Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Econometrics 141 (2007) 283–319
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Nonparametric frontier estimation via local linear regression

Carlos Martins-Filho[a,*], Feng Yao[b]

[a]Department of Economics, Oregon State University, Ballard Hall 303, Corvallis, OR 97331-3612, USA
[b]Department of Economics, University of North Dakota, P.O. Box 8369, Grand Forks, ND 58202-8369, USA

Available online 20 February 2007

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.

JEL classification: C14

Keywords: Nonparametric regression frontier; Local linear estimation; U statistics

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 $(y, x) \in \Re_+ \times \Re_+^D$ where y describes the output of a production

*Corresponding author. Tel.: +1 541 737 1476; fax: +1 541 737 5917.
E-mail addresses: carlos.martins@orst.edu (C. Martins-Filho), feng.yao@mail.business.und.edu (F. Yao).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.01.005

unit and x describes the D inputs used in production. The production technology is given by the set $T = \{(y, x) \in \Re_+ \times \Re_+^D : x \text{ can produce } y\}$ and the production function or frontier associated with T is $\rho(x) = \sup\{y \in \Re_+ : (y, x) \in T\}$ for all $x \in \Re_+^D$. Let $(y_0, x_0) \in T$ characterize the performance of a production unit and define $0 \leqslant R_0 \equiv \frac{y_0}{\rho(x_0)} \leqslant 1$ 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 $N_n \equiv \{(Y_t, X_t)\}_{t=1}^n$ that share a technology T, to obtain estimates of $\rho(\cdot)$ and by extension $R_t = \frac{Y_t}{\rho(X_t)}$ for $t = 1, \ldots, n$. 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 largely in econometrics by Charnes et al. (1978) data envelopment analysis (DEA) and Deprins et al. (1984) free disposal hull (FDH), is based on the assumption that all observed data lie in the technology set T, i.e., $\mathrm{P}((Y_t, X_t) \in T) = 1$ for all t. The stochastic approach, pioneered by Aigner et al. (1977) and Meeusen and van den Broeck (1977), allows for random shocks in the production process and consequently $\mathrm{P}((Y_t, X_t) \notin T) > 0$. 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 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 a flexible nonparametric structure, the appeal of DEA and FDH estimators has increased since Gijbels et al. (1999) and Park et al. (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, Knight (2001) has proposed a local polynomial frontier estimator that envelops the data as a smooth function of input usage, not a discontinuous or piecewise linear function as in the case of FDH and DEA estimators, respectively. Unfortunately, Knight's estimator has no explicit asymptotic distribution making it difficult to conduct inference and to correct its inherent bias. Similarly, the piecewise polynomial estimator of Hall et al. (1998) does not have an explicit asymptotic distribution rendering difficult its practical use. Cazals et al. (2002) have proposed an estimator based on the joint survivor function that is more robust to extreme values and outliers than

[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 et al. (1995) for some preliminary asymptotic results.

DEA, FDH and Knight's estimators, and does not suffer from their inherent biasedness.[3] Girard and Jacob (2004) proposed a smooth kernel estimator of the frontier based on a convolution representation that extends the estimator of Geffroy (1964). The asymptotic properties of their estimator, however, are obtained for the case where the data N_n is a Poisson process, rather than under the more conventional setting where N_n is a random sample.

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. 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 et al. (2000), but in addition our estimation procedure has some general properties that can prove desirable vis-a-vis the estimators currently available in the literature. First, as in Cazals et al. (2002) and Girard and Jacob (2004), the estimator we propose is more robust to extreme values and outliers than those proposed by Hall et al. (1998) and Knight (2001) as well as DEA and FDH. Second, our frontier estimator is a smooth function of input usage. Third, the construction of our estimator is fairly simple as it is in essence a local linear kernel estimator, avoiding the constrained optimization problems in Knight (2001) and Hall et al. (1998). Fourth, although our estimator envelops the data, it is not intrinsically biased as DEA/FDH, Knight's estimator or the piecewise polynomial estimator of Hall et al. (1998), therefore no bias correction is necessary. In addition to these general properties, we are able to establish the asymptotic normality and consistency of our production frontier and efficiency estimators under assumptions that are fairly standard in the nonparametric statistics literature. Contrary to Girard and Jacob (2004) our asymptotic normality result is obtained for the case where N_n is a random sample rather than a Poisson process. We view our proposed estimator not necessarily as a substitute to 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 popular bias corrected FDH estimator of Park et al. (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 DGP 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 $D+1$-dimensional random vector with common density g for all $t \in \{1, 2, \ldots\}$ and that $\{Z_t\}$ forms an independently distributed sequence. We assume there are

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

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 $\Re^D_+$, $\sigma(\cdot):\Re^D_+ \to (0,\infty)$ 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: $\mathrm{E}(R_t|X_t = x) \equiv \mu_R$ 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 Eq. (1) imply that $\mathrm{E}(Y_t|X_t = x) = \frac{\mu_R}{\sigma_R}\sigma(x)$ and $V(Y_t|X_t = x) = \sigma^2(x)$. The model can therefore be rewritten as

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

where $b = \frac{\mu_R}{\sigma_R}$, $\varepsilon_t = \frac{R_t - \mu_R}{\sigma_R}$, $m(X_t) = b\sigma(X_t)$, $\mathrm{E}(\varepsilon_t|X_t = x) = 0$ and $V(\varepsilon_t|X_t = x) = 1$.[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; Hadri, 1999) on production frontiers. Fourth, the structure of the proposed frontier allows for the convenient separation of its shape, represented by $\sigma(\cdot)$, and location represented by σ_R in the estimation process. 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 $x \in \Re^D_+$ we first obtain $\hat{m}(x; h_n) \equiv \hat{\alpha}$ where

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

$K(\cdot):\Re^D \to \Re$ 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 Stone (1977) and Fan (1992) with regressand Y_t and regressors X_t. In the second stage, we follow Hall and Carroll (1989) and

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

Fan and Yao (1998) by defining $e_t \equiv (Y_t - \hat{m}(X_t; h_n))^2$ to obtain $\hat{\sigma}^2(x; h_n) \equiv \hat{\alpha}_1$, where

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

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

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

is obtained. We emphasize that the estimator s_R depends on the bandwidth h_n through $\hat{\sigma}(X_t; h_n)$. Furthermore, in what follows it is desirable to distinguish the bandwidth used in the first two stages of estimation, which we will denote by h_n, from that used in defining s_R, which we will denote by g_n, where $0 < g_n \to 0$ as $n \to \infty$. Therefore, we represent the production frontier estimator at $x \in \Re^D$ by $\hat{\rho}(x; h_n, g_n) = \frac{\hat{\sigma}(x; h_n)}{s_R(g_n)}$. Note that by construction, provided that the chosen kernel K is smooth, $\hat{\rho}(x; h_n, g_n)$ is a smooth estimator that envelops the data (no observed pair (Y_t, X_t) lies above $(\hat{\rho}(X_t; h_n, g_n), X_t))$ but may lie above or below the true frontier $\rho(X_t)$. As such, our estimator does not suffer from the inherent downward bias of DEA/FDH, the estimator proposed by Knight (2001) or Hall et al. (1998).

As pointed out above, in model (2) the parameter σ_R provides the location of the production frontier, whereas its shape is given by $\sigma(\cdot)$. Since besides the conditional moment restrictions on R_t there are no 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, creating a difficulty in locating the frontier, i.e., estimating σ_R. The estimation of σ_R by s_R implies that there exists *one* observed production unit whose production plan lies on the estimated frontier, and by consequence the forecasted value for R_t associated with this unit is identically one.

The problem of locating the production frontier is also inherent in obtaining the estimator proposed by Knight (2001), as well as DEA and FDH estimators. By construction, these estimators require that *multiple* production units 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 in the case of DEA and FDH) that envelop 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, these estimators will provide a downwardly biased location for the frontier. It is this dependency on boundary data points that makes these estimators highly susceptible to extreme values, not only as determinants of the location but also the shape of the frontier. 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, if estimated efficiency levels are defined as $\hat{R}_t = \frac{s_R(g_n) Y_t}{\hat{\sigma}(X_t; h_n)}$, the efficiency ranking of firms, as well as their estimated relative efficiency $\frac{\hat{R}_t}{\hat{R}_\tau}$ for $t, \tau = 1, 2, \ldots, n$ 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

3.1. Asymptotic normality

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 et al. (2000) to obtain the asymptotic distribution of DEA and FDH estimators. We focus on DEA and FDH because these are by far the most used deterministic frontier estimators in applied econometrics and productivity analysis.

Assumption A1. (1) $Z_t = (X_t, R_t)'$ for $t = 1, 2, \ldots, n$ 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) $0 < \underline{B}_{g_X} \leqslant g_X(x) \leqslant \bar{B}_{g_X} < \infty$ for all $x \in G$, G a compact subset of $\Theta = \times_{t=1}^{D}(0, \infty)$, 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) $0 < \underline{B}_\sigma \leqslant \sigma(x) \leqslant \bar{B}_\sigma < \infty$ for all $x \in \Theta$. (5) $\sigma^2(\cdot) : \Theta \to \Re$ is a measurable twice continuously differentiable function in Θ with second derivative denoted by $\sigma^{2(2)}(x)$. (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 $\phi(y, x)$. That $\{(Y_t, X_t)\}_{t=1}^n$ forms an iid sequence corresponds to assumption AI in Park et al. (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 et al. (2000). A2(6) is implied by A1 in Gijbels et al. (1999) and AIII in Park et al. (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. $K(x) : S_D \to \Re$ is a symmetric density function with bounded support $S_D \subset \Re^D$ satisfying: (1) $\int xK(x)\,dx = 0$. (2) $\int x^2 K(x)\,dx = \sigma_K^2$. (3) for all $x \in \Re^D$, $|K(x)| < B_K < \infty$. (4) for all $x, x' \in \Re^D$, $|K(x) - K(x')| < m\|x - x'\|$ for some $0 < m < \infty$, where $\|\cdot\|$ is the Euclidean norm.

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 et al. (2000). We note that obtaining consistency as well as the asymptotic distribution of DEA and FDH estimators for the production frontier and associated firm efficiency depends crucially on the assumption (AII in Park et al., 2000) that the joint density $\phi(y, x)$ of (Y_t, X_t) is positive at the frontier.[5] In reality, there might be situations in which this assumption is too strong. In contrast, we assume that R_t takes values in the entire interval [0, 1], but there is no need for the joint 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

[5] By consequence this assumption is also crucial in obtaining the asymptotic distribution of the estimator proposed by Cazals et al. (2002), as verified in their Theorem 3.2.

convergence of $\max_{1\leqslant t\leqslant n} R_t$ 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 of Y_t. 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 $\hat{\sigma}(x;h_n)$ and by consequence those of $\frac{\hat{\sigma}(x;h_n)}{s_R(g_n)}$ derive from the fact that $\hat{\sigma}(x;h_n)$ is based on regressands that are themselves residuals from a first stage nonparametric regression. This problem is in great part handled by the use of our Lemma 3 on U statistics, which generalizes Lemma 3.1 in Powell et al. (1989) for the case where the U-statistic's kernel is of dimension greater than two. 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 $D=1$. For $D>1$ all of the results hold with appropriate adjustments on the relative speed of n, h_n^D and g_n^D.[6]

Lemma 1 establishes the order in probability of certain linear combinations of kernel functions that appear repeatedly in component expressions of our estimators. The proofs of the lemmas and theorems that follow rely on repeated use of Lebesgue's dominated convergence theorem, which we will refer to often. All proofs are collected in Appendix.

