04004	1.0977
42	1.13907	1.15362	1.15265	1.15362
43	1.48239	1.58842	0.999783	1
44	0.740396	1	0.990089	1
45	1.65527	1.65741	1.65737	1.65741
46	1.46069	1.91869	1.88454	1.91869
47	1.01266	1.02647	1.01839	1.01851
48	3.81032	4.00288	1.69004	1.69503
49	2.71042	2.91595	1.86435	1.86779

Table 2: *Marginal and conditional measures for PFT schools (CCR data).*

	$\widehat{\lambda}_{m,i}$	$\widehat{\lambda}_i$	$\widehat{\lambda}_{m,i\|\widehat{V}_i}$	$\widehat{\lambda}_{i\|\widehat{V}_i}$
50	1.43774	1.88853	1.44641	1.46122
51	3.5477	3.81683	0.999897	1
52	0.933291	1	0.999257	1
53	2.26933	2.62223	1.44959	1.48997
54	0.987192	1.29244	0.996324	1
55	1.69485	1.80408	1.01931	1.0201
56	2.28675	2.4019	1	1
57	1.7241	1.9741	1.35572	1.37368
58	1.17219	1.18716	1.17951	1.18716
59	2.26034	2.37417	1.00531	1.00535
60	5.1889	5.45115	1.07953	1.07953
61	1.8608	1.8632	1	1
62	2.87898	3.02397	1.27674	1.2805
63	1.80784	1.9378	1.09535	1.09571
64	2.67332	2.71177	1.20177	1.20182
65	2.83509	3.74253	3.60318	3.74253
66	1.98713	2.61513	2.54451	2.61513
67	1.55118	1.66269	1.33944	1.34145
68	4.80663	5.04957	1	1
69	3.2116	3.71103	0.999383	1

Table 3: *Marginal and conditional measures for NFT schools (CCR data).*

	$n = 100$		$n = 200$		$n = 400$		$n = 1000$	
col#	1	2	3	4	5	6	7	8
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
FDH	0.4134	0.1881	0.4845	0.2447	0.5350	0.2930	0.5947	0.3579
MCstd	(0.0059)	(0.0047)	(0.0045)	(0.0043)	(0.0037)	(0.0039)	(0.0046)	(0.0055)
FDH $\|\, V$	0.1316	0.0614	0.1260	0.0520	0.1001	0.0344	0.0836	0.0205
MCstd	(0.0094)	(0.0037)	(0.0085)	(0.0035)	(0.0070)	(0.0024)	(0.0082)	(0.0022)
Order-m	0.3731	0.1485	0.3823	0.1501	0.3840	0.1497	0.3906	0.1535
MCstd	(0.0043)	(0.0031)	(0.0028)	(0.0021)	(0.0021)	(0.0016)	(0.0021)	(0.0017)
Order-$m\|V$	0.1482	0.0540	0.1138	0.0345	0.0838	0.0211	0.0587	0.0108
MCstd	(0.0080)	(0.0035)	(0.0066)	(0.0024)	(0.0053)	(0.0016)	(0.0061)	(0.0011)

Table 4: *Monte-Carlo results for Example 1 based on 500 MC replications (for* $n = 1000$*, only 200 MC).*

6 Conclusions

In this paper we proposed a model that allows to handle unobserved heterogeneity, when this heterogeneity is linked to some particular input (or output). We can identify this latent

variable through a nonseparable nonparametric model. Then by using the technique of conditional efficiency measures and their robust order-m versions, we are able to measure the technical efficiency of firms that take this latent factor into account.

The asymptotic properties of our estimators are derived and we illustrate through some simulated data how the procedure works. In particular, we see that we are able to recover by appropriate plots the effect of the unobserved factor on the production process, including the case where it has no effect.

We first illustrate our method by means of a real data set on the activity of Banks already used in the literature (Simar and Wilson, 2007) then we also analyze the popular data set on schools from Charnes et al. (1981). Finally, a Monte-Carlo experiment illustrates how the estimator behaves for finite samples.

Open issues on our research agenda include the adaptation to our setup of the separability test proposed by Daraio et al. (2010), and extension to multivariate output Y and multivariate latent factor V.

A Appendix A: Bandwidth Selection for Conditional Efficiency Scores: A New Approach

In this appendix, we extend previous results from Bădin et al. (2010) in order to achieve estimators of the conditional efficiency scores that share some desirable monotonicity properties. It also turns out that this procedure is less numerically demanding than the original method suggested in Bădin et al. (2010). The statistical theory of bandwidth selection for conditional densities and distributions can be found in Hall et al. (2004) and Li and Racine (2008).

