Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 137 (2007) 1–27
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Nonparametric stochastic frontiers: A local maximum likelihood approach

Subal C. Kumbhakar[a,*], Byeong U. Park[b,1], Léopold Simar[c,2], Efthymios G. Tsionas[d]

[a]*Department of Economics, State University of New York at Binghamton, USA*
[b]*Department of Statistics, Seoul National University, Republic of Korea*
[c]*Institut de Statistique, Université Catholique de Louvain, Belgium*
[d]*Department of Economics, Athens University of Economics and Business, Greece*

Available online 2 May 2006

Abstract

This paper proposes a new approach to handle nonparametric stochastic frontier (SF) models. It is based on local maximum likelihood techniques. The model is presented as encompassing some anchorage parametric model in a nonparametric way. First, we derive asymptotic properties of the estimator for the general case (local linear approximations). Then the results are tailored to a SF model where the convoluted error term (efficiency plus noise) is the sum of a half normal and a normal random variable. The parametric anchorage model is a linear production function with a homoscedastic error term. The local approximation is linear for both the production function and the parameters of the error terms. The performance of our estimator is then established in finite samples using simulated data sets as well as with a cross-sectional data on US commercial banks. The methods appear to be robust, numerically stable and particularly useful for investigating a production process and the derived efficiency scores.

JEL classification: C13; C14

Keywords: Stochastic cost frontier; Banking; Nonparametric; Local maximum likelihood

*Corresponding author. Tel.: +1 607 777 4762; fax: +1 607 777 2681.
E-mail addresses: kkar@binghamton.edu (S.C. Kumbhakar), bupark@stats.snu.ac.kr (B.U. Park), simar@stat.ucl.ac.be (L. Simar), tsionas@aueb.gr (E.G. Tsionas).
[1]Research of B. U. Park was supported by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.
[2]Research support from the "Interuniversity Attraction Pole", Phase V (No. P5/24) from the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.03.006

1. Introduction

The economic function of a firm is to bid resources away from alternative uses. As a result of such resource transfer, aggregate output may be increased or decreased. If inefficiency exists, an increase in output can be achieved by reallocating resources to more efficient uses. Given the seriousness of the issue (viz., the economic, political and social implications of inefficiency), it is essential that the measurement of efficiency/performance should be theoretically valid and subject to unambiguous interpretations.

Econometric measurement of efficiency of firms goes back to Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). When analyzing the performance of a firm that is observed to produce an output level $y \in \mathbb{R}$ using input quantities $x \in \mathbb{R}^d$, one typically compares the observed output level with the maximum possible output that can be obtained from the production frontier (defined as $f(x) = \max\{y : y \in P(x)\}$ where $P(x)$ describes the set of outputs that are feasible for each input vector $x \in \mathbb{R}^d$). The estimation of the production frontier is obtained from a random sample of observed firms $\{(X_i, Y_i) \mid i = 1, \ldots, n\}$. Then an efficiency score for a given production plan (x, y) is derived from the distance of this point to the estimated production frontier. The same approach can also be applied when a cost frontier is analyzed. In the latter we seek the minimal cost achievable for a given level of output(s).

Since the publication of the seminal papers by Aigner et al. (1977) and Meeusen and van den Broeck (1977), econometric estimation of parametric stochastic frontier (SF) functions has become a standard practice in efficiency measurement studies. However, in this approach, the estimation relies heavily on the particular choices of the functional form of the production/cost frontier (Cobb–Douglas, Translog, etc.) as well as the specific distributional assumptions on the error term (a convolution of a one-sided inefficiency term and a two-sided noise term). Typically, these parametric models are written as

$$Y_i = \beta_0 + \beta^{\mathrm{T}} X_i - u_i + v_i, \quad i = 1, \ldots, n,$$

where $u_i > 0$ is the inefficiency term and $v_i \in \mathbb{R}$ represents random noise. The estimation technique is straightforward and is primarily based on the maximum likelihood principle.[3] Of course, in practice we cannot be confident about the validity of these parametric assumptions that are used to estimate the model. The parametric form of the frontier function might be wrong due to several reasons. For example, the parametric functional form might be wrong, the stochastic specifications of the error components (particularly for the inefficiency component) might be wrong, among others.

An alternative to the parametric SF is the deterministic nonparametric approach where no specific parametric assumptions are made on the model. The frontier is defined as the upper boundary of the attainable set, say $\Psi = \{(x, y) \mid x \text{ can produce } y\}$. In these nonparametric approaches the statistical properties of envelopment estimators like DEA and FDH[4] (Farrell, 1957; Charnes et al., 1978; Deprins et al., 1984), rely on the so-called "deterministic" assumption, viz.,

$$\mathrm{Prob}((X_i, Y_i) \in \Psi) = 1. \tag{1.1}$$

This assumption implies that no noise is allowed in these deterministic frontier (DF) models. The introduction of noise in a full nonparametric setup is problematic due to

[3] Other methods, such as COLS and MOLS (see Kumbhakar and Lovell, 2000), are often used.

[4] DEA and FDH are acronyms for Data Envelopment Analysis and Free Disposal Hull, respectively.

identification problems (see Hall and Simar, 2002). Statistical inference is now available in these nonparametric DF models (see Simar and Wilson, 2000, for a recent survey) but assumption (1.1) is too strong in many practical situations where we might expect measurement error, random shocks, etc. Recently Cazals et al. (2002), Aragon et al. (2002) and Daouia and Simar (2004) have proposed robust versions of the FDH estimator, robust to extremes values and/or outliers since they do not envelop all the data. But these approaches still rely heavily on the deterministic assumption (1.1), where no noise is allowed.

In the presence of panel data, Park and Simar (1994), Park et al. (1998, 2003, 2006), in a series of papers, consider the semiparametric estimation of SF panel models under various assumptions on the joint distribution of the random firm effects and the regressors and on various dynamic specifications. The nonparametric part of these models concerns the distribution of the inefficiency terms. However, the estimators in these panel models are based on the linearity of the efficient frontier.

Fan et al. (1996) propose a two-step pseudo-likelihood estimator in a semiparametric model where the production frontier is not specified, but distributional assumptions are imposed on the stochastic components as in Aigner et al. (1977). An average production frontier is then estimated through standard kernel methods, the shift for the frontier is obtained through a moment condition, as in the MOLS approach (see Kumbhakar and Lovell, 2000) and the remaining parameters of the stochastic components are estimated by maximizing a pseudolikelihood function.

Our purpose in this paper is to propose a new approach to handle nonparametric SF models. The method is based on the local maximum likelihood principle (see Tibshirani and Hastie, 1987, or Fan and Gijbels, 1996), which is nonparametric in the sense that the parameters of a given local polynomial model are localized with respect to the covariates of the model. Our approach extends and generalizes Fan et al. (1996), not only by considering more general order-m local polynomial estimators but also by localizing the parameters of the stochastic component of the model. Note that our estimator is obtained through a one-step maximization procedure. As pointed out by Gozalo and Linton (2000), localizing can be viewed as a way of nonparametrically encompassing a parametric "anchorage" model. The idea to use local likelihood method for SF models was first suggested by Kumbhakar and Tsionas (2002) for a particular case of the model proposed here. In this paper we develop the general theory along with the asymptotic properties. We also investigate how the procedure works in practice with some simulated data sets. Finally, we provide a banking application using a cross-sectional data on 500 US banks for the year 2000.

The paper is organized as follows. Section 2 presents the model and the theory with the main asymptotic results. Section 3 analyzes the practical side, viz., how to compute the local estimators and determine the bandwidth. Section 4 reports results from both simulated and real data sets, and finally Section 5 concludes the paper. Regularity conditions and proofs are given in Section 6.

2. Main results

2.1. The model

We consider a set of i.i.d. random variables (X_i, Y_i), for $i = 1, \ldots, n$ with $X_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$. The joint pdf of (X, Y) is decomposed into a marginal pdf for X: $\mathrm{pdf}(x) = p(x)$

and a conditional pdf for Y given X: $\mathrm{pdf}(y|x) = g(y, \theta(x))$, where $\theta(x) \in \mathbb{R}^k$ is unknown and has to be estimated. The function g is assumed to be known.

The localized version of the parametric model in Aigner et al. (1977) is a particular case of our model. In this case, the conditional pdf for Y given $X = x$ would, for instance, be characterized through

$$Y = r(X) - u + v,$$

where $r(x)$ is the frontier function, $u|X = x \sim |\mathcal{N}(0, \sigma_u^2(x))|$ and $v|X = x \sim \mathcal{N}(0, \sigma_v^2(x))$, u and v being independent conditionally on X. Here, $\theta(x) = (r(x), \sigma_u^2(x), \sigma_v^2(x))^{\mathrm{T}}$ is a three-dimensional local parameter. In our approach here we will consider local polynomial approximations for $\theta(x)$. For simplicity of presentation, we treat only the cases where the orders of local polynomials are equal for all the components of $\theta(\cdot)$. In practice, one may prefer to use different orders of polynomials for different components. The theory for the latter cases may be obtained along the same lines of development as for the equal order cases.

The conditional log-likelihood can thus be written as

$$L(\theta) = \sum_{i=1}^{n} \log g(Y_i, \theta(X_i)).$$

In the next section we consider the order-m local polynomial estimator of $\theta(x)$ when x is univariate and then, in the following section, we derive the local linear estimator of $\theta(x)$ when x is multivariate.

2.2. Univariate case

Here, we consider the case when $d = 1$. Let x be a fixed interior point in the support of $p(x)$ and let $q \equiv \log g$. Denote, $\theta_j \equiv (\theta_{j1}, \ldots, \theta_{jk})^{\mathrm{T}}$, for $j = 0, 1, \ldots, m$. Then the conditional local log-likelihood for the mth order local polynomial fit is given by

$$L_n(\theta_0, \theta_1, \ldots, \theta_m) = \sum_{i=1}^{n} q(Y_i, \theta_0 + \theta_1(X_i - x) + \cdots + \theta_m(X_i - x)^m)K_h(X_i - x),$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$, K being a kernel function and h the appropriate bandwidth. Thus, the log-likelihood depends on the local x. The local polynomial estimator $\widehat{\theta}(x)$ is given by $\widehat{\theta}(x) = \widehat{\theta}_0(x)$ where

$$(\widehat{\theta}_0(x), \ldots, \widehat{\theta}_m(x)) = \arg \max_{\theta_0, \ldots, \theta_m} L_n(\theta_0, \theta_1, \ldots, \theta_m).$$

We give the asymptotic distribution of $\widehat{\theta}(x)$. For this we need to introduce some additional notations. We define for $v \in \mathbb{R}^k$,

$$\begin{aligned}
q_1(y, v) &= \frac{\partial}{\partial v} q(y, v), \\
q_2(y, v) &= \frac{\partial^2}{\partial v \partial v^{\mathrm{T}}}\, q(y, v), \\
\rho(x) &= -\mathrm{E}[q_2(Y_1, \theta(X_1))|X_1 = x].
\end{aligned}$$

For $\mu_j = \int u^j K(u)\,\mathrm{d}u$, $\kappa_j = \int u^j K^2(u)\,\mathrm{d}u$ and $\theta_j^{(m+1)}(x) = (\partial^{m+1}/\partial x^{m+1})\theta_j(x)$, we also introduce the following matrices and vectors:

$$N = \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_m \\ \mu_1 & \mu_2 & \cdots & \mu_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_m & \mu_{m+1} & \cdots & \mu_{2m} \end{pmatrix}, \quad S = \begin{pmatrix} \kappa_0 & \kappa_1 & \cdots & \kappa_m \\ \kappa_1 & \kappa_2 & \cdots & \kappa_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_m & \kappa_{m+1} & \cdots & \kappa_{2m} \end{pmatrix},$$

$$\theta^{(m+1)}(x) = \begin{pmatrix} \theta_1^{(m+1)}(x) \\ \vdots \\ \theta_k^{(m+1)}(x) \end{pmatrix}, \quad \gamma = \begin{pmatrix} \mu_{m+1} \\ \vdots \\ \mu_{2m+1} \end{pmatrix}.$$

Finally, we define

$$v(x) = \mathrm{E}[q_1(Y_1, \theta(X_1))q_1^{\mathrm{T}}(Y_1, \theta(X_1))|X = x].$$

Using this notation, we now state our theorem for the univariate case.

