Local Exponential Frontier Estimation*

Carlos Martins-Filho**
Hudson S. Torrent***
Flavio A. Ziegelmann****

Abstract

In this paper we propose a local exponential estimator for a multiplicative nonparametric frontier model first introduced by Martins-Filho and Yao (2007). We improve their estimation procedure by adopting a variant of the local exponential smoothing introduced in Ziegelmann (2002). Our estimator is shown to be consistent and asymptotically normal under mild regularity conditions. In addition, due to local exponential smoothing, potential negativity of conditional variance functions that may hinder the use of Martins-Filho and Yao's estimator is avoided. A Monte Carlo study is performed to shed light on the finite sample properties of the estimator and to contrast its performance with that of the estimator proposed in Martins-Filho and Yao (2007). We also conduct an empirical exercise in which a production function and associated efficiencies for branches of financial institutions in the United States are estimated.

Keywords: Nonparametric Frontier Models, Local Exponential Smoothing, Local Exponential Regression.

JEL Codes: C14, C21.

---

*Submitted in June 2014. Revised in July 2014.

**Department of Economics, University of Colorado, Boulder, CO 80309-0256, USA, and IFPRI, 2033 K Street NW, Washington, DC 20006-1002, USA. E-mail: `carlos.martins@colorado.edu` and `c.martins-filho@cgiar.org`

***Department of Statistics – PPGE, Federal University of Rio Grande do Sul, Porto Alegre, RS 90040-000, Brazil. The author wishes to thank CNPq (process 485561/2013-1) for financial support. E-mail: `hudson.torrent@ufrgs.br`

****Department of Statistics – PPGE and PPGA, Federal University of Rio Grande do Sul, Porto Alegre, RS 90040-000, Brazil. The author wishes to thank CNPq (processes 305290/2012-6 and 485561/2013-1) and FAPERGS (process 1994/12-6) for financial support. E-mail: `flavioz@ufrgs.br`

1. Introduction

Economists have been concerned with the theoretical and empirical measurement of productive efficiency for at least fifty years, *viz.*, the seminal work of Koopmans (1951), Debreu (1951) and Farrell (1957). The microeconomic theoretical foundations for such measurement is relatively well established (Färe et al. (1985), Färe et al. (2008)). However, the specification of flexible statistical models and accompanying estimators of production frontiers that permit robust empirical investigations of efficiency is the object of a much more recent and developing literature in Econometrics and Statistics.[1]

From an econometric perspective, the main objective in this literature can be stated simply. Let $\Psi = \{(x, y) \in \Re_+^{p+1} : x \text{ can produce } y\}$ be a technology where $x \in \Re_+^p$ is a vector of inputs used to produce an output $y \in \Re_+$. The production frontier associated with Ψ is defined as $\rho(x) = \sup\{y \in \Re_+ : (x, y) \in \Psi\}$ for all $x \in \Re_+^p$. Given a sample of n realized production plans (or production units) $\chi_n = \{(X_i, Y_i)\}_{i=1}^n$, which share the technology Ψ, the principal goal of this literature is to estimate $\rho(x)$ for any $x \in \Re_+^p$. For an arbitrary production plan $(X_i, Y_i) \in \Psi$, we define its (inverse) Farrell efficiency as $0 \leq R_i = \frac{Y_i}{\rho(X_i)} \leq 1$. Once an estimate of ρ is available, estimated efficiencies can be readily obtained. Estimated efficiencies can then be used to construct relative and absolute efficiency rankings for the observed production plans or units. It is hard to overstate the empirical relevance of constructing efficiency rankings. They are used by managers to allocate resources within organizations and consequently establish the natural boundaries of a firm or by policy makers to determine the most efficient allocation of public resources in education, health care, pollution abatement, etc. Fried (2008) provides a comprehensive survey of the empirical/applied use of such rankings.

There exists two main statistical approaches for modeling production frontiers. The deterministic approach is based on the assumption that all observed data lie in Ψ, i.e., $P((X_i, Y_i) \in \Psi) = 1$ for all i, where P is a probability measure. In these models, any deviation of realized output Y_i from $\rho(X_i)$ is attributable to unobserved inefficiencies of the production plan i. The stochastic approach allows for random shocks to the production process. As a result, observed output Y_i at any input level can be smaller or larger than $\rho(X_i)$. As a result, it may be that $P((X_i, Y_i) \notin \Psi) > 0$ for some i. Although more appealing from an econometric perspective, separating inefficiency and random shock in stochastic frontier models requires strong parametric assumptions on the joint density of (X_i, Y_i) (Aigner et al. (1977), Fan et al. (1996), Kumbhakar et al. (2007), Martins-Filho and Yao (2014)). In contrast, deterministic frontier models can be estimated under much milder restrictions on the stochastic process generating χ_n.

Estimation and inference for deterministic frontier models has been largely

[1] See Simar and Wilson (2008) for a recent comprehensive review of the latest developments.

conducted using DEA (data envelopment analysis) and FDH (free disposal hull) estimators (Charnes et al. (1978), Deprins et al. (1984)). Although the asymptotic properties of DEA (Kneip et al. (2008)) and FDH (Park et al. (2000)) are now well known, these estimators are not robust to extreme values, are inherently biased downward and generate estimated frontiers that are either non-smooth or discontinuous. To remedy these problems a number of alternative nonparametric frontier specification and estimation procedures have been proposed (Aragon et al. (2005), Martins-Filho and Yao (2007, 2008), Daouia and Simar (2007), Daouia et al. (2009, 2010, 2012)). In this paper we add to this literature by considering a novel estimator for the multiplicative nonparametric frontier model first proposed in Martins-Filho and Yao (2007). We assume that output Y_i is generated by

$$Y_i = \frac{\sigma(X_i)}{\sigma_R} R_i \text{ for } i = 1, 2, \cdots, n \tag{1}$$

where R_i is an unobserved random variable representing efficiency and taking values in the interval $[0, 1]$, X_i is an observed random vector representing inputs taking values in $\Re^p_+$, $\sigma(x) : \Re^p_+ \to (0, \infty)$ is a measurable function, σ_R is an unknown parameter and the production frontier is given by $\rho(x) \equiv \frac{\sigma(x)}{\sigma_R}$. In this model R_i has the effect of contracting output from optimal levels that lie on the production frontier. The larger R_i the more efficient the production unit because the closer the realized output is to that on the production frontier. We assume that $E(R_i|X_i = x) \equiv \mu_R$ where $0 < \mu_R < 1$ and $V(R_i|X_i = x) \equiv \sigma_R^2$. Here, the parameter μ_R is interpreted as a mean efficiency given input usage and the common technology Ψ, whereas σ_R is a scale parameter for the conditional distribution of R_i that also locates the production frontier. Its shape is captured by $\sigma(x)$. These conditional moment restrictions together with equation (1) imply that $E(Y_i|X_i = x) = \frac{\mu_R}{\sigma_R}\sigma(x)$ and $V(Y_i|X_i = x) = \sigma^2(x)$. The model can therefore be rewritten as,

$$Y_i = b\sigma(X_i) + \sigma(X_i)\frac{(R_i - \mu_R)}{\sigma_R} = m(X_i) + \sigma(X_i)\epsilon_i \tag{2}$$

where $b = \frac{\mu_R}{\sigma_R}$, $\epsilon_i = \frac{R_i - \mu_R}{\sigma_R}$, $m(X_i) = b\sigma(X_i)$, $E(\epsilon_i|X_i = x) = 0$ and $V(\epsilon_i|X_i = x) = 1$.[2]

Given the location-scale nature of (2), we follow Fan and Yao (1998) and propose an estimation procedure that consists of three stages: first, $m(x)$ is estimated using the local linear estimator of Fan (1992); second, squared residual from the first stage are used in a local exponential procedure to estimate the conditional variance $\sigma^2(x)$ as in Ziegelmann (2002); third, the estimated conditional variance from stage 2 is used to estimate σ_R based on an anchoring assumption to be

[2]For simplicity, we will henceforth write $E(\cdot|X_i = x)$ or $V(\cdot|X_i = x)$ simply as $E(\cdot|X_i)$ or $V(\cdot|X_i)$.

discussed in Section 2. The estimator is fairly easy to implement as it involves standard nonparametric procedures. In addition, the frontier estimator has a number of desirable characteristics: first, contrary to the frontier estimators in Aragon et al. (2005), Daouia et al. (2009) and Martins-Filho and Yao (2008), it is a smooth function of input; second, although the frontier estimator envelops the data, it is not intrinsically biased as the popular DEA (data envelopment analysis) and FDH (free disposal hull) estimators, therefore no bias correction is needed; third, the estimator is fairly robust to outliers and extreme values. In addition, our estimation procedure leads to a frontier estimator that is consistent and asymptotically normal when suitably centered and normalized. Lastly, our estimation procedure improves on the estimator developed in Martins-Filho and Yao (2007) in that our procedure assures that the estimated conditional variance function (and estimated frontier) is always positive. Potential negativity of the estimated variance may be a major impediment in empirical studies that use Martins-Filho and Yao (2007). Our proposed estimator is also shown to have desirable small sample properties as revealed by a Monte Carlo study which provides both evidence on the estimator's finite sample behavior and its performance relative to the estimator proposed in Martins-Filho and Yao (2007).

Besides this introduction, our paper has five more sections. Section 2 presents the deterministic frontier model under consideration, lists the assumptions on the data generating process and gives a detailed description of the estimator. Section 3 provides the main theorems which characterize the asymptotic behavior of the estimator. Section 4 contains a Monte Carlo simulation and in Section 5 we apply our methodology to construct an efficiency ranking for branches of financial institutions in the United States. Finally, Section 6 provides a summary and conclusions.

2. Statistical Model and Estimation Procedure

In this section we provide a full specification of the statistical model under consideration and give a detailed description of the estimation procedure. We start by listing a set of assumptions that are sufficient to establish the main asymptotic results in Section 3.

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

Assumption A2. 1. $Y_i = \sigma(X_i)\frac{R_i}{\sigma_R}$. 2. $R_i \in [0, 1]$, $X_i \in S_X$. 3. $E(R_i|X_i) = \mu_R$, $V(R_i|X_i) = \sigma_R^2$. 4. $\sigma^2(x) = \exp(f(x))$ where $f(x)$ is everywhere differentiable with derivatives of order $d = 1, 2$ denoted by $f^{(d)}(x)$. 5. $\sigma(x) \leq \bar{B}_\sigma < \infty$ for all $x \in S_X$. 6. We denote the first and second derivatives of $\sigma^2(\cdot) : S_X \to \Re$ by $\sigma^{2(1)}(x)$ and $\sigma^{2(2)}(x)$ and assume that $|\sigma^{2(2)}(x)| < \bar{B}_{2\sigma}$ for all $x \in S_X$.

Assumptions A1 and A2 imply that $\{(Y_i, X_i)\}_{i=1}^n$ is an iid sequence of random vectors, which is a typical assumption in the deterministic frontier literature. Contrary to Park et al. (2000) and Kneip et al. (2008) we do not need to assume that the joint density of (Y_i, X_i) is positive at the frontier, which can be too restrictive in some settings. In contrast, for asymptotic normality we require, as in Martins-Filho and Yao (2007), that $max_{1 \le i \le n} R_i$ approaches 1 at a suitable rate when $n \to \infty$ (see Theorem 3 below). Assumption A2.4 assures that for any unknown and arbitrary $f(x)$, we have $\sigma(x) > 0$.

The following assumption is standard in nonparametric estimation and involves only the kernel K. We observe that A3 is satisfied by commonly used kernels such as the Epanechnikov, Biweight and others. Assumption A4 is a Lipschitz condition on the marginal density of X which can be relaxed (Mynbaev and Martins-Filho (2010)) at the expense of greater mathematical complexity.

Assumption A3. $K(x) : \times_{i=1}^p [-1, 1] \to \Re$ is a symmetric density function with bounded support satisfying: 1. $\int x_i K(x) dx = 0$ for $i = 1, \cdots, p$. 2. $\int x_i x_j K(x) dx = \sigma_K^2$ for $i = j$, and 0 for $i \neq j$ and $i, j = 1, \cdots, p$. 3. for all $x \in \Re^p$, $|K(x)| < B_K < \infty$. 4. for all $x, x' \in \Re^p$, $|K(x) - K(x')| < m||x - x'||$ 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 $0 < m_g < \infty$.

We propose the following three stage estimation procedure. First, for any $x \in \Re_+^p$ we obtain $\hat{m}(x; h_n) \equiv \hat{\alpha}$ where

$$(\hat{\alpha}, \hat{\beta}) = \text{argmin}_{\alpha,\beta} \sum_{i=1}^n (Y_i - \alpha - \beta(X_i - x))^2 K\left(\frac{X_i - x}{h_n}\right).$$

The bandwidth h_n satisfies $0 < h_n \to 0$ as $n \to \infty$. This is the local linear kernel estimator of Stone (1977) and Fan (1992) with regressand Y_i and regressors X_i. In the second stage, we follow Ziegelmann (2002) by defining $e_i \equiv (Y_i - \hat{m}(X_i; h_n))^2$ and obtain $\hat{\sigma}_e^2(x; h_n) \equiv \exp(\hat{\theta}_1)$, where

$$(\hat{\theta}_1, \hat{\theta}_2) = \text{argmin}_{\theta_1,\theta_2} \sum_{i=1}^n (e_i - \exp(\theta_1 + \theta_2(X_i - x)))^2 K\left(\frac{X_i - x}{h_n}\right).$$

This provides an estimator $\hat{\sigma}(x; h_n) = \left(\hat{\sigma}_e^2(x; h_n)\right)^{1/2}$. In the third stage, an estimator for σ_R is obtained by defining

$$s_R(h_n) = \left(\max_{1 \le i \le n} \frac{Y_i}{\hat{\sigma}(X_i; h_n)}\right)^{-1}.$$

As observed in Martins-Filho and Yao (2007) the estimation of σ_R by s_R is justified by assuming that there exists *one* observed production unit whose production plan lies on the estimated frontier. This is the anchoring assumption we referred to in the introduction. As a consequence the forecasted value for R_i associated with this unit is identically one. We emphasize that the estimator s_R depends on the bandwidth h_n through $\hat{\sigma}(X_i; h_n)$. 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 $x \in \Re^p$ 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_i, X_i) lies above $(\hat{\rho}(X_i; h_n, g_n), X_i))$ but may lie above or below the true frontier $\rho(X_i)$, therefore avoiding the inherent bias of DEA and FDH estimators.

