Journal of Multivariate Analysis 100 (2009) 309–333

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonparametric regression estimation with general parametric error covariance

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, West Virginia University, P.O. Box 6025, Morgantown, WV 26506-6025, USA*

ARTICLE INFO

Article history:
Received 11 July 2007
Available online 29 April 2008

AMS 2000 subject classifications:
62G08
62G20

Keywords:
Local linear estimation
Asymptotic normality
Mixing processes

ABSTRACT

The asymptotic distribution for the local linear estimator in nonparametric regression models is established under a general parametric error covariance with dependent and heterogeneously distributed regressors. A two-step estimation procedure that incorporates the parametric information in the error covariance matrix is proposed. Sufficient conditions for its asymptotic normality are given and its efficiency relative to the local linear estimator is established. We give examples of how our results are useful in some recently studied regression models. A Monte Carlo study confirms the asymptotic theory predictions and compares our estimator with some recently proposed alternative estimation procedures.

1. Introduction

Recently there has been a growing interest in the specification of nonparametric regression models in which the regression errors' correlation structure can be described parametrically. For example, Xiao et al. [38] consider a nonparametric regression with stationary error terms that have an invertible linear process representation which encompasses all finite order ARMA(p, q) processes; Vilar-Fernández and Francisco-Fernández [32] consider a fixed design nonparametric regression whose errors follow an AR(1) process; Lin and Carroll [18], Ruckstuhl et al. [25] and Wang [34] consider a nonparametric regression for panel/clustered data where the error term covariance structure follows a pre-specified parametric structure; Fan et al. [11] consider a nonparametric regression frontier model with errors whose covariance structure follows a parametric specification proposed by Aigner et al. [1] and Smith and Kohn [28] consider the estimation of a finite set of nonparametric regressions whose error structure follows the parametric seemingly unrelated structure proposed by Zellner [39].

These models can be viewed as extensions of the regression literature in two related but distinct ways. First, they represent an extension of the vast Generalized Least Squares (GLS) linear and nonlinear parametric regression literatures [13, 37] to the nonparametric regression setting, and as such they represent improvements on the modeling of (un)conditional expectations. Second, they can be viewed as extensions of the nonparametric regression literature from the typical case where regression errors are independent and identically distributed (iid) to cases where specific parametric structures for correlation and heteroscedasticity are allowed [27]. In either case, the usefulness of these extensions in econometric and statistical practice is well recognized and documented [23,10]. In their most general form, these regression models can be

* Corresponding author.
E-mail addresses: carlos.martins@oregonstate.edu (C. Martins-Filho), yaofengyf@gmail.com (F. Yao).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2008.04.013

written as

$$Y_i = m(X_i) + U_i, \quad i = 1, 2, \ldots \tag{1}$$

where X_i is a vector of regressors, Y_i is a regressand and the error U_i is such that

$$E(U_i) = 0 \quad \text{for all } i = 1, 2, \ldots, \qquad E(U_i U_j) = \omega_{ij}(\theta_0), \quad \theta_0 \in \Re^p, p < \infty. \tag{2}$$

The important characteristic of (2) is that each element of the error covariance can be expressed as a function $\omega_{ij}(\theta)$ of a finite set of parameters θ_0. Previous works on the estimation of these models have had two main objectives. The first is to establish the asymptotic properties of well known nonparametric regression estimators such as local polynomial and Nadaraya–Watson estimators under the assumed error correlation structure [38,32]. Although progress in this direction has been made, it is unfortunate that most asymptotic results for traditional estimators are specific to the assumed covariance structure and lack the generality that would allow their applicability under alternative parametric structures for the error correlation. A more general result under covariance structure (2) for the local linear estimator seems to be especially useful as this estimator has a number desirable properties, such as design adaptability, reduced bias (as compared to Nadaraya–Watson estimators), good boundary properties and mini–max efficiency [7–9]. The first contribution of this paper is to provide a set of sufficient conditions under which the asymptotic normality of the local linear estimator can be established when the error correlation structure has the general parametric structure in (2). These conditions encompass a number of models proposed so far in the nonparametric literature as well as other structures that have been popular in the GLS parametric literature [20].

The second objective of the existing literature is to propose estimators that by incorporating the information contained in the error covariance structure will lead to better performance – asymptotically or in finite sample – *vis a vis* the traditional estimators [27,18,25,34]. How to best incorporate the error covariance matrix information into local polynomial nonparametric regression estimators is still an open question. Lin and Carroll [18] show that in typical random effects panel data models, when a standard kernel based estimator is used, it is better to estimate the regression by ignoring the correlation structure within a cluster — the "working independence" approach. An alternative kernel smoothing method proposed by Wang [34] achieves smaller variance when the correlation structure is taken into account. However, it is not clear how to generalize this approach to the case of a general error covariance. A particularly promising approach has been the *pre-whiten* method proposed by Ruckstuhl et al. [25] and adopted by Xiao et al. [38]. However, as in the case of the local linear estimator, the asymptotic properties of this *pre-whiten* estimator have been established only for specific parametric structures of the error covariance (random effects panel data and autocorrelated errors). In fact, as will be argued below, establishing the asymptotic normality of the *pre-whiten* estimator in general settings could be quite difficult. Hence, in the second part of this paper we propose a new two-step estimator, inspired by Ruckstuhl et al. [25], that incorporates information contained in the error covariance structure and is asymptotically normal under fairly mild restrictions on the parametric structure of the covariance (see Assumptions A6 and TA 4.1–4.3). Our estimator is an improvement over the traditional local linear estimator in that its bias is of the same order but its asymptotic distribution has strictly smaller variance.

Our results are useful from at least two perspectives. First, since our results hold for generally specified parametric covariances, they eliminate the need to repeatedly establish asymptotic normality for both estimators – local linear and the two-step procedure proposed herein – under specific structures of $\omega_{ij}(\theta_0)$. Second, because the two estimators are asymptotically normal and converge at similar rates, establishing relative efficiency is facilitated. At their technical core, both contributions in this paper can be viewed as extensions to the results of Mack and Silverman [19] and Masry and Fan [22]. These extensions are made possible by relying on inequalities for non-stationary processes provided by Doukhan [6] and Volkonskii and Rozanov [33]. The rest of the paper is organized as follows. Section 2 provides the general characteristics of the regression model that we consider, defines the local linear estimator, and gives a list of assumptions and the two main theorems necessary for establishing the properties of the local linear estimator for model (1)–(2). In Section 3 we define a new two-step estimator based on the knowledge of $\omega_{ij}(\theta_0)$ and give sufficient conditions for obtaining its asymptotic normality. We then obtain the asymptotic equivalence of the two-step estimator based on $\omega_{ij}(\theta_0)$ and its feasible version based on an estimator $\omega_{ij}(\dot{\theta})$, where $\dot{\theta} - \theta_0 = o_p(1)$. These results are obtained for a special class of parametric covariances as specified by our Assumptions A6 and TA 4.1–4.3 in Theorem 4. Section 4 gives two applications of our results that illustrate how our theorems encompass and extend previous results in the literature. Section 5 contains a Monte Carlo study that implements our two-step estimator, sheds some light on its finite sample properties, and compares its performance to that of existing estimators. Section 6 provides a summary of the paper.

2. A nonparametric regression model with general parametric covariance

Suppose there are n observations $\vec{y} = (Y_1, \ldots, Y_n)'$, $\vec{x} = (X_1, \ldots, X_n)'$ on the regressand and regressors for the model (1)–(2). The objective is to estimate the regression function $m(x)$ at some point $x \in \Re^D$, $D < n$.[1] There is a vast literature [15]

[1] In what follows we proceed for simplicity with the assumption that $D = 1$. *Mutatis mutandis* all results follow for $D > 1$.

on how to proceed with estimation of m. Here, we focus our attention on the local linear estimator (LLE) which was popularized by Fan [7] due to its well known desirable properties. Furthermore, our results for the LLE are easily extended to the also popular Nadaraya–Watson estimator. Let $e' = (1, 0)$, $1_n' = (1, \dots, 1)$, a vector of ones of length n, and $h_n > 0$ a sequence of bandwidths; then the LLE is defined as

$$\check{m}(x) = e' \left(R_x' K_x R_x\right)^{-1} R_x' K_x \vec{y} \tag{3}$$

where $R_x = (1_n, \vec{x} - 1_n x)$, $K_x = \operatorname{diag}\left\{K\left(\frac{X_i - x}{h_n}\right)\right\}_{i=1}^n$. It will be convenient for our purposes to rewrite (3) as $\check{m}(x) = \frac{1}{nh_n}\sum_{i=1}^n W_n\left(\frac{x_i - x}{h_n}, x\right) Y_i$, where $W_n(z, x) = e' S_n^{-1}(x)(1, z)' K(z)$ and

$$S_n(x) = \frac{1}{nh_n}\begin{pmatrix} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) & \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right) \\ \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right) & \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^2 \end{pmatrix} = \begin{pmatrix} s_{n,0}(x) & s_{n,1}(x) \\ s_{n,1}(x) & s_{n,2}(x) \end{pmatrix}.$$

To establish the asymptotic normality of $\check{m}(x)$ for model (1)–(2) we follow the traditional approach of breaking the problem into two parts. First, we establish the uniform convergence in probability of the components of $R_x' K_x R_x$ after a suitable normalization. This is accomplished as an application of Theorem 1 which is given below. Second, we establish the asymptotic distribution of the $R_x' K_x \vec{y}$ vector (and of the estimator itself) in Theorem 2. We now provide a list of general assumptions that will be selectively adopted in these theorems and introduce some notation. In what follows C always denotes a generic constant that may take different values in $\Re$ and the sequence of bandwidths h_n is such that $h_n \to 0$ and $nh_n^2 \to \infty$ as $n \to \infty$.

