Semiparametric stochastic frontier estimation via profile likelihood [1]

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu
Voice: + 1 303 492 4599

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org
Voice: + 1 202 862 8144

and

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 2937867

December, 2011

Abstract. We consider the estimation of a nonparametric stochastic frontier model with composite error density which is known up to a finite parameter vector. Our primary interest is on the estimation of the parameter vector, as it provides the basis for estimation of firm specific (in)efficiency. Our frontier model is similar to that of Fan et al. (1996), but here we extend their work in that: a) we establish the asymptotic properties of their estimation procedure, and b) propose and establish the asymptotic properties of an alternative estimator based on the maximization of a conditional profile likelihood function. The estimator proposed in Fan et al. (1996) is asymptotically normally distributed but has bias which does not vanish as the sample size $n \to \infty$. In contrast, our proposed estimator is asymptotically normally distributed and correctly centered at the true value of the parameter vector. In addition, our estimator is shown to be efficient in a broad class of semiparametric estimators. Our estimation procedure provides a fast converging alternative to the recently proposed estimator in Kumbhakar et al. (2007). A Monte Carlo study is performed to shed light on the finite sample properties of these competing estimators.

Keywords and phrases. stochastic frontier models; nonparametric frontiers; profile likelihood estimation.

JEL Classifications. C14, C22

[1]We thank Daniel Henderson, Peter C. B. Phillips and participants in the XX New Zealand Econometrics Study Group and Midwest Econometrics Group Meetings for helpful comments. We also thank Essie Maasoumi, an Associate Editor and two referees for comments that improved the paper substantially. Any remaining errors are the authors' responsibility.

1 Introduction

There exists a vast literature on the estimation of production, cost and profit frontiers. Simar and Wilson (2008) provide a comprehensive review of the main statistical models and estimators that have been developed in the last four decades of research.

Recently a large number of articles (see, *inter alia*, Gijbels et al. (1999), Cazals et al. (2002), Aragon et al. (2005), Daouia and Simar (2007), Martins-Filho and Yao (2007, 2008) and Daouia et al. (2009)) have appeared in an attempt to improve and refine the estimation of deterministic frontiers models. Interestingly, theoretical developments and improvements for stochastic frontier models, pioneered by Aigner et al. (1977) and Meeusen and van den Broeck (1977), have not appeared with similar vigor. This is in spite of the great success these stochastic frontier models have had among empirical researchers. Greene (1993), Coelli (1995) and Kumbhakar and Lovell (2000) provide extensive reviews of applications and empirical uses of these models. Perhaps, one of the great disadvantages of stochastic frontier models has been their reliance on very tight parametric specifications for *both* the frontier to be estimated and the conditional density of the regressand of interest (output, cost or profit). Take, for example, the original stochastic production frontier model proposed by Aigner et al. They assume that an observed output-input pair $(y_i, x_i) \in \Re \times \Re^D$ for $i = 1, \cdots, n$ is the realization of independent and identically distributed process with density $f(y, x)$. The process is such that

$$y_i = g(x_i) + \varepsilon_i$$

where $g(x_i) = x_i'\beta$, $\varepsilon_i = v_i - u_i$ where u_i and v_i are independent unobserved random variables such that $v_i \sim N(0, \sigma_v^2)$, $u_i \sim |N(0, \sigma_u^2)|$ and x_i is independent of ε_i. Their formulation leads to a conditional density function of y_i given x_i, denoted by $f_{y|x}(y)$, which takes the structure

$$f_{y|x}(y; \theta, x'\beta) \equiv f_\varepsilon(y - x'\beta; \theta) = \frac{2}{\sqrt{\sigma_u^2 + \sigma_v^2}} \phi\left(\frac{y - x'\beta}{\sqrt{\sigma_u^2 + \sigma_v^2}}\right)\left(1 - \Phi\left(\frac{\sqrt{\sigma_u^2/\sigma_v^2}}{\sqrt{\sigma_u^2 + \sigma_v^2}}(y - x'\beta)\right)\right) \quad (1)$$

where $\theta = (\sigma_u^2, \sigma_v^2)' \in \Theta \subset (0, \infty) \times (0, \infty)$, $\phi(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}x^2\right)$, $\Phi(x) = \int_{-\infty}^{x} \phi(z)dz$, the frontier is given by $g(x) = x'\beta$ and $f_\varepsilon(\varepsilon; \theta)$ denotes that density of ε_i. Although there have been variations on this model, e.g. Greene (1990), the specification of stochastic frontiers has largely relied on a full parametric specification of the conditional density $f_{y|x}$, and as one would expect, estimation of the parameters has been conducted by maximizing the induced likelihood function.

Recently, a number of papers have attempted to provide much needed flexibility to the basic stochastic frontier model. Fan et al. (1996) retain much of the structure adopted by Aigner et al. However, they allow the frontier to belong to a much broader class of functions. Instead of the parametric $x'\beta$, their frontier is a smooth nonparametric function $g(x) : \Re^D \to \Re$. Since in their model $E(y|x) = g(x) - \sqrt{\frac{2}{\pi}\sigma_u^2}$ they propose a two step likelihood type estimator for θ (section 2 of this paper provides a complete description of the estimation procedure) based on a kernel (Nadaraya-Watson) estimator for $E(y|x)$. Although promising and intuitive in its construction, the authors did not investigate the asymptotic properties of their proposed estimator. A brief simulation provided in the paper provides what seems to be desirable experimental properties, but the results are in their totality rather incomplete.

Kumbhakar et al. (2007) take a different approach. Instead of the semiparametric model proposed by Fan et al. they consider a localized version of Aigner et al. where all "parameters" $(g(x), \sigma_u^2(x), \sigma_v^2(x))$ of the likelihood function depend on x. In this context, they adopt the local likelihood estimation approach pioneered by Staniswalis (1989) and also explored by Fan et al. (1995). Although their approach is quite general, allowing for example for conditional heterocedasticity, there are two undesirable features of this fully localized model that can potentially be avoided in a semiparametric specification. First, since all "parameters" are local, the rate of convergence of their proposed estimator is rather slow when the number of conditioning variables (inputs, in the case of a production function) is large. This is the well known curse of dimensionality that afflicts multidimensional kernel based nonparametric estimation. Since it is quite common in frontier models to have a large number of conditioning variables relative to the sample size, the accuracy of the asymptotic approximations can be rather poor. For example, in the empirical exercise conducted by Kumbhakar et al., a sample of 500 banks is used with 9 conditioning variables, calling into question the accuracy of the asymptotic approximation and the reliability of the resulting efficiency rankings. Second, since all parameters of the model depend on x, both expected population efficiency and firm (production plan) specific efficiencies, however calculated, depend on x. As a result, for any specific sample, uncountably many expected population efficiencies and efficiency firm rankings can be produced, a rather unsatisfying byproduct for the empirical user.

More recently, Kuosmanen and Kortelainen (2011) propose a two step estimator that combines a constrained (convex) nonparametric least squares procedure for the estimation of a nonparametric frontier

$g(x_i)$ with a method of moments or pseudo maximum likelihood estimator for θ and Simar and Zeleynuk (2011) consider a stochastic DEA/FDH estimator that expands on Simar (2007) and Kumbhakar et al. (2007). However, neither of these papers provide the asymptotic properties of their proposed estimators.

In this paper we contribute to the stochastic frontier literature in two ways. First, as in Fan et al. (1996), we consider a semiparametric model. We let the frontier be fully nonparametric ($g(x)$), but we consider conditional densities that allow for a parametric expression of the expected value of inefficiency. A special case of this structure is the model adopted in Fan et al. We study the estimation procedure proposed by Fan et al. in this *broader* class of models and show that their proposed estimator for the parameters of the model is consistent. Furthermore, we show that the two stage estimator they propose for θ is asymptotically normally distributed with parametric convergence rate $\sqrt{n}$. However, their proposed estimation procedure produces a bias that does not decay to zero when normalized by $\sqrt{n}$ under optimal smoothing for the estimator for $g(x)$. In addition, we show that the variance matrix associated with the asymptotic distribution is not the inverse of the Fisher information. These results, although new, are not unexpected in light of the work by Stein (1954), Severini (2000) and Severini and Wong (1992).

The second contribution we make in this paper, as it relates to the estimation of stochastic frontiers, is to define an alternative frontier estimation procedure that is inspired on the procedure described in Severini and Wong (1992). We show that our proposed estimator for θ is consistent and $\sqrt{n}$ asymptotically normal. Furthermore, contrary to the procedure proposed by Fan et al., our estimator carries no asymptotic bias and is efficient in a class of semiparametric estimators defined in Severini and Wong (1992).[1] Although our approach still relies on a partially parametric model, the efficient estimators we produce are free from the slow convergence rates described above and no parametric structure is imposed on the frontier. From a more technical perspective, the main result in this paper is the fact that under fairly mild conditions local linear regression estimation can be used to estimate least favorable curves in conditionally parametric models. This extension of Severini and Wong (1992) is not obvious and is embedded in the proof of our Lemma 2.

Besides this introduction, the paper has five more sections. Section 2 provides a description of our model and the estimators we study. Section 3 gives the derived asymptotic properties of the estimators

[1] See also van der Vaart (1999).

under study and lists a collection of assumptions that are sufficient for the results. A discussion of these assumptions is also provided. In section 4, a small Monte Carlo study is provided to shed some light on the finite sample properties of the estimators. The study seems to confirm the results suggested by the asymptotic theory. Section 5 gives an empirical application using an extensively used data set provided by Greene (1990). Lastly, section 6 provides a summary and some concluding remarks. All proofs, tables and graphs are relegated to appendices.

2 A semiparametric stochastic frontier

Let $\left\{\begin{pmatrix} y_i \\ x_i \end{pmatrix}\right\}_{i=1,2,\cdots}$ be a collection of independent and identically distributed (i.i.d.) random vectors taking values in $\Re^{D+1}$ with $y_i \in \Re$ and $x_i \in \Re^D$. In the case of production frontiers we take y_i to be a measure of output and x_i to be an input vector, but other frontiers (profit, cost) can also be considered provided that y_i is taken to be a scalar.[2] We assume that the density function of $\begin{pmatrix} y_i \\ x_i \end{pmatrix}$ exists and is denoted by $f(y, x)$ when evaluated at $\begin{pmatrix} y \\ x \end{pmatrix}$. We denote the marginal density of x_i by $f_x(x)$ when evaluated at x. For all $x \in \Re^D$ such that $f_x(x) \neq 0$ we denote the conditional density function of y_i given x_i by $f_{y|x}(y)$. Throughout, we assume that $f_{y|x}(y)$ belongs to a family of densities which is known up to a parameter $\theta \in \Theta \subset \Re^P$, P a positive integer, and a function $g(x) : \Re^D \to \Re$ belonging to a class $\mathcal{G}$. The true values of θ and $g(x)$ will be denoted by θ_0 and $g_0(x)$, and our main objective is to estimate θ_0 and $g_0(x)$ based on a sample $\chi_n = \left\{\begin{pmatrix} y_i \\ x_i \end{pmatrix}\right\}_{i=1}^{n}$. We follow Severini and Wong (1992) and assume that $f(y, x) = f_{y|x}(y; \theta, g(x))f_x(x)$, i.e., the parameter θ and the function $g(x)$ enter f only through $f_{y|x}$. This type of semiparametric structure has been called conditionally parametric models, since conditional on a particular value x the conditional density is parametrized by a finite number of parameters. In addition, as a direct link to the stochastic frontier framework, we restrict the class of conditional densities to those that satisfy $E(y|x) \equiv m(x; \theta, g) = g(x) - \gamma(\theta)$ where $\gamma(\theta) : \Theta \to \Re$ and $V(y|x) = v(\theta)$ where $v(\theta) : \Theta \to (0, \infty)$. Again, for the case of production frontiers, $g(x)$ is interpreted as the systematic portion of the frontier and $\gamma(\theta_0)$ denotes the expected reduction in expected output due to inefficiencies. It is easy to verify that the conditional density associated with the stochastic frontier model considered

[2] It should be noted that the case of multiple outputs can be accommodated in our framework by adopting the polar coordinate representation of $\begin{pmatrix} y_i \\ x_i \end{pmatrix}$ given in Simar and Zelenyuk (2011). See their equations (2.5) and (2.6).

by Aigner et al. (1977) and Fan et al. (1996) is a special case of this structure. Since in their case $f_{y|x}$ is given by equation (1), then $E(y|x) = g(x) - \gamma(\theta)$ where $\gamma(\theta) = \sqrt{\frac{2}{\pi}\sigma_u^2}$ and $V(y|x) = v(\theta) = \frac{\pi-2}{\pi}\sigma_u^2 + \sigma_v^2$.

Given this stochastic structure we consider maximum likelihood (ML) type estimators of θ_0 and $g_0(x)$ based on the following log-likelihood function

$$\bar{l}_n(\theta, g) = \frac{1}{n}\sum_{i=1}^{n} log f_{y|x}(y_i; \theta, g(x_i)) = \frac{1}{n}\sum_{i=1}^{n} log f_{\varepsilon}(y_i - m(x_i; \theta, g) - \gamma(\theta); \theta). \quad (2)$$

We investigate two alternative ML procedures. The first, motivated by Fan et al. (1996), is based on the fact that if g_0 were known, a parametric ML estimator for θ_0 could be obtained in a routine manner by maximizing $\bar{l}_n(\theta, g_0) = \frac{1}{n}\sum_{i=1}^{n} log f_{\varepsilon}(y_i - m(x_i; \theta, g_0) - \gamma(\theta); \theta)$ over the set Θ. Since g_0 is unknown, $\bar{l}_n(\theta, g_0)$ can be approximated by $\bar{l}_n(\theta, \hat{m} + \gamma(\theta)) = \frac{1}{n}\sum_{i=1}^{n} log f_{\varepsilon}(y_i - \hat{m}(x_i) - \gamma(\theta); \theta)$ where $\hat{m}(x_i)$ is an estimator for $m(x_i; \theta, g_0)$. Then, we define the estimator

$$\hat{\theta} \equiv \underset{\theta}{\operatorname{argmax}}\ \bar{l}_n(\theta, \hat{m} + \gamma(\theta)). \quad (3)$$

The second, motivated by Severini and Wong (1992), involves "joint" estimation of θ_0 and $g_0(x)$. To this end define,

$$\bar{\ell}_n(\theta, g_x) = \frac{1}{n}\sum_{i=1}^{n} log f_{\varepsilon}(y_i - g_x(x_i); \theta)\frac{1}{h_n}K\left(\frac{x_i - x}{h_n}\right) \quad (4)$$

where $g_x(x_i) = \alpha(x) + \beta(x)(x_i - x)$, K is a kernel function and h_n is a bandwidth. The estimation procedure involves two-steps. First, for fixed x and θ define $\hat{\alpha}_\theta(x)$ and $\hat{\beta}_\theta(x)$ as

$$(\hat{\alpha}_\theta(x), \hat{\beta}_\theta(x)) \equiv \underset{\alpha(x),\beta(x)}{argmax}\ \bar{\ell}_n(\theta, g_x). \quad (5)$$

Second, we define

$$\tilde{\theta} \equiv \underset{\theta}{argmax}\ \bar{l}_n(\theta, \hat{\alpha}_\theta) = \underset{\theta}{argmax}\frac{1}{n}\sum_{i=1}^{n} log f_{\varepsilon}(y_i - \hat{\alpha}_\theta(x_i); \theta). \quad (6)$$

The estimator for $g_0(x)$ is then given by $\tilde{g}(x) \equiv \hat{\alpha}_{\tilde{\theta}}(x)$. In equations (4), (5) and (6) as well as in the rest of the paper we assume for simplicity that $D = 1$. The asymptotic results we obtain for $\hat{\theta}$ and $\tilde{\theta}$ are not impacted by D provided that the speed at which $h_n \to 0$ is suitably adjusted.

From a computational perspective, the estimators are fairly easy to implement. In our Monte Carlo study we provide a discussion of bandwidth selection and the description of an algorithm for obtaining $\tilde{\theta}$ using $D > 1$. In the next section we study some of the asymptotic properties for $\hat{\theta}$ and $\tilde{\theta}$.

3 Asymptotic theory

3.1 The estimator $\hat{\theta}$

We start by listing assumptions that will be used throughout the paper.

ASSUMPTION A1. 1. $\theta_0 \in int(\Theta)$, where $int(\Theta)$ denotes the interior of the compact set $\Theta \subset \Re^P$; 2. The class $\mathcal{G}$, to which g_0 belongs, is given by $\mathcal{G} = \{g(x) : G \to \mathcal{H}\}$ where G a compact subset of $\Re^D$, $\mathcal{H}$ a compact subset of $\Re$, $g(x) \in int(\mathcal{H})$ for all x in G and $g(x)$ is twice continuously differentiable; 3. $E(y|x) \equiv m(x; \theta_0, g_0) = g_0(x) - \gamma(\theta_0)$ where $\gamma(\theta) : \Theta \to \Re$ which is twice continuously differentiable in Θ; 4. $V(y|x) = v(\theta_0)$ where $v(\theta) : \Theta \to (0, \infty)$.

An important consequence of assumption A1.3 is that, for any $\theta \in \Theta$, $x \in G$ and $g \in \mathcal{G}$ we have that $g(x) - g_0(x) = m(x; \theta, g) - m(x; \theta, g_0)$. As such, although $m(x; \theta, g)$ depends on θ, the difference $|m(x; \theta, g) - m(x; \theta, g_0)|$ does not.

ASSUMPTION A2. 1. $f_x(x) \in [B_L, B_U]$, $B_L, B_U \in (0, \infty)$ for all $x \in G$; 2. For all $x, s \in G$ we have that $|f_x(x) - f_x(s)| \leq C||x - s||_E$ for some $C \in (0, \infty)$, where $||x||_E$ denotes the Euclidean norm of x.