For simplicity of notation, we will omit the unobserved variable V in what follows, since it is not essential for explaining the method. Using the notations introduced in Section 2, we have to select a bandwidth for computing an estimate of $F_{Y|X,Z}(y|X \leq x, Z = z)$. The kernel estimator is given in (2.10). Bădin et al. (2010) suggested to estimate $F_{Y|X,Z}(y|X \leq x, Z = z)$ directly by kernel methods and so, they derive an optimal bandwidth $h_z(x)$ which depends on the current value of x. They adapt for this the techniques proposed by Hall et al. (2004) and Li and Racine (2008).

We notice that this has to be computed at many points of interest (often the n data points observed in the sample) and this can be numerically very demanding if n is large. The resulting bandwidth has the appropriate size $h_z(x) = O(n^{-1/(d+4)})$, but we are not sure

that the resulting estimator $\widehat{\lambda}(x, y|z)$ has all the desirable monotonicity properties. Indeed, it is easy to show that $\widehat{\lambda}(x, y|z)$ is monotone decreasing with y (for fixed x and z), but it is not true that $\widehat{\lambda}(x, y|z)$ is monotone decreasing with x (when y and z are being held fixed) due to the fact that the bandwidth may change with the value of x. However this is a desirable property, because reaching the same y with conditions z with less inputs x should produce more efficient firms.

We improve the method suggested by Bădin et al. (2010) to obtain an estimator of the conditional efficiency score having all the desired monotonicity properties. It turns out that the resulting bandwidth estimator has the optimal size, is less numerically demanding and does not depend on the chosen orientation. The idea is not to estimate the conditional distribution $F_{Y|X,Z}(y|X \leq x, Z = z)$ directly, but rather the conditional probabilities $H_{XY|Z}(x, y|Z = z)$, requiring to determine the optimal bandwidth independently of x (and of y). The nonparametric estimator is defined by

$$\widehat{H}_{XY|Z}(x, y|Z = z) = \frac{\sum_{i=1}^{n} \mathbb{1}(X_i \leq x, Y_i \geq y) K_{h_z}(Z_i - z)}{\sum_{i=1}^{n} K_{h_z}(Z_i - z)}, \tag{A.1}$$

where the notations were introduced in Section 2. Now we define

$$\widehat{\lambda}(x, y|z) = \sup\{\lambda > 0 \mid \widehat{H}_{XY|Z}(x, \lambda y|Z = z) > 0\}. \tag{A.2}$$

Since for fixed (y, z), $\widehat{H}_{XY|Z}(x, y|Z = z)$ cannot decrease with x, so will $\widehat{\lambda}(x, y|z)$. We also remark that the optimal bandwidth obtained when estimating $H_{XY|Z}(x, y|Z = z)$ is not depending on the chosen orientation.

The cross-validation procedure is explained in Li and Racine (2008). First, cross-validation is applied to an estimator of the conditional density function $f(x, y|z)$, which leads to a bandwidth $\tilde{h}_z$ say.[9] Then, $\tilde{h}_z$ has to be rescaled to reach the appropriate optimal size, namely we define $h_z = \tilde{h}_z \, n^{-(p+1)/((d+4)(d+p+5))}$, which can be used to define the estimators of the conditional efficiency scores. In our Monte-Carlo experiment we used both approaches to determine the bandwidth and the one presented in this appendix produced better results. The results in Table 5 are therefore based on the latter bandwidth selector.

[9]This can be obtained by the package 'np' in R, see Hayfield and Racine (2008). It involves, as proposed by Hall et al. (2004), the estimation of bandwidths for (x, y, z), even if at the end we only are interested in a bandwidth for z, say $\tilde{h}_z$. This bandwidth has to be rescaled since we do not estimate a density but a distribution function. Bădin et al. (2010) pointed out an unfortunate typo in Li and Racine (2007), page 183, when defining the rescaling factor.

B Appendix B: Proofs of the Asymptotic Results

Assumptions

(A1) (a) $nh_v h_z^d \to \infty$, $nh_w^q h_v^2 (\log n)^{-1} \to \infty$, and $h_w^2 = o(h_v)$.
(b) $nh_v h_z^d \max(h_z, h_v, h_w)^{2(p+1)} \to 0$.
(c) $nh_v h_z^d \max(h_z, h_v, h_w)^4 \to 0$.

(A2) k is a symmetric kernel with support $[-1, 1]$, and k is twice continuously differentiable.