Theorem 2.1. *Under regularity conditions* (*see Section* 6.1), *it follows that*

$$\sqrt{nh}[(e_0^{\mathrm{T}} N^{-1} S N^{-1} e_0)\rho(x)^{-1} v(x)\rho(x)^{-1}/p(x)]^{-1/2}$$
$$\times \left(\widehat{\theta}(x) - \theta(x) - \frac{h^{m+1}}{(m+1)!}(e_0^{\mathrm{T}} N^{-1}\gamma)\theta^{(m+1)}(x)\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, I_k),$$

where $e_0 = (1, 0, \ldots, 0)^{\mathrm{T}}$ *is the* $(m+1)$-*dimensional unit vector*.

If m is even and K is symmetric, the constant factor $e_0^{\mathrm{T}} N^{-1}\gamma$ of the bias turns out to be zero. In that case, the order of the bias can be shown to be h^{m+2} with a stronger smoothness assumption on θ. However, it is widely accepted that in local polynomial estimation odd order fits are preferable to even order (see Section 3.3 of Fan and Gijbels, 1996 for a discussion on choosing the polynomial order). As a special case where $m = 1$ (local linear estimator), we have

$$e_0^{\mathrm{T}} N^{-1}\gamma = \frac{\mu_2^2 - \mu_1\mu_3}{\mu_0\mu_2 - \mu_1^2} = \mu_2,$$
$$e_0^{\mathrm{T}} N^{-1} S N^{-1} e_0 = \frac{\kappa_0\mu_2^2 - 2\kappa_1\mu_1\mu_2 + \kappa_2\mu_1^2}{(\mu_0\mu_2 - \mu_1^2)^2} = \kappa_0,$$

where the two equalities on the right are valid for appropriate symmetric kernels, with $\mu_0 = 1$ and $\mu_1 = \mu_3 = 0$.

We notice, as for most of kernel based nonparametric estimators, the role of the bandwidth h to balance the bias and the variance of the estimator. An optimal asymptotic value of h can be derived from the above theorem[5] (e.g., minimizing the asymptotic MSE would lead to $h \asymp n^{-1/(2m+3)}$ for an odd m).

[5]In practice the bandwidth choice may be based on cross-validation. We address this issue in Section 3.

2.3. Multivariate case

In this case, $\theta(x)$ is a function of $x \in \mathbb{R}^d$ and we consider only the local linear fit. The results can be extended to higher order local polynomial fits with more complicated notations. For many applications the local linear fit would suffice as we see in our numerical examples presented in Section 4. For simplicity of presentation, we assume further that the multivariate kernel K satisfies

$$\int K(u)\,\mathrm{d}u = 1 \quad \text{and} \quad \int uu^{\mathrm{T}} K(u)\,\mathrm{d}u = \mu_2 I_d.$$

In the multiple covariate case, the conditional local linear log-likelihood is given by

$$L_n(\theta_0, \Theta_1) = \sum_{i=1}^{n} q(Y_i, \theta_0 + \Theta_1(X_i - x))K_H(X_i - x).$$

Here θ_0 is a $k \times 1$ vector, Θ_1 is a $k \times d$ matrix, H is a bandwidth matrix which we assume to be positive definite and symmetric, and $K_H(u) = |H|^{-1}K(H^{-1}u)$. For instance we could choose a multivariate product kernel, viz.,

$$K(u) = K_1(u_1) \cdots K_1(u_d),$$

where $K_1(\cdot)$ is a symmetric univariate probability density. In this case

$$\int uu^{\mathrm{T}} K(u)\,\mathrm{d}u = \left(\int u_1^2 K_1(u_1)\,\mathrm{d}u_1\right) I_d.$$

The local linear estimator $\widehat{\theta}(x)$ is given by

$$\widehat{\theta}(x) = \widehat{\theta}_0(x),$$

where $\widehat{\theta}_0(x)$ and $\widehat{\Theta}_1(x)$ maximize $L_n(\theta_0, \Theta_1)$ with respect to θ_0 and Θ_1.

Define $\rho(\cdot)$ and $v(\cdot)$ as in the case $d = 1$. Let

$$B_H(x) = \begin{pmatrix} \mathrm{tr}(\theta_1''(x)H^2) \\ \vdots \\ \mathrm{tr}(\theta_k''(x)H^2) \end{pmatrix},$$

where $\theta_j(x)$ is the jth component of $\theta(x)$ and $\theta_j''(x)$ is the $(d \times d)$ Hessian matrix of $\theta_j(x)$. Now we can state our theorem (the proof is omitted).

Theorem 2.2. *Under the regularity conditions* (*see Section* 6.1), *it follows that*

$$(n|H|)^{1/2}\left[\frac{1}{p(x)}\int K^2(u)\,\mathrm{d}u \cdot \rho(x)^{-1} v(x) \rho(x)^{-1}\right]^{-1/2} \times \left(\widehat{\theta}(x) - \theta(x) - \frac{1}{2}\mu_2 B_H(x)\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, I_k).$$

3. Practical computations

3.1. Computation of the local linear estimator

We will illustrate how the computations could be performed in the case of a local linear fit, namely, choosing for the local convolution of the noise and the inefficiency term, the convolution of a normal and a half normal, in the spirit of Aigner et al. (1977) and Meeusen and van den Broeck (1977). Other choices are also available, such as the convolution of a normal and an exponential, but to save space, we present the normal-half normal case. So the parametric "anchorage" model is the following:

$$Y = \beta_0 + \beta^{\mathrm{T}} X - u + v, \tag{3.1}$$

where $u \sim |\mathcal{N}(0, \sigma_u^2)|$ and $v \sim \mathcal{N}(0, \sigma_v^2)$, u and v being mutually independent and both independent of X.

Our nonparametric localized model is tailored as

$$Y = r(X) - u + v,$$

where $u|X = x \sim |\mathcal{N}(0, \sigma_u^2(x))|$ and $v|X = x \sim \mathcal{N}(0, \sigma_v^2(x))$, u and v being independent conditionally on X. The conditional probability density function of $\varepsilon = v - u$ is given by[6]

$$f(\varepsilon|X = x) = \frac{2}{\sigma(x)}\, \varphi\left(\frac{\varepsilon}{\sigma(x)}\right) \Phi\left(-\varepsilon\, \frac{\lambda(x)}{\sigma(x)}\right),$$

where $\sigma^2(x) = \sigma_u^2(x) + \sigma_v^2(x)$ and $\lambda(x) = \sigma_u(x)/\sigma_v(x)$. Finally, $\varphi(\cdot)$ and $\Phi(\cdot)$ are the pdf and CDF of a standard normal variable.

In order to avoid nonnegativity restrictions[7] on the variance functions $\sigma^2(x)$ and $\lambda(x)$ in the local linear approximations, we choose rather the following coordinate system: $\theta(x) = (r(x), \widetilde{\sigma}^2(x), \widetilde{\lambda}(x))^{\mathrm{T}}$ where $\widetilde{\sigma}^2(x) = \log(\sigma^2(x))$ and $\widetilde{\lambda}(x) = \log(\lambda(x))$.

Thus the conditional pdf of Y given X is

$$g(y; \theta(x)) = \frac{2}{\exp(0.5\widetilde{\sigma}^2(x))}\, \varphi\left(\frac{y - r(x)}{\exp(0.5\widetilde{\sigma}^2(x))}\right) \Phi(-(y - r(x)) \exp(\widetilde{\lambda}(x) - 0.5\widetilde{\sigma}^2(x))),$$

and the conditional local log-likelihood is given by

$$L(\theta) \propto \sum_{i=1}^{n} \left[-\frac{1}{2}\, \widetilde{\sigma}^2(X_i) - \frac{1}{2} \frac{(Y_i - r(X_i))^2}{\exp(\widetilde{\sigma}^2(X_i))} \right. \\ \left. + \log \Phi\left(-(Y_i - r(X_i)) \exp\left(\widetilde{\lambda}(X_i) - \frac{1}{2}\, \widetilde{\sigma}^2(X_i) \right) \right) \right],$$

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

$$f(\varepsilon|X = x) = \frac{2}{\sigma(x)}\, \varphi\left(\frac{\varepsilon}{\sigma(x)}\right) \Phi\left(\varepsilon\, \frac{\lambda(x)}{\sigma(x)}\right).$$

[7]We thank an anonymous Associate Editor and a referee for this suggestion.

where the constants have been eliminated. The conditional local log-likelihood for the local linear fit at the point x, is finally given by

$$
\begin{aligned}
L_n(\theta_0, \Theta_1) \propto \sum_{i=1}^{n} \Bigg[& -\frac{1}{2}\,(\widetilde{\sigma}_0^2 + \widetilde{\sigma}_1^{2\mathrm{T}}(X_i - x)) - \frac{1}{2}\,\frac{(Y_i - r_0 - r_1^{\mathrm{T}}(X_i - x))^2}{\exp(\widetilde{\sigma}_0^2 + \widetilde{\sigma}_1^{2\mathrm{T}}(X_i - x))} \\
& + \log \Phi\Bigg(- (Y_i - r_0 - r_1^{\mathrm{T}}(X_i - x)) \\
& \times \exp\left(\widetilde{\lambda}_0 + \widetilde{\lambda}_1^{\mathrm{T}}(X_i - x) - \frac{1}{2}(\widetilde{\sigma}_0^2 + \widetilde{\sigma}_1^{2\mathrm{T}}(X_i - x))\right)\Bigg)\Bigg] K_H(X_i - x).
\end{aligned} \tag{3.2}
$$

Here, the local parameters are the (3×1) vector $\theta_0 = (r_0, \widetilde{\sigma}_0^2, \widetilde{\lambda}_0)^{\mathrm{T}}$ and the $(3 \times d)$ matrix $\Theta_1^{\mathrm{T}} = (r_1,\ \widetilde{\sigma}_1^2, \widetilde{\lambda}_1)$ with $r_1, \widetilde{\sigma}_1^2$ and $\widetilde{\lambda}_1$ being $(d \times 1)$ vectors. The local linear estimator of the model is given by $\widehat{\theta}_0(x)$ where

$$
(\widehat{\theta}_0(x), \widehat{\Theta}_1(x)) = \arg\max_{\theta_0, \Theta_1} L_n(\theta_0, \Theta_1). \tag{3.3}
$$

From $\widehat{\widetilde{\sigma}}{}_0^2(x)$ and $\widehat{\widetilde{\lambda}}_0(x)$, values of $\widehat{\sigma}_0^2(x)$, $\widehat{\lambda}_0(x)$, $\widehat{\sigma}_u^2(x)$, $\widehat{\sigma}_v^2(x)$ can also be derived.