Since we will be considering kernel based nonparametric estimators, we will make the following standard assumption on the kernel function K. As in assumption A2, throughout the paper C will represent an arbitrary positive real number.

ASSUMPTION A3. 1. $K(\phi_1, \cdots, \phi_D) : S_D \subset \Re^D \to \Re$ is a symmetric density function with S_D a compact set; 2. $\int \phi_i K(\phi_1, \cdots, \phi_D) d(\phi_1, \cdots, \phi_D) = 0$, $\int \phi_i \phi_j K(\phi_1, \cdots, \phi_D) d(\phi_1, \cdots, \phi_D) = \sigma_K^2 > 0$ if $i = j$, otherwise $\sigma_K^2 = 0$ for all i and j; 3. For all $x \in S_D$ we have $K(x) \leq C$; 4. For all $x, s \in S_D$ we have $|K(x) - K(s)| \leq C||x - s||_E$ for some C.

Assumption A3 is satisfied by many commonly used kernels, including Epanechnikov and biweight.

ASSUMPTION A4. 1. For all $\theta \in \Theta$ we have that if $\theta \neq \theta_0$ then $f_\varepsilon(y - g_0(x); \theta) \neq f_\varepsilon(y - g_0(x); \theta_0)$ for all (y, x); 2. If $\{\theta_i\}_{i=1,2,\cdots}$ is a sequence in Θ such that $\theta_i \to \theta$ as $i \to \infty$, then $log f_\varepsilon(y - g_0(x); \theta_i) \to log f_\varepsilon(y - g_0(x); \theta)$ as $i \to \infty$ for all $\theta \in \Theta$; 3. $E\left(sup_{\theta \in \Theta} |log f_\varepsilon(y - g_0(x); \theta)|\right) < \infty$; 4. For all (y, x), $g \in \mathcal{G}$ and $\theta \in \Theta$, $|log f_\varepsilon(y - g(x); \theta) - log f_\varepsilon(y - g_0(x); \theta)| \leq b(y, x, \theta)|g(x) - g_0(x)| = b(y, x, \theta)|m(x; \theta, g) - m(x; \theta, g_0)|$ with $b(y, x, \theta) > 0$, and $E\left(sup_{\theta \in \Theta} b(y, x, \theta)\right) < \infty$.

Assumptions A4.1 and A4.3 guarantee that $E(log f_\varepsilon(y - g_0(x); \theta))$ has a unique maximum at θ_0.

Assumption A5. 1. For all $\eta = g(x) \in \mathcal{H}$, $logf_\varepsilon(y-\eta;\theta)$ is twice continuously differentiable with respect to θ and $f_\varepsilon(y-\eta;\theta) > 0$ on some open ball $S_{0,\theta} = S(\theta_0, d(\theta_0))$ of θ_0 with $S_{0,\theta} \subset \Theta$ and $d(\theta_0)$ the radius of the ball; 2. $E\left(sup_{\theta \in S_{0,\theta}} \left|\frac{\partial^2}{\partial\theta_j \partial\theta_k} logf_\varepsilon(y-g_0(x);\theta)\right|\right) < \infty$ for $k, j = 1, \cdots, P$; 3. We denote by $\frac{\partial}{\partial\eta}$ the partial derivative operator with respect to η, the argument of f_ε that immediately follows the negative sign, and assume $\frac{\partial}{\partial\eta} logf_\varepsilon(y-g_0(x);\theta)$ is continuously differentiable in $S_{0,\theta}$. Furthermore, $\int \sup_{\theta \in S_{0,\theta}} ||\frac{\partial^2}{\partial\theta\partial\eta} logf_\varepsilon(y-g_0(x);\theta)||_E f_\varepsilon(y-g_0(x);\theta_0)dy < \infty$ and

$$E\left(\sup_{\theta \in S_{0,\theta}} ||\frac{\partial^2}{\partial\theta\partial\eta} logf_\varepsilon(y-g_0(x);\theta)||_E |g_0^{(2)}(x)|\right) < \infty;$$

4. The matrix

$$\bar{H} = E\left(\frac{\partial^2}{\partial\theta\partial\theta'} logf_\varepsilon(y-g_0(x);\theta_0)\right) + \frac{\partial}{\partial\theta}\gamma(\theta_0) E\left(\frac{\partial^2}{\partial\theta\partial\eta} logf_\varepsilon(y-g_0(x);\theta_0)\right)' + E\left(\frac{\partial^2}{\partial\theta\partial\eta} logf_\varepsilon(y-g_0(x);\theta_0)\right)\frac{\partial}{\partial\theta}\gamma(\theta_0)'$$

exists and is nonsingular; 5. Let $\eta_0 = g_0(x)$ for any $x \in G$ and put $S_{0,\eta} = S(\eta_0, d(\eta_0))$. $logf_\varepsilon(y-\eta;\theta)$ is continuously differentiable on $S_{0,\eta}$, an open interval of $\mathcal{H}$, $E\left(\sup_{\eta \in S_{0,\eta}} \left|\frac{\partial}{\partial\eta} logf_\varepsilon(y-\eta;\theta)\right|\right) < \infty$ and for all $x \in G$, $\frac{\partial}{\partial\eta} E\left(logf_\varepsilon(y-g_0(x);\theta_0)|x\right) = 0$; 6. For all $\theta \in \Theta$, $\frac{\partial}{\partial\eta} logf_\varepsilon(y-g_0(x);\theta)$ is continuous at x,

$$E\left(\sup_{\theta\in\Theta}\left(\frac{\partial}{\partial\eta} logf_\varepsilon(y-\eta;\theta)\right)^2\right) < \infty \text{ and } E\left(\sup_{\theta\in\Theta}\left(\frac{\partial}{\partial\eta} logf_\varepsilon(y-\eta;\theta)\right)^2 y^2\right) < \infty.$$

If $\mathcal{G}$ is a normed linear space and $T(g) : \mathcal{G} \to \Re$ is a functional, we denote the Fréchet differential of T at g with increment $h \in \mathcal{G}$ of order $i = 1, 2$ by $\delta_F^i T(g;h)$. Note that if the Fréchet differentials of order $i = 1, 2$ of T exist at g, they coincide with the Gateaux differentials of order $i = 1, 2$ at g, denoted by $\delta_G^i T(g;h) = \frac{d^i}{d\alpha} T(g+\alpha h)\mid_{\alpha=0}$ (see Luenberger (1969) and Lusternik and Sobolev (1964)). Furthermore, there exists a Taylor's Theorem (Graves (1927)) such that we can write, $T(g+h) = T(g) + \delta_F^1 T(g;h) + \int_0^1 \delta_F^2 T(g+th;h)(1-t)dt = T(g) + \frac{d_F}{dg} T(g)h + \frac{h^2}{2}\int_0^1 \frac{d_F^2}{dg^2} T(g+th)(1-t)dt$ where $\frac{d_F}{dg}T(g)$ and $\frac{d_F^2}{dg^2}T(g)$ are called the first and second order Fréchet derivatives of T at g. We take the norm in $\mathcal{G}$ to be $sup_{x\in\Re^D}|g(x)|$.

Assumption A6. 1. $logf_\varepsilon(y-g_0(x);\theta)$ is twice Fréchet differentiable at g_0 with increment $h(x) = g(x) - g_0(x)$ and denote the Fréchet derivatives of order $i = 1, 2$ at g_0 by $\frac{d_F^i}{dg^i} logf_\varepsilon(y-g_0(x);\theta)$; 2.

$\frac{d_F}{dg} log f_\varepsilon(y - g_0(x);\theta_0)$ is continuous at every $x \in G$; 3. We assume that the matrix

$$\sigma_F^2 = E\left(\left(\frac{\partial}{\partial\theta} log f_\varepsilon(y_i - g_0(x_i);\theta_0) + (y_i - m(x_i;\theta_0, g_0))\int \frac{\partial^2}{\partial\theta\partial\eta} f_\varepsilon(y - g_0(x_i);\theta_0) f_\varepsilon(y - g_0(x_i);\theta_0)dy\right)\right.$$
$$\left.\times\left(\frac{\partial}{\partial\theta} log f_\varepsilon(y_i - g_0(x_i);\theta_0) + (y_i - m(x_i;\theta_0, g_0))\int \frac{\partial^2}{\partial\theta\partial\eta} f_\varepsilon(y - g_0(x_i);\theta_0) f_\varepsilon(y - g_0(x_i);\theta_0)dy\right)'\right).$$

exists and is positive definite.

We observe that Fréchet derivatives are, in this case, bounded linear operators from $\mathcal{G}$ to $\Re$.

Assumption A7. 1. $\sup_{\theta\in S_{0,\theta}} \left|E\left(\frac{\partial}{\partial\eta} log f_\varepsilon(y - g_0(x_i);\theta)|x_i\right)\right| < \infty$; 2. $\frac{\partial^2}{\partial\theta_i\partial\eta} log f_\varepsilon(y - g_0(x_i);\theta)$ is continuously differentiable in $S_{0,\theta}$ and $E\left(\sup_{\theta\in S_{0,\theta}}\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\eta} log f_\varepsilon(y - g_0(x_i);\theta)\right| |x_i\right) < \infty$ for all $x_i \in G$ almost surely; 3. $\sup_{\theta\in S_{0,\theta}}\left|E\left(\frac{\partial^2}{\partial\theta_i\partial\eta} log f_\varepsilon(y - g_0(x_i);\theta)\right| |x_i\right) < \infty$; 4. $E\left(\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\eta} log f_\varepsilon(y - g_0(x_i);\theta)|x_i\right)$ is continuous in $S_{0,\theta}$ almost surely; 5. $E(|y_i - g_0(x_i)|) < \infty$; 6. $\frac{\partial^2}{\partial\theta_i\partial\theta_j} E\left(g_0^{(2)}(x_i)\frac{\partial}{\partial\eta} log f_{y|x}(y;\theta, g_0(x_i))\right)$ is continuous in $S_{0,\theta}$ almost surely.

We now define a specific estimator $\hat{m}$ to be used in equation (3). For any $x \in G$ we define $\hat{m}(x) \equiv \hat{\alpha}(x)$ where

$$(\hat{\alpha}(x), \hat{\beta}(x)) = \underset{\alpha(x),\beta(x)}{argmin} \sum_{i=1}^{n} (y_i - \alpha(x) - \beta(x)(x_i - x))^2 K\left(\frac{x_i - x}{h_n}\right) \quad (7)$$

where $0 < h_n \to 0$ as $n \to \infty$ is a bandwidth. This is the local linear estimator of Fan (1992, 1993). We note that the Nadaraya-Watson estimator used in Fan et al. (1996) as a special case of the local linear estimator ($\beta(x) = 0$). In addition, contrary to the Nadaraya-Watson estimator, the local linear estimator is design-adaptive, has good boundary properties and is mini-max efficient.[3] The next two theorems establish the consistency and $\sqrt{n}$ asymptotic normality of $\hat{\theta}$ after suitable centering.[4]

Theorem 1 *Given assumptions A1.1-3, A2, A3, A4, and the estimator $\hat{m}(x)$ defined in (7), if $\frac{nh_n^3}{log(n)} \to \infty$ as $n \to \infty$ then $\hat{\theta} - \theta_0 = o_p(1)$.*

Theorem 2 *Given assumptions A1-A7, and the estimator $\hat{m}(x)$ defined in (7), if $\frac{nh_n^3}{log(n)} \to \infty$ as $n \to \infty$ and $h_n = O(n^{-1/5})$ then $\sqrt{n}(\hat{\theta} - \theta_0 - B_n) \xrightarrow{d} N(0, \bar{H}^{-1}\sigma_F^2\bar{H}^{-1})$, where $B_n = -\frac{h_n^2}{2}\sigma_K^2\bar{H}^{-1}M + o_p(h_n^2)$, M is a P-vector with p^{th} element given by $M_p = E\left(g_0^{(2)}(x_i)\frac{\partial^2}{\partial\theta_p\partial\eta} f_\varepsilon(y_i - g_0(x_i);\theta_0)\right)$, $\bar{H}$ is as defined in A5.4 and σ_F^2 is given in A6.3.*

We note that theorems 1 and 2 require $\underset{x\in G}{sup}|\hat{m}(x) - m(x)| = O_p\left(\left(\frac{log\, n}{nh_n}\right)^{1/2} + h_n^2\right)$. Since both the local

[3] See Fan (1993) and Li and Racine (2007).

[4] The proofs for theorems 1 and 2 as well as the proof for Lemma 1 in section 3.2 can be found in Martins-Filho and Yao (2011).

linear and Nadaraya-Watson estimators for m share this uniform convergence rate under our assumptions, either could be used to attain the stated asymptotic characterization for $\hat{\theta}$.

Practical use of theorems 1 and 2 requires the verification of assumptions A1.3-4 and A4-A7 for a chosen density f_ε. Although it is possible to find specific classes of densities that do not satisfy our assumptions, a more useful exercise is to verify that such assumptions are met by commonly considered stochastic frontier specifications. Simar and Wilson (2010) observe that most applied stochastic frontier models rely on the density given by (1), which indeed satisfies all required assumptions. This is verified in Martins-Filho and Yao (2011) where in addition the structure of the matrices which appear in Theorem 2 are also obtained. Specifically, it is of great practical interest to obtain expressions for σ_F^2 and $\bar{H}$. These expressions allow for the construction of confidence intervals and asymptotically valid hypothesis testing. Let $s^2 \equiv \sigma_u^2 + \sigma_v^2$, $\lambda = \sqrt{\sigma_u^2/\sigma_v^2}$, $w = \frac{1}{2\lambda s^3}$, $w_1 = -\frac{1}{2\lambda}\frac{\sigma_u^2(\sigma_u^2+2\sigma_v^2)}{(\sigma_v^2)^2 s^3}$, $I = \int \frac{e\frac{\sqrt{2}}{s\pi^{3/2}}}{1-erf(e\frac{\lambda}{s\sqrt{2}})} exp(-e^2(\frac{\lambda^2}{s^2}+\frac{1}{2s^2}))de$, $I_1 = \int \frac{e^2\frac{\sqrt{2}}{s\pi^{3/2}}}{1-erf(e\frac{\lambda}{s\sqrt{2}})} exp(-e^2(\frac{\lambda^2}{s^2}+\frac{1}{2s^2}))de$, $C_1 = \frac{\gamma(\sigma_u^2,\sigma_v^2)}{s^4} + \frac{\lambda}{s}wI - (\frac{\lambda}{s})^2 w\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1} + w\sqrt{\frac{2}{\pi(\lambda^2+1)}}$, $C_2 = \frac{\gamma(\sigma_u^2,\sigma_v^2)}{s^4} + \frac{\lambda}{s}w_1 I - (\frac{\lambda}{s})^2 w_1\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1} + w_1\sqrt{\frac{2}{\pi(\lambda^2+1)}}$ where erf is the Gaussian error function. If we denote the (i,j) element of σ_F^2 by $\sigma_{F(i,j)}^2$, then

$$\begin{aligned}
\sigma_{F(1,1)}^2 &= \frac{1}{2s^4} + w^2 I_1 + C_1^2(s^2 - \gamma(\sigma_u^2,\sigma_v^2)^2) + \frac{1}{s^4}C_1(-3\sigma_v^2\sqrt{2\sigma_u^2/\pi} - (2\sigma_u^2)^{3/2}(1/\sqrt{\pi}) + \gamma(\sigma_u^2,\sigma_v^2)s^2) \\
&- 2wC_1\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1} \\
\sigma_{F(1,2)}^2 &= \frac{1}{2s^4} + ww_1 I_1 + C_1C_2(s^2 - \gamma(\sigma_u^2,\sigma_v^2)^2) + (C_1+C_2)(-3\sigma_v^2\sqrt{2\sigma_u^2/\pi} - (2\sigma_u^2)^{3/2}(1/\sqrt{\pi}) \\
&+ \gamma(\sigma_u^2,\sigma_v^2)s^2) - (wC_2 + w_1C_1)\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1} \\
\sigma_{F(2,2)}^2 &= \frac{1}{2s^4} + w_1^2 I_1 + C_2^2(s^2 - \gamma(\sigma_u^2,\sigma_v^2)^2) + \frac{1}{s^4}C_2(-3\sigma_v^2\sqrt{2\sigma_u^2/\pi} - (2\sigma_u^2)^{3/2}(1/\sqrt{\pi}) + \gamma(\sigma_u^2,\sigma_v^2)s^2) \\
&- 2C_2w_1\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1}
\end{aligned}$$

and $\bar{H} = \begin{pmatrix} \bar{H}_{11} & \bar{H}_{12} \\ \bar{H}_{21} & \bar{H}_{22} \end{pmatrix}$ where $\bar{H}_{11} = -\frac{1}{2s^4} - w^2 I_1 + 2\frac{\partial}{\partial\sigma_u^2}\gamma(\sigma_u^2,\sigma_v^2)C_1$, $\bar{H}_{12} = -\frac{1}{2s^4} - ww_1 I_1 + 2\frac{\partial}{\partial\sigma_u^2}\gamma(\sigma_u^2,\sigma_v^2)C_2$, and $\bar{H}_{22} = -\frac{1}{2s^4} - w_1^2 I_1$. Given that $\hat{\theta}$ is a consistent estimator for θ_0, consistent estimators for σ_F^2 and $\bar{H}$ can be obtained using the above expression and a numerical evaluation of the integrals in I and I_1.

