A Nonparametric Model of Frontiers

Carlos Martins-Filho
Department of Economics
Oregon State University
Ballard Hall 303
Corvallis, OR 97331-3612 USA
email: carlos.martins@orst.edu
Voice: + 1 541 737 1476
Fax: + 1 541 737 5917

and

Feng Yao
Department of Economics
Oregon State University
Ballard Hall 303
Corvallis, OR 97331-3612 USA
email: yaof@onid.orst.edu
Voice: + 1 541 737 2321
Fax: + 1 541 737 5917

December, 2003

Abstract. In this paper we propose a nonparametric regression frontier model that assumes no specific parametric family of densities for the unobserved stochastic component that represents efficiency in the model. Nonparametric estimation of the regression frontier is obtained using a local linear estimator that is shown to be consistent and $\sqrt{nh_n}$ asymptotically normal under standard assumptions. The estimator we propose envelops the data but is not inherently biased as Free Disposal Hull - FDH or Data Envelopment Analysis - DEA estimators. It is also more robust to extreme values than the aforementioned estimators. A Monte Carlo study is performed to provide preliminary evidence on the estimator's finite sample properties and to compare its performance to a bias corrected FDH estimator.

Keywords and Phrases. nonparametric regression frontier, local linear estimation, U statistics.

JEL Classifications. C14

1 Introduction

The specification and estimation of production frontiers and the measurement of the associated efficiency level of production units has been the subject of a vast and growing literature since the seminal work of Farrell(1957). The main objective of this literature can be stated simply. Consider $(y, x) \in \Re_+ \times \Re_+^K$ where y describes the output of a production unit and x describes the K inputs used in production. The production technology is given by the set $T = \{(y, x) \in \Re_+ \times \Re_+^K : x \text{ can produce } y\}$ and the production function or frontier associated with T is $\rho(x) = sup\{y \in \Re_+ : (y, x) \in T\}$ for all $x \in \Re_+^K$. Let $(y_0, x_0) \in T$ characterize the performance of a production unit and define $0 \leq R_0 \equiv \frac{y_0}{\rho(x_0)} \leq 1$ to be this unit's (inverse) Farrell output efficiency measure. The main objective in production and efficiency analysis is, given a random sample of production units $\{(Y_t, X_t)\}_{t=1}^n$ that share a technology T, to obtain estimates of $\rho(\cdot)$ and by extension $R_t = \frac{Y_t}{\rho(X_t)}$ for $t = 1, \cdots, n$. Secondary objectives, such as efficiency rankings and relative performance of production units, can be subsequently obtained.

There exists in the current literature two main approaches for the estimation of $\rho(\cdot)$. The deterministic approach, represented by Charnes et al.(1978) data envelopment analysis (DEA) and Deprins et al.(1984) free disposal hull (FDH) estimators, is based on the assumption that all observed data lies in the technology set T, i.e., $P((Y_t, X_t) \in T) = 1$ for all t. The stochastic approach, pioneered by Aigner, Lovell and Schmidt(1977) and Meeusen and van den Broeck(1977), allows for random shocks in the production process and consequently $P((Y_t, X_t) \notin T) > 0$. Although more appealing from an econometric perspective, it is unfortunate that identification of stochastic frontier models requires strong parametric assumptions on the joint distribution of (Y_t, X_t) and/or $\rho(\cdot)$. These parametric assumptions may lead to misspecification of $\rho(\cdot)$ and invalidate any optimal derived properties of the proposed estimators (generally maximum likelihood) and consequently lead to erroneous inference. In addition, as recently pointed out by Baccouche and Kouki(2003), estimated inefficiency levels and firm efficiency rankings are sensitive to the specification of the joint density of (Y_t, X_t). Hence, different density specifications can lead to different conclusions regarding technology and efficiency from the same random sample. Such deficiencies of stochastic frontier models have contributed to the popularity of deterministic frontiers.[1]

Deterministic frontier estimators, such as DEA and FDH, have gained popularity among applied researchers because their construction relies on very mild assumptions on the technology T. Specifically, there is no need to assume any restrictive parametric structure on $\rho(\cdot)$ or the joint density of (Y_t, X_t). In addition to the flexible nonparametric structure, the appeal of these estimators has increased since Gijbels et al.(1999) and Park, Simar and Weiner(2000) have obtained their asymptotic distributions under some fairly reasonable assumptions.[2] Although much progress has been made in both estimation and inference in the deterministic frontier literature, we believe that alternatives to DEA and FDH estimators may be desirable. Recently, Cazals et al.(2002) have proposed a new estimator based on the joint survivor function that is more robust to extreme values and outliers than DEA and FDH estimators and does not suffer from their inherent biasedness.[3]

In this paper we propose a new deterministic production frontier regression model and estimator that can be viewed as an alternative to the methodologies currently available, including DEA and FDH estimators and the estimator of Cazals et al.(2002). Our frontier model shares the flexible nonparametric structure that characterizes the data generating processes (DGP) underlying the results in Gijbels et al.(1999) and Park, Simar and Weiner(2000) but in addition our estimation procedure has some general properties that can prove desirable *vis a vis* DEA and FDH. First, as in Cazals et al.(2002), the estimator we propose is more robust to extreme values and outliers; second, our frontier estimator is a smooth function of input usage, not a discontinuous or piecewise linear function (as in the case of FDH and DEA estimators, respectively); third, the construction of our estimator is fairly simple as it is in essence a local linear kernel estimator; and fourth, although our estimator envelops the data, it is not intrinsically biased and therefore no bias correction is necessary. In addition to these general properties we are able to establish the asymptotic distribution and consistency of the production frontier and efficiency estimators under assumptions that are fairly standard in the nonparamteric statistics literature. We view our proposed estimator not necessarily as a substitute to

[1]See Seifford(1996) for an extensive literature review that illustrates the widespread use of deterministic frontiers.

[2]See the earlier work of Banker(1993) and Korostelev, Simar and Tsybakov(1995) for some preliminary asymptotic results.

[3]Bias corrected FDH and DEA estimators are available but their asymptotic distributions are not known. Again, see Gijbels et al.(1999) and Park, Simar and Weiner(2000)

estimators that are currently available but rather as an alternative that can prove more adequate in some contexts.

In addition to this introduction, this paper has five more sections. Section 2 describes the model in detail, contrasts its assumptions with those in the past literature and describes the estimation procedure. Section 3 provides supporting lemmas and the main theorems establishing the asymptotic behavior of our estimators. Section 4 contains a Monte Carlo study that implements the estimator, sheds some light on its finite sample properties and compares its performance to the bias corrected FDH estimator of Park, Simar and Weiner(2000). Section 5 provides a conclusion and some directions for future work.

2 A Nonparametric Frontier Model

The construction of our frontier regression model is inspired by data generating processes for multiplicative regression. Hence, rather than placing primitive assumptions directly on (Y_t, X_t) as it is common in the deterministic frontier literature, we place primitive assumptions on (X_t, R_t) and obtain the properties of Y_t by assuming a suitable regression function. We assume that $Z_t \equiv (X_t, R_t)'$ is a $K+1$-dimensional random vector with common density g for all $t \in \{1, 2, ...\}$ and that $\{Z_t\}$ forms an independently distributed sequence. We assume there are observations on a random variable Y_t described by

$$Y_t = \sigma(X_t)\frac{R_t}{\sigma_R}. \tag{1}$$

R_t is an unobserved random variable, X_t is an observed random vector taking values in $\Re^K_+$, $\sigma(\cdot) : \Re^K_+ \to (0, \infty)$ is a measurable function and σ_R is an unknown parameter. In the case of production frontiers we interpret Y_t as output, $\rho(\cdot) \equiv \frac{\sigma(\cdot)}{\sigma_R}$ as the production frontier with inputs X_t and R_t as efficiency with values in $[0, 1]$. R_t has the effect of contracting output from optimal levels that lie on the production frontier. The larger R_t the more efficient the production unit because the closer the realized output is to that on the production frontier. In section 3 we provide a detailed list of assumptions that is used in obtaining the asymptotic properties of our estimator, however in defining the elements of the model and the estimator, two important conditional moment restrictions on R_t must be assumed; $E(R_t|X_t = x) \equiv \mu_R$ where $0 < \mu_R < 1$ and $V(R_t|X_t = x) \equiv \sigma^2_R$. It should be noted that by construction $0 < \sigma^2_R < \mu_R < 1$. The parameter μ_R is interpreted as a mean efficiency given input usage and the common technology T and σ_R is a scale parameter for the conditional distribution of R_t that also locates the production frontier. These conditional moment restrictions together with equation (1) imply that $E(Y_t|X_t = x) = \frac{\mu_R}{\sigma_R}\sigma(x)$ and $V(Y_t|X_t = x) = \sigma^2(x)$. The model can therefore be rewritten as,

$$Y_t = \sigma(X_t)\frac{R_t}{\sigma_R} = b\sigma(X_t) + \sigma(X_t)\frac{(R_t - \mu_R)}{\sigma_R} = m(X_t) + \sigma(X_t)\epsilon_t \tag{2}$$

where $b = \frac{\mu_R}{\sigma_R}$, $\epsilon_t = \frac{R_t-\mu_R}{\sigma_R}$, $m(X_t) = b\sigma(X_t)$, $E(\epsilon_t|X_t = x) = 0$ and $V(\epsilon_t|X_t = x) = 1$.[4]

The frontier model described in (2) has a number of desirable properties. First, the frontier $\rho(\cdot) \equiv \frac{\sigma(\cdot)}{\sigma_R}$ is not restricted to belong to a known parametric family of functions and therefore there is no *a priori* undue restriction on the technology T. Second, although the existence of conditional moments are assumed for R_t, no specific parametric family of densities is assumed, therefore bypassing a number of potential problems arising from misspecification. Third, the model allows for conditional heteroscedasticity of Y_t as has been argued for in previous work (Caudill et al., 1995 and Hadri, 1999) on production frontiers. Finally, the structure of (2) is similar to regression models studied by Fan and Yao(1998), therefore lending itself to similar estimation *via* kernel procedures. This similarity motivates the estimation procedure that is described below.

The nonparametric local linear frontier estimation we propose can be obtained in three easily implementable stages. For any $x \in \Re^K_+$ we first obtain $\hat{m}(x) \equiv \hat{\alpha}$ where

$$(\hat{\alpha}, \hat{\beta}) = argmin_{\alpha,\beta} \sum_{t=1}^{n} (Y_t - \alpha - \beta(X_t - x))^2 K\left(\frac{X_t - x}{h_n}\right)$$

$K(\cdot) : \Re^K \to \Re$ is a density function and $0 < h_n \to 0$ as $n \to \infty$ is a bandwidth. This is the local linear kernel estimator of Fan(1992) with regressand Y_t and regressors X_t. In the second stage, we follow Hall and Carroll(1989) and Fan and Yao(1998) by defining $e_t \equiv (Y_t - \hat{m}(X_t))^2$ to obtain $\hat{\alpha}_1 \equiv \hat{\sigma}^2(x)$, where

$$(\hat{\alpha}_1, \hat{\beta}_1) = argmin_{\alpha_1,\beta_1} \sum_{t=1}^{n} (e_t - \alpha_1 - \beta_1(X_t - x))^2 K\left(\frac{X_t - x}{h_n}\right)$$

[4]For simplicity in notation, we will henceforth write $E(\cdot|X_t = x)$ or $V(\cdot|X_t = x)$ simply as $E(\cdot|X_t)$ or $V(\cdot|X_t)$.

which provides an estimator $\hat{\sigma}(x) = \left(\hat{\sigma}^2(x)\right)^{1/2}$. In the third stage an estimator for σ_R,

$$s_R = \left(max_{1\le t\le n} \frac{Y_t}{\hat{\sigma}(X_t)} \right)^{-1}$$

is obtained. Hence, a production frontier estimator at $x \in \Re^K$ is given by $\hat{\rho}(x) = \frac{\hat{\sigma}(x)}{s_R}$. We note that by construction, provided that the chosen kernel K is smooth, $\hat{\rho}(x)$ is a smooth estimator that envelops the data (no observed pair Y_t lies above $\hat{\rho}(X_t)$) but may lie above or below the true frontier $\rho(X_t)$.

In our model, the parameter σ_R provides the location of the production frontier, whereas its shape is provided by $\sigma(\cdot)$. Since besides the conditional moment restrictions on R_t there are no other restrictions other than $R_t \in [0, 1]$, the observed data $\{(Y_t, X_t)\}_{t=1}^n$ may or may not be dispersed close to the frontier, hence the estimation of σ_R requires an additional normalization assumption. We assume that there exists one observed production unit that is efficient, in that the forecasted value for R_t associated with this unit is identically one. This normalization provides the motivation for the above definition of s_R. The problem of locating the production frontier is also inherent in obtaining DEA and FDH estimators. The normalization in these cases involves a number of production units being forced by construction to be efficient, i.e., lie on the frontier. This results from the fact that these estimators are defined to be minimal functions (with some stated properties, e.g., convexity and monotonicity) that envelope the data. Hence, if the stochastic process that generates the data is such that (Y_t, X_t) lie away from the true frontier, e.g., μ_R and σ_R are small, DEA and FDH will provide a downwardly biased location for the frontier. It is this dependency on boundary data points that makes these estimators highly succeptible to extreme values. This is in contrast with the estimator we propose which by construction is not a minimal enveloping function of the data. Furthermore, we note that although the location of the frontier in our model depends on the estimator s_R and its inherent normalization, if estimated efficiency levels are defined as $\hat{R}_t = \frac{s_R Y_t}{\hat{\sigma}(X_t)}$, the efficiency ranking of firms, as well as their estimated relative efficiency $\frac{\hat{R}_t}{\hat{R}_\tau}$ for $t, \tau = 1, 2, \cdots, n$ are entirely independent of the estimator s_R. In the next section we investigate the asymptotic properties of our estimators.

3 Asymptotic Characterization of the Estimators

In this section we establish the asymptotic properties of the frontier estimator described above. We first provide a sufficient set of assumptions for the results we prove below and provide some contrast with the assumptions made in Gijbels et al. (1999) and Park, Simar and Wiener(2000) to obtain the asymptotic distribution of DEA and FDH estimators.

Assumption A1. 1. $Z_t = (X_t, R_t)'$ for $t = 1, 2, \cdots, n$ is an independent and identically distributed sequence of random vectors with density g. We denote by $g_X(x)$ and $g_R(r)$ the common marginal densities of X_t and R_t respectively, and by $g_{R|X}(r; X)$ the common conditional density of R_t given X. 2. $0 < \underline{B}_{g_X} \leq g_X(x) \leq \bar{B}_{g_X} < \infty$ for all $x \in G$, G a compact subset of $\Theta = \times_{t=1}^{K}(0, \infty)$, which denotes the Cartesian product of the intervals $(0, \infty)$.