The estimation of the individual efficiency score for a particular point (x, y) might be obtained from the Jondrow et al. (1982) procedure. It can be shown that

$$
u|\varepsilon, X = x \sim \mathcal{N}^{+}(\mu^*(x), \sigma^{2*}(x)),
$$

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

$$
\mu^*(x) = \frac{-\varepsilon \sigma_u^2(x)}{\sigma^2(x)},
$$

$$
\sigma^{2*}(x) = \frac{\sigma_u^2(x)\sigma_v^2(x)}{\sigma^2(x)}.
$$

In particular, we can compute

$$
\mathrm{E}(u|\varepsilon, X = x) = \frac{\sigma(x)\lambda(x)}{1 + \lambda^2(x)} \left[\frac{\varphi(-\varepsilon\lambda(x)/\sigma(x))}{\Phi(-\varepsilon\lambda(x)/\sigma(x))} - \frac{\varepsilon\lambda(x)}{\sigma(x)} \right].
$$

As in Jondrow et al. (1982) a point estimator of the individual efficiency score for an observation (X_i, Y_i) could be obtained from

$$
\widehat{u}_i = \frac{\widehat{\sigma}_0(X_i)\widehat{\lambda}_0(X_i)}{1 + \widehat{\lambda}_0^2(X_i)} \left[\frac{\varphi(-\widehat{\varepsilon}_i\widehat{\lambda}_0(X_i)/\widehat{\sigma}_0(X_i))}{\Phi(-\widehat{\varepsilon}_i\widehat{\lambda}_0(X_i)/\widehat{\sigma}_0(X_i))} - \frac{\widehat{\varepsilon}_i\widehat{\lambda}_0(X_i)}{\widehat{\sigma}_0(X_i)} \right], \tag{3.4}
$$

where $\widehat{\varepsilon}_i = Y_i - \widehat{r}_0(X_i)$. We know of course, as in the full parametric case, that this is a rather poor predictor of u_i, since it is based on a single "noisy" observation $\widehat{\varepsilon}_i$. As usual in frontier models, if the variables are measured in logs, a point estimate of the efficiency is then provided by $\widehat{\mathrm{eff}}_i = \exp(-\widehat{u}_i) \in [0, 1]$.

Solving (3.3) requires iterative algorithms, starting with some initial values. For example, we could choose $\Theta_1^{\mathrm{T}} = (\widehat{\beta}, 0_d, 0_d)$ and $\theta_0 = (\widehat{\beta}_0, \log(\widehat{\sigma}^2), \log(\widehat{\lambda}))$, the parametric maximum likelihood estimator from model (3.1), as the initial values of Θ_1 and θ_0.

The following is an alternative way to choose the starting values. Start with the local linear least squares estimator $\widehat{r}_0(x)$ and $\widehat{r}_1(x)$ of Gozalo and Linton (2000) and then correct the local intercept $\widehat{r}_0(x)$ for the moment condition along the lines of the parametric MOLS estimators (see, e.g., Kumbhakar and Lovell, 2000). Here, we suggest the use of the global parametric MLE $\widehat{\sigma}^2$ and $\widehat{\lambda}$ for the moment correction. Since the local linear $\widehat{r}_0(x)$ can be viewed as an estimator of $r_0(x) - \mathrm{E}(u|X=x) = r_0(x) - \sqrt{2\sigma_u^2(x)/\pi}$, a sensible moment corrected estimator for the local intercept is obtained through $\widehat{r}_0^{\mathrm{MOLS}}(x) = \widehat{r}_0(x) + \sqrt{2\widehat{\sigma}_u^2/\pi}$, where $\widehat{\sigma}_u^2 = \widehat{\sigma}^2\widehat{\lambda}^2/(1+\widehat{\lambda}^2)$. In this case, the initial values for solving (3.3) are finally given by $\Theta_1^{\mathrm{T}} = (\widehat{r}_1(x), 0_d, 0_d)$ and $\theta_0 = (\widehat{r}_0^{\mathrm{MOLS}}(x), \widehat{\sigma}^2, \widehat{\lambda})$. In the optimization algorithms used in the numerical illustrations below, the latter choice of starting values proved to be very efficient and numerically stable.

3.2. Bandwidth selection

In practice we could choose a product kernel with a bandwidth $H = hV$ where V is the empirical variance–covariance matrix of the d covariates X_i. The choice of the bandwidth is then reduced to the selection of the scalar h. An alternative is to use the product kernel $h^{-d}\prod_{j=1}^{d} K(h^{-1}(x_j))$, so $H = hI_d$. In the numerical illustrations below we choose the more sensible d-dimensional vector of bandwidths h as

$$h = h_{\mathrm{base}} s_X n^{-1/5},$$

where s_X is the vector of empirical standard deviations of the d components of X. So the bandwidth is adjusted for different scales of the variables and different sample sizes. Then the cross-validation criterion is evaluated for a grid of values for h_{base}.

The cross-validation proceeds as follows. For a given value of h_{base}, we compute

$$CV(h_{\mathrm{base}}) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - (\widehat{r}_0^{(i)}(X_i) - \widehat{u}_i^{(i)})]^2, \tag{3.5}$$

where $\widehat{r}_0^{(i)}$ and $\widehat{u}^{(i)}$ are the leave-one-out version of the local linear estimators derived above. The optimal value for h_{base} is then easily found by an appropriate grid search.

If n is large, the evaluation in (3.5) could be performed on a random subsample of m units, where $m \ll n$ to reduce the computational burden. Also, a trimmed version of the average of (3.5) might be useful in the sense that it is less sensitive to potential numerical problems when computing the local ML many times.

4. Numerical illustrations

4.1. Some simulated data sets

In the following examples with simulated data sets, we investigate how our procedure is flexible and robust to deviations from the parametric anchorage model (3.1), either because the frontier function is not linear, or because the efficiency term is not homoscedastic or for both reasons. In all the computations below we choose the Gaussian kernel.

4.1.1. Example 1: a linear homoscedastic model

In this first case, we choose a standard linear model where the anchorage parametric model is true. We simulated a sample of size 100 from the model:

$$Y = 5 + 5X - u + v,$$

where $X \sim \mathrm{U}(0,1)$, $u \sim |\mathcal{N}(0, \sigma_u^2 = 1)|$ and $v \sim \mathcal{N}(0, \sigma_v^2)$, where $\sigma_v = 0.75 \times \mathrm{std}(u) = 0.75 \times \sigma_u \sqrt{(\pi - 2)/\pi}$. So the noise to signal ratio is 0.75, a pretty high value in this framework. The cross-validation, as expected, gave a flat value of $CV(h)$ for values of h_{base} greater than, say, 6. Indeed, larger values of h_{base} makes our local maximum likelihood estimator equal to the full parametric MLE, as it should be in this particular example. A trimmed version (deleting the upper 5% of the values in the average of (3.5)) gave very similar results. The fit proposed in Fig. 1 was obtained for a value of $h_{\mathrm{base}} = 21$ and is very good. The statistical noise introduced by encompassing the true linear model in the nonparametric world did not affect the estimation.

4.1.2. Example 2: a linear model with heteroscedasticity

This is the same model as in Example 1, but we introduce heteroscedasticity in the distribution of the inefficiencies. Here, we have $u|X = x \sim |\mathcal{N}(0, \sigma_u^2 \times (1+2x)^2)|$, where as above $\sigma_u = 1$ and $\sigma_v = 0.75 \times \sigma_u \sqrt{(\pi - 2)/\pi}$. Note that the heterogeneity in the distribution of the inefficiency term u is rather large: $\mathrm{E}(u|x)$ and $\mathrm{std}(u|x)$ vary by a factor of 3 over the range of X.

The results are still very good and similar to those obtained in the preceding example: a large optimal value for the bandwidth and a very good fit with $h_{\mathrm{base}} = 21$ (see Fig. 2). So the procedure is robust to the heteroscedasticity introduced here.

4.1.3. Example 3: a nonlinear exponential homoscedastic model

In this example we simulated a data set with a nonlinear model, so that the anchorage model is no longer the true one for the frontier, but still is for the stochastic part. We simulated a sample of size 100 from the model:

$$Y = 10 - 5\exp(-X) - u + v,$$

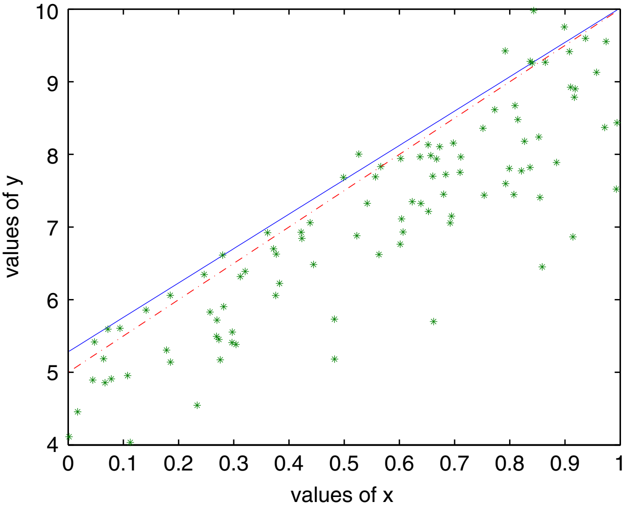


Fig. 1. Example 1, linear homoscedastic model: the true frontier (dashed-dotted) and the local maximum likelihood fit (solid).

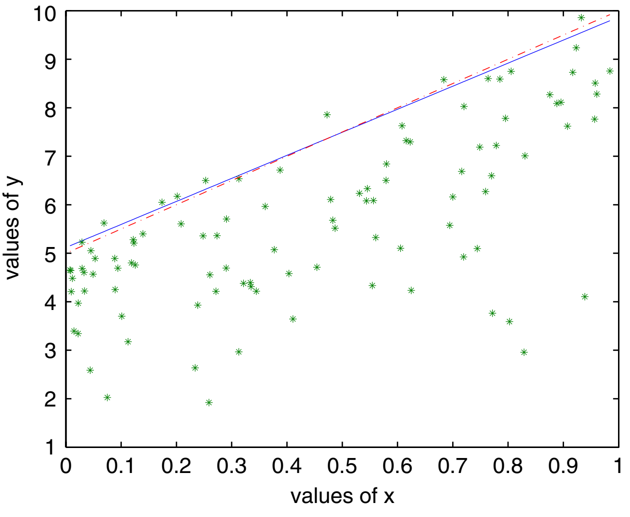


Fig. 2. Example 2, linear model with heteroscedasticity: the true frontier (dashed-dotted) and the local maximum likelihood fit (solid).

where $X \sim \mathrm{U}(0,3)$, $u \sim |\mathcal{N}(0, \sigma_u^2 = 2^2)|$ and $v \sim \mathcal{N}(0, \sigma_v^2)$, where $\sigma_v = 0.50 \times \mathrm{std}(u)$. The cross-validation provided $h_{\mathrm{base}} = 1.6$ and the resulting fit is shown in Fig. 3. We see that the localized version of the MLE is able to detect the curvature of the true model.