(A3) The support of W, respectively Z, is a compact subset of $\mathbb{R}^q$, respectively $\mathbb{R}^d$.

(A4) $F_{X^1|W}(x^1|w)$ is twice continuously differentiable with respect to the components of w, and the partial derivatives of order 1 and 2 are all bounded. Moreover, the density f_W exists and $\inf_w f_W(w) > 0$.

(A5) $F_{Y|X,V,Z}(y|X \leq x, v, z)$ is twice continuously differentiable with respect to the components of v and z, and the partial derivatives of order 1 and 2 are all bounded.

(A6) The densities $f_{V,Z}(v, z)$ and $f_{X,Y,V,Z}(x, y, v, z)$ exist and are continuous on their support, and $f_{V,Z}(v, z)$ is bounded away from zero.

We also need to introduce a set of notations, which are coming from Jeong, Park and Simar (2010). Let $f^h_{X,Y|V,Z}(\cdot, \cdot|v, z)$ be the conditional density of (X, Y) given that $\|Z - z\| \leq h_z$ and $|V - v| \leq h_v$, and let $\Psi^h(v, z) = \big\{(x, y) \in \mathbb{R}^{p+1}_+ | f^h_{X,Y|V,Z}(x, y|v, z) > 0\big\}$. Moreover, let $\lambda^h(x, y|v, z) = \sup\big\{\lambda | (x, \lambda y) \in \Psi^h(v, z)\big\}$.

For a fixed x, y, v and z, we need to introduce the following regularity conditions. They are required to apply Theorem 1 in Jeong, Park and Simar (2010).

(C1) $\Psi(v, z)$ and $\Psi^h(v, z)$ are free disposable.

(C2) $\lambda^h(x, y|v, z) - \lambda(x, y|v, z) = o\big((nh_v h_z^d)^{-1/(p+1)}\big)$.

(C3) For $(\bar{v}, \bar{z})$ in a neighborhood of (v, z), the conditional density $f_{X,Y|V,Z}(\cdot, \cdot|\bar{v}, \bar{z})$ of (X, Y) given $(V, Z) = (\bar{v}, \bar{z})$ exists and it satisfies $f_{X,Y|V,Z}\big(\bar{x}, \lambda(\bar{x}, \bar{y}|\bar{v}, \bar{z})\bar{y}|\bar{v}, \bar{z}\big) > 0$ for all $(\bar{x}, \bar{y}, \bar{v}, \bar{z})$ in a neighborhood of (x, y, v, z). Moreover, $f^h_{X,Y|V,Z}(\cdot, \cdot|\bar{v}, \bar{z})$ converges to $f_{X,Y|V,Z}(\cdot, \cdot|\bar{v}, \bar{z})$ as $h \to 0$.

(C4) $\lambda(\cdot, \cdot|v, z)$ and $\lambda^h(\cdot, \cdot|v, z)$ are continuously differentiable in a neighborhood of (x, y), and their first partial derivatives at (x, y) are all nonzero.

Proof of Theorem 3.1. We restrict attention to the estimator $\widehat{\phi}(x,v,z)$, since both estimators only differ by a scalar constant from each other. Moreover, in order to make the notations less heavy we omit the vector Z in what follows (it can be treated in the same way as V). Let $a_n = C_1(nh_w^q)^{-1/2}(\log n)^{1/2} + C_2 h_w^2$ for some $C_1, C_2 > 0$. Then, by assumption (A1), $a_n = o(h_v)$ and hence there exists a sequence $\delta_n \to 0$ such that $a_n = h_v\delta_n$. Note that the estimator $\widehat{V}_i$ satisfies $\max_{i=1,\dots,n}|\widehat{V}_i - V_i| = O_P(a_n) = o_P(a_n\delta_n^{-1/2}) = o_P(h_v\delta_n^{1/2})$.