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

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

(a) *If* $nh_n^3\to\infty$ *then* $\sup_{x\in G}|s_j(x)-\mathrm{E}(s_j(x))|=\mathrm{O_p}((\frac{\ln(n)}{nh_n})^{1/2})$.
(b) *If* $\frac{nh_n^3}{\ln(n)}\to\infty$ *then* $\sup_{x\in G}\frac{1}{h_n}|s_j(x)-\mathrm{E}(s_j(x))|=\mathrm{o_p}(1)$.

Part (b) of Lemma 1 is a direct consequence of part (a), and in combination with Assumption A4 can be used to easily show that $s_0(x)-g_X(x)=\mathrm{O_p}(h_n)$, $s_1(x)=\mathrm{O_p}(h_n)$ and $s_2(x)-g_X(x)\sigma_K^2=\mathrm{O_p}(h_n)$ uniformly in G by taking $f(x,r)=1$. These uniform boundedness results are used to prove the following lemma.

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

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

[6] If different bandwidths $h_1,\dots,h_D,g_1,\dots,g_D$ are used, a more extensive adjustment of the relative speed assumptions of n, h_i, g_i are necessary, but with no qualitative consequence to the results obtained.

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

Arguments similar to those used in the proof of Lemma 2 can be used to establish that

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

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

Lemmas 2 and 3 (which appears in Appendix) 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, part (a), establishes the order in probability of the difference between $\hat{\sigma}^2(x; h_n) - \sigma^2(x)$ and $\frac{1}{nh_n g_X(x)}\sum_{t=1}^n K(\frac{X_t-x}{h_n})r_t^*$ uniformly in G. This result permits, under suitable normalization, the investigation of the asymptotic properties of $\hat{\sigma}^2(x; h_n) - \sigma^2(x)$ by restricting attention to $\frac{1}{nh_n g_X(x)}\sum_{t=1}^n K(\frac{X_t-x}{h_n})r_t^*$. The second, part (b), establishes the $\sqrt{nh_n}$-asymptotic normality of $\frac{1}{nh_n g_X(x)}\sum_{t=1}^n K(\frac{X_t-x}{h_n})r_t^*$, and uses this result to obtain the asymptotic normality of $\sqrt{nh_n}(\hat{\sigma}(x; h_n) - \sigma(x))$. The proof of Theorem 1 is similar to that provided by Fan and Yao (1998), but there are two main differences. First, our results are for iid variables; second, and most importantly, their proof as stated is incorrect. Specifically, the inequality involving the term they label I_{33} following their Eq. (A2.6) is incorrect.[7]

Some of the assumptions in the following theorems are made for convenience on ε_t, rather than R_t. Since $\varepsilon_t = \frac{R_t - \mu_R}{\sigma_R}$ these assumptions have a direct counterpart for R_t. Specifically we have $\mathrm{E}(\varepsilon_t^4|X_t) = \mu_4(X_t) \Rightarrow \mathrm{E}(R_t^4|X_t)$ exists as a function of X_t and $\mathrm{E}(|\varepsilon_t||X_t) = \mu_1(X_t)$ is uniformly bounded in G, which implies that $\mathrm{E}(|R_t - \mu_R||X_t)$ exists as a uniformly bounded function of X_t. We note that although Assumption A2(3) implies that $\mathrm{E}(\varepsilon_t|X_t) = 0$ is not dependent on X_t, it does not, in general, imply that $\mathrm{E}(|\varepsilon_t||X_t)$ is independent of X_t.

Theorem 1. *Suppose that Assumptions* A1–A4 *are holding. In addition assume that* $\mathrm{E}(|\varepsilon_t||X_t) = \mu_1(X_t)$ *is a uniformly bounded function of* $X_t \in G$ *a compact subset of* $(0, \infty)$. *If* $h_n \to 0$, $\frac{nh_n^3}{\ln(n)} \to \infty$, *then for every* $x \in G$

(a) $\sup_{x\in G}|\hat{\sigma}^2(x; h_n) - \sigma^2(x) - \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K(\frac{X_t-x}{h_n})r_t^*| = \mathrm{O_p}(h_n^3) + \mathrm{O_p}((\frac{h_n \ln(n)}{n})^{1/2})$.

(b) *If, in addition, we assume that* $\mathrm{E}(\varepsilon_t^4|X_t = x) = \mu_4(x)$ *is continuous in* $(0, \infty)$, $h_n^2 \ln(n) \to 0$ *and* $nh_n^5 = \mathrm{O}(1)$ *then*

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

for all $x \in G$ *where* $B_{1n} = \frac{h_n^2\sigma_K^2}{4\sigma(x)}\sigma^{2(2)}(x) + \mathrm{o_p}(h_n^2)$.

[7]See Fan and Yao (1998, p. 658).

The results in Theorem 1 refer to the estimator $\hat{\sigma}(x;h_n)$, but since our main interest lies on $\hat{\rho}(x;h_n,g_n) \equiv \frac{\hat{\sigma}(x;h_n)}{s_R(g_n)}$, a complete characterization of the asymptotic behavior of the frontier estimator requires a characterization of the asymptotic behavior of $s_R(g_n)$, and how it combines with the results obtained in Theorem 1 for $\hat{\sigma}(x;h_n)$. We first establish in part (a) of Theorem 2 a general result regarding the order in probability of $s_R(g_n) - \sigma_R$. It states that if the estimator $\hat{\sigma}(x;g_n)$ used to obtain s_R is $O_p(L_n)$, where L_n is an arbitrary nonstochastic sequence such that $0 < L_n \to 0$ as $n \to \infty$, and if $1 - \max_{1 \leqslant t \leqslant n} R_t = O_p(L_n)$, then $s_R(g_n) - \sigma_R = O_p(L_n)$. The result is useful in that from part (a) of Theorem 1, if $\frac{ng_n^5}{\ln(n)} \to \infty$, then $\hat{\sigma}(x;g_n) - \sigma(x) = O_p(g_n^2)$. Hence, together with the assumption that $1 - \max_{1 \leqslant t \leqslant n} R_t = O_p(g_n^2)$ we obtain $s_R(g_n) - \sigma_R = O_p(g_n^2)$. It should be noted that the required boundedness in probability of $1 - \max_{1 \leqslant t \leqslant n} R_t$ is not necessary to establish the consistency of $s_R(g_n)$, which results directly from part (a) of Theorem 1. Its use is confined to part (b) of Theorem 2, where we use the result on the order of $s_R(g_n)$ to obtain the asymptotic normality of $\hat{\rho}(x;h_n,g_n)$ under a suitable normalization.

Theorem 2. *Let L_n be a nonstochastic sequence such that $0 < L_n \to 0$ as $n \to \infty$ and suppose that* (1) $\hat{\sigma}(x;g_n) - \sigma(x) = O_p(L_n)$ *uniformly in G, and* (2) $1 - \max_{1 \leqslant t \leqslant n} R_t = O_p(L_n)$. *Then,*

(a) $s_R(g_n) - \sigma_R = O_p(L_n)$,
(b) *Under the assumptions in Theorem* 1 *part* (b), *if* $\frac{ng_n^5}{\ln(n)} \to \infty$, $nh_n^5 = o(1)$, *and* $nh_n g_n^4 = O(1)$ *then*

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

where $B_{2n} = O_p(g_n^2)$.

Assumption (2) in Theorem 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 $\varepsilon > 0$ as $n \to \infty$, $F_R^n(1-\varepsilon) \to 0$ sufficiently fast. Mathematically, the importance of assumption (2) lies in controlling the order in probability of the term $\hat{\sigma}(x;h_n)(s_R^{-1}(g_n) - \sigma_R^{-1})$, and by consequence controlling the bias introduced by the estimation of σ_R in the estimated frontier.

The conditions on the order of the bandwidths h_n and g_n are also crucial for asymptotic normality of the estimated frontier. In particular, they imply that the bandwidth h_n, used in the first and second stages of the estimation, must satisfy $nh_n^5 = o(1)$, which represents an undersmoothing in the estimation $\hat{\sigma}(x;h_n)$. In addition, the bandwidth g_n used to obtain s_R in the third stage must converge to zero slower than h_n, i.e., $\frac{h_n}{g_n} \to 0$ as $n \to \infty$ at suitable speed. The requirement $ng_n^5 \to \infty$ in the estimation of s_R is necessary only in that it provides a convenient order for B_{2n}.

A sharper result on the bias term B_{2n} can be obtained by assuming that $1 - \max_{1 \leqslant t \leqslant n} R_t = o_p(g_n^2)$. In this case part (b) of Theorem 2 can be extended to give

$$\sqrt{nh_n}\left(\frac{\hat{\sigma}(x;h_n)}{s_R(g_n)} - \frac{\sigma(x)}{\sigma_R} - B_{3n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{4\sigma_R^2 g_X(x)}(\mu_4(x)-1)\int K^2(y)\,dy\right), \quad (3)$$

where $B_{3n} = \frac{g_n^2 \sigma(x)\sigma_K^2}{4\sigma_R} \sup_{x \in G, R \in [0,1]}\left(-\frac{\sigma^{2(2)}(x)R}{\sigma^2(x)}\right) + o_p(g_n^2)$. We note that this increased precision in the expression of the bias is unnecessary for inference purposes, since it is normally conducted under the assumption that $nh_ng_n^4 \to 0$, in which case $\sqrt{nh_n}B_{3n} \to 0$ as $n \to \infty$. It is also clear that since $\frac{\hat{\sigma}(x;h_n)}{s_R(g_n)} - \frac{\sigma(x)}{\sigma_R} = \hat{\sigma}(x;h_n)(\frac{1}{s_R(g_n)} - \frac{1}{\sigma_R}) + \frac{1}{\sigma_R}(\hat{\sigma}(x;h_n) - \sigma(x))$ and $\hat{\sigma}(x;h_n) = \mathrm{O_p}(1)$ an immediate consequence of Theorem 2 is that $\frac{\hat{\sigma}(x;h_n)}{s_R(g_n)} - \frac{\sigma(x)}{\sigma_R} = \mathrm{o_p}(1)$, establishing consistency of the frontier estimator.

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 $\hat{R}_0 - R_0 = \mathrm{o_p}(1)$ and

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

where $B_{4n} = \frac{g_n^2\sigma_K^2R_0}{4}\sup_{x\in G, R\in[0,1]}(-\frac{\sigma^{2(2)}(x)R}{\sigma^2(x)}) + \mathrm{o_p}(g_n^2)$.