Assumption A1. 1. Let $f_i(x)$ be the marginal density of X_i evaluated at x, with $f_i(x) < C$ for all i and x; 2. $f_i^{(d)}(x)$ is the dth-order derivative of $f_i(x)$ evaluated at x and we assume that $|f_i^{(1)}(x)| < C$; 3. $|f_i(x) - f_i(x')| \le C|x - x'|$ for all x, x'; 4. $f_{lkijmo}(x_l, \dots, x_o)$ denotes the joint density of $X_l, \dots, X_o$ evaluated at $x_l, \dots, x_o$ and we assume that $f_{lkijmo}(x_l, \dots, x_o) < C$ for all $x_l, \dots, x_o$. 5. $\bar{f}_n(x) = n^{-1}\sum_{i=1}^n f_i(x) \to \bar{f}(x)$ as $n \to \infty$ where $0 < \bar{f}(x) < \infty$; 6. as $n \to \infty$ $0 < \inf_{x \in G}|\bar{f}_n(x)| < C$ for $x \in G$ a compact set.

Assumption A2. $K(x) : \Re \to \Re$ is a symmetric bounded function with compact support S_K such that: 1. $\int K(x)\mathrm{d}x = 1$; 2. $\int xK(x)\mathrm{d}x = 0$; 3. $\int x^2 K(x)\mathrm{d}x = \sigma_K^2$; 4. for all $x, x' \in S_K$ we have $|K(x) - K(x')| \le C|x - x'|$.

Assumption A3. $\omega_{ij}(\theta_0)$ is the (i, j) element of $\Omega = E(UU')$ with $|\omega_{ij}(\theta_0)| < C$ for all i, j, $\bar{\omega}_n(\theta) = n^{-1}\sum_{i=1}^n \omega_{ii}(\theta) \to \bar{\omega}(\theta)$ as $n \to \infty$ where $0 < \bar{\omega}(\theta) < \infty$ for every θ and $\bar{\omega}_{fn}(x, \theta) = n^{-1}\sum_{i=1}^n \omega_{ii}(\theta) f_i(x) \to \bar{\omega}_f(x, \theta)$ as $n \to \infty$ where $0 < \bar{\omega}_f(x, \theta) < \infty$ for every x and θ.

Let $\{R_t\}$ be a sequence of random variables defined in a probability space (S, F, P) and $\Im_a^b$ be the σ-algebra of events generated by the random variables $\{R_t : a \le t \le b\}$; then $\alpha(\Im_a^b, \Im_c^d) = \sup_{A \in \Im_a^b, B \in \Im_c^d} |P(A \cap B) - P(A)P(B)|$ and $\alpha(m) = \sup_t \alpha(\Im_{-\infty}^t, \Im_{t+m}^\infty)$. A stochastic process is said to be α-mixing if process $\alpha(m) \to 0$ as $m \to \infty$. Then we assume:

Assumption A4. 1. $\{(X_i, U_i)'\}_{i=1,2,\dots}$ is an α-mixing process of size -2, which implies that $\sum_{j=1}^\infty j^a \alpha(j)^{1-\frac{2}{\delta}} < \infty$ for $\delta > 2$ and $a > 1 - 2/\delta$; 2. we denote the joint density of $(X_i, U_i)'$ by $f_{X_i,U_i}(x_i, u_i)$, the density of X_i conditional on U_i by $f_{X_i|U_i}(x)$ with $f_{X_i|U_i}(x) < C$ and the conditional density of X_i, X_j given U_i, U_j by $f_{X_iX_j|U_iU_j}(x_i, x_j)$ with $f_{X_iX_j|U_iU_j}(x_i, x_j) < C$ for all x_i, x_j; 3. there exists a sequence of positive integers satisfying $s_n \to \infty$ and $s_n = o((nh_n)^{1/2})$ such that $\left(\frac{n}{h_n}\right)^{1/2}\alpha(s_n) \to 0$ as $n \to \infty$.

Assumption A5. $m^{(d)}(x) < C$ for all x and $d = 1, 2$, where $m^{(d)}(x)$ is the dth-order derivative of $m(x)$ evaluated at x.

Our Assumption A1 requires the densities of regressor X_i to be smooth and bounded functions, and in the case where the X_i come from heterogeneous distributions, the average of the densities must converge. This is automatically satisfied if the X_i come from the same distribution, or the X_i are part of a strictly stationary sequence. Assumption A2 is a standard assumption for the kernel functions in the nonparametric regression estimation. Assumption A3 ensures that the weighted average of the diagonal terms of the error covariance converges as $n \to \infty$, which is trivially met when there is a homoscedastic error structure. Under the mixing conditions imposed in Assumption A4, the dependence among $\{(X_i, U_i)'\}$ will diminish as the distance between indices increases, which is general enough to include many interesting cases like panel data models or autoregressive models of order (p) (see Section 4), while still allowing a central limit theorem to apply on the standardized summation. We impose a smoothness condition on $m(x)$ in Assumption A5 so the standard Taylor approximations could carry through.

We now state Theorem 1 which is a supporting result for the main theorems that follow. All proofs are provided in Appendix A.

Theorem 1. *Let $\{(X_i, U_i)\}_{i=1}^n$ be a stochastic sequence of vectors, $\{v_i\}_{i=1}^n$ be a uniformly bounded non-stochastic sequence in $\Re$ and define*

$$s_j(x) = (nh_n)^{-1}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\left(\frac{X_i - x}{h_n}\right)^j g(U_i)v_i \quad \text{with } j = 0, 1, 2$$

where $g : \Re \to \Re$ is measurable. Assume that: 1. $E(|g(U_i)|^{2+\theta}) < C$ *for some $\theta > 0$ and all i;* 2. $\sup_{x\in G}\int |g(U_i)|^a f_{X_i,U_i}(x, U_i)\mathrm{d}U_i < \infty$ *for some $a > 1$;* 3. *Assumptions A2 and A4. For G a compact subset of $\Re$ we have*

$$\sup_{x\in G} |s_j(x) - E(s_j(x))| = O_p\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}\right) \tag{4}$$

provided that $s, \beta > 2$ and we have that $n^{(\theta+1/s)(\beta+1.5)+1.25-\beta/2}h_n^{-1.75-\beta/2}(\ln(n))^{0.25+\beta/2} \to 0$.

By taking $v_i = 1$ and $g(x) = 1$ for all i and x in Theorem 1 we have that $\sup_{x\in G} |s_{n,j}(x) - E(s_{n,j}(x))| = o_p(h_n^p)$ for $p > 0$ and $j = 0, 1, 2$ provided that $\frac{nh_n^{2p+1}}{\ln(n)} \to \infty$. The last condition is consistent with $n^{(\theta+1/s)(\beta+1.5)+1.25-\beta/2}h_n^{-1.75-\beta/2}(\ln(n))^{0.25+\beta/2} \to 0$ as $n \to \infty$ for $\theta > 0$ and $s > 2$. Consequently, if $p = 1$, $\frac{nh_n^3}{\ln(n)} \to \infty$ we have that $\sup_{x\in G}\frac{1}{h_n}|s_{n,j}(x) - E(s_{n,j}(x))| = o_p(1)$.

The next theorem establishes the asymptotic $\sqrt{nh_n}$- normality for the local linear estimator under general parametric covariance structure. We stress that the importance of the result lies in the fact that the regression errors are not restricted to being (iid) or even weakly stationary. We do assume, however, that $\{X_i\}_{i=1,2,\ldots}$ and $\{U_i\}_{i=1,2,\ldots}$ are independent processes.