For fixed $t > 0$ and for $b_n = (nh_v)^{-1/(p+1)}$, consider

$$
\begin{aligned}
&P\big(b_n^{-1}(\phi(x,v) - \widehat{\phi}(x,v)) \le t\big)\\
&= 1 - P\big(b_n^{-1}(\phi(x,v) - \widehat{\phi}(x,v)) > t, \max_{i=1,\dots,n}|\widehat{V}_i - V_i| \le h_v\delta_n^{1/2}\big)\\
&\quad - P\big(b_n^{-1}(\phi(x,v) - \widehat{\phi}(x,v)) > t, \max_{i=1,\dots,n}|\widehat{V}_i - V_i| > h_v\delta_n^{1/2}\big)\\
&= 1 - P\big(\max\{Y_i : X_i \le x, |\widehat{V}_i - v| \le h_v\} < \phi(x,v) - b_n t, \max_{i=1,\dots,n}|\widehat{V}_i - V_i| \le h_v\delta_n^{1/2}\big)\\
&\quad + o(1)\\
&\ge 1 - P\big(\max\{Y_i : X_i \le x, |V_i - v| \le h_v(1-\delta_n^{1/2})\} < \phi(x,v) - b_n t,\\
&\qquad \max_{i=1,\dots,n}|\widehat{V}_i - V_i| \le h_v\delta_n^{1/2}\big) + o(1)\\
&\ge 1 - P\big(\max\{Y_i : X_i \le x, |V_i - v| \le h_v(1-\delta_n^{1/2})\} < \phi(x,v) - b_n t\big) + o(1). \qquad \text{(B.1)}
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
&P\big(b_n^{-1}(\phi(x,v) - \widehat{\phi}(x,v)) \le t\big)\\
&\le 1 - P\big(\max\{Y_i : X_i \le x, |V_i - v| \le h_v(1+\delta_n^{1/2})\} < \phi(x,v) - b_n t,\\
&\qquad \max_{i=1,\dots,n}|\widehat{V}_i - V_i| \le h_v\delta_n^{1/2}\big) + o(1)\\
&\le P\big(\max\{Y_i : X_i \le x, |V_i - v| \le h_v(1+\delta_n^{1/2})\} \ge \phi(x,v) - b_n t\big)\\
&\quad + P\big(\max_{i=1,\dots,n}|\widehat{V}_i - V_i| > h_v\delta_n^{1/2}\big) + o(1)\\
&= 1 - P\big(\max\{Y_i : X_i \le x, |V_i - v| \le h_v(1+\delta_n^{1/2})\} < \phi(x,v) - b_n t\big) + o(1). \quad \text{(B.2)}
\end{aligned}
$$

In what follows, we will work out the upper bound (B.2). The lower bound (B.1) can be developed in exactly the same way, and these two bounds together will give at the end the limiting distribution. Write

$$
\begin{aligned}
&P\big(b_n^{-1}(\phi(x,v) - \widehat{\phi}(x,v)) \le t\big)\\
&\le 1 - \prod_{i=1}^{n} P\big(X_i > x \text{ or } |V_i - v| > h_v(1+\delta_n^{1/2}) \text{ or } Y_i < \phi(x,v) - b_n t\big) + o(1)\\
&= 1 - (\pi_n)^n + o(1) \quad \text{(say)}.
\end{aligned}
$$

Note that

$$\begin{aligned}\pi_n &= 1 - P\big(X \le x, Y \ge \phi(x,v) - b_n t \,\big|\, |V-v| \le h_v(1+\delta_n^{1/2})\big)\, P\big(|V-v| \le h_v(1+\delta_n^{1/2})\big) \\ &= 1 - 2p_n h_v(1+\delta_n^{1/2}) \quad \text{(say)},\end{aligned}$$

since $P\big(|V-v| \le h_v(1+\delta_n^{1/2})\big) = 2h_v(1+\delta_n^{1/2})$. It can be seen that $p_n = \mu_{xv}(b_n t)^{p+1}/2$ for some $\mu_{xv} > 0$ using condition (A6) and Theorem 1 in Jeong, Park and Simar (2010), and hence

$$\begin{aligned}&P\big(b_n^{-1}(\phi(x,v) - \widehat{\phi}(x,v)) \le t\big) \\ &\le 1 - (\pi_n)^n + o(1) = 1 - \big(1 - \mu_{xv} b_n^{p+1} t^{p+1} h_v\big)^n + o(1) \\ &= 1 - \Big(1 - \frac{\mu_{xv} t^{p+1}}{n}\Big)^n + o(1) \to 1 - \exp(-\mu_{xv} t^{p+1}).\end{aligned}$$

In a similar way we can show that

$$P\big(b_n^{-1}(\phi(x,v) - \widehat{\phi}(x,v)) \le t\big) \ge 1 - \exp(-\mu_{xv} t^{p+1}) + o(1),$$

which finishes the proof. □

Proof of Theorem 3.2. Let us start with the proof of (i). We will show that $\widehat{\lambda}_m(x,y|v,z) - \widetilde{\lambda}_m(x,y|v,z) = o_P\big((nh_v h_z^d)^{-1/2}\big)$. Consider the following expansion:

$$\begin{aligned}&\widehat{\lambda}_m(x,y|v,z) - \widetilde{\lambda}_m(x,y|v,z) \\ &= -\int_0^\infty \Big[\widehat{F}_{Y|X,\widehat{V},Z}\left(uy|X \le x,v,z\right)^m - \widehat{F}_{Y|X,V,Z}\left(uy|X \le x,v,z\right)^m\Big] du \\ &= -m\int_0^\infty \Big[\widehat{F}_{Y|X,\widehat{V},Z}\left(uy|X \le x,v,z\right) - \widehat{F}_{Y|X,V,Z}\left(uy|X \le x,v,z\right)\Big] \\ &\qquad \times \widehat{F}_{Y|X,V,Z}\left(uy|X \le x,v,z\right)^{m-1} du\,(1+o_P(1)) \\ &= -\int_0^\infty \Big[\widehat{F}_{Y|X,\widehat{V},Z}\left(uy|X \le x,v,z\right) - \widehat{F}_{Y|X,V,Z}\left(uy|X \le x,v,z\right)\Big]\widehat{g}(u)du\,(1+o_P(1)) \\ &= -\int_0^\infty \Big[\widehat{F}_{Y|X,\widehat{V},Z}\left(uy|X \le x,v,z\right) - \widehat{F}_{Y|X,V,Z}\left(uy|X \le x,v,z\right)\Big]g(u)du\,(1+o_P(1)), \quad \text{(B.3)}\end{aligned}$$

where $\widehat{g}(u) = m\widehat{F}_{Y|X,V,Z}\left(uy|X \le x,v,z\right)^{m-1}$ and $g(u) = mF_{Y|X,V,Z}\left(uy|X \le x,v,z\right)^{m-1}$. Here, the second equality above holds true since straightforward calculations show that $\sup_u |\widehat{F}_{Y|X,\widehat{V},Z}\left(uy|X \le x,v,z\right) - \widehat{F}_{Y|X,V,Z}\left(uy|X \le x,v,z\right)| = o_P(1)$, and the third equality follows from the fact that it can be easily seen that $\sup_u |\widehat{g}(u) - g(u)| = o_P(1)$.

We can write (B.3) as $\big(\widehat{N}/\widehat{D} - \widetilde{N}/\widetilde{D}\big)(1+o_P(1))$, where

$$\widehat{N} = \widehat{f}_{V,Z}^{-1}(v,z)\frac{1}{n}\sum_{j=1}^{n}\left[\int_0^\infty \mathbb{1}(Y_j \geq uy)g(u)du\right]\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)K_{h_v}(v-\widehat{V}_j)$$

$$\widehat{D} = \widehat{f}_{V,Z}^{-1}(v,z)\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)K_{h_v}(v-\widehat{V}_j)$$

$$\widetilde{N} = \widehat{f}_{V,Z}^{-1}(v,z)\frac{1}{n}\sum_{j=1}^{n}\left[\int_0^\infty \mathbb{1}(Y_j \geq uy)g(u)du\right]\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)K_{h_v}(v-V_j)$$

$$\widetilde{D} = \widehat{f}_{V,Z}^{-1}(v,z)\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)K_{h_v}(v-V_j),$$

and where $\widehat{f}_{V,Z}(v,z) = n^{-1}\sum_{j=1}^n K_{h_z}(z-Z_j)K_{h_v}(v-V_j)$. We also let $D = F_{X|V,Z}(X \leq x|v,z)$ and $N = \int_0^\infty H_{X,Y|V,Z}(x,uy|v,z)g(u)du$, where $H_{X,Y|V,Z}(x,y|v,z) = P(X \leq x, Y \geq y|v,z)$. For sake of simplicity, we omit the dependence on (y,x,v,z) in the notations above.

We have

$$\frac{\widehat{N}}{\widehat{D}} - \frac{\widetilde{N}}{\widetilde{D}} \;=\; \frac{\widehat{N}-\widetilde{N}}{\widehat{D}} - \widetilde{N}\left(\frac{\widehat{D}-\widetilde{D}}{\widehat{D}\widetilde{D}}\right) = \left(\frac{\widehat{N}-\widetilde{N}}{D} - \frac{N}{D^2}\left(\widehat{D}-\widetilde{D}\right)\right)(1+o_P(1))$$