3.2. Bandwidth selection

The asymptotic normality result in Eq. (3) can be used to provide guidance on bandwidth selection. We follow standard bandwidth selection methods (see Fan and Gijbels, 1995; Ruppert et al., 1995) by considering the minimization of an asymptotic approximation of the estimator's weighted mean integrated squared error (AMISE). The minimization of the AMISE in our case is not standard, since the bias and variance of the frontier estimator depend on different bandwidths. To make the minimization of AMISE amenable to standard solutions, we consider $g_n = n^{\gamma}h_n$, where $0 < \gamma < \frac{1}{6}$ which guarantees that all conditions required in Theorem 2 on the relative speed of h_n and g_n are met. Hence, we write

$$\begin{aligned} AMISE(\hat{\rho}(h_n)) = {} & \frac{1}{nh_n}\int \frac{\sigma^2(x)}{4\sigma_R^2}(\mu_4(x) - 1)\,\mathrm{d}x \int K^2(\phi)\,\mathrm{d}\phi \\ & + \frac{h_n^4 n^{4\gamma}(\sigma_K^2)^2}{16\sigma_R^2}\left(\sup_{x\in G, R\in[0,1]}\left(-\frac{\sigma^{2(2)}(x)R}{\sigma^2(x)}\right)\right)^2 \int \sigma^2(\phi)g_X(\phi)\,\mathrm{d}\phi, \end{aligned}$$

which is a function only of h_n. Using standard calculus we find that bandwidth h_n that minimizes AMISE is given by

$$\begin{aligned} h_n^* &= \left(\frac{\int K^2(\phi)\,\mathrm{d}\phi \int \sigma^2(x)(\mu_4(x) - 1)\,\mathrm{d}x}{(\sigma_K^2)^2\left(\sup_{x\in G, R\in[0,1]}\left(-\frac{\sigma^{2(2)}(x)R}{\sigma^2(x)}\right)\right)^2 \int \sigma^2(x)g_X(x)\,\mathrm{d}x}\right)^{1/5} n^{-(1+4\gamma)/5} \\ &= Cn^{-(1+4\gamma)/5}. \end{aligned} \tag{5}$$

The practical use of h_n^* requires the estimation of the unknowns appearing in C, as in traditional plug-in bandwidth selection methods. In the next section, we provide an easily implementable estimation procedure for these unknowns. We perform a simulation study that sheds some light on our estimator's 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 study. For comparison purposes, we also include in the study the bias corrected FDH estimator described in Park et al. (2000). Our simulations are based on model (1), i.e.,

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

We generate data with the following characteristics. The X_t are pseudorandom variables from a uniform distribution with support given by $[a_l, b_u]$. $R_t = \exp(-Z_t)$ where Z_t are pseudorandom variables from an exponential distribution with parameter $\beta > 0$, therefore R_t has support in $(0, 1]$. We consider two specifications for $\sigma(\cdot)$:

$$\sigma_1(x) = \sqrt{x} \quad \text{with } x \in [a_l, b_u] = [10, 100] \quad \text{and} \quad \sigma_2(x) = 3(x - 1.5)^3 + 0.25x + 1.125 \quad \text{with } x \in [a_l, b_u] = [1, 2]$$

which are associated with convex and nonconvex production technologies, respectively. $\sigma_2(x)$ is also considered in the simulations conducted by Park et al. (2000). Five parameters for the exponential distribution were considered: $\beta_1 = 3$, $\beta_2 = 1.5$, $\beta_3 = 1$, $\beta_4 = \frac{2}{3}$, $\beta_5 = \frac{1}{3}$. These choices of parameters produce, respectively, the following values for the parameters of $g_{R|X}$: $(\mu_R, \sigma_R^2) = (0.25, 0.08)$, $(0.4, 0.09)$, $(0.5, 0.08)$, $(0.6, 0.07)$ and $(0.75, 0.04)$. Three sample sizes $n = 200, 300, 400$ 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 $(y_0, x_0) = (3, 1.25), (3, 1.5), (3, 1.75)$ for $\sigma_2(x)$. The values of X correspond to the 25th, 50th and 75th percentile of its support and the values of Y are arbitrarily chosen output levels below the frontier. An important aspect in the implementation of our frontier estimator is bandwidth selection. We consider the following *rule-of-thumb* bandwidth:

$$\hat{h}_{\text{ROT}} = \left(\frac{\int K^2(\psi)\,\mathrm{d}\psi(\hat{\mu}_4(\lambda_n) - 1)\int \dot{\sigma}^2(x)\,\mathrm{d}x}{(\sigma_K^2)^2 \left(\max_{1 \leqslant t \leqslant n} \left(\frac{\dot{\sigma}^{2(2)}(x_t)\dot{R}_t}{\dot{\sigma}^2(x)} \right) \right)^2 \frac{1}{n}\sum_{t=1}^n \dot{\sigma}^2(x_t)} \right)^{1/5} n^{-(1+4\gamma)/5},$$

where γ is set to be 0.11 in all experiments, which satisfies the requirements in Theorem 2, $K(\cdot)$ is an Epanechnikov kernel and $\hat{g}_{\text{ROT}} = n^{\gamma}\hat{h}_{\text{ROT}}$. The sequence $\{\dot{\sigma}^2(X_t)\}_{t=1}^n$ is estimated with an ordinary least square quartic regression of $\{\hat{\varepsilon}_t^2\}_{t=1}^n$ on $\{X_t\}_{t=1}^n$, with $\hat{\varepsilon}_t = Y_t - \hat{m}(X_t)$, where $\hat{m}(X_t)$ is estimated via local linear regression with a rule-of-thumb bandwidth as in Ruppert et al. (1995). $\{\dot{\sigma}^2(X_t)\}_{t=1}^n$ is then used to construct $\int \dot{\sigma}^2(x)\,\mathrm{d}x$, $\max_{1 \leqslant t \leqslant n}(\frac{\dot{\sigma}^{2(2)}(x_t)\dot{R}_t}{\dot{\sigma}^2(x)})$ and $\frac{1}{n}\sum_{t=1}^n \dot{\sigma}^2(x_t)$. $\hat{\mu}_4(\lambda_n) = \frac{1}{n}\sum_{t=1}^n (\frac{Y_t}{\hat{\sigma}(X_t;\lambda_n)} - \hat{b})^4$, where $\hat{b} = \frac{\sum_{t=1}^n \hat{\sigma}(X_t;\lambda_n) Y_t}{\sum_{t=1}^n \hat{\sigma}^2(X_t;\lambda_n)}$ is an estimator for $b = \mu_R/\sigma_R$. Consistency of $\hat{b}$ is established in Lemma 4 that appears in Appendix.[8] $\{\hat{\sigma}^2(X_t;\lambda_n)\}_{t=1}^n$ in $\hat{\mu}_4$ is estimated via local linear regression of

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

$\{\hat{\varepsilon}_t^2\}_{t=1}^n$ on $\{X_t\}_{t=1}^n$, with a rule-of-thumb bandwidth λ_n as in Ruppert et al. (1995) and Fan and Yao (1998).

Given the convergence in (4) asymptotic confidence intervals for the efficiency R_0 can be constructed. To construct a $1-\alpha$ confidence interval for R_0, we obtain a bandwidth $\hat{h}_u = \hat{h}_{\text{ROT}} n^{-\delta}$ for $\hat{\sigma}(x;\hat{h}_u)$ for some positive δ such that the term $\sqrt{n\hat{h}_u}B_{4n} \to 0$ as $n \to \infty$, and here δ is picked to be 0.02. Similarly $\hat{g}_u = \hat{h}_u n^{\gamma}$. Hence, for quantiles $Z_{\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$ of a standard normal distribution we have

$$\lim_{n\to\infty} \mathrm{P}\left(\hat{R}_0 - \left(\sqrt{n\hat{h}_u}\right)^{-1}\hat{\sigma}_0(x_0,R_0)Z_{1-\frac{\alpha}{2}} \leqslant R_0 \leqslant \hat{R}_0 - \left(\sqrt{n\hat{h}_u}\right)^{-1}\hat{\sigma}_0(x_0,R_0)Z_{\frac{\alpha}{2}}\right) = 1-\alpha,$$

where $\hat{\sigma}_0^2(x_0,R_0) = \frac{\hat{R}_0^2}{4\hat{g}_X(x_0)}(\hat{\mu}_4(\hat{h}_u)-1)\int K^2(y)\,\mathrm{d}y$, $\hat{R}_0 = \frac{y_0}{\hat{\sigma}(x_0;\hat{h}_u)}s_R(\hat{g}_u)$, $\hat{g}_X(x_0)$ is the Rosenblatt kernel density estimator. Confidence intervals for R_0 using the bias corrected FDH estimator are given in Park et al. (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:

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

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 $R_{\text{mag}} = \frac{1}{n}\sum_{t=1}^n (\hat{R}_t - R_t)^2$ 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 $R_{\text{rel}} = \frac{1}{n}\sum_{t=1}^n \left|\frac{\hat{R}_t}{\hat{R}_i} - \frac{R_t}{R_i}\right|$, where i is the position index for $R_i = \max_{1\leqslant t\leqslant n} R_t$, and $\hat{R}_i$ is the ith 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 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 1–4 and illustrated with Figs. 1–5. 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. To help interpret the results, we illustrate the relative performance of the two frontier estimators in terms of MSE in Fig. 1. Fig. 2 shows kernel density estimates for the two frontier estimators around the true value evaluated at $x = 55$ based on 1000 simulations, $\mu_R = 0.25$ and $\sigma_1(x)$, for $n = 200$ and 400. 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

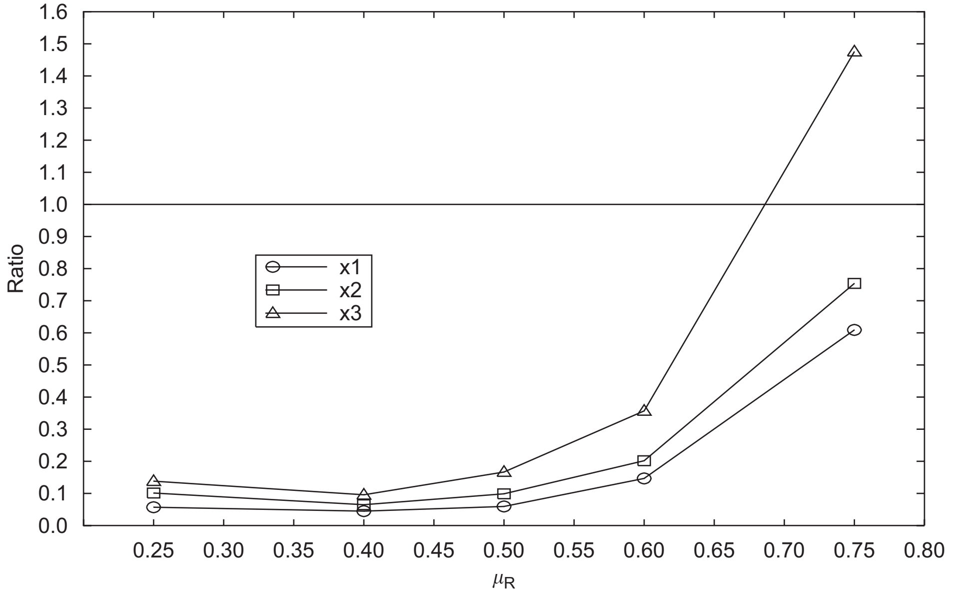

Fig. 1. Relative performance of NP and FDH estimators: $n = 400$, $\sigma_1(x)$, $Ratio = \frac{NP's\ MSE}{FDH's\ MSE}$.