Theorem 2. *Let $\{(X_i, U_i)\}_{i=1}^n$ be a stochastic sequence of vectors and assume that $Y_i = m(X_i) + U_i$ for $i = 1, 2, \ldots$, $\{X_i\}_{i=1,2,\ldots}$ and $\{U_i\}_{i=1,2,\ldots}$ are independent with $E(U_i) = 0$ for all $i = 1, 2, \ldots$, $E(U_iU_j) = \omega_{ij}(\theta_0)$ $\theta_0 \in \Re^p$, $p < \infty$. If we assume that Assumptions A1–A5 are met and $E(|U_i|^{2+\theta}) < C$ for some $\theta > 0$ and all i, then*

$$(nh_n)^{1/2}(\check{m}(x) - m(x) - B_{n,1}(x)) \xrightarrow{d} N\left(0, \frac{\bar{\omega}_f(x,\theta_0)}{\bar{f}^2(x)}\int K^2(\phi)\mathrm{d}\phi\right) \tag{5}$$

where $B_{n,1}(x) = \frac{h_n^2}{2}\sigma_K^2 m^{(2)}(x) + o_p(h_n^2)$, provided $\frac{\ln(n)}{nh_n^3} \to 0$ and $h_n^2\ln(n) \to 0$.

In the case where $\{(X_i, U_i)'\}$ is an iid sequence with $f(x)$ being the marginal density for X_i and $\omega(\theta)$ the variance of U_i, the asymptotic variance is simplified to being $\frac{\omega(\theta)}{f(x)}\int K^2(\phi)\mathrm{d}\phi$. Theorem 2 can therefore be seen as a generalization of the classic asymptotic normality result for local linear estimation under the iid assumption. Examples in Section 4 illustrate the applicability of this general result in regression models where the error covariance has a random effects panel data structure, and an AR(p) structure.

3. Two-step estimation — Asymptotic normality

The estimator $\check{m}(x)$ studied in the previous section has the desirable property of being $\sqrt{nh_n}$-asymptotically normal. However, the fact that none of the information provided by the error covariance structure is used in its construction suggests that alternative estimators can provide improved performance. How to incorporate the covariance structure in defining an alternative estimator has been the subject of various papers (see *inter alia* [27,18], but one promising approach has been a two-step procedure that transforms the model to yield spherical regression errors. The motivation behind the procedure is quite simple. Let $\Omega(\theta_0)$ be an $n \times n$ matrix with the (i, j) element given by $\omega_{ij}(\theta_0)$, $P^{-1}(\theta_0)$ an $n \times n$ matrix with the (i, j) element given by $v_{ij}(\theta_0)$ and $P(\theta_0)$ an $n \times n$ matrix with the (i, j) element given by $p_{ij}(\theta_0)$ such that $\Omega(\theta_0) = P(\theta_0)P(\theta_0)'$. Let $\vec{m}' = (m(X_1), \ldots, m(X_n))$, $U' = (U_1, \ldots, U_n)$, I_n be the identity matrix of size n and define $Z = P^{-1}(\theta_0)\vec{y} + (I_n - P^{-1}(\theta_0))\vec{m}$. Then,

$$Z = \vec{m} + P^{-1}(\theta_0)U = \vec{m} + \varepsilon. \tag{6}$$

Given that the components of the stochastic process $\{U_i\}_{i=1,2,\ldots}$ can be written as $U_i = \sum_{j=1}^{q} p_{ij}\varepsilon_j$ where $q = 1, 2, \ldots, n$, if $\{\varepsilon_i\}_{i=1,2,\ldots}$ is an independent identically distributed process with zero mean and variance σ^2 then the model described in (6) is the standard nonparametric regression model with spherical errors. The difficulty in dealing with such a model stems from the fact that the regressand Z is not observed since $\vec{m}$ and the components of $P^{-1}(\theta_0)$ are generally unknown – since θ_0 is unknown – and must be substituted by suitable estimates. Hence, implementation normally requires a first-stage estimation in which $\check{m}(x)$ and estimators for the elements of $P^{-1}(\theta_0)$, say $P^{-1}(\dot{\theta})$ (normally using residuals $\check{U}_i = Y_i - \check{m}(X_i)$), are obtained, and a second stage in which the regressand $\hat{Z} = P^{-1}(\dot{\theta})\vec{y} + (I_n - P^{-1}(\dot{\theta}))\check{\vec{m}}$ is used in (6). The asymptotic properties of the resulting estimator are not known in general, but Xiao et al. [38] have obtained $\sqrt{nh_n}$-asymptotic normality for a stationary error structure that has an invertible linear process representation $U_t = \sum_{j=0}^{\infty} c_j\varepsilon_{t-j}$. A key feature of their structure is that the diagonal elements of $P^{-1}(\theta_0)$ are all equal to 1, a property that we will see below has important consequences in establishing the asymptotic normality of the estimator. Since this cannot be generally assumed we will propose a slightly different estimator that circumvents the difficulties that we encountered with the estimator for general models.

In what follows we will restrict ourselves to stochastic processes $\{U_i\}_{i=1,2,\ldots}$ that can be constructed from linear transformations of iid processes. Hence, we assume:

Assumption A6. The components of the stochastic process $\{U_i\}_{i=1,2,...}$ can be written as $U_i = \sum_{j=1}^{q} p_{ij}\varepsilon_j$ where $q = 1, 2, \ldots, n$ and $\{\varepsilon_i\}_{i=1,2,...}$ is an independent identically distributed process with zero mean and unit variance.

For economy of notation we also write p_{ij}, v_{ij}, P and P^{-1} where it is well understood that all of these variables depend on θ. Let $H = \text{diag}\{v_{ii}^{-1}\}_{i=1}^{n}$ and define $Z = HP^{-1}\vec{y} + (I_n - HP^{-1})\vec{m}$. Then,

$$Z = \vec{m} + HP^{-1}U = \vec{m} + \gamma. \tag{7}$$

Given Assumption A6, $\{\gamma_i\}_{i=1,2,...}$ is an independent heterogeneous sequence with $E(\gamma) = 0$ and $E(\gamma\gamma') = H^2 = \text{diag}\{v_{ii}^{-2}\}_{i=1}^{n}$.

As above, the regression error γ_i in the transformed regression (7) is independent and heteroscedastic, but the vector of regressands is unknown. If $m(X_i)$ is estimated at a first stage by $\check{m}(X_i)$, then the only source of ignorance about Z is due to P^{-1} or the fact that θ_0 is unknown. In Theorem 3 below we focus on establishing the asymptotic normality of the estimator

$$\hat{m}(x) = e'\left(R_x'K_xR_x\right)^{-1}R_x'K_x\check{Z} \tag{8}$$

where $\check{Z} = HP^{-1}\vec{y} - (HP^{-1} - I_n)\check{m}$, $\check{m}' = (\check{m}(X_1), \ldots, \check{m}(X_n))$ and for the moment we assume that θ_0, and therefore P^{-1} (and consequently H), is known.

Theorem 3. *Let $\{(X_i, U_i)\}_{i=1}^{n}$ be a stochastic sequence of vectors and assume that $Y_i = m(X_i) + U_i$ for $i = 1, 2, \ldots$, $\{X_i\}_{i=1,2,...}$ and $\{U_i\}_{i=1,2,...}$ are independent with $E(U_i) = 0$ for all $i = 1, 2, \ldots$, $E(U_iU_j) = \omega_{ij}(\theta_0)$ $\theta_0 \in \Re^p$, $p < \infty$. Consider the estimator $\hat{m}(x)$ described above, such that h_n is the bandwidth used in the first-stage estimation and g_n is the bandwidth used in the second stage of the estimation. If we assume that Assumptions A1–A6 are met and $E(|U_i|^{2+\theta}) < C$ for some $\theta > 0$ and all i, then,*

$$(ng_n)^{1/2}(\hat{m}(x) - m(x) - B_{n,1}(x)) \xrightarrow{d} N\left(0, \frac{\bar{\omega}_f(x, \theta_0)}{\bar{f}^2(x)}\int K^2(\phi)\mathrm{d}\phi\right) \tag{9}$$

where $B_{n,1}(x) = \frac{g_n^2}{2}\sigma_K^2 m^{(2)}(x) + o_p(g_n^2)$, $\bar{\omega}_f(x, \theta_0) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f_i(x)v_{ii}^{-2}$ provided that: 1. $\frac{h_n}{g_n} \to 0$ and $g_n = O(n^{-1/5})$; 2. $\sup_i \sum_{j=1,j\neq i}^{n}\frac{|v_{ij}|}{|v_{ii}|} = O(1)$ and $\sup_i \sum_{j=1,j\neq i}^{n}\frac{|v_{ji}|}{|v_{jj}|} = O(1)$.