3. Asymptotic Characterization of the Estimator

Due to the similarity between our proposed estimation strategy and that proposed in Martins-Filho and Yao (2007), most of our focus will be on establishing the asymptotic properties of the second stage estimator under exponential smoothing. For simplicity, but without loss of generality, all of our results are for the case where $p = 1$. For the case where $p > 1$, all results hold with appropriate adjustments on the relative speed of n, h_n^p and g_n^p. The proofs for all results are provided in Appendix 2.

We start by noting that $\sigma^{2(1)}(x) = \exp(f(x)) f^{(1)}(x)$ and therefore a local linear approximation for $\sigma^2(X_i)$ is given by $L(X_i - x, \theta(x)) = \exp(\theta_1(x) + \theta_2(x)(X_i - x))$, where $\theta(x) = (f(x), f^{(1)}(x))' = (\theta_1(x), \theta_2(x))'$. It is easily verifiable that $L(0, \theta(x)) = \exp(\theta_1(x))$, $L^{(1)}(0, \theta(x)) = \exp(\theta_1(x))\theta_2(x)$ and $L^{(2)}(X_i - x, \theta(x)) = \theta_2(x)^2 \exp(\theta_1(x) + \theta_2(x)(X_i - x))$. When necessary we will denote by $\theta^0(x) = (\theta_1^0(x), \theta_2^0(x))$ the true values of $f(x)$ and $f^{(1)}(x)$. Since we have defined $e_i = (Y_i - \hat{m}(X_i; h_n))^2$ we write the second stage estimator as

$$(\hat{\theta}_1(x), \hat{\theta}_2(x)) \equiv \underset{\theta_1, \theta_2}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} (e_i - L(X_i - x, \theta))^2 \frac{1}{h_n} K\left(\frac{X_i - x}{h_n}\right).$$

Furthermore,

$$\begin{aligned}(\hat{\theta}_1(x), \hat{\theta}_2(x)) = \underset{\theta_1, \theta_2}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \Bigg(& e_i - \exp(\theta_1) \\ & - \theta_2 \exp(\theta_1)(X_i - x) - \frac{1}{2} L^{(2)}(\lambda_i(X_i - x), \theta)(X_i - x)^2 \Bigg)^2 \\ & \times \frac{1}{h_n} K\left(\frac{X_i - x}{h_n}\right)\end{aligned}$$

where $\lambda_i \in [0,1]$. Now, suppose $\hat{\theta}_1(x)$, and $\hat{\theta}_2(x)$ are uniformly consistent estimators of $\theta_1^0(x)$ and $\theta_2^0(x)$ in a compact set G and put $\hat{\varepsilon}_i = \frac{1}{2}L^{(2)}(\lambda_i(X_i - x), \hat{\theta}(x))(X_i - x)^2$. We will first provide the asymptotic properties of the estimator $\gamma_1^*(x)$ defined by

$$(\gamma_1^*(x;h_n), \gamma_2^*(x;h_n)) \equiv \underset{\gamma_1,\gamma_2}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}(e_i - \hat{\varepsilon}_i - \gamma_1 - \gamma_2(X_i - x))^2 \frac{1}{h_n} K\left(\frac{X_i - x}{h_n}\right), \tag{3}$$

where $\gamma_1 = \exp(\theta_1)$ and $\gamma_2 = \theta_2 \exp(\theta_1)$. To this end we first obtain the following auxiliary lemma.

Lemma 3.1 *Assume A1-A4. If* $h_n \to 0$, $\frac{nh_n^3}{\ln(n)} \to \infty$, *then for every* $x \in G$ *a compact subset of* $(0,\infty) \times [0,1]$ *we have*

$$\begin{aligned} \gamma_1^*(x;h_n) \quad &- \quad \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} \\ & \qquad K\left(\frac{X_i - x}{h_n}\right)\left(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)\right) \\ &= \quad O_p(R_{n,1}(x)) \end{aligned}$$

uniformly in G, with

$$\begin{aligned} R_{n,1}(x) = \frac{1}{n}\Bigg\{&\left|\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)\right)\right| \\ &+ \left|\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)\left(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)\right)\right|\Bigg\}. \end{aligned}$$

Lemma 1 reveals that to ascertain the uniform order in probability of

$$\gamma_1^*(x;h_n) - \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x))$$

in a compact set G, it suffices to investigate the order of the absolute value of the terms

$$c_1(x) = \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x))$$

and

$$c_2(x) = \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)).$$

However, given assumption A3 of compact support for the kernel K, it suffices to investigate the order of $|c_1(x)|$.[3] In Theorem 1 we provide the exact order of

$$\gamma_1^*(x;h_n) - \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x))$$

and establish that under suitable normalization and centering $\gamma_1^*(x;h_n)$ is asymptotically normally distributed.

Theorem 3.2 *Suppose that assumptions A1-A4 hold. In addition assume that $E(|\epsilon_i||X_i) = \mu_1(X_i)$ is a uniformly bounded function of $X_i \in G$, 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) $\sup_{x\in G}|\gamma_1^*(x;h_n) - \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)$ $(X_i - x))| = O_p(h_n^3) + O_p((\frac{h_n \ln(n)}{n})^{1/2})$,
b) If, in addition, we assume that $E(\epsilon_i^4|X_i = x) = \mu_4(x)$ is continuous in $(0,\infty)$, $h_n^2 \ln(n) \to 0$ and $nh_n^5 = O(1)$ then for every $x \in G$

$$\sqrt{nh_n}(\gamma_1^*(x) - \sigma^2(x) - B_{1n}) \xrightarrow{d} N\left(0, \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right),$$

where $B_{1n} = \frac{h_n^2\sigma_K^2}{2}(\sigma^{2(2)}(x) - L^{(2)}(0,\theta^0(x))) + o_p(h_n^2)$ with $\theta^0(x) = (f(x), f^{(1)}(x))$ uniquely defined by $\sigma^{2(i)}(x) = L^{(i)}(0,\theta^0), i = 0,1$.

It is a direct consequence of Theorem 1 and the equality

$$\sqrt{nh_n}\left(\sqrt{\gamma_1^*(x)} - \sigma(x) - \frac{1}{2\sigma(x)}B_{1n}(x) + \left(\frac{1}{2\sigma(x)} - \frac{1}{2\sigma_b(x)}\right)B_{1n}(x)\right)$$
$$= \frac{1}{2\sqrt{\sigma_b^2(x)}}\sqrt{nh_n}(\gamma_1^* - \sigma^2(x) - B_{1n})$$

$$\sqrt{nh_n}\left(\sqrt{\gamma_1^*(x)} - \sigma(x) - B_{2n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{4g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right),$$

[3]It should be emphasized that kernels with non-compact support could also be accommodated, provided that their rate of tail decay is sufficiently fast, but this would involve much longer proofs.

where $B_{2n} = \frac{h_n^2 \sigma_K^2}{4\sigma(x)}(\sigma^{2(2)}(x) - L^{(2)}(0, \theta^0)) + o_p(h_n^2)$.

Theorem 1 relies on the uniform consistency of $\hat{\theta}(x)$ as an estimator of $\theta^0(x)$ in the compact set G. The next theorem establishes the desired uniform consistency of $\hat{\theta}(x)$.

Theorem 3.3 *Assume that A1-A4 hold and that for all x fixed in a compact subset G of $(0, \infty)$ we have that $(\hat{\theta}_1(x), \hat{\theta}_2(x))$ satisfy the following estimating equations*

$$\begin{pmatrix} \frac{1}{nh_n}\sum_{i=1}^n (e_i - L(X_i - x, \hat{\theta}(x)))L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right) \\ \frac{1}{nh_n}\sum_{i=1}^n (e_i - L(X_i - x, \hat{\theta}(x)))L(X_i - x, \hat{\theta}(x))(X_i - x)K\left(\frac{X_i - x}{h_n}\right) \end{pmatrix} = 0.$$

Furthermore, assume that for any fixed x, $\theta^0(x)$ is in the interior of a compact set $\bar{\Theta} \subset \Re^2$. Then, $\hat{\theta}(x) - \theta^0(x) = o_p(1)$ uniformly on a compact set G of $(0, \infty)$, where $\hat{\theta}(x) = (\hat{\theta}_1(x), \hat{\theta}_2(x))'$.

It is a direct consequence of Theorem 2 and the second part of the proof of Theorem 1 in Hall et al. (1999) that $\exp(\hat{\theta}_1(x)) - \gamma_1^*(x) = o_p(h_n^2)$. Combined with Theorem 1 we have,

$$\sqrt{nh_n}\left(\sqrt{\exp(\hat{\theta}_1(x))} - \sigma(x) - B_{2n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{4g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right),$$

where $B_{2n} = \frac{h_n^2 \sigma_K^2}{4\sigma(x)}(\sigma^{2(2)}(x) - L^{(2)}(0, \theta^0)) + o_p(h_n^2)$. The results in Theorems 1 and 2 refer to the estimator $\hat{\sigma}(x; h_n) = \sqrt{\exp(\hat{\theta}_1(x))}$, 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 from Theorem 1 for $\hat{\sigma}(x; h_n)$. The following Theorem 3 is presented without proof, as it can be obtained directly from Martins-Filho and Yao (2007) in combination with Theorems 1 and 2 given above. Part a) of Theorem 3 is 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 $1 - max_{1 \leq t \leq n} R_t = O_p(L_n)$, 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 $1 - max_{1 \leq i \leq n} R_i = O_p(g_n^2)$ we obtain $s_R(g_n) - \sigma_R = O_p(g_n^2)$. It should be noted that the required boundedness in probability of $1 - max_{1 \leq i \leq n} R_i$ 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 3, 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 3.4 *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) $1 - max_{1\leq i\leq n} R_i = O_p(L_n)$. 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 $B_{2n} = O_p(g_n^2)$.

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. 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 $1 - \max_{1\leq i\leq n} R_i = o_p(g_n^2)$. In this case part (b) of Theorem 2 can be extended to give

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

where $B_{3n} = \frac{g_n^2\sigma(x)\sigma_k^2}{4\sigma_R}\sup_{x\in G, R\in[0,1]}\left(\frac{[\sigma^{2(2)}(x) - L^{(2)}(0,\theta^0)]R}{\sigma^2(x)}\right) + o_p(g_n^2)$. 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 $nh_n g_n^4 \to 0$, in which case $\sqrt{nh_n}B_{3n} \to 0$ as $n \to \infty$. If we compare the preceding result to that obtained from Theorem 2 in Martins-Filho and Yao, we can see that the two estimators have exactly the same asymptotic variance (resulting in the same efficiency) but a different bias. The difference is governed by the term $L^{(2)}(0, \theta^0)$. As mentioned in Ziegelmann (2002), since $L^{(2)}(0, \theta^0)$ is a nonnegative quantity, we conclude that the bias of the estimator we propose can be smaller than that of the local linear estimator if $\sigma^{2(2)}(x)$ is nonnegative and greater than $L^{(2)}(0, \theta^0)$.

Our results show that a local exponential estimator can be incorporated into the second stage estimation replacing the local linear estimator without loss of consistency or asymptotic normality, previously established under the assumptions of Martins-Filho and Yao (2007).

4. Monte Carlo Study

In this section we investigate some of the finite sample properties of our estimator, henceforth referred to as NPE, via a Monte Carlo study. For comparison purposes, we also include in the study the local linear frontier estimator proposed in Martins-Filho and Yao (2007), referred to as NP. We note that in Martins-Filho and Yao (2007) an extensive Monte Carlo study was performed comparing their estimator to the bias-corrected FDH estimator. They find that in most experiments considered the NP estimator outperforms bias-corrected FDH in terms of bias, MSE and the various efficiency criterion measures considered in their paper and herein (see discussion below). Given the relative performance of NPE and NP discussed below, we do not report results on the relative performance of the NPE estimator and the bias corrected FDH.

Our simulations are based on model (1), i.e., $Y_i = \frac{\sigma(X_i)R_i}{\sigma_R}$, with $p = 1$. We generate data with the following characteristics. The X_i are pseudorandom variables from a uniform distribution with support given by $[a_l, b_u]$. $R_i = \exp(-Z_i)$, where Z_i are pseudorandom variables from an exponential distribution with parameter $\beta > 0$, therefore R_i has support on $(0, 1]$. We consider three specifications for $\sigma(x)$:

$$\sigma_1(x) = \sqrt{x}, \text{ with } x \in [a_l, b_u] = [10, 100];$$
$$\sigma_2(x) = 3(x - 1.5)^3 + 0.25x + 1.125, \text{ with } x \in [a_l, b_u] = [1, 2] \text{ and}$$
$$\sigma_3(x) = x^2 \text{ with } x \in [a_l, b_u] = [1, 2].$$

These functions are associated with concave, non-concave nor convex and convex production frontiers, respectively. Two parameters for the exponential distribution are considered: $\beta_1 = 3$ and $\beta_2 = 1/3$. These choices of parameters produce, respectively, the following values for the parameters of $g_{R|X} : (\mu_R, \sigma_R^2) = (0.25, 0.08)$ and $(0.75, 0.04)$. Three sample sizes $n = 200, 400, 600$ were used.