From a practical perspective, theorems 1 and 2 provide asymptotic results that justify the extension of the stochastic frontier model of Aigner et al. to the case where the frontier is fully nonparametric. The results seem to be of immediate interest to empirical researchers given the enormous popularity of the model first proposed by Aigner at al. From a technical perspective, it is not surprising that in theorem 2 a parametric estimator - $\hat{\theta}$ - that is based on averages of a nonparametrically estimated curve converges at a parametric $\sqrt{n}$ rate (see Doksum and Samarov (1995)). Additionally, theorem 2 shows that $\hat{\theta}$ carries an asymptotic bias that does not decay to zero when normalized by $\sqrt{n}$. The presence of this bias is precisely the motivation for the generalized profile likelihood estimator proposed by Severini and Wong (1992) for conditionally parametric models.[5] Hence, we now turn our attention to the estimator $\tilde{\theta}$.

3.2 The estimator $\tilde{\theta}$

Suppose we consider a reparametrization of $f_\varepsilon(y - g(x);\theta)$ given by $f_\varepsilon(y - \alpha_\theta(x);\theta)$, where for every $x \in G$, $\alpha_\theta(x) : \Theta \to \mathcal{H}$ with $\alpha_{\theta_0}(x) = g_0(x)$. The parametric submodel described by the curve $\alpha_\theta(x)$ has Fisher information given by

$$\mathcal{I}_0\left(\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)\right) = E\left(\frac{\partial}{\partial\theta}logf_\varepsilon(y-g_0(x);\theta_0) + \frac{d_F}{dg}logf_\varepsilon(y-g_0(x);\theta_0)\frac{\partial}{\partial\theta}\alpha_{\theta_0(x)}\right)\left(\frac{\partial}{\partial\theta}logf_\varepsilon(y-g_0(x);\theta_0)\right. \\ \left. + \frac{d_F}{dg}logf_\varepsilon(y-g_0(x);\theta_0)\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)\right)'$$

where $\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)$ is the tangent vector associated with $\alpha_\theta(x)$ evaluated at θ_0. As argued in van der Vaart (1999), it is desirable to minimize the information associated with the parametric submodel induced by $\alpha_\theta(x)$. Since the information depends only on the tangent vector it is natural to define[6]

$$\mathcal{I}_{\theta_0} = \inf_{\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)} \mathcal{I}_0\left(\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)\right). \tag{8}$$

Bickel et al. (1993) show that provided $E((\frac{d_F}{dg}logf_\varepsilon(y-g_0(x);\theta_0)^2|x) > 0$ the minimizer for (8), say $\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)^*$, satisfies

$$\frac{\partial}{\partial\theta_k}\alpha_{\theta_0}(x)^* = -\frac{E\left(\frac{\partial}{\partial\theta_k}logf_\varepsilon(y-g_0(x);\theta_0)\frac{d_F}{dg}logf_\varepsilon(y-g_0(x);\theta_0)|x\right)}{E((\frac{d_F}{dg}logf_\varepsilon(y-g_0(x);\theta_0)^2|x)} \text{ for } k = 1,\cdots,P. \tag{9}$$

The tangent vector $\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)^*$ is called a least favorable direction. Interestingly, if we let $\alpha_\theta(x) \in \mathcal{G}$ be the unique maximizer of $E(logf_\varepsilon(y-g(x);\theta)|x)$ for fixed x and θ, using a Taylor's expansion around θ_0 shows that $\alpha_\theta(x)$ minimizes Fisher's information, provided $E(logf_\varepsilon(y-g(x);\theta)|x) < E(logf_\varepsilon(y-g_0(x);\theta_0)|x)$

[5] Naturally, the asymptotic bias associated with $\hat{\theta}$ can be eliminated by choosing a non optimal bandwidth decay rate (undersmoothing) for the estimator $\hat{m}(x)$.

[6] We follow the usual practice of defining, for any two squared matrices A and B, $A \leq B$ if, and only if, $B - A$ is positive semidefinite.

whenever $\theta \neq \theta_0$. As such, Severini and Wong define a least favorable curve for a conditional parametric model as $\alpha_\theta(x) : \Theta \rightarrow \mathcal{H}$ that for every $x \in G$ satisfies the following: (1) For each $x \in G$, $\alpha_{\theta_0}(x) = g_0(x)$; (2) For each $x \in G$, $\frac{\partial}{\partial\theta}\alpha_\theta(x)$ and $\frac{\partial^2}{\partial\theta\partial\theta'}\alpha_\theta(x)$ exist and $\sup_{x\in G}|\frac{\partial}{\partial\theta_p}\alpha_\theta(x)| < \infty$, and $\sup_{x\in G}|\frac{\partial^2}{\partial\theta_p\partial\theta_m}\alpha_\theta(x)| < \infty$ for all $m, p = 1, ..., P$; (3) For each $x \in G$, $\frac{\partial}{\partial\theta}\alpha_{\theta_0}(x)$ minimizes the Fisher information for the parametric sub-model described by $log f_\varepsilon(y - \alpha_\theta(x); \theta)$.

Now, consider an arbitrary estimator for $\alpha_\theta(x)$ given by $\dot{\alpha}_\theta(x)$ and define

$$\dot{\theta} \equiv \underset{\theta}{argmax}\ \bar{l}_n(\theta, \dot{\alpha}_\theta) = \underset{\theta}{argmax}\frac{1}{n}\sum_{i=1}^{n} log_\varepsilon(y_i - \dot{\alpha}_\theta(x_i); \theta). \tag{10}$$

The following theorem establishes the asymptotic normality and consistency of $\dot{\theta}$ conditional on the asymptotic properties of $\dot{\alpha}_\theta(x)$ and some regularity conditions on the class containing f_ε.

Theorem 3 *(Severini and Wong (1992))* $\dot{\theta} - \theta = o_p(1)$ *and* $\sqrt{n}(\dot{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}_{\theta_0}^{-1})$ *provided the following assumptions hold.*

PA1: 1. For fixed (but arbitrary) $\theta' \in \Theta$ *and* $\eta' \in \mathcal{H}$ *and for all* $\theta \in \Theta$ *and* $\eta \in \mathcal{H}$*, let* $\rho(\theta, \eta) = \int log f_\varepsilon(y - \eta; \theta) f_\varepsilon(y - \eta'; \theta')dy$*. If* $\theta \neq \theta'$ *then* $\rho(\theta, \eta) < \rho(\theta', \eta')$*; 2.* $\tilde{I}_\theta(\theta, \eta) > 0$ *for all* $\theta \in \Theta$ *and* $\eta \in \mathcal{H}$ *where*

$$\begin{aligned}\tilde{I}_\theta(\theta, \eta) &= E\left(\frac{\partial}{\partial\theta} log f_\varepsilon(y - \eta; \theta)\frac{\partial}{\partial\theta} log f_\varepsilon(y - \eta; \theta)'\right) \\ &- \frac{E(\frac{\partial}{\partial\eta} log f_\varepsilon(y - \eta; \theta)\frac{\partial}{\partial\theta} log f_\varepsilon(y - \eta; \theta))E(\frac{\partial}{\partial\eta} log f_\varepsilon(y - \eta; \theta)\frac{\partial}{\partial\theta'} log f_\varepsilon(y - \eta; \theta))}{E((\frac{\partial}{\partial\eta} log f_\varepsilon(y - \eta; \theta)^2)}\end{aligned}$$

PA2: For $r, s = 0, 1, 2, 3, 4$ *and* $r + s \leq 4$, $\frac{\partial^{r+s}}{\partial\theta_p^r\partial\eta^s} log f_\varepsilon(y - \eta; \theta)$ *exists for* $p = 1, \cdots, P$ *and*

$$E\left(\sup_{\theta\in\Theta, \eta\in\mathcal{H}}\left|\frac{\partial^{r+s}}{\partial\theta_p^r\partial\eta^s} log f_\varepsilon(y - \eta; \theta)\right|^2\right) < \infty.$$

PA3: 1. For all $x \in G$ *and* $\theta \in \Theta$, $q_\theta(x) \equiv \dot{\alpha}_\theta(x) - \alpha_\theta(x) = o_p(1)$*. For each* $\theta \in \Theta$ *and for all* $r, s = 0, 1, 2$ *with* $r + s \leq 2$ *we have that* $\frac{\partial^{r+s}}{\partial x^r\partial\theta_p^s}\alpha_\theta(x)$ *and* $\frac{\partial^{r+s}}{\partial x^r\partial\theta_p^s}\dot{\alpha}_\theta(x)$ *exist; 2.* $\sup_{x\in G}|q_{\theta_0}(x)| = o_p(n^{-a})$ *and* $\sup_{x\in G}|\frac{\partial}{\partial\theta_p} q_{\theta_0}(x)| = o_p(n^{-b})$ *with* $a + b \geq 1/2$ *and* $a \geq 1/4$*; 3.* $\sup_{\theta\in\Theta}\sup_{x\in G}|q_\theta(x)| = o_p(1)$, $\sup_{\theta\in\Theta}\sup_{x\in G}\left|\frac{\partial}{\partial\theta_p} q_\theta(x)\right| = o_p(1)$, $\sup_{\theta\in\Theta}\sup_{x\in G}\left|\frac{\partial^2}{\partial\theta_m\partial\theta_p} q_\theta(x)\right| = o_p(1)$ *for all* $m, p = 1, \cdots, P$*; 4. For some* $\delta > 0$, $\sup_{x\in G}|\frac{\partial}{\partial x} q_{\theta_0}(x)| = o_p(n^{-\delta})$ *and* $\sup_{x\in G}|\frac{\partial^2}{\partial x\partial\theta_p} q_{\theta_0}| = o_p(n^{-\delta})$.

The importance of Theorem 3 rests principally on the fact that if an estimator for the least favorable curve can be found to satisfy PA3, then an estimator for θ that satisfies equation (10) has the stated asymptotic properties. In contrast with the estimator described in Theorem 2, $\dot{\theta}$ is not biased asymptotically and is efficient in the sense that it is based on a suitable estimator $\dot{\alpha}_\theta(x)$ for the least favorable curve.

We now show that the estimator defined in equation (5), i.e., $\hat{\alpha}_\theta(x)$ satisfies PA3. This result extends Lemma 5 in Severini and Wong, where it is shown that if the least favorable curve is estimated using a local constant approximation, i.e., in equation (5) $g_x(x_i) = \alpha(x)$ ($\beta(x) = 0$) PA3 is met. It is convenient for our purposes to establish the following Lemma 1, which will be used repeatedly in the proof of Lemma 2, which verifies $PA3$.

Lemma 1 *Assume that PA1 and PA2 hold. If $nh_n^3 \to \infty$ as $n \to \infty$ we have*

$$\sup_{x\in G,\theta\in\Theta,\eta\in\mathcal{H}} \left| \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}\frac{\partial}{\partial\eta} log f_\varepsilon(y_i-\eta;\theta) K\left(\frac{x_i-x}{h_n}\right)\left(\begin{array}{c}1\\ x_i-x\end{array}\right) - \frac{1}{h_n}E\left(\frac{\partial}{\partial\eta} log f_\varepsilon(y_i-\eta;\theta) K\left(\frac{x_i-x}{h_n}\right)\left(\begin{array}{c}1\\ x_i-x\end{array}\right)\right)\right| = \left(\begin{array}{c} O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right)\\ O_p\left(h_n\left(\frac{log(n)}{nh_n}\right)^{1/2}\right)\end{array}\right).$$

Lemma 2 *Assume that $\alpha_\theta(x)$ satisfies $\frac{\partial}{\partial\eta}E(log f_\varepsilon(y-\alpha_\theta(x);\theta)) = 0$ and is the unique maximizer of $E(log f_\varepsilon(y-\alpha_\theta(x);\theta))$. Define for some $\bar{d}_\theta(x) = \left(\begin{array}{c}\bar{d}_{0,\theta}(x)\\ \bar{d}_{1,\theta}(x)\end{array}\right) \in \Re^2$ the function*

$$G_\theta(\bar{d}_\theta(x)) = f_x(x)\int_{S_D} E\left(\frac{\partial}{\partial\eta} log f_\varepsilon(y_i-\alpha_\theta(x)-\bar{d}_{0,\theta}(x)-\bar{d}_{1,\theta}(x)h_n\psi;\theta)|x\right)K(\psi)\left(\begin{array}{c}1\\ h_n\psi\end{array}\right)d\psi$$

and assume that for all $\epsilon > 0$ there exists $\delta > 0$ such that whenever $||\sup_{\theta\in\Theta,x\in G} G_\theta(\bar{d}_\theta(x))|| \le \delta$ we have $\sup_{\theta\in\Theta,x\in G}|\bar{d}_{0,\theta}(x)| < \epsilon$ and $\sup_{\theta\in\Theta,x\in G}|\bar{d}_{1,\theta}(x)| < \epsilon$. Then under the assumptions in Lemma 1 and regularity conditions PB, PC and PD we have that the estimator $\hat{\alpha}_\theta(x) = \hat{\eta}_0$ obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}\frac{\partial}{\partial\eta} log f_\varepsilon(y_i-\eta_0-\eta_1(x_i-x);\theta)K\left(\frac{x_i-x}{h_n}\right)\left(\begin{array}{c}1\\ x_i-x\end{array}\right) = \left(\begin{array}{c}0\\ 0\end{array}\right)$$

with respect to η_0 and η_1 for each x and θ satisfies:

a) $\sup_{\theta\in\Theta,x\in G}|\hat{\alpha}_\theta(x)-\alpha_\theta(x)| = o_p(1)$ *and* $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)-\frac{\partial}{\partial x}\alpha_\theta(x)| = o_p(1)$

b) $\sup_{\theta\in\Theta,x\in G}|\hat{\alpha}_\theta(x)-\alpha_\theta(x)| = O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O_p\left(h_n^2\right)$ *and*

$\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)-\frac{\partial}{\partial x}\alpha_\theta(x)| = O_p\left(\left(\frac{log(n)}{nh_n^3}\right)^{1/2}\right) + O_p\left(h_n\right)$

c) $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_j}\hat{\alpha}_\theta(x)-\frac{\partial}{\partial\theta_j}\alpha_\theta(x)| = O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O_p\left(h_n^2\right)$ *and*

$\sup_{\theta\in\Theta,x\in G}|\frac{\partial^2}{\partial\theta_j\partial x}\hat{\alpha}_\theta(x)-\frac{\partial^2}{\partial\theta_j\partial x}\alpha_\theta(x)| = O_p\left(\left(\frac{log(n)}{nh_n^3}\right)^{1/2}\right) + O_p\left(h_n\right)$

d) $\sup_{\theta\in\Theta,x\in G}|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\hat{\alpha}_\theta(x)-\frac{\partial^2}{\partial\theta_j\partial\theta_k}\alpha_\theta(x)| = O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O_p\left(h_n^2\right)$ *and*

$$\sup_{\theta\in\Theta,x\in G}|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial x}\hat{\alpha}_\theta(x)-\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial x}\alpha_\theta(x)| = O_p\left(\left(\frac{log(n)}{nh_n^3}\right)^{1/2}\right)+O_p(h_n) \text{ provided that } nh_n^3 \to \infty \text{ as } n \to \infty.$$

A direct consequence of Lemma 2 and Theorem 3 is that $\tilde{\theta}$ as defined by equations (5) and (6) has the following asymptotic properties

$$\tilde{\theta} - \theta_0 = o_p(1) \text{ and } \sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}_0^{-1}). \tag{11}$$

The constraints imposed on f_ε in Theorem 2 and Lemma 2 go beyond those required in assumptions A4-A7 on the previous section. As a result they must be checked for any assumed specific conditional density f_ε. Most importantly, we have verified that all assumptions placed on f_ε in Lemma 2 are satisfied by the conditional density given in equation (1).[7] Of particular interest is the exact form that $\mathcal{I}_0$ takes when f_ε in equation (1) is used in estimation. In this case the $\mathcal{I}_0$ matrix has (i, j) elements with $i, j = 1, 2$ given by

$$\begin{aligned}\mathcal{I}_{\theta_0}(1,1) = &\frac{1}{2s^4} + w^2 I_1 + \alpha'^2_{\sigma_u^2}(1/s^2 + (\lambda/s)^2 I_2) + \frac{1}{s^4}\left(\alpha'_{\sigma_u^2}\gamma(\sigma_u^2,\sigma_v^2) + \alpha'_{\sigma_u^2}\frac{\lambda}{s}\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1}\right)\\ &- \frac{1}{s^2}\alpha'_{\sigma_u^2}\sqrt{\frac{2}{\pi(\lambda^2+1)}}\left(\frac{\lambda}{s} + 2w\frac{s^2}{\lambda^2+1}\right) - 2\alpha'_{\sigma_u^2}\frac{\lambda}{s}wI + \frac{1}{s^6}\alpha'_{\sigma_u^2}\left(-3\sigma_v^2\sqrt{2/\pi\sigma_u^2} - \frac{(2\sigma_u^2)^{3/2}}{\sqrt{\pi}}\right)\end{aligned}$$

$$\begin{aligned}\mathcal{I}_{\theta_0}(1,2) = \mathcal{I}_{\theta_0}(2,1) = &\frac{1}{2s^4} + ww_1 I_1 + \alpha'_{\sigma_u^2}\alpha'_{\sigma_v^2}(1/s^2 + (\lambda/s)^2 I_2) + \frac{1}{2s^4}\left((\alpha'_{\sigma_u^2} + \alpha'_{\sigma_v^2})(\gamma(\sigma_u^2,\sigma_v^2)\right.\\ &\left. + \frac{\lambda}{s}\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1})\right)\\ &- \frac{1}{s^2}\sqrt{\frac{2}{\pi(\lambda^2+1)}}\left(\frac{\lambda}{2s}(\alpha'_{\sigma_u^2} + \alpha'_{\sigma_v^2}) + (w\alpha'_{\sigma_v^2} + w_1\alpha'_{\sigma_u^2})\frac{s^2}{\lambda^2+1}\right) - (w\alpha'_{\sigma_v^2} + w_1\alpha'_{\sigma_u^2})\frac{\lambda}{s}I\\ &+ \frac{1}{2s^6}(\alpha'_{\sigma_u^2} + \alpha'_{\sigma_v^2})\left(-3\sigma_v^2\sqrt{2/\pi\sigma_u^2} - \frac{(2\sigma_u^2)^{3/2}}{\sqrt{\pi}}\right)\end{aligned}$$

$$\begin{aligned}\mathcal{I}_{\theta_0}(2,2) = &\frac{1}{2s^4} + w_1^2 I_1 + \alpha'^2_{\sigma_v^2}(1/s^2 + (\lambda/s)^2 I_2) + \frac{1}{s^4}\left(\alpha'_{\sigma_v^2}\gamma(\sigma_u^2,\sigma_v^2) + \alpha'_{\sigma_v^2}\frac{\lambda}{s}\sqrt{\frac{2}{\pi(\lambda^2+1)}}\frac{s^2}{\lambda^2+1}\right)\\ &- \frac{1}{s^2}\alpha'_{\sigma_v^2}\sqrt{\frac{2}{\pi(\lambda^2+1)}}\left(\frac{\lambda}{s} + 2w_1\frac{s^2}{\lambda^2+1}\right) - 2\alpha'_{\sigma_v^2}\frac{\lambda}{s}w_1 I\\ &+ \frac{1}{s^6}\alpha'_{\sigma_v^2}\left(-3\sigma_v^2\sqrt{2/\pi\sigma_u^2} - \frac{(2\sigma_u^2)^{3/2}}{\sqrt{\pi}}\right)\end{aligned}$$

where $\alpha'_{\sigma_u^2} = C_1(\frac{1}{s^2} + (\frac{\lambda}{s})^2 I_2)^{-1}$, $\alpha'_{\sigma_v^2} = C_2(\frac{1}{s^2} + (\frac{\lambda}{s})^2 I_2)^{-1}$ and all remaining constants are as defined following Theorem 2. Given that $\tilde{\theta}$ is a consistent estimator for θ_0, consistent estimators for $\mathcal{I}_{\theta_0}$ can be obtained given the above expression.