We note that difference between the variances of the asymptotic distributions of $\check{m}(x)$ and $\hat{m}(x)$ is given by

$$\lim_{n\to\infty}\frac{1}{n\bar{f}(x)^2}\sum_{i=1}^{n} f_i(x)\left(\omega_{ii}(\theta_0) - \frac{1}{v_{ii}^2}\right)\int K^2(\phi)\mathrm{d}\phi. \tag{10}$$

By Theorem 12.2.10 in [14] we have that $p_{ii}v_{ii} \geq 1$. Consequently,

$$p_{ii}^2 \geq \frac{1}{v_{ii}^2} \Rightarrow \omega_{ii}(\theta_o) = p_{ii}^2 + \sum_{j=1,j\neq i}^{n} p_{ij}^2 \geq \frac{1}{v_{ii}^2}$$

which establishes that $\hat{m}(x)$ is *efficient* relative to $\check{m}(x)$. The improvement over local linear estimation is obtained even though $\hat{m}(x)$ ignores the heteroscedastic structure of the error.

Notice also that we impose two more assumptions in Theorem 3. The first one relates to undersmoothing in the first-stage regression so that the magnitude of the bias created by $\hat{m}(x)$ will be smaller than the leading bias term in the second stage. This assumption is common in two-stage nonparametric regression estimation, e.g., Assumption 7 in [38], Assumption B5 in [29] and Remark 1 in [34]. The second assumption is essentially uniform summability of the rows of error covariance, which is a sufficient condition used in the proof of Theorem 3 to control the order of magnitude for summation terms showing up in the second stage. Similar assumptions have been used in the literature, i.e., Assumption A.3 in [12] and Assumption 5 in [38].

An important part in the proof of Theorem 3 (Appendix A) is that $\check{Z}_i = m(X_i) - \sum_{j=1,j\neq i}^{n}\frac{v_{ij}}{v_{ii}}(\check{m}(X_j) - m(X_j)) + \gamma_i$. If instead we were considering the estimator $\tilde{m}(x) = e'\left(R_x'K_xR_x\right)^{-1}R_x'K_x\check{Z}$ where $\check{Z} = P^{-1}\vec{y} - (P^{-1} - I_n)\check{m}$, then $\check{Z}_i = m(X_i) + \varepsilon_i - \sum_{j=1}^{n} v_{ij}(\check{m}(X_j) - m(X_j)) + (\check{m}(X_i) - m(X_i))$ and $B_n(x) = \frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n} K\left(\frac{X_i - x}{g_n}\right)\check{Z}_i^*$, which is asymptotically equivalent to $\tilde{m}(x) - m(x)$, would have an extra term given by $\frac{1}{\bar{f}_n(x)}\frac{1}{ng_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{g_n}\right)(\check{m}(X_i) - m(X_i))$ which cannot easily be shown to be $o_p((ng_n)^{-1/2})$ under the general conditions that we consider. By construction, whenever the diagonal elements of P^{-1} are equal to 1 this extra term does not appear even when $\check{Z} = P^{-1}\vec{y} - (P^{-1} - I_n)\check{m}$. Hence, we have the following result which we state as a Corollary to Theorem 3.

Corollary 1. *Let $\{(X_i, U_i)\}_{i=1}^{n}$ be a stochastic sequence of vectors and assume that $Y_i = m(X_i) + U_i$ for $i = 1, 2, \ldots$, $\{X_i\}_{i=1,2,...}$ and $\{U_i\}_{i=1,2,...}$ are independent with $E(U_i) = 0$ for all $i = 1, 2, \ldots$, $E(U_iU_j) = \omega_{ij}(\theta_0)$ $\theta_0 \in \Re^p$, $p < \infty$. Consider the estimator $\tilde{m}(x)$ described above, such that h_n is the bandwidth used in the first-stage estimation and g_n is the bandwidth used in the second stage of the estimation. If we assume that Assumptions A1–A6 are met and $E(|U_i|^{2+\theta}) < C$ for some $\theta > 0$ and all i, then,*

$$(ng_n)^{1/2}(\tilde{m}(x) - m(x) - B_{n,1}(x)) \xrightarrow{d} N\left(0, \frac{1}{\bar{f}(x)}\int K^2(\phi)\mathrm{d}\phi\right) \tag{11}$$

provided that: 1. $\frac{h_n}{g_n} \to 0$ *and* $g_n = O(n^{-1/5})$; 2. $\sup_i \sum_{j=1,j\neq i}^{n} |v_{ij}| = O(1)$ *and* $\sup_i \sum_{j=1,j\neq i}^{n} |v_{ji}| = O(1)$; 3. $P^{-1}(\theta_0)$ *is such that* $v_{ii}(\theta_0) = 1$ *for all i.*

The use of Theorem 3 and its Corollary is restricted in practice due to the fact that the parameter θ used in defining P is generally unknown and must be estimated. Hence, we now turn our attention to a feasible estimator $\dot{m}(x) = e'\left(R_x'K_xR_x\right)^{-1} R_x'K_x\dot{Z}$ where $\dot{Z} = H(\dot{\theta})P^{-1}(\dot{\theta})\vec{y} - (H(\dot{\theta})P^{-1}(\dot{\theta}) - I_n)\check{m}$ and for which $\dot{\theta} - \theta_0 = o_p(1)$. The next theorem provides sufficient conditions under which $\sqrt{ng_n}(\dot{m}(x) - \hat{m}(x)) = o_p(1)$. As such, it gives conditions under which the feasible estimator is asymptotically equivalent to $\hat{m}(x)$, therefore inheriting its desirable properties, namely asymptotic normality and efficiency relative to the LLE. The theorem can be viewed as an extension of the theorem in [20] to the case of nonparametric regression.

Theorem 4. *Suppose that all assumptions in Theorem 3 are holding and assume in addition that:*

TA 4.1: $H(\theta)P^{-1}(\theta)$ has at most $W < \infty$ distinct nonzero elements for every n, denoted by $g_{wn}(\theta)$ for $w = 1, 2, \ldots, W$. That is, there are $n^2 - W$ elements that are either zero or duplicates of other nonzero elements in $H(\theta)P^{-1}(\theta)$. For each w, $g_{wn}(\theta)$ converges uniformly as $n \to \infty$ to a real valued function $g_w(\theta)$ on an open set O containing θ_0, where g_w is continuous at θ_0.

TA 4.2: The number of nonzero elements in each column (and row) of $H(\theta)P^{-1}(\theta)$ is uniformly bounded by $\aleph$ as $n \to \infty$.

TA 4.3: There exists $C < \infty$ such that $\sum_{i=1}^{n} |\omega_{ij}(\theta)| < C$ for every $n = 1, 2, \ldots$ and $j = 1, 2, \ldots$.

If $\dot{\theta} - \theta_0 = o_p(1)$ then we have

$$\sqrt{ng_n}(\hat{m}(x) - \dot{m}(x)) = o_p(1).$$

4. Selected applications

In this section we provide two applications for the results that we have obtained. The first deals with clustered or panel data models. Here, the asymptotic normality result that we obtain for the local linear and the two-stage estimators is novel. The second application is for nonparametric regression models with autoregressive errors of order p, which have been studied by Vilar-Fernández and Francisco-Fernández [32] for the case where $p = 1$ under fixed design regressors. The examples illustrate the applicability of our theorems to popular nonparametric models and reveal the ease of verifying the conditions listed in Theorems 3 and 4.

4.1. Clustered or panel data models

We focus on the regression models for clustered data proposed by Ruckstuhl et al. [25] and also studied by Wang [34]. The model is a direct extension to the nonparametric regression setting of the one-way *random effects* model that is popular in the panel data literature [2]. Consider

$$Y_{ij} = m(X_{ij}) + \alpha_i + \varepsilon_{ij} \quad i = 1, \ldots, N; j = 1, \ldots, J, \tag{12}$$

where $\{\alpha_i\}_{i=1,2,\ldots}$ are independent with $E(\alpha_i) = 0$ and $V(\alpha_i) = \sigma_\alpha^2$ for all i; $\{\varepsilon_{ij}\}_{i,j=1,2,\ldots}$ are independent with $E(\varepsilon_{ij}) = 0$ and $V(\varepsilon_{ij}) = \sigma_\varepsilon^2$ for all i, j and the processes $\{\alpha_i\}_{i=1,2,\ldots}$ and $\{\varepsilon_{ij}\}_{i,j=1,2,\ldots}$ are independent. Ruckstuhl et al. [25] assume that $\{X_i\}_{i=1,2,\ldots}$ where $X_i' = (X_{i1}, \ldots, X_{iJ})$ is an independent and identically distributed vector sequence with the marginal density of X_{ij} given by f_j.