under condition (A5). Next, write

$$\begin{aligned}
&\big[\widehat{N}-\widetilde{N}\big]\widehat{f}_{V,Z}(v,z)\\
&= \frac{1}{n}\sum_{j=1}^{n}\left[\int_0^\infty \mathbb{1}(Y_j \geq uy)g(u)du\right]\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)\Big[K_{h_v}(v-\widehat{V}_j) - K_{h_v}(v-V_j)\Big]\\
&= -\frac{1}{n}\sum_{j=1}^{n}\left[\int_0^{Y_j/y} g(u)du\right]\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)h_v^{-1}K'_{h_v}(v-V_j)(\widehat{V}_j - V_j)(1+o_P(1))\\
&\big[\widehat{D}-\widetilde{D}\big]\widehat{f}_{V,Z}(v,z)\\
&= \frac{1}{n}\sum_{j=1}^{n}\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)\Big[K_{h_v}(v-\widehat{V}_j) - K_{h_v}(v-V_j)\Big]\\
&= -\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)h_v^{-1}K'_{h_v}(v-V_j)(\widehat{V}_j - V_j)(1+o_P(1)).
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
&\widehat{\lambda}_m(x,y|v,z) - \widetilde{\lambda}_m(x,y|v,z)\\
&= \Big[\frac{\widehat{N}}{\widehat{D}} - \frac{\widetilde{N}}{\widetilde{D}}\Big](1+o_P(1))\\
&= \Big[-n^{-1}\sum_{j=1}^{n}\frac{1}{D}\mathbb{1}(X_j \leq x)K_{h_z}(z-Z_j)h_v^{-1}K'_{h_v}(v-V_j)(\widehat{V}_j - V_j)\Big(\int_0^{Y_j/y} g(u)du - \frac{N}{D}\Big)\Big]\\
&\qquad\times(1+o_P(1)).
\end{aligned}$$

We have that

$$\begin{aligned}\widehat{V}_j - V_j &= \frac{n^{-1}\sum_{k=1}^n \Big[I(X_k^1 \le X_j^1) - F_{X^1|W}(X_j^1|W=W_j)\Big]K_{h_w}(W_j - W_k)}{n^{-1}\sum_{k=1}^n K_{h_w}(W_j - W_k)} \\ &= \frac{n^{-1}\sum_{k=1}^n \Big[I(X_k^1 \le X_j^1) - F_{X^1|W}(X_j^1|W=W_j)\Big]K_{h_w}(W_j - W_k)}{f_W(W_j)} \\ &\quad \times(1+o_P(1)),\end{aligned}$$

uniformly in j, and hence $\widehat{\lambda}_m(x,y|v,z) - \widetilde{\lambda}_m(x,y|v,z)$ can be written as a V-statistic of order 2, ignoring the factors of order $(1+o_P(1))$ in what follows. Since a V-statistic can be written as a U-statistic plus negligible terms, we can focus attention on

$$\begin{aligned}&n^{-2}\sum_{j\ne k} \mathbb{1}(X_j \le x)K_{h_z}(z-Z_j)h_v^{-1}K'_{h_v}(v-V_j)\Big(\int_0^{Y_j/y} g(u)du - \frac{N}{D}\Big) \\ &\quad\times\Big[I(X_k^1 \le X_j^1) - F_{X^1|W}(X_j^1|W=W_j)\Big]\frac{K_{h_w}(W_j-W_k)}{f_W(W_j)} \\ &= n^{-2}\sum_{j\ne k} Q(T_j,T_k) \qquad \text{(say)}.\end{aligned}$$

To obtain a linear expansion for this U-statistic, we will use Lemma 3.1 in Powell, Stock and Stoker (1989), who give conditions under which a Hoeffding type decomposition is valid when the kernel of the U-statistic depends on n. For this, note that straightforward calculations show that

$$\begin{aligned}n^{-1}\sum_{j=1}^n \big\{E[Q(T_j,T_k)|T_j] - E[Q(T_j,T_k)]\big\} &= O_P\big(h_w^2(nh_v^3h_z^d)^{-1/2}\big) \\ n^{-1}\sum_{k=1}^n \big\{E[Q(T_j,T_k)|T_k] - E[Q(T_j,T_k)]\big\} &= O_P\big(n^{-1/2}\big) \\ E[Q(T_1,T_2)] &= O\big(h_z^2+h_v^2+h_w^2\big) \\ E[Q^2(T_1,T_2)] &= O\big(h_z^{-d}h_v^{-3}h_w^{-q}\big).\end{aligned}$$

Lemma 3.1 in Powell, Stock and Stoker (1989) requires that the U-statistic is standardized in such a way that $E[Q^2(T_j,T_k)] = o(n)$. Hence, in order to apply the latter result, we need to multiply our U-statistic by $(h_z^d h_v^3 h_w^q)^{1/2} n^{1/2}\delta_n$ for some $\delta_n \to 0$, and we choose $\delta_n = (nh_v^2h_w^q)^{-1/2}$ which goes to zero by assumption (A1). Then, we find that

$$\begin{aligned}&n^{-2}\sum_{j\ne k} Q(T_j,T_k) \\ &= O_P\big(h_w^2(nh_v^3h_z^d)^{-1/2} + n^{-1/2} + h_z^2 + h_v^2 + h_w^2\big) + o_P\big(n^{-1}(h_v^3h_z^dh_w^q)^{-1/2}\delta_n^{-1}\big) \\ &= o_P\big((nh_vh_z^d)^{-1/2}\big)\end{aligned}$$

again by assumption (A1). This ends the proof of (i).