4.1.4. *Example 4: a nonlinear exponential model with heteroscedasticity*

We choose the same exponential model as above but we allow heteroscedasticity in the inefficiency term u which again allows the conditional mean and standard deviation of $u|x$ to change by a factor of 3 over the range of X. We now have $u|X = x \sim |\mathcal{N}(0, \sigma_u^2 \times (1 + x/3)^2)|$. The cross-validation provided $h_{\mathrm{base}} = 1.15$ and the resulting fit is shown in Fig. 4. We see again that the local MLE is able to detect the curvature of the true model and that the fit looks pretty well. Now, the chosen parametric anchorage model is wrong both for the frontier model and for the stochastic errors, but its localized estimated version fits the model pretty well. This illustrates the "nonparametric" nature of our estimator.

4.1.5. *Example 5: a quadratic homoscedastic model*

We choose here an alternative nonlinear model adding a quadratic term in the frontier model. We simulate a sample of size $n = 100$ with:

$$Y = 10 + 30X - 8X^2 - u + v,$$

where $X \sim \mathrm{U}(1,2)$, $u \sim |\mathcal{N}(0, \sigma_u^2 = 1)|$ and $v \sim \mathcal{N}(0, \sigma_v^2)$, where $\sigma_v = 0.75 \times \mathrm{std}(u)$. As shown in Fig. 5, the local MLE procedure ($h_{\mathrm{base}} = 1.45$) still performs very well.

4.1.6. *Example 6: a quadratic model with heteroscedasticity*

Here, we simulate the data ($n = 100$) with the same quadratic model with heteroscedasticity for the efficiency term, $u|X = x \sim |\mathcal{N}(0, \sigma_u^2 \times (1 + x/2)^2)|$, with $\sigma_u^2 = 1$. Here, the conditional mean and the standard deviation of u vary by a factor of 2 over the

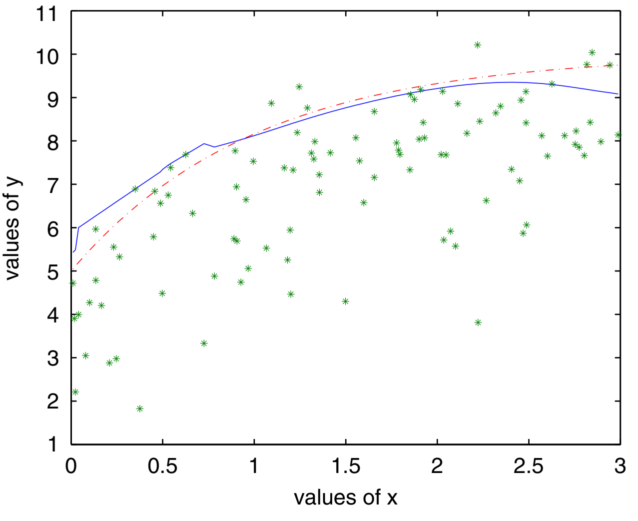


Fig. 3. Example 3, nonlinear model: the true frontier (dashed-dotted) and the local maximum likelihood fit (solid).

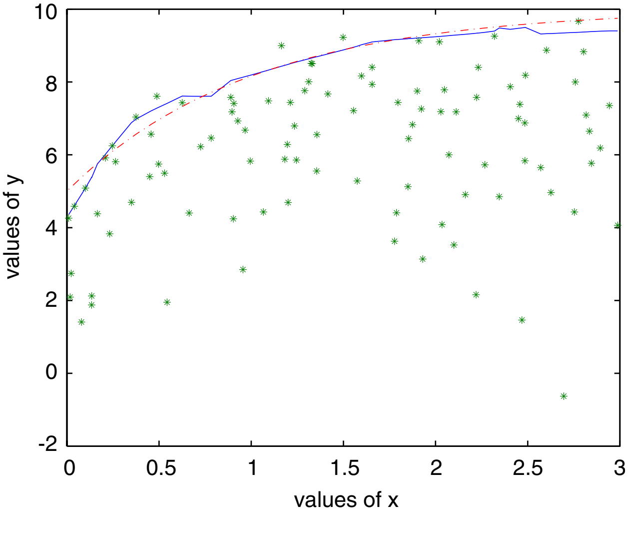


Fig. 4. Example 4, nonlinear model with heteroscedasticity: the true frontier (dashed-dotted) and the local maximum likelihood fit (solid).

range of X. The bandwidth selection procedure provided an optimal value of $h_{\text{base}} = 1.80$ and the excellent fit ($h_{\text{base}} = 1.65$) is displayed in Fig. 2. The localized ML is here again able to capture the curvature and is not perturbed by the heteroscedasticity (Fig. 6).

4.1.7. *Example 7: a multivariate quadratic model with heteroscedasticity*

This is a bivariate extension of the preceding example. The model is written as

$$Y = 5 + 10X_1 + 5X_2 - 5X_1X_2 - u + v,$$

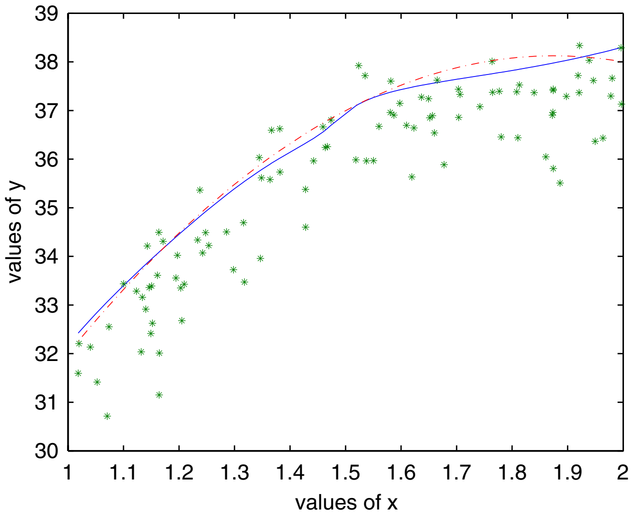


Fig. 5. Example 5, quadratic model: the true frontier (dashed-dotted) and the local maximum likelihood fit (solid).

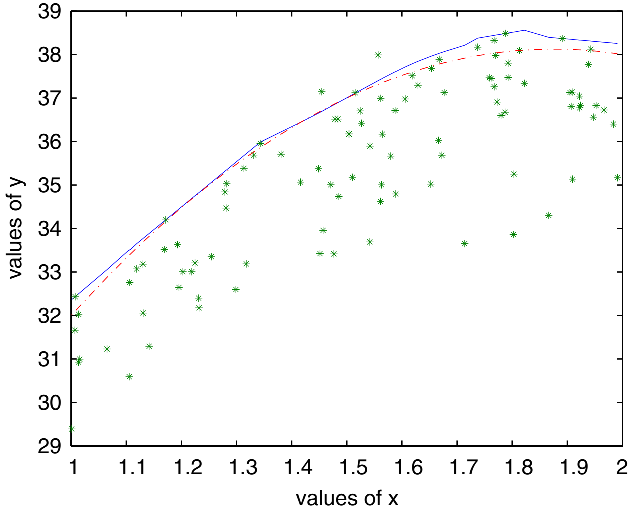
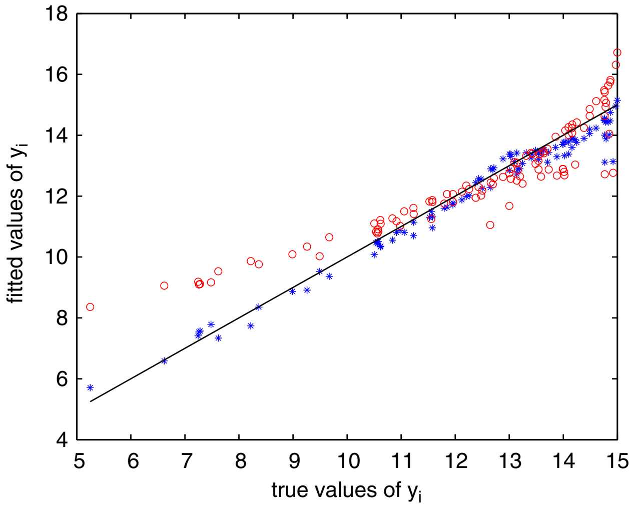


Fig. 6. Example 6, quadratic model with heteroscedasticity: the true frontier (dashed-dotted) and the local maximum likelihood fit (solid).

where $X_1 \sim \mathrm{U}(0,1)$, $X_2 \sim \mathrm{U}(0,2)$, $u|X = x \sim |\mathcal{N}(0, \sigma_u^2 \times (1 + x_1 + x_2/2)^2)|$ and $v \sim \mathcal{N}(0, \sigma_v^2)$, where $\sigma_u = 1$ and $\sigma_v = 0.5 \times \sigma_u \sqrt{(\pi - 2)/\pi}$. We notice that here the conditional mean and standard deviation of u can vary by a factor of 3 over the range of X. Here, we give the results for $n = 100$.

The optimal bandwidth obtained by cross-validation is around $h_{\text{base}} = 3.15$. To give an idea of the quality of the fit in this model, the top panel of Fig. 7 displays the plot of the true frontier value against the fitted values. The bottom panel illustrates the quality of the



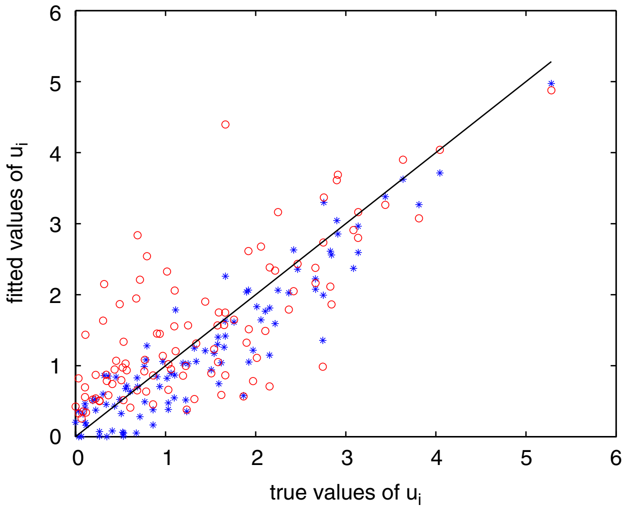


Fig. 7. Example 7, a bivariate model with heteroscedasticity: true against estimated frontier (top panel) and true against predicted efficiencies (bottom panel), for $i = 1, \ldots, 100$. The asterisks "*" are for the local MLE and the circles "o" are for the MLE of the parametric anchorage model.

prediction of the efficiency terms by showing the plot of the true simulated u_i against its predictor $\widehat{u}_i$, as provided by (3.4). Surprisingly, we can see that the difficult task of predicting one random variable u_i, conditionally on one observation of a noisy residual $\widehat{\varepsilon}(X_i)$ (see (3.4)) works pretty well globally. Of course, some individual prediction might be far from the truth.