We define $Y_i' = (Y_{i1}, \ldots, Y_{iJ})$, $\vec{y} = (Y_1', \ldots, Y_N')'$, $X_i' = (X_{i1}, \ldots, X_{iJ})$, $\vec{x} = (X_1', \ldots, X_N')'$ and $U_{ij} = \alpha_i + \varepsilon_{ij}$. Then, given the assumptions on α_i and ε_{ij} we have that for $U_i' = (U_{i1}, \ldots, U_{iJ})$, $E(U_iU_i') = \Sigma = \sigma_\varepsilon^2 I_J + \sigma_\alpha^2 1_J 1_J'$ and if $U = (U_1', \ldots, U_N')'$, $E(UU') = I_N \otimes \Sigma = \Omega(\sigma_\varepsilon^2, \sigma_\alpha^2)$. In this context we have that $\check{m}(x) = e'\left(\bar{R}_x'\bar{K}_x\bar{R}_x\right)^{-1} \bar{R}_x'\bar{K}_x\vec{y}$ where $\bar{R}_x = (1_{NJ}, \vec{x} - 1_{NJ}x)$, $\bar{K}_x = \operatorname{diag}\left\{K\left(\frac{X_{ij}-x}{h_n}\right)\right\}_{i=1,j=1}^{N,J}$. Let $n = NJ$; then the LLE estimator can be written as $\check{m}(x) = \frac{1}{nh_n}\sum_{i=1}^{N}\sum_{j=1}^{J} W_n\left(\frac{X_{ij}-x}{h_n}, x\right) Y_{ij}$.

We assume Assumption A1.1–4 and verify that Assumption A1.5–6 hold since $\bar{f}_n(x) = \frac{1}{J}\sum_{j=1}^{J} f_j(x)$ and as assumed in Ruckstuhl et al. [25] if $0 < f_j(x) < C$ we have $0 < \bar{f}_n(x) < B$. Assumption A3 is verified since $0 < \sigma_\alpha^2, \sigma_\varepsilon^2 < C$ and consequently $\frac{1}{n}\sum_{i=1}^{n} \omega_{ii}(\sigma_\alpha^2, \sigma_\varepsilon^2) = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and $\bar{\omega}_f(x, \sigma_\alpha^2, \sigma_\varepsilon^2) = (\sigma_\alpha^2 + \sigma_\varepsilon^2)\bar{f}_n(x)$. Now, since the process $\{X_i\}$ is independent and identically distributed, $\{X_{ij}\}$ is such that $\alpha(t) = 0$ for all $t \geq J$. Similarly, since $\{\alpha_i\}$ is independent and $\{\varepsilon_{ij}\}$ is independent, we have that U_{ij} and $U_{i'j'}$ is independent for all $i \neq i'$ for all j, j' and therefore $\alpha(t) = 0$ for all $t \geq J$, verifying Assumption A4 given the independence of $\{X_i\}$ and $\{U_{ij}\}$. Assumption A6 is easily verified by the independence of $\{\alpha_i\}$ and $\{\varepsilon_{ij}\}$ and noting that $U = Pv$ where v is a vector of iid random variables with $E(v_i) = 0$ and $V(v_i) = 1$. Hence, we conclude that

$$\sqrt{ng_n}\left(\check{m}(x) - m(x) - \left(\sigma_K^2\frac{m^{(2)}(x)}{2}g_n^2 + o_p(g_n^2)\right)\right) \xrightarrow{d} N\left(0, \frac{\sigma_\varepsilon^2 + \sigma_\alpha^2}{\frac{1}{J}\sum_{j=1}^{J} f_j(x)}\int K^2(\phi)\mathrm{d}\phi\right). \tag{13}$$

From [35] we have that $P^{-1}(\sigma_\alpha^2, \sigma_\varepsilon^2) = I_N \otimes V^{-1/2}$ where

$$V^{-1/2} = v_d \begin{pmatrix} 1 & \frac{v_0}{v_d} & \cdots & \frac{v_0}{v_d} \\ \frac{v_0}{v_d} & 1 & \cdots & \frac{v_0}{v_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{v_0}{v_d} & \frac{v_0}{v_d} & \cdots & 1 \end{pmatrix} \tag{14}$$

where $v_d = \frac{1}{\sigma_\varepsilon} - \left(1 - \frac{\sigma_\varepsilon}{\sigma_1}\right)\frac{1}{J\sigma_\varepsilon}$ and $v_0 = -\left(1 - \frac{\sigma_\varepsilon}{\sigma_1}\right)\frac{1}{J\sigma_\varepsilon}$ and $\sigma_1 = \sqrt{J\sigma_\alpha^2 + \sigma_\varepsilon^2}$. Hence, since $0 < \sigma_\alpha^2, \sigma_\varepsilon^2 < C$ and J is finite, we have that the sum of the elements in every row and column of HP^{-1} (excluding the diagonals) is $(J-1)\frac{v_0}{v_d} < C$, which satisfies condition 2 in Theorem 3. TA 4.1 is met with $W = 2$, $g_1(\sigma_\alpha^2, \sigma_\varepsilon^2) = v_0/v_d$ and $g_2(\sigma_\alpha^2, \sigma_\varepsilon^2) = 1$ the uniform convergence is trivial as neither function depends on n and the continuity is easily verified. TA 4.2 is met with $\aleph = J$ and TA 4.3 is met since $\sum_{i=1}^n |\omega_{ij}(\theta_0)| \le J\sigma_\alpha^2 + \sigma_\varepsilon^2$.

Consistent estimators for σ_α^2 and σ_ε^2 are given by $\dot\sigma_\varepsilon^2 = \frac{1}{N(J-1)}\sum_{i=1}^N \sum_{j=1}^J (Y_{ij} - \check m(X_{ij}) - (\bar Y_i - \bar m_i))^2$ and $\dot\sigma_\alpha^2 = \frac{1}{N}\sum_{i=1}^N (\bar Y_i - \bar m_i)^2 - \frac{1}{J}\dot\sigma_\varepsilon^2$, where $\bar Y_i = \frac{1}{J}\sum_{j=1}^J Y_{ij}$ and $\bar m_i = \frac{1}{J}\sum_{j=1}^J \check m(X_{ij})$.Thus, we conclude that

$$\sqrt{ng_n}\left(\dot m(x) - m(x) - \left(\sigma_K^2 \frac{m^{(2)}(x)}{2} g_n^2 + o_p(g_n^2)\right)\right) \xrightarrow{d} N\left(0, \frac{\sigma_\varepsilon^2\left(1 - \frac{1}{J}(1 - \frac{\sigma_\varepsilon}{\sigma_1})\right)^{-2}}{\frac{1}{J}\sum\limits_{j=1}^J f_j(x)} \int K^2(\phi)\mathrm{d}\phi\right). \tag{15}$$

4.2. Nonparametric regression with AR(p) errors

We now consider

$$Y_i = m(X_i) + U_i \quad \text{for } t = 1, \ldots, n \tag{16}$$

where $\{X_i\}$ is independent of $\{U_i\}$, satisfies Assumptions A1, A3 and is α-mixing of size -2. U_i is strictly stationary with $U_i = r_1 U_{i-1} + r_2 U_{i-2} + \cdots + r_p U_{i-p} + v_i$ for $i = 0, \pm 1, \pm 2, \ldots$ where $v_i \sim \text{iid}(0, \sigma^2)$ with probability density function $f_v(x)$. Then $\{U_i\}$ satisfies the relevant portions of Assumption A3. Pham and Tran [24] show that $\{U_i\}$ is α-mixing with $\alpha(j) \to 0$ exponentially as $j \to \infty$, which gives that $\{U_i\}$ is of size $-a$ for all $a \in \Re^+$, therefore satisfying Assumption A4.1. Hence,

$$\sqrt{ng_n}\left(\check m(x) - m(x) - \left(\sigma_K^2 \frac{m^{(2)}(x)}{2} g_n^2 + o_p(g_n^2)\right)\right) \xrightarrow{d} N\left(0, \frac{\gamma(0)}{\bar f} \int K^2(\phi)\mathrm{d}\phi\right) \tag{17}$$

where $\gamma(0)$ is the variance of the AR(p) process.