Consider now the proof of (ii). The idea of the proof is to adapt Theorem 3.1 in Cazals, Florens and Simar (2002), who developed a similar result in the input-oriented setting. Consider the operator T defined by:

$$T(F)(x,y,v,z) = \int_0^\infty \Big(1 - F_{Y|X,V,Z}\left(uy|X \leq x,(V,Z)=(v,z)\right)^m\Big)du. \tag{B.4}$$

The operator T associates a real value to any distribution function F. The operator is Fréchet differentiable with respect to the sup-norm and following Cazals, Florens and Simar (2002), we can compute its Fréchet derivative:

$$T(\widehat{F}) - T(F) = dT(F)(\widehat{F} - F) + o_P(\|\widehat{F} - F\|).$$

Then, applying this expansion to $(nh_vh_z^d)^{1/2}\big(\widetilde{\lambda}_m(x,y|v,z) - \lambda_m(x,y|v,z)\big)$, we obtain the following leading term:

$$\begin{aligned}
&\frac{(nh_vh_z^d)^{1/2}}{n}\sum_{i=1}^n \frac{m}{F_{X|V,Z}(x|v,z)f_{V,Z}(v,z)}K_{h_v}(v-V_i)K_{h_z}(z-Z_i)\,\mathbb{1}(X_i \leq x)\\
&\quad\times\left\{\int_0^\infty \Big(1 - F_{Y|X,V,Z}(uy|X\leq x,v,z)^{m-1}\,\mathbb{1}(Y_i \leq uy)\Big)du - \lambda_m(x,y|v,z)\right\}\\
&= \frac{(nh_vh_z^d)^{1/2}}{n}\sum_{i=1}^n A(X_i,Y_i;x,y,v,z)K_{h_v}(v-V_i)K_{h_z}(z-Z_i),
\end{aligned}$$

where

$$\begin{aligned}
A(X,Y;x,y,v,z) \;=\; & \frac{m}{F_{X|V,Z}(x|v,z)f_{V,Z}(v,z)}\,\mathbb{1}(X\leq x)\\
&\times\left\{\int_0^\infty \Big(1 - F_{Y|X,V,Z}(uy|X\leq x,v,z)^{m-1}\,\mathbb{1}(Y_i \leq uy)\Big)du - \lambda_m(x,y|v,z)\right\}.
\end{aligned}$$

As $o_P(\|\widehat{F} - F\|) = o_P\big((nh_vh_z^d)^{-1/2}\big)$, the residual term converges to 0. The result now follows from the central limit theorem for triangular arrays applied to the leading term.

References

[1] Aigner, D.J. and Chu, S.F., 1968, On estimating the industry production function, American Economic Review 58, 826–839.

[2] Aigner, D.J., Lovell, C.A.K. and Schmidt, P., 1977, Formulation and estimation of stochastic frontier models, Journal of Econometrics 6, 21–37.

[3] Aly, H.Y., Grabowski, R., Pasurka, C. and Rangan, N., 1990, Technical, scale, and allocative efficiencies in U.S. banking: An empirical investigation, Review of Economics and Statistics 72, 211–218.

[4] Angrist, J.D. and Krueger, A.B., 1991, Does compulsory school attendance affect schooling and earnings?, Quaterly Journal of Economics 106, 979–1014.

[5] Angrist, J.D. and Krueger, A.B., 2001, Instrumental variables and the search for identification: from supply and demand to natural experiments, The Journal of Economic Perspectives 15, 69–85.

[6] Aragon, Y., Daouia, A. and Thomas-Agnan, C., 2005, Nonparametric frontier estimation: a conditional quantile-based approach, Econometric Theory 21, 358–389.

[7] Bădin, L., Daraio, C. and Simar, L., 2010, Optimal bandwidth selection for conditional efficiency measures: a data-driven approach, European Journal of Operational Research 201, 633–640.