Assumption A2. 1. $Y_t = \sigma(X_t)\frac{R_t}{\sigma_R}$; 2. $R_t \in [0, 1]$, $X_t \in \Theta$; 3. $E(R_t|X_t) = \mu_R$, $V(R_t|X_t) = \sigma_R^2$; 4. $0 < \underline{B}_\sigma \leq \sigma(x) \leq \bar{B}_\sigma < \infty$ for all $x \in \Theta$; 5. $\sigma^2(\cdot) : \Theta \to \Re$ is a measurable twice continuously differentiable function in Θ; 6. $|\sigma^{2(2)}(x)| < \bar{B}_{2\sigma}$ for all $x \in \Theta$

Assumptions A1.1 and A2 imply that $\{(Y_t, X_t)\}_{t=1}^n$ forms an iid sequence of random variables with some joint density $\phi(y, x)$. This corresponds to assumption AI in Park, Simar and Wiener(2000) and is also assumed in Gijbels et al.(1999). Given that $0 < \sigma_R < 1$, A2.4 and A2.5 are implied by assumption AIII in Park, Simar and Wiener(2000). A2.6 is implied by A2 in Gijbels et al.(1999) and AIII in Park, Simar and Wiener(2000). The following assumption A3 is standard in nonparametric estimation and involves only the kernel $K(\cdot)$. We observe that A3 is satisfied by commonly used kernels such as Epanechnikov, Biweight and others.

Assumption A3. $K(x) : \Re \to \Re$ is a symmetric density function with bounded support $S_K \subset \Re^K$ satisfying:

1. $\int xK(x)dx = 0$; 2. $\int x^2K(x)dx = \sigma_K^2$; 3. for all $x \in \Re^K$, $|K(x)| < B_K < \infty$; 4. for all $x, x' \in \Re^K$, $|K(x) - K(x')| < m||x - x'||$ for some $0 < m < \infty$;

ASSUMPTION A4. For all $x, x' \in \Theta$, $|g_X(x) - g_X(x')| < m_g||x - x'||$ for some $0 < m_g < \infty$.

A Lipschitz condition such as A4 is also assumed in Park, Simar and Wiener(2000). We note that consistency and asymptotic normality of the DEA and FDH estimators for the production frontier and associated firm efficiency depends crucially on the assumption (AII in Park, Simar and Wiener(2000)) that the joint density $\phi(y, x)$ of (Y_t, X_t) is positive at the frontier.[5] At an intuitive level this means that the data generating process (DGP)cannot be one that repeatedly gives observations that are bounded away from the frontier. In reality, there might be situations in which this assumption can be too strong. Consider, for example, the analysis of efficiency in a domestic industry or sector of an economy that is institutionally protected from foreign - potentially more efficient - competition. Unless there is an institutional change (open markets) it seems unreasonable to assume that the DGP is one that would produce efficient production units. In contrast, we assume that R_t takes values in the entire interval $[0, 1]$, but there is no need for the density of the data to be positive at the frontier to obtain consistency or asymptotic normality of the frontier estimator. However, asymptotic normality of the frontier, as is made explicit in Theorem 2 requires a particular assumption on the speed of convergence of $max_{1 \leq t \leq n} R_t$ to 1 as $n \to \infty$, which clearly implies some restriction on the shape of g_R.

Lastly, we make some general comments on our assumptions. As alluded to before the assumption that Z_t are iid does not prevent the model from allowing for conditional heteroscedasticity. Also, we do not assume that X_t and R_t are contemporaneously independent as it is usually done in stochastic frontier models. All that is assumed here is that conditional first and second centered moments are independent of input usage.

The main difficulties in obtaining the asymptotic properties of $\hat{\sigma}$ and by consequence those of $\frac{\hat{\sigma}}{s_R}$ derive from the fact that $\hat{\sigma}$ is based on regressands that are themselves residuals from a first stage nonparametric

[5] By consequence this assumption is also crucial in obtaining the asymptotic distribution of the estimator proposed by Cazals et al.(2002).

regression. This problem is in great part handled by the use of Lemma 3 on U statistics that appears in the appendix. Although we need only deal with U-statistics of dimension 2, Lemma 3 generalizes to $k \leq n$ Lemma 3.1 in Powell et al.(1989) where the case for $k = 2$ is proven. This lemma is of general interest and can be used whenever there is a need to analyze some specific linear combinations of nonparametric kernel estimators. For simplicity, but without loss of generality, all of our proofs are for $K = 1$. For $K > 1$ all of the results hold with appropriate adjustments on the relative speed of n and h_n^K.[6]

Lemma 1 below establishes the order in probability of certain linear combinations of kernel functions that appear repeatedly in component expressions of our estimators. The proof of the lemmas and theorems that follow depend on the repeated application of a version of Lebesgue's dominated convergence theorem which can be found in Pagan and Ullah(1999, p.362) and Prakasa-Rao (1983, p.35). Henceforth, we refer to this result as *the proposition of Prakasa-Rao.*

Lemma 1 *Assume A1, A2, A3 and suppose that* $f(x,r) : (0,\infty) \times [0,1] \rightarrow \Re$ *is a continuous function in* G *a compact subset of* $(0,\infty)$ *with* $|f(x,r)| < B_f < \infty$. *Let*

$$s_j(x) = (nh_n)^{-1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^j f(X_t, R_t) \text{ with } j = 0, 1, 2.$$

a) If $nh_n^2 \rightarrow \infty$ *then* $sup_{x\in G}|s_j(x) - E(s_j(x))| = O_p\left(\frac{ln(n)}{nh_n}\right)$.

b) If $nh_n^{2p+1}(ln(h_n))^{-1} \rightarrow \infty$ *for* $p > 0$, *then* $sup_{x\in G}|s_j(x) - E(s_j(x))| = o_p(h_n^p)$

Proof [Lemma 1] a) We prove the case where $j = 0$. Similar arguments can be used for $j = 1, 2$. Let $B(x_0, r) = \{x \in \Re : |x - x_0| < r\}$ for $r \in \Re^+$. G compact implies that there exists $x_0 \in G$ such that $G \subseteq B(x_0, r)$. Therefore for all $x, x' \in G$ $|x - x'| < 2r$. Let $h_n > 0$ be a sequence such that $h_n \rightarrow 0$ as $n \rightarrow \infty$ where $n \in \{1, 2, 3 \cdots\}$. For any n, by the Heine-Borel theorem there exists a finite collection of sets $\left\{B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)\right\}_{k=1}^{l_n}$ such that $G \subset \cup_{k=1}^{l_n} B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)$ for $x_k \in G$ with $l_n < \left(\frac{n}{h_n^2}\right)^{1/2} r$. For

[6] If different bandwidths $h_1, \cdots, h_K$ are used, a more extensive adjustment of the relative speed assumptions of n and h_i are necessary, but with no qualitative consequence to the results obtained.

$x \in B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)$,

$$|s_0(x) - s_0(x_k)| \le (nh_n)^{-1}\sum_{t=1}^{n} m|h_n^{-1}(x_k - x)|B_f < B_f m(nh_n^2)^{-1/2} \text{ and}$$

$$|E(s_0(x_k)) - E(s_0(x))| < B_f m(nh_n^2)^{-1/2}.$$

Hence,

$$|s_0(x) - E(s_0(x))| \le |s_0(x_k) - E(s_0(x_k))| + 2B_f m(nh_n^2)^{-1/2} \text{ and}$$

$$sup_{x\in G}|s_0(x) - E(s_0(x))| \le max_{1\le k\le l_n}|s_0(x_k) - E(s(x_k))| + 2B_f m(nh_n^2)^{-1/2}.$$

If $nh_n^2 \to \infty$, then to prove a) it suffices to show that there exists a constant $\Delta > 0$ such that for all $\epsilon > 0$ there exists N such that for all $n > N$,

$$P\left(\frac{nh_n}{ln(n)} max_{1\le k\le l_n}|s_0(x) - E(s_0(x))| \ge \Delta\right) \le \epsilon.$$

Let $\varepsilon_n = \frac{ln(n)}{nh_n}\Delta$. Then, for every n,

$$P\left(max_{1\le k\le l_n}|s_0(x_k) - E(s_0(x_k))| \ge \varepsilon_n\right) \le \sum_{k=1}^{l_n} P\left(|s_0(x_k) - E(s_0(x_k))| \ge \varepsilon_n\right)$$

But $|s_0(x_k) - E(s_0(x_k))| = |\frac{1}{n}\sum_{t=1}^{n} W_{tn}|$ where $W_{tn} = \frac{1}{h_n}K(\frac{X_t - x_k}{h_n})f(X_t, R_t) - \frac{1}{h_n}E\left(K(\frac{X_t - x_k}{h_n})f(X_t, R_t)\right)$ with $E(W_{tn}) = 0$ and $|W_{tn}| \le \frac{2B_K B_f}{h_n} = \frac{B_W}{h_n}$. Since $\{W_{tn}\}_{t=1}^{n}$is an independent sequence, by Bernstein's inequality

$$P\left(|s_0(x_k) - E(s_0(x_k))| \ge \varepsilon_n\right) < 2exp\left(\frac{-nh_n\varepsilon_n^2}{2h_n\bar{\sigma}^2 + \frac{2B_W\varepsilon_n}{3})}\right)$$

where $\bar{\sigma}^2 = n^{-1}\sum_{t=1}^{n} V(W_{tn}) = h_n^{-2}E\left(K^2\left(\frac{X_t - x_k}{h_n}\right)f^2(X_t, R_t)\right) - \left(h_n^{-1}E\left(K\left(\frac{X_t - x_k}{h_n}\right)f(X_t, R_t)\right)\right)^2$. Under assumptions A1 and A3 and the fact that $f(x,r)$ and $g(x,r)$ are continuous in G we have that $h_n\bar{\sigma}^2 \to B_{\bar{\sigma}^2}$ by the proposition of Prakasa-Rao. Hence, for any $n > N$ there exists a constant $B_c > 0$ such that, $2h_n\bar{\sigma}^2 + \frac{2}{3}B_W\varepsilon_n \le B_c\varepsilon_n$. Hence, $\frac{-nh_n\varepsilon_n^2}{2h_n\bar{\sigma}^2 + \frac{2B_W\varepsilon_n}{3}} \le \frac{-nh_n\varepsilon_n^2}{B_c\varepsilon_n} = \frac{-\Delta ln(n)}{B_c}$. Hence, for any $\epsilon > 0$ there exists N such that for all $n > N$,

$$P\left(max_{1\le k\le l_n}|s_0(x_k) - E(s_0(x_k))| \ge \varepsilon_n\right) < 2l_n n^{-\Delta/B_c} < 2\left(\frac{n}{h_n^2}\right)^{-1/2} rn^{-\Delta/B_c} < 2\left(nh_n^2\right)^{1/2} r < \epsilon$$

provided $\Delta > B_c$.

b) As in part a) define a collection of sets $\{B(x_k, h_n^a)\}_{k=1}^{l_n}$ such that $G \subset \cup_{k=1}^{l_n} B(x_k, h_n^a)$ for $x_k \in G$ with $l_n < h_n^{-a} r$ for $a \in (0, \infty)$. By assumption $|s_0(x) - s_0(x_k)| \leq (nh_n)^{-1} \sum_{t=1}^{n} m|h_n^{-1}(x_k - x)|B_f < B_f m h_n^{a-2}$ for $x \in B(x_k, h_n^a)$. Similarly, $|E(s_0(x_k)) - E(s_0(x))| < B_f m h_n^{a-2}$ for $x \in B(x_k, h_n^a)$. Hence, $|s_0(x) - E(s_0(x))| \leq |s_0(x_k) - E(s_0(x_k))| + 2B_f m h_n^{a-2}$ for $x \in B(x_k, h_n^a)$ and

$$sup_{x\in G}|s_0(x) - E(s_0(x))| \leq max_{1\leq k\leq l_n}|s_0(x_k) - E(s(x_k))| + 2B_f m h_n^{a-2}.$$

To show that $lim_{n\to\infty} P\left(sup_{x\in G}|s_0(x) - E(s_0(x))| \geq h_n^p \epsilon\right) = 0$ for $p > 0$ we need $h_n^{a-p-2} \to 0$ as $n \to \infty$ and $lim_{n\to\infty} P\left(max_{1\leq k\leq l_n}|s_0(x) - E(s_0(x))| \geq h_n^p \epsilon\right) = 0$. But

$$P\left(max_{1\leq k\leq l_n}|s_0(x_k) - E(s_0(x_k))| \geq h_n^p \epsilon\right) \leq \sum_{k=1}^{l_n} P\left(|s_0(x_k) - E(s_0(x_k))| \geq h_n^p \epsilon\right)$$

Using Bernstein's inequality as in a), we have

$$P\left(|s_0(x_k) - E(s_0(x_k))| \geq h_n^p \epsilon\right) < 2exp\left(\frac{-nh_n^{2p}\epsilon^2}{2(\bar{\sigma}^2 + B_W \frac{h_n^p \epsilon}{3})}\right).$$

Hence for the desired result the righthand side of the inequality must approach zero as $n \to \infty$. It suffices to show that $\frac{nh_n^{2p}\epsilon^2}{2\bar{\sigma}^2 + 2/3 B_W h_n^p \epsilon} + aln(h_n) \to \infty$, which given that $\bar{\sigma}^2 = O(1)$ will result if $\frac{nh_n^{2p+1}}{ln(h_n)} \to \infty$.

Comment. An important special case of part b) in Lemma 1 occurs when $f(x, r) \equiv 1$ for all x, r and $p = 1$. In this case we have

$$sup_{x\in G}\frac{1}{h_n}\left|(nh_n)^{-1}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^j - E(s_j(x))\right| = o_p(1)$$

for $j = 0, 1, 2$. This result in combination with assumption A4 can be used to show that $s_0(x) - g_X(x) = O_p(h_n)$, $s_1(x) = O_p(h_n)$ and $s_2(x) - g_X(x)\sigma_K^2 = O_p(h_n)$ uniformly in G. These uniform boundedness results are used to prove Lemma 2.

Lemma 2 *Assume A1, A2, A3 and A4. If* $h_n \to 0$ *and* $\frac{nh_n^3}{ln(h_n)} \to \infty$, *then for every* $x \in G$ *the compact set described Lemma 1, we have*

$$\hat{\sigma}^2(x) - \sigma^2(x) = \frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\hat{r}_t - \sigma^2(x) - \sigma^{2(1)}(x)(X_t - x)\right) + O_p(R_{n,1}(x))$$

uniformly in G, where $\hat{r}_t = \sigma^2(X_t)\epsilon_t^2 + (m(X_t) - \hat{m}(X_t))^2 + 2(m(X_t) - \hat{m}(X_t))\sigma(X_t)\epsilon_t$, $\sigma^{2(1)}(x)$ *is the first derivative of* $\sigma^2(x)$, $R_{n,1}(x) = n^{-1}\left(\left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)r_t^*\right| + \left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)r_t^*\right|\right)$ *and* $r_t^* = \hat{r}_t - \sigma^2(x) - \sigma^{2(1)}(x)(X_t - x)$.