An important aspect in the implementation of our frontier estimator is bandwidth selection. We consider the following rule-of-thumb bandwidth.

$$\hat{h}_{ROT} = \left(\frac{\int K^2(\phi)d\phi(\hat{\mu}_4(\lambda_n) - 1) \int \dot{\sigma}^2(x)dx}{(\sigma_K^2)^2 \left(\max_{1 \le i \le n} \left(\frac{(\dot{\sigma}^{2(2)}(x_i) - \dot{\beta}^2 e^{\dot{\alpha}})\dot{R}_i}{\dot{\sigma}^2(x_i)} \right) \right)^2 \frac{1}{n} \sum_{i=1}^n \dot{\sigma}^2(x_i)} \right)^{1/5} n^{-1/5}$$

The sequence $\{\dot{\sigma}^2(X_i)\}_{i=1}^n$ is estimated with an ordinary least square quartic regression of $\{\hat{\epsilon}_i^2\}_{i=1}^n$ on $\{X_i\}_{i=1}^n$, with $\hat{\epsilon}_i = Y_i - \hat{m}(X_i)$, where $\hat{m}(X_i)$ is estimated via local linear regression with a rule-of-thumb bandwidth as in Ruppert et al. (1995).

$\{\dot{\sigma}^2(X_i)\}_{i=1}^n$ is then used to construct $\int \dot{\sigma}^2(x)dx$, $\max_{1\leq i\leq n}\left(\frac{(\dot{\sigma}^{2(2)}(x_i)-\dot{\beta}^2 e^{\dot{\alpha}})\dot{R}_i}{\dot{\sigma}^2(x_i)}\right)$ and $\frac{1}{n}\sum_{i=1}^n \dot{\sigma}^2(x_i)$. In particular $\dot{\beta}^2 e^{\dot{\alpha}}$ is estimated by $(\dot{\sigma}^{2(1)}(x))^2/\dot{\sigma}^2(x)$. $\hat{\mu}_4(\lambda_n) = \frac{1}{n}\sum_{i=1}^n\left(\frac{Y_i}{\hat{\sigma}(X_i,\lambda_n)} - \hat{b}\right)^4$, where $\hat{b} = \frac{\sum_{i=1}^n \hat{\sigma}(X_i,\lambda_n)Y_i}{\sum_{i=1}^n \hat{\sigma}(X_i,\lambda_n)}$ is an estimator for $b = \mu_R/\sigma_R$. $\{\hat{\sigma}^2(X_i,\lambda_n)\}_{i=1}^n$ in $\hat{\mu}_4$ is estimated via local linear regression of $\{\hat{\epsilon}_i^2\}_{i=1}^n$ on $\{X_i\}_{i=1}^n$, with a rule-of-thumb bandwidth λ_n as in Ruppert et al. (1995) and Fan and Yao (1998).

The results of our simulations are summarized in Figures 2-19 that appear in Appendix 1. Whenever negative estimates for $\sigma^2(\cdot)$ occur in the case of NP, the sample is discarded. In this case, another sample is generated until 1000 *valid* repetitions are obtained. Figures 2-19 give boxplots of MSE for the frontier function estimator $\hat{\rho}(\cdot)$, shape function estimator $\hat{\sigma}(\cdot)$, location parameter estimator s_R, and efficiency estimator $\hat{R}_i$. Each boxplot is constructed from 1000 points (repetitions), where each point corresponds to a sample draw and is calculated as the squared Euclidean distance between the estimate and true value of $\rho(\cdot)$, $\sigma(\cdot)$, σ_R and R_i. The thick horizontal line inside the rectangle in each boxplot corresponds to the median of the distribution, and the rectangle height corresponds to interquartile range. Consequently 50% of data is represented by the rectangle. The two thin horizontal lines below and above the rectangle are the whiskers. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range.

General regularities: As expected from the asymptotic results of Section 3, as the sample size n increases, the boxplots show that MSE decreases for the totality of simulations for all estimators and values for μ_R considered.

We now turn to the impact of different values of μ_R on the performance of NPE and NP. Regarding s_R, $\hat{\sigma}(x)$, the frontier estimator and the efficiency, the best performance in terms of MSE occurs when $\mu_R = 0.25$. The relative diminished performance when $\mu_R = 0.75$ is most likely explained by the fact that for this DGP σ_R^2 is half of its value in other DGP, contributing to its higher variance as suggested in Theorem 2.

Remark 1. It is worth noting that most of the frontier estimators available in literature present better performance as the concentration of firms near to the frontier increases. Since NP and NPE estimators are based on conditional variance, their performance are disregard of that concentration. Therefore they may be valuable alternatives to estimating production frontiers in situations such that the majority of firms are not close to the frontier.

Relative performance of estimators: For $\beta_1 = 3$ there are no great differences between NP and NPE. Some exceptions occur when we use the DGP with $\sigma_1(x)$. In this case, NP performs better than NPE with some exception regarding dispersion of the MSE in frontier function estimator. See Figures 2-10. The main differences in performance occur when we use $\beta_2 = 1/3$. In this case, on estimating the production frontier (Figures 11-19) there seems to be evidence that NPE dominates NP in terms of MSE in all cases considered.The gain of NPE seems to be on estimating σ_R, since that NPE outperforms NP on estimating σ_R, while NP does a slightly better job than NPE on estimating frontier shape function.

Remark 2. NP and NPE performances are quite similar in the most favorable case for both of them, that is, with $\beta_1 = 3$. Nevertheless, NPE performs better in the hardest design. Therefore it seems to be a valuable tool in estimating the frontier if compared with the NP estimator. Furthermore, in the next section we shall see that preventing negative estimates for the variance may be important in empirical work.

5. Empirical exercise – The Case of a Production Frontier for Bank Branches in the United States

We illustrate the use of our methodology by analyzing United States (US) bank data. The goal is to estimate a production frontier for bank branches in the US territory using cross sectional data from 2009. The data source is the Federal Deposit Insurance Corporation (FDIC) and all data are publicly available from FDIC's website (http://www2.fdic.gov/SDI/main.asp).

We only consider one input and one output. If multiple inputs are considered, one can avoid slow convergence rates due to increased number of regressors (curse of dimensionality) by reducing the number of conditioning variables via principal components analysis, for instance. In order to measure branch output we use *net loans and leases (nll)*, whereas to measure branch inputs we consider *total deposits (td)*. We restrict our sample to branches working with *total deposits* between US$ 10,000.00 and US$ 1,000,000.00 corresponding to *net loans and leases* between US$ 10.00 and US$ 1,000,000.00. Moreover, we project the inputs into the interval [0,1] to facilitate the bandwidth choice. For sake of comparison with our proposed estimator, we include in our analysis both the FDH estimator and the nonparametric linear estimator of Martins-Filho and Yao (2007).

Figure 1 shows the estimated frontiers for US territory and Table 1 presents the efficiency rank for the 30 most efficient branches. The smoothness level for NP is the highest we can get without obtaining negative scale estimates using a non variable bandwidth. Therefore, the NP estimator fails in providing nonnegative estimates for the conditional variance using larger bandwidths. Such a result advocates in favour of the NPE estimator, which possesses the natural nonnegativity property.

Our analysis here amounts to descriptive comments of the empirical results and should obviously be complemented by a more in depth knowledege of the banking industry and state regulatory environment. Our goal was simply to illustrate the use of our estimation method and how it can be useful in analyzing the efficiency of a particular industry.

6. Summary and Conclusions

In this paper we use the idea of local exponential smoothing to improve the nonparametric frontier estimator proposed by Martins-Filho and Yao (2007). Their estimation strategy suffered from the undesirable property of potentially generating negative estimated conditional variances. Local exponential smoothing prevents this problem. In addition, there seems to be finite sample gains in adopting exponential smoothing. These gains are particularly large in the estimation of the location parameter in the frontier model. Our simulation results confirm and give added support to those in Ziegelmann (2002). We also illustrate our approach via an empirical data frontier analysis, offering the practitioner an applied viewpoint.

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.

Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, 21:358–389.

Charnes, A., Cooper, W., & Rohdes, E. (1978). Measuring the efficiency of decision making units. *Europena Journal of Operational Research*, 2:429–444.

Daouia, A., Flores, J.-P., & Simar, L. (2010). Frontier estimation and extreme value theory. *Bernoulli*, 16:1039–1063.

Daouia, A., Flores, J.-P., & Simar, L. (2012). Regularization of nonparametric frontier estimators. *Journal of Econometrics*, 168:285–299.

Daouia, A., Gardes, L., & Girard, S. (2009). Large sample approximation of the distribution for smoothed monotone frontier estimators. Technical report, INRIA.

Daouia, A. & Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140:375–400.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19:273–292.

Deprins, D., Simar, L., & Tulkens, H. (1984). *Measuring labor inefficiencies in post offices*. North Holland, Amsterdam.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Satistical Association*, 87:998–1004.

Fan, J. & Yao, Q. (1998). Efficient Estimation of Conditional Variance Functions in Stochastic regression. *Biometrika*, 85:645–660.

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

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). *The measurement of efficiency of production*. Kluwer-Nijhoff. Boston.

Färe, R., Grosskopf, S., & Margaritis, D. (2008). Productivity and efficiency: Malmquist and more. In Fried, H., Lovell, C. A. K., & Schmidt, S. S., editors, *The measurement of productive efficiency*, chapter 5, pages 522–621. Oxford University Press, 2nd edition.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society A*, 120:253–281.

Fried, H. (2008). Efficiency and productivity. In Fried, H., Lovell, C. A. K., & Schmidt, S. S., editors, *The Measurement of Productive Efficiency and Productivity Growth*, chapter 1, pages 3–91. Oxford University Press, 2nd edition.

Hall, P., Wolff, R., & Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94:154–163.

Kneip, A., Simar, L., & Wilson, P. (2008). Asymptotics and consistent Bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24:1663–1697.

Koopmans, T. (1951). An analysis of production as an efficient combination of activities. In Koopmans, T. C., editor, *Activity Analysis of Production and Allocation*, volume 13 of *Cowles Commission for Research in Economics*. Joh Wiley and Sons, New York, NY.

Kumbhakar, S., Park, B. U., Simar, L., & Tsionas, E. (2007). Nonparametric stochastic frontiers: a local maximum likelihood approach. *Journal of Econometrics*, 137:1–27.

Martins-Filho, C. & Yao, F. (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics*, 141:283–319.

Martins-Filho, C. & Yao, F. (2008). A smoothed conditional quantile frontier estimator. *Journal of Econometrics*, 143:317–333.

Martins-Filho, C. & Yao, F. (2014). Nonparametric stochastic frontier estimation via profile likelihood. *Econometric Reviews*, forthcoming.

Mynbaev, K. & Martins-Filho, C. (2010). Bias reduction in kernel density estimation via Lipschitz condition. *Journal of Nonparametric Statistics*, 22:219–235.

Park, B. U., Simar, L., & Weiner, C. (2000). The FDH estimator for productivity efficiency scores: asymptotic properties. *Econometric Theory*, 16:855–877.

Ruppert, D., Sheather, S., & Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90:1257–1270.

Simar, L. & Wilson, P. (2008). Statistical inference in nonparametric frontier models: recent developments and perspectives. In Fried, H., Lovell, C. A. K., & Schmidt, S. S., editors, *The Measurement of Productive Efficiency*, chapter Chapter 4. Oxford University Press, Oxford, UK.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, 5:595–620.

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge, UK.

Ziegelmann, F. (2002). Nonparametric estimation of volatility functions: the local exponential approach. *Econometric Theory*, 18:985–991.