[7] See Martins-Filho and Yao (2011).

4 Monte Carlo study

4.1 Implementation

The estimator $\tilde{\theta}$ described by (5) and (6) is implemented using the following algorithm:

Step 1: Obtain an initial maximum likelihood estimate $\check{\theta}$ based on a linear parametric specification for $g(x) = x'\beta$ and define an initial value $\theta^{(k)} \equiv \check{\theta}$, where $k = 0$.

Step 2: For each x_j, $j = 1, \cdots, n$ in the sample, maximize the local log-likelihood function based on $\theta^{(k)}$

$$\left(\hat{\alpha}_{\theta^{(k)}}(x_j), \hat{\beta}_{\theta^{(k)}}(x_j)\right) = argmax_{\{\alpha(x_j),\beta(x_j)\}} \frac{1}{n}\sum_{i=1}^{n} \log f_\varepsilon(y_i - \alpha(x_j) - \beta(x_j)(x_i - x_j); \theta^{(k)}) K\left(\frac{x_i - x_j}{h_n}\right) \tag{12}$$

Step 3: Based on $\hat{\alpha}_{\theta^{(k)}}(x_i)$ for $i = 1, \cdots, n$ from step 2, maximize the global log-likelihood function with respect to θ, and obtain

$$\theta^{(k+1)} = argmax_\theta \frac{1}{n}\sum_{i=1}^{n} \log f_\varepsilon(y_i - \hat{\alpha}_{\theta^{(k)}}(x_i); \theta) \tag{13}$$

Step 4: Using $\theta^{(k+1)}$ repeat step 2 and then step 3. Continue this cycle until k is such that $||\theta^{(k+1)} - \theta^{(k)}||_E < \epsilon$. We set $\epsilon = 0.001$.

Step 5: Fix the estimator $\tilde{\theta}$ at the value obtained from the last cycle of step 4 and put $\tilde{g}(x) = \hat{\alpha}_{\tilde{\theta}}(x)$ in step 2.

As pointed out by Lam and Fan (2008), this algorithm is equivalent to a Newton-Raphson procedure but (13) incorporates the functional dependence of $\hat{\alpha}_\theta(x)$ on θ by using the value of $\theta^{(k)}$ from the previous step as a proxy for θ. Specifically, the algorithm treats the $\frac{d}{d\theta}\hat{\alpha}_\theta(x)$ and $\frac{d^2}{d\theta d\theta'}\hat{\alpha}_\theta(x)$ in the Newton-Raphson procedure as zeros and computes $\hat{\alpha}_\theta(x)$ using the values of $\theta^{(k)}$ in the previous iteration. Thus the maximization is easier to carry out. We recommend calculating $(\hat{\alpha}_{\theta^{(k)}}(x), \frac{\partial \hat{\alpha}_{\theta^{(k)}}(x)}{\partial x})$ in step 2 at a fixed but fine grid of points of x. Then use linear interpolation to calculate the other values of $\hat{\alpha}_{\theta^{(k)}}(x)$.

We could utilize the consistent plug-in estimator $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_v^2)$ (Fan et al.(1996)) in Step 1 of our algorithm but we avoid doing so in the simulation to provide a fair performance comparison of $(\hat{\sigma}_u^2, \hat{\sigma}_v^2)$ and $(\tilde{\sigma}_u^2, \tilde{\sigma}_v^2)$. Given the strict concavity of $log f_\epsilon(y - g(x)); \sigma_u^2, \sigma_v^2)$ the initial estimates are important only for computational speed. Given that the linear specification is a simple and popular alternative, we use the maximum likelihood estimates based on a linear parametric specification for $g(x)$ as the initial values for both $(\hat{\sigma}_u^2, \hat{\sigma}_v^2)$ and $(\tilde{\sigma}_u^2, \tilde{\sigma}_v^2)$.

Implementation of our estimator requires the selection of a bandwidth h_n. Since $g_0(x) = E(y|x)+\gamma(\theta_0)$ where $\gamma(\theta)$ is a non-stochastic function of θ, we use the data driven rule-of-thumb bandwidth $\hat{h}_{ROT}$ of Ruppert et al. (1995). We observe that $\hat{h}_{ROT}/h_n^* - 1 = o_p(1)$ where h_n^* is the bandwidth that minimizes the local linear regression estimator's asymptotic mean squared error (AMISE). Since, $h_n^* = O(n^{-\frac{1}{5}})$ we have that $h_n^* \to 0$ at a speed that is consistent with that required by the asymptotic theory.

4.2 Data Generation and Results

In this section, we perform a Monte Carlo study which implements our profile likelihood semi-parametric stochastic frontier estimator and provides evidence on its finite sample performance. We consider the stochastic production frontier model of Fan et al. (1996) where output-input pairs (y_i, x_i) are generated in accordance with assumptions A1, A2 and the conditional density given by (1). We consider four different functional forms for $g(x)$: $g_1(x) = 1 + x$, $g_2(x) = 1 + ln(1 + x)$, $g_3(x) = 1 - \frac{1}{1+x}$ and $g_4(x) = 1 + 0.5arctan(20(x - 0.5))$. The first three functions are considered in Fan et al. (1996) and we introduce the last one, which exhibits more pronounced nonlinearity. We generate the univariate input x_i from an uniform distribution on $[0, 1]$. To facilitate comparison, the parameter $\theta = (\sigma_u^2, \sigma_v^2)$ is set at $\theta^{(1)} = (1.379, 0.501)$, $\theta^{(2)} = (0.988, 0.642)$ and $\theta^{(3)} = (0.551, 0.799)$ to coincide with the values considered by Aigner et al. (1977) and Fan et al. (1996). Note that the ratio $(\frac{\sigma_u^2}{\sigma_v^2})$ decreases from $\theta^{(1)}$ to $\theta^{(3)}$. Figure 1 provides a plot of a typical simulated sample for the production frontier $g_4(x)$, where the sample was generated with $(\sigma_u^2, \sigma_v^2) = (1.379, 0.501)$. Superimposed on the plot are the true production frontier, estimated frontiers with both a profile likelihood (PL) estimator $(\tilde{\theta})$ and a plug in (PI) estimator $(\hat{\theta})$. The figure suggest that both estimators seem to capture fairly well the shape of the underlying production frontier.

The PL estimator is implemented using the algorithm described above and the bandwidth $\hat{h}_{ROT}$. We use the Epanechnikov kernel in the estimation, which satisfies assumption A3. For comparison purpose we also include in the study the PI estimator and a parametric maximum likelihood (ML) estimator constructed under the correct specification of the production frontier. The parametric ML estimator is, under these circumstances, expected to outperform the PL and PI estimators. The PI estimator is implemented as described by Fan et al. (1996), but here the conditional expectation $m_0(x)$ is estimated via a local linear estimator, so that the asymptotic characterization given in section 3 is applicable. We set the sample sizes at $n = 300$, 600 and 900, and perform 500 replications for each experimental design. We investigate the performances of PL, PI and ML in estimating the global parameter θ. The performance of the estimators is summarized by their bias, standard deviation and root mean squared error, which are provided in Tables 1-4.

As suggested by the asymptotic theory, the performance of all three estimators in terms of bias, standard deviation, and root mean squared error generally improves as n increases, with a few exceptions for the bias. All three estimators generally exhibit a negative bias in estimating σ_u^2 and a positive bias in estimating σ_v^2 with a few exceptions for small samples when $(\sigma_u^2, \sigma_v^2) = (0.551, 0.799)$. It is also clear that it is harder to estimate σ_u^2 than to estimate σ_v^2, as the bias, standard deviation and root mean squared error for all estimators of σ_v^2 are smaller than those of σ_u^2. The above observations are consistent with the results in the Monte Carlo studies of Fan et al. (1996) and Aigner et al. (1977). General conclusions regarding the relative performance of the estimators are unambiguous and conform with our expectations. The parametric ML estimator performs best since it is based on a correct specification of the production frontier and the distribution of the composite error term. Among the two semiparametric estimators that relax the parametric assumption on the production frontier, the PL estimator we propose outperforms the PI estimator of Fan et al. across almost all experimental designs, and the improvement is significant. For example, in the simulation with $g_1(x)$ as production frontier and $(\sigma_u^2, \sigma_v^2) = (1.379, 0.501)$, the reduction in the root mean squared error from the PL estimator over that of the PI estimator are about 5% in estimating both σ_u^2 and σ_v^2. A few exceptions occur for the smallest sample when $(\sigma_u^2, \sigma_v^2) = (0.551, 0.799)$, which corresponds to the case where the variance of the efficiency term $(\frac{\pi-2}{\pi}\sigma_u^2)$ is significantly smaller than that of the noise in the production frontier (σ_v^2). The result is generally consistent with the fact that asymptotically the PL estimator reaches a semiparametric efficiency bound, while PI does not.

Overall our simulations seem to indicate that our proposed estimator can outperform the estimator proposed by Fan et al. (1996) in finite samples.

5 An empirical application

In this section we provide an empirical application of the semiparametric profile likelihood and plug-in estimation (PL and PI) using data on the U.S. electricity industry. The data have been used by Christensen and Green (1976), Gijbels et al. (1999), Martins-Filho and Yao (2008) and are provided in Green (1990). The model fitted in Green (1990) is a restricted specification of the cost function,

$$Ln(Cost/P_f) = \beta_0 + \beta_1 LnQ + \beta_2 Ln^2 Q + \beta_3 Ln(P_l/P_f) + \beta_4 Ln(P_k/P_f) + \epsilon. \tag{14}$$

The output (Q) is a function of three factors, labor (l), capital (k), and fuel (f). The three factor prices are P_l, P_k, and P_f. The restriction of linear homogeneity in the factor price has been imposed on the cost function. For detailed description of the data set and analysis, see Christensen and Greene (1976) and Greene(1990). Since we estimate a cost frontier rather than a production frontier, equation (1) is slightly modified and written as

$$f_{y|x}(y;\sigma_u^2,\sigma_v^2,g(x)) \equiv f_\varepsilon(y-g(x);\sigma_u^2,\sigma_v^2) = \frac{2}{\sqrt{\sigma_u^2+\sigma_v^2}}\phi\left(\frac{y-g(x)}{\sqrt{\sigma_u^2+\sigma_v^2}}\right)\left(1-\Phi\left(-\frac{\sqrt{\sigma_u^2/\sigma_v^2}}{\sqrt{\sigma_u^2+\sigma_v^2}}(y-g(x))\right)\right).$$

Since the parametric specification of the cost frontier might be restrictive, we utilize the semiparametric PL and PI approaches to estimate the frontier and analyze the efficiency levels of firms in the electric utility industry. We implement the PL estimator $(\tilde{\sigma}_u^2, \tilde{\sigma}_v^2)$ as described in section 4 using a gaussian product kernel and a bandwidth given by $h_l = c\sigma_{x_l} n^{-\frac{1}{4+p}}$ with $l = 1, 2, 3$ (see Härdle (1990) and Fan et al. (1996)), where c is a constant set to be 1.25, and σ_{x_l} is the sample standard deviation of x_l, $x_1 = LnQ$, $x_2 = Ln(P_l/P_f)$, $x_3 = Ln(P_k/P_f)$ and $n = 123$ is the sample size. We implement the PI estimator $(\hat{\sigma}_u^2, \hat{\sigma}_v^2)$ of Fan et al. (1996) using a local linear estimator for the conditional mean.

The PL estimation results give $\tilde{\sigma}_u^2 = 0.0010$, $\tilde{\sigma}_v^2 = 0.0103$ with $\tilde{\sigma}_u^2$ accounting for only 3.37% of the estimated conditional variance of $Ln(Cost/P_f)$. The PI estimation gives even smaller variance estimate for the one-sided error with $\hat{\sigma}_u^2 = 0.00001$ and $\hat{\sigma}_v^2 = 0.0108$. In contrast, as provided in Greene (1990) where the linear cost frontier and normal half-normal composite error model is fitted with maximum likelihood (ML) estimation $(\check{\sigma}_u^2, \check{\sigma}_v^2)$, the estimates are $\check{\sigma}_u^2 = 0.0241$, $\check{\sigma}_v^2 = 0.0115$ with $\check{\sigma}_u^2$ accounting for 43.2% of the estimated conditional variance of $Ln(Cost/P_f)$. The changes in the estimated parameters and changes in the allocation of total variance of the disturbance to the inefficiency term, similar in the direction but smaller in magnitude, are also observed in Fan et al. for Quebec dairy farm data and Green (1990) for gamma-distributed frontier model. Relatively small estimates of the one-sided components of the disturbance are also obtained in the empirical examples in Aigner et al. Given that the signal-to-noise ratio or the efficiency scores from the semiparametric estimations are fairly small, one might wonder whether the residuals display skewness in the "right" direction as discussed in Simar and Wilson (2010) and Almanidis and Sickles (2011), though they argue that the "wrong" skewness (negative skewness in the case of estimating cost frontier) is either a finite sample problem or associated with potential mis-specifications in the efficiency distribution. The empirical skewness of PL, PI and ML residuals are 0.0005, -0.0002, and 0.0105 respectively, indicating the residuals from PL and ML exhibit skewness in the "right" directions. We observe that the semiparametric estimation gives estimation results quite different from the parametric approach, suggesting that the cost frontier may be nonlinear in x_l.

We plot the cost frontier estimated by PL, PI and ML against the observed cost in Figure 2. The semiparametric estimates seem to be slightly more concentrated around the line of equality between estimated frontier and observed cost than the ML estimates, although we do not observe a substantial difference between the quality of fit using these approaches. To compare the difference in these three estimates, we provide the plot of estimated marginal cost against the output in Figure 3. The ML procedure assumes a parametric frontier that implies a marginal cost given by $\beta_1 + 2\beta_2 LnQ$. We obtain the marginal cost under PL by maximizing the local likelihood function in equation (12) at different sample values of $Ln(Q)$ with $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ reported above, fixing $Ln(P_l/P_f)$ and $Ln(P_k/P_f)$ at the sample mean values. The marginal cost under PI is obtained in an analogous fashion. The marginal costs obtained with PL and PI are quite similar in magnitudes, but they differ substantially from those of ML over the range of $Ln(Q)$, indicating that the assumption that marginal cost is linear in $Ln(Q)$ might be too restrictive.

Based on the frontier estimation results, we evaluate firm specific efficiency levels. We follow Jondrow et al. (1982) and obtain firm-specific efficiency as

$$e_i^f = \frac{(\sigma_u^2 + \sigma_v^2)^{\frac{1}{2}} \sqrt{\sigma_u^2/\sigma_v^2}}{1 + \sigma_u^2/\sigma_v^2} \left[\frac{\phi(-\frac{\sqrt{\sigma_u^2/\sigma_v^2}}{\sqrt{\sigma_u^2+\sigma_v^2}}\epsilon_i)}{1 - \Phi(-\frac{\sqrt{\sigma_u^2/\sigma_v^2}}{\sqrt{\sigma_u^2+\sigma_v^2}}\epsilon_i)} + \frac{\sqrt{\sigma_u^2/\sigma_v^2}}{\sqrt{\sigma_u^2 + \sigma_v^2}}\epsilon_i \right].$$

The unknown parameters are replaced with their estimates and ϵ_i is replaced with $\hat{\epsilon}_i$, the difference between y_i and the estimated frontier evaluated at x_i. Since the data are presented in logs, the efficiencies are calculated as $exp(\hat{e}_i^f)$.

Estimated efficiencies based on PL fall between 1.0171 and 1.0370 and the average efficiency score is 1.0251. Thus, on average, the cost of the U.S. electricity utility industry is increased by 2.5% due to inefficiency. The average efficiency score is roughly the same as that obtained in a COLS/gamma estimate provided in Table 2 of Green (1990). With a much smaller estimate of σ_u^2, estimated efficiencies of PI are between 1.0024 and 1.0026, and the average efficiency score is 1.0025. In contrast, the estimated efficiencies with ML range from 1.0308 to 1.4794 with an average efficiency score of 1.1338. Both PL and ML's estimated efficiency densities are skewed to the right, while PI's density seems to be skewed slightly to the left. Figure 4 shows the plot of estimated efficiencies with the three approaches against the observed cost. It seems that ML predicts much higher inefficiencies than the semiparametric PL and PI approaches. This is consistent with the general observations made in Kumbhakar et al. (2007) using a local maximum likelihood approach. The high estimated inefficiency might be attributed to a misspecification of the frontier function.