Proof [Lemma 2] Let $R_n \equiv (1_n, \vec{x} - 1_n x)$, $P_n \equiv diag\left\{K\left(\frac{X_t-x}{h_n}\right)\right\}_{t=1}^n$, $\hat{r}' = (\hat{r}_1, ..., \hat{r}_n)$ with $\hat{r}_t = \sigma^2(X_t)\epsilon_t^2 + (m(X_t) - \hat{m}(X_t))^2 + 2(m(X_t) - \hat{m}(X_t))\sigma(X_t)\epsilon_t$, $1_n = (1, \cdots, 1)'$, $\vec{x} = (X_1, \cdots, X_n)'$,

$$S_n(x) = (nh_n)^{-1}\begin{pmatrix} \sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right) & \sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{X_t-x}{h_n} \\ \sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{X_t-x}{h_n} & \sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)^2 \end{pmatrix} \text{ and } S(x) = \begin{pmatrix} g_X(x) & 0 \\ 0 & g_X(x)\sigma_K^2 \end{pmatrix}.$$

Then, $\hat{\sigma}^2(x) - \sigma^2(x) = \frac{1}{nh_n}\sum_{t=1}^n W_n\left(\frac{X_t-x}{h_n}, x\right)r_t^*$ where $W_n(z,x) = (1,0)S_n^{-1}(x)(1,z)'K(z)$ and $r_t^* = \hat{r}_t - \sigma^2(x) - \sigma^{2(1)}(x)(X_t - x)$. Let $A_n(x) \equiv \hat{\sigma}^2(x) - \sigma^2(x) - \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)r_t^*$, then

$$\begin{aligned}
|A_n| &= \frac{1}{nh_n}\left|\sum_{t=1}^n\left(W_n\left(\frac{X_t-x}{h_n}, x\right) - \frac{1}{g_X(x)}K\left(\frac{X_t-x}{h_n}\right)\right)r_t^*\right| \\
&= \frac{1}{nh_n}\left|(1,0)(S_n^{-1}(x) - S^{-1}(x))\begin{pmatrix} \sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)r_t^* \\ \sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{X_t-x}{h_n}r_t^* \end{pmatrix}\right| \\
&\leq \frac{1}{h_n}\left((1,0)(S_n^{-1}(x) - S^{-1}(x))^2(1,0)'\right)^{1/2}\frac{1}{n}\left(\left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)r_t^*\right| + \left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{X_t-x}{h_n}r_t^*\right|\right)
\end{aligned}$$

By the comment following Lemma 1, $B_n(x) \equiv \frac{1}{h_n}\left((1,0)(S_n^{-1}(x) - S^{-1}(x))^2(1,0)'\right)^{1/2} = O_p(1)$ uniformly in G. Hence if we put $R_{n,1}(x) \equiv n^{-1}\left(\left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)r_t^*\right| + \left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{X_t-x}{h_n}r_t^*\right|\right)$ and the proof is complete.

Comment. Similar arguments can be used to prove that,

$$\hat{m}(x) - m(x) = \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\left(Y_t - m(x) - m^{(1)}(x)(X_t - x)\right) + O_p(R_{n,2}(x))$$

where $R_{n,2}(x) = n^{-1}\left(\left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)Y_t^*\right| + \left|\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)Y_t^*\right|\right)$ and $Y_t^* = Y_t - m(x) - m^{(1)}(x)(X_t - x)$.

Lemmas 2 and 3 are used to prove Theorem 1, which is the basis for establishing uniform consistency and asymptotic normality of the frontier estimator. Theorem 1 contains two results. The first (a) shows that the difference $\hat{\sigma}^2(x) - \sigma^2(x)$ is $h_n^{3-\delta}$ uniformly asymptotically equivalent to $\frac{1}{nh_n g_X(x)} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*$ in G for $\delta > 0$. Hence, we can investigate the asymptotic properties of $\hat{\sigma}^2(x) - \sigma^2(x)$ by restricting attention to $\frac{1}{nh_n g_X(x)} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*$. The second (b) establishes the asymptotic normality of a suitable normalization of $\frac{1}{nh_n g_X(x)} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*$.

Some of the assumptions in the following theorems are made for convenience on ϵ_t rather than R_t. Since $\epsilon_t = \frac{R_t - \mu_R}{\sigma_R}$ these assumptions have a direct counterpart for R_t. Specifically we have $E(\epsilon_t^4|X_t) = \mu_4(X_t) \Rightarrow E(R_t^4|X_t)$ exists as a function of X_t and $E(|\epsilon_t||X_t) = \mu_1 \Rightarrow E(|R_t - \mu_R||X_t)$ exists as a function of X_t.

Theorem 1 *Suppose that assumptions A1, A2, A3 and A4 are holding. In addition assume that* $E(|\epsilon_t||X_t) = \mu_1$ *for* $X_t \in G$ *a compact subset of* $(0, \infty)$. *If* $h_n \to 0$, $\frac{nh_n^3}{ln(h_n)} \to \infty$ *and* $\frac{nh_n^3}{ln(n)} \to C$ *where* C *is a constant, then for every* $x \in G$

a) $sup_{x \in G} \left| \hat{\sigma}^2(x) - \sigma^2(x) - \frac{1}{nhg_X(x)} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^* \right| = O_p(h_n^3)$

b) if in addition we assume that 1. $E(\epsilon_t^4|X_t = x) = \mu_4(x)$ *is continuous in* $(0, \infty)$ *then*

$$\sqrt{nh_n}\left(\hat{\sigma}^2(x) - \sigma^2(x) - B_{0n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right),$$

for all $x \in G$ *where* $B_{0n} = \frac{h_n^2 \sigma_K^2}{2} \sigma^{2(2)}(x) + o_p(h_n^2)$.

Proof [Theorem 1] (a) Given the upperbound $\bar{B}_{g_X}$ and Lemma 2

$$\begin{aligned}\left|\hat{\sigma}^2(x) - \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*\right| &\leq \bar{B}_{g_X} B_n(x) h_n \left(\frac{1}{nh_n g_X(x)}\left(\left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*\right|\right.\right. \\ &+ \left.\left.\left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right) r_t^*\right|\right)\right) \\ &= \bar{B}_{g_X} B_n(x) h_n \left(|c_1(x)| + |c_2(x)|\right).\end{aligned}$$

Since $B_n(x) = O_p(1)$ uniformly in G from the comment following Lemma 1, it suffices to investigate the order in probability of $|c_1(x)|$ and $|c_2(x)|$. Here, we establish the order of $c_1(x)$ noting that the proof for $c_2(x)$ follows a similar argument given assumption A3. We write $c_1(x) = I_{1n} + I_{2n} - I_{3n} + I_{4n}$ where

$$I_{1n}(x) = \frac{1}{nh_n g_X(x)} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\left(\sigma^2(X_t) - \sigma^2(x) - \sigma^{2(1)}(X_t - x)\right)$$

$$
\begin{aligned}
I_{2n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \sigma^2(X_t)(\epsilon_t^2 - 1) \\
I_{3n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \sigma(X_t)\epsilon_t(\hat{m}(X_t) - m(X_t)) \\
I_{4n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) (m(X_t) - \hat{m}(X_t))^2
\end{aligned}
$$

and examine each term separately. $I_{1n}(x)$: by Taylor's theorem there exists $X_{tb} = \lambda X_t + (1-\lambda)x$ for some $\lambda \in [0,1]$ such that $I_{1n} = \frac{h_n}{2ng_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 \sigma^{2(2)}(X_{tb})$. Given A1.2 and A2.6 we have

$$
\begin{aligned}
sup_{x\in G}|I_{1n}(x)| &\leq \frac{\bar{B}_{2\sigma}\underline{B}_{g_X}^{-1}}{2}\left(h_n^2 sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 - E\left(\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2\right)\right|\right. \\
&+ \left. h_n^2 sup_{x\in G} E\left(\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2\right)\right) \\
&= \frac{\bar{B}_{2\sigma}\underline{B}_{g_X}^{-1}}{2}\left(h_n^3 o_p(1) + h_n^2 O(1)\right) = O_p(h_n^2) \text{ by part b) of Lemma 1 with } p = 1.
\end{aligned}
$$

$I_{2n}(x)$: Note that by assumption A1.2

$$
\begin{aligned}
|I_{2n}(x)| &\leq \underline{B}_{g_X}^{-1} sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2 - 1)\right| \text{ and} \\
sup_{x\in G}|I_{2n}(x)| &\leq \underline{B}_{g_X}^{-1} sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2 - 1)\right| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right)
\end{aligned}
$$

where the last equality follows from part a) in Lemma 1 with $f(X_t, R_t) = \sigma^2(X_t)(\epsilon_t^2 - 1)$, which is bounded in G by assumptions A2.2 and A2.4.

$I_{3n}(x)$: From the comment following Lemma 2 and by Taylor's theorem there exists $X_{kt} = \lambda X_k + (1-\lambda)X_t$ for some $\lambda \in [0,1]$ such that $I_{3n}(x) = I_{31n}(x) + I_{32n}(x) + I_{33n}(x)$, where

$$
\begin{aligned}
I_{31n}(x) &= \frac{2}{n^2 h_n^2 g_X(x)} \sum_{t=1}^{n}\sum_{k=1}^{n} \frac{1}{g_X(X_t)} K\left(\frac{X_t - x}{h_n}\right) K\left(\frac{X_k - X_t}{h_n}\right) \sigma(X_t)\sigma(X_k)\epsilon_t\epsilon_k \\
I_{32n}(x) &= \frac{h_n^2}{n^2 h_n^2 g_X(x)} \sum_{t=1}^{n}\sum_{k=1}^{n} \frac{1}{g_X(X_t)} K\left(\frac{X_t - x}{h_n}\right) K\left(\frac{X_k - X_t}{h_n}\right)\left(\frac{X_k - X_t}{h_n}\right)^2 \sigma(X_t)\epsilon_t m^{(2)}(X_{kt}) \\
I_{33n}(x) &= \frac{2}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)\epsilon_t\left(\hat{m}(X_t) - m(X_t) - \frac{1}{nh_n g_X(X_t)}\sum_{t=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) Y_k^*\right)
\end{aligned}
$$

We now examine each of these terms separately. Note that,

$$|I_{31n}(x)| \leq 2\underline{\mathrm{B}}_{g_X}^{-1}\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| sup_{x\in G}\frac{1}{nh_n}\left|\sum_{k=1}^{n}K\left(\frac{X_k-x}{h_n}\right)\sigma(X_k)\epsilon_k\right|.$$

Since $|\sigma(X_k)\epsilon_k| < C$ for a generic constant C. If $nh_n^2 \to \infty$ we have by part a) of Lemma 1,

$$sup_{x\in G}\frac{1}{nh_n}\left|\sum_{k=1}^{n}K\left(\frac{X_k-x}{h_n}\right)\sigma(X_k)\epsilon_k\right| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right).$$

Therefore,

$$sup_{x\in G}|I_{31n}(x)| \leq 2\underline{\mathrm{B}}_{g_X}^{-1}O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right)sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right|$$

Since $\left|\frac{\sigma(X_t)|\epsilon_t|}{g_X(X_t)}\right| < C$,

$$\begin{aligned}
& sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right| \leq sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| - \right.\right.\\
& \left.\left. E\left(\frac{1}{g_X(X_t)}h_n^{-1}K(\frac{X_t-x}{h_n})\sigma(X_t)|\epsilon_t|\right)\right)\right| + sup_{x\in G}E\left(\frac{1}{g_X(X_t)}\frac{1}{h_n}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right)\\
= \; & O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) + \frac{1}{h_n}sup_{x\in G}E\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right),
\end{aligned}$$

by part a) of Lemma 1. Now, $\frac{1}{h_n}E\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right) = \int K(\phi)\sigma(x+h_n\phi)\mu_1 d\phi$ and by the proposition in Prakasa-Rao,

$$\frac{1}{h_n}sup_{x\in G}E\left(\frac{1}{g_X(X_t)}K(\frac{X_t-x}{h_n})\sigma(X_t)|\epsilon_t|\right) \leq \mu_1\int K(\phi)d\phi sup_{x\in G}\sigma(x) \leq C$$

given assumption A2.4 and $E\left(|\epsilon_t||X_t\right) = \mu_1$. Therefore, $\frac{1}{h_n}sup_{x\in G}E\left(\frac{1}{g_X(X_t)}K(\frac{X_t-x}{h_n})\sigma(X_t)|\epsilon_t|\right) = O(1)$ and consequently $sup_{x\in G}|I_{31n}(x)| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right)$.

Now,by assumptions A2.1 and A2.6

$$|I_{32n}(x)| \leq \underline{\mathrm{B}}_{g_X}^{-1}bB_{2\sigma}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| sup_{x\in G}\left|\frac{1}{n}\sum_{k=1}^{n}K\left(\frac{X_k-x}{h_n}\right)\left(\frac{X_k-x}{h_n}\right)^2\right|.$$

From the analysis of I_{1n}, $sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)^2\right| = O_p(h_n)$ and by using part b) of Lemma 1 $sup_{x\in G}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| = O_p(h_n)$, which gives $sup_{x\in G}|I_{32n}| = O_p(h_n^2)$. From the comments following Lemma 2

$$D_n(X_t) \equiv \left|\hat{m}(X_t) - m(X_t) - \frac{1}{nh_ng_X(x)}\sum_{k=1}^{n}K\left(\frac{X_t-x}{h_n}\right)Y_k^*\right| \leq B_n(X_t)R_{n,2}(X_t),$$

hence $|I_{33n}(x)| \leq O_p(1)\frac{2}{nh_n g_X(x)}\sum_{k=1}^n K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|R_{n,2}(X_t)$. Now, we can write

$$R_{n,2}(X_t) \leq |R_{11}(X_t)| + |R_{12}(X_t)| + |R_{21}(X_t)| + |R_{22}(X_t)|,$$

where $R_{11}(X_t) = \frac{1}{n}\sum_{k=1}^n K\left(\frac{X_k-X_t}{h_n}\right)\sigma(X_k)\epsilon_k$, $R_{12}(X_t) = \frac{1}{2n}\sum_{k=1}^n K\left(\frac{X_k-X_t}{h_n}\right)(X_k - X_t)^2 m^{(2)}(X_{kt})$, $R_{21}(X_t) = \frac{1}{n}\sum_{k=1}^n K\left(\frac{X_k-X_t}{h_n}\right)\left(\frac{X_k-X_t}{h_n}\right)^2\sigma(X_k)\epsilon_k$ and $R_{22}(X_t) = \frac{h_n^2}{2n}\sum_{k=1}^n K\left(\frac{X_k-X_t}{h_n}\right)\left(\frac{X_k-X_t}{h_n}\right)^3 m^{(2)}(X_{kt})$. By part b) of Lemma 1 $sup_{X_t\in G}|R_{11}(X_t)| = o_p(h_n^2)$ and by the analysis of I_{32n} we have that $sup_{X_t\in G}|R_{12}(X_t)| = O_p(h_n^3)$. Again by Lemma 1 and the fact that $E(\epsilon_t|X_t) = 0$ we have that $sup_{X_t\in G}|R_{21}(X_t)| = o_p(h_n^2)$. Finally, given that K is defined on a bounded support, by Lemma 1 and A2.6 we obtain $sup_{X_t\in G}|R_{22}(X_t)| = O_p(h_n^3)$. Hence, $sup_{X_t\in G}R_{n,2}(X_t) = o_p(h_n^2)$ and

$$|I_{33n}(x)| \leq 2\underline{\mathrm{B}}_{g_x}^{-1}O_p(1)o_p(h_n^2)\frac{1}{nh_n}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| = 2\underline{\mathrm{B}}_{g_x}^{-1}o_p(h_n^2)I_{331n}.$$

By Lemma 1, $sup_{x\in G}I_{331n} = o_p(h_n) + O(1)$ and therefore $sup_{x\in G}|I_{33n}| = o_p(h_n^2)$. Combining all results we have $sup_{x\in G}|I_{3n}| = O_p(h_n^2) + O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right)$.