Appendix 1: Tables and Graphics

Table 1
NPE Efficiency Rank - US Territory - inputs in $[10,000, 1,000,000]$

	Agency	City	State	$\hat{R}$
1	Wright Express Financial Services Corporation	Salt Lake City	UT	1.00
2	Glacier Bank	Kalispell	MT	0.98
3	Northern Bank & Trust Company	Woburn	MA	0.96
4	Access National Bank	Reston	VA	0.96
5	The Needham Bank	Needham	MA	0.93
6	Atlantic Capital Bank	Atlanta	GA	0.93
7	Parke Bank	Sewell	NJ	0.93
8	Quad City Bank and Trust Company	Bettendorf	IA	0.93
9	Two Rivers Bank & Trust	Burlington	IA	0.92
10	Community West Bank, National Association	Goleta	CA	0.92
11	Southern First Bank, National Association	Greenville	SC	0.92
12	Bank of Washington	Washington	MO	0.92
13	First Bank Richmond, National Association	Richmond	IN	0.91
14	Jefferson Bank and Trust Company	Eureka	MO	0.91
15	First Financial Bank	El Dorado	AR	0.90
16	The Park Bank	Madison	WI	0.90
17	The Foster Bank	Chicago	IL	0.90
18	Integrity Bank	Camp Hill	PA	0.90
19	Wainwright Bank & Trust Company	Boston	MA	0.89
20	Mountain West Bank	Coeur D Alene	ID	0.89
21	Monarch Bank	Chesapeake	VA	0.89
22	Horicon Bank	Horicon	WI	0.88
23	Metropolitan National Bank	New York	NY	0.88
24	Republic Bank	Bountiful	UT	0.88
25	Adams Bank & Trust	Ogallala	NE	0.87
26	Farmers & Merchants Bank	Timberville	VA	0.87
27	Centennial Bank	Fountain Valley	CA	0.87
28	Citizens National Bank of Texas	Waxahachie	TX	0.87
29	Kansas State Bank of Manhattan	Manhattan	KS	0.87
30	Medallion Bank	Salt Lake City	UT	0.86

Figure 1
Frontier Estimation - US Territory - $x \in [10,000, 1,000,000]$

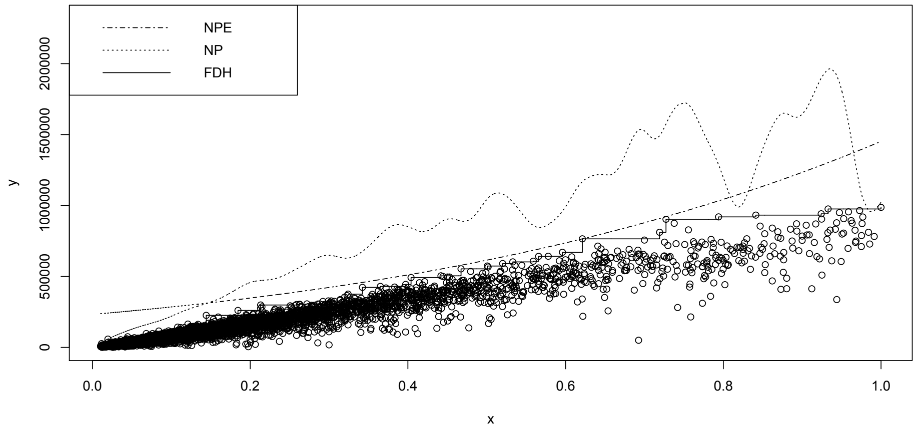


Figure 2
Frontier I - Boxplot of Estimators - $n = 200$ - $\mu_R = 0.25$

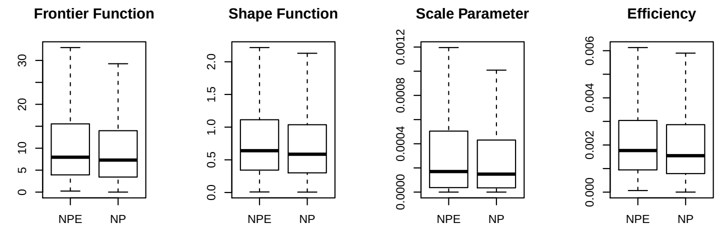


Figure 3
Frontier I - Boxplot of Estimators - $n = 400$ - $\mu_R = 0.25$

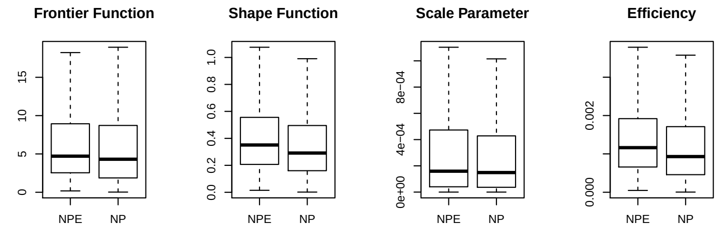


Figure 4
Frontier I - Boxplot of Estimators - $n = 600$ - $\mu_R = 0.25$

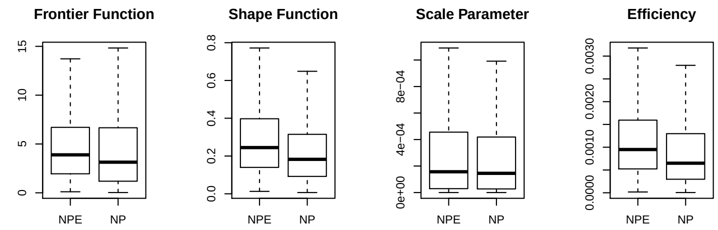


Figure 5
Frontier II - Boxplot of Estimators - $n = 200$ - $\mu_R = 0.25$

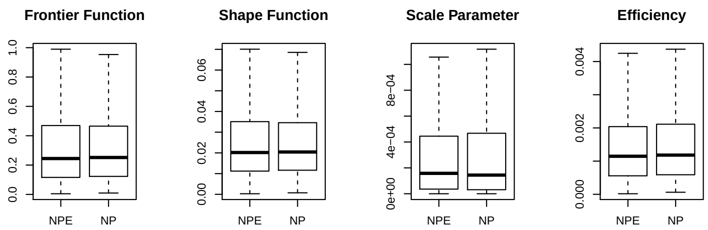


Figure 6
Frontier II - Boxplot of Estimators - $n = 400$ - $\mu_R = 0.25$

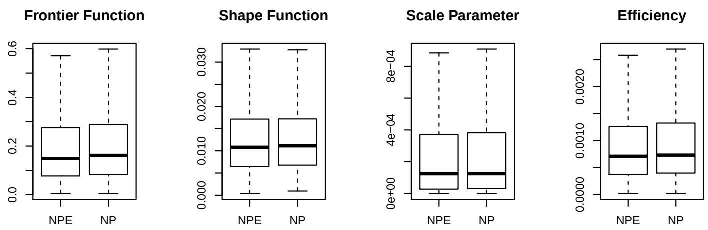


Figure 7
Frontier II - Boxplot of Estimators - $n = 600$ - $\mu_R = 0.25$

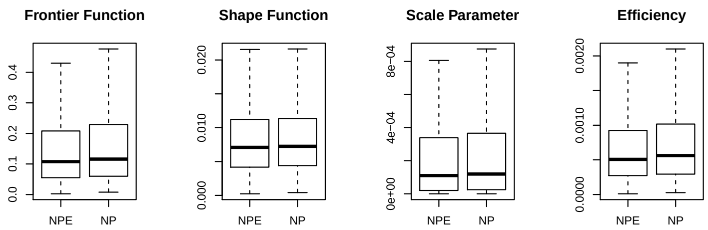


Figure 8
Frontier III - Boxplot of Estimators - $n = 200$ - $\mu_R = 0.25$

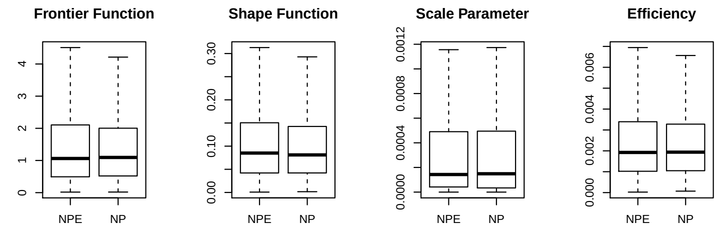


Figure 9
Frontier III - Boxplot of Estimators - $n = 400$ - $\mu_R = 0.25$

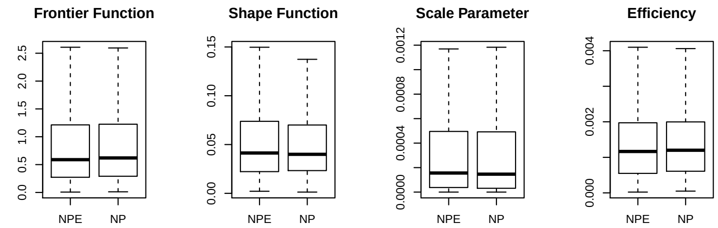


Figure 10
Frontier III - Boxplot of Estimators - $n = 600$ - $\mu_R = 0.25$

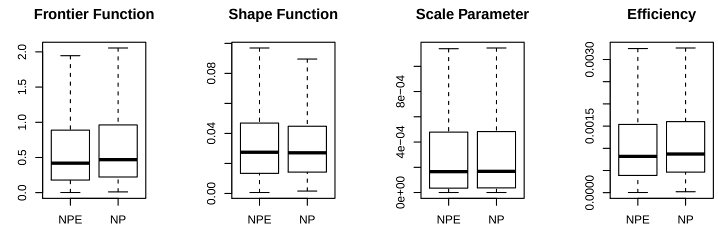


Figure 11
Frontier I - Boxplot of Estimators - $n = 200$ - $\mu_R = 0.75$

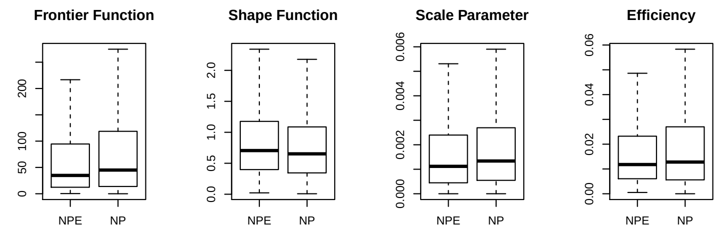


Figure 12
Frontier I - Boxplot of Estimators - $n = 400$ - $\mu_R = 0.75$

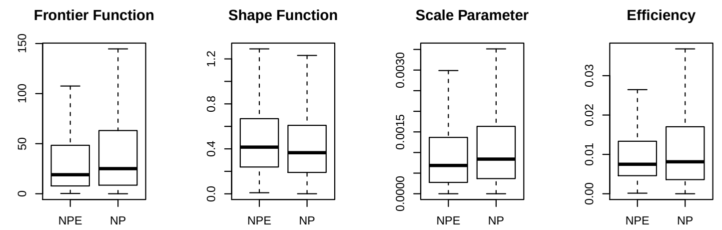


Figure 13
Frontier I - Boxplot of Estimators - $n = 600$ - $\mu_R = 0.75$

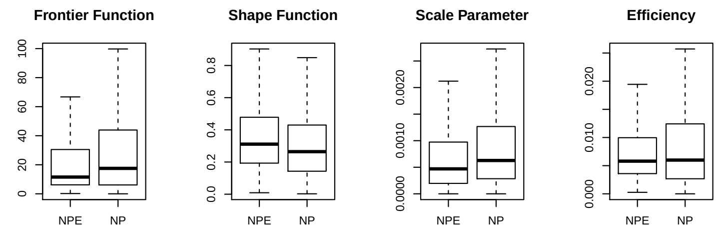


Figure 14
Frontier II - Boxplot of Estimators - $n = 200$ - $\mu_R = 0.75$

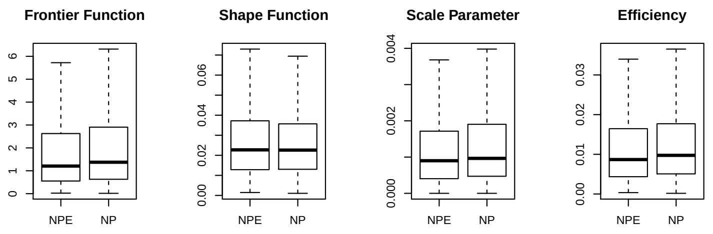


Figure 15
Frontier II - Boxplot of Estimators - $n = 400$ - $\mu_R = 0.75$

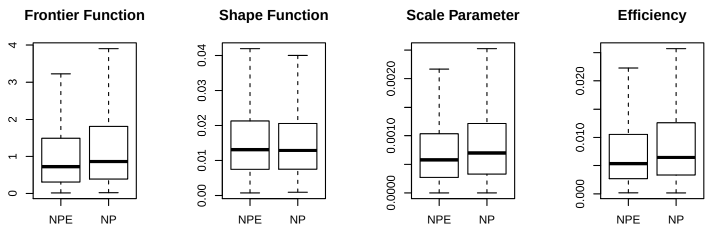


Figure 16
Frontier II - Boxplot of Estimators - $n = 600$ - $\mu_R = 0.75$

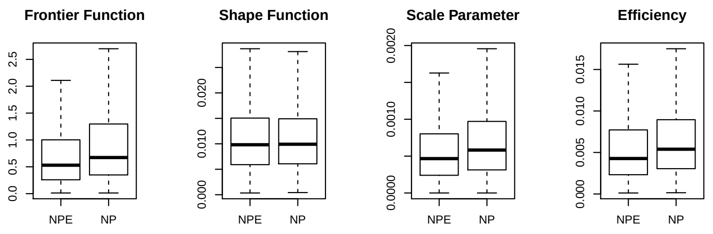


Figure 17
Frontier III - Boxplot of Estimators - $n = 200$ - $\mu_R = 0.75$

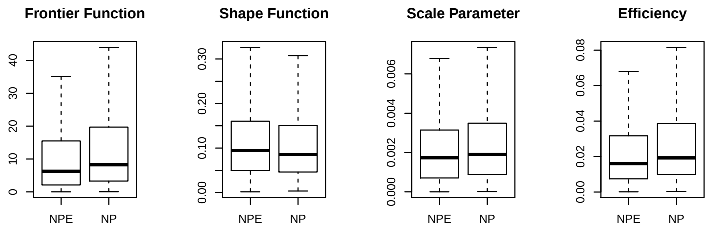


Figure 18
Frontier III - Boxplot of Estimators - $n = 400$ - $\mu_R = 0.75$

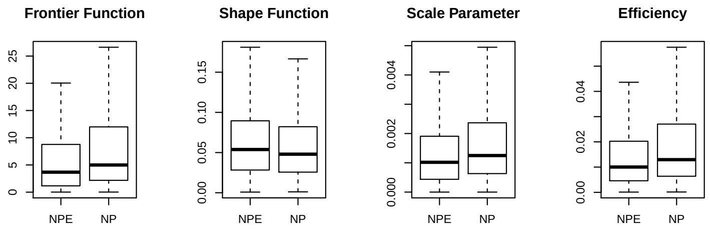
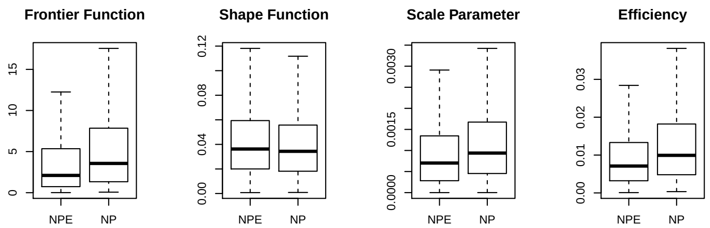


Figure 19
Frontier III - Boxplot of Estimators - $n = 600$ - $\mu_R = 0.75$



Appendix 2: Proofs

Proof of Lemma 1. Given the algebraic structure of the optimand in equation (3) we can write

$$A_n \equiv \gamma_1^*(x) - \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)).$$

Letting $S(x) = \begin{pmatrix} g_X(x) & 0 \\ 0 & g_X(x)\sigma_K^2 \end{pmatrix}$ we have by the Cauchy-Schwarz inequality that

$$|A_n| \leq \frac{1}{h_n}\left((1,0)(S_n^{-1}(x) - S^{-1}(x))^2(1,0)'\right)^{1/2} R_{n,1}(x).$$

From part (b) of Lemma 1 in Martins-Filho and Yao (2007),

$$B_n(x) \equiv h_n^{-1}\left((1,0)(S_n^{-1}(x) - S^{-1}(x))^2(1,0)'\right)^{1/2} = O_p(1)$$

uniformly in G, therefore completing the proof.