[8] Bădin, L., Daraio, C. and Simar, L., 2012, How to measure the impact of environmental factors in a nonparametric production model? European Journal of Operational Research, 223, 818–833.

[9] Card, D., 1995, Using geographic variation in college proximity to estimate the return to schooling, In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp, Toronto: University of Toronto Press.

[10] Cazals, C., Florens, J.P. and Simar, L., 2002, Nonparametric frontier estimation: a robust approach. Journal of Econometrics 106, 1–25.

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

[12] Charnes, A., Cooper, W.W. and Rhodes, E., 1981, Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through, Management Science 27, 668–697.

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

[14] Daraio, C. and Simar, L., 2005, Introducing environmental variables in nonparametric frontier models: a probabilistic approach. Journal of Productivity Analysis 24, 93–121.

[15] Daraio, C. and Simar, L., 2007, Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications, Springer, New-York.

[16] Daraio, C., Simar, L. and Wilson, P.W., 2010, Testing whether two-stage estimation is meaningful in non-parametric models of production, Discussion Paper DP1031, ISBA.

[17] Debreu, G., 1951, The coefficient of resource utilization, Econometrica 19:3, 273–292.

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

[19] Farrell, M.J., 1957, The measurement of productive efficiency, Journal of the Royal Statistical Society - Series A 120, 253–281.

[20] Florens, J.P. and Simar, L., 2005, Parametric approximations of nonparametric frontier, Journal of Econometrics 124, 91–116.

[21] Florens, J.P., Simar, L. and Van Keilegom, I., 2013, Frontier estimation in nonparametric location-scale models, Journal of Econometrics 178, 456–470.

[22] Greene, W.H., 1980, Maximum likelihood estimation of econometric frontier, Journal of Econometrics 13, 27–56.

[23] Hall, P., Racine, J.S. and Li, Q., 2004, Cross-validation and the estimation of conditional probability densities, Journal of the American Statistical Association 99, 486, 1015–1026.

[24] Hayfield, T. and Racine, J.S., 2008, Nonparametric econometrics: the np package, Journal of Statistical Software 27,(5).

[25] Imbens, G.W. and Newey, W.K., 2009, Identification and estimation of triangular simultaneous equations models without additivity, Econometrica 77, 1481–1512.

[26] Jeong, S.O., Park, B.U. and Simar, L., 2010, Nonparametric conditional efficiency measures: asymptotic properties, Annals of Operations Research 173, 105–122.

[27] Koopmans, T.C., 1951, An analysis of production as an efficient combination of activities, in Koopmans, T.C. (ed) Activity Analysis of Production and Allocation, Cowles Commission for Research in Economics, Monograph 13, John-Wiley, New-York.

[28] Li, Q. and Racine, J., 2007, Nonparametric Econometrics: Theory and Practice. Princeton University Press.

[29] Li, Q. and Racine, J., 2008, Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data, Journal of Business & Economic Statistics 26, 423–434.

[30] Mammen, E., Rothe, C. and Schienle, M., 2012, Nonparametric regression with nonparametrically generated regressors, Annals of Statistics 40, 1132–1170.

[31] Matzkin, R.L., 2003, Nonparametric estimation of nonadditive random functions, Econometrica 71, 1339–1375.

[32] Meeusen, W. and Van den Broek, J., 1977, Efficiency estimation from Cobb-Douglas production functions with composed errors, International Economic Review 18, 435–444.

[33] Newey, W.K., Powell, J.L. and Vella, F., 1999, Nonparametric estimation of triangular simultaneous equations models, Econometrica 67, 565–603.

[34] Park, B.U., Simar, L. and Weiner, C., 2000, The FDH estimator for productivity efficiency scores: asymptotic properties, Econometric Theory 16, 855–877.

[35] Powell, J.L., Stock, J.H. and Stoker, T.M., 1989, Semiparametric estimation of index coefficients, Econometrica 57, 1403–1430.

[36] Simar, L. and Wilson, P.W., 2007, Estimation and inference in two-stage, semiparametric models of production processes, Journal of Econometrics 136, 31–64.

[37] Simar, L. and Wilson, P.W., 2011, Two-stage DEA: Caveat Emptor, Journal of Productivity Analysis 36, 205–218.

[38] Simar, L. and Wilson, P.W., 2013, Estimation and inference in nonparametric frontier models: Recent developments and perspectives, Foundations and Trends® in Economics, Vol. 5: No 3-4, 183–337.

[39] Vanhems, A. and Van Keilegom, I., 2013, Semiparametric transformation model with endogeneity: a control function approach, Journal of Econometrics (under revision).