$I_{4n}(x)$: We write $I_{4n} = I_{41n}(x) + I_{42n}(x) + I_{43n}(x) + I_{44n}(x) + I_{45n}(x) + I_{46n}(x)$ where

$$\begin{aligned}
I_{41n}(x) &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{1}{n^2h_n^2g_X^2(X_t)}\sum_{k=1}^n\sum_{l=1}^n K\left(\frac{X_k-X_t}{h_n}\right)K\left(\frac{X_l-X_t}{h_n}\right)\sigma(X_t)\sigma(X_l)\epsilon_k\epsilon_l\\
I_{42n}(x) &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{1}{4n^2h_n^2g_X^2(X_t)}\sum_{k=1}^n\sum_{l=1}^n K\left(\frac{X_k-X_t}{h_n}\right)(X_k-X_t)^2K\left(\frac{X_l-X_t}{h_n}\right)\\
&\times\ (X_l-X_t)^2m^{(2)}(X_{kt})m^{(2)}(X_{lt})\\
I_{43n}(x) &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)D_n^2(X_t)\\
I_{44n}(x) &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{1}{n^2h_n^2g_X^2(X_t)}\sum_{k=1}^n\sum_{l=1}^n K\left(\frac{X_k-X_t}{h_n}\right)K\left(\frac{X_l-X_t}{h_n}\right)(X_l-X_t)^2\\
&\times\ m^{(2)}(X_{lt})\sigma(X_k)\epsilon_k\\
I_{45n}(x) &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{2D_n(X_t)}{nh_ng_X(X_t)}\sum_{k=1}^n K\left(\frac{X_k-X_t}{h_n}\right)\sigma(X_k)\epsilon_k\\
I_{46n}(x) &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\frac{D_n(X_t)}{nh_ng_X(X_t)}\sum_{k=1}^n K\left(\frac{X_k-X_t}{h_n}\right)m^{(2)}(X_{kt})(X_k-X_t)^2
\end{aligned}$$

where $D_n(X_t) = \hat{m}(X_t) - m(X_t) - \frac{1}{nh_ng_X(X_t)}\sum_{k=1}^n K\left(\frac{X_k-X_t}{h_n}\right)Y_k^*$ where Y_k^* is defined as in the comment following Lemma 2. We now examine each term separately. First,

$$
\begin{aligned}
I_{41n}(x) &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\left(\frac{1}{nhg_X(X_t)}\sum_{l=1}^{n} K\left(\frac{X_l-X_t}{h_n}\right)\sigma(X_l)\epsilon_l\right)^2 \\
&= \frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)(I_{411}(X_t))^2
\end{aligned}
$$

where $I_{411}(X_t) = \frac{1}{nhg_X(X_t)}\sum_{l=1}^{n} K\left(\frac{X_l-X_t}{h_n}\right)\sigma(X_l)\epsilon_l$. But,

$$
\begin{aligned}
sup_{X_t\in G}|I_{411}(X_t)| &\leq \underline{\mathrm{B}}_{g_X}^{-1} h_n \frac{1}{h_n} sup_{X_t\in G}\left|\frac{1}{nh_n}\sum_{l=1}^{n} K\left(\frac{X_l-x}{h_n}\right)\sigma(X_l)\epsilon_l\right| \\
&= \underline{\mathrm{B}}_{g_X}^{-1} h_n o_p(1) \text{ by part b) of Lemma 1.}
\end{aligned}
$$

Hence, $sup_{X_t\in G}|I_{411}(X_t)| = o_p(h_n)$ and $sup_{X_t\in G}(I_{411})^2 = o_p(h_n^2)$ and

$$
sup_{X_t\in G}|I_{41n}(x)| \leq o_p(h_n^2) sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\right| = o_p(h_n^2).
$$

Now,

$$
\begin{aligned}
|I_{42n}(x)| &= \frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\left(\frac{1}{2nh_n g_X(X_t)}\sum_{t=1}^{n} K\left(\frac{X_k-X_t}{h_n}\right)m^{(2)}(X_{kt})(X_k-X_t)^2\right)^2 \\
&= \frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)(I_{421}(X_t))^2
\end{aligned}
$$

where $I_{421}(X_t) = \frac{1}{2nh_n g_X(X_t)}\sum_{t=1}^{n} K\left(\frac{X_k-X_t}{h_n}\right)m^{(2)}(X_{kt})(X_k-X_t)^2$. But $|I_{421}(X_t)| \leq \underline{\mathrm{B}}_{g_X}^{-1} h_n^{-1}|R_{12}(X_t)|$ and since $sup_{X_t\in G}|R_{12}(X_t)| = O_p(h_n^3)$ from above, we have that $sup_{X_t\in G}(I_{421}(X_t))^2 = O_p(h_n^4)$. Since $\frac{1}{nh_n}\left|\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\right| = O_p(1)$ we have $sup_{x\in G}|I_{42n}| = O_p(h_n^4)$.

For the $I_{43n}(x)$ we first observe that from our analysis of I_{33n} we have that $sup_{X_t\in G}|D_n(X_t)| = o_p(h_n^2)$ hence $|I_{43n}(x)| \leq \underline{\mathrm{B}}_{g_x}^{-1} o_p(h_n^4)\left|\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\right|$ and consequently $sup_{x\in G}|I_{43}(x)| = o_p(h_n^4)$ since $\frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right) = O_p(1)$ uniformly in G.

Now,

$$
|I_{44n}(x)| \leq \left|\frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\right| sup_{X_t\in G}|I_{441}(X_t)| sup_{X_t\in G}|I_{442}(X_t)|, \text{ where}
$$

$I_{441}(X_t) = \frac{1}{nh_n g_X(X_t)}\sum_{k=1}^{n} K\left(\frac{X_k-X_t}{h_n}\right)\sigma(X_k)\epsilon_k$, $I_{442}(X_t) = \frac{1}{nh_n g_X(X_t)}\sum_{k=1}^{n} K\left(\frac{X_k-X_t}{h_n}\right)(X_k-X_t)^2 m^{(2)}(X_{kt})$.

But given that $sup_{X_t\in G} I_{441}(X_t) \leq \underline{\mathrm{B}}_{g_X}^{-1} h_n^{-1} sup_{X_t\in G}|R_{11}(X_t)| = o_p(h_n)$ and

$$
sup_{X_t\in G} I_{442}(X_t) \leq 2\underline{\mathrm{B}}_{g_X}^{-1} h_n^{-1} sup_{X_t\in G}|R_{12}(X_t)| = O_p(h_n^2),
$$

we have $sup_{x\in G} I_{44n}(x) = o_p(h_n^3)$. Finally,

$$|I_{45n}(x)| \leq \frac{2}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) sup_{X_t\in G}|D_n(X_t)| sup_{X_t\in G}|I_{441}(X_t)|$$

which implies from above that $sup_{x\in G}|I_{45n}(x)| = o_p(h_n^3)$ and

$$|I_{46n}(x)| \leq \frac{2}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) sup_{X_t\in G}|D_n(X_t)| sup_{X_t\in G}|I_{421}(X_t)|$$

which from above gives $sup_{x\in G}|I_{46n}(x)| = o_p(h_n^4)$, hence $sup_{x\in G}|I_{4n}| = o_p(h_n^2)$. Combining all terms we have that $sup_{x\in G}|c_1(x)| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) + O_p(h_n^2)$ and also $sup_{x\in G}|c_2(x)| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) + O_p(h_n^2)$. Provided that $\frac{nh_n^3}{ln(n)} \to C$ for some constant C we have that $sup_{x\in G}|c_1(x)|, sup_{x\in G}|c_2(x)| = O_p(h_n^2)$. Consequently,

$$\left|\hat{\sigma}^2(x) - \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) r_t^*\right| \leq O_p(h_n^3).$$

(b) From part a) $I_{1n}(x) = \frac{1}{2}\frac{h_n}{n}\frac{1}{g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)^2 \sigma^{2(2)}(X_{tb})$, and given A1,

$$E\left(\frac{I_{1n}(x)}{h_n^2}\right) = \frac{1}{2g_X(x)} \int \phi^2 K(\phi) \sigma^{2(2)}(x + h_n\theta\phi) g_X(x + h_n\phi) d\phi \text{ and}$$

$$V\left(\frac{I_{1n}(x)}{h_n^2}\right) = \frac{1}{4g_X(x)^2}\left(\frac{1}{nh_n^2} E\left(K^2\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^4 (\sigma^{2(2)}(X_{tb}))^2\right) - \frac{1}{n}\left(\frac{1}{h_n} E\left(K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 \sigma^{2(2)}(X_{tb})\right)\right)^2\right)$$

for $|\theta| \leq 1$. Given assumptions A1,A2.5 and A3 and by the proposition of Prakasa-Rao,

$$E\left(\frac{I_{1n}(x)}{h_n^2}\right) \to \frac{1}{2}\sigma^{2(2)}(x)\sigma_K^2 \text{ and } V\left(\frac{I_{1n}(x)}{h_n^2}\right) \to 0$$

hence by Chebyshev's inequality $\frac{I_{1n}(x)}{h_n^2} - \frac{1}{2}\sigma^{2(2)}(x)\sigma_K^2 = o_p(1)$.

We now establish that $\sqrt{nh} I_{2n} \xrightarrow{d} N\left(0, \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right)$. To this end, note that

$$\sqrt{nh} I_{2n} = \sum_{t=1}^{n} \frac{1}{\sqrt{nh} g_X(x)} K\left(\frac{X_t - x}{h_n}\right) \sigma^2(X_t)(\epsilon_t^2 - 1) = \sum_{t=1}^{n} Z_{tn}$$

where $\{Z_{tn} : t = 1, \cdots, n; n = 1, 2, \cdots\}$ forms an independent triangular array with $E(Z_{tn}) = 0$ and

$$\begin{aligned} s_n^2 = \sum_{t=1}^{n} E(Z_{tn}^2) &= \frac{1}{nh_n g_X^2(x)} \sum_{t=1}^{n} E\left(K^2\left(\frac{X_t - x}{h_n}\right)\sigma^4(X_t)(\epsilon_t^2 - 1)^2\right) \\ &= \frac{1}{h_n g_X^2(x)} E\left(K^2\left(\frac{X_t - x}{h_n}\right)\sigma^4(X_t)(\mu_4(X_t) - 1)\right) \end{aligned}$$

where $\mu_4(X_t) = E\left(\epsilon_t^4|X_t\right)$. By the proposition of Prakasa-Rao and the continuity of $\mu_4(X_t)$, $s_n^2 \rightarrow \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x)-1)\int K^2(\phi)d\phi$. By Liapounov's central limit theorem $\sum_{t=1}^n \frac{Z_{tn}}{s_n} \xrightarrow{d} N(0,1)$ provided that $lim_{n\rightarrow\infty}\sum_{t=1}^n E\left|\frac{Z_{tn}}{s_n}\right|^{2+\delta} = 0$ for some $\delta > 0$. Now,

$$\begin{aligned}\sum_{t=1}^n E\left|\frac{Z_{tn}}{s_n}\right|^{2+\delta} &= (s_n^2)^{-1-\delta/2}\sum_{t=1}^n E\left|Z_{tn}\right|^{2+\delta} \\ &= (s_n^2)^{-1-\delta/2}\frac{g_X(x)^{-2-\delta}}{(nh_n)^{\delta/2}}\frac{1}{h_n}E\left|K\left(\frac{X_t-x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2-1)\right|^{2+\delta}\end{aligned}$$

But,

$$\begin{aligned}\frac{1}{h_n}E\left|K\left(\frac{X_t-x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2-1)\right|^{2+\delta} &= \frac{1}{h_n}E\left(K^{2+\delta}\left(\frac{X_t-x}{h_n}\right)(\sigma^2(X_t))^{2+\delta}E(|\epsilon_t^2-1|^{2+\delta}|X_t)\right) \\ &\leq \frac{C}{h_n}E\left(K^{2+\delta}\left(\frac{X_t-x}{h_n}\right)\right) \rightarrow Cg_X(x)\int K^{2+\delta}(x)dx\end{aligned}$$

where the inequality follows from the existence of $\mu_4(X_t)$, A1, A2.4 and A3.

We now examine $I_{3n}(x)$. As in part a) we write $I_{3n}(x) = I_{31n}(x) + I_{32n}(x) + I_{33n}(x)$ and look at each term separately. Using the notation of Lemma 3 in the appendix,

$$I_{31n}(x) = \frac{2K(0)}{n^2h_n^2g_X(x)}\sum_{t=1}^n K\left(\frac{X_t-x}{h_n}\right)\sigma^2(X_t)\frac{\epsilon_t^2}{g_X(X_t)} + \frac{n-1}{n}\left(\begin{array}{c}n\\2\end{array}\right)^{-1}\sum_{t<k}\psi_n(Z_t,Z_k) = I_{311} + \frac{n-1}{n}I_{312}$$

where, $\psi_n(Z_t,Z_k) = h_{tk} + h_{kt}$, $h_{tk} = \frac{1}{g_X(x)h_n^2}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)K\left(\frac{X_k-X_t}{h_n}\right)\sigma(X_t)\sigma(X_k)\epsilon_t\epsilon_k$, $Z_t = (X_t,\epsilon_t)$. Given our assumptions,

$$\begin{aligned}E\left(\sqrt{nh_n}I_{311}\right) &= \frac{2K(0)}{\sqrt{nh_n}g_X(x)}\frac{1}{h_n}\int K\left(\frac{z-x}{h_n}\right)\frac{\sigma^2(z)}{g_X(z)}g_X(z)dz, \\ E_{g_X}\left(V\left(\sqrt{nh_n}I_{311}|\vec{x}\right)\right) &= \frac{4K(0)}{n^2h_n^2g_X^2(x)}\frac{1}{h_n}\int K^2\left(\frac{z-x}{h_n}\right)\frac{\sigma^4(z)}{g_X^2(z)}(\mu^4(z)-1)g_X(z)dz\end{aligned}$$

and

$$\begin{aligned}V_{g_X}\left(E\left(\sqrt{nh_n}I_{311}|\vec{x}\right)\right) &= \frac{4K^2(0)}{nh_ng_X^2(x)}\left(\frac{1}{nh_n}\frac{1}{h_n}\int K^2\left(\frac{z-x}{h_n}\right)\frac{\sigma^4(z)}{g_X^2(z)}g_X(z)dz\right. \\ &- \left.\frac{1}{n}\left(\frac{1}{h_n}\int K\left(\frac{z-x}{h_n}\right)\frac{\sigma^2(z)}{g_X(z)}g_X(z)dz\right)^2\right).\end{aligned}$$