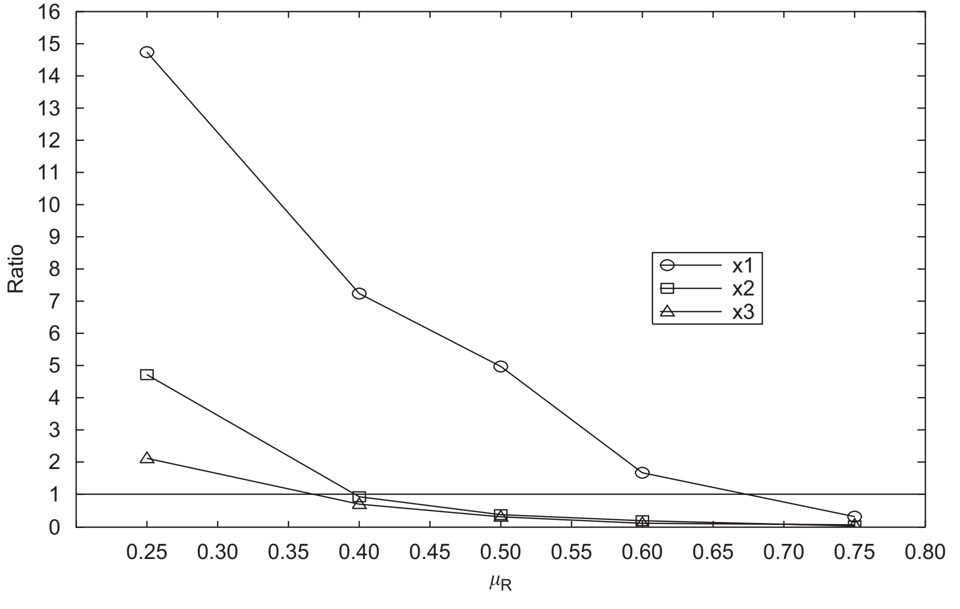

Fig. 2. Relative performance of NP and FDH estimators: $n = 400$, $\sigma_2(x)$, $Ratio = \frac{FDH's\ MSE}{NP's\ MSE}$.

Table 4 gives the overall performance of the efficiency estimators according to the measures described above. For comparison purposes, we provide in Figs. 4 and 5 a plot of the NP and bias corrected FDH frontier estimates. The jagged appearance of the graph for FDH (B–C FDH in Figs. 4 and 5) is due to the bias correction. We also include in the plots

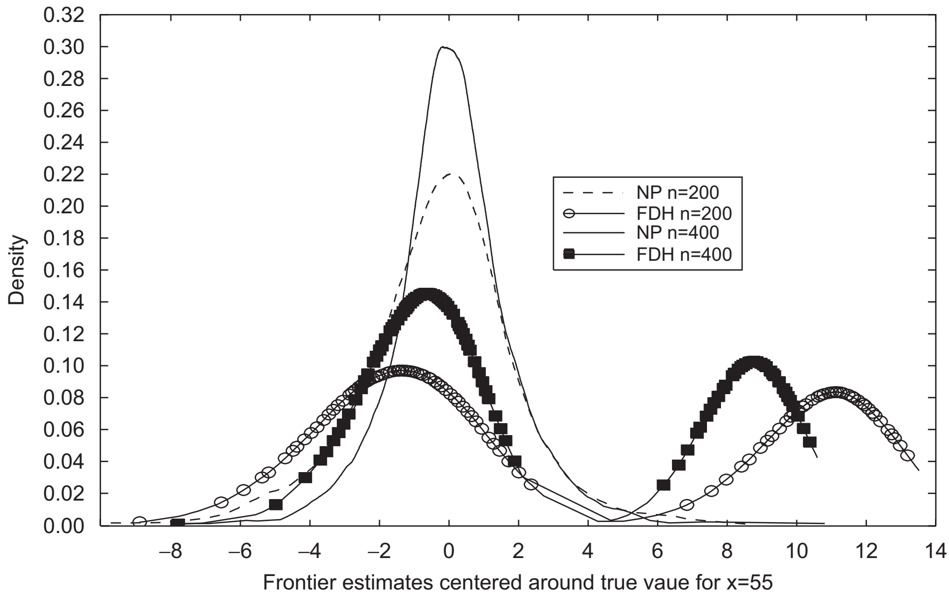

Fig. 3. Density estimates for NP and FDH estimators.

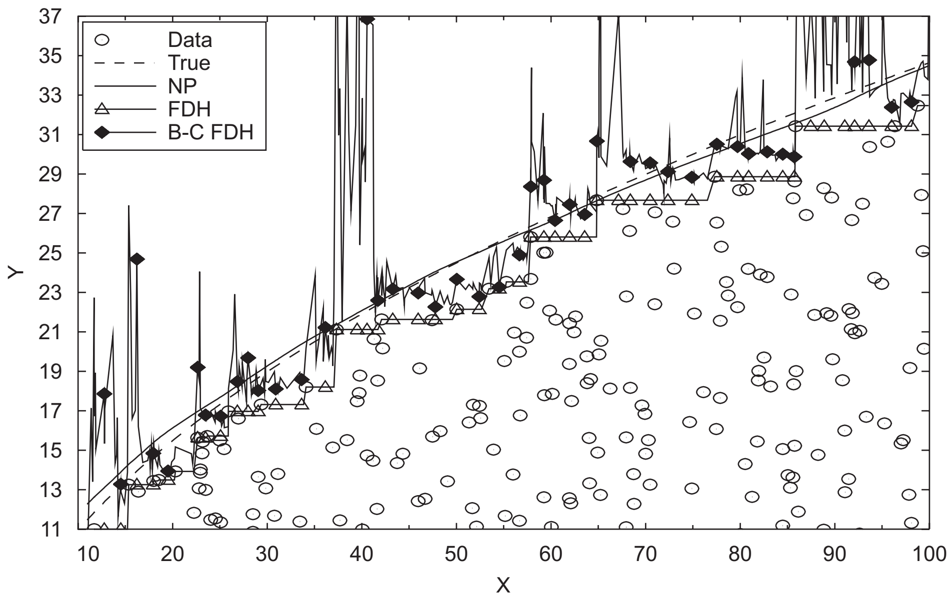

Fig. 4. Frontier estimates for NP and FDH estimators: $n = 400$, $\mu_R = 0.5$ and $\sigma_1(x)$.

a graph for the uncorrected FDH frontier estimates. The graphs are for $\sigma_1(x)$ and $\sigma_2(x)$ with $\mu_R = 0.5$ and $n = 400$. We first identify some general regularities on estimation performance.

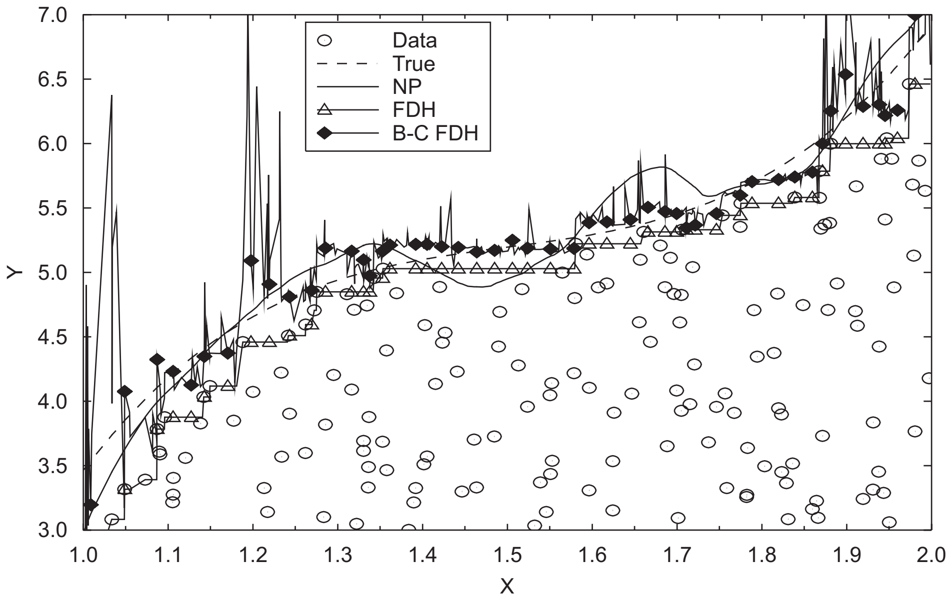

Fig. 5. Frontier estimates for NP and FDH estimators: $n = 400$, $\mu_R = 0.5$ and $\sigma_2(x)$.

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%, while that for FDH is usually below 95%. For both estimators there is no clear evidence that their empirical coverage probabilities get closer to 95% as n increases. Regarding the measures of overall performance for efficiency estimators mentioned above, both estimators perform better 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 of s_R increases but MSE oscillates, with the bias being negative. The bias of $\hat{\sigma}(x; h_n)$, which is negative for most experiments considered, does not 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 $\hat{\sigma}(x; h_n)$ oscillates with μ_R, which reflects the fact that the variance of $\hat{\sigma}(x; h_n)$ depends on μ_R in a nonlinear fashion, as indicated by Theorem 1. Following the prediction in Theorem 2, the bias of the NP frontier estimator is generally positive, except for small μ_R and $n = 100$, and the bias has a pattern of increasing with μ_R and MSE tend to 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 occurring when $\sigma(x) = \sigma_2(x)$. 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

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

$\sigma_1(x)$	n	S_R		$\hat{\sigma}(x_1): x_1 = 32.5$		$\hat{\sigma}(x_2): x_2 = 55$		$\hat{\sigma}(x_3): x_3 = 77.5$	
		Bias ($\times 10^{-1}$)	MSE ($\times 10^{-3}$)	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_R = 0.25$	200	−0.025	0.289	−0.057	0.253	−0.154	0.489	−0.085	0.573
	300	−0.052	0.225	−0.020	0.176	−0.139	0.353	−0.054	0.406
	400	−0.049	0.188	−0.026	0.123	−0.100	0.252	−0.017	0.286
$\mu_R = 0.4$	200	−0.067	0.196	−0.081	0.128	−0.148	0.264	−0.102	0.279
	300	−0.062	0.156	−0.039	0.090	−0.077	0.152	−0.053	0.224
	400	−0.060	0.118	−0.020	0.062	−0.059	0.113	−0.046	0.157
$\mu_R = 0.5$	200	−0.087	0.223	−0.055	0.105	−0.118	0.226	−0.104	0.254
	300	−0.075	0.156	−0.030	0.078	−0.074	0.140	−0.061	0.173
	400	−0.073	0.134	−0.023	0.059	−0.066	0.113	−0.053	0.152
$\mu_R = 0.6$	200	−0.118	0.318	−0.061	0.128	−0.108	0.240	−0.080	0.333
	300	−0.099	0.210	−0.048	0.103	−0.100	0.175	−0.045	0.225
	400	−0.092	0.173	−0.039	0.071	−0.082	0.145	−0.060	0.184
$\mu_R = 0.75$	200	−0.144	0.375	−0.061	0.216	−0.144	0.447	−0.057	0.596
	300	−0.124	0.275	−0.012	0.172	−0.118	0.296	−0.067	0.418
	400	−0.099	0.188	−0.027	0.129	−0.088	0.220	−0.020	0.356

$\sigma_2(x)$	n	S_R		$\hat{\sigma}(x_1): x_1 = 1.25$		$\hat{\sigma}(x_2): x_2 = 1.5$		$\hat{\sigma}(x_3): x_3 = 1.75$	
		Bias ($\times 10^{-1}$)	MSE ($\times 10^{-3}$)	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_R = 0.25$	200	−0.028	0.277	−0.022	0.019	−0.029	0.027	−0.017	0.028
	300	−0.057	0.251	−0.026	0.013	−0.031	0.017	−0.011	0.019
	400	−0.070	0.199	−0.022	0.011	−0.024	0.015	−0.007	0.014
$\mu_R = 0.4$	200	−0.084	0.234	−0.035	0.012	−0.026	0.013	−0.020	0.015
	300	−0.083	0.173	−0.026	0.007	−0.023	0.009	−0.009	0.010
	400	−0.079	0.150	−0.016	0.005	−0.020	0.007	−0.005	0.007
$\mu_R = 0.5$	200	−0.106	0.272	−0.034	0.010	−0.025	0.012	−0.017	0.012
	300	−0.103	0.218	−0.023	0.007	−0.025	0.009	−0.008	0.008
	400	−0.096	0.173	−0.017	0.005	−0.019	0.007	−0.013	0.007
$\mu_R = 0.6$	200	−0.151	0.410	−0.033	0.010	−0.036	0.016	−0.018	0.016
	300	−0.124	0.276	−0.026	0.008	−0.023	0.010	−0.011	0.010
	400	−0.116	0.223	−0.023	0.007	−0.020	0.008	−0.010	0.008
$\mu_R = 0.75$	200	−0.168	0.447	−0.027	0.018	−0.040	0.023	−0.006	0.024
	300	−0.140	0.304	−0.023	0.013	−0.030	0.016	−0.000	0.019
	400	−0.122	0.237	−0.023	0.012	−0.009	0.013	0.001	0.015

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 measure considered is R_{rank}, while in the case of R_{mag} and R_{rel}, FDH performs better as μ_R increases.