We can also compare in the two panels of Fig. 7 how we improve the fit over the usual parametric MLE (with the anchorage model). The top panel clearly indicates that the homoscedastic linear model is evidently unable to detect the curvature of the frontier and the bottom panel shows the dramatic consequence when estimating the efficiency of each unit.

This example shows that our procedure is not only robust to reasonable departures from the anchorage model, but it also indicates that results obtained by the local likelihood method should always be preferred or at least compared to those obtained by traditional MLE methods with the anchorage models the practitioner has in mind. This is particularly important when working with real data.

4.2. A real data example

As a final illustration, we worked out an example with a real data set. The analysis is based on a cross-sectional random sample of 500 US commercial banks for the year 2000 (this data set has been analyzed by Kumbhakar and Tsionas, 2005). Following Kumbhakar and Tsionas we assume the outputs of these banks are: installment loans (to individuals for personal/household expenses) (q_1), real estate loans (q_2), business loans (q_3), federal funds sold and securities purchased under agreements to resell (q_4), miscellaneous assets (other assets that cannot be properly included in any other asset items in the balance sheet) (q_5). The inputs are: labor (x_1), capital (x_2), purchased funds (x_3), interest-bearing deposits in total transaction accounts (x_4) and interest-bearing deposits in total nontransaction accounts (x_5). We also have information on the total expenses of each input, so we can derive the input price ($w_1 - w_5$) by dividing total expenses on it by the corresponding input quantity. Total cost is then defined as the sum of expenses on these five inputs. To impose the linear homogeneity (on input prices) restrictions, we normalize total cost and the input prices with respect to w_5.

In banking studies it is standard to use a cost function approach, so we analyze a cost function and the anchorage parametric model is a Cobb–Douglas:

$$\begin{aligned} TC = {} & \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \beta_3 W_3 + \beta_4 W_4 + \beta_5 Q_1 + \beta_6 Q_2 + \beta_7 Q_3 \\ & + \beta_8 Q_4 + \beta_9 Q_5 + u + v, \end{aligned}$$

where TC is the log of total cost (normalized by price of fifth input), W_1 to W_4 are the logs of input prices (relative to W_5) and Q_1 to Q_5 are the outputs in log. The stochastic specification is as in (3.1).

The results from the parametric maximum likelihood method are given in Table 1. The covariance matrix of the estimators are computed from the inverse of the computed Hessian at the optimal values.

The estimated parameters are reported in Table 1, where all the coefficients appear to have the right sign, all the coefficient being highly significant, under the chosen parametric specification. We observe a value for $\hat{\lambda} = 1.88$ but with a large standard deviation. The localized version of these estimators requires the selection of the bandwidth. The CV criterion was selected at $m = 100$ randomly chosen points with the leave-one-out formula given in (3.5). We started with a crude grid of values for h and then we approach the optimal value of h_{base} by a final search over a finer grid. The final result is shown in Fig. 8: the optimal value is found to be around $h_{\text{base}} = 5.5$.

Fig. 9, illustrates the variation of the localized estimates of the cost function $r_0(x)$ and of the variance functions $\sigma_0^2(x)$ and $\lambda_0(x)$, evaluated at the 500 data points by plotting their density estimates.[8] The density of $\widehat{\sigma}_0^2(x)$ (negative values for $\widehat{\sigma}_0^2(x)$ are only due to the

[8]The computing time for all these estimations was 640 s on a Pentium M, Centrino 2 GHz with a cache-memory of 2 Mb.

Table 1
Parametric MLE for the banking data

Parameter	Estimates	Std error
Const	0.7072	0.3049
p_1	0.2635	0.0452
p_2	0.0484	0.0163
p_3	0.1561	0.0266
p_4	0.0656	0.0256
q_1	0.0654	0.0142
q_2	0.4440	0.0157
q_3	0.1847	0.0213
q_4	0.0594	0.0078
q_5	0.1911	0.0218
σ^2	0.0974	0.0118
λ	1.8806	0.3610

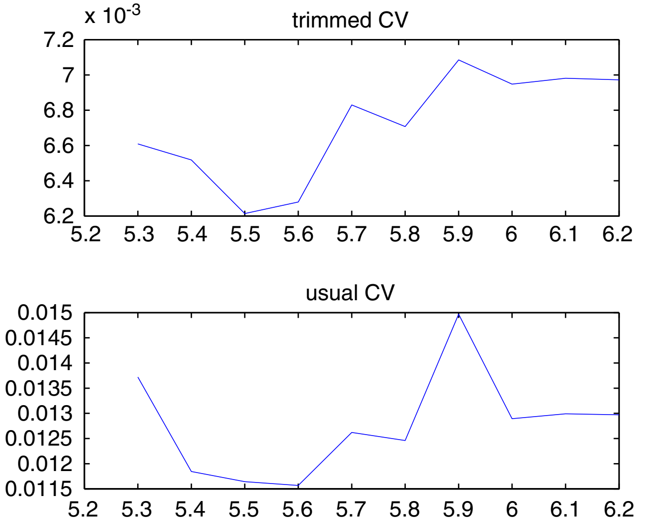


Fig. 8. Bank data set: CV criterion.

smoothing of the histogram) and particularly the density of $\widehat{\lambda}_0(x)$ clearly indicate the presence of heteroscedasticity with an important heterogeneity in the shares of efficiency to noise (from 1 to 4).

The different histograms of the values $\hat{r}_{1j}(x_i)$, $j = 1, \ldots, 9$ and $i = 1, \ldots, 500$ are also illuminating. They are given in Fig. 10. It indicates that these observation-specific coefficients vary across different banks, showing that the parametric CD model (assuming constant coefficients) might be wrong. Indeed, if $r_{1j}(\cdot)$ would be constant, then $\hat{r}_{1j}(x_i)$ would be very near the parametric MLE $\widehat{\beta}_j$ for all i. Thus, in that case, most of the $\hat{r}_{1j}(x_i)$ values would fall in the range of length $4\sigma(\widehat{\beta}_j)$. If we compare the values of $4\sigma(\widehat{\beta}_j)$ obtained from Table 1 with the histograms in Fig. 10, we see that this is not the case for any $\hat{r}_{1j}$. The

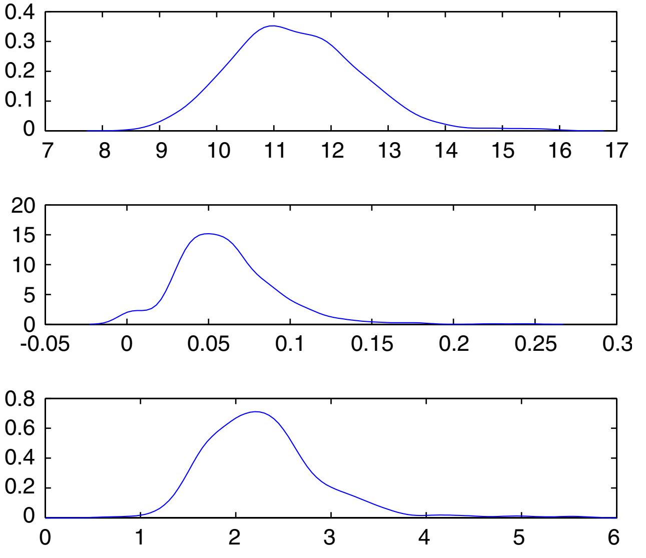


Fig. 9. Distribution of estimates over the 500 observed points. From top to bottom: cost function $r_0(x)$, and variance functions $\sigma_0^2(x)$, $\lambda(x)$.

comparison of the empirical standard deviations of $\hat{r}_{1j}(x_i)$ with $\sigma(\widehat{\beta}_j)$, for $j = 1, \ldots, 9$, which is reported in Table 2, also confirms this.

Note also that, as expected, most of the output coefficient are positive ($\hat{r}_{15}$ to $\hat{r}_{19}$). For the interpretation of $\hat{r}_{11}$ to $\hat{r}_{14}$, it has to be noted that the prices are relative to the price of the fifth input.

In Table 3 we report the detailed results for 25 banks randomly drawn from the full data set. The last column reports the efficiency scores. Since we are estimating a cost function, and not a production function, the value of the fitted cost here is computed at the value at the frontier $\widehat{TC}_i$ plus the estimated inefficiency term $\hat{u}_i$, for $i = 1, \ldots, 500$.[9] The efficiency scores are given by $\exp(\hat{u}_i) \geqslant 1$. The histograms of these 500 values are provided in Fig. 11. It seems that the marginal distribution of u is more truncated normal than half normal. Note also that in this example the average efficiency score over the 500 banks is 1.1723. Thus, on average, cost for these bank is increased by 17% due to inefficiency.

To compare our results with the parametric likelihood method with the anchorage model, we can look at Figs. 12 and 13. The left panels of these two figures are for our local likelihood approach and the right panels for the usual parametric MLE. We do not see substantial differences between the two approaches for the quality of the fit (although the localized approach seems to provide a more concentrated cloud of points near the straight line of the equality between TC_i and $\widehat{TC}_i + \hat{u}_i$), but the difference is much more important in the right panels for the efficiency scores. It seems that the standard parametric approach predicts much higher inefficiencies than the more flexible localized version. This overestimation of the inefficiency might be due to error in the specification of the model or in the implicit homoscedastic nature of the standard MLE.

[9]More specifically, $100(\exp(\hat{u}_i) - 1)$ is the percent by which cost is increased due to inefficiency.