Following Mandy and Martins-Filho [20] we note that since $0 < \sigma^2 < C$ we can find a $p \times p$ lower triangular matrix A such that

$$AE((u_1, \ldots, u_p)'(u_1, \ldots, u_p))A' = \sigma^2 I_p \quad \text{and}$$

$$P^{-1}(\theta_0) = \left(\begin{array}{ccc|cccc} & A & & 0 & \cdots & \cdots & 0 \\ ---- & ---- & ---- & ---- & ---- & ---- & \vdots \\ -r_p & \cdots & -r_1 & 1 & 0 & \cdots & 0 \\ 0 & -r_p & \cdots & -r_1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & 0 & & -r_p & \cdots & -r_1 \;\; 1 \end{array}\right) \tag{18}$$

where $\theta_0 = (r_1, r_2, \ldots, r_p, \sigma^2)$. Since there are a finite number of bounded nonzero elements in each column and row of $P^{-1}(\theta_0)$, condition 2 in Theorem 3 is automatically met. Also, P^{-1} is a lower triangular matrix where all elements that lie more than p positions away from the main diagonal are zero, verifying TA 4.2 with $\aleph = p + 1$. Also, there are at most $W = p(p+1)/2 + (p+1)$ distinct functions in P^{-1}, all of which are independent of n for $n \ge W$ (implying uniform convergence trivially) and continuous at θ_0 since the operations involved in obtaining A are continuous when $0 < \sigma^2 < C$. This verifies TA 4.1.

To verify TA 4.3 we note that an $AR(p)$ process can be written as a p-dimensional VAR(1) process $e_i = Re_{i-1} + \varepsilon_i$, where $e_i = (U_{i-p+1} \ldots U_i)'$, $\varepsilon_i = (0, \ldots 0, v_i)'$, and

$$R = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ & & \ddots & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \\ r_p & r_{p-1} & \cdots & \cdots & r_2 & r_1 \end{pmatrix}. \tag{19}$$

If the process is strictly stationary then the absolute eigenvalues of R are less than 1, and also $E(e_i e_j') = R^{|i-j|}E(e_t e_t')$ for arbitrary t. From the definition of e_i, the sum $\sum_{i=1}^n |E(U_i U_j)|$ is the lower right element of $\sum_{i=1}^n |E(e_i e_j')|$ where the absolute value is taken elementwise. But,

$$\sum_{i=1}^n |E(e_i e_j')| \le 2\sum_{i=1}^n |E(e_i e_0')| \le 2\left(\sum_{i=0}^n |R^i|\right)|E(e_0 e_0')|$$

and re-writing $|R^i|$ in Jordan canonical form yields

$$\sum_{i=1}^n |E(e_i e_j')| \le 2|J|\left(\sum_{i=0}^n |\Lambda^i|\right)|J^{-1}||E(e_0 e_0')|$$

where Λ is a diagonal matrix involving the eigenvalues of R and J is a fixed matrix depending only on R. Since the absolute eigenvalues are less than one $\sum_{i=0}^\infty |\Lambda_i|$ converges, which verifies TA 4.3.

Consistent estimators $\dot{r}_i$ for r_i, $i = 1, \dots, p$, can be obtained (see [32] by defining residuals $\check{U}_i = Y_i - \check{m}(X_i)$ and performing least squares estimation on the following artificial regression:

$$\check{U}_i = r_1\check{U}_{i-1} + r_2\check{U}_{i-2} + \cdots + r_p\check{U}_{i-p} + \check{v}_i \quad \text{for } i = p+1, p+2, \dots.$$

where $\check{v}_i$ is an arbitrary regression error. Hence, we conclude

$$\sqrt{ng_n}\left(\dot{m}(x) - m(x) - \left(\sigma_K^2 \frac{m^{(2)}(x)}{2} g_n^2 + o_p(g_n^2)\right)\right) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\bar{f}}\int K^2(\phi)\mathrm{d}\phi\right). \tag{20}$$

5. Monte Carlo study

In this section, we perform a Monte Carlo study to implement our two-step estimator, henceforth referred to as 2SLL, and illustrate its finite sample performance. We consider a one-way random effects panel data and an AR(2) parametric covariance structure, under which the asymptotic properties of 2SLL and of LLE are provided in the previous section.

For panel data structure, the data generating process (DGP) is given by (12), where the univariate pseudo-random variable X_{ij} is generated independently from an uniform distribution with support $[-2, 2]$. The pseudo-random variable α_i is independently generated from a normal distribution with zero mean and variance $\sigma_\alpha^2 = 4$, and ϵ_{ij} is independently generated from a standard normal distribution. We investigate three function specifications for $m(x)$: $m_1(x) = \sin(0.75x)$, $m_2(x) = 0.5 + \frac{\exp(-4x)}{1+\exp(-4x)}$ and $m_3(x) = 1 - 0.9\exp(-2x^2)$. $m_1(x)$ was used by Fan [7] to illustrate the advantage of LLE over Nadaraya–Watson and Gasser–Müller estimators, and $m_2(x)$ and $m_3(x)$ were used by Martins-Filho and Yao [21] to model the volatility of financial asset returns. All specifications for $m(\cdot)$ are nonlinear and twice differentiable. We fix $J = 2$, and consider three sample sizes $N = 100$, 150 and 200.

For the AR(2) structure, the DGP is given by (16), where the univariate pseudo-random variable X_i is generated independently from a uniform distribution with support $[-2, 2]$. For the error $U_i = r_1U_{i-1} + r_2U_{i-2} + v_i$, we set $r_1 = 0.5$, $r_2 = -0.4$ and generate the pseudo-random variable v_i independently from a standard normal distribution. It is straightforward to verify that for this choice of parameters $\{U_i\}$ is a stationary process. The same three functional forms for $m(\cdot)$ as were given above are adopted. We consider three sample sizes $n = 100$, 200, and 400.

The implementation of our 2SLL estimator requires the selection of bandwidth sequences h_n and g_n. We select the bandwidth $\hat{g}_n$ using the *rule-of-thumb* data driven plug-in method of Ruppert et al. [26] and let $\hat{h}_n = (NJ)^{-\frac{1}{10}}\hat{g}_n$ in the panel data model and $\hat{h}_n = n^{-\frac{1}{10}}\hat{g}_n$ in the AR(2) model. An Epanechnikov kernel is utilized throughout the simulations. We note that the choice of bandwidth and kernel satisfies the requirements in Theorems 2 and 3.

For comparison purposes, we include in our simulations several estimators proposed in the extant literature. Ullah and Roy [31], Lin and Carroll [18] and Henderson and Ullah [16] consider the panel data model and local linear estimators based on transformed observations to incorporate the information contained in error covariance structure in a specific fashion. Their estimators are defined as

$$\hat{\delta}_i(x) = e'(R_x'W_r(x)R_x)^{-1}R_x'W_r(x)\vec{y}$$

for $i = 1, 2$ and $W_1(x) = (P^{-1})'K_xP^{-1}$ and $W_2(x) = K_x^{-\frac{1}{2}}\Omega^{-1}K_x^{-\frac{1}{2}}$. Essentially, $\hat{\delta}_1(x)$ is a LLE on the transformed observations $K_x^{-\frac{1}{2}}P^{-1}\vec{y}$ and $K_x^{-\frac{1}{2}}P^{-1}\vec{x}$, while $\hat{\delta}_2(x)$ is obtained using transformed observations $P^{-1}K_x^{-\frac{1}{2}}\vec{y}$ and $P^{-1}K_x^{-\frac{1}{2}}\vec{x}$. These are also among the estimators considered by Welsh and Yee [36] in the context of a seemingly unrelated regression (SUR) and vector measurement error (VME) model (see their equation (5)). They focus on $\hat{\delta}_2(x)$ as all other estimators that they consider are generally inconsistent in the presence of nonzero correlation in the panel model.[2] As observed by Lin and Carroll [18], Ruckstuhl et al. [25] and Su and Ullah [30], for the clustered/panel data model of Section 4.1, $\hat{\delta}_i(x)$ cannot achieve asymptotic improvement over LLE, but we include both in our simulation to verify their finite sample performance relative to 2SLL. Henderson and Ullah [16] provide feasible versions of $\hat{\delta}_i(x)$ by estimating the unknowns in Ω consistently. Henceforth, we refer to $\hat{\delta}_i(x)$ as HUi and their feasible versions as FHUi. We note that their estimators for the parameters in the covariance matrix coincide with those provided in Section 4.1. For the panel data structure, we also consider the two-step estimator proposed by Ruckstuhl et al. [25], henceforth referred to as RWC, which is more efficient than the local linear estimator, and follow their suggestion to set $\tau = \sigma_\epsilon$. Note that if we set $\tau = \frac{1}{v_d}$, then RWC coincides with 2SLL. Alternatively, as proposed by Su and Ullah [30], the RWC estimator can be constructed with an optimal τ that minimizes an asymptotic approximation for the mean squared error. In their simulation, the optimal τ is selected via a grid search over the interval $[0, \sigma_\epsilon^2 + \sigma_\alpha^2]$ and is approximately $\frac{1}{v_d}$. Hence, the performance of their estimator is similar to ours.[3] The unknown parameters in Ω are estimated as described in Section 4.1.