Proof of Theorem 1. a) Given the comments following Lemma 1, it suffices to investigate the order of $|c_1(x)|$. After substituting e_i, we write $c_1(x) = I_{1n}(x) + I_{2n}(x) + I_{3n}(x) + I_{4n}(x) - I_{5n}(x)$, where

$$\begin{aligned}
I_{1n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\sigma^2(X_i) - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)\right) \\
I_{2n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)(\epsilon_i^2 - 1)\sigma^2(X_i) \\
I_{3n}(x) &= \frac{2}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\sigma(X_i)\epsilon_i(\hat{m}(X_i; h_n) - m(X_i)) \\
I_{4n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)(\hat{m}(X_i; h_n) - m(X_i))^2 \\
I_{5n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\hat{\varepsilon}_i
\end{aligned}$$

The uniform order in probability of $I_{jn}(x)$ for $j = 2, 3, 4$ on the set G is given in Martins-Filho and Yao (2007) Theorem 1, part (a). Here we study the order of

$I_{1n}(x) - I_{5n}(x)$. Note that

$$\begin{aligned}
I_{1n}(x) - I_{5n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \Bigg(\sigma^2(X_i) \\
&\quad - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x) \\
&\quad - \frac{1}{2} L^{(2)}(\lambda_i(X_i - x), \hat{\theta}(x))(X_i - x)^2\Bigg) \\
&= \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \Bigg(\frac{1}{2}\sigma^{2(2)}(\lambda_i^{'}(X_i - x) + x)(X_i - x)^2 \\
&\quad - \frac{1}{2} L^{(2)}(\lambda_i(X_i - x), \hat{\theta}(x))(X_i - x)^2\Bigg) \\
&= \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) R_i (X_i - x)^2 \\
&= \frac{h_n}{n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \left(\frac{X_i - x}{h_n}\right)^2 R_i.
\end{aligned}$$

where $R_i = \frac{1}{2}\Bigg(\sigma^{2(2)}(\lambda_i^{'}(X_i - x) + x) - L^{(2)}(\lambda_i(X_i - x), \hat{\theta}(x))\Bigg)$ and $\lambda_i' \in [0,1]$. Now, rewriting $R_i = \frac{1}{2}\Bigg(\sigma^{2(2)}(\lambda_i^{'}(X_i - x) + x) - L^{(2)}(0, \theta^0)\Bigg) + \frac{1}{2}\Bigg(L^{(2)}(0, \theta^0) - L^{(2)}(\lambda_i(X_i - x), \hat{\theta}(x))\Bigg)$ we have

$$\begin{aligned}
I_{1n}(x) - I_{5n}(x) &= \frac{h_n}{2n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \left(\frac{X_i - x}{h_n}\right)^2 \\
&\quad \times \Bigg(\sigma^{2(2)}(\lambda_i^{'}(X_i - x) + x) - L^{(2)}(0, \theta^0)\Bigg) \\
&\quad + \frac{h_n}{2n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \left(\frac{X_i - x}{h_n}\right)^2 \\
&\quad \times \Bigg(L^{(2)}(0, \theta^0) - L^{(2)}(\lambda_i(X_i - x), \hat{\theta}(x))\Bigg) \\
&= J_{1n}(x) + J_{2n}(x).
\end{aligned}$$

Since $L^{(2)}(0, \theta) = \exp(\theta_1(x))(\theta_2(x))^2 = \sigma^2(x)(f^{(1)}(x))^2$, we have that $L^{(2)}(0, \theta) < C$ provided $|f^{(1)}(x)| < B_f$ and $0 < \sigma^2(x) < \bar{B}_\sigma^2$. Also, since $|\sigma^{2(2)}(x)| < \bar{B}_{2\sigma}$ for all x by the same argument in (Martins-Filho and Yao, 2007, p. 307), we have that $\sup_{x\in G}|J_{1n}(x)| \leq O_p(h_n^2)$. Now,

$$
\begin{aligned}
J_{2n}(x) = &-\frac{h_n}{2ng_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \Big((\hat{\theta}_2(x)^2\exp(\hat{\theta}_1(x)) \\
&- \theta_2(x)^2\exp(\theta_1(x)))\exp(\hat{\theta}_2(x)(X_i-x)\lambda_i) \\
&+ \theta_2^2(x)\exp(\theta_1(x))(\exp(\hat{\theta}_2(x)(X_i-x)\lambda_i)-1)\Big) \\
= &(\hat{\theta}_2(x)^2\exp(\hat{\theta}_1(x)) - \theta_2(x)^2\exp(\theta_1(x))\frac{-h_n}{2ng_X(x)} \\
&\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \exp(\hat{\theta}_2(x)(X_i-x)\lambda_i) \\
&- \theta_2^2(x)\exp(\theta_1(x))\frac{h_n}{2ng_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \\
&\times (\exp(\hat{\theta}_2(x)(X_i-x)\lambda_i)-1).
\end{aligned}
$$

Note that whenever $\left|\frac{X_i-x}{h_n}\right| > 1$ we have $K\left(\frac{X_i-x}{h_n}\right) = 0$. Hence, consider

$$
M_n(x) = \frac{-h_n}{ng_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \exp(\hat{\theta}_2(x)(X_i-x)\lambda_i).
$$

All terms in the sum are positive, and since the exponential function is everywhere increasing $\exp(\hat{\theta}_2(x)(X_i-x)\lambda_i) \leq \exp(|\hat{\theta}_2(x)|h_n)$ since $\lambda_i \in [0,1]$ and $|X_i-x| \leq h_n$, otherwise $K\left(\frac{X_i-x}{h_n}\right)=0$. Therefore,

$$
\begin{aligned}
|M_n(x)| &\leq \frac{h_n e^{|\hat{\theta}_2(x)|h_n}}{n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \\
&\leq \underline{\mathrm{B}}_{gX}^{-1} e^{|\hat{\theta}_2(x)|h_n} \frac{h_n^2}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \\
&= \underline{\mathrm{B}}_{gX}^{-1} e^{|\hat{\theta}_2(x)|h_n} h_n^2 \Bigg\{ \frac{1}{nh_n} \sum_{i=1}^{n} \Bigg[K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \\
&\quad - \frac{1}{h_n} E\Bigg(K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \Bigg) \Bigg] \\
&\quad + \frac{1}{h_n} E\Bigg(K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \Bigg) \Bigg\}
\end{aligned}
$$

and consequently

$$
\begin{aligned}
\sup_{x\in G}|M_n(x)| \leq \underline{\mathrm{B}}_{gX}^{-1} \sup_{x\in G} e^{|\hat{\theta}_2(x)|h_n} h_n^2 \Bigg\{ \sup_{x\in G} \Bigg| \frac{1}{nh_n} \sum_{i=1}^{n} \Bigg[K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \\
- \frac{1}{h_n} E\Bigg(K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \Bigg) \Bigg] \Bigg| \\
+ \sup_{x\in G} \frac{1}{h_n} E\Bigg(K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \Bigg) \Bigg\}.
\end{aligned}
$$

From (Martins-Filho and Yao, 2007, p. 306),

$$
\begin{aligned}
\sup_{x\in G} \quad & \Bigg| \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \\
& - \frac{1}{h_n} E\Bigg(K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \Bigg) \Bigg| = o_p(h_n),
\end{aligned}
$$

and $\sup_{x\in G} \frac{1}{h_n} E\Bigg(K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \Bigg) = O(1)$. Furthermore, given that $\hat{\theta}_2(x)$ is an uniformly consistent estimator for $\theta_2(x)$, we have $e^{|\hat{\theta}_2(x)|} = e^{|\hat{\theta}_2(x)-\theta_2(x)+\theta_2(x)|h_n} \leq e^{|\hat{\theta}_2(x)-\theta_2(x)|h_n + |\theta_2(x)|h_n} \xrightarrow{p} 1$ uniformly in G. Hence, $\sup_{x\in G}|M_n(x)| \leq \underline{\mathrm{B}}_{gX}^{-1} h_n^2 (h_n o_p(1) + O(1)) = \underline{\mathrm{B}}_{gX}^{-1} h_n^3 o_p(1) + \underline{\mathrm{B}}_{gX}^{-1} O(h_n^2) = O_p(h_n^2)$. Since $(\hat{\theta}(x) - \theta(x)) = o_p(1)$ uniformly in G, we have by Slutsky Theorem that $\hat{\theta}_2(x)^2 e^{\hat{\theta}_1(x)} - \theta_2^2(x) e^{\theta_1(x)} = o_p(1)$. Similarly,

$$
-\theta_2^2(x) e^{\theta_1(x)} \frac{h_n}{2n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 (e^{\hat{\theta}_2(x)(X_i-x)\lambda_i} - 1) = o_p(h_n^2).
$$

Hence, $\sup_{x\in G}|J_{2n}(x)| = o_p(h_n^2)$. In all, $\sup_{x\in G}|I_{1n}(x) - I_{5n}(x)| = O_p(h_n^2)$, and using the results in Martins-Filho and Yao (2007) for $I_{2n}(x)$, $I_{3n}(x)$ and $I_{4n}(x)$ we have

$$sup_{x\in G}\left|\gamma_1^*(x) - \sigma^2(x) - \frac{1}{nh_n g_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\right.$$

$$\left.\left(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)\right)\right| \leq O_p(h_n^3) + O_p\left(\left(\frac{h_n \ln(n)}{n}\right)^{1/2}\right),$$

which completes the proof of part a).
b) A direct consequence of a) is the fact that

$$\begin{aligned}\sqrt{nh_n}(\gamma_1^*(x) - \sigma^2(x)) - \frac{1}{\sqrt{nh_n}g_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \\ \times\left(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)\right) \\ = \sqrt{nh_n^7}O_p(1) + (h_n^2\ln(n))^{1/2}O_p(1).\end{aligned}$$

Hence, provided $h_n^2\ln(n) \to 0$ as $n \to \infty$, the asymptotic distribution of $\sqrt{nh_n}$ $(\gamma_1^*(x) - \sigma^2(x))$ is the same as that of

$$\begin{aligned}\frac{1}{\sqrt{nh_n}g_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) & \left(e_i - \hat{\varepsilon}_i - \sigma^2(x) - \sigma^{2(1)}(x)(X_i - x)\right) \\ = & \sqrt{nh_n}c_1(x),\end{aligned}$$

which can be written as

$$\sqrt{nh_n}c_1(x) = \sqrt{nh_n}(I_{1n}(x) - I_{5n}(x) + I_{2n}(x) + I_{3n}(x) + I_{4n}(x)).$$

From Martins-Filho and Yao (2007) we have that

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

Also, $\sqrt{nh_n}I_{3n}(x) = \sqrt{nh_n}$ $\left(\frac{1}{\sqrt{n}}o_p(1) + h_n^2 o_p(1)\right) = \sqrt{h_n}o_p(1) + \sqrt{nh_n^5}o_p(1)$. Hence, provided that $nh_n^5 = O(1)$, $\sqrt{nh_n}I_{3n}(x) = o_p(1)$. Moreover, $\sqrt{nh_n}I_{4n}(x) = \sqrt{nh_n}(h_n^2 o_p(1)) = \sqrt{nh_n^5}o_p(1) = o_p(1)$. We now focus on

$$\begin{aligned}\sqrt{nh_n}(I_{1n}(x) - I_{5n}(x)) &= \sqrt{nh_n}\frac{1}{nh_n g_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2 R_i \\ &= \sqrt{nh_n}B_n(x).\end{aligned}$$

Note that we can write,

$$\begin{aligned}
\sqrt{nh_n}B_n(x) &= \sqrt{nh_n}\Bigg(\frac{h_n}{n}\frac{1}{2g_X(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2 \\
&\quad (\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x)-L^{(2)}(0,\theta^0))\Bigg) \\
&+ \sqrt{nh_n}\Bigg(\frac{h_n}{n}\frac{1}{2g_X(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2(L^{(2)}(0,\theta^0)) \\
&- L^{(2)}(\lambda_i(X_i-x),\hat{\theta}(x))\Bigg).
\end{aligned}$$