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

Table 2
Bias and MSE of nonparametric and FDH frontier estimators

$\sigma_1(x)$	n		$x_1 = 32.5$		$x_2 = 55$		$x_3 = 77.5$	
			NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	200	Bias	0.004	3.294	−0.290	3.347	0.029	3.203
		MSE	2.886	51.535	5.017	52.270	7.208	50.809
	300	Bias	0.327	2.642	0.010	2.723	0.415	3.010
		MSE	2.511	37.088	3.972	37.962	5.166	39.146
	400	Bias	0.279	2.418	0.122	2.527	0.517	2.521
		MSE	1.710	29.942	2.989	29.616	4.162	30.097
$\mu_R = 0.4$	200	Bias	0.174	2.700	0.074	2.766	0.350	2.575
		MSE	1.394	38.684	2.344	38.978	3.098	37.866
	300	Bias	0.280	2.049	0.274	1.947	0.458	2.034
		MSE	1.153	25.241	1.770	24.740	2.902	25.122
	400	Bias	0.326	1.535	0.309	1.765	0.451	1.863
		MSE	0.822	18.432	1.270	19.500	1.958	20.480
$\mu_R = 0.5$	200	Bias	0.440	2.226	0.399	2.002	0.613	2.123
		MSE	1.629	30.063	2.770	28.396	3.760	30.273
	300	Bias	0.431	1.738	0.429	1.868	0.613	1.772
		MSE	1.183	20.460	1.705	22.225	2.602	21.745
	400	Bias	0.443	1.573	0.444	1.427	0.624	1.382
		MSE	1.012	16.985	1.488	15.082	2.518	15.118
$\mu_R = 0.6$	200	Bias	0.825	1.839	0.957	1.939	1.322	1.727
		MSE	2.936	26.026	4.992	26.047	7.189	23.261
	300	Bias	0.686	1.559	0.740	1.607	1.179	1.385
		MSE	2.073	18.826	3.051	18.698	5.235	15.789
	400	Bias	0.655	1.021	0.726	1.191	1.020	1.213
		MSE	1.534	10.479	2.493	12.354	4.373	12.247
$\mu_R = 0.75$	200	Bias	2.107	1.766	2.362	1.660	3.459	1.740
		MSE	11.584	23.577	19.298	22.047	32.202	23.588
	300	Bias	2.010	1.474	2.055	1.446	2.834	1.328
		MSE	9.792	16.881	14.699	16.291	22.364	15.168
	400	Bias	1.481	1.028	1.643	1.201	2.402	1.053
		MSE	5.975	9.805	9.374	12.431	16.541	11.207

$\sigma_2(x)$	n		$x_1 = 1.25$		$x_2 = 1.5$		$x_3 = 1.75$	
			NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	200	Bias	−0.020	0.836	−0.043	0.458	0.008	0.204
		MSE	0.243	3.303	0.308	1.458	0.389	0.897
	300	Bias	0.013	0.506	0.002	0.391	0.089	0.174
		MSE	0.163	1.855	0.178	1.041	0.282	0.645
	400	Bias	0.048	0.540	0.047	0.317	0.127	0.131
		MSE	0.116	1.710	0.161	0.759	0.248	0.525
$\mu_R = 0.4$	200	Bias	0.020	0.466	0.059	0.176	0.093	0.012
		MSE	0.123	1.640	0.137	0.277	0.196	0.198
	300	Bias	0.045	0.297	0.067	0.130	0.128	0.012
		MSE	0.067	0.911	0.102	0.165	0.154	0.160
	400	Bias	0.075	0.177	0.070	0.084	0.134	−0.009

Table 2 (*continued*)

$\sigma_2(x)$	n		$x_1 = 1.25$		$x_2 = 1.5$		$x_3 = 1.75$	
			NP	FDH	NP	FDH	NP	FDH
		MSE	0.064	0.463	0.077	0.071	0.116	0.080
$\mu_R = 0.5$	200	Bias	0.066	0.238	0.114	0.123	0.161	−0.033
		MSE	0.109	0.751	0.142	0.145	0.212	0.085
	300	Bias	0.099	0.161	0.104	0.079	0.185	−0.025
		MSE	0.082	0.330	0.102	0.039	0.169	0.068
	400	Bias	0.107	0.150	0.113	0.064	0.150	−0.036
		MSE	0.064	0.318	0.080	0.029	0.122	0.036
$\mu_R = 0.6$	200	Bias	0.203	0.179	0.210	0.113	0.317	−0.024
		MSE	0.219	0.448	0.252	0.101	0.404	0.054
	300	Bias	0.164	0.154	0.197	0.060	0.272	−0.040
		MSE	0.140	0.323	0.170	0.029	0.287	0.041
	400	Bias	0.158	0.108	0.186	0.049	0.252	−0.030
		MSE	0.123	0.205	0.133	0.023	0.224	0.022
$\mu_R = 0.75$	200	Bias	0.552	0.174	0.527	0.076	0.792	−0.028
		MSE	0.895	0.426	0.927	0.042	1.595	0.059
	300	Bias	0.445	0.100	0.450	0.047	0.669	−0.031
		MSE	0.610	0.169	0.645	0.014	1.158	0.029
	400	Bias	0.365	0.090	0.479	0.034	0.580	−0.003
		MSE	0.472	0.143	0.606	0.009	0.907	0.045

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 $\mu_R = 0.25$, 0.4, 0.5 and 0.6, with exceptions in cases where $\sigma(x) = \sigma_2(x)$, while FDH is better with $\mu_R = 0.75$ (see Figs. 1 and 2). Specifically, when DGP uses $\sigma_1(x)$, NP outperforms FDH in almost all experiment designs, with a few exceptions when $\mu_R = 0.75$. When $\sigma_2(x)$ is used in the DGP with $\mu_R = 0.6$, FDH is better with exceptions where the frontier is estimated at the 25th percentile of X, in which case the NP outperforms FDH. The dominance of FDH over NP when $\mu_R = 0.75$ is most likely explained in this DGP by the fact that in this case $\sigma_R^2 = 0.04$—roughly half of its values in other DGPs—contributing to a higher variance of the NP estimator as suggested by Theorem 2. The relative performance of both frontier estimators in terms of MSE is illustrated in Figs. 1 and 2, where for different points and $n = 400$, the ratio of NP's MSE over FDH's MSE is plotted against μ_R for $\sigma_1(x)$, and the ratio of FDH's MSE over NP's MSE is plotted against μ_R for $\sigma_2(x)$ (similar graphs result when we examine the cases where $n = 200$ and 300).

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 $\mu_R = 0.75$. In terms of R_{mag}, NP generally outperforms FDH when $\mu_R = 0.25, 0.4, 0.5$, while FDH is better when $\mu_R = 0.6, 0.75$, with exceptions in $\mu_R = 0.6$ and $\sigma_1(x)$. Based on these results, it seems reasonable to conclude that when we are dealing

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

$\sigma_1(x)$	n	$x_1 = 32.5,\ y_1 = 10$		$x_2 = 55,\ y_2 = 10$		$x_3 = 77.5,\ y_3 = 10$	
		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	200	0.937	0.760	0.944	0.758	0.939	0.764
	300	0.929	0.751	0.925	0.755	0.940	0.786
	400	0.934	0.757	0.937	0.800	0.943	0.791
$\mu_R = 0.4$	200	0.942	0.808	0.947	0.814	0.955	0.804
	300	0.929	0.809	0.939	0.793	0.931	0.810
	400	0.922	0.784	0.915	0.802	0.934	0.816
$\mu_R = 0.5$	200	0.943	0.829	0.938	0.822	0.947	0.801
	300	0.934	0.829	0.941	0.825	0.948	0.810
	400	0.913	0.838	0.915	0.839	0.921	0.817
$\mu_R = 0.6$	200	0.943	0.824	0.925	0.825	0.946	0.833
	300	0.926	0.809	0.929	0.830	0.923	0.854
	400	0.935	0.826	0.902	0.828	0.909	0.853
$\mu_R = 0.75$	200	0.947	0.832	0.922	0.836	0.945	0.832
	300	0.914	0.847	0.890	0.847	0.919	0.831
	400	0.936	0.854	0.908	0.832	0.911	0.824

$\sigma_2(x)$	n	$x_1 = 1.25,\ y_1 = 3$		$x_2 = 1.5,\ y_2 = 3$		$x_3 = 1.75,\ y_3 = 3$	
		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	200	0.931	0.857	0.912	0.904	0.904	0.831
	300	0.931	0.816	0.943	0.917	0.905	0.819
	400	0.949	0.831	0.928	0.911	0.906	0.779
$\mu_R = 0.4$	200	0.938	0.871	0.940	0.939	0.935	0.792
	300	0.957	0.870	0.928	0.935	0.900	0.769
	400	0.944	0.849	0.927	0.945	0.904	0.763
$\mu_R = 0.5$	200	0.968	0.872	0.949	0.941	0.926	0.786
	300	0.950	0.896	0.945	0.953	0.918	0.737
	400	0.960	0.888	0.930	0.961	0.905	0.719
$\mu_R = 0.6$	200	0.966	0.878	0.935	0.961	0.916	0.761
	300	0.973	0.900	0.929	0.949	0.904	0.730
	400	0.962	0.901	0.929	0.949	0.915	0.747
$\mu_R = 0.75$	200	0.978	0.893	0.948	0.960	0.948	0.757
	300	0.973	0.902	0.942	0.955	0.915	0.740
	400	0.965	0.902	0.909	0.948	0.920	0.784

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 Fig. 3. The figure shows kernel density estimates for the frontier around the true value evaluated at $x = 55$ for NP ($\frac{\hat{\sigma}(x;h_n)}{s_R(g_n)} - \frac{\sigma(x)}{\sigma_R}$) and FDH ($\hat{\rho}_{\mathrm{FDH}}(x) - \frac{\sigma(x)}{\sigma_R}$) based on 1000 simulations, $\mu_R = 0.25$ and