6 Summary and conclusions

In this paper we consider the estimation of a semiparametric stochastic frontier model. We study two estimators for the parameters of the model. We first establish the asymptotic properties (until now unknown) of an estimator proposed by Fan et al. (1996). The estimator is shown to be consistent and asymptotically normal, however the asymptotic distribution is incorrectly centered. We then propose a new estimator based on a profile likelihood procedure for conditionally parametric models first suggested by Severini and Wong (1992). We show that our estimator is consistent, asymptotically normal and efficient in a suitably defined class of semiparametric estimators. Practical use of the estimators requires the specification of a conditional density that must meet some regularity conditions. We verify that the density used in Aigner et al. (1977) and Fan et al. (1996) satisfy all of the stated regularity conditions. However, future work should investigate whether broader classes of densities that may potentially be used by applied researchers in efficiency and productivity studies meet such regularity conditions.

Appendix 1 - Assumptions and proofs

ASSUMPTION PB: 1. $\sup_{x\in G} f_x(x) < C$; 2. $\sup_{x\in G} f_x^{(1)}(x) < C$; 3. $\sup_{x\in G} f_x^{(2)}(x) < C$; 4. $\inf_{x\in G} f_x(x) > 0$; 5. $\sup_{x\in G,\theta\in\Theta} |\frac{\partial}{\partial x}\alpha_\theta(x)| < C$.

ASSUMPTION PC: $\int \sup_{x\in G} \left| \frac{\partial}{\partial x}\left(\frac{\partial^{s_1+s_2+r}}{\partial\theta_k^{s_1}\partial\theta_j^{s_2}\partial\eta^r} log f_\varepsilon(y-\alpha_\theta(x);\theta) f_\varepsilon(y-g_0(x);\theta)\right)\right| dy < c$, with $s_1, s_2, r \geq 0$, where $(s_1, s_2, r) = (0,0,1), (1,0,1), (0,1,1), (1,1,1)$.

ASSUMPTION PD: $\sup_{x\in G,\theta\in\Theta,\eta\in\mathcal{H}} \left| \int \frac{\partial^{s_1+s_2+r}}{\partial\theta_k^{s_1}\partial\theta_j^{s_2}\partial\eta^r} log f_\varepsilon(y-\eta;\theta) \frac{\partial^j}{\partial x^j} f_{y|x}(y-g_0(x);\theta_0)dy\right| < C$ with $s_1, s_2, r \geq 0$, for j=0, $s_1+s_2+r \leq 4$, and (s_1, s_2, r) = (0,0,5); for j=1, (s_1, s_2, r) = (0,0,1), (1,0,2), (0,1,2), (1,0,3), (0,1,3), (0,0,4), (0,0,3), (1,1,2), (1,1,3), (0,1,4), (1,0,4), (0,0,5) and for j=2, (s_1, s_2, r) = (0,0,1), (0,0,2), (0,0,3), (1,0,1), (0,1,1), (1,0,2), (0,1,2), (1,0,3), (0,1,3), (0,0,4), (1,1,1), (1,1,2), (1,1,3), (0,1,4), (1,0,4), (0,0,5).

Lemma 2: *Proof.* a) For fixed x and θ we define $(\hat{\eta}_0, \hat{\eta}_1) = \underset{\eta_0,\eta_1}{\text{argmax}} \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n} log f_\varepsilon(y_i - \eta_0 - \eta_1(x_i - x);\theta) K\left(\frac{x_i-x}{h_n}\right)$ where $\hat{\eta}_0 = \hat{\alpha}_\theta(x)$, $\hat{\eta}_1 = \frac{\partial}{\partial x}\hat{\alpha}_\theta(x)$ satisfy first order conditions. Put $\hat{d}_{0,\theta}(x) = \hat{\alpha}_\theta(x) - \alpha_\theta(x)$, $\hat{d}_{1,\theta}(x) = \frac{\partial}{\partial x}\hat{\alpha}_\theta(x) - \frac{\partial}{\partial x}\alpha_\theta(x)$ and $\eta_\theta(x, x_i) = \alpha_\theta(x) + \frac{\partial}{\partial x}\alpha_\theta(x)(x_i - x)$. Let $\hat{d}_\theta(x) = \begin{pmatrix} \hat{d}_{0,\theta}(x) \\ \hat{d}_{1,\theta}(x) \end{pmatrix}$ and write

$$G_{n\theta}(\hat{d}_\theta(x)) = \begin{pmatrix} G_{n\theta_0}(\hat{d}_\theta(x)) \\ G_{n\theta_1}(\hat{d}_\theta(x)) \end{pmatrix} = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n}\frac{\partial}{\partial\eta} log f_\varepsilon(y_i - \eta_\theta(x, x_i) - \hat{d}_{0,\theta}(x) - \hat{d}_{1,\theta}(x)(x_i - x);\theta) K\left(\frac{x_i - x}{h_n}\right) \times \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Letting $z' = (0,0)$ we have by Taylor's Theorem that for $d^*_{j,\theta}(x) = (1-\lambda_j)\hat{d}_{j,\theta}(x)$ and $j = 0,1$ we have $G_{n\theta}(\hat{d}_\theta(x)) = G_{n\theta}(z) + H_{n\theta}(d^*_\theta(x))\hat{d}_\theta(x) = 0$ where

$$H_{n\theta}(d^*_\theta(x)) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n}\frac{\partial^2}{\partial\eta^2} log f_\varepsilon(y_i - \eta_\theta(x, x_i) - d^*_{0,\theta}(x) - d^*_{1,\theta}(x)(x_i - x);\theta) K\left(\frac{x_i - x}{h_n}\right)\begin{pmatrix} 1 & x_i - x \\ x_i - x & (x_i - x)^2 \end{pmatrix}.$$

We now define $s_{jn}(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n}\frac{\partial^2}{\partial\eta^2} log f_\varepsilon(y_i - \eta_\theta(x, x_i) - d^*_{0,\theta}(x) - d^*_{1,\theta}(x)(x_i - x);\theta) K\left(\frac{x_i-x}{h_n}\right)(x_i - x)^j$ for $j = 0,1,2$. We note that under our assumptions $\alpha_\theta(x)$ is an unique maximum and satisfies $\frac{\partial}{\partial\eta} E\left(log f_\varepsilon(y_i - \alpha_\theta(x);\theta)|x\right) = 0$. Given that we can interchange the partial derivative with the expectation $\frac{\partial}{\partial\eta} E\left(log f_\varepsilon(y_i - \alpha_\theta(x);\theta)|x\right) = E\left(\frac{\partial}{\partial\eta} log f_\varepsilon(y_i - \alpha_\theta(x);\theta)|x\right) = 0$. Hence, for some $\bar{d}_\theta(x) = \begin{pmatrix} \bar{d}_{0,\theta}(x) \\ \bar{d}_{1,\theta}(x) \end{pmatrix} \in \Re^2$ we have from the definition of G_θ that $G_\theta(\bar{d}_\theta(x)) = 0$ will have a unique solution at $\bar{d}_\theta(x) = 0$. Now, we have assumed that for all $\epsilon > 0$ there exists a $\delta > 0$ such that $|| \sup_{\theta\in\Theta,x\in G} |G_\theta(\bar{d}_\theta(x))| ||_E \leq \delta$ implies $\sup_{\theta\in\Theta,x\in G} |\bar{d}_{i,\theta}(x)| \leq \epsilon$ for $i = 0,1$. Now, $P(\sup_{\theta\in\Theta,x\in G} |\hat{d}_{0,\theta}(x)| > \epsilon) \leq$

$P(||\sup_{\theta\in\Theta,x\in G}|G_\theta(\hat{d}_\theta(x))|||_E > \epsilon) = P(||\sup_{\theta\in\Theta,x\in G}|G_\theta(\hat{d}_\theta(x)) - G_{n\theta}(\hat{d}_\theta(x))|||_E > \epsilon)$ since $G_{n\theta}(\hat{d}_\theta(x)) = 0$. By the c_r inequality we have

$$\begin{aligned} P(\sup_{\theta\in\Theta,x\in G}|\hat{\alpha}_\theta(x) - \alpha_\theta(x)| > \epsilon) &\leq P(\sup_{\theta\in\Theta,x\in G}|G_{\theta,0}(\hat{d}_\theta(x)) - G_{n\theta_0}(\hat{d}_\theta(x))| > \epsilon/2) \\ &+ P(\sup_{\theta\in\Theta,x\in G}|G_{\theta,1}(\hat{d}_\theta(x)) - G_{n\theta_1}(\hat{d}_\theta(x))| > \epsilon/2) \end{aligned}$$

and we now show that $\sup_{\theta\in\Theta,x\in G}|G_{\theta,j}(\hat{d}_\theta(x)) - G_{n\theta_j}(\hat{d}_\theta(x))| = o_p(1)$ for $j = 0, 1$.

$$\begin{aligned} \sup_{\theta\in\Theta,x\in G}|G_{\theta,j}(\hat{d}_\theta(x)) - G_{n\theta_j}(\hat{d}_\theta(x))| &\leq \sup_{\theta\in\Theta,x\in G}|G_{n\theta_j}(\hat{d}_\theta(x)) - E(G_{n\theta_j}(\hat{d}_\theta(x)))| \\ &+ \sup_{\theta\in\Theta,x\in G}|E(G_{n\theta_j}(\hat{d}_\theta(x))) - G_{\theta,j}(\hat{d}_\theta(x))| = I_1 + I_2. \end{aligned}$$

Now, observe that $I_1 \leq \sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}\left|\frac{1}{n}\sum_{i=1}^n\left(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta;\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j - E\left(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta;\theta)\right.\right.\right.$ $\left.\left.\left.\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j\right)\right)\right| = o_p(1)$ by Lemma 1. Now, we can write

$$\begin{aligned} I_2 = \sup_{\theta\in\Theta,x\in G}&\left|\frac{1}{n}\sum_{i=1}^n\left(E\left(\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \eta_\theta(x,x_i) - \hat{d}_{0,\theta}(x) - \hat{d}_{1,\theta}(x)(x_i-x);\theta)K\left(\frac{x_i-x}{h_n}\right)\right.\right.\right. \\ \times\ & (x_i-x)^j) - f_x(x)\int_G E(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \alpha_\theta(x) - \hat{d}_{0,\theta}(x) - \hat{d}_{1,\theta}(x)(x_i-x);\theta)|x) \\ \times\ & \left.\left.\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j dx_i\right)\right| \end{aligned}$$

where $\hat{d}_{0,\theta}(x) + \hat{d}_{1,\theta}(x)(x_i-x) \in \mathcal{H}_1$ a compact subset of $\Re$, and we immediately have

$$\begin{aligned} I_2 \leq \sup_{\theta\in\Theta,\eta\in\mathcal{H}_1,x\in G}&\left|E\left(\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \eta_\theta(x,x_i) - \eta;\theta)K\left(\frac{x_i-x}{h_n}\right)\right.\right. \\ \times\ & (x_i-x)^j) - f_x(x)\int_G E(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \alpha_\theta(x) - \eta;\theta)|x) \\ \times\ & \left.\left.\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j dx_i\right)\right| = \sup_{\theta\in\Theta,\eta\in\mathcal{H}_1,x\in G}|I_{21} - I_{22}|. \end{aligned}$$

Using the fact that $f_\epsilon(y_i-g_0(x_i);\theta_0) = f_\epsilon(y_i-g_0(x);\theta_0)+\frac{\partial}{\partial x}f_\epsilon(y_i-g_0(x);\theta_0)(x_i-x)+\frac{\partial^2}{\partial x^2}f_\epsilon(y_i-g_0(x^*);\theta_0)$ $(x_i-x)^2$ for $x^* \in L(x_i, x)$ we can write

$$\begin{aligned} I_{21} &= E\left(\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \eta_\theta(x,x_i) - \eta;\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j\right) \\ &+ E\left(\int\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \eta_\theta(x,x_i) - \eta;\theta)\frac{\partial}{\partial x}f_\varepsilon(y_i - g_0(x);\theta_0)dy_iK\left(\frac{x_i-x}{h_n}\right)(x_i-x)^{j+1}\right) \\ &+ E\left(\int\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \eta_\theta(x,x_i) - \eta;\theta)\frac{\partial^2}{\partial x^2}f_\varepsilon(y_i - g_0(x);\theta_0)dy_iK\left(\frac{x_i-x}{h_n}\right)(x_i-x)^{j+2}\right) \\ &= I_{211} + I_{212} + I_{213}. \end{aligned}$$

Given that $\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i)-\eta;\theta) = \frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\alpha_\theta(x)-\eta;\theta)+\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x)-\eta;\theta)\frac{\partial}{\partial x}\alpha_\theta(x)(x_i-x) + \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i - \eta^*_\theta(x) - \eta;\theta)\left(\frac{\partial}{\partial x}\alpha_\theta(x)\right)^2(x_i-x)^2$ where $\eta^*_\theta(x) \in L(\alpha_\theta(x), \eta_\theta(x,x_i))$ we can show that under regularity conditions PB and PD we have $\sup_{\theta\in\Theta,\eta\in\mathcal{H}_1,x\in G}|I_{211}-I_{22}|=O(h_n^2)$, $\sup_{\theta\in\Theta,\eta\in\mathcal{H}_1,x\in G}|I_{212}|=O(h_n^2)$ and $\sup_{\theta\in\Theta,\eta\in\mathcal{H}_1,x\in G}|I_{213}|=O(h_n^2)$. Therefore, $I_2\leq O(h_n^2)$ which combined with the order of I_1 gives $\sup_{\theta\in\Theta,x\in G}|G_{\theta,j}(\hat{d}_\theta(x))-G_{n\theta_j}(\hat{d}_\theta(x))|=o_p(1)$ for $j=0$. The case for $j=1$ can be treated in analogous manner. Similarly, we obtain the proof of $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)-\frac{\partial}{\partial x}\alpha_\theta(x)|=o_p(1)$.

b) Recall from part a) that $\hat{d}_\theta(x)=-H_{n\theta}(d^*_\theta(x))^{-1}G_{n\theta}(z)=-\begin{pmatrix} s_{0n} & s_{1n}\\ s_{1n} & s_{2n}\end{pmatrix}^{-1}G_{n\theta}(z)$. Note that by Taylor's Theorem we can write

$$
\begin{aligned}
s_{jn}(x)&=\frac{1}{n}\sum_{i=1}^n\frac{1}{h_n}\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j\\
&+\frac{1}{n}\sum_{i=1}^n\frac{1}{h_n}\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta^*_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^jd^*_{0,\theta}(x)\\
&+\frac{1}{n}\sum_{i=1}^n\frac{1}{h_n}\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta^*_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^{j+1}d^*_{1,\theta}(x)\\
&=I_1+I_2+I_3
\end{aligned}
$$

where $\eta^*_\theta(x,x_i)=\eta_\theta(x,x_i)+\lambda(d^*_{0,\theta}(x)+d^*_{1,\theta}(x)(x_i-x))$ for $\lambda\in[0,1]$. We now define,

$$
\begin{aligned}
s_0&=f_x(x)E\left(\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)|x\right)\\
s_1&=h_n^2\sigma_K^2f_x^{(1)}(x)E\left(\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)|x\right)\\
&+h_n^2\sigma_K^2f_x(x)\int\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)\frac{\partial}{\partial x}f_\varepsilon(y_i-\alpha_\theta(x);\theta)dy_i\\
&+h_n^2\sigma_K^2f_x(x)E\left(\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)|x\right)\frac{\partial}{\partial x}\alpha_\theta(x)=s_{11}+s_{12}+s_{13}\\
s_2&=h_n^2\sigma_K^2s_0.
\end{aligned}
$$

We will show that $I_2,\ I_3=o_p(1)$ and that I_1 converges to s_j uniformly in G and Θ. From the proof of part a) we know that $\underset{x\in G,\theta\in\Theta}{sup}|d^*_{0,\theta}(x)|$, $\underset{x\in G,\theta\in\Theta}{sup}|d^*_{1,\theta}(x)|=o_p(1)$. Hence, for $\underset{x\in G,\theta\in\Theta}{sup}|I_2|=o_p(1)$ we need to show that $I_{21}=\frac{1}{n}\sum_{i=1}^n\frac{1}{h_n}\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta^*_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j=O_p(1)$ uniformly in G and Θ.

$$
\begin{aligned}
\underset{x\in G,\theta\in\Theta}{sup}|I_{21}|\leq&\underset{x\in G,\theta\in\Theta,\eta\in\mathcal{H}}{sup}|\frac{1}{n}\sum_{i=1}^n\frac{1}{h_n}\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j\\
&-E(\frac{1}{h_n}\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)|\\
&+\underset{x\in G,\theta\in\Theta,\eta\in\mathcal{H}}{sup}|E(\frac{1}{h_n}\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)|=I_{211}+I_{212}.
\end{aligned}
$$