But using the definition of $J_{2n}(x)$ given in part a) as well as its order in probability we have that the last term is $\sqrt{nh_n}J_{2n}(x)=\sqrt{nh_n}(h_n^2 o_p(1))=\sqrt{nh_n^5}o_p(1)=o_p(1)$ provided $nh_n^5=O(1)$. Now,

$$\begin{aligned}
&E\Bigg(\frac{1}{h_n^2}\frac{h_n}{n}\frac{1}{2g_X(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2[\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x) \\
&\quad -L^{(2)}(0,\theta^0)]\Bigg) \\
&= \frac{1}{h_n^2}\Bigg(\frac{h_n}{n}\frac{1}{2g_X(x)}n\int K(\phi)\phi^2[\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x)-L^{(2)}(0,\theta^0)] \\
&\quad g_X(x+h_n\phi)h_n d\phi\Bigg) \\
&= \frac{1}{2g_X(x)}\Bigg(\int K(\phi)\phi^2 h^{(2)}(x+\lambda_i^{'}h_n\phi)g(x+h_n\phi)h_n d\phi \\
&\quad -L^{(2)}(0,\theta^0)\int K(\phi)\phi^2 g_X(x+h_n\phi)d\phi\Bigg)\rightarrow \frac{1}{2}h^{(2)}(x)\sigma_K^2-\frac{1}{2}L^{(2)}(0,\theta^0)\sigma_K^2.
\end{aligned}$$

as $n \to \infty$ by Lebesgue's dominated convergence Theorem. Also,

$$V\Bigg(\frac{1}{h_n^2}\frac{h_n}{n}\frac{1}{2g_X(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2$$
$$\times[\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x)-L^{(2)}(0,\theta^0)]\Bigg)$$
$$=\frac{1}{4g_X^2(x)}\frac{1}{n^2h_n^2}nVK\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2[\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x)$$
$$-L^{(2)}(0,\theta^0)]$$
$$=\frac{1}{4g_X^2(x)}\frac{1}{nh_n^2}\Bigg\{E\Bigg(K^2\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^4[\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x)$$
$$-L^{(2)}(0,\theta^0)]^2\Bigg)$$
$$-\Bigg(E\Bigg(K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)^2[\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x)$$
$$-L^{(2)}(0,\theta^0)]\Bigg)\Bigg)^2\Bigg\}$$
$$=\frac{1}{h_n^2}\Bigg(\frac{h_n}{n}\frac{1}{2g_X(x)}n\int K(\phi)\phi^2[\sigma^{2(2)}(\lambda_i^{'}(X_i-x)+x)-L^{(2)}(0,\theta^0)]$$
$$g_X(x+h_n\phi)h_nd\phi\Bigg)$$
$$=\frac{1}{4g_X^2(x)}\Bigg\{\frac{1}{nh_n^2}\int K^2(\phi)\phi^4[h^{(2)}(x+\lambda_i^{'}h_n\phi)$$
$$-L^{(2)}(0,\theta^0)]^2g(x+h_n\phi)h_nd\phi$$
$$-\frac{1}{n}\Bigg(\frac{1}{h_n}\int K(\phi)\phi[h^{(2)}(x+\lambda_i^{'}h_n\phi)$$
$$-L^{(2)}(0,\theta^0)]g_X(x+h_n\phi)h_nd\phi\Bigg)^2\Bigg\}.$$

Observe that

$$\frac{-1}{4g_X^2(x)}\frac{1}{n}\left(\int K(\phi)\phi[h^{(2)}(x+\lambda_i^{'}h_n\phi)-L^{(2)}(0,\theta^0)]g_X(x+h_n\phi)d\phi\right)^2\to 0$$

as $n \to \infty$ and

$$\frac{1}{4g_X^2(x)}\frac{1}{nh_n^2}\int K^2(\phi)\phi^4[h^{(2)}(x+\lambda_i^{'}h_n\phi)-L^{(2)}(0,\theta^0)]^2g(x+h_n\phi)d\phi\to 0$$

provided that $nh_n \to \infty$ and given that $\int K^2(\phi)\phi^4 d\phi < C$. In all,

$$\frac{h_n}{n}\frac{1}{2g_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2 [\sigma^{2(2)}(\lambda_i^{'}(X_i - x) + x) - L^{(2)}(0, \theta^0)]$$
$$= \frac{1}{2}h_n^2(h^{(2)}(x) - L^{(2)}(0, \theta^0))\sigma_K^2 + o_p(h_n^2),$$

and

$$\frac{h_n}{n}\frac{1}{2g_X(x)}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2 [L^{(2)}(0, \theta^0) - L^{(2)}(\lambda_i(X_i - x), \hat{\theta}(x))]$$
$$= o_p(h_n^2).$$

Hence, we conclude that

$$\sqrt{nh_n} \quad \left(\gamma_1^*(x) - \sigma^2(x) - \frac{1}{2}h_n^2\sigma_K^2(\sigma^{2(2)}(x) - L^{(2)}(0, \theta^0)) + o_p(h_n^2)\right) \xrightarrow{d}$$
$$N\left(0, \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right),$$

which completes the proof.

Proof of Theorem 2. Since $\hat{\theta}(x) \equiv \begin{pmatrix} \hat{\theta}_1(x) \\ \hat{\theta}_2(x) \end{pmatrix}$ is in the interior of some compact subset $\bar{\Theta}$ of $\Re^2$ and satisfies

$$D_n(x, \hat{\theta}(x)) =$$
$$\begin{pmatrix} \frac{1}{nh_n}\sum_{i=1}^{n}(e_i - L(X_i - x, \hat{\theta}(x)))L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right) \\ \frac{1}{nh_n}\sum_{i=1}^{n}(e_i - L(X_i - x, \hat{\theta}(x)))L(X_i - x, \hat{\theta}(x))(X_i - x)K\left(\frac{X_i - x}{h_n}\right) \end{pmatrix} = 0$$

it suffices to show that there exists $D(x, \theta)$ such that $D(x, \theta^0(x)) = 0$ and $\underset{\theta \in \bar{\Theta}, x \in G}{sup}$ $||D_n(x, \theta(x)) - D(x, \Theta)|| = o_p(1)$ (van der Vaart (1998)). We focus on $D_{n,1}(x, \theta)$, the first element of $D_n(x, \theta)$ and set $D_1(x, \theta) = g_X(x)L(0, \theta)(\sigma^2(x) - L(0, \theta))$.

Given that $e_i = (Y_i - \hat{m}(X_i; h_n))^2$ we can write

$$\frac{1}{nh_n}\sum_{i=1}^{n}(\sigma^2(X_i)\epsilon_i^2 - \sigma^2(X_i) + \sigma^2(X_i) - L(X_i - x, \hat{\theta}(x)))$$
$$L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right)$$
$$+ \frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i) - \hat{m}(X_i; h_n))^2 L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right)$$
$$- \frac{2}{nh_n}\sum_{i=1}^{n}(\sigma^2(X_i))^{1/2}\epsilon_i L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right)(\hat{m}(X_i; h_n)$$
$$- m(X_i)) = 0.$$

or $D_{n,1}(x, \hat{\theta}) \equiv I_{1n}^*(x) + I_{2n}^*(x) - I_{3n}^*(x) + I_{4n}^*(x) = 0$ where

$$I_{1n}^*(x) = \frac{1}{nh_n}\sum_{i=1}^{n}(\sigma^2(X_i) - L(X_i - x, \hat{\theta}(x)))L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right)$$
$$I_{2n}^*(x) = \frac{1}{nh_n}\sum_{i=1}^{n}\sigma^2(X_i)(\epsilon_i^2 - 1)L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right)$$
$$I_{3n}^*(x) = \frac{2}{nh_n}\sum_{i=1}^{n}(\sigma^2(X_i))^{1/2}\epsilon_i L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right)(\hat{m}(X_i) - m(X_i))$$
$$I_{4n}^*(x) = \frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i) - \hat{m}(X_i; h_n))^2 L(X_i - x, \hat{\theta}(x))K\left(\frac{X_i - x}{h_n}\right)..$$

We will show that $\sup_{x\in G}\sup_{\theta\in\bar{\Theta}}\left|D_{n,1}(x,\theta) - g_X(x)(\sigma^2(x) - L(0,\theta))L(0,\theta)\right|$ $= o_p(1)$, and since $g_X(x)(\sigma^2(x) - L(0,\theta^0))L(0,\theta^0) = 0$ it follows that $\hat{\theta}(x)$ is a uniformly consistent estimator for θ^0. We start by considering $I_{2n}^*(x)$.

$$I_{2n}^*(x) = \frac{1}{nh_n}\sum_{i=1}^{n}\sigma^2(X_i)(\epsilon_i^2 - 1)[L(0,\theta) + L^{(1)}(0,\theta)(X_i - x)$$
$$\exp(\theta_2\lambda_i(X_i - x))]K\left(\frac{X_i - x}{h_n}\right)$$
$$= L(0,\theta)\frac{1}{nh_n}\sum_{i=1}^{n}\sigma^2(X_i)(\epsilon_i^2 - 1)K\left(\frac{X_i - x}{h_n}\right) + L^{(1)}(0,\theta)\frac{h_n}{nh_n}\sum_{i=1}^{n}\sigma^2(X_i)$$
$$\times\left(\frac{X_i - x}{h_n}\right)\exp(\theta_2\lambda_i(X_i - x))K\left(\frac{X_i - x}{h_n}\right)$$
$$= I_{21,n}^*(x) + I_{22,n}^*(x)$$

Now,

$$\sup_{x\in G}|I^*_{21,n}(x)| \leq |L(0,\theta)|\bar{B}_{g_X}\sup_{x\in G}\left|\frac{h_n}{nh_n}\sum_{i=1}^{n}\sigma^2(X_i)(\epsilon_i^2-1)K\left(\frac{X_i-x}{h_n}\right)\right|$$
$$= |L(0,\theta)|\bar{B}_{g_X}O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right),$$

and $\sup_{\theta\in\Theta}\sup_{x\in G}|I^*_{21,n}(x)| \leq \sup_{\theta\in\Theta}|L(0,\theta)|\bar{B}_{g_X}O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right)$. For $I^*_{22,n}(x)$ we have

$$|I^*_{22,n}(x)| \leq h_n|L^{(1)}(0,\theta)|\bar{B}_{g_X}\frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}\sigma^2(X_i)|\epsilon_i^2-1|\left|\frac{X_i-x}{h_n}\right|$$
$$\times\exp(\theta_2\lambda_i(X_i-x))K\left(\frac{X_i-x}{h_n}\right)$$
$$\leq h_n|L^{(1)}(0,\theta)|\bar{B}_{g_X}\frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}\sigma^2(X_i)|\epsilon_i^2-1|\left|\frac{X_i-x}{h_n}\right|$$
$$\times\exp(|\theta_2|h_n)K\left(\frac{X_i-x}{h_n}\right)$$

and therefore

$$\sup_{x\in G}|I^*_{22,n}(x)| \leq h_n|L^{(1)}(0,\theta)|\bar{B}_{g_X}e^{|\theta_2|h_n}\sup_{x\in G}\frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}\sigma^2(X_i)|\epsilon_i^2-1|$$
$$\times\left|\frac{X_i-x}{h_n}\right|\exp(|\theta_2|h_n)K\left(\frac{X_i-x}{h_n}\right).$$

Now, note that $\sigma^2(X_i)|\epsilon_i^2-1| = |(R_i-\mu_R)^2-\sigma_R^2|\sigma_R^{-2}\sigma^2(X_i) \leq C$ given that $R_i\in[0,1]$, and

$$E\left(\frac{1}{h_n}K\left(\frac{X_i-x}{h_n}\right)h(X_i)|\epsilon_i^2-1|\right) = \frac{1}{h_n}\int K\left(\frac{X_i-x}{h_n}\right)\sigma^2(X_i)$$
$$\times\frac{|(R_i-\mu_R)^2-\sigma_R^2|}{\sigma_R^2}g_X(X_i)g_R(R_i|X_i)dX_idR_i$$
$$= \frac{1}{\sigma_R^2h_n}\int K\left(\frac{X_i-x}{h_n}\right)\sigma^2(X_i)$$
$$\times\int|(R_i-\mu_R)^2$$
$$-\sigma_R^2|g_R(R_i|X_i)dR_ig_X(X_i)dX_i.$$

Now, since $\int |(R_i-\mu_R)^2-\sigma_R^2|g_R(R_i|X_i)dR_i \leq \int |(R_i-\mu_R)^2|g_R(R_i|X_i)dR_i+\sigma_R^2 \leq 2\sigma_R^2$, we will denote this integral by $\eta(X_i)$. So,

$$E\left(\frac{1}{h_n}K\left(\frac{X_i-x}{h_n}\right)\sigma^2(X_i)|\epsilon_i^2-1|\right)=\frac{1}{h_n\sigma_R^2}\int K(\phi)\sigma^2(x+h_n\phi)\eta(x+h_n\phi) \times g_X(x+h_n\phi)h_n d\phi$$

with

$$\sup_{x\in G}E\left(\frac{1}{h_n}K\left(\frac{X_i-x}{h_n}\right)\sigma^2(X_i)|\epsilon_i^2-1|\right)=\frac{1}{\sigma_R^2}\int K(\phi)d\phi\sup_{x\in G}\sigma^2(x)\sup_{x\in G}\eta(x) \times \sup_{x\in G}g_X(x)\leq C.$$

where C is an arbitrary constant. By Lemma 1 - part (a) in Martins-Filho and Yao (2007)

$$\sup_{x\in G}\frac{1}{nh_n g_X(x)}\sum_{i=1}^{n}\sigma^2(X_i)|\epsilon_i^2-1|K\left(\frac{X_i-x}{h_n}\right)=O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right)+O(1),$$

and consequently

$$\sup_{x\in G}|I^*_{22,n}(x)|\leq h_n|L^{(1)}(0,\theta)|\bar{B}_{gX}\exp(|\theta_2|h_n)\left[O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right)+O(1)\right]$$
$$\sup_{\theta\in\Theta}\sup_{x\in G}|I^*_{22,n}(x)|\leq \sup_{\theta\in\Theta}h_n|L^{(1)}(0,\theta)|\bar{B}_{gX}\exp(|\theta_2|h_n)\left[O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right)+O(1)\right].$$