Table 4
Overall efficiency estimators by nonparametric and FDH

$\sigma_1(x)$	n	R_{rank}		R_{mag} $(\times 10^{-2})$		R_{rel} $(\times 10^{-1})$	
		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	200	0.997	0.981	0.253	0.844	0.247	1.440
	300	0.998	0.987	0.171	0.621	0.208	1.233
	400	0.999	0.989	0.108	0.503	0.184	0.994
$\mu_R = 0.4$	200	0.994	0.970	0.235	0.776	0.293	1.492
	300	0.996	0.978	0.139	0.563	0.236	1.126
	400	0.997	0.983	0.083	0.439	0.197	0.915
$\mu_R = 0.5$	200	0.990	0.962	0.313	0.680	0.345	1.174
	300	0.993	0.973	0.161	0.477	0.289	0.950
	400	0.995	0.979	0.161	0.368	0.247	0.837
$\mu_R = 0.6$	200	0.976	0.955	0.696	0.554	0.478	1.024
	300	0.984	0.969	0.278	0.384	0.371	0.890
	400	0.988	0.977	0.193	0.283	0.327	0.687
$\mu_R = 0.75$	200	0.909	0.943	1.595	0.373	0.772	0.846
	300	0.927	0.962	0.780	0.236	0.639	0.583
	400	0.943	0.972	0.541	0.171	0.533	0.466

$\sigma_2(x)$	n	R_{rank}		R_{mag} $(\times 10^{-2})$		R_{rel} $(\times 10^{-1})$	
		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	200	0.996	0.985	0.251	0.594	0.294	0.946
	300	0.998	0.989	0.155	0.432	0.244	0.854
	400	0.998	0.991	0.124	0.362	0.213	0.661
$\mu_R = 0.4$	200	0.993	0.979	0.260	0.521	0.340	0.920
	300	0.996	0.985	0.164	0.374	0.276	0.843
	400	0.997	0.988	0.107	0.285	0.249	0.608
$\mu_R = 0.5$	200	0.987	0.974	0.956	0.458	0.409	0.933
	300	0.991	0.982	0.210	0.309	0.339	0.700
	400	0.993	0.986	0.151	0.239	0.299	0.563
$\mu_R = 0.6$	200	0.970	0.970	0.796	0.371	0.545	0.708
	300	0.979	0.979	1.120	0.255	0.451	0.556
	400	0.984	0.984	0.293	0.192	0.403	0.474
$\mu_R = 0.75$	200	0.889	0.959	1.304	0.274	0.880	0.596
	300	0.913	0.973	0.914	0.177	0.732	0.434
	400	0.926	0.979	0.831	0.129	0.654	0.407

$\sigma(x) = \sqrt{x}$, for $n = 200$ and 400. 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. Fig. 3 also shows that the estimated densities become tighter with more acute spikes as the sample size increases, as expected from the available asymptotic results.[9]

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, FDH as well as other 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 estimators of Cazals et al. (2002), Girard and Jacob (2004) and Knight (2001) 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. Also, it seems desirable to derive minimax convergence rates for our model.

Acknowledgments

We would like to thank the editors and two anonymous referees for their excellent comments and suggestions. We also thank R. Färe and S. Grosskopf for helpful discussions. The authors retain responsibility for any remaining errors.

Appendix

Proof of 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 $h_n\to 0$ as $n\to\infty$ where $n\in\{1,2,3,\ldots\}$. For any n, by the

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

Heine–Borel theorem there exists a finite collection of sets $\{B(x_k, (\frac{n}{h_n^3})^{-1/2})\}_{k=1}^{l_n}$ such that $G \subset \bigcup_{k=1}^{l_n} B(x_k, (\frac{n}{h_n^3})^{-1/2})$ for $x_k \in G$ with $l_n < (\frac{n}{h_n^3})^{1/2} r$. For $x \in B(x_k, (\frac{n}{h_n^3})^{-1/2})$,

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

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

Hence,

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

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

Since, $(\frac{nh_n}{\ln(n)})^{1/2} 2B_f m(nh_n)^{-1/2} \to 0$, then to prove (a) it suffices to show that there exists a constant $\Delta > 0$ such that for all $\varepsilon > 0$ there exists N such that for all $n > N$, $\mathrm{P}\Big((\frac{nh_n}{\ln(n)})^{1/2} \max_{1 \leqslant k \leqslant l_n} |s_0(x) - \mathrm{E}(s_0(x))| \geqslant \Delta\Big) \leqslant \varepsilon$. Let $\varepsilon_n = (\frac{\ln(n)}{nh_n})^{1/2}\Delta$. Then, for every n,

$$\mathrm{P}\left(\max_{1 \leqslant k \leqslant l_n} |s_0(x_k) - \mathrm{E}(s_0(x_k))| \geqslant \varepsilon_n\right) \leqslant \sum_{k=1}^{l_n} \mathrm{P}(|s_0(x_k) - \mathrm{E}(s_0(x_k))| \geqslant \varepsilon_n).$$

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

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

where $\bar{\sigma}^2 = n^{-1}\sum_{t=1}^{n} V(W_{tn}) = h_n^{-2}\mathrm{E}(K^2(\frac{X_t - x_k}{h_n}) f^2(X_t, R_t)) - (h_n^{-1}\mathrm{E}(K(\frac{X_t - x_k}{h_n}) f(X_t, R_t)))^2$. 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 Lebesgue's dominated convergence theorem, for some constant $B_{\bar{\sigma}^2}$. Let $c_n = 2h_n\bar{\sigma}^2 + \frac{2}{3}B_W\varepsilon_n$. Then, $\frac{-nh_n\varepsilon_n^2}{2h_n\bar{\sigma}^2 + \frac{2B_W\varepsilon_n}{3}} = \frac{-\Delta^2 \ln(n)}{c_n}$. Hence, for any $\varepsilon > 0$ there exists N such that for all $n > N$,

$$\mathrm{P}\left(\max_{1 \leqslant k \leqslant l_n} |s_0(x_k) - \mathrm{E}(s_0(x_k))| \geqslant \varepsilon_n\right) < 2l_n n^{-\Delta^2/c_n} < 2\left(\frac{n}{h_n^3}\right)^{1/2} r n^{-\Delta^2/c_n} < 2(nh_n^3)^{-1/2} r < \varepsilon$$

since $c_n \to 2B_{\bar{\sigma}^2}$ and therefore there exists $\Delta^2 > 2B_{\bar{\sigma}^2}$.

(b) The result follows directly from part (a) and the assumption that $\frac{nh_n^3}{\ln(n)} \to \infty$. □

Proof of Lemma 2. Let $\hat{r}' = (\hat{r}_1, \ldots, \hat{r}_n)$ with $\hat{r}_t = \sigma^2(X_t)\varepsilon_t^2 + (m(X_t) - \hat{m}(X_t; h_n))^2 + 2(m(X_t) - \hat{m}(X_t; h_n))\sigma(X_t)\varepsilon_t$,

$$S_n(x) = (nh_n)^{-1}\begin{pmatrix} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) & \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\frac{X_t - x}{h_n} \\ \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\frac{X_t - x}{h_n} & \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 \end{pmatrix}$$

and $S(x) = \begin{pmatrix} g_X(x) & 0 \\ 0 & g_X(x)\sigma_K^2 \end{pmatrix}$.

Then, $\hat{\sigma}^2(x;h_n) - \sigma^2(x) = \frac{1}{nh_n}\sum_{t=1}^n W_n(\frac{X_t-x}{h_n}, x)r_t^*$ 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 $A_n(x) \equiv \hat{\sigma}^2(x;h_n) - \sigma^2(x) - \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K(\frac{X_t-x}{h_n})r_t^*$, then

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

where the inequality follows from the Cauchy–Schwarz Inequality and the fact that for a set a_i, $i = 1, \ldots, n$ of positive numbers $\sum_{i=1}^n a_i^2 \leqslant (\sum_{i=1}^n a_i)^2$. By part (b) of Lemma 1, $B_n(x) \equiv \frac{1}{h_n}((1,0)(S_n^{-1}(x) - S^{-1}(x))^2(1,0)')^{1/2} = O_p(1)$ uniformly in G. Hence, if we put $R_{n,1}(x) \equiv n^{-1}(|\sum_{t=1}^n K(\frac{X_t-x}{h_n})r_t^*| + |\sum_{t=1}^n K(\frac{X_t-x}{h_n})\frac{X_t-x}{h_n}r_t^*|)$ the proof is complete. □

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

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

Since $B_n(x) = O_p(1)$ uniformly in G, from part (b) of 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

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

$$I_{3n}(x) = \frac{2}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)\varepsilon_t(\hat{m}(X_t; h_n) - m(X_t)),$$
$$I_{4n}(x) = \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)(m(X_t) - \hat{m}(X_t; h_n))^2$$

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 $I_{1n} = \frac{h_n}{2ng_X(x)} \sum_{t=1}^{n} K(\frac{X_t - x}{h_n})(\frac{X_t - x}{h_n})^2 \sigma^{2(2)}(X_{tb})$. Given A1(2) and A2(6) we have

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

$I_{2n}(x)$: note that by Assumption A1(2)

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

where the last equality follows from part (a) in Lemma 1 with $f(X_t, R_t) = \sigma^2(X_t)(\varepsilon_t^2 - 1)$, 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

$$I_{31n}(x) = \frac{2}{n^2 h_n^2 g_X(x)} \sum_{t=1}^{n}\sum_{k=1}^{n} \frac{1}{g_X(X_t)} K\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_k - X_t}{h_n}\right)\sigma(X_t)\sigma(X_k)\varepsilon_t\varepsilon_k,$$

$$\begin{aligned}I_{32n}(x) = &\frac{h_n^2}{n^2 h_n^2 g_X(x)} \sum_{t=1}^{n}\sum_{k=1}^{n} \frac{1}{g_X(X_t)} K\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_k - X_t}{h_n}\right)\left(\frac{X_k - X_t}{h_n}\right)^2\\ &\times\sigma(X_t)\varepsilon_t m^{(2)}(X_{kt}),\end{aligned}$$

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

where $Y_k^* = Y_k - m(X_t) - m^{(1)}(X_t)(X_k - X_t)$. We now examine each of these terms separately. Note that,

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

Since $|\sigma(X_k)\varepsilon_k| < C$ for a generic constant C. If $nh_n^3 \to \infty$ we have by part (a) of Lemma 1,

$$\sup_{x\in G} \frac{1}{nh_n}\left|\sum_{k=1}^{n} K\left(\frac{X_k - x}{h_n}\right)\sigma(X_k)\varepsilon_k\right| = \mathrm{O_p}\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}\right).$$

Therefore, $\sup_{x\in G}|I_{31n}(x)| \leqslant 2\underline{B}_{g_X}^{-1}\mathrm{O_p}((\frac{nh_n}{\ln(n)})^{-1/2})\sup_{x\in G}|\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K(\frac{X_t-x}{h_n})\sigma(X_t)|\varepsilon_t||$.