Since, $V\left(\sqrt{nh_n}I_{311}\right) = E_{g_X}\left(V\left(\sqrt{nh_n}I_{311}|\vec{x}\right)\right) + V_{g_X}\left(E\left(\sqrt{nh_n}I_{311}|\vec{x}\right)\right)$, provided that $nh_n \rightarrow \infty$ a direct application of the proposition of Prakasa-Rao gives, $E\left(\sqrt{nh_n}I_{311}\right), V\left(\sqrt{nh_n}I_{311}\right) \rightarrow 0$ and consequently by Chebyshev's inequality we have $I_{311} = o_p((nh_n)^{-1/2})$. Given our assumptions it is easily verified that $E(\psi_n(Z_t, Z_k)) = 0$ and $\psi_{1n}(Z_t) = 0$. Hence, by direct use of Lemma 3, we have $\sqrt{n} I_{312} = o_p(1)$ provided that $E(\psi_n^2(Z_t, Z_k)) = o(n)$. We now turn to verifying that $E(\psi_n^2(Z_t, Z_k)) = o(n)$. Note that,

$$\begin{aligned}
\frac{1}{n}E(\psi_n^2(Z_t, Z_k)) &= \frac{1}{ng_X^2(x)h_n^4}E\left(K^2\left(\frac{X_t - X_j}{h_n}\right)K^2\left(\frac{X_t - x}{h_n}\right)\sigma^2(X_t)\sigma^2(X_j)\epsilon_t^2\epsilon_j^2\frac{1}{g_X^2(X_t)}\right) \\
&+ \frac{1}{ng_X^2(x)h_n^4}E\left(K^2\left(\frac{X_t - X_j}{h_n}\right)K^2\left(\frac{X_j - x}{h_n}\right)\sigma^2(X_t)\sigma^2(X_j)\epsilon_t^2\epsilon_j^2\frac{1}{g_X^2(X_j)}\right) \\
&+ \frac{2}{ng_X^2(x)h_n^4}E\left(K^2\left(\frac{X_t - X_j}{h_n}\right)K\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_j - x}{h_n}\right)\sigma^2(X_t)\sigma^2(X_j)\epsilon_t^2\epsilon_j^2\frac{1}{g_X(X_j)g_X(X_t)}\right) \\
&= U_1 + U_2 + U_3
\end{aligned}$$

We focus on the first term - U_1. Since, $t \neq j$ we have that,

$$E(U_1|\vec{x}) = \frac{1}{ng_X^2(x)h_n^4}K^2\left(\frac{X_t - X_j}{h_n}\right)\sigma^2(X_t)\sigma^2(X_j)K^2\left(\frac{X_t - x}{h_n}\right)\frac{1}{g_X^2(X_t)} \text{ and}$$

$$E(U_1) = \frac{1}{ng_X^2(x)h_n^4}\int\int K^2\left(\frac{X_t - X_j}{h_n}\right)\sigma^2(X_t)\sigma^2(X_j)K^2\left(\frac{X_t - x}{h_n}\right)\frac{1}{g_X^2(X_t)}g_X(X_t)g_X(X_j)dX_t dX_j.$$

Given our assumptions, if $nh_n^2 \to \infty$, by Lebesgue's Dominated Convergence theorem we have $E(U_1) \to 0$. We omit the analysis of U_2 and U_3 which can be treated similarly. Hence, combining the results on I_{311} and I_{312} we have that $\sqrt{nh_n} I_{31n} = o_p(1)$. Now we turn to the analysis of $I_{32n}(x)$. Using the notation of Lemma 3 we have $I_{32n}(x) = \frac{n-1}{2n}\frac{1}{g_X(x)}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{t<k}\psi_n(Z_t, Z_k)$ where $\psi_n(Z_t, Z_k) = h_{tk} + h_{kt}$ and

$$h_{tk} = K\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_t - X_k}{h_n}\right)\left(\frac{X_t - X_k}{h_n}\right)^2 m^{(2)}(X_{tk})\frac{\sigma(X_t)\epsilon_t}{g_X(X_t)}$$

and $Z_t = (X_t, \epsilon_t)$. Given our assumptions $E(\psi_n(Z_t, Z_k)) = 0$ and

$$\psi_{1n}(Z_t) = K\left(\frac{X_t - x}{h_n}\right)\frac{\sigma(X_t)\epsilon_t}{g_X(X_t)}E\left(K\left(\frac{X_t - X_k}{h_n}\right)\left(\frac{X_t - X_k}{h_n}\right)^2 m^{(2)}(X_{tk})|Z_t\right)$$

Hence, using the notation in Lemma 3, $\sqrt{n}\hat{u}_n = \frac{2}{\sqrt{n}}\sum_{t=1}^n \phi_{1n}(Z_t)$, with $E(\sqrt{n}\hat{u}_n) = 0$ and

$$V(\sqrt{n}\hat{u}_n) = 4E\left(K^2\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_t - X_k}{h_n}\right)\left(\frac{X_t - X_k}{h_n}\right)^2\frac{\sigma^2(X_t)}{g_X^2(X_t)}m^{(2)}(X_{tk}) \times \right.$$
$$\left. K\left(\frac{X_t - X_l}{h_n}\right)\left(\frac{X_t - X_l}{h_n}\right)^2 m^{(2)}(X_{tl})\right)$$

Using the proposition of Prakasa-Rao we have $V(\sqrt{n}\hat{u}_n) \to 0$ and consequently by Lemma 3, $\sqrt{n}I_{32n} = o_p(1)$ provided that $E(\psi_n^2(Z_t, Z_k)) = o(n)$. Now,

$$\frac{1}{n}E(\psi_n^2(Z_t,Z_j)) = \frac{1}{4nh_n^4}\int\int K^2\left(\frac{X_t-X_j}{h_n}\right)K^2\left(\frac{X_t-x}{h_n}\right)(X_t-X_j)^4\frac{\sigma^2(X_t)\epsilon_t^2 m^{(2)2}(X_{tj})}{g_X^2(X_t)}g_X(X_t)g_X(X_j)dX_tdX_j$$
$$+\frac{1}{4nh_n^4}\int\int K^2\left(\frac{X_t-X_j}{h_n}\right)K^2\left(\frac{X_j-x}{h_n}\right)(X_t-X_j)^4\frac{\sigma^2(X_j)\epsilon_j^2 m^{(2)2}(X_{tj})}{g_X^2(X_j)}g_X(X_t)g_X(X_j)dX_tdX_j+$$
$$\frac{2}{4nh_n^4}\int\int K^2\left(\frac{X_t-X_j}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)K\left(\frac{X_t-x}{h_n}\right)(X_t-X_j)^4\frac{\sigma(X_t)\sigma(X_j)\epsilon_t\epsilon_j m^{(2)}(X_{tj})m^{(2)}(X_{tj})}{g_X(X_t)g_X(X_j)}\times$$
$$g_X(X_t)g_X(X_j)dX_tdX_j = U_1+U_2+U_3$$

Given our assumptions, a direct application of Lebesgue's dominated convergence theorem gives $U_1, U_2, U_3 \to 0$. Since from part a) $I_{33n} = o_p(h_n^2)$ we have that by combining all terms $I_{3n}(x) = o_p(n^{-1/2}) + o_p(h_n^2)$. Finally, since we have already established in part a) that $I_{4n}(x) = o_p(h_n^2)$, combining all convergence results for $I_{1n}(x), I_{2n}(x), I_{3n}(x)$ and $I_{4n}(x)$ we have,

$$\sqrt{nh_n}\left(\hat{\sigma}^2(x)-\sigma^2(x)-B_{0n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x)-1)\int K^2(y)dy\right),$$

for all $x \in G$ where $B_{0n} = \frac{h_n^2\sigma_K^2}{2}\sigma^{2(2)}(x) + o_p(h_n^2)$, which completes the proof.

It is a direct consequence of part a) in Theorem 1 that $sup_{x\in G}\left|\hat{\sigma}^2(x)-\sigma^2(x)\right| = O_p(h_n^2)$ which implies that $\hat{\sigma}^2(x)-\sigma^2(x) \xrightarrow{p} 0$ uniformly G. We now use the former result to show that $\hat{\sigma}(x)-\sigma(x) = O_p(h_n^2)$ uniformly in G and consequently obtain $\hat{\sigma}(x)-\sigma(x) = o_p(1)$ uniformly in G.

Corollary 1 *Assume A1, A2, A3, A4 and that* $h_n \to 0$, $\frac{nh^3}{ln(h_n)} \to \infty$ *and* $\hat{\sigma}^2(x)-\sigma^2(x) = O_p(h_n^2)$ *uniformly in* G *a compact subset of* $(0,\infty)$. *Then, for all* $\epsilon, \delta > 0$ *there exists* $N_{\epsilon,\delta}$ *such that for* $n > N_{\epsilon,\delta}$, $P\left(\left\{\underline{B}_\sigma^2 > inf_{x\in G}|\hat{\sigma}^2(x)|\right\}\right) < \delta$ *and* $\hat{\sigma}(x)-\sigma(x) = O_p(h_n^2)$ *uniformly in* G

Proof [Corollary 1] Fix $\epsilon, \delta > 0$. Then for all $x \in G$ $|\hat{\sigma}^2(x)| \leq |\hat{\sigma}^2(x)-\sigma^2(x)| + \bar{B}_\sigma$. Therefore, $sup_G|\hat{\sigma}^2(x)| \leq sup_G|\hat{\sigma}^2(x)-\sigma^2(x)| + \bar{B}_\sigma$ and $P\left(\epsilon+\bar{B}_\sigma < sup_G|\hat{\sigma}^2(x)|\right) \leq P\left(sup_G|\hat{\sigma}^2(x)-\sigma^2(x)| > \epsilon\right) < \delta$ for $n > N_{\epsilon,\delta}$. Also, for all $x \in G$ $\underline{B}_\sigma^2 - |\hat{\sigma}^2(x)| \leq |\sigma^2(x)-\hat{\sigma}^2(x)|$ and $\underline{B}_\sigma^2 - inf_G|\hat{\sigma}^2(x)| \leq sup_G|\sigma^2(x)-\hat{\sigma}^2(x)|$ which gives

$$P\left(\left\{inf_G|\hat{\sigma}^2(x)| < \underline{B}_\sigma^2 - \epsilon\right\}\right) \leq P\left(\left\{sup_G|\sigma^2(x)-\hat{\sigma}^2(x)| > \epsilon\right\}\right) < \delta.$$

for $n > N_{\epsilon,\delta}$. By the mean value theorem and A2, there exists $\sigma_b^2(x) = \theta\sigma^2(x) + (1-\theta)\hat{\sigma}^2(x)$ for some $0 \leq \theta \leq 1$ and $\forall x \in G$ such that

$$\begin{aligned} sup_G|\hat{\sigma}(x) - \sigma(x)| &= \frac{1}{2} sup_G \left| \frac{1}{\sqrt{\sigma_b^2(x)}} \right| sup_G|\hat{\sigma}^2(x) - \sigma^2(x)| \\ &= \frac{1}{2} \left(inf_G|\sigma_b^2(x)| \right)^{-1/2} sup_G|\hat{\sigma}^2(x) - \sigma^2(x)| \end{aligned}$$

Note that $inf_G|\sigma_b^2(x)| \geq \theta\underline{\mathrm{B}}_\sigma^2 + (1-\theta) inf_G|\hat{\sigma}^2(x)|$ and therefore

$$P\left(\left\{\underline{\mathrm{B}}_\sigma^2 > inf_G|\sigma_b^2(x)|\right\}\right) \leq P\left(\left\{\underline{\mathrm{B}}_\sigma^2 > inf_G|\hat{\sigma}^2(x)|\right\}\right) < \delta$$

for $n > N_{\epsilon,\delta}$. Hence, $\sigma_b^2(x)^{-1/2} = O_p(1)$ uniformly in G which combines with $\hat{\sigma}^2(x) - \sigma^2(x) = O_p(h_n^2)$ uniformly in G from the comment following Theorem 1 to give $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$ uniformly in G.

The asymptotic normality of $\hat{\sigma}(x)$ is easily obtained from part b) of Theorem 1 by noting that

$$\sqrt{nh_n}\left(\hat{\sigma}(x) - \sigma(x) - \frac{1}{2\sigma(x)}B_{0n} + \left(\frac{1}{2\sigma(x)} - \frac{1}{2\sigma_b(x)}\right)B_{0n}\right) = \frac{1}{2\sqrt{\sigma_b^2(x)}}\sqrt{nh_n}(\hat{\sigma}^2(x) - \sigma^2(x) - B_{0n})$$

and given the uniform consistency of $\hat{\sigma}(x)$ in G from the corollary we have that,

$$\sqrt{nh_n}\left(\hat{\sigma}(x) - \sigma(x) - \frac{1}{2\sigma(x)}B_{0n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{4g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right)$$

The results in Theorem 1 and its corollary refer to the estimator $\hat{\sigma}(x)$, but since our main interest lies on $\hat{\rho}(x) \equiv \frac{\hat{\sigma}}{s_R}$, a complete characterization of the asymptotic behavior of the frontier estimator requires that we provide convergence results on s_R. Theorem 2 below shows that given that $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$, we are able to show that $s_R - \sigma_R = O_p(h_n^2)$ provided that $max_{1 \leq t \leq n} R_t$ converges to 1 sufficiently fast. It should be noted that the required speed of convergence on $max_{1 \leq t \leq n} R_t$ is not necessary to establish the consistency of s_R, which results directly from $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$. As made explicit below, its use is necessary only in obtaining asymptotic distributional results on $\hat{\rho}(x)$.