We now turn our attention to $I^*_{3n}(x)$, which can be written as

$$\begin{aligned}
I^*_{3n}(x) &= \frac{2}{nh_n}\sum_{i=1}^{n}(\sigma^2(X_i))^{1/2}\epsilon_i K\left(\frac{X_i-x}{h_n}\right)(m(X_i)-\hat{m}(X_i;h_n)) \\
&\quad \times \left[L(0,\theta)+L^{(1)}(\lambda_i(X_i-x),\theta)\right] \\
&= \frac{2}{nh_n}\sum_{i=1}^{n}(\sigma^2(X_i))^{1/2}\epsilon_i K\left(\frac{X_i-x}{h_n}\right)(m(X_i)-\hat{m}(X_i;h_n))\Big[L(0,\theta) \\
&\quad + L^{(1)}(0,\theta)(X_i-x)\exp(\theta_2\lambda_i(X_i-x))\Big] \\
&= L(0,\theta)\frac{2}{nh_n}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_n))(\sigma^2(X_i))^{1/2}\epsilon_i K\left(\frac{X_i-x}{h_n}\right) \\
&\quad + L^{(1)}(0,\theta)\frac{2}{nh_n}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_n))(\sigma^2(X_i))^{1/2}\epsilon_i K\left(\frac{X_i-x}{h_n}\right) \\
&\qquad \times (X_i-x)\exp(\theta_2\lambda_i)(X_i-x)) \\
&= I^*_{31,n}(x)+I^*_{32,n}(x).
\end{aligned}$$

We write $I^*_{31,n}(x) = L(0,\theta)g_X(x)\frac{2}{nh_n g_X(x)}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_n))$ $(\sigma^2(X_i))^{1/2}\epsilon_i K\left(\frac{X_i-x}{h_n}\right)$. From Martins-Filho and Yao (2007) we have that

$$\begin{aligned}
\sup_{x\in G} \quad & \Bigg|L(0,\theta)\frac{2}{nh_n}\sum_{i=1}^{n}(m(X_i) \\
- \quad & \hat{m}(X_i;h_n))(\sigma^2(X_i))^{1/2}\epsilon_i K\left(\frac{X_i-x}{h_n}\right)\Bigg| \\
= \quad & O_p(h_n^2) \\
+ \quad & O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right).
\end{aligned}$$

Hence, $sup_{x\in G}|I^*_{31,n}(x)| \le L(0,\theta)\bar{B}_{g_X}\left[O_p(h_n^2)+O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right)\right]$, and $\sup_{\theta\in\bar{\Theta}}\sup_{x\in G}|I^*_{31,n}(x)| \le C\bar{B}_{g_X}\left[O_p(h_n^2)+O_p\left(\left(\frac{\ln(n)}{nh_n}\right)^{1/2}\right)\right]$, since $\sup_{\theta\in\bar{\Theta}}$

$L(0,\theta) = C$, given that $\bar{\Theta}$ is compact. $I^*_{32,n}(x)$ can be written as,

$$\begin{aligned}
I^*_{32,n}(x) &= 2L^{(1)}(0,\theta)g_X(x)\frac{h_n}{nh_n g_X(x)}\sum_{i=1}^{n}(m(X_i) \\
&- \hat{m}(X_i;h_n))(\sigma^2(X_i))^{1/2}\epsilon_i K\left(\frac{X_i - x}{h_n}\right) \\
&\times \left(\frac{X_i - x}{h_n}\right)\exp(\theta_2\lambda_i(X_i - x)) \\
&= 2L^{(1)}(0,\theta)g_X(x)h_n I^*_{321,n}(x) \\
|I^*_{32,n}(x)| &= 2|L^{(1)}(0,\theta)|g_X(x)h_n|I^*_{321,n}(x)|. \\
|I^*_{321,n}(x)| &\leq \frac{1}{nh_n g_X(x)}\sum_{i=1}^{n}|m(X_i) \\
&- \hat{m}(X_i,h_n)|(\sigma^2(X_i))^{1/2}|\epsilon_i|K\left(\frac{X_i - x}{h_n}\right) \\
&\times \left|\frac{X_i - x}{h_n}\right|\exp(\theta_2\lambda_i(X_i - x)).
\end{aligned}$$

Since, we have that if $\left|\frac{X_i - x}{h_n}\right| > 1$ then $K\left(\frac{X_i - x}{h_n}\right) = 0$ we can write that

$$\begin{aligned}
|I^*_{321,n}(x)| &\leq \frac{1}{nh_n g_X(x)}\sum_{i=1}^{n}|m(X_i) - \hat{m}(X_i;h_n)|(\sigma^2(X_i))^{1/2}|\epsilon_i|K\left(\frac{X_i - x}{h_n}\right) \\
&\times \left|\frac{X_i - x}{h_n}\right| e^{|\theta_2|h_n}
\end{aligned}$$

since $e^{\theta_2\lambda_i(X_i - x)} \leq e^{|\theta_2|h_n}$ given that $0 \leq \lambda_i \leq 1$. Now, note that from Fan and Yao (1998) and arguments similar to those used to establish Lemma 1, we have that for a bandwidth h_1 used in the first stage estimation, we have

$$\begin{aligned}
\hat{m}(X_i;h_1) - m(X_i) = \frac{1}{nh_1 g_X(X_i)}\sum_{t=1}^{n}K\left(\frac{X_t - X_i}{h_1}\right)\Big[Y_t - m(X_i) \\
- m^{(1)}(X_i)(X_t - X_i)\Big] + O_p(R_{n,2}(X_i))
\end{aligned}$$

where $R_{n,2}(X_i) = \frac{1}{n}\left|\sum_{t=1}^{n}K\left(\frac{X_t - X_i}{h_1}\right)Y^*_t\right| + \frac{1}{n}\left|\sum_{t=1}^{n}K\left(\frac{X_t - X_i}{h_1}\right)\left(\frac{X_t - X_i}{h_1}\right)Y^*_t\right|$ and $Y^*_t = Y_t - m(x) - m^{(1)}(x)(X_t - x) = \frac{1}{2}m^{(2)}(X_{ti})(X_t - X_i)^2 + (\sigma^2(X_t))^{1/2}\epsilon_t$, for $X_{ti} = \phi X_t + (1-\phi)X_i$ for some $\phi \in [0,1]$. Thus, we can write

$$\begin{aligned}\hat{m}(X_i;h_1) - m(X_i) &= \frac{h_1^2}{nh_1 g_X(X_i)} \sum_{t=1}^{n} K\left(\frac{X_t - X_i}{h_1}\right)\left(\frac{X_t - X_i}{h_1}\right)^2 \frac{1}{2} m^{(2)}(X_{ti}) \\ &+ \frac{1}{nh_1 g_X(X_i)} \sum_{t=1}^{n} K\left(\frac{X_t - X_i}{h_1}\right)(h(X_t))^{1/2}\epsilon_t \\ &+ O_p(R_{n,2}(X_i))\end{aligned}$$

and

$$\begin{aligned}|\hat{m}(X_i,h_1) - m(X_i)| &\leq \frac{h_1^2}{nh_1 g_X(X_i)} \sum_{t=1}^{n} K\left(\frac{X_t - X_i}{h_1}\right)\left(\frac{X_t - X_i}{h_1}\right)^2 \\ &\quad \times \frac{1}{2}|m^{(2)}(X_{ti})| \\ &+ \frac{1}{nh_1 g_X(X_i)} \sum_{t=1}^{n} K\left(\frac{X_t - X_i}{h_1}\right)(h(X_t))^{1/2}|\epsilon_t| \\ &+ O_p(R_{n,2}(X_i)).\end{aligned}$$

Consequently, we have

$$
\begin{aligned}
|I^*_{321,n}(x)| &\leq e^{|\theta_2|h_n}\frac{1}{nh_n g_X(x)}\sum_{i=1}^{n}\Bigg[\frac{h_1^2}{nh_1 g_X(X_i)}\sum_{t=1}^{n}K\left(\frac{X_t-X_i}{h_1}\right)\left(\frac{X_t-X_i}{h_1}\right)^2 \\
&\times \frac{1}{2}|m^{(2)}(X_{ti})| \\
&+ \frac{1}{nh_1 g_X(X_i)}\sum_{t=1}^{n}K\left(\frac{X_t-X_i}{h_1}\right)(h(X_t))^{1/2}|\epsilon_t| + O_p(R_{n,2}(X_i))\Bigg] \\
&\times (h(X_t))^{1/2}|\epsilon_t|K\left(\frac{X_i-x}{h_n}\right) \\
&= e^{|\theta|h_n}\frac{1}{nh_n g_X(x)}\frac{h_1^2}{nh_1}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{1}{g_X(X_i)} \\
&\times K\left(\frac{X_t-X_i}{h_1}\right)\left(\frac{X_t-X_i}{h_1}\right)^2 K\left(\frac{X_i-x}{h_n}\right) \\
&\times (\sigma^2(X_t))^{1/2}|\epsilon_t|\frac{1}{2}|m^{(2)}(X_{ti})| \\
&+ e^{|\theta|h_n}\frac{1}{nh_n g_X(x)}\frac{1}{nh_1}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{1}{g_X(X_i)}K\left(\frac{X_t-X_i}{h_1}\right)(\sigma^2(X_t))^{1/2}|\epsilon_t| \\
&\times (h(X_i))^{1/2}|\epsilon_i|K\left(\frac{X_i-x}{h_n}\right) \\
&+ e^{|\theta|h_n}\frac{1}{nh_n g_X(x)}\sum_{i=1}^{n}(\sigma^2(X_i))^{1/2}|\epsilon_i|K\left(\frac{X_i-x}{h_n}\right)O_p(R_{n,2}(X_i)) \\
&= L_{n,1}(x)+L_{n,2}(x)+L_{n,3}(x).
\end{aligned}
$$

We investigate each of these terms separately. First, we write

$$L_{n,1}(x) \leq e^{|\theta_2|h_n} \underline{\mathrm{B}}_{g_X}^{-1} \frac{1}{nh_n} \frac{h_1^2}{nh_1} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \times \frac{\sigma^2(X_i)^{1/2}}{g_X(X_i)} |\epsilon_i| \times \frac{1}{nh_1} K\left(\frac{X_t - X_i}{h_1}\right)\left(\frac{X_t - X_i}{h_1}\right)^2$$
$$\leq e^{|\theta_2|h_n} \underline{\mathrm{B}}_{g_X}^{-1} \frac{h_1^2}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \frac{\sigma^2(X_i)^{1/2}}{g_X(X_i)} |\epsilon_i| \times \frac{1}{h_1} \sup_{x \in G} \frac{1}{n} \sum_{t=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2.$$

By (Martins-Filho and Yao, 2007, p. 307) we have that

$$\frac{1}{n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \frac{\sigma^2(X_i)^{1/2}}{g_X(X_i)} |\epsilon_i| = O_p(h_n)$$

and

$$\sup_{x \in G} \frac{1}{n} \sum_{t=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2 = O_p(h_1).$$

Hence, $\sup_{\theta \in \bar{\Theta}} \sup_{x \in G} L_{n,1}(x) \leq \sup_{\theta \in \bar{\Theta}} e^{|\theta_2|h_n} \underline{\mathrm{B}}_{g_X}^{-1} h_1^2 O_p(1)$. Second, we write

$$L_{n,2}(x) = e^{|\theta|h_n} \frac{1}{nh_n g_X(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \frac{\sigma^2(X_i)^{1/2}}{g_X(X_i)} |\epsilon_i| \times \sum_{t=1}^{n} \frac{1}{nh_1} K\left(\frac{X_t - X_i}{h_1}\right) \sigma^2(X_t)^{1/2} |\epsilon_t|$$
$$\leq e^{|\theta_2|h_n} \underline{\mathrm{B}}_{g_X}^{-1} \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \frac{\sigma^2(X_i)^{1/2}}{g_X(X_i)} |\epsilon_i| \times \sup_{x \in G} \frac{1}{nh_1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_1}\right) \sigma^2(X_t)^{1/2} |\epsilon_t|.$$

Now, observe that

$$E\left(\frac{1}{h_1}K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}|\epsilon_t|\right) = \frac{1}{h_1}\int K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}\frac{1}{\sigma_R}$$
$$\times |R_t-\mu_R|g_X(X_t)\times g_{R|X}(R_t;X_t)dX_tdR_t$$
$$=\frac{1}{\sigma_R}\frac{1}{h_1}\int K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}$$
$$\times\int|R_t-\mu_R|\times g_{R|X}(R_t;X_t)dR_tg_X(X_t)dX_t$$
$$=\frac{1}{\sigma_R}\frac{1}{h_1}\int K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}$$
$$\times\mu_1(X_t)g_X(X_t)dX_t$$
$$=\frac{1}{\sigma_R}\frac{1}{h_1}\int K(\phi)(\sigma^2(x+h_1\phi))^{1/2}$$
$$\times\mu_1(x+h_1\phi)h_1g_X(x+h_1\phi)d\phi.$$

Hence,

$$\sup_{x\in G}E\left(\frac{1}{h_1}K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}|\epsilon_t|\right)\leq\frac{1}{\sigma_R}\int K(\phi)d\phi\sup_{x\in G}\sigma^2(x))^{1/2}\sup_{x\in G}g_X(x)$$
$$\times\sup_{x\in G}\mu_1(x)\leq C,$$

since $(\sigma^2(X_t))^{1/2}|\epsilon_t|=\sigma^2(X_t)^{1/2}\frac{1}{\sigma_R}|R_t-\mu_R|<C$, by Lemma 1 in Martins-Filho and Yao (2007), part (a), if $nh_n^3\to\infty$. Therefore, we can write

$$\sup_{x\in G}\frac{1}{nh_1}\sum_{t=1}^{n}K\left(\frac{X_t-x}{h_1}\right)$$
$$\times(\sigma^2(X_t))^{1/2}|\epsilon_t|\leq\sup_{x\in G}\left|\frac{1}{nh_1}\sum_{t=1}^{n}K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}|\epsilon_t|\right.$$
$$\left.-E\left(\frac{1}{h_1}K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}|\epsilon_t|\right)\right|$$
$$+\sup_{x\in G}\frac{1}{h_1}E\left(K\left(\frac{X_t-x}{h_1}\right)(\sigma^2(X_t))^{1/2}|\epsilon_t|\right)$$
$$=O_p\left(\left(\frac{\ln(n)}{nh_1}\right)^{1/2}\right)+O(1).$$

Hence, $\sup_{\theta\in\bar\Theta}\sup_{x\in G}L_{n,2}(x)\leq\sup_{\theta\in\Theta}e^{|\theta_2|h_n}\underline{\text{B}}_{gX}^{-1}O_p(1)$. Lastly, by (Martins-Filho and Yao, 2007, p. 308) $\sup_{\theta\in\Theta}\sup_{x\in G}L_{n,3}(x)\leq\sup_{\theta\in\bar\Theta}e^{|\theta_2|h_n}\underline{\text{B}}_{gX}^{-1}h_n^2o_p(1)$.