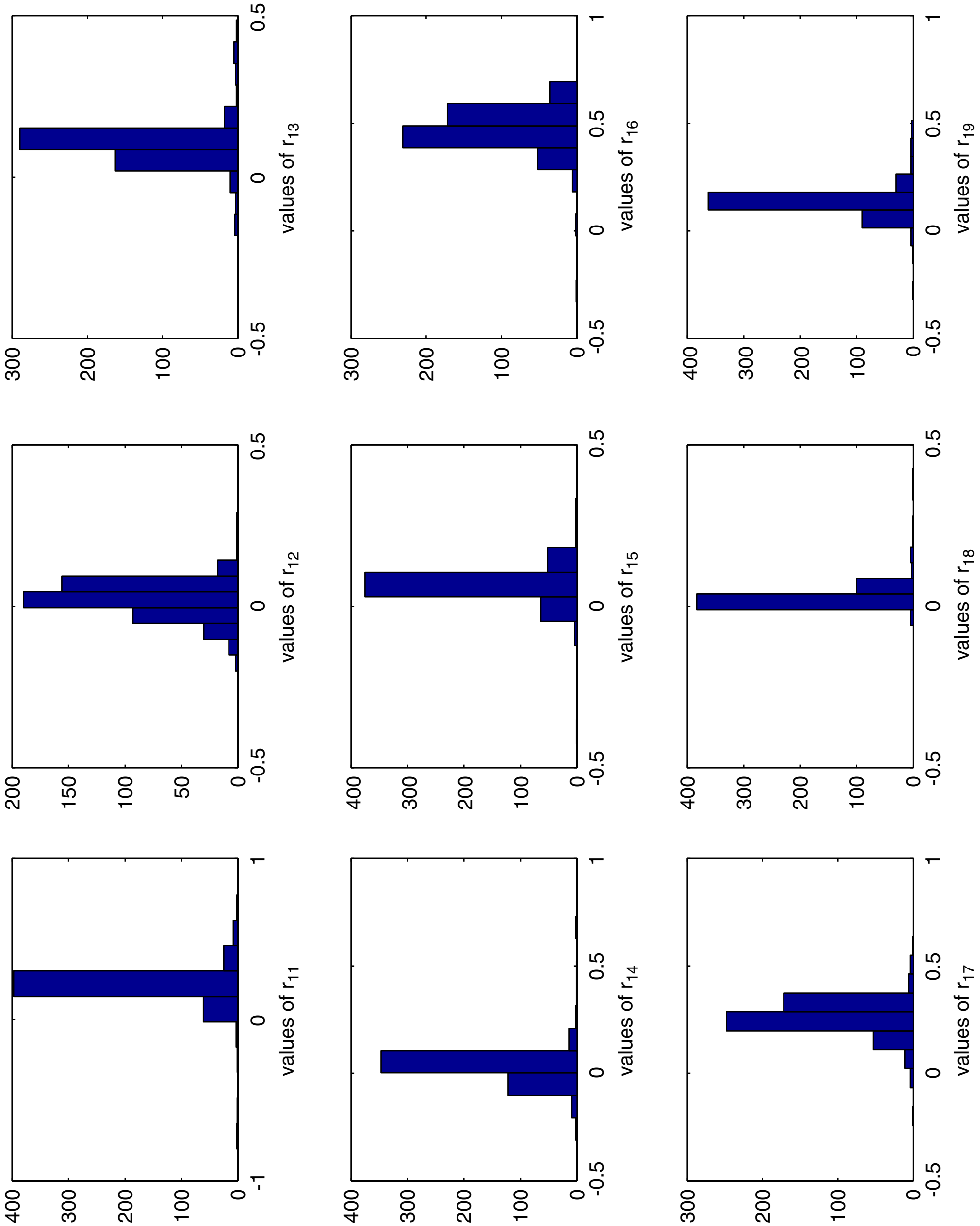


Fig. 10. Distribution of localized estimates of elasticities over the 500 observed points.

Table 2
Empirical standard deviation of $\hat{r}_{1j}(x_i)$, compared with the parametric standard deviation of the MLE of β_j

Variable	$\mathrm{Std}(\hat{r}_{1j}(x_i))$	$\hat{\sigma}(\hat{\beta}_j)$
p_1	0.1106	0.0452
p_2	0.0525	0.0163
p_3	0.0594	0.0266
p_4	0.0693	0.0256
q_1	0.0449	0.0142
q_2	0.0901	0.0157
q_3	0.0712	0.0213
q_4	0.0349	0.0078
q_5	0.0579	0.0218

5. Conclusions

In this paper we have provided a new approach for estimating SF functions nonparametrically. The approach is based on local maximum likelihood techniques. The model can be presented as encompassing some anchorage parametric model in a nonparametric way. The estimation here is obtained by localizing the likelihood function to be maximized. The asymptotic properties of our estimator are established for the general setup of local linear approximations.

We tailored the results to a stochastic production frontier model where the convoluted/composed error term (inefficiency plus noise) is the sum of a half normal and a normal random variable. The parametric anchorage model is a linear production function and a homoscedastic error term.

The performance of the estimator is first illustrated with some simulated data sets. Then we apply it to a real data set (US commercial banks) to test the flexibility of our method. We find enough variability in the estimated coefficient to suggest that the underlying production technology is heterogeneous so far as the output elasticities and the estimates of the variances of the composed errors are concerned. Based on this result, we conclude that estimated technology from a parametric model and elasticities, efficiencies, etc., derived therefrom might be wrong, especially if the technology is heterogeneous.

All these numerical illustrations indicate that the methods are robust, numerically stable and particularly useful for investigating a production process and the derived efficiency scores.

The analysis is proposed in a cross-sectional framework, it is a challenge for future research to extend the approach to panel models. This extension would be complementary to Park et al. (1998, 2003).

6. Regularity conditions and proof of Theorem 2.1

6.1. Regularity conditions

Define $z(u) = (1, u, \ldots, u^m)^{\mathrm{T}}$, $\psi(t|x) = \mathrm{E}\{q_1(Y_1, \theta(x) + t)|X_1 = x\}$ and

$$Q(a) = p(x) \int z(u) \otimes \psi(a_0 + a_1 u + \cdots + a_m u^m | x) K(u)\,\mathrm{d}u,$$

Table 3
Table of some fitted values

Units	W_1	W_2	W_3	W_4	Q_1	Q_2	Q_3	Q_4	Q_5	Cost	Fitted cost	$\hat{u}_i$	Efficiencies
1	6.5966	1.7829	−0.1313	−0.4929	9.2551	10.1758	9.3858	7.3460	7.0630	11.0795	11.1971	0.0829	1.0864
2	6.9129	1.3529	−1.0427	−0.8361	9.4112	11.1363	10.5474	9.6934	7.7790	12.1166	12.1116	0.1320	1.1411
3	6.9674	2.5524	−0.3873	−0.8174	7.9186	8.2782	8.8932	5.9915	6.6425	9.9978	10.0653	0.1060	1.1118
4	6.7498	1.0411	−0.2340	−0.3768	9.0120	9.8810	10.1022	8.4160	7.3914	11.2389	11.2868	0.0745	1.0774
5	6.7008	2.8757	−0.4098	−0.9645	10.3653	11.1503	10.9163	9.9988	7.9040	12.5539	12.5015	0.2087	1.2321
6	6.6795	2.6082	−0.5432	−0.7307	6.9266	8.9525	9.1494	8.3526	6.4831	10.3847	10.4012	0.1368	1.1466
7	7.2480	2.0128	0.8435	1.0546	9.0860	11.4115	10.7046	6.3509	8.5599	12.6403	12.6403	0.0068	1.0068
8	6.6397	2.3575	−0.1957	−0.4287	9.4590	10.5429	9.9740	9.5541	8.0074	12.3938	12.1838	0.5148	1.6732
9	6.2022	2.1152	−0.3518	−0.6966	6.5323	7.6158	7.6212	5.4337	5.4205	9.3176	9.2644	0.2283	1.2564
10	6.5531	1.4028	−0.2706	−1.6217	9.1369	10.6247	10.8835	6.9575	7.4164	11.6774	11.7040	0.0621	1.0640
11	6.6039	1.3790	−0.4982	−0.9325	9.5613	10.3416	9.8936	7.7075	6.9068	11.4081	11.3997	0.1115	1.1179
12	6.8066	2.6768	−0.3423	−0.7454	7.0876	8.4964	8.3916	5.8861	5.6525	9.5742	9.7066	0.0592	1.0610
13	6.8697	1.6097	−0.3267	−0.9636	8.6380	12.1599	11.3077	9.9665	8.0548	12.9121	12.8970	0.1352	1.1448
14	6.2881	1.4901	−0.1081	−0.4522	8.6980	10.9591	9.3496	8.4919	7.2944	11.6220	11.6063	0.1499	1.1617
15	6.9144	1.6299	0.4353	−0.6377	8.2597	9.1669	10.4917	8.3894	7.8296	11.4570	11.3914	0.2826	1.3265
16	6.3287	2.2669	−0.0698	−0.6538	6.5751	8.3807	7.8192	7.5256	5.6904	9.8814	9.8412	0.2138	1.2383
17	7.0549	2.0778	−0.4106	0.2150	9.1180	13.0627	11.3933	9.3102	9.1156	13.6900	13.6815	0.1209	1.1285
18	6.7379	1.7127	−0.2584	−0.6250	7.8629	9.3493	10.2826	7.7363	7.4006	11.0253	11.0510	0.1111	1.1175
19	6.5680	1.2609	0.2738	−0.7842	9.4957	11.1189	9.9293	7.5137	7.5104	12.1546	12.0791	0.2865	1.3317
20	6.8310	1.1302	−0.0275	−1.2897	8.4011	10.6982	9.4122	8.8793	8.3102	11.4811	11.6546	0.0887	1.0928
21	6.4870	0.4366	−0.5490	−1.0479	8.2620	10.3070	9.7662	7.4384	6.9027	10.9640	11.0269	0.0356	1.0363
22	6.4722	1.4614	−0.0992	−0.6478	8.7883	10.2242	9.4014	8.1970	7.4366	11.2746	11.2885	0.1423	1.1529
23	7.0401	1.9240	0.1283	−1.0937	8.9091	8.3008	9.2534	8.3309	6.2086	10.7290	10.6633	0.3473	1.4152
24	6.8392	2.0559	0.4172	−0.5878	9.6402	10.9646	10.0582	0.0000	7.4176	11.8415	11.8415	0.0007	1.0007
25	6.9965	2.0503	−0.4884	−0.3147	6.9985	8.0401	7.7706	6.9078	4.7095	9.3526	9.3540	0.0874	1.0914

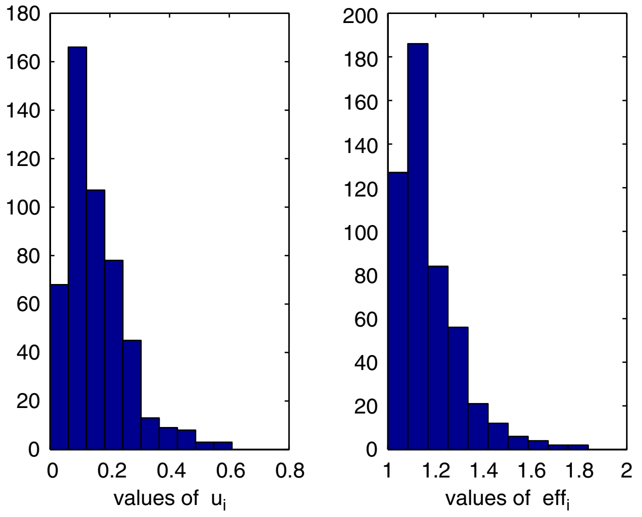


Fig. 11. Distribution of efficiencies over the 500 observed points. Left panel $\hat{U}_i$, right panel $\widehat{eff}_i$.