From Lemma 1, given PA2 and $nh_n^3 \to \infty$, we have that $I_{211} = O_p\left(h_n^j \left(\frac{log(n)}{nh_n}\right)^{1/2}\right)$. Now observe that

$$E(\frac{1}{h_n}\frac{\partial^3}{\partial\eta^3} log f_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j) =$$

$$\int\int \frac{1}{h_n}\frac{\partial^3}{\partial\eta^3} log f_\varepsilon(y_i-\eta;\theta) f_\varepsilon(y_i-g_0(x_i);\theta_0)dy_i K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j f_x(x_i)dx_i$$

and since $f_\varepsilon(y_i-g_0(x_i);\theta_0) = f_\varepsilon(y_i-g_0(x);\theta_0) + \frac{\partial}{\partial x} f_\varepsilon(y_i-g_0(x^*);\theta_0)(x_i-x)$ for $x^* \in L(x_i,x)$ and $f_x(x_i) = f_x(x) + f_x^{(1)}(x)(x_i-x) + (1/2)f_x^{(2)}(x^*)(x_i-x)^2$ for $x^* \in L(x_i,x)$ we can write

$$\begin{aligned} I_{212} &\leq \sup_{x\in G,\theta\in\Theta,\eta\in\mathcal{H}} \left|\int\int \frac{1}{h_n}\frac{\partial^3}{\partial\eta^3} log f_\varepsilon(y_i-\eta);\theta) f_\varepsilon(y_i-g_0(x_i);\theta_0)dy_i K\left(\frac{x_i-x}{h_n}\right)\right. \\ &\times \left.(f_x(x)(x_i-x)^j + f_x^{(1)}(x)(x_i-x)^{j+1} + (1/2)f_x^{(2)}(x^*)(x_i-x)^{j+2})dx_i\right| \\ &+ \sup_{x\in G,\theta\in\Theta,\eta\in\mathcal{H}} \left|\int\int \frac{1}{h_n}\frac{\partial^3}{\partial\eta^3} log f_\varepsilon(y_i-\eta;\theta)\frac{\partial}{\partial x} f_\varepsilon(y_i-g_0(x^*);\theta_0)dy_i K\left(\frac{x_i-x}{h_n}\right)\right. \\ &\times \left.(f_x(x)(x_i-x)^{j+1} + f_x^{(1)}(x)(x_i-x)^{j+2} + (1/2)f_x^{(2)}(x^*)(x_i-x)^{j+3})dx_i\right| \\ &= I_{2121} + I_{2122} \end{aligned}$$

Since the kernel K is a bounded function with compact support and given regularity conditions PB and PD, we have that for $j=0$, $I_{2121} \leq \sup_{x\in G,\theta\in\Theta,\eta\in\mathcal{H}} |E(\frac{\partial^3}{\partial\eta^3} log f_\varepsilon(y_i-\eta;\theta)|x)|(\sup_{x\in G} f_x(x) + h_n^2\sigma_K^2 \sup_{x\in G}|f_x^{(2)}(x)|) = O(1)$, for $j=1$, $I_{2121} \leq \sup_{x\in G,\theta\in\Theta,\eta\in\mathcal{H}} |E(\frac{\partial^3}{\partial\eta^3} log f_\varepsilon(y_i-\eta;\theta)|x)|(h_n^2\sigma_K^2 \sup_{x\in G}|f_x^{(1)}(x)| + Ch_n^3 \sup_{x\in G}|f_x^{(2)}(x)|) = O(h_n^2)$, and for $j=2$, $I_{2121} \leq \sup_{x\in G,\theta\in\Theta,\eta\in\mathcal{H}} |E(\frac{\partial^3}{\partial\eta^3} log f_\varepsilon(y_i-\eta;\theta)|x)|(h_n^2\sigma_K^2 \sup_{x\in G}|f_x(x)| + Ch_n^3 \sup_{x\in G}|f_x^{(1)}(x)| + Ch_n^4 \sup_{x\in G}|f_x^{(2)}(x)|) = O(h_n^2)$ Using similar arguments we can establish $I_{2122} = O(h_n)$ if $j=0$, $I_{2122} = O(h_n^2)$ if $j=1$ and $I_{2122} = O(h_n^3)$ if $j=2$. Combining the orders of I_{2121}, I_{2122} and I_{211} we have $\sup_{x\in G,\theta\in\Theta} |I_{21}| =$

$$\begin{cases} O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O(1) \text{ for } j=0 \\ O_p\left(h_n\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O(h_n^2) \text{ for } j=1 \\ O_p\left(h_n^2\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O(h_n^2) \text{ for } j=2 \end{cases} \text{ and } \sup_{x\in G,\theta\in\Theta}|I_2| = \begin{cases} o_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + o(1) \text{ for } j=0 \\ o_p\left(h_n^j\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + o(h_n^2) \text{ for } j=1,2. \end{cases}$$

Following analogous arguments and manipulations we obtain,

$$\sup_{x\in G,\theta\in\Theta}|I_3| = \begin{cases} o_p\left(h_n\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + o(h_n^2) \text{ for } j=0 \\ o_p\left(h_n^{j+1}\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + o(h_n^{j+1}) \text{ for } j=1,2. \end{cases}$$

We now focus on I_1. Note that

$$\begin{aligned} I_1 - s_j &= \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{h_n}\frac{\partial^2}{\partial \eta^2} log f_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j \\ &- E(\frac{1}{h_n}\frac{\partial^2}{\partial \eta^2} log f_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)) \\ &+ E(\frac{1}{h_n}\frac{\partial^2}{\partial \eta^2} log f_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)) - s_j \\ &= I_{11} + I_{12} - s_j \end{aligned}$$

and $\sup_{x\in G,\theta\in\Theta}|I_{11}| \leq O_p\left(h_n^j\left(\frac{log(n)}{nh_n}\right)^{1/2}\right)$ by Lemma 1, given PA2 and $nh_n^3 \to \infty$. By use of the expansions $f_x(x_i) = f_x(x) + f_x^{(1)}(x)(x_i-x) + (1/2)f_x^{(2)}(x^*)(x_i-x)^2$ for $x^* \in L(x_i,x)$,

$$f_\varepsilon(y_i-g_0(x_i);\theta_0) = f_\varepsilon(y_i-g_0(x);\theta_0) + \frac{\partial}{\partial x}f_\varepsilon(y_i-g_0(x);\theta_0)(x_i-x) + (1/2)\frac{\partial^2}{\partial x^2}f_\varepsilon(y_i-g_0(x^*);\theta_0)(x_i-x)^2$$

for $x^* \in L(x_i,x)$ and $\frac{\partial^2}{\partial\eta^2}log f_\varepsilon(y_i-\eta_\theta(x,x_i);\theta) = \frac{\partial^2}{\partial\eta^2}log f_\varepsilon(y_i-\alpha_\theta(x);\theta) + \frac{\partial^3}{\partial\eta^3}log f_\varepsilon(y_i-\alpha_\theta(x);\theta)\frac{\partial}{\partial x}\alpha_\theta(x)$ $(x_i-x) + \frac{\partial^4}{\partial\eta^4}log f_\varepsilon(y_i-\eta_\theta^*(x);\theta)\left(\frac{\partial}{\partial x}\alpha_\theta(x)\right)^2(x_i-x)^2$ for $\eta_\theta^*(x) \in L(\alpha_\theta(x),\eta_\theta(x,x_i))$ together with regularity conditions PB and PD and the fact that the kernel K is a bounded function on a compact support gives $\sup_{x\in G,\theta\in\Theta}|I_1-s_j| = \begin{cases} O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O(h_n^2) \text{ for } j=0 \\ O_p\left(h_n^j\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + O(h_n^3) \text{ for } j=1,2. \end{cases}$ In all, combining the results for I_1, I_2 and I_3 we have that $\sup_{x\in G,\theta\in\Theta}|s_{jn}-s_j| = \begin{cases} O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + o(1) \text{ for } j=0 \\ O_p\left(h_n\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + o(h_n^2) \text{ for } j=1 \\ O_p\left(h_n^2\left(\frac{log(n)}{nh_n}\right)^{1/2}\right) + o(h_n^2) \text{ for } j=2 \end{cases}$ We now write $\hat{d}_\theta(x) = -\left(\begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n}\end{pmatrix}^{-1} - \begin{pmatrix} s_0 & s_1 \\ s_1 & s_2\end{pmatrix}^{-1}\right)G_{n\theta}(z) - \begin{pmatrix} s_0 & s_1 \\ s_1 & s_2\end{pmatrix}^{-1}G_{n\theta}(z)$. We will show that $\sup_{\theta\in\Theta,x\in G}\hat{d}_\theta(x)$ has the stated order in probability. To that end note that simple algebra manipulations reveal that the existence of $\begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n}\end{pmatrix}^{-1} - \begin{pmatrix} s_0 & s_1 \\ s_1 & s_2\end{pmatrix}^{-1}$ depends on $\liminf_n \inf_{\theta\in\Theta}\inf_{x\in G}|h_n^{-4}(s_{0n}s_{2n}-s_{1n}^2)(s_0s_2-s_1^2)| > 0$ in probability. Given the order in probability results we have obtained for $\sup_{x\in G,\theta\in\Theta}|s_{jn}-s_j|$ we write $h_n^{-4}(s_{0n}s_{2n}-s_{1n}^2)(s_0s_2-s_1^2) = \left(\frac{1}{h_n^2}s_0s_2 + o(1)\right)^2$ and check that $\inf_{x\in G,\theta\in\Theta}|s_0| > 0$ and $\inf_{x\in G,\theta\in\Theta}\frac{1}{h_n^2}|s_2| > 0$ with $\inf_{x\in G,\theta\in\Theta}f_x(x) >$ from PB. If, for fixed x and θ, $E(log f_\varepsilon(y_i-\eta;\theta)|x)$ has a unique maximum at η_0 which satisfies $\frac{\partial}{\partial\eta}E(log f_\varepsilon(y_i-\eta_0;\theta)|x) = 0$ and $\frac{\partial^2}{\partial\eta^2}E(log f_\varepsilon(y_i-\eta_0;\theta)|x) < 0$, then we know from the implicit function theorem that if $\frac{\partial}{\partial\eta}E(log f_\varepsilon(y_i-\eta;\theta)|x)$ is continuous on a neighborhood of x, θ and η_0 and $|\frac{\partial}{\partial\eta}(\frac{\partial}{\partial\eta}E(log f_\varepsilon(y_i-\eta_0;\theta)|x))| \neq 0$, there exists a unique $\alpha_\theta(x)$ in the neighborhood of η_0 such that $\frac{\partial}{\partial\eta}E(log f_\varepsilon(y_i-\alpha_\theta(x);\theta)|x) = 0$ and $-\frac{\partial^2}{\partial\eta^2}E(log f_\varepsilon(y_i-\alpha_\theta(x);\theta)|x) > 0$ for all x and

θ. Hence, we have that $\inf_{x\in G,\theta\in\Theta}|s_0| \geq \inf_{x\in G,\theta\in\Theta} f_x(x) \inf_{x\in G,\theta\in\Theta} -\frac{\partial^2}{\partial\eta^2}E(logf_\varepsilon(y_i-\alpha_\theta(x);\theta)|x) > 0$ provided that $\frac{\partial^2}{\partial\eta^2}E(logf_\varepsilon(y_i-\alpha_\theta(x);\theta)|x) = E(\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)|x)$, which is verified given that $\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)$ is continuously differentiable on an open set of $\alpha_\theta(x)$ and $E(\sup_{\eta\in\mathcal{H}}|\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)|$ $|x) < \infty$. In a similar fashion we verify that $\inf_{x\in G,\theta\in\Theta}\frac{1}{h_n^2}|s_2| > 0$. Thus, we obtain as a direct consequence of the order in probability of $\sup_{x\in G,\theta\in\Theta}|s_{jn}-s_j|$ and the preceding discussion that $\begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n}\end{pmatrix}^{-1} - \begin{pmatrix} s_0 & s_1 \\ s_1 & s_2\end{pmatrix}^{-1} = \begin{pmatrix} o_p(1) & o_p(1) \\ o_p(1) & o_p(h_n^{-2})\end{pmatrix}$ and $\begin{pmatrix} s_0 & s_1 \\ s_1 & s_2\end{pmatrix}^{-1} = \begin{pmatrix} O(1) & O(1) \\ O(1) & O(h_n^{-2})\end{pmatrix}$ uniformly in Θ and G.

We now show that each element in $G_{n\theta}(z)$ given by

$$I = \sup_{\theta\in\Theta,x\in G}|\frac{1}{n}\sum_{i=1}^n \frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j| = O_p(h_n^j(log(n)/nh_n)^{1/2})+O(h_n^{2+j})$$

for $j=0,1$. Since, by Lemma 1

$$\begin{aligned} I &\leq \sup_{\theta\in\Theta,x\in G}|\frac{1}{n}\sum_{i=1}^n \frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j \\ &- E(\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)| \\ &+ \sup_{\theta\in\Theta,x\in G}|E(\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)| \\ &= O_p(h_n^j(log(n)/nh_n)^{1/2}) + \sup_{\theta\in\Theta,x\in G}|E(\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)| \end{aligned}$$

it suffices to establish that $I_1 = \sup_{\theta\in\Theta,x\in G}|E(\frac{1}{h_n}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)| = O(h_n^{2+j})$. Once again, we use Taylor's theorem to expand $f_\varepsilon(y_i-g_0(x_i);\theta_0)$ around x and $\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)$ around $\alpha_\theta(x)$ and write for some $\eta_\theta^*(x)\in L(\eta_\theta(x,x_i),\alpha_\theta(x))$ and $x^*\in L(x_i,x)$

$$\begin{aligned} I_1 &= \sup_{\theta\in\Theta,x\in G}|\int\int(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)+\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)\frac{\partial}{\partial x}\alpha_\theta(x)(x_i-x) \\ &+ \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta^*(x);\theta)(\frac{\partial}{\partial x}\alpha_\theta(x))^2(x_i-x)^2)(f_\varepsilon(y_i-g_0(x);\theta_0)+\frac{\partial}{\partial x}f_\varepsilon(y_i-g_0(x);\theta_0)(x_i-x) \\ &+ \frac{\partial^2}{\partial x^2}f_\varepsilon(y_i-g_0(x^*);\theta_0)(x_i-x)^2)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}f_x(x_i)(x_i-x)^j dx_i| \\ &= \sup_{\theta\in\Theta,x\in G}|I_{11}+\cdots+I_{19}|. \end{aligned}$$

We investigate each of these terms separately:

$$\begin{aligned} \sup_{\theta\in\Theta,x\in G}|I_{11}| &= \sup_{\theta\in\Theta,x\in G}|\int\int\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)f_\varepsilon(y_i-g_0(x);\theta_0)dy_iK\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}f_x(x_i) \\ &\times (x_i-x)^j dx_i| = 0 \end{aligned}$$

since $E(\frac{\partial}{\partial \eta} logf_\varepsilon(y_i - \alpha_\theta(x);\theta)|x) =$ for all θ and x;

$$\begin{aligned}\sup_{\theta\in\Theta, x\in G}|I_{12}+I_{13}| &= \sup_{\theta\in\Theta,x\in G}|\int\int (f_\varepsilon(y_i-g_0(x);\theta_0)\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)\frac{\partial}{\partial x}\alpha_\theta(x)(x_i-x) \\ &+ \frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)\frac{\partial}{\partial x}f_\varepsilon(y_i-g_0(x);\theta_0)dy_i(x_i-x))K\left(\frac{x_i-x}{h_n}\right) \\ &\times \frac{1}{h_n}f_x(x_i)(x_i-x)^j dx_i| = 0\end{aligned}$$

given regularity condition PD and the fact that $\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)f_\varepsilon(y_i-g_0(x);\theta_0)$ is continuously differentiable at x. Given regularity conditions PB and PD we show that

$$\begin{aligned}\sup_{\theta\in\Theta, x\in G}|I_{14}| &= \sup_{\theta\in\Theta,x\in G}|\int\int \frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)\frac{\partial}{\partial x}\alpha_\theta(x)\frac{\partial}{\partial x}f_\varepsilon(y_i-g_0(x);\theta_0)dy_iK\left(\frac{x_i-x}{h_n}\right) \\ &\times \frac{1}{h_n}f_x(x_i)(x_i-x)^{j+2}dx_i| = O(h_n^{2+j}),\end{aligned}$$

$$\begin{aligned}\sup_{\theta\in\Theta, x\in G}|I_{15}| &= \sup_{\theta\in\Theta,x\in G}|\int\int \frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\alpha_\theta(x);\theta)\frac{\partial}{\partial x}\alpha_\theta(x)\frac{\partial^2}{\partial x^2}f_\varepsilon(y_i-g_0(x^*);\theta_0)dy_i(x_i-x)^{j+3} \\ &\times K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}f_x(x_i)dx_i| = O(h_n^{3+j}),\end{aligned}$$

$$\begin{aligned}\sup_{\theta\in\Theta, x\in G}|I_{16}| &= \sup_{\theta\in\Theta,x\in G}|\int\int \frac{\partial}{\partial\eta}logf_\varepsilon(y_i;\theta,\alpha_\theta(x))\frac{\partial^2}{\partial x^2}f_\varepsilon(y_i-g_0(x^*);\theta_0)dy_i(x_i-x)^{j+2} \\ &\times K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}f_x(x_i)dx_i| = O(h_n^{2+j}),\end{aligned}$$

$$\begin{aligned}\sup_{\theta\in\Theta, x\in G}|I_{17}| &= \sup_{\theta\in\Theta,x\in G}|\int\int \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta^*(x);\theta)(\frac{\partial}{\partial x}\alpha_\theta(x))^2f_\varepsilon(y_i-g_0(x);\theta_0)dy_i \\ &\times K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}f_x(x_i)(x_i-x)^{2+j}dx_i| = O(h_n^{2+j}),\end{aligned}$$

$$\begin{aligned}\sup_{\theta\in\Theta, x\in G}|I_{18}| &= \sup_{\theta\in\Theta,x\in G}|\int\int \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta^*(x);\theta)(\frac{\partial}{\partial x}\alpha_\theta(x))^2\frac{\partial}{\partial x}f_\varepsilon(y_i-g_0(x);\theta_0)dy_i \\ &\times K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}f_x(x_i)(x_i-x)^{3+j}dx_i| = O(h_n^{3+j}),\end{aligned}$$

$$\begin{aligned}\sup_{\theta\in\Theta, x\in G}|I_{19}| &= \sup_{\theta\in\Theta,x\in G}|\int\int \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta^*(x);\theta)(\frac{\partial}{\partial x}\alpha_\theta(x))^2\frac{\partial^2}{\partial x^2}f_\varepsilon(y_i-g_0(x^*);\theta_0)dy_i \\ &\times K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}f_x(x_i)(x_i-x)^{4+j}dx_i| = O(h_n^{4+j}).\end{aligned}$$

c) Since $\begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n}\end{pmatrix}\hat{d}_\theta(x) = G_{n\theta}(z)$ and taking a partial derivative $\frac{\partial}{\partial\theta_k}$ on both sides gives

$$\begin{pmatrix} \frac{\partial}{\partial\theta_k}s_{0n} & \frac{\partial}{\partial\theta_k}s_{1n} \\ \frac{\partial}{\partial\theta_k}s_{1n} & \frac{\partial}{\partial\theta_k}s_{2n}\end{pmatrix}\hat{d}_\theta(x) + \begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n}\end{pmatrix}\frac{\partial}{\partial\theta_k}\hat{d}_\theta(x) = \frac{\partial}{\partial\theta_k}G_{n\theta}(z)$$

and consequently $\frac{\partial}{\partial\theta_k}\hat{d}_\theta(x) = \begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n} \end{pmatrix}^{-1} \frac{\partial}{\partial\theta_k}G_{n\theta}(z) - \begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial}{\partial\theta_k}s_{0n} & \frac{\partial}{\partial\theta_k}s_{1n} \\ \frac{\partial}{\partial\theta_k}s_{1n} & \frac{\partial}{\partial\theta_k}s_{2n} \end{pmatrix} \hat{d}_\theta(x)$.
Given the results in parts a) and b), if (i) $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}s_{jn}| = O_p(h_n^j)$ for $j = 0, 1, 2$ and (ii)