Combining the results on $L_{n,1}(x)$, $L_{n,2}(x)$ and $L_{n,3}(x)$ we have, together with compactness of $\bar{\Theta}$, that $\sup_{\theta\in\bar{\Theta}}\sup_{x\in G}|I^*_{32,n}(x)| = o_p(1)$. Now, consider

$$\begin{aligned}
I^*_{4n}(x) &= \frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_1))^2\Big[L(0,\theta) \\
&\quad + L^{(1)}(0,\theta)(X_i-x)e^{\theta_2\lambda_i(X_i-x)}\Big]K\left(\frac{X_i-x}{h_n}\right) \\
&= L(0,\theta)\frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_1))^2K\left(\frac{X_i-x}{h_n}\right) \\
&\quad + L^{(1)}(0,\theta)\frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i) \\
&\quad - \hat{m}(X_i;h_1))^2(X_i-x)e^{\theta_2\lambda_i(X_i-x)}K\left(\frac{X_i-x}{h_n}\right) \\
&= I^*_{41,n}(x)+I^*_{42,n}(x).
\end{aligned}$$

We consider each of these terms separately.

$$\begin{aligned}
I^*_{41,n}(x) &= L(0,\theta)g_X(x)\frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}(m(X_i) \\
&\quad - \hat{m}(X_i;h_n))^2K\left(\frac{X_i-x}{h_n}\right) \text{ which gives} \\
\sup_{x\in G}|I^*_{41,n}(x)| &\leq |L(0,\theta)|\bar{B}_{g_X}\sup_{x\in G}\left|\frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}(m(X_i)\right. \\
&\quad \left. - \hat{m}(X_i;h_n))^2K\left(\frac{X_i-x}{h_n}\right)\right|
\end{aligned}$$

By (Martins-Filho and Yao, 2007, p. 310)

$$\sup_{x\in G}\left|\frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_n))^2K\left(\frac{X_i-x}{h_n}\right)\right| = o_p(h_n^2) \text{ and consequently}$$

$\sup_{\theta\in\bar{\Theta}}\sup_{x\in G}|I^*_{41,n}(x)|\leq \sup_{\theta\in\bar{\Theta}}|L(0,\theta)|\bar{B}_{gX}h_n^2 o_p(1)$. Now,

$$I^*_{42,n}(x) = L^{(1)}(0,\theta)h_n\frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_1))^2K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right) \times e^{\theta_2\lambda_i(X_i-x)} \text{ and}$$

$$|I^*_{42,n}(x)| \leq |L^{(1)}(0,\theta)|h_n\frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i) - \hat{m}(X_i;h_1))^2K\left(\frac{X_i-x}{h_n}\right)\left|\frac{X_i-x}{h_n}\right|e^{\theta_2\lambda_i(X_i-x)}$$

Again, by compactness of the support of K, we have

$$|I^*_{42,n}(x)| \leq |L^{(1)}(0,\theta)|h_n\frac{1}{nh_n}\sum_{i=1}^{n}(m(X_i)-\hat{m}(X_i;h_1))^2K\left(\frac{X_i-x}{h_n}\right) \times \left|\frac{X_i-x}{h_n}\right|e^{|\theta_2|h_n} \text{ and}$$

$$\sup_{x\in G}|I^*_{42,n}(x)| \leq h_n\bar{B}_{gX}|L^{(1)}(0,\theta)|e^{|\theta_2|h_n}o_p(h_n^2).$$

Consequently, $\sup_{\theta\in\bar{\Theta}}\sup_{x\in G}|I^*_{42,n}(x)| \leq h_n^3\bar{B}_{gX}\sup_{\theta\in\Theta}|L^{(1)}(0,\theta)|e^{|\theta_2|h_n}$. We now examine $I^*_{1n}(x)$. We start by noting that

$$\begin{aligned}\sigma^2(X_i)-L(X_i-x,\hat{\theta}(x)) &= \sigma^2(X_i)-L(0,\theta)-L^{(1)}(0,\theta)(X_i-x)e^{\theta_2\lambda_i(X_i-x)}\\ &= \sigma^2(x)+\sigma^{2(1)}(\lambda_i^{'}(X_i-x)+x)(X_i-x)-L(0,\theta)\\ &\quad - L^{(1)}(0,\theta)(X_i-x)e^{\theta_2\lambda_i(X_i-x)}.\end{aligned}$$

Therefore,

$$\begin{aligned}I^*_{1n}(x) &= g_X(x)\frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}\Big[\sigma^2(x)-L(0,\theta)+\sigma^{2(1)}(\lambda_i^{'}(X_i-x)+x)(X_i-x)\\ &\quad - L^{(1)}(0,\theta)(X_i-x)e^{\theta_2\lambda_i(X_i-x)}\Big]K\left(\frac{X_i-x}{h_n}\right)L(X_i-x,\theta)\\ &= g_X(x)\Bigg\{\frac{1}{nh_ng_X(x)}(\sigma^2(x)-L(0,\theta))\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)L(X_i-x,\theta)\\ &\quad + \frac{1}{nh_ng_X(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)\Big[\sigma^{2(1)}(\lambda_i^{'}(X_i-x)+x)(X_i-x)\\ &\quad - L^{(1)}(0,\theta)(X_i-x)e^{\theta_2\lambda_i(X_i-x)}\Big]L(X_i-x,\theta)\Bigg\}\\ &= g_X(x)(I^*_{11,n}(x)+I^*_{12,n}(x)).\end{aligned}$$

We now look at $I^*_{11,n}(x)$, and $I^*_{12,n}(x)$ in isolation.

$$\begin{aligned}
I^*_{11,n}(x) &= [\sigma^2(x) - L(0,\theta)]\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) L(X_i - x, \theta) \\
&= [\sigma^2(x) - L(0,\theta)]\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \Big(L(0,\theta) \\
&\quad - L^{(1)}(0,\theta)(X_i - x)e^{\theta_2\lambda_i(X_i - x)}\Big) \\
&= [\sigma^2(x) - L(0,\theta)]\Big\{L(0,\theta)\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) - h_n L^{(1)}(0,\theta) \\
&\quad \times \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right) e^{\theta_2\lambda_i(X_i - x)}\Big\}.
\end{aligned}$$

From Lemma 1 in Martins-Filho and Yao (2007) we have that $\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)$ converges uniformly to $g_X(x)$ on a compact set G, hence the first term converges uniformly to $g_X(x)L(0,\theta)$ on G. By arguments made earlier in the proof we have that $\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right) \times e^{\theta_2\lambda_i(X_i - x)}$ is uniformly bounded in probability on G, hence the last term is $o_p(1)$ uniformly in G. Hence, we gave $I^*_{11,n}(x) \xrightarrow{p} g_X(x)[h(x) - L(0,\theta)]L(0,\theta)$. Now we treat $I^*_{12,n}(x)$. Note that,

$$\begin{aligned}
I^*_{12,n}(x) &= \frac{h_n}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)\Big(\sigma^{2(1)}(\lambda_i'(X_i - x) + x) - L^{(1)}(0,\theta) \\
&\quad \times e^{\theta_2\lambda_i(X_i - x)}\Big)\Big(L(0,\theta) - h_n L^{(1)}(0,\theta)\left(\frac{X_i - x}{h_n}\right) e^{\theta_2\lambda_i(X_i - x)}\Big) \\
&= h_n\frac{1}{nh_n}L(0,\theta)\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)\Big(\sigma^{2(1)}(\lambda_i'(X_i - x) + x) \\
&\quad - L^{(1)}(0,\theta)e^{\theta_2\lambda_i(X_i - x)}\Big) \\
&\quad - h_n^2\frac{1}{nh_n}L^{(1)}(0,\theta)\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2 \\
&\quad \times e^{\theta_2\lambda_i(X_i - x)}\Big(\sigma^{2(1)}(\lambda_i'(X_i - x) + x) - L^{(1)}(0,\theta)e^{\theta_2\lambda_i(X_i - x)}\Big)
\end{aligned}$$

$$= h_n \frac{1}{nh_n} L(0,\theta) \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)\Bigg(\sigma^{2(1)}(\lambda_i^{'}(X_i - x) + x)$$

$$- L^{(1)}(\lambda_i(X_i - x), \theta)\Bigg) - h_n^2 \frac{1}{nh_n} L^{(1)}(0,\theta) \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2$$

$$\times e^{\theta_2 \lambda_i (X_i - x)}\Bigg(\sigma^{2(1)}(\lambda_i^{'}(X_i - x) + x) - L^{(1)}(\lambda_i(X_i - x), \theta)\Bigg)$$

$$= I^*_{121,n}(x) - I^*_{122,n}(x).$$

Observe that

$$I^*_{121,n}(x) = h_n \frac{1}{nh_n} L(0,\theta) \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)[\sigma^{2(1)}(\lambda_i^{'}(X_i - x) + x)$$

$$- L^{(1)}(0,\theta)]$$

$$+ h_n \frac{1}{nh_n} L(0,\theta) \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)[L^{(1)}(0,\theta)$$

$$- L^{(1)}(\lambda_i(X_i - x), \theta)],$$

where $L^{(1)}(0,\theta) = \sigma^2(x) < C$. Also, $|\sigma^{2(1)}(x)| \leq \sigma^2(x)|f^{(1)}(x)| < C$ provided $|f^{(1)}(x)| < B_f$. Hence,

$$|I^*_{121,n}(x)| \leq h_n \underline{\mathrm{B}}_{gX}^{-1} C \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left|\frac{X_i - x}{h_n}\right|$$

$$+ h_n \underline{\mathrm{B}}_{gX}^{-1} C \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left|\frac{X_i - x}{h_n}\right|[L^{(1)}(0,\theta)$$

$$- L^{(1)}(\lambda_i(X_i - x), \theta)].$$

Since, $K\left(\frac{X_i - x}{h_n}\right) = 0$ whenever $\left|\frac{X_i - x}{h_n}\right| < 1$,

$$|I^*_{121,n}(x)| \leq h_n \underline{\mathrm{B}}_{gX}^{-1} C \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)$$

$$+ h_n \underline{\mathrm{B}}_{gX}^{-1} C \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)|L^{(1)}(0,\theta) - L^{(1)}(\lambda_i(X_i - x), \theta)|.$$

Now, $|L^{(1)}(0,\theta) - L^{(1)}(\lambda_i(X_i - x), \theta)| = \sigma^2(x)|1 - e^{\theta_2 \lambda_i (X_i - x)}| \leq \sigma^2(x)(1 + e^{h_n|\theta_2|})$. Hence, $|I^*_{121,n}(x)| \leq h_n \underline{\mathrm{B}}_{gX}^{-1} C \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) + h_n \sigma^2(x)(1 + e^{h_n|\theta_2|})\ \underline{\mathrm{B}}_{gX}^{-1} C \frac{1}{nh_n}$

$\times \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)$. Again, using Lemma 1 in Martins-Filho and Yao (2007),

$$\sup_{x\in G} |I^*_{121,n}(x)| \leq h_n \underline{\mathrm{B}}_{g_X}^{-1} B_h C[g_X(x) + O_p(h_n)] + h_n h(x)(1 + e^{h_n|\theta_2|})\underline{\mathrm{B}}_{g_X}^{-1} B_h [g_X(x) + O_p(h_n)],$$

which gives $\sup_{\theta\in\bar{\Theta}, x\in G} |I^*_{121,n}(x)| = o_p(1)$. Similar arguments show that $\sup_{\theta\in\bar{\Theta}, x\in G} |I^*_{122,n}(x)| = o_p(1)$ completing the proof for $D_{n,1}(x)$.

For the second element $D_{n,2}(x,\theta)$ of the vector $D_n(x,\theta)$ we put $D_2(x,\theta) = 0$ and note that by assumption A3 it can be verified, given the arguments used above, that $\sup_{\theta\in\bar{\Theta}, x\in G} |D_{n,2}(x,\theta)| = o_p(1)$.