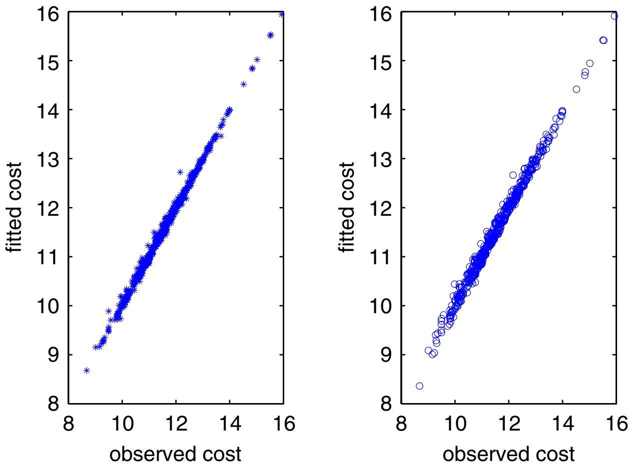


Fig. 12. Quality of the fit over the 500 observed points. Left panel local likelihood method, right panel parametric likelihood.

where $\otimes$ is the Kronecker product between vectors and matrices and $a = (a_0^{\mathrm{T}}, \ldots, a_m^{\mathrm{T}})^{\mathrm{T}}$. Note that each a_i is a k-dimensional vector and thus a is $k(m+1)$-dimensional. The assumptions for Theorem 2.1 are as follows. Below, for a matrix $A = (a_{ij})$, $|A|$ denotes $\sqrt{\sum_i \sum_j a_{ij}^2}$.

(A1) $Q(a) = 0$ has the unique solution $a = 0$;
(A2) $\sup_{|t| \leqslant C} |\psi(t|x+z) - \psi(t|x)| \to 0$ as $z \to 0$ for $C > 0$;

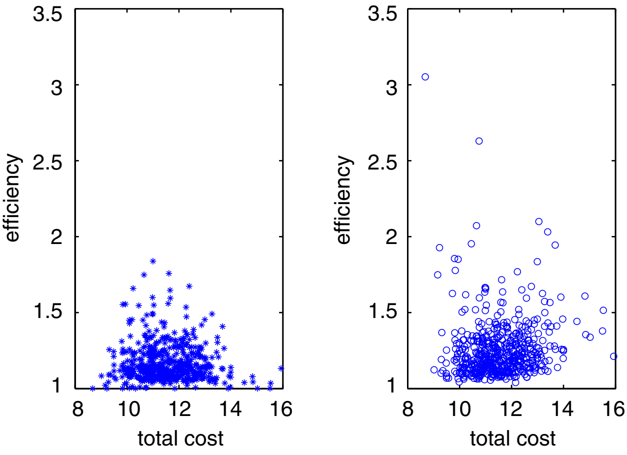


Fig. 13. Efficiencies versus total cost over the 500 observed points. Left panel local likelihood method, right panel parametric likelihood.

(A3) for any compact set $\mathscr{D}$ there exists a function U_1 such that $\sup_{|\theta| \in \mathscr{D}} |q_1(y, \theta)| \leqslant U_1(y)$ and $\sup_{|z-x| \leqslant \varepsilon} \mathrm{E}(U_1^{2+\delta}(Y)|X = z) < \infty$ for some $\varepsilon, \delta > 0$;
(A4) $q_2(y, \theta)$ is continuous in θ for each y, and for any compact set $\mathscr{D}$ there exists a function U_2 such that $\sup_{|\theta| \in \mathscr{D}} |q_2(y, \theta)| \leqslant U_2(y)$ and $\sup_{|z-x| \leqslant \varepsilon} \mathrm{E}(U_2^2(Y)|X = z) < \infty$ for some $\varepsilon > 0$;
(A5) $p(x) > 0$, and the matrices $\rho(x)$ and $v(x)$ are positive definite;
(A6) p, all entries of ρ and v are continuous at x;
(A7) K is nonnegative, bounded and supported on $[-1, 1]$;
(A8) All entries of θ have $(m+1)$th continuous derivatives at x;
(A9) $h \to 0$ and $nh \to \infty$ as $n \to \infty$, and $nh^{2m+3} < C$ for some positive constant C.

For Theorem 2.2 we need (A1)–(A7) plus the following assumptions:

(A8′) All entries of θ are twice partially continuously differentiable at x;
(A9′) All entries of the bandwidth matrix H tend to zero with $n|H|$ increasing to infinity as n goes to infinity, and satisfies $n|H|^5 < C$ for some positive constant C.

We remark that the compact support assumption on the kernel K is made for concise presentation of the proof. With more deliberate arguments the theorems can be shown to remain valid for kernels with exponentially decaying tails such as the Gaussian which we use in our numerical study in Section 4.

6.2. Proof of Theorem 2.1

Before we get into the proof, we point out that the standard technique based on a uniform quadratic approximation of the target function $L_n(\theta)$ (see Fan et al., 1995) cannot be used here since it relies heavily on the convexity of L_n. In our setting, we do not assume convexity of L_n, but only ask uniqueness of the solution of the equation $Q(a) = 0$, which is weaker than the convexity assumption. Our approach is to first establish the consistency of the estimator and then get a stochastic expansion to prove the theorem.

We introduce additional notations:

$$\widetilde{\theta}(u) = \theta(x) + \theta'(x)(u - x) + \cdots + \theta^{(m)}(x)(u - x)^m/m!,$$

$$\widehat{a}_j \equiv \widehat{a}_j(x) = h^j\left(\widehat{\theta}_j(x) - \frac{\theta^{(j)}(x)}{j!}\right), \quad j = 0, \ldots, m,$$

where

$$\theta^{(j)}(x) = \begin{pmatrix} \theta_1^{(j)}(x) \\ \vdots \\ \theta_k^{(j)}(x) \end{pmatrix} \quad \text{and} \quad \theta_i^{(j)}(x) = \frac{\partial^j}{\partial x^j}\theta_i(x).$$

Also, let

$$Z_i = \left(1, \left(\frac{X_i - x}{h}\right), \ldots, \left(\frac{X_i - x}{h}\right)^m\right)^{\mathrm{T}}.$$

Then, $\widehat{a} \equiv (\widehat{a}_0^{\mathrm{T}}, \ldots, \widehat{a}_m^{\mathrm{T}})^{\mathrm{T}}$ is the solution of the equation $Q_n(a) = 0$, where

$$Q_n(a) = n^{-1}\sum_{i=1}^{n} Z_i \otimes q_1\left(Y_i, \widetilde{\theta}(X_i) + a_0 + a_1\left(\frac{X_i - x}{h}\right) + \cdots + a_m\left(\frac{X_i - x}{h}\right)^m\right)K_h(X_i - x).$$

At the end of this section we will prove for any compact set $\mathscr{A}$

$$\sup_{a\in\mathscr{A}} |Q_n(a) - \mathrm{E}Q_n(a)| = \mathrm{O_p}\left(\sqrt{(\log n)/(nh)}\right), \tag{6.1}$$

$$\sup_{a\in\mathscr{A}} |\mathrm{E}Q_n(a) - Q(a)| = \mathrm{o}(1). \tag{6.2}$$

Also, we will show that for any $\varepsilon > 0$

$$\sup_{|a|\leqslant\varepsilon} |S_n(a) - \mathrm{E}S_n(a)| = \mathrm{O_p}\left(\sqrt{(\log n)/(nh)}\right), \tag{6.3}$$

where

$$S_n(a) = n^{-1}\sum_{i=1}^{n}(Z_i Z_i^{\mathrm{T}}) \otimes q_2\left(Y_i, \widetilde{\theta}(X_i) + a_0 + a_1\left(\frac{X_i - x}{h}\right) + \cdots + a_m\left(\frac{X_i - x}{h}\right)^m\right)K_h(X_i - x).$$

The foregoing results (6.1) and (6.2) imply $\sup_{a\in\mathscr{A}}|Q_n(a) - Q(a)| = \mathrm{o_p}(1)$ for any compact set $\mathscr{A}$. By assumption (A1) this establishes the consistency result

$$\widehat{a} = \mathrm{o_p}(1). \tag{6.4}$$

Now, a Taylor expansion of $Q_n(a)$ gives

$$0 = Q_n(\widehat{a}) = Q_n(0) + S_n(a^*)\widehat{a}, \tag{6.5}$$

where a^* is a random vector such that $|a^*| \leqslant |\widehat{a}|$. The consistency of $\widehat{a}$ and result (6.3) enable us to approximate $S_n(a^*)$ by $\mathrm{E}S_n(0)$. To see this, we first observe that (6.3) and (6.4) imply

$$S_n(a^*) - [\mathrm{E}S_n(a)]_{a=a^*} = \mathrm{o_p}(1). \tag{6.6}$$

Also, by assumption (A4) it follows that for a given $\eta > 0$ there exists $\delta > 0$ such that, for sufficiently large n, $|\mathrm{E}S_n(a) - \mathrm{E}S_n(0)| \leqslant \eta$ for all a with $|a| \leqslant \delta$. This and the consistency (6.4) together with (6.6) establish

$$S_n(a^*) - \mathrm{E}S_n(0) = \mathrm{o_p}(1).$$

Furthermore, one can show that

$$\begin{aligned} \mathrm{E}S_n(0) &= -\mathrm{E}[(Z_1 Z_1^{\mathrm{T}}) \otimes \rho(X_1)]K_h(X_1 - x) + \mathrm{o}(1) \\ &= -[N \otimes \rho(x)]p(x). \end{aligned}$$

Here, the first equation follows from the dominated convergence theorem with assumption (A4) again and the second is a consequence of assumption (A6). Since $Q_n(0) = \mathrm{O_p}(n^{-1/2}h^{-1/2})$ as shown below, we obtain from (6.5) the following expansion for $\widehat{a}$:

$$\widehat{a} = p(x)^{-1}[N \otimes \rho(x)]^{-1} Q_n(0) + \mathrm{o_p}(n^{-1/2}h^{-1/2}). \tag{6.7}$$

We now prove $Q_n(0) = \mathrm{O_p}(n^{-1/2}h^{-1/2})$ and the asymptotic normality of $Q_n(0)$. For this, we approximate $\mathrm{E}Q_n(0)$ by

$$\begin{aligned} \mathrm{E}Q_n(0) &= \mathrm{E}[\{Z_1 \otimes q_1(Y_1, \widetilde{\theta}(X_1))\}K_h(X_1 - x)] \\ &= \mathrm{E}[\{Z_1 \otimes q_2(Y_1, \theta(X_1))\}\{\widetilde{\theta}(X_1) - \theta(X_1)\}K_h(X_1 - x)] + \mathrm{o}(h^{m+1}) \\ &= \mathrm{E}[\{Z_1 \otimes \rho(X_1)\}\{\theta(X_1) - \widetilde{\theta}(X_1)\}K_h(X_1 - x)] + \mathrm{o}(h^{m+1}) \\ &= \frac{1}{(m+1)!}\{\gamma \otimes \rho(x)\}\theta^{(m+1)}(x)p(x)h^{m+1} + \mathrm{o}(h^{m+1}). \end{aligned} \tag{6.8}$$

The second equality follows from the fact $Q(0) = 0$ and assumption (A4), and the last equality is obtained by using assumptions (A6) and (A8). Similarly, we approximate the variance of $Q_n(0)$ by

$$\begin{aligned} Var[Q_n(0)] &= n^{-1} Var[\{Z_1 \otimes q_1(Y_1, \widetilde{\theta}(X_1))\}K_h(X_1 - x)] \\ &= n^{-1}h^{-1}\mathrm{E}[\{(Z_1 Z_1^{\mathrm{T}}) \otimes (q_1(Y_1, \widetilde{\theta}(X_1))q_1^{\mathrm{T}}(Y_1, \widetilde{\theta}(X_1)))\}(K^2)_h(X_1 - x)] \\ &\quad + \mathrm{O}(n^{-1}h^{2m+2}) \\ &= n^{-1}h^{-1}\mathrm{E}[\{(Z_1 Z_1^{\mathrm{T}}) \otimes v(X_1)\}(K^2)_h(X_1 - x)] + \mathrm{o}(n^{-1}h^{-1}) \\ &= n^{-1}h^{-1}\{S \otimes v(x)\}p(x) + \mathrm{o}(n^{-1}h^{-1}), \end{aligned} \tag{6.9}$$

where we use assumption (A3) for the third equality. Since $nh^{2m+3} = \mathrm{O}(1)$ by (A9), the above calculations establish that $Q_n(0) = \mathrm{O_p}(n^{-1/2}h^{-1/2})$.