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i;\theta,\eta_\theta(x,x_i)))K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j = O_p(h_n^j(log(n)/nh_n)^{1/2}) + O(h_n^{2+j})$$

uniformly in G and Θ for $j = 0, 1$ we have the desired order for $\frac{\partial}{\partial\theta_k}\hat{d}_\theta(x)$. We start by establishing (i), but for that purpose we obtain two related results, $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\hat{\alpha}_\theta(x)| = O_p(1)$ and $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)| = O_p(1)$. First, note that $\begin{pmatrix} \frac{\partial}{\partial\theta_k}\hat{\alpha}_\theta(x) \\ \frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\hat{\alpha}_\theta(x) \end{pmatrix} = -\begin{pmatrix} s_{0n} & s_{1n} \\ s_{1n} & s_{2n} \end{pmatrix}^{-1}\begin{pmatrix} T_{0n} \\ T_{1n} \end{pmatrix} = -\begin{pmatrix} O_p(1) & O_p(1) \\ O_p(1) & O_p(h_n^{-2}) \end{pmatrix}^{-1}$ $\begin{pmatrix} T_{0n} \\ T_{1n} \end{pmatrix}$ uniformly in G and Θ, where $T_{jn} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i - \hat{\alpha}_\theta(x) - \frac{\partial}{\partial x}\hat{\alpha}_\theta(x)(x_i-x);\theta)\frac{1}{h_n}$ $K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j$ for $j = 0, 1$. Now, we write

$$\begin{aligned} T_{jn} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j \\ &+ \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta^*_\theta(x,x_i);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j\hat{d}_{0,\theta}(x) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta^*_\theta(x,x_i);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^{j+1}\hat{d}_{1,\theta}(x) = T_{1j}+T_{2j}+T_{3j} \end{aligned}$$

where $\eta^*_\theta(x,x_i) \in L(\eta_\theta(x,x_i), \hat{\alpha}_\theta(x) + \frac{\partial}{\partial x}\hat{\alpha}_\theta(x)(x_i-x))$. Now we can write

$$\begin{aligned} \sup_{\theta\in\Theta,x\in G}|T_{1j}| &\le \sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta;\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j \\ &- E(\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta;\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)| \\ &+ \sup_{\theta\in\Theta,x\in G}|E(\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)| \\ &= O_p(h_n^j(log(n)/nh_n)^{1/2}) + T_{11j} \text{ by Lemma 1.} \end{aligned}$$

When $j = 0$, it follows directly from from conditions PD that $T_{11j} = O(1)$. When $j = 1$, $T_{111} = \sup_{\theta\in\Theta,x\in G}|E(\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x))|$ and by expanding $\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)$ around $\alpha_\theta(x)$ and $f_\varepsilon(y_i - g_0(x_i);\theta_0)$ around x we write

$$\begin{aligned} T_{111} &= \sup_{\theta\in\Theta,x\in G}|\int\int(\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\alpha_\theta(x);\theta) + \frac{\partial}{\partial\theta_k}\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta^*_\theta(x);\theta)\frac{\partial}{\partial x}\alpha_\theta(x)(x_i-x)) \\ &\times (f_\varepsilon(y_i-g_0(x);\theta_0) + \frac{\partial}{\partial x}f_\varepsilon(y_i-g_0(x^*);\theta_0)(x_i-x))dy_i\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)f_x(x_i)dx_i|. \end{aligned}$$

Given the boundedness of K and its compact support, it follows from conditions PB and PD that

$T_{111} = O(h_n^2)$. Hence, combining the orders obtained we have that

$$\sup_{\theta\in\Theta,x\in G}|T_{1j}| = \begin{cases} O_p((log(n)/nh_n)^{1/2}) + O(1) \text{ if } j=0 \\ O_p(h_n(log(n)/nh_n)^{1/2}) + O(h_n^2) \text{ if } j=1. \end{cases}$$

For the term T_{2j}, we note that from our previous results that $\sup_{\theta\in\Theta,x\in G}|\hat{d}_{0\theta}(x)| = O_p((log(n)/nh_n)^{1/2}) + O_p(h_n^2) = O_p(h_n^2)$. Hence to obtain $\sup_{\theta\in\Theta,x\in G}|T_{2j}| = \begin{cases} O_p(1) \text{ if } j=0 \\ O_p(h_n^2) \text{ if } j=1 \end{cases}$, we need only show that

$$\sup_{\theta\in\Theta,x\in G}|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_k}\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta_\theta^*(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j| = O_p(1).$$

But this follows from an application of Lemma 1 and condition PD. Similar arguments give $\sup_{\theta\in\Theta,x\in G}|T_{3j}| = \begin{cases} O_p(1) \text{ if } j=0 \\ O_p(h_n^2) \text{ if } j=1 \end{cases}$. Combining the orders of T_{1j}, T_{2j}, T_{3j} we conclude that $\sup_{\theta\in\Theta,x\in G}|T_{0n}| = O_p(1)$ and $\sup_{\theta\in\Theta,x\in G}|T_{1n}| = O_p(h_n^2)$, which are sufficient to establish that $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\hat{\alpha}_\theta(x)| = O_p(1)$ and $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)| = O_p(1)$. Now, since $s_{jn} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta_\theta(x,x_i)-d_{0,\theta}^*(x)-d_{1,\theta}^*(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j$ for $j=0,1,2$ where $d_{0,\theta}^*(x) = (1-\lambda_0)(\hat{\alpha}_\theta(x)-\alpha_\theta(x))$ and $d_{1,\theta}^*(x) = (1-\lambda_1)(\frac{\partial}{\partial x}\hat{\alpha}\theta(x)-\frac{\partial}{\partial x}\alpha_\theta(x))$ and taking partial derivatives we obtain,

$$\begin{aligned}
\frac{\partial}{\partial\theta_k}s_{jn} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta_k\partial\eta^2}logf_\varepsilon(y_i-\eta_\theta(x,x_i)-d_{0,\theta}^*(x)-d_{1\theta}^*(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta(x,x_i)-d_{0,\theta}^*(x)-d_{1,\theta}^*(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j \\
&\times (\frac{\partial}{\partial\theta_k}\alpha_\theta(x)+\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\alpha_\theta(x)(x_i-x)+(1-\lambda_0)(\frac{\partial}{\partial\theta_k}\hat{\alpha}_\theta(x)-\frac{\partial}{\partial\theta_k}\alpha_\theta(x)) \\
&+ (1-\lambda_1)(\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)-\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\alpha_\theta(x)))(x_i-x)) \\
&= I_{1j}+I_{2j}.
\end{aligned}$$

We now write,

$$\begin{aligned}
\sup_{\theta\in\Theta,x\in G}|I_{1j}| \le & \sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta_k\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j \\
& - E(\frac{\partial^3}{\partial\theta_k\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)| \\
& + \sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}E(\frac{\partial^3}{\partial\theta_k\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j) \\
& = O_p\left(h_n^j(log(n)/nh_n)^{1/2}\right) + \sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}E(\frac{\partial^3}{\partial\theta_k\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)\times\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j)
\end{aligned}$$

where the last equality follows from Lemma 1. In addition, using regularity conditions PB and PD we have that $\sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}E(\frac{\partial^3}{\partial\theta_k\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j) = O(h_n^j)$ and consequently

$\sup_{\theta\in\Theta,x\in G}|I_{1j}| = O_p(h_n^j)$. The term I_{2j} can be written as,

$$
\begin{aligned}
I_{2j} &= \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta(x,x_i)+d^*_{0,\theta}(x)+d^*_{1,\theta}(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j(1-\lambda_0)\frac{\partial}{\partial\theta_k}\hat{\alpha}_\theta(x)\\
&+ \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta(x,x_i)+d^*_{0\theta}(x)+d^*_{1,\theta}(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^{j+1}(1-\lambda_1)\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)\\
&+ \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta(x,x_i)+d^*_{0\theta}(x)+d^*_{1,\theta}(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j\lambda_0\frac{\partial}{\partial\theta_k}\alpha_\theta(x)\\
&+ \frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta(x,x_i)+d^*_{0\theta}(x)+d^*_{1,\theta}(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^{j+1}\lambda_1\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\hat{\alpha}_\theta(x).
\end{aligned}
$$

Since we have already established that $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\hat{\alpha}_\theta(x)| = O_p(1)$ and $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\hat{\alpha}_\theta(x)| = O_p(1)$ and given that $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\alpha_\theta(x)| < C$ and $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\alpha_\theta(x)| < C$, we need only establish the order of $\sup_{\theta\in\Theta,x\in G}|\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta(x,x_i)+d^*_{0\theta}(x)+d^*_{1\theta}(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j|$ and

$$\sup_{\theta\in\Theta,x\in G}|\frac{\partial^3}{\partial\eta^3}logf_\varepsilon(y_i-\eta_\theta(x,x_i)+d^*_{0\theta}(x)+d^*_{1\theta}(x)(x_i-x);\theta)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^{j+1}|.$$

It follows directly from Lemma 1, and conditions PB and PD that these terms are of order $O_p(h_n^j)$ and $O_p(h_n^{j+1})$. As such, we conclude that $\sup_{\theta\in\Theta,x\in G}|I_{2j}| = O_p(h_n^j)$ and consequently $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}s_{jn}| = O_p(h_n^j)$.
Finally, to complete the proof of c) we must establish (ii), that is, we must show that

$$I = \frac{1}{n}\sum_{i=1}^n\frac{\partial}{\partial\theta_k}(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta))K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j = O_p(h_n^j(log(n)/nh_n)^{1/2}) + O(h_n^{2+j})$$

uniformly in G and Θ for $j = 0, 1$. Observe that we can write

$$
\begin{aligned}
I &= \frac{1}{n}\sum_{i=1}^n\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta))K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j\\
&+ \frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta))K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j(\frac{\partial}{\partial\theta_k}\alpha_\theta(x)+\frac{\partial}{\partial\theta_k}\frac{\partial}{\partial x}\alpha_\theta(x)(x_i-x))
\end{aligned}
$$

and using Lemma 1 we can immediately conclude that

$$
\begin{aligned}
\sup_{\theta\in\Theta,x\in G}|I| &\le O_p(h_n^j(log(n)/nh_n)^{1/2}) + \sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\alpha_\theta(x)|\\
&\times \sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}|\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j\\
&- E(\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j)|\\
&+ \sup_{\theta\in\Theta,x\in G}|\frac{\partial^2}{\partial\theta_k\partial x}\alpha_\theta(x)|\sup_{\theta\in\Theta,\eta\in\mathcal{H},x\in G}|\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^{j+1}\\
&- E(\frac{\partial^2}{\partial\eta^2}logf_\varepsilon(y_i-\eta;\theta)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^{j+1})|\\
&+ \sup_{\theta\in\Theta,x\in G}|E(\frac{\partial}{\partial\theta_k}(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j)|.
\end{aligned}
$$

Since we assume $\sup_{\theta\in\Theta,x\in G}|\frac{\partial}{\partial\theta_k}\alpha_\theta(x)|<C$ and $\sup_{\theta\in\Theta,x\in G}|\frac{\partial^2}{\partial\theta_k\partial x}\alpha_\theta(x)|<C$, we have by another application of Lemma 1 that $\sup_{\theta\in\Theta,x\in G}|I|\leq O_p(h_n^j(log(n)/nh_n)^{1/2})+\sup_{\theta\in\Theta,x\in G}|E(\frac{\partial}{\partial\theta_k}(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)$ $\frac{1}{h_n}(x_i-x)^j)|$. Repeated use of Taylor's Theorem together with the assumption that $\frac{\partial}{\partial\eta}logf_\varepsilon(y-\alpha_\theta(x);\theta)$ is continuously differentiable at θ and $\frac{\partial}{\partial\theta_k}(\frac{\partial}{\partial\eta}logf_\varepsilon(y-\alpha_\theta(x);\theta)f_\varepsilon(y-g_0(x);\theta_0)$ is continuously differentiable at x gives $\sup_{\theta\in\Theta,x\in G}|E(\frac{\partial}{\partial\theta_k}(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)K\left(\frac{x_i-x}{h_n}\right)\frac{1}{h_n}(x_i-x)^j)|=O_p(h_n^{2+j})$ for $j=0,1$, which concludes the proof of part c).

d) We note that

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_l\theta_k}\hat{d}_\theta(x) &= \begin{pmatrix} s_{0n} & s_{1n}\\ s_{1n} & s_{2n}\end{pmatrix}^{-1}\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\theta_l\partial\theta_k}\left(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)\right)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)\begin{pmatrix}1\\ x_i-x\end{pmatrix}\\
&-\begin{pmatrix} s_{0n} & s_{1n}\\ s_{1n} & s_{2n}\end{pmatrix}^{-1}\begin{pmatrix}\frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{0n} & \frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{1n}\\ \frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{1n} & \frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{2n}\end{pmatrix}\hat{d}_\theta(x)\\
&-\begin{pmatrix} s_{0n} & s_{1n}\\ s_{1n} & s_{2n}\end{pmatrix}^{-1}\begin{pmatrix}\frac{\partial}{\partial\theta_k}s_{0n} & \frac{\partial}{\partial\theta_k}s_{1n}\\ \frac{\partial}{\partial\theta_k}s_{1n} & \frac{\partial}{\partial\theta_k}s_{2n}\end{pmatrix}\frac{\partial}{\partial\theta_l}\hat{d}_\theta(x)\\
&-\begin{pmatrix} s_{0n} & s_{1n}\\ s_{1n} & s_{2n}\end{pmatrix}^{-1}\begin{pmatrix}\frac{\partial}{\partial\theta_l}s_{0n} & \frac{\partial}{\partial\theta_l}s_{1n}\\ \frac{\partial}{\partial\theta_l}s_{1n} & \frac{\partial}{\partial\theta_l}s_{2n}\end{pmatrix}\frac{\partial}{\partial\theta_k}\hat{d}_\theta(x).
\end{aligned}$$

Given the results in b) and c) it suffices to establish that $\begin{pmatrix}\sup_{x\in G,\theta\in\Theta}|\frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{0n}| & \sup_{x\in G,\theta\in\Theta}|\frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{1n}|\\ \sup_{x\in G,\theta\in\Theta}|\frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{1n}| & \sup_{x\in G,\theta\in\Theta}|\frac{\partial^2}{\partial\theta_l\partial\theta_k}s_{2n}|\end{pmatrix}=$ $\begin{pmatrix}O_p(1) & O_p(h_n)\\ O_p(h_n) & O_p(h_n^2)\end{pmatrix}$ and that for $j=0,1$ $\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\theta_l\partial\theta_k}\left(\frac{\partial}{\partial\eta}logf_\varepsilon(y_i-\eta_\theta(x,x_i);\theta)\right)\frac{1}{h_n}K\left(\frac{x_i-x}{h_n}\right)(x_i-x)^j=O_p(h_n^j(log(n)/nh_n)^{1/2})+O(h_n^{2+j})$. Then, $\sup_{\theta\in\Theta,x\in G}|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\hat{\alpha}_\theta(x)-\frac{\partial^2}{\partial\theta_j\partial\theta_k}\alpha_\theta(x)|=O_p\left(\left(\frac{log(n)}{nh_n}\right)^{1/2}\right)+O_p\left(h_n^2\right)$ and furthermore $\sup_{\theta\in\Theta,x\in G}|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial x}\hat{\alpha}_\theta(x)-\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial x}\alpha_\theta(x)|=O_p\left(h_n^{-1}\left(\frac{log(n)}{nh_n}\right)^{1/2}\right)+O_p\left(h_n\right)$. We omit the rest of the proof since it follows arguments that mimic those used in part c). However, in addition to conditions PB and PD, conditions PC must be assumed to complete the proof.