Theorem 2 *Suppose that (1) $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$ uniformly in G and that (2) for all $\delta > 0$ there exists a constant $\Delta > 0$ such that for all $n > N_\delta$ we have that $P\left(max_{1 \leq t \leq n} R_t > 1 - h_n^2\Delta\right) > 1 - \delta$. Then, $s_R - \sigma_R = O_p(h_n^2)$.*

Proof [Proof of Theorem 2] We start by noting that $|s_R - \sigma_R| = s_R\sigma_R|s_R^{-1} - \sigma_R^{-1}|$. By Corollary 1 $(sup_{X_t\in G}\hat{\sigma}(X_t))^{-1} = O_p(1)$, hence by definition $s_R \leq (sup_{X_t\in G}\hat{\sigma}(X_t))^{-1}(max_{1\leq t\leq n}Y_t)^{-1} = O_p(1)$. Hence, to obtain the desired result it suffices to show that $s_R^{-1} - \sigma_R^{-1} = O_p(h_n^2)$. Since, $|s_R^{-1} - \sigma_R^{-1}| = \sigma_R^{-1}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1|$ we need only show that $max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1 = O_p(h_n^2)$. Note that for some $\Delta', \Delta > 0$,

$$P\left(h_n^{-2}sup_{X_t\in G}\left|\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} - 1\right| < \Delta\right) \geq P\left(h_n^{-2}sup_{X_t\in G}|\sigma(X_t) - \hat{\sigma}(X_t)| < \Delta'\right).$$

Therefore, given supposition (1) in the statement of the theorem, for all $\delta > 0$ there exists $\Delta > 0$ such that for all $n > N_\delta$,

$$P\left(h_n^{-2}sup_{X_t\in G}\left|\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} - 1\right| < \Delta\right) > 1 - \delta. \quad (3)$$

Now suppose that $max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1 \geq 0$. Then,

$$\begin{aligned} |max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1| &= max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1 \\ &\leq max_{1\leq t\leq n}R_t sup_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} - 1 \\ &\leq sup_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} - 1 \end{aligned}$$

Hence, $h_n^{-2}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1| \leq h_n^{-2}|sup_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} - 1|$ and by inequality (3)

$$P\left(h_n^{-2}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1| < \Delta\right) > 1 - \delta$$

Now suppose that $max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1 < 0$. Then,

$$\begin{aligned} |max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1| &= 1 - max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} \\ &\leq max_{1\leq t\leq n}R_t inf_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} \end{aligned}$$

and

$$P\left(h_n^{-2}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1| < \Delta\right) \geq P\left(max_{1\leq t\leq n}R_t inf_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} > 1 - h_n^2\Delta\right).$$

By inequality (3) and assumption (2) in the statement of the theorem, for all $\delta > 0$ there is some $\Delta_1, \Delta > 0$ such that whenever $n > N_\delta$, $P\left(inf_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} > 1 - h_n^2\Delta\right) > 1-\delta$ and $P\left(max_{1\leq t\leq n}R_t > 1 - h_n^2\Delta_1\right) > 1-\delta$. Hence, for all $\delta > 0$ there is some $\Delta_2 > 0$ such that whenever $n > N_\delta$

$$P\left(max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} > 1 - h_n^2\Delta_2\right) > 1 - \delta.$$

which completes the proof.

Since, $\frac{\hat{\sigma}(x)}{s_R} - \frac{\sigma(x)}{\sigma_R} = \hat{\sigma}(x)\left(\frac{1}{s_R} - \frac{1}{\sigma_R}\right) + \frac{1}{\sigma_R}\left(\hat{\sigma}(x) - \sigma(x)\right)$ and $\hat{\sigma}(x) = O_p(1)$ an immediate consequence of Theorem 2 is that $\frac{\hat{\sigma}(x)}{s_R} - \frac{\sigma(x)}{\sigma_R} = o_p(1)$, establishing consistency of the frontier estimator. The asymptotic distribution of $\hat{\rho}(x)$ can be easily obtained from Theorems 1 and 2 by first noting that from Theorem 1 we have that,

$$\sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{\sigma_R} - \frac{\sigma(x)}{\sigma_R} - \frac{1}{2\sigma(x)\sigma_R}B_{0n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{4\sigma_R^2 g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right).$$

Also, since $\sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{\sigma_R} - \frac{\sigma(x)}{\sigma_R} - \frac{1}{2\sigma(x)\sigma_R}B_{0n}\right) \equiv \sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{s_R} - \frac{\sigma(x)}{\sigma_R} - \hat{\sigma}(x)\left(s_R^{-1} - \sigma_R^{-1}\right) - \frac{1}{2\sigma(x)\sigma_R}B_{0n}\right)$ we have by Theorem 2 that $\hat{\sigma}(x)\left(s_R^{-1} - \sigma_R^{-1}\right) = O_p(h_n^2)$ and consequently we can write,

$$\sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{s_R} - \frac{\sigma(x)}{\sigma_R} - \frac{1}{2\sigma(x)\sigma_R}B_{1n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{4\sigma_R^2 g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right) \tag{4}$$

where $B_{1n} = \frac{h_n^2\sigma_K^2\sigma^{(2)2}(x)}{4\sigma_R\sigma(x)} + O_p(h_n^2)$. The asymptotic properties of the frontier estimator can be used directly to obtain the properties of the implied inverse Farrell efficiency. If (y_0, x_0) is a production plan with $x_0 \in G$, then $\hat{R}_0 - R_0 = o_p(1)$ and

$$\sqrt{nh_n}\left(\hat{R}_0 - R_0 + B_{2n}\right) \xrightarrow{d} N\left(0, \frac{R_0^2}{4g_X(x_0)}(\mu_4(x_0) - 1)\int K^2(y)dy\right) \tag{5}$$

where $B_{2n} = \frac{h_n^2\sigma_K^2 R_0\sigma^{2(2)}(x_0)}{4\sigma^2(x_0)} + O_p(h_n^2)$.

The importance of Theorem 2, and in particular its assumption (2), in establishing the asymptotic normality of the frontier and efficiency estimators lies in establishing that the term $\hat{\sigma}(x)\left(s_R^{-1} - \sigma_R^{-1}\right)$ is of the same order as B_{0n}. This allows us to combine the asymptotic biases introduced by the local linear nonparametric estimation and the term introduced by the estimation of σ_R. Assumption (2) places an additional constraint on the DGP that goes beyond those in A1, A2 and A4. Informally, the assumption can be interpreted as a shape restriction on the marginal distribution - $F_R(r)$ of R_t that guarantees that for all $\epsilon > 0$ as $n \to \infty$, $F_R^n(1-\epsilon) \to 0$ sufficiently fast.

Given the results described in theorems 1 and 2, standard bandwidth selection methods (Fan and Gijbels, 1995; Ruppert, Sheather and Wand, 1995) can be used to obtain a data driven h_n. These data driven bandwidth selection methods are asymptotically equivalent to an optimal bandwidth which is $O(n^{-1/5})$.

In addition, as is typical in nonparametric regression, if there is undersmoothing the bias terms vanish asymptotically. In the next section we perform a simulation study that sheds some light on the estimators finite sample performance and compares it to the bias corrected FDH estimator of Park et al.(2000).

4 Monte Carlo Study

In this section we investigate some of the finite sample properties of our estimator, henceforth referred to as NP *via* a Monte Carlo. For comparison purposes, we also include in the study the bias corrected FDH estimator described in Park, Simar and Weiner (2000). Our simulations are based on model (1), i.e.,

$$Y_t = \frac{\sigma(X_t)}{\sigma_R} R_t \text{ with } K = 1.$$

We generate data with the following characteristics. The X_t are pseudo random variables from a uniform distribution with support given by $[10, 100]$. $R_t = exp(-Z_t)$ where Z_t are pseudo random variables from an exponential distribution with parameter $\beta > 0$, therefore R_t has support in $(0, 1]$. We consider two specifications for $\sigma(\cdot)$: $\sigma_1(x) = \sqrt{x}$ and $\sigma_2(x) = 0.0015x^2$, which are associated with production functions that admit decreasing and increasing returns to scale respectively. Three parameters for the exponential distribution were considered: $\beta_1 = 3$, $\beta_2 = 1$, $\beta_3 = 1/3$. These choices of parameters produce, respectively, the following values for the parameters of $g_{R|X}$: $(\mu_R, \sigma_R^2) = (0.25, 0.08), (0.5, 0.08)$, and $(0.75, 0.04)$. Three sample size $n = 100, 300, 600$ are considered and 1000 repetitions are performed for each alternative experimental design. We evaluate the frontiers and construct confidence intervals for efficiency at $(y_0, x_0) = (10, 32.5), (10, 55), (10, 77.5)$ for $\sigma_1(x)$ and at $(y_0, x_0) = (2.5, 32.5), (2.5, 55), (2.5, 77.5)$ for $\sigma_2(x)$. The values of X correspond to the 25^{th},50^{th} and 75^{th} percentile of its support and the values of Y are arbitrarily chosen output levels below the frontier.

Given the convergence in (5) asymptotic confidence intervals for efficiency R_0 can be constructed. To construct a $1-\alpha$ confidence interval for R_0, we obtain a bandwidth h_n for $\hat{\sigma}(x)$ such that $nh_n^5 \to 0$ as $n \to \infty$ (undersmoothing) which eliminates the asymptotic bias. Hence, for quantiles $Z_{\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$ of a standard normal distribution we have

$$\lim_{n\to\infty}\{P(\hat{R}_0-(\sqrt{nh})^{-1}\hat{\sigma}_0(x_0,R_0)Z_{1-\frac{\alpha}{2}}\le R_0\le \hat{R}_0-(\sqrt{nh})^{-1}\hat{\sigma}_0(x_0,R_0)Z_{\frac{\alpha}{2}})\}=1-\alpha$$

where $\hat{\sigma_0}^2(x_0,R_0)=\frac{\hat{R}_0^2}{4\hat{g}_X(x_0)}(\hat{\mu}_4(x_0)-1)\int K^2(y)dy$, $K(\cdot)$ is the Epanechnikov kernel, $\hat{R}_0=\frac{y_0}{\hat{\sigma}(x_0)}s_R$, $\hat{g}_X(x_0)$ is the Rosenblatt kernel density estimator and $\hat{\mu}_4=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t}{\hat{\sigma}(X_t)}-\hat{b}\right)^4$. The estimator $\hat{\mu}_4$ depends on an estimator for b which we define as $\hat{b}=\frac{\sum_{t=1}^{n}\hat{\sigma}(X_t)Y_t}{\sum_{t=1}^{n}\hat{\sigma}^2(X_t)}$. Consistency of this estimator is proved in Lemma 4 in the appendix.[7]

Confidence intervals for R_0 using the bias corrected FDH estimator are given in Park, Simar and Wiener(2000). We follow their suggestion and choose their constant C to be 1 and select their bandwidth (ξ) to be proportional to $n^{-1/3}$.

The evaluation of the overall performance of the efficiency estimator was based on three different measures. First, we consider the correlation between the efficiency rankings produced by the estimator and the true efficiency rankings:

$$R_{rank}=\frac{cov(rank(\hat{R}_t),rank(R_t))}{\sqrt{var(rank(\hat{R}_t))\,var(rank(R_t))}}=\frac{\sum_{t=1}^{n}(rank(\hat{R}_t)-\overline{rank(\hat{R}_t)})(rank(R_t)-\overline{rank(R_t)})}{\sqrt{\sum_{t=1}^{n}(rank(\hat{R}_t)-\overline{rank(\hat{R}_t)})^2\,\sum_{t=1}^{n}(rank(R_t)-\overline{rank(R_t)})^2}}$$

where $rank(R_t)$ gives the ranking index according to the magnitude of R_t and $\overline{rank(R_t)}$ is the mean of $rank(R_t)$. The closer R_{rank} for $\hat{R}_t$ is to 1, the higher the correlation between the true R_t and $\hat{R}_t$, thus the better the estimator $\hat{R}_t$. The second measure we consider is

$$R_{mag}=\frac{1}{n}\sum_{t=1}^{n}(\hat{R}_t-R_t)^2$$

which is simply the squared Euclidean distance between the estimated vector of efficiencies and the true vector of efficiencies. The third measure we use is,

$$R_{rel}=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{R}_t}{\hat{R}_i}-\frac{R_t}{R_i}\right|,$$

where i is the position index for $R_i=\max_{1\le t\le n}R_t$, and $\hat{R}_i$ is the i^{th} corresponding element in $\{\hat{R}_t\}_{t=1}^{n}$, which may or may not be the maximum of $\hat{R}_t$. Hence R_{rank}, R_{mag} summarize the performance of the

[7] Note that together, the consistency of s_R from Theorem 2 and Lemma 4 can be used to define a consistent estimator for μ_R, $\hat{\mu}_R=\hat{b}s_R$.

estimator $\hat{R}_t$ in ranking and calculating the magnitude of efficiency. R_{rel} captures the relative efficiency. In our simulations we consider estimates $\hat{R}_t$ based on both our estimator and the bias corrected FDH estimator.

The results of our simulations are summarized in Tables 1,2,3 and 4. Table 1 provides the bias and mean squared error - MSE of s_R and $\hat{\sigma}(x)$ at three different values of x. Table 2 gives the bias and MSE of our estimator (NP) as well as those of the bias corrected FDH frontier estimator. Table 3 gives the empirical coverage probability (the frequency that the estimated confidence interval contains the true efficiency in 1000 repetitions) for efficiency for both estimators and Table 4 gives the overall performance of the efficiency estimators according to the measures described above. We first identify some general regularities on estimation performance.

General Regularities. As expected from the asymptotic results of section 3, as the sample size n increases, the bias and the MSE for s_R, $\hat{\sigma}(x)$, and the frontier estimator based on NP generally decrease, with some exceptions when it comes to the bias. The frontier estimator based on the bias corrected FDH also exhibits decreasing MSE and bias, with a number of exceptions in the latter case. We observe that the empirical coverage probability for NP is close to the true 95% and generally approaches 95% as n increases with exceptions for small μ_R, while that for FDH is usually below 95% and there is no clear evidence that they get closer to 95% as n increases. The asymptotics of both estimators seem to be confirmed in general terms as their performances improve with large n.

We now turn to the impact of different values of μ_R on the performance of NP and FDH. As μ_R increases, the bias and MSE of s_R increase, with the bias being generally negative except for small μ_R and small sample($n = 100$). The bias of $\hat{\sigma}(x)$, which is negative for $\sigma_1(x)$ and mostly positive for $\sigma_2(x)$, doesn't seem to be impacted by μ_R. Note that the sign of these biases is in accordance to what the asymptotic results predict due to the presence of $\sigma^{2(2)}(x)$ in the bias term. Also, in accordance to the asymptotic results derived in section 3, the MSE for $\hat{\sigma}(x)$ oscillates with μ_R, which reflects the fact that the variance of $\hat{\sigma}(x)$ depends on μ_R in a nonlinear fashion, as indicated by Theorem 1. The bias of the NP frontier estimator is generally positive, except for small μ_R and $n = 100$, and generally increases with μ_R except for the case where $n = 100$, whereas its MSE oscillates with μ_R. In general, the FDH frontier estimator has a positive bias, which together with MSE decreases with μ_R in most experiments, exceptions occuring when $\sigma(x) = 0.0015x^2$. No clear pattern is discerned from the impact of larger μ_R on the empirical coverage probability for NP, but there is weak evidence that FDH is improved. Regarding the measures of overall performance for the efficient estimator described above, the NP estimator seems to perform worse when μ_R is larger for R_{rank}, R_{mag} and R_{rel}. The FDH estimator performs worse when μ_R is larger and the performance measured considered is R_{rank}, while in the case of R_{mag} and R_{rel}, FDH performs better as μ_R increases for the case of $\sigma_1(x)$, but the performance oscillates when we consider $\sigma_2(x)$.

Lastly, as one would expect from the NP estimation procedure, the experimental results indicate that as measured by bias and MSE, the estimation of the NP frontier is less accurate and precise than that of $\sigma(x)$, since the NP frontier estimator involves the estimation of both $\sigma(x)$ and σ_R.