Since $|\frac{\sigma(X_t)|\varepsilon_t|}{g_X(X_t)}| < C$,

$$\begin{aligned}
&\sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)|\varepsilon_t|\right| \\
&\leqslant \sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)|\varepsilon_t|\right.\right. \\
&\quad \left.\left. -\mathrm{E}\left(\frac{1}{g_X(X_t)}h_n^{-1}K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)|\varepsilon_t|\right)\right)\right| \\
&\quad + \sup_{x\in G}\mathrm{E}\left(\frac{1}{g_X(X_t)}\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)|\varepsilon_t|\right) \\
&= \mathrm{O_p}\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}\right) + \frac{1}{h_n}\sup_{x\in G}\mathrm{E}\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)|\varepsilon_t|\right),
\end{aligned}$$

by part (a) of Lemma 1. Now, $\frac{1}{h_n}\mathrm{E}(\frac{1}{g_X(X_t)}K(\frac{X_t-x}{h_n})\sigma(X_t)|\varepsilon_t|) = \int K(\phi)\sigma(x + h_n\phi)\mu_1(x + h_n\phi)\,\mathrm{d}\phi$ and by Lebesgue's dominated convergence theorem,

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

given Assumption A2(4) and the fact that $\mu_1(X_t)$ is uniformly bounded in G. Therefore,

$$\frac{1}{h_n}\sup_{x\in G}\mathrm{E}\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)|\varepsilon_t|\right) = \mathrm{O}(1)$$

and consequently $\sup_{x\in G}|I_{31n}(x)| = \mathrm{O_p}((\frac{nh_n}{\ln(n)})^{-1/2})$. Now, by Assumptions A2(1) and A2(6)

$$\begin{aligned}
|I_{32n}(x)| &\leqslant \underline{B}_{g_X}^{-1} b B_{2\sigma}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)|\varepsilon_t| \\
&\quad \times \sup_{x\in G}\left|\frac{1}{n}\sum_{k=1}^{n}K\left(\frac{X_k - x}{h_n}\right)\left(\frac{X_k - x}{h_n}\right)^2\right|.
\end{aligned}$$

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

$O_p(h_n^2)$. From Lemma 2

$$|D_n(X_t)| \equiv \left| \hat{m}(X_t; h_n) - m(X_t) - \frac{1}{nh_n g_X(X_t)} \sum_{k=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) Y_k^* \right| \leqslant B_n(X_t) R_{n,2}(X_t).$$

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

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

where $R_{11}(X_t) = \frac{1}{n}\sum_{k=1}^{n} K(\frac{X_k - X_t}{h_n})\sigma(X_k)\varepsilon_k$, $R_{12}(X_t) = \frac{1}{2n}\sum_{k=1}^{n} K(\frac{X_k - X_t}{h_n})(X_k - X_t)^2 m^{(2)}(X_{kt})$, $R_{21}(X_t) = \frac{1}{n}\sum_{k=1}^{n} K(\frac{X_k - X_t}{h_n})(\frac{X_k - X_t}{h_n})\sigma(X_k)\varepsilon_k$ and $R_{22}(X_t) = \frac{h_n^2}{2n}\sum_{k=1}^{n} K(\frac{X_k - X_t}{h_n})(\frac{X_k - X_t}{h_n})^3 m^{(2)}(X_{kt})$.

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 $E(\varepsilon_t|X_t) = 0$ 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

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

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 $\sup_{x \in G} |I_{3n}| = O_p(h_n^2) + O_p((\frac{nh_n}{\ln(n)})^{-1/2})$.

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

$$I_{41n}(x) = \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{1}{n^2 h_n^2 g_X^2(X_t)} \sum_{k=1}^{n}\sum_{l=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) \times K\left(\frac{X_l - X_t}{h_n}\right) \sigma(X_t)\sigma(X_k)\varepsilon_k\varepsilon_l,$$

$$I_{42n}(x) = \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{1}{4n^2 h_n^2 g_X^2(X_t)} \sum_{k=1}^{n}\sum_{l=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right)(X_k - X_t)^2 \times K\left(\frac{X_l - X_t}{h_n}\right)(X_l - X_t)^2 m^{(2)}(X_{kt}) m^{(2)}(X_{lt}),$$

$$I_{43n}(x) = \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) D_n^2(X_t),$$

$$I_{44n}(x) = \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{1}{n^2 h_n^2 g_X^2(X_t)} \sum_{k=1}^{n}\sum_{l=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) K\left(\frac{X_l - X_t}{h_n}\right) \times (X_l - X_t)^2 m^{(2)}(X_{lt})\sigma(X_k)\varepsilon_k,$$

$$I_{45n}(x)=\frac{1}{nh_n g_X(x)}\sum_{t=1}^{n}K\left(\frac{X_t-x}{h_n}\right)\frac{2D_n(X_t)}{nh_n g_X(X_t)}\sum_{k=1}^{n}K\left(\frac{X_k-X_t}{h_n}\right)\sigma(X_k)\varepsilon_k,$$
$$I_{46n}(x)=\frac{1}{nh_n g_X(x)}\sum_{t=1}^{n}K\left(\frac{X_t-x}{h_n}\right)\frac{D_n(X_t)}{nh_n g_X(X_t)}\sum_{k=1}^{n}K\left(\frac{X_k-X_t}{h_n}\right)m^{(2)}(X_{kt})(X_k-X_t)^2.$$

We now examine each term separately. First,

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

where $I_{411}(X_t)=\frac{1}{nh_n g_X(X_t)}\sum_{l=1}^{n}K(\frac{X_l-X_t}{h_n})\sigma(X_l)\varepsilon_l$. But,

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

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

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

Now,

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

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

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

Now,

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

$I_{441}(X_t) = \frac{1}{nh_n g_X(X_t)} \sum_{k=1}^{n} K(\frac{X_k - X_t}{h_n})\sigma(X_k)\varepsilon_k$, $I_{442}(X_t) = \frac{1}{nh_n g_X(X_t)} \sum_{k=1}^{n} K(\frac{X_k - X_t}{h_n})(X_k - X_t)^2 m^{(2)}$ (X_{kt}). But given that $\sup_{X_t \in G} I_{441}(X_t) \leqslant \underline{\mathrm{B}}_{g_X}^{-1} h_n^{-1} \sup_{X_t \in G} |R_{11}(X_t)| = \mathrm{o_p}(h_n)$ and

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

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

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

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

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

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

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

(b) From part (a), provided that $h_n^2 \ln(n) \to 0$ we can concentrate on $\frac{1}{\sqrt{nh_n}\, g_X(x)} \sum_{t=1}^{n} K(\frac{X_t - x}{h_n}) r_t^*$ to obtain the asymptotic distribution of $\sqrt{nh_n}(\hat{\sigma}^2(x; h_n) - \sigma^2(x))$. $I_{1n}(x) = \frac{1}{2} \frac{h_n}{n} \frac{1}{g_X(x)} \sum_{t=1}^{n} K(\frac{X_t - x}{h_n})(\frac{X_t - x}{h_n})^2 \sigma^{2(2)}(X_{tb})$, and given A1,

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

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

for $|\theta| \leqslant 1$. Given Assumptions A1, A2(5) and A3 and by Lebesgue's dominated convergence theorem,

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

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

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

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

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

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

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

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

But,

$$\begin{aligned} &\frac{1}{h_n}\mathrm{E}\left|K\left(\frac{X_t - x}{h_n}\right)\sigma^2(X_t)(\varepsilon_t^2 - 1)\right|^{2+\delta} \\ &\quad = \frac{1}{h_n}\mathrm{E}\left(K^{2+\delta}\left(\frac{X_t - x}{h_n}\right)(\sigma^2(X_t))^{2+\delta}\mathrm{E}(|\varepsilon_t^2 - 1|^{2+\delta}|X_t)\right) \\ &\quad \leqslant \frac{C}{h_n}\mathrm{E}\left(K^{2+\delta}\left(\frac{X_t - x}{h_n}\right)\right) \to C g_X(x)\int K^{2+\delta}(x)\,\mathrm{d}x, \end{aligned}$$

where the inequality follows from A1, A2(2), 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 Appendix,

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

where $\psi_n(Z_t, Z_k) = h_{tk} + h_{kt}$, $h_{tk} = \frac{1}{g_X(x)h_n^2}\frac{1}{g_X(X_t)} K(\frac{X_t - x}{h_n})K(\frac{X_k - X_t}{h_n})\sigma(X_t)\sigma(X_k)\varepsilon_t\varepsilon_k$, $Z_t = (X_t, \varepsilon_t)$. Letting $\vec{x} = (X_1, \ldots, X_n)$ and given our assumptions,

$$\mathrm{E}(\sqrt{nh_n}I_{311}) = \frac{2K(0)}{\sqrt{nh_n}g_X(x)}\frac{1}{h_n}\int K\left(\frac{z - x}{h_n}\right)\frac{\sigma^2(z)}{g_X(z)} g_X(z)\,\mathrm{d}z,$$

$$\mathrm{E}_{g_X}(V(\sqrt{nh_n}I_{311}|\vec{x})) = \frac{4K(0)}{n^2h_n^2g_X^2(x)}\frac{1}{h_n}\int K^2\left(\frac{z-x}{h_n}\right)\frac{\sigma^4(z)}{g_X^2(z)}(\mu^4(z)-1)g_X(z)\,\mathrm{d}z$$

and

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

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

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

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

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

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

Given our assumptions, if $nh_n^2 \to \infty$, by Lebesgue's dominated convergence theorem we have $\mathrm{E}(U_1) \to 0$. 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 $\sqrt{nh_n}I_{31n} = \mathrm{o_p}(1)$. Now we turn to the analysis of $I_{32n}(x)$. Using the notation of Lemma 3 we have $I_{32n}(x) = \frac{n-1}{2n}\frac{1}{g_X(x)}\binom{n}{2}^{-1}\sum_{t<k}\psi_n(Z_t,Z_k)$ where $\psi_n(Z_t,Z_k) = h_{tk} + h_{kt}$ and

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

and $Z_t = (X_t, \varepsilon_t)$. Given our assumptions $\mathrm{E}(\psi_n(Z_t, Z_k)) = 0$ and

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

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

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

Using Lebesgue's dominated convergence theorem we have $V(\sqrt{n}\hat{u}_n) \to 0$ and consequently by Lemma 3, $\sqrt{n}I_{32n} = \mathrm{o_p}(1)$ provided that $\mathrm{E}(\psi_n^2(Z_t, Z_j)) = \mathrm{o}(n)$. Now,

$$\begin{aligned}
&\frac{1}{n}\mathrm{E}(\psi_n^2(Z_t, Z_j)) \\
&= \frac{1}{4nh_n^4}\int\int K^2\left(\frac{X_t - X_j}{h_n}\right)K^2\left(\frac{X_t - x}{h_n}\right)(X_t - X_j)^4 \\
&\quad \times \frac{\sigma^2(X_t)\varepsilon_t^2 m^{(2)2}(X_{tj})}{g_X^2(X_t)}\, g_X(X_t)g_X(X_j)\,\mathrm{d}X_t\,\mathrm{d}X_j \\
&\quad + \frac{1}{4nh_n^4}\int\int k^2\left(\frac{X_t - X_j}{h_n}\right)K^2\left(\frac{X_j - x}{h_n}\right)(X_t - X_j)^4 \\
&\quad \times \frac{\sigma^2(X_j)\varepsilon_j^2 m^{(2)2}(X_{tj})}{g_X^2(X_j)}\, g_X(X_t)g_X(X_j)\,\mathrm{d}X_t\,\mathrm{d}X_j \\
&\quad + \frac{2}{4nh_n^4}\int\int K^2\left(\frac{X_t - X_j}{h_n}\right)K\left(\frac{X_j - x}{h_n}\right)K\left(\frac{X_t - x}{h_n}\right)(X_t - X_j)^4 \\
&\quad \times \frac{\sigma(X_t)\sigma(X_j)\varepsilon_t\varepsilon_j m^{(2)}(X_{tj})m^{(2)}(X_{tj})}{g_X(X_t)g_X(X_j)} \times g_X(X_t)g_X(X_j)\,\mathrm{d}X_t\,\mathrm{d}X_j \\
&= U_1 + U_2 + U_3.
\end{aligned}$$