Appendix 2 - Tables and Figures

Table 1 Bias($\times 10^{-1}$)(B), Standard deviation(S) and Root Mean Squared Error(R) for ML, PL and PI estimators with $g_1(x) = 1 + x$

	n		$(\sigma_u^2, \sigma_v^2) = (1.379, 0.501)$ B	S	R	$(0.988, 0.642)$ B	S	R	$(0.551, 0.799)$ B	S	R
σ_u^2	300	ML	-0.468	0.466	0.467	-0.672	0.484	0.488	-0.561	0.469	0.472
		PL	-0.797	0.532	0.537	-1.193	0.576	0.587	-0.380	0.521	0.522
		PI	-1.374	0.552	0.568	-1.655	0.582	0.605	-0.457	0.510	0.511
	600	ML	-0.013	0.302	0.301	-0.048	0.337	0.337	-0.473	0.367	0.370
		PL	-0.233	0.375	0.375	-0.459	0.426	0.428	-0.504	0.431	0.434
		PI	-0.542	0.405	0.408	-0.891	0.467	0.475	-0.470	0.435	0.437
	900	ML	-0.204	0.240	0.240	-0.243	0.281	0.282	-0.594	0.335	0.340
		PL	-0.366	0.282	0.285	-0.576	0.365	0.369	-0.620	0.402	0.407
		PI	-0.576	0.304	0.309	-0.929	0.414	0.423	-0.668	0.418	0.423
σ_v^2	300	ML	0.065	0.146	0.146	0.192	0.177	0.178	0.086	0.179	0.179
		PL	0.076	0.170	0.170	0.286	0.202	0.204	-0.067	0.189	0.189
		PI	0.286	0.178	0.180	0.460	0.203	0.208	-0.029	0.185	0.185
	600	ML	0.051	0.098	0.098	0.002	0.120	0.120	0.104	0.136	0.136
		PL	0.077	0.122	0.122	0.102	0.149	0.149	0.074	0.156	0.156
		PI	0.190	0.133	0.134	0.266	0.164	0.166	0.074	0.158	0.158
	900	ML	0.064	0.075	0.075	0.071	0.097	0.097	0.170	0.124	0.125
		PL	0.083	0.089	0.089	0.163	0.127	0.128	0.164	0.145	0.146
		PI	0.159	0.096	0.097	0.298	0.145	0.148	0.191	0.151	0.152

Table 2 Bias($\times 10^{-1}$)(B), Standard deviation(S) and Root Mean Squared Error(R) for ML, PL and PI estimators with $g_2(x) = 1 + ln(1 + x)$

	n		$(\sigma_u^2, \sigma_v^2) = (1.379, 0.501)$ B	S	R	$(0.988, 0.642)$ B	S	R	$(0.551, 0.799)$ B	S	R
σ_u^2	300	ML	-0.220	0.429	0.429	-0.763	0.507	0.513	-0.290	0.465	0.466
		PL	-1.026	0.519	0.529	-1.972	0.607	0.638	-0.704	0.509	0.513
		PI	-1.800	0.558	0.585	-2.432	0.613	0.659	-0.764	0.490	0.495
	600	ML	-0.207	0.294	0.294	-0.199	0.350	0.350	0.038	0.366	0.366
		PL	-0.835	0.391	0.400	-1.320	0.463	0.481	-0.774	0.421	0.428
		PI	-1.258	0.430	0.448	-1.904	0.507	0.541	-1.076	0.434	0.447
	900	ML	-0.079	0.237	0.237	-0.090	0.278	0.278	-0.363	0.327	0.329
		PL	-0.425	0.277	0.280	-1.033	0.379	0.392	-1.325	0.385	0.407
		PI	-0.750	0.323	0.331	-1.630	0.447	0.475	-1.603	0.406	0.436
σ_v^2	300	ML	-0.083	0.137	0.137	0.227	0.171	0.173	-0.023	0.180	0.180
		PL	0.076	0.170	0.170	0.544	0.210	0.217	0.032	0.191	0.191
		PI	0.355	0.185	0.188	0.717	0.213	0.224	0.065	0.186	0.186
	600	ML	0.066	0.094	0.094	0.041	0.119	0.119	-0.066	0.139	0.139
		PL	0.218	0.127	0.129	0.383	0.161	0.165	0.178	0.158	0.159
		PI	0.372	0.141	0.146	0.603	0.178	0.188	0.297	0.163	0.166
	900	ML	0.030	0.077	0.077	0.053	0.093	0.093	0.133	0.122	0.122
		PL	0.098	0.089	0.089	0.343	0.132	0.136	0.451	0.142	0.149
		PI	0.214	0.105	0.107	0.567	0.157	0.167	0.561	0.148	0.159

Table 3 Bias($\times 10^{-1}$)(B), Standard deviation(S) and Root Mean Squared Error(R) for ML, PL, and PI estimators with $g_3(x) = 1 - \frac{1}{1+x}$

	n		$(\sigma_u^2, \sigma_v^2) = (1.379, 0.501)$ B	S	R	$(0.988, 0.642)$ B	S	R	$(0.551, 0.799)$ B	S	R
σ_u^2	300	ML	-0.139	0.466	0.466	-0.479	0.498	0.500	0.203	0.481	0.481
		PL	-1.096	0.582	0.592	-1.782	0.593	0.619	-0.546	0.531	0.533
		PI	-1.755	0.600	0.625	-2.427	0.597	0.644	-0.912	0.525	0.532
	600	ML	-0.136	0.282	0.282	-0.466	0.354	0.357	-0.509	0.373	0.376
		PL	-0.891	0.383	0.393	-1.927	0.476	0.513	-1.681	0.412	0.445
		PI	-1.369	0.439	0.459	-2.558	0.513	0.573	-1.977	0.424	0.467
	900	ML	-0.122	0.226	0.226	0.029	0.278	0.278	-0.371	0.322	0.323
		PL	-0.743	0.293	0.302	-1.307	0.397	0.417	-1.407	0.366	0.391
		PI	-1.104	0.334	0.352	-1.986	0.459	0.500	-1.753	0.384	0.422
σ_v^2	300	ML	0.005	0.137	0.137	0.111	0.174	0.174	-0.115	0.175	0.175
		PL	0.219	0.181	0.182	0.469	0.209	0.214	0.060	0.193	0.193
		PI	0.453	0.190	0.195	0.707	0.213	0.225	0.198	0.191	0.191
	600	ML	0.015	0.088	0.088	0.084	0.125	0.126	0.104	0.135	0.135
		PL	0.203	0.125	0.127	0.539	0.168	0.176	0.477	0.149	0.157
		PI	0.377	0.145	0.150	0.773	0.183	0.199	0.596	0.155	0.166
	900	ML	-0.004	0.071	0.071	-0.013	0.094	0.094	0.117	0.118	0.118
		PL	0.151	0.095	0.096	0.411	0.139	0.144	0.457	0.134	0.141
		PI	0.277	0.110	0.113	0.663	0.164	0.177	0.592	0.142	0.154

Table 4 Bias($\times 10^{-1}$)(B), Standard deviation(S) and Root Mean Squared Error(R) for ML, PL, and PI estimators with $g_4(x) = 1 + 0.5arctan(20(x - 0.5))$

	n		$(\sigma_u^2, \sigma_v^2) = (1.379, 0.501)$ B	S	R	$(0.988, 0.642)$ B	S	R	$(0.551, 0.799)$ B	S	R
σ_u^2	300	ML	-0.190	0.437	0.437	-0.080	0.473	0.473	-0.128	0.453	0.453
		PL	-0.543	0.509	0.511	-0.209	0.533	0.533	-0.275	0.507	0.508
		PI	-1.228	0.538	0.552	-0.813	0.542	0.548	-0.373	0.484	0.485
	600	ML	-0.222	0.314	0.314	-0.447	0.344	0.347	-0.290	0.372	0.372
		PL	-0.399	0.366	0.368	-0.622	0.419	0.423	-0.427	0.427	0.429
		PI	-0.791	0.394	0.401	-0.977	0.447	0.457	-0.530	0.428	0.431
	900	ML	-0.218	0.246	0.246	-0.399	0.300	0.303	-0.305	0.309	0.310
		PL	-0.301	0.269	0.271	-0.735	0.365	0.372	-0.433	0.353	0.356
		PI	-0.547	0.289	0.294	-1.217	0.425	0.441	-0.749	0.381	0.388
σ_v^2	300	ML	0.025	0.137	0.137	-0.051	0.167	0.167	-0.040	0.175	0.175
		PL	0.056	0.160	0.159	-0.101	0.188	0.188	-0.069	0.192	0.192
		PI	0.298	0.171	0.174	0.115	0.195	0.195	-0.038	0.184	0.184
	600	ML	0.029	0.095	0.095	0.133	0.120	0.121	0.020	0.143	0.143
		PL	0.037	0.113	0.113	0.154	0.143	0.143	0.024	0.160	0.160
		PI	0.177	0.123	0.124	0.289	0.154	0.156	0.058	0.160	0.160
	900	ML	0.048	0.079	0.079	0.127	0.105	0.106	0.086	0.117	0.117
		PL	0.041	0.088	0.088	0.211	0.128	0.129	0.103	0.131	0.132
		PI	0.129	0.095	0.096	0.390	0.150	0.155	0.220	0.141	0.143

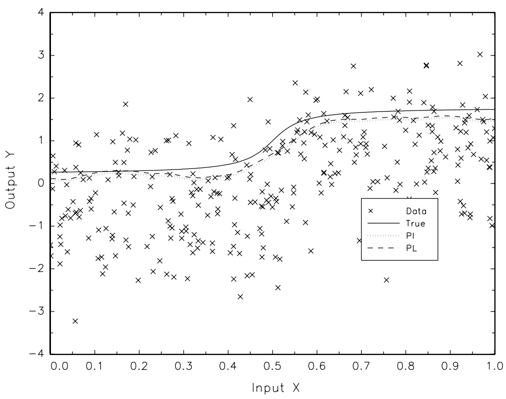

Figure 1: Simulated dataset $n = 300$, $g_4(x) = 1 + 0.5arctan(20(x - 0.5))$ and $(\sigma_u^2, \sigma_v^2) = (1.379, 0.501)$.

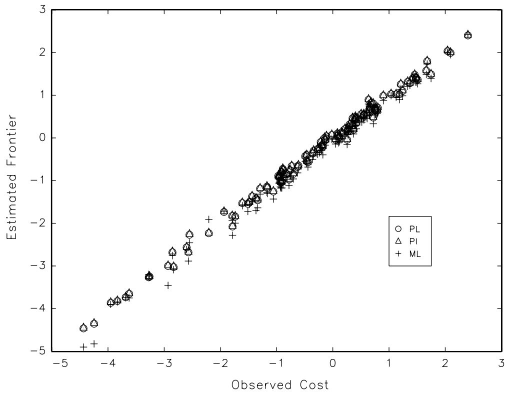

Figure 2: Cost frontier estimated with PL, PI, and ML.

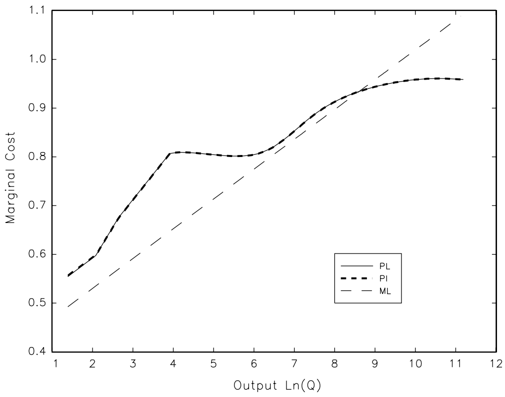

Figure 3: Marginal cost estimated with PL, PI, and ML.

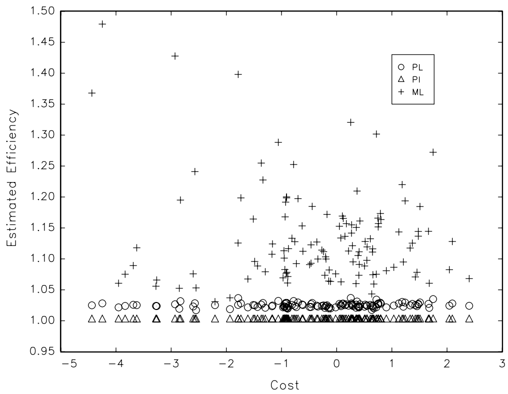

Figure 4: Plot of estimated efficiencies with PL, PI, and ML against cost $Ln(C/P_f)$.

References

[1] Almanidis, P., and R. C. Sickles, 2011, Skewness issues in stochastic frontier models: fact or fiction? In I. van Keilegom and P. W. Wilson (Eds), Exploring Research Frontiers in Contemporary Statistics and Econometrics, 201-227, Springer-Verlag, Berlin Heidelberg.

[2] Aigner, D., C. A. K. Lovell and P. Schmidt, 1977, Formulation and estimation of stochastic frontiers production function models. Journal of Econometrics, 6, 21-37.

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

[4] Bickel, P., C. A. J. Klaassen, Y. Ritov and J. Wellner, 1993, Efficient and adaptive estimation for semiparametric models. Springer, New York.

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

[6] Christensen, L.R. and W. H. Greene, 1976, Economies of scale in U.S. electric power generation. Journal of Political Economy, 84, 655-676.

[7] Coelli, T., 1995, Estimators and hypothesis tests for a stochastic frontier function. Journal of Productivity Analysis, 6, 247-268.

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

[9] Daouia, A., L. Gardes and S. Girard, 2009, Large sample approximation of the distribution for smoothed monotone frontier estimators. INRIA working paper.

[10] Doksum, K. and A. Samarov, 1995, Nonparametric estimation of of global functionals and a measure of the explanatory power of covariates in regression. The Annals of Statistics, 23, 1443-1473.

[11] Fan, J., 1992, Design adaptive nonparametric regression. Journal of the American Statistical Association, 87, 998-1004.

[12] Fan, J., 1993, Local linear regression smoothers and their mini-max efficiencies. Annals of Statistics, 21, 196-216.

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

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

[15] Gijbels, I., E. Mammen, B. Park and L. Simar, 1999, On estimation of monotone and concave frontier functions. Journal of the American Statistical Association, 94, 220-228.

[16] Graves, L. M., 1927, Riemann integration and Taylor's Theorem in general analysis. Transactions of the American Mathematical Society, 29,163-177.

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

[18] Greene, W., H., 1993, The econometric approach to efficiency analysis. In: H. Fried, C. A. K. Lovell, and S. S. Schmidt, (Eds.), The Measurement of Productive Efficiency. Oxford University Press, Oxford.

[19] Härdle, W., 1990, Applied non-parametric regression. Cambridge University Press, Cambridge, UK.

[20] Jondrow, J., C. A. K. Lovell, I. S. Materov and P. Schmidt, 1982, On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19, 233-238.

[21] Kumbhakar, S. C. and Lovell, C. A. K., 2000, Stochastic frontier analysis. Cambridge university Press, Cambridge, UK.

[22] Kumbhakar, S. C., B. U. Park, L. Simar and E. Tsionas, 2007, Nonparametric stochastic frontiers: a local maximum likelihood approach. Journal of Econometrics, 137, 1-27.

[23] Kuosmanen, T. and M. Kortelainen, 2011, Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. Journal of Productivity Analysis, forthcoming.

[24] Li, Q., and J. Racine, 2007, Nonparametric Econometrics. Princeton University Press, Princeton, NJ.

[25] Lam, C., and J. Fan, 2008, Profile kernel likelihood inference with diverging number of parameters. Annals of Statistics, 36, 2232-2260.

[26] Luenberger, D., 1969, Optimization by vector space methods. John Wiley and Sons, New York.

[27] Lusternik, L. A. and V. J. Sobolev, 1964, Elements of functional analysis. Hindustan Publishing, Delhi.

[28] Martins-Filho, C. and F. Yao, 2007, Nonparametric frontier estimation via local linear regression. Journal of Econometrics, 141, 283-319.

[29] Martins-Filho, C. and F. Yao, 2008, A smoothed conditional quantile frontier estimator. Journal of Econometrics, 143, 317-333.

[30] Martins-Filho, C. and F. Yao, 2011, Technical supplement to "Semiparametric stochastic frontier estimation via profile likelihood." http://spot.colorado.edu/∼martinsc/Research.html

[31] Meeusen, W. and J. van den Broeck, 1977, Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18, 435-444.

[32] Ruppert, D., S. Sheather and M. P. Wand, 1995, An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association, 90, 1257-1270.

[33] Severini, T., 2000, Likelihood methods in statistics. Oxford University Press, Oxford.

[34] Severini, T. and W. H. Wong, 1992, Profile likelihood and conditionally parametric models. Annals of Statistics, 20, 1768-1802.

[35] Simar, L., 2007, How to improve the performances of DEA/FDH estimators in the presence of noise? Journal of Productivity Analysis 28, 183-201.

[36] Simar, L. and P. Wilson, 2008, Statistical inference in nonparametric frontier models: recent developments and perspectives. In: H. Fried, C. A. K. Lovell, and S. S. Schmidt, (Eds.), The Measurement of Productive Efficiency, 2nd edition. Oxford University Press, Oxford.

[37] Simar, L. and P. Wilson, 2010, Inferences from cross-sectional, stochastic frontier models. Econometric Reviews, 29, 62-98.

[38] Simar, L. and V. Zelenyuk, 2011, Stochastic FDH/DEA estimators for frontier analysis. Journal of Productivity Analysis, forthcoming.

[39] Staniswalis, J., 1989, On the kernel estimate of a regression functionin likelihood based models. Journal of the American Statistical Association, 84, 276-283.

[40] Stein, C., 1954, Efficient nonparametric testing and estimation. In: Neyman, J. (Ed.), Proceedings of the third Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley.

[41] van der Vaart, A., 1999, Semiparametric statistics. Lecture notes, Vrije Universiteit Amsterdam.