For the AR(2) error structure, we consider the two-step estimator proposed in [32], henceforth referred to as VFF. Their estimator is defined for the AR(1) model and they show that under fixed design, VFF outperforms the LLE for finite samples. We consider VFF under a random design with an AR(2) covariance structure, where

$$P^{-1} = \begin{pmatrix} \left(\frac{(1+r_2)(1+r_1-r_2)(1-r_1-r_2)}{1-r_2}\right)^{\frac{1}{2}} & 0 & 0 & \cdots & \cdots & 0 \\ -\frac{r_1\sqrt{1-r_2^2}}{1-r_2} & \sqrt{1-r_2^2} & 0 & 0 & \cdots & 0 \\ -r_2 & -r_1 & 1 & 0 & \cdots & 0 \\ 0 & -r_2 & -r_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & & -r_2 & -r_1 & 1 \end{pmatrix}.$$

Since H in 2SLL is a diagonal matrix with the diagonal element being the reciprocal of that in P^{-1}, we observe that VFF differs from 2SLL only in the treatment of the first two observations; hence the estimators are asymptotically equivalent. Hence, we expect the estimators to have similar finite sample performances, which is confirmed in the Monte Carlo study. Although HUi were initially proposed for a panel data error structure, it is straightforward to adapt to the AR(2) structure. We follow the procedures in Section 4.2 to estimate the unknown parameters in Ω.

In total, for the panel data structure we consider nine estimators: LLE, four infeasible estimators where we utilize the true covariance matrix parameters which are available in the simulation study – HU1, HU2, RWC, 2SLL, and four feasible estimators – FHU1, FHU2, FRWC, and F2SLL, where we attach the letter "F" in front of the acronyms to indicate that the unknown parameters in the covariance matrix are estimated. For the AR(2) error structure we consider nine estimators: LLE, HU1, HU2, VFF, 2SLL, FHU1, FHU2, FVFF and F2SLL. All the estimators, except 2SLL and F2SLL, are implemented with bandwidth $\hat{g}_n$ described previously. For each experiment design, we perform 1000 repetitions, evaluate $m(x)$ at twenty equally spaced points over the support interval for the regressor (X) and obtain the average bias, standard deviation and root mean squared error of each estimator. To avoid evaluation over areas of support where data are sparse, we exclude the lower and upper 5% of the support interval. The results are reported in Tables 1 and 2 (Appendix B) for the panel data error structure and AR(2) structure, respectively.

As the sample size increases, across all experiment designs, all estimators generally perform better in terms of averaged standard deviation, root mean squared error and bias, where exceptions occur in bias, whose magnitude is much smaller. This confirms the asymptotic results in Section 4, and agrees with the consistency of the alternative estimators. In terms of the relative performance measured by standard deviation and root mean squared error, when panel data and infeasible estimators are considered, we observe that 2SLL consistently performs the best, followed closely by the RWC estimator. For all three functional forms considered, we notice that the reduction of standard deviation and root mean squared error by 2SLL and RWC over LLE are well over 15%. These results are consistent with our Theorem 3, as well as Theorem 4 in [25], which suggests that two-step estimation properly accounting for the covariance information can improve upon the classical local linear estimator. LLE carries similar standard deviation and root mean squared error to HU2, but both LLE and HU2 always outperform the HU1 estimator. Hence, HUi estimators do not seem to provide gains in terms of efficiency over LLE, at least under the panel data error specification.

When the AR(2) model is considered, across all specifications for $m(x)$, VFF and 2SLL perform similarly and outperform all the other alternatives. The improvement in efficiency from both estimators against LLE is over 10%. Again this is consistent

[2] See [36, p. 3016].

[3] We have investigated the relative performance of their estimator and our 2SLL proposed here by selecting τ in accordance with their equation (3.2). The simulation results (available upon request) show that across different experiment designs, the average RMSE for their estimator decreases with n, and is smaller than that of LLE but still larger than that of our 2SLL.

with our Theorem 3 as well as the comments above regarding the similarity of the two estimators. In addition, our results indicate that the simulation results in [32] carry through in the case of the DGP that we specify.

For the AR(2) error structure, both HU1 and HU2 estimators outperform the LLE, with HU1 outperforming HU2. The asymptotic distributions for the HU*i* estimators under an AR(p) structure are unknown, but on the basis of our simulations these might be viable alternatives. As we expected, the feasible estimators perform slightly worse than the infeasible estimators, where exceptions occur for the HU*i* estimators under the panel data error structure. We notice that the extra burden in computing the unknown parameter is minimal since the increase in magnitude of average standard deviation and root mean squared error is small. Consequently, the observations regarding the relative performances among alternative estimators are largely maintained as those for their infeasible versions. This observation gives support for our Theorem 4 in that feasible 2SLL, obtained by estimating the unknown parameters of the covariance matrix, is asymptotically equivalent to its infeasible version and outperforms the traditional LLE.

6. Summary

In this paper we provide sufficient conditions for the asymptotic normality of the local linear estimator proposed by Fan [7] in regression models where the regression error has a non-spherical parametric covariance structure and the regressors are dependent and heterogeneously distributed. In this context, it seems natural to define an alternative estimator that incorporates the parametric covariance structure in an attempt to reduce the variance of the asymptotic distribution. We propose a two-step estimator that incorporates the parametric information given by the error covariance and provide sufficient conditions for obtaining its asymptotic distribution. A feasible version of the two-step estimator that substitutes true parameter values with consistent estimators is shown to be $\sqrt{ng_n}$ asymptotically equivalent in probability to the two-step estimator under some easily verified conditions. A Monte Carlo study reveals that the asymptotic results for our estimator are confirmed for finite samples and that our estimator can outperform previously proposed estimators.

Appendix A

Proof of Theorem 1. 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 such that $h_n \to 0$ as $n \to \infty$ where $n \in \{1, 2, 3 \ldots\}$. For any n by the Heine–Borel Theorem there exists a finite collection of sets $\left\{B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)\right\}_{k=1}^{l_n}$ such that $G \subset \cup_{k=1}^{l_n} B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)$ for $x_k \in G$ with $l_n < \left(\frac{n}{h_n^2}\right)^{1/2} r$. The proof has three steps.

(1) We show that

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

(2) Let $s_0^B(x) = (nh_n)^{-1} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) g(U_i)v_i I(|g(U_i)| \le B_n)$ where $B_1 \le B_2 \le \cdots$, such that $\sum_{i=1}^\infty B_i^{-s} < \infty$ for some $s > 0$ and $I(\cdot)$ is the indicator function. We show that

$$\sup_{x\in G} |s_0(x) - s_0^B(x) - E(s_0(x) - s_0^B(x))| = O_{as}(B_n^{1-s}).$$

(3) Let $0 < \Delta < \infty$, $\beta > 2$ and $\varepsilon_n = \left(\frac{nh_n}{\ln(n)}\right)^{-1/2} \Delta$; we show that

$$P\left(\max_{1\le k\le l_n} \left|s_0^B(x_k) - E(s_0^B(x_k))\right| \ge \varepsilon_n\right) = O(B_n^{\beta+1.5} n^{1.25-\beta/2} h_n^{-1.75-\beta/2} (\ln(n))^{0.25+\beta/2}).$$

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

$$\begin{aligned} |s_0(x) - s_0(x_k)| &= \left|\frac{1}{nh_n} \sum_{i=1}^n \left(K\left(\frac{X_i - x}{h_n}\right) - K\left(\frac{X_i - x_k}{h_n}\right)\right) g(U_i)v_i\right| \\ &\le \frac{1}{nh_n} \sum_{i=1}^n C\left|\frac{x_k - x}{h_n}\right| |g(U_i)v_i| \quad \text{by Assumption A2.4.} \\ &\le \frac{1}{h_n^2} C\left(\frac{n}{h_n^2}\right)^{-1/2} \frac{1}{n} \sum_{i=1}^n |g(U_i)v_i| \le C(nh_n^2)^{-1/2} \frac{1}{n} \sum_{i=1}^n |g(U_i)|. \end{aligned}$$

By the measurability of g and Assumption A4, $\{|g(U_i)|\}_{i=1,2,\ldots}$ is α-mixing of size -2. Furthermore, given that $E(|U_i|^{2+\theta}) < C$ for some $\theta > 0$ and all i, we have from McLeish's LLN (see [37], p. 49) that $\frac{1}{n}\sum_{i=1}^n |g(Y_i)| - \frac{1}{n}\sum_{i=1}^n E(|g(Y_i)|) = o_p(1)$ and since $\frac{1}{n}\sum_{i=1}^n E(|g(U_i)|) < C$ we have $|s_0(x) - s_0(x_k)| \le C(nh_n^2)^{-1/2}$ and similarly, $E(|s_0(x) - s_0(x_k)|) \le C(nh_n^2)^{-1/2}$. Combining the two results, $\sup_{x\in G} |s_0(x) - E(s_0(x))| \le \max_{1\le k\le l_n} |s_0(x_k) - E(s(x_k))| + 2C(nh_n^2)^{-1/2}$.