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) $I_{33n} = \mathrm{o_p}(h_n^2)$ we have that by combining all terms $I_{3n}(x) = \mathrm{o_p}(n^{-1/2}) + \mathrm{o_p}(h_n^2)$. Finally, since we have already established in part (a) that $I_{4n}(x) = \mathrm{o_p}(h_n^2)$, combining all convergence results for $I_{1n}(x), I_{2n}(x), I_{3n}(x)$ and $I_{4n}(x)$ we have that if $nh_n^5 = \mathrm{O}(1)$, then

$$\sqrt{nh_n}(\hat{\sigma}^2(x; h_n) - \sigma^2(x) - B_{0n}) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma^4(x)}{g_X(x)}\,(\mu_4(x) - 1)\int K^2(y)\,\mathrm{d}y\right), \tag{6}$$

for all $x \in G$ where $B_{0n} = \frac{h_n^2 \sigma_K^2}{2} \sigma^{2(2)}(x) + o_p(h_n^2)$. It is a direct consequence of part (a) that $\hat{\sigma}(x; h_n)$ is uniformly consistent. Hence, noting that

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

for $\sigma_b^2(x) = \theta\sigma^2(x) + (1-\theta)\hat{\sigma}^2(x;h_n)$ for some $0 \leqslant \theta \leqslant 1$ we have that by (3) and the uniform consistency of $\hat{\sigma}(x;h_n)$ in G

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

where $B_{1n} = \frac{h_n^2 \sigma_K^2}{4\sigma(x)} \sigma^{2(2)}(x) + o_p(h_n^2)$. □

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

$$P\left(L_n^{-1} \sup_{X_t \in G} \left|\frac{\sigma(X_t)}{\hat{\sigma}(X_t;g_n)} - 1\right| < \Delta\right) \geqslant P\left(L_n^{-1} \sup_{X_t \in G} |\sigma(X_t) - \hat{\sigma}(X_t;g_n)| < \Delta'\right).$$

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

$$P\left(L_n^{-1} \sup_{X_t \in G} \left|\frac{\sigma(X_t)}{\hat{\sigma}(X_t;g_n)} - 1\right| < \Delta\right) > 1 - \delta. \tag{7}$$

Now suppose that $\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} - 1 \geqslant 0$. Then, $|\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} - 1| \leqslant \sup_{X_t \in G} \frac{\sigma(X_t)}{\hat{\sigma}(X_t;g_n)} - 1$ and $L_n^{-1} |\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} - 1| \leqslant L_n^{-1} |\sup_{X_t \in G} \frac{\sigma(X_t)}{\hat{\sigma}(X_t;g_n)} - 1|$. By inequality (7)

$$P\left(L_n^{-1} \left|\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} - 1\right| < \Delta\right) > 1 - \delta.$$

Now suppose that $\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} - 1 < 0$. Then, $|\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} - 1| \leqslant 1 - \max_{1 \leqslant t \leqslant n} R_t \inf_{X_t \in G} \frac{\sigma(X_t)}{\hat{\sigma}(X_t;g_n)}$ and

$$P\left(L_n^{-1} \left|\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} - 1\right| < \Delta\right) \geqslant P\left(\max_{1 \leqslant t \leqslant n} R_t \inf_{X_t \in G} \frac{\sigma(X_t)}{\hat{\sigma}(X_t;g_n)} > 1 - L_n\Delta\right).$$

By inequality (7) and assumption (2) in the statement of the theorem, for all $\delta > 0$ there is some $\Delta_1, \Delta > 0$ such that whenever $n > N_\delta$, $P(\inf_{X_t \in G} \frac{\sigma(X_t)}{\hat{\sigma}(X_t;g_n)} > 1 - L_n\Delta) > 1 - \delta$ and $P(\max_{1 \leqslant t \leqslant n} R_t > 1 - L_n\Delta_1) > 1 - \delta$. Hence, for all $\delta > 0$ there is some $\Delta_2 > 0$ such that whenever $n > N_\delta$ $P(\max_{1 \leqslant t \leqslant n} \frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t;g_n)} > 1 - L_n\Delta_2) > 1 - \delta$.

(b) Note that $\sqrt{nh_n}(\frac{\hat{\sigma}(x;h_n)}{\sigma_R}-\frac{\sigma(x)}{\sigma_R}-\frac{B_{1n}}{\sigma_R}) \equiv \sqrt{nh_n}(\frac{\hat{\sigma}(x;h_n)}{s_R(g_n)}-\frac{\sigma(x)}{\sigma_R}-\hat{\sigma}(x;h_n)(s_R(g_n)^{-1}-\sigma_R^{-1})-\frac{B_{1n}}{\sigma_R})$. From part (b) of Theorem 1 we have that $\sqrt{nh_n}(\frac{\hat{\sigma}(x;h_n)}{\sigma_R}-\frac{\sigma(x)}{\sigma_R}-\frac{B_{1n}}{\sigma_R}) \xrightarrow{\text{d}} \text{N}(0, \frac{\sigma^2(x)}{4\sigma_R^2 g_X(x)}(\mu_4(x)-1)\int K^2(y)\,\text{d}y)$, and from part (a), provided that $\frac{ng_n^5}{\ln(n)} \to \infty$ we have that $\hat{\sigma}(x;h_n)(s_R(g_n)^{-1}-\sigma_R^{-1}) = \text{O}_\text{p}(g_n^2)$. Hence, given that $nh_n^5 \to 0$ and $nh_ng_n^4 = \text{O}(1)$

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

where $B_{2n} = \text{O}_\text{p}(g_n^2)$. □

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

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

$$V(n^{1/2}(u_n-\hat{u}_n)) = n\text{E}\left(\left(\sum_{j=2}^k \binom{k}{j} H_n^{(j)}\right)^2\right) = n\text{E}\left(\sum_{j'=2}^k\sum_{j=2}^k \binom{k}{j}\binom{k}{j'} H_n^{(j)} H_n^{(j')}\right)$$
$$= n\sum_{j=2}^k \binom{k}{j}^2 \binom{n}{j}^{-1} \text{E}(h_n^{(j)}(Z_1,\ldots,Z_j)^2),$$

where the last equality follows from Theorem 3 in Lee (1990, p. 30). By Chebyshev's inequality, for all $\varepsilon > 0$, $\text{P}(|n^{1/2}(u_n-\hat{u}_n)| \geqslant \varepsilon) \leqslant n\text{E}((u_n-\hat{u}_n)^2)/\varepsilon^2$. Therefore, it suffices to show that

$$n\sum_{j=2}^k \binom{k}{j}^2 \binom{n}{j}^{-1} \text{E}(h_n^{(j)}(Z_1,\ldots,Z_j)^2) = \text{o}(1).$$

If for all $j = 2,\ldots,k$

$$\text{E}((h_n^{(j)}(Z_1,\ldots,Z_k))^2) = \text{O}(\text{E}(\psi_n^2(Z_1,\ldots,Z_k))) \tag{8}$$

then for some $\Delta > 0$,

$$n\text{E}((u_n-\hat{u}_n)^2) \leqslant n\sum_{j=2}^k \binom{k}{j}^2 \binom{n}{j}^{-1} \Delta\text{E}(\psi_n^2(Z_1,\ldots,Z_k))$$
$$= n^2\sum_{j=2}^k \binom{k}{j}^2 \frac{(n-j)!j!}{n!} n^{-1}\Delta\text{E}(\psi_n^2(Z_1,\ldots,Z_k)).$$

Since $E(\psi_n^2(Z_1,\ldots,Z_k)) = o(n)$ 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 $\Delta' > 0$, $nE((u_n - \hat{u}_n)^2) \leqslant \Delta' n^2 \binom{k}{2}^2 \frac{(n-2)!2!}{n!} \frac{E(\psi_n^2(Z_1,\ldots,Z_k))}{n} \leqslant O(1)o(1)$. We now use induction to prove that $E((h_n^{(j)}(Z_1,\ldots,Z_k))^2) = O(\psi_n^2(Z_1,\ldots,Z_k))$. Note that

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

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

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

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

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

and by Theorem 3 in Lee (1990)

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

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 $O(E(\psi_n^2(Z_1,\ldots,Z_k)))$. Second, again by Theorem 3 in Lee (1990)

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

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

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

Proof. We write $\hat{b} - b = \theta_1 - \theta_2 + \theta_3 + \theta_4 - \theta_5$, where

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

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

$$\theta_3 = \frac{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)\varepsilon_t}{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)},$$

$$\theta_4 = \frac{\frac{1}{n}\sum_{t=1}^n \sigma(X_t)(\hat{\sigma}(X_t;h_n) - \sigma(X_t))\varepsilon_t}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t;h_n)} \quad \text{and,}$$

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

Under Assumptions A1–A4 a routine application of Kolmogorov's law of large numbers gives $\theta_3 = \mathrm{o_p}(1)$. Now,

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

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

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

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

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 $D_{2n}\leqslant n^{-1}\sum_{t=1}^n\frac{|Y_t|}{2\sigma(X_t)}\sup_{X_t\in G}|\hat{\sigma}^2(X_t;h_n)-\sigma^2(X_t)|=o_p(h_n)n^{-1}\sum_{t=1}^n\frac{|Y_t|}{2\sigma(X_t)}$ $=o_p(h_n)n^{-1}\sum_{t=1}^n\frac{1}{2}|b+\varepsilon_t|=o_p(h_n)$, where the last equality follows from $n^{-1}\sum_{t=1}^n\frac{1}{2}|b+\varepsilon_t|=O_p(1)$ by Chebyshev's inequality. Hence, $\theta_1+\theta_4=o_p(h_n)$. Now, $|\theta_2|\leqslant\left|\frac{b}{\sum_{t=1}^n\hat{\sigma}^2(X_t;h_n)}\right||n^{-1}\sum_{t=1}^n(\hat{\sigma}^2(X_t;h_n)-\sigma^2(X_t))|=O_p(1)o_p(h_n)=o_p(h_n)$ 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 $\hat{b}-b=o_p(1)$. □

References

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

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

Banker, R., 1993. Maximum likelihood, consistency and data development analysis: a statistical foundation. Management Science 39, 1265–1273.

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

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

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

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

Fan, J., 1992. Design adaptive nonparametric regression. Journal of the American Statistical Association 87, 998–1004.

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

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660.

Farrell, M., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society A 120, 253–290.

Geffroy, J., 1964. Sur un problème d'estimation géométrique. Publications de l'Institut de Statistique de l'Université de Paris XIII, pp. 191–200.

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

Girard, S., Jacob, P., 2004. Extreme values and kernel estimates of point processes boundaries. ESAIM: Probability and Statistics 8, 150–168.

Hadri, K., 1999. Estimation of a doubly heteroscedastic stochastic frontier cost function. Journal of Business and Economic Statistics 17, 359–363.

Hall, P., Carroll, R., 1989. Variance function estimation in regression: the effect of estimation of the mean. Journal of the Royal Statistical Society B 51, 3–14.

Hall, P., Park, B., Stern, S., 1998. On polynomial estimators of frontiers and boundaries. Journal of Multivariate Analysis 66, 71–98.

Hoeffding, W., 1961. The strong law of large numbers for U-statistics (mimeo series 302). Institute of Statistics, University of North Carolina, Chapel Hill, NC.

Knight, K., 2001. Limiting distributions of linear programming estimators. Extremes 4, 87–103.

Korostelev, A.P., Simar, L., Tsybakov, A.B., 1995. Efficient estimation of monotone boundaries. Annals of Statistics 23, 476–489.

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

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

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

Powell, J., Stock, J., Stoker, T., 1989. Semiparametric estimation of index coefficients. Econometrica 57, 1403–1430.

Ruppert, D., Sheather, S., Wand, M., 1995. An effective bandwidth selector for least squares regression. Journal of the American Statistical Association 90, 1257–1270.

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

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.

Stone, C.J., 1977. Consistent nonparametric regression. Annals of Statistics 5, 595–620.