Relative Performance of Estimators. On estimating the production frontier (Table 2) there seems to be evidence that NP dominates FDH in terms of bias and MSE when $\mu_R = 0.25$ and $\mu_R = 0.5$, with exceptions in cases where $\sigma(x) = 0.0015x^2$, while FDH is better with $\mu_R = 0.75$. Regarding the empirical coverage probabilities (Table 3), the NP estimator is superior in most experiments, i.e., NP estimates are much closer to the intended probability $1-\alpha = 95\%$. When the different measures of overall performance we considered are analyzed (Table 4), we observe that the NP estimator outperforms FDH in terms of R_{rank} and R_{rel}, except when $\mu_R = 0.75$ and $\sigma(x) = \sqrt{x}$. In terms of R_{mag}, NP generally outperforms FDH when $\mu_R = 0.25, 0.5$, while FDH is better when $\mu_R = 0.75$. Based on these results, it seems reasonable to conclude that when we are dealing with DGPs that produce inefficient and mediocre firms with large probability, then the fact that the NP estimator is impacted to a lesser degree by extreme values results in better performance *vis a vis* the FDH estimator, whose construction depends heavily on boundary points. This improved performance is easily perceived in Figure 1. The figure shows kernel density estimates for the frontier around the true value evaluated at $x = 55$ for NP $\left(\frac{\hat{\sigma}(x)}{s_R} - \frac{\sigma(x)}{\sigma_R}\right)$ and FDH $\left(\hat{\rho}_{FDH}(x) - \frac{\sigma(x)}{\sigma_R}\right)$ based on 1000 simulations, $\mu_R = 0.25$ and $\sigma(x) = \sqrt{x}$, for $n = 100$ and 600. The kernel density estimates were calculated using an Epanechnikov kernel and bandwidths were selected using the *rule-of-thumb* of Silverman(1986). We observe that the NP estimator is more tightly centered around the true frontier and shows the familiar symmetric bell shape, while that of FDH is generally bimodal with greater variability. Figure 1 also shows that the estimated densities become tighter with more acute spikes as the sample size increases, as expected from the available asymptotic results.[8]

5 Conclusion

In this paper we proposed a new nonparametric frontier model together with estimators for the frontier and associated efficiency levels of production units or plans. Our estimator can be viewed as an alternative to DEA and FDH estimators that are popular and have been widely used in the empirical literature. The estimator is easily implementable, as it is in essence a local linear kernel estimator, and we show that it is consistent and asymptotically normal when suitably normalized. Efficiency rankings and relative efficiency of firms are estimated based only on some rather parsimonious restrictions on conditional moments. The assumptions required to obtain the asymptotic properties of the estimator are standard in nonparametric statistics and are flexible enough to preserve the desirable generality that has characterized nonparametric deterministic frontier estimators. In contrast to DEA and FDH estimators, our estimator is not intrinsically biased but it does envelop the data, in the sense that no observation can lie above the estimated frontier. The small Monte Carlo study we perform seems to confirm the asymptotic results we have obtained and also seems to indicate that for a number of DGPs our proposed estimator can outperform bias corrected FDH according to various performance measures.

Our estimator together with DEA, FDH and the recently proposed estimator of Cazals et al.(2002) forms a set of procedures that can be used for estimating nonparametric deterministic frontiers and for which asymptotic distributional results are available. Future research on the relative performance of all of these alternatives under various DGPs would certainly be desirable from a theoretical and practical viewpoints. Furthermore, extensions of all such models and estimators to accommodate stochastic frontiers with minimal additional assumptions that result in identification is also desirable. Lastly, with regards to our estimator, an extension to the case of multiple outputs should be accomplished.

[8]Similar graphs but with less dramatic differences between the NP and FDH estimators are obtained when $\mu_R = 0.5$.

Appendix

Lemma 3 *Let* $\{Z_i\}_{i=1}^n$ *be a sequence of i.i.d. random variables and* $\psi_n(Z_1, \cdots, Z_k)$ *be a symmetric function with* $k \le n$*. Let* $u_n = \begin{pmatrix} n \\ k \end{pmatrix}^{-1} \sum_{(n,k)} \psi_n(Z_{i_1}, \cdots, Z_{i_k})$ *and* $\hat{u}_n = \frac{k}{n}\sum_{i_1=1}^n (\psi_{1n}(Z_i) - \theta_n) + \theta_n$*, where* $\sum_{(n,k)}$ *denotes a sum over all subsets* $1 \le i_1 < i_2 < \cdots < i_k \le n$ *of* $\{1, 2, \cdots, n\}$*,* $\psi_{1n}(Z_i) = E\left(\psi_n(Z_1, \cdots, Z_k)|Z_i\right)$*,* $\theta_n = E\left(\psi_n(Z_1, \cdots, Z_k)\right)$*. If* $E\left(\psi_n^2(Z_1, \cdots, Z_k)\right) = o(n)$ *then* $\sqrt{n}(u_n - \hat{u}_n) = o_p(1)$*.*

Proof Using Hoeffding's(1961) decomposition for U-statistics we write, $u_n = \theta_n + \sum_{j=1}^k \begin{pmatrix} k \\ j \end{pmatrix} H_n^{(j)}$ where $H_n^{(j)} = \begin{pmatrix} n \\ j \end{pmatrix}^{-1} \sum_{(n,j)} h_n^{(j)}(Z_{v_1}, \cdots, Z_{v_j})$, $h_n^{(1)}(Z_{v_1}) = \psi_{1n}(Z_{v_1}) - \theta_n$, $h_n^{(c)}(Z_{v_1}, \cdots, Z_{v_c}) = \psi_{cn}(Z_{v_1}, \cdots, Z_{v_c}) - \sum_{j=1}^c \sum_{(c,j)} h_n^{(j)}(Z_{i_1}, ..., Z_{i_j}) - \theta_n$ where $\psi_{cn}(Z_{v_1}, \cdots, Z_{v_c}) = E(\psi_n(Z_1, \cdots, Z_k)|Z_1, \cdots, Z_c)$ and $c = 2, ..., k$. Then, $u_n - \hat{u}_n = \sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix} H_n^{(j)}$ and it is straightforward to show that $E(u_n - \hat{u}_n) = 0$. Also,

$$\begin{aligned} V(n^{1/2}(u_n - \hat{u}_n)) &= nE\left(\left(\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix} H_n^{(j)}\right)^2\right) = nE\left(\sum_{j'=2}^k \sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix} \begin{pmatrix} k \\ j' \end{pmatrix} H_n^{(j)} H_n^{(j')}\right) \\ &= n\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix}^2 \begin{pmatrix} n \\ j \end{pmatrix}^{-1} E\left(h_n^{(j)}(Z_1, ..., Z_j)^2\right) \end{aligned}$$

where the last equality follows from theorem 3 in Lee(1990, p.30). By Chebyshev's inequality, for all $\epsilon > 0$, $P\left(|n^{1/2}(u_n - \hat{u}_n)| \ge \epsilon\right) \le nE((u_n - \hat{u}_n)^2)/\epsilon^2$. Therefore, it suffices to show that

$$n\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix}^2 \begin{pmatrix} n \\ j \end{pmatrix}^{-1} E\left(h_n^{(j)}(Z_1, ..., Z_j)^2\right) = o(1).$$

If for all $j = 2, ..., k$

$$E\left((h_n^{(j)}(Z_1, ..., Z_k))^2\right) = O\left(E\left(\psi_n^2(Z_1, ..., Z_k)\right)\right) \tag{6}$$

then for some $\Delta > 0$,

$$nE((u_n - \hat{u}_n)^2) \le n\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix}^2 \begin{pmatrix} n \\ j \end{pmatrix}^{-1} \Delta E(\psi_n^2(Z_1, ..., Z_k)) =$$

$$n^2\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix}^2 \frac{(n-j)!j!}{n!} n^{-1}\Delta E(\psi_n^2(Z_1, ..., Z_k)).$$

Since $E(\psi_n^2(Z_1,...,Z_k))=o(n)$ by assumption, for fixed k, there are a finite number of terms in $\sum_{j=2}^k$, the magnitude determined by $j=2$. For some $\Delta'>0$, $nE((u_n-\hat{u}_n)^2)\leq \Delta' n^2\left(\begin{array}{c}k\\2\end{array}\right)^2\frac{(n-2)!2!}{n!}\frac{E(\psi_n^2(Z_1,...,Z_k))}{n}\leq O(1)o(1)$. We now use induction to prove that $E\left((h_n^{(j)}(Z_1,...,Z_k))^2\right)=O\left(\psi_n^2(Z_1,...,Z_k)\right)$. Note that

$$h_n^{(j)}(Z_1,...,Z_j)=\psi_{jn}(Z_1,...,Z_j)+\sum_{d=1}^{j-1}(-1)^d\sum_{(j,j-d)}\psi_{(j-d)n}(Z_{i_1},...,Z_{i_{j-d}})+(-1)^j\theta_n \text{ for } j=2,...,m.$$

We first establish the result for $j=2$.

$$\begin{aligned}(h_n^{(2)}(Z_1,Z_2))^2 &= \psi_{2n}^2(Z_1,Z_2)-\psi_{1n}^2(Z_1)-\psi_{1n}^2(Z_2)+\theta_n^2-2\psi_{2n}(Z_1,Z_2)\psi_{1n}(Z_1)-2\psi_{2n}(Z_1,Z_2)\psi_{1n}(Z_2)\\ &+ 2\psi_{2n}(Z_1,Z_2)\theta_n+2\psi_{1n}(Z_1)\psi_{1n}(Z_2)-2\psi_{1n}(Z_1)\theta_n-2\psi_{1n}(Z_2)\theta_n\end{aligned}$$

By Cauchy-Schwarz's inequality, the expected value of each term on the righthand side can be shown to be less than $E\left(\psi_n^2(Z_1,Z_2)\right)$. Since there are a finite number of terms $E\left((h_n^{(2)}(Z_1,Z_2))^2\right)=O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right)$. Now suppose that the statement is true for all $2\leq j\leq k-1$. For $j=k$

$$E(h_n^{(k)}(Z_1,...,Z_k)^2)=E\left(\psi_n(Z_1,...,Z_k)^2\right)+E\left(\left(\sum_{j=1}^{k-1}\sum_{(k,j)}h_n^{(j)}(Z_{i_1},...,Z_{i_j})\right)^2\right)+\theta_n^2-$$

$$2\sum_{j=1}^{k-1}\sum_{(k,j)}E\left(h_n^{(j)}(Z_{i_1},...,Z_{i_j})\psi_n(Z_1,...,Z_k)\right)-2E\left(\psi_n(Z_1,...,Z_k)\theta_n\right)+2\theta_n\sum_{j=1}^{k-1}\sum_{(k,j)}E\left(h^{(j)}(Z_{i_1},...,Z_{i_j})\right)$$

and by Theorem 3 in Lee(1990)

$$\begin{aligned}E\left(\left(\sum_{j=1}^{k-1}\sum_{(k,j)}h_n^{(j)}(Z_{i_1},...,Z_{i_j})\right)^2\right) &= \sum_{j=1}^{k-1}\sum_{(k,j)}\sum_{j'=1}^{k-1}\sum_{(k,j')}E\left(h_n^{(j)}(Z_{i_1},...,Z_{i_j})h_n^{(j')}(Z_{i_1},...,Z_{i_{j'}})\right)\\ &= \sum_{j=1}^{k-1}\sum_{(k,j)}E\left((h_n^{(j)}(Z_{i_1},...,Z_{i_j}))^2\right)\end{aligned}$$

Given that this sum has a finite number of terms and the induction hypothesis we have that the left-hand side of the last equality is $O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right)$. Second, again by Theorem 3 in Lee(1990)

$$E\left(h_n^{(j)}(Z_{i_1},...,Z_{i_j})\psi_n(Z_1,...,Z_k)\right)=E\left((h_n^{(j)}(Z_{i_1},...,Z_{i_j}))^2\right),$$

therefore by the induction hypothesis $\sum_{j=1}^{k-1}\sum_{(k,j)} E\left((h_n^{(j)}(Z_{i_1},...,Z_{i_j}))^2\right) = O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right)$. Finally, $E\left(\psi_n(Z_1,...,Z_k)\theta_n\right) = \theta_n^2 \leq E(\psi_n^2(Z_1,...,Z_k))$ and the last term is zero. Hence, $E(h_n^{(k)}(Z_1,...,Z_k)^2) = O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right)$ for all $j = 2,...,k$.

Lemma 4 *Assume that A1, A2, A3 and A4. If $h_n \to 0$, $\frac{nh_n^3}{ln(h_n)} \to \infty$, and $X_t \in G$ a compact subset of $\Re$, then $\hat{b} - b = o_p(1)$*

Proof [Lemma 4] We write $\hat{b} - b = \theta_1 - \theta_2 + \theta_3 + \theta_4 - \theta_5$, where

$$\theta_1 = \frac{b}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)}\left(\frac{1}{n}\sum_{t=1}^n \sigma(X_t)(\hat{\sigma}(X_t) - \sigma(X_t))\right), \theta_2 = \frac{b}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)}\left(\frac{1}{n}\sum_{t=1}^n (\hat{\sigma}^2(X_t) - \sigma^2(X_t))\right),$$

$$\theta_3 = \frac{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)\epsilon_t}{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)}, \theta_4 = \frac{\frac{1}{n}\sum_{t=1}^n \sigma(X_t)(\hat{\sigma}(X_t) - \sigma(X_t))\epsilon_t}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)} \text{ and,}$$

$$\theta_5 = \frac{\frac{1}{n}\sum_{t=1}^n (\hat{\sigma}^2(X_t) - \sigma^2(X_t))}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)} \frac{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)\epsilon_t}{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)}.$$