Next, for the asymptotic normality of $Q_n(0)$ we note that

$$[Var(Q_n(0))]^{-1/2}(Q_n(0) - \mathrm{E}Q_n(0)) = \sum_{i=1}^{n} \zeta_{ni},$$

where ζ_{ni} for $1 \leqslant i \leqslant n$ are independent and identically distributed for each n, and $|\zeta_{ni}| \leqslant (\text{const.})\ (h/n)^{1/2} U_1(Y_i) K_h(X_i - x)$ for sufficiently large n. The latter property follows from assumption (A3). A direct application of the Lindeberg–Feller theorem with the condition $\sup_{|z-x| \leqslant \varepsilon} \mathrm{E}(U_1^{2+\delta}(Y)|X = z) < \infty$ for some $\varepsilon > 0$ as in (A3) establishes the asymptotic normality.

Let $D = [N \otimes \rho(x)]p(x)$. Expansion (6.7) and the above asymptotic normality of $Q_n(0)$ imply

$$[D^{-1} Var(Q_n(0)) D^{-1}]^{-1/2}(\widehat{a} - D^{-1}\mathrm{E}Q_n(0)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, I_{k(m+1)}).$$

Since $\widehat{a}_0 = \widehat{\theta}(x) - \theta(x) = (e_0^{\mathrm{T}} \otimes I_k)\widehat{a}$, we obtain

$$[(e_0^{\mathrm{T}} \otimes I_k) D^{-1} Var(Q_n(0)) D^{-1}(e_0 \otimes I_k)]^{-1/2}(\widehat{\theta}(x) - \theta(x) - (e_0^{\mathrm{T}} \otimes I_k) D^{-1}\mathrm{E}Q_n(0)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, I_k).$$

Using the properties of Kronecker product, viz., $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ and $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, it can be seen from (6.8) and (6.9) that

$$\begin{aligned}&(nh)(e_0^{\mathrm{T}} \otimes I_k) D^{-1} Var(Q_n(0)) D^{-1}(e_0 \otimes I_k)\\ &\quad = (e_0^{\mathrm{T}} N^{-1} S N^{-1} e_0)\rho(x)^{-1} v(x) \rho(x)^{-1}/p(x) + \mathrm{o}(1),\end{aligned}$$

$$(e_0^{\mathrm{T}} \otimes I_k) B^{-1}\mathrm{E}Q_n(0) = \frac{h^{m+1}}{(m+1)!}(e_0^{\mathrm{T}} N^{-1}\gamma)\theta^{(m+1)}(x) + \mathrm{o}(h^{m+1}).$$

Proof of (6.1). Write $Q_n(a) = n^{-1}\sum_{i=1}^n \xi_i(a)$. Define

$$\xi_i^*(a) = \xi_i(a) I(|\xi_i(a)| \leqslant \sqrt{n}) - \mathrm{E}[\xi_i(a) I(|\xi_i(a)| \leqslant \sqrt{n})].$$

By (A3) there exists a constant $c > 0$ such that

$$\sup_{a \in \mathcal{A}} \left| n^{-1} \sum_{i=1}^n \xi_i(a) I(|\xi_i(a)| > \sqrt{n}) \right| \leqslant c n^{-1} \sum_{i=1} U_1(Y_i) I(U_1(Y_i) > c^{-1}\sqrt{n}) K_h(X_i - x).$$

The expectation of the right-hand side of the above inequality is bounded by

$$c^2 n^{-1/2} \mathrm{E} U_1(Y_1)^2 K_h(X_1 - x) = \mathrm{O}(n^{-1/2}).$$

This shows

$$\sup_{a \in \mathcal{A}} \left| n^{-1} \sum_{i=1}^n \xi_i(a) I(|\xi_i(a)| > \sqrt{n}) \right| = \mathrm{O_p}(n^{-1/2}).$$

Similarly, we can prove

$$\sup_{a \in \mathcal{A}} |\mathrm{E}\xi_1(a) I(|\xi_1(a)| > \sqrt{n})| = \mathrm{O}(n^{-1/2}).$$

These reduce the proof of (6.1) to that of

$$\sup_{a \in \mathcal{A}} \left| n^{-1} \sum_{i=1}^n \xi_i^*(a) \right| = \mathrm{O_p}\left(\sqrt{(\log n)/(nh)}\right). \tag{6.10}$$

From (A4) one can show that for $a,\ b \in \mathscr{A}$

$$n^{-1}\sum_{i=1}^{n} |\xi_i^*(a) - \xi_i^*(b)| \leqslant R_n|a-b|,$$

where $R_n = \mathrm{O_p}(1)$ and R_n does not depend on a and b. Thus result (6.10) is proved if we show that there exists a positive constant α, which does not depend on a and M, such that for each $a \in \mathscr{A}$

$$\mathrm{P}\left[\left|n^{-1}\sum_{i=1}^{n} \xi_i^*(a)\right| > M\sqrt{(\log n)/(nh)}\right] \leqslant 2n^{\alpha - M}. \tag{6.11}$$

We prove (6.11) by a simple application of the Markov inequality as follows:

$$\begin{aligned}
&\mathrm{P}\left[n^{-1}\sum_{i=1}^{n}\xi_i^*(a) > M\sqrt{(\log n)/(nh)}\right] \\
&\quad = \mathrm{P}\left[\sqrt{(h\log n)/n}\sum_{i=1}^{n}\xi_i^*(a) > M\log n\right] \\
&\quad \leqslant n^{-M}\mathrm{E}\left[\exp\left(\sqrt{(h\log n)/n}\sum_{i=1}^{n}\xi_i^*(a)\right)\right] \\
&\quad \leqslant n^{-M}\prod_{i=1}^{n}\left[1 + n^{-1}(\log n)h\mathrm{E}\xi_i^*(a)^2\exp\left(2n^{1/2}\sqrt{(h\log n)/n}\right)\right] \\
&\quad \leqslant n^{-M}\exp\left[(\log n)h\mathrm{E}\xi_1^*(a)^2\exp\left(2\sqrt{h\log n}\right)\right] \\
&\quad \leqslant n^{\alpha - M}
\end{aligned}$$

for sufficiently large n, where α is a constant such that $h\mathrm{E}\xi_1^*(a)^2 \leqslant \alpha$. □

Proof of (6.2). By assumption (A2), one can see

$$\begin{aligned}
\mathrm{E}Q_n(a) &= \mathrm{E}\left[Z_1 \otimes \psi\left(\widetilde{\theta}(X_1) - \theta(X_1) + a_0 + a_1\left(\frac{X_1 - x}{h}\right)\right.\right. \\
&\qquad \left.\left. + \cdots + a_m\left(\frac{X_1 - x}{h}\right)^m \middle| X_1\right)K_h(X_1 - x)\right] \\
&= \mathrm{E}\left[Z_1 \otimes \psi\left(\widetilde{\theta}(X_1) - \theta(X_1) + a_0 + a_1\left(\frac{X_1 - x}{h}\right)\right.\right. \\
&\qquad \left.\left. + \cdots + a_m\left(\frac{X_1 - x}{h}\right)^m \middle| x\right)K_h(X_1 - x)\right] + \mathrm{o}(1)
\end{aligned}$$

uniformly for $a \in \mathscr{A}$. Since $\psi(t|x)$ is continuous in t by assumption (A4), the expectation on the right-hand side of the second equation equals $Q(a) + \mathrm{o}(1)$ uniformly for $a \in \mathscr{A}$. □

Proof of (6.3). The proof is similar to that of (6.1), hence it is omitted. □

References

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

Aragon, Y., Daouia, A., Thomas-Agnan, C., 2002. Nonparametric frontier estimation: a conditional quantile-based approach. Discussion paper, GREMAQ et LSP, Université de Toulouse ⟨http://www.univ-tlse1.fr/GREMAQ/Statistique/adt1202.pdf⟩.

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

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the inefficiency of decision making units. European Journal of Operational Research 2, 429–444.

Daouia, A., Simar, L., 2004. Nonparametric efficiency analysis: a multivariate conditional quantile approach. Journal of Econometrics, forthcoming.

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

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman & Hall, London.

Fan, J., Heckman, N.E., Wand, M.P., 1995. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. Journal of the American Statistical Association 90, 141–150.

Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business and Economic Statistics 14, 460–468.

Farrell, M.J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society A120, 253–281.

Gozalo, P., Linton, O., 2000. Local nonlinear least squares: using parametric information in nonparametric regression. Journal of Econometrics 99, 63–106.

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

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

Kumbhakar, S.C., Lovell, C.A.K., 2000. Stochastic Frontier Analysis. Cambridge University Press, New York.

Kumbhakar, S.C., Tsionas, E.G., 2002. Nonparametric stochastic frontier models. Manuscript presented at the NAPWII, June 2002, Schenectady, New York.

Kumbhakar, S.C., Tsionas, E.G., 2005. The joint measurement of technical and allocative inefficiencies: an application of bayesian inference in nonlinear random-effects models. Journal of the American Statistical Association 100, 736–747.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb–Douglas production function with composed error. International Economic Review 8, 435–444.

Park, B.U., Simar, L., 1994. Efficient semiparametric estimation in a stochastic frontier model. Journal of the American Statistical Association 89, 929–936.

Park, B.U., Sickles, R.C., Simar, L., 1998. Stochastic panel frontiers: a semiparametric approach. Journal of Econometrics 84, 273–301.

Park, B.U., Sickles, R.C., Simar, L., 2003. Semiparametric efficient estimation of AR(1) panel data models. Journal of Econometrics 117, 279–309.

Park, B.U., Sickles, R.C., Simar, L., 2006. Semiparametric efficient estimation in dynamic panel data models. Discussion 0315, Institut de Statistique, UCL, Belgium ⟨http://www.stat.ucl.ac.be⟩; Journal of Econometrics, to be published.

Simar, L., Wilson, P.W., 2000. Statistical inference in nonparametric frontier models: the state of the art. Journal of Productivity Analysis 13, 49–78.

Tibshirani, R., Hastie, T.J., 1987. Local likelihood estimation. Journal of the American Statistical Association 82, 559–567.