Under assumptions A1-A4 a routine application of Kolmogorov's law of large numbers gives $\theta_3 = o_p(1)$. Now,

$$\begin{aligned}\theta_1 + \theta_4 &= \frac{1}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)} n^{-1}\sum_{t=1}^n (\hat{\sigma}(X_t) - \sigma(X_t))Y_t = \frac{1}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)} n^{-1}\sum_{t=1}^n \left(\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right) \\ &\times (\hat{\sigma}^2(X_t) - \sigma^2(X_t))Y_t + \frac{1}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)} n^{-1}\sum_{t=1}^n \frac{1}{2\sigma(X_t)}(\hat{\sigma}^2(X_t) - \sigma^2(X_t))Y_t \\ &= \frac{1}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)}(D_{1n} + D_{2n}),\end{aligned}$$

where $\sigma_b^2(X_t) = \theta\sigma^2(X_t) + (1-\theta)\hat{\sigma}^2(X_t)$ for some $0 \leq \theta \leq 1$ and for all $X_t \in G$. Since $\frac{1}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)} = O_p(1)$ from Theorem 1, it suffices to consider D_{1n} and D_{2n}. We first consider D_{1n}. It is easy to see that if $\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}} = o_p(h_n)$ uniformly in G and $n^{-1}\sum_{t=1}^n |\hat{\sigma}^2(X_t) - \sigma^2(X_t)||Y_t| = o_p(h_n)$, then $D_{1n} = o_p(h_n^2)$. Now, $\left|\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right| \leq \frac{1}{2}\underline{\text{B}}_\sigma^{-1}\frac{1}{\sqrt{\sigma_b^2(X_t)}}|\sigma(X_t) - \sigma_b(X_t)|$ and since $\sigma^2(X_t) - \sigma_b^2(X_t) = (1-\theta)(\sigma^2(X_t) - \hat{\sigma}^2(X_t))$ we have by Theorem 1 that $\sigma^2(X_t) - \sigma_b^2(X_t) = o_p(h_n)$ uniformly for $X_t \in G$. From Corollary 1 it follows that $\sigma(X_t) - \sigma_b(X_t) = o_p(h_n)$ and $\frac{1}{\sqrt{\sigma_b^2(X_t)}} = O_p(1)$ uniformly in G. Hence,

$$sup_{X_t \in G}\left|\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right| = o_p(h_n)$$

and

$$|D_{1n}| \leq n^{-1}\sum_{t=1}^{n}|Y_t|sup_{X_t\in G}\left|\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right|sup_{X_t\in G}|\hat{\sigma}^2(X_t)-\sigma^2(X_t)| \leq o_p(h_n^2)n^{-1}\sum_{t=1}^{n}|Y_t| = o_p(h_n^2)$$

where the last equality follows from the fact that $n^{-1}\sum_{t=1}^{n}|Y_t| = O_p(1)$ by Chebyshev's inequality. Now $D_{2n} \leq n^{-1}\sum_{t=1}^{n}\frac{|Y_t|}{2\sigma(X_t)}sup_{X_t\in G}|\hat{\sigma}^2(X_t) - \sigma^2(X_t)| = o_p(h_n)n^{-1}\sum_{t=1}^{n}\frac{|Y_t|}{2\sigma(X_t)} = o_p(h_n)n^{-1}\sum_{t=1}^{n}\frac{1}{2}|b+\epsilon_t| = o_p(h_n)$, where the last equality follows from $n^{-1}\sum_{t=1}^{n}\frac{1}{2}|b+\epsilon_t| = O_p(1)$ by Chebyshev's inequality. Hence, $\theta_1 + \theta_4 = o_p(h_n)$. Now, $|\theta_2| \leq \left|\frac{b}{\sum_{t=1}^{n}\hat{\sigma}^2(X_t)}\right|\left|n^{-1}\sum_{t=1}^{n}(\hat{\sigma}^2(X_t) - \sigma^2(X_t))\right| = O_p(1)o_p(h_n) = o_p(h_n)$ by Theorem 1. Finally, $|\theta_5| = o_p(h_n)$ by the results from the analysis of θ_2 and θ_3. Combining all the convergence results $\hat{b} - b = o_p(1)$.

Acknowledgments. We thank R. Färe and S. Grosskopf for helpful discussions.

References

[1] 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.

[2] Baccouche, R. and M. Kouki, 2003, Stochastic production frontier and technical inefficiency: a sensitivity analysis, Econometric Reviews, 22, 79-91.

[3] Banker, R., 1993, Maximum likelihood, consistency and data development analysis: a statistical foundation, management Science, 39, 1265-1273.

[4] Caudill, S., J. Ford, D. Gropper, 1995, Frontier estimation and firm specific inefficiency measures in the presence of heterocedasticity, Journal of Business and Economic Statistics, 13, 105-111.

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

[6] Charnes, A., W. Cooper and E. Rhodes, 1978, Measuring the efficiency of decision making units, European Journal of Operational Research, 429-444.

[7] Deprins, D., L. Simar and H. Tulkens, 1984, Measuring labor inefficiency in post offices. In M. Marchand, P. Pestiau and H. Tulkens(Eds.), The performance of public enterprises: concepts and measurements. North Holland, Amsterdam.

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

[9] Fan, J. and I. Gijbels, 1995, Data driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation, Journal of the Royal Statistical Society B 57, 371-394.

[10] Fan, J., and Q. Yao, 1998, Efficient estimation of conditional variance functions in stochastic regression, Biometrika, 85, 645-660.

[11] Farrell, M., 1957, The measurement of productive efficiency, Journal of the Royal Statistical Society A, 120, 253-290.

[12] 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.

[13] Hadri, K., 1999, Estimation of a doubly heterocedastic stochastic frontier cost function, Journal of Business and Economic Statistics, 17, 359-363.

[14] Hall, P. and R. Carroll, 1989, Variance function estimation in regression: the effect of estimation of the mean. journal of the Royal Statistical Society B, 51, 3-14.

[15] Hoeffding, W., 1961, The strong law of large numbers for U-statistics, Institute of Statistics, Mimeo Series 302, University of North Carolina, Chapel Hill, NC.

[16] Korostelev, L. Simar and Tsybakov, 1995, Efficient estimation of monotone boundaries, Annals of Statistics, 23, 476-489.

[17] Lee, A.J., 1990, U-Statistics. Marcel Dekker, New York.

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

[19] Pagan, A. and A. Ullah, 1999, Nonparametric econometrics. Cambridge University Press, Cambridge, UK.

[20] Park, B., L. Simar and Ch. Weiner, 2000, The FDH estimator for productivity efficient scores: asymptotic properties, Econometric Theory, 16, 855-877.

[21] Powell, J., J. Stock, T. Stoker, 1989, Semiparametric estimation of index coefficients, Econometrica, 57, 1403-1430.

[22] Prakasa-Rao, B.L.S., 1983, Nonparametric functional estimation. John Wiley and Sons, New York.

[23] Ruppert D., S. Sheather, M. Wand, 1995, An Effective Bandwidth Selector for Least Squares Regression, Journal of the American Statistical Association, 90, 1257-1270.

[24] Seifford, L., 1996, Data envelopment analysis: the evolution of the state of the art(1978-1995), Journal of Productivity Analysis 7, 99-137.

[25] Silverman, B.W,, 1986, Density estimation for statistics and data analysis. Chapman and Hall, London.

Table 1: bias and MSE for S_R and $\hat{\sigma}(x)$

$\sigma_1(x) = \sqrt{x}$	n	S_R bias	S_R MSE($\times 10^{-1}$)	$\hat{\sigma}(x_1): x_1 = 32.5$ bias	$\hat{\sigma}(x_1): x_1 = 32.5$ MSE	$\hat{\sigma}(x_2): x_2 = 55$ bias	$\hat{\sigma}(x_2): x_2 = 55$ MSE	$\hat{\sigma}(x_3): x_3 = 77.5$ bias	$\hat{\sigma}(x_3): x_3 = 77.5$ MSE
$\mu_R = 0.25$	100	0.010	0.005	-0.237	0.963	-0.322	1.608	-0.490	2.440
	300	-0.010	0.004	-0.080	0.281	-0.113	0.469	-0.137	0.698
	600	-0.012	0.003	-0.032	0.151	-0.057	0.240	-0.077	0.350
$\mu_R = 0.5$	100	-0.026	0.012	-0.205	0.422	-0.284	0.761	-0.334	1.075
	300	-0.018	0.005	-0.070	0.139	-0.115	0.211	-0.078	0.295
	600	-0.014	0.003	-0.042	0.065	-0.064	0.110	-0.061	0.154
$\mu_R = 0.75$	100	-0.046	0.026	-0.230	0.974	-0.304	1.601	-0.355	2.420
	300	-0.027	0.009	-0.124	0.334	-0.127	0.531	-0.115	0.758
	600	-0.019	0.005	-0.029	0.171	-0.044	0.247	-0.066	0.380

$\sigma_2(x) = 0.0015x^2$	n	S_R bias	S_R MSE($\times 10^{-1}$)	$\hat{\sigma}(x_1): x_1 = 32.5$ bias	$\hat{\sigma}(x_1): x_1 = 32.5$ MSE	$\hat{\sigma}(x_2): x_2 = 55$ bias	$\hat{\sigma}(x_2): x_2 = 55$ MSE	$\hat{\sigma}(x_3): x_3 = 77.5$ bias	$\hat{\sigma}(x_3): x_3 = 77.5$ MSE
$\mu_R = 0.25$	100	0.016	0.008	0.027	0.104	-0.121	0.722	-0.402	2.643
	300	-0.005	0.003	0.065	0.036	0.006	0.227	-0.070	0.849
	600	-0.010	0.003	0.058	0.020	0.028	0.123	-0.016	0.451
$\mu_R = 0.5$	100	-0.024	0.014	0.036	0.046	-0.057	0.307	-0.272	1.386
	300	-0.015	0.004	0.057	0.017	0.001	0.102	-0.031	0.373
	600	-0.013	0.003	0.045	0.010	0.025	0.052	-0.025	0.213
$\mu_R = 0.75$	100	-0.050	0.031	0.036	0.106	-0.166	0.782	-0.503	2.781
	300	-0.026	0.009	0.082	0.041	0.007	0.226	-0.108	0.899
	600	-0.020	0.005	0.057	0.022	0.028	0.128	-0.029	0.488

Table 2: Bias and MSE of Nonparametric and FDH frontier Estimators

			$x_1 = 32.5$		$x_2 = 55$		$x_3 = 77.5$	
$\sigma_1(x) = \sqrt{x}$	n		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	bias	-1.422	4.446	-1.909	4.673	-2.639	4.246
		MSE	12.895	89.907	20.873	95.434	32.012	90.600
	300	bias	0.486	2.939	0.604	2.998	0.713	3.225
		MSE	3.916	39.069	6.751	39.532	10.751	40.645
	600	bias	0.797	2.174	0.984	2.314	1.127	2.197
		MSE	2.854	21.894	4.650	23.276	6.273	22.135
$\mu_R = 0.5$	100	bias	1.290	3.432	1.611	3.522	1.929	3.046
		MSE	8.920	62.745	15.201	62.717	22.018	56.067
	300	bias	1.047	1.689	1.282	1.624	1.736	1.848
		MSE	3.243	20.337	5.218	19.569	8.239	21.934
	600	bias	0.835	0.999	1.052	1.257	1.303	1.094
		MSE	1.594	8.937	2.720	11.349	3.857	9.727
$\mu_R = 0.75$	100	bias	8.552	3.030	10.899	2.834	12.826	2.932
		MSE	255.763	53.209	325.684	50.862	349.823	50.477
	300	bias	4.126	1.577	5.633	1.397	6.881	1.362
		MSE	30.484	17.467	59.484	15.665	85.023	15.498
	600	bias	3.075	0.884	3.978	0.766	4.639	0.839
		MSE	16.200	7.512	26.989	6.595	38.007	7.520

			$x_1 = 32.5$		$x_2 = 55$		$x_3 = 77.5$	
$\sigma_2(x) = 0.0015x^2$	n		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	bias	-0.184	0.535	-1.214	-1.561	-2.967	-4.712
		MSE	1.361	6.059	9.930	11.365	37.537	41.254
	300	bias	0.353	0.537	0.356	-0.810	0.407	-2.649
		MSE	0.604	2.780	3.327	4.591	11.454	15.303
	600	bias	0.429	0.456	0.729	-0.389	1.170	-1.983
		MSE	0.449	1.615	2.429	2.275	7.705	8.328
$\mu_R = 0.5$	100	bias	0.695	1.093	1.359	0.191	2.078	-1.566
		MSE	1.617	6.090	9.213	6.484	34.783	12.922
	300	bias	0.527	0.649	0.904	0.434	1.668	-0.762
		MSE	0.507	2.496	2.353	2.767	8.032	4.871
	600	bias	0.425	0.594	0.843	0.355	1.393	-0.450
		MSE	0.299	1.592	1.498	1.842	4.792	2.680
$\mu_R = 0.75$	100	bias	3.378	1.148	7.713	0.434	14.211	-1.152
		MSE	20.865	5.903	128.855	6.508	456.643	11.019
	300	bias	1.832	0.768	3.879	0.514	6.943	-0.496
		MSE	5.067	2.563	26.146	2.572	89.554	3.999
	600	bias	1.278	0.539	2.883	0.465	5.230	-0.125
		MSE	2.387	1.410	13.516	1.721	47.090	2.072

Table 3: Empirical coverage probability for $\hat{R}$ by nonparametric and FDH for $1-\alpha = 95\%$

		$x_1 = 32.5, y_1 = 10$		$x_2 = 55, y_2 = 10$		$x_3 = 77.5, y_3 = 10$	
$\sigma_1(x) = \sqrt{x}$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.958	0.748	0.957	0.748	0.965	0.727
	300	0.984	0.776	0.981	0.771	0.976	0.792
	600	0.964	0.787	0.970	0.779	0.972	0.785
$\mu_R = 0.5$	100	0.994	0.810	0.987	0.825	0.996	0.801
	300	0.964	0.830	0.967	0.812	0.955	0.831
	600	0.946	0.827	0.952	0.846	0.951	0.839
$\mu_R = 0.75$	100	0.999	0.836	1.000	0.811	1.000	0.845
	300	0.979	0.855	0.966	0.843	0.968	0.836
	600	0.939	0.837	0.939	0.832	0.932	0.834
		$x_1 = 32.5, y_1 = 2.5$		$x_2 = 55, y_2 = 2.5$		$x_3 = 77.5, y_3 = 2.5$	
$\sigma_2(x) = 0.0015x^2$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.925	0.616	0.932	0.763	0.945	0.490
	300	0.921	0.680	0.970	0.766	0.975	0.557
	600	0.911	0.691	0.947	0.800	0.965	0.561
$\mu_R = 0.5$	100	0.957	0.735	0.985	0.757	0.993	0.773
	300	0.903	0.712	0.961	0.777	0.980	0.773
	600	0.879	0.756	0.944	0.730	0.956	0.782
$\mu_R = 0.75$	100	0.995	0.756	0.999	0.782	0.999	0.785
	300	0.945	0.777	0.984	0.767	0.979	0.774
	600	0.918	0.780	0.953	0.753	0.954	0.785

Table 4: Overall Measures of Efficiency Estimators by nonparametric and FDH

		R_{rank}		R_{mag}		R_{rel}	
$\sigma_1(x) = \sqrt{x}$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.990	0.966	0.014	0.014	0.054	0.200
	300	0.997	0.986	0.002	0.006	0.026	0.120
	600	0.999	0.992	0.001	0.004	0.019	0.078
$\mu_R = 0.5$	100	0.966	0.934	0.034	0.012	0.074	0.200
	300	0.990	0.973	0.003	0.005	0.036	0.101
	600	0.996	0.986	0.001	0.003	0.024	0.067
$\mu_R = 0.75$	100	0.785	0.893	0.148	0.008	0.161	0.133
	300	0.893	0.962	0.017	0.002	0.086	0.059
	600	0.938	0.981	0.009	0.001	0.059	0.039
		R_{rank}		R_{mag}		R_{rel}	
$\sigma_2(x) = 0.0015x^2$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.987	0.944	0.014	0.026	0.059	0.314
	300	0.996	0.976	0.008	0.013	0.033	0.202
	600	0.998	0.987	0.006	0.008	0.026	0.148
$\mu_R = 0.5$	100	0.956	0.830	0.072	0.033	0.091	0.427
	300	0.983	0.919	0.011	0.016	0.052	0.262
	600	0.990	0.951	0.009	0.009	0.039	0.189
$\mu_R = 0.75$	100	0.747	0.641	0.097	0.029	0.185	0.332
	300	0.863	0.841	0.056	0.011	0.111	0.176
	600	0.906	0.912	0.042	0.005	0.080	0.126

Figure 1: Density Estimates for NP and FDH estimators