Step 2: $\sup_{x\in G}|s_0(x)-s_0^B(x)-E(s_0(x)-s_0^B(x))|\leq T_1+T_2$, where $T_1=\sup_{x\in G}|s_0(x)-s_0^B(x)|$ and $T_2=\sup_{x\in G}|E(s_0(x)-s_0^B(x))|$. We show that $T_1=o_{as}(1)$ and $T_2=O(B_n^{1-s})$ for $s>0$. $T_1=\sup_{x\in G}\left|(nh_n)^{-1}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)g(U_i)v_iI(|g(U_i)|>B_n)\right|$. By the Borel–Cantelli Lemma, for any $\epsilon>0$ and for all m satisfying $m'<m<n$ we have $P(|g(U_m)|\leq B_n)>1-\epsilon$ and by Chebyshev's Inequality and the increasing nature of the B_i sequence, for $n>N\in\mathfrak{R}$ we have $P(|g(U_i)|<B_n)>1-\epsilon$ for $i<m'$. Hence, for $n>\max\{N,m\}$ we have that for all $i\leq n$, $P(|g(U_i)|<B_n)>1-\epsilon$ and therefore $I(|g(U_i)|>B_n)=0$ with probability 1, which gives $T_1=o_{as}(1)$.

$$\begin{aligned}E(s_0(x)-s_0^B(x)) &= \frac{1}{nh_n}\sum_{i=1}^n\int\int_{|g(U_i)|>B_n}K\left(\frac{X_i-x}{h_n}\right)g(U_i)v_if_{X_i,U_i}(X_i,U_i)\mathrm{d}X_i\mathrm{d}U_i\\ &\leq \frac{C}{n}\sum_{i=1}^n\sup_{x\in G}\int_{|g(U_i)|>B_n}|g(U_i)|f_{X_i,U_i}(x,U_i)\mathrm{d}U_i.\end{aligned}$$

By Hölder's inequality, for $s>1$,

$$\int_{|g(U_i)|>B_n}|g(U_i)|f_{X_i,U_i}(x,U_i)\mathrm{d}U_i\leq\left(\int|g(U_i)|^sf_{X_i,U_i}(x,U_i)\mathrm{d}U_i\right)^{1/s}\left(\int I(|g(U_i)|>B_n)f_{X_i,U_i}(x,U_i)\mathrm{d}U_i\right)^{1-1/s}$$

where the first integral after the inequality is uniformly bounded by assumption and since $f_{X_i|U_i}(x)<C$, we have by Chebyshev's Inequality $\left(\int I(|g(U_i)|>B_n)f_{X_i,U_i}(x,U_i)\mathrm{d}U_i\right)^{1-1/s}\leq C(P(|g(U_i)|>B_n))^{1-1/s}\leq CB_n^{1-s}$. Hence, $T_2=O(B_n^{1-s})$.

Step 3: $P\left(\max_{1\leq k\leq l_n}\left|s_0^B(x_k)-E(s_0^B(x_k))\right|\geq\varepsilon_n\right)\leq\sum_{i=1}^{l_n}P\left(\left|s_0^B(x_k)-E(s_0^B(x_k))\right|\geq\varepsilon_n\right)$ and let $s_0^B(x_k)-E(s_0^B(x_k))=\frac{1}{n}\sum_{i=1}^n Z_i$ where

$$Z_i=\frac{1}{h_n}K\left(\frac{X_i-x_k}{h_n}\right)g(U_i)v_iI(|g(U_i)|\leq B_n)-E\left(\frac{1}{h_n}K\left(\frac{X_i-x_k}{h_n}\right)g(U_i)v_iI(|g(U_i)|\leq B_n)\right).$$

By the uniform bound on v_i, Assumption A2 and $|g(U_i)|I(|g(U_i)|\leq B_n)\leq B_n$ we have that $|Z_i|\leq Ch_n^{-1}B_n$. Let $\|Z_i\|_\infty=\inf\{a:P(Z_i>a)=0\}$; then $\sup_{1\leq i\leq n}\|Z_i\|_\infty\leq C\frac{B_n}{h_n}$. Then, from Theorem 1.3 in [4] we have that for each $q=1,2,\ldots,[n/2]$

$$P\left(\frac{1}{n}\left|\sum_{i=1}^n Z_i\right|>\varepsilon_n\right)\leq 4\exp\left(\frac{-\varepsilon_n^2q}{8v^2(q)}\right)+22\left(1+\frac{4CB_n}{\varepsilon_nh_n}\right)^{1/2}q\alpha\left(\left[\frac{n}{2q}\right]\right)$$

where $v^2(q)=\frac{2}{p^2}\sigma^2(q)+\frac{CB_n\varepsilon_n}{2h_n}$, $p=n/2q$,

$$\sigma^2(q)=\max_{0\leq j\leq 2q-1}E((([jp]+1-jp)Z_{[jp]+1}+Z_{[jp]+2}+\cdots+Z_{[(j+1)p]}+((j+1)p-[(j+1)p])Z_{[(j+1)p+1]})^2)$$

and $[a]$ denotes the integer part of $a\in\mathfrak{R}$. We first note that $\frac{h_n}{p}\sigma^2(q)=O(1)$. To see this note that

$$\sigma^2(q)\leq\max_{0\leq j\leq 2q-1}\left(\sum_{[jp]<i\leq[(j+1)p+1]}E(Z_i^2)+2\sum_{\substack{[jp]+1\leq l\leq[(j+1)p]\\ l<i}}\sum_{[jp]+1<i\leq[(j+1)p+1]}|E(Z_lZ_i)|\right).$$

Given Assumption A4.2 and $E(|g(U_i)|^{2+\theta})<C$ for some $\theta>0$ and all i we have after some simple algebra

$$\sum_{[jp]<i\leq[(j+1)p+1]}E(Z_i^2)\leq O(p/h_n).$$

Using Theorem(3)1 in [6], for $\delta>2$ we have that $|E(Z_iZ_l)|\leq Ch_n^{-2+2/\delta}(\alpha(i-l))^{1-2/\delta}$. Now, for any l such that $[jp]+1\leq l\leq[(j+1)p]$ we have that $\sum_{[jp]+1<i\leq[(j+1)p+1]}|E(Z_lZ_i)|\leq\sum_{i=1}^{p*-1}|E(Z_lZ_{l+i})|+\sum_{i=1}^{p*-1}|E(Z_lZ_{l-i})|$ where $p*=[(j+1)p+1]-[jp]+1$. Letting d_n be a sequence of integers such that $d_nh_n\to 0$ we can write

$$\sum_{i=1}^{p*-1}|E(Z_lZ_{l+i})|=\sum_{i=1}^{d_n}|E(Z_lZ_{l+i})|+\sum_{i=d_n+1}^{p*-1}|E(Z_lZ_{l+i})|=J_1+J_2$$

and it can be easily shown that $J_1=o(h_n^{-1})$ and $J_2=O(h_n^{-1})$. Similarly we obtain $\sum_{i=1}^{p*-1}|E(Z_lZ_{l-i})|=O(h_n^{-1})$. Combining the results on the variance and covariances we have that $\frac{h_n}{p}\sigma^2(q)\leq C$ for n sufficiently large. Hence, we have that $ph_nv^2(q)\leq C+CpB_n\varepsilon_n$ and choosing $p=(B_n\varepsilon_n)^{-1}$ we have that for n sufficiently large $ph_nv^2(q)\leq C$. Then, $4\exp\left(\frac{-\varepsilon_n^2q}{8v^2(q)}\right)\leq 4\exp\left(\frac{-\varepsilon_n^2nh_n}{16C}\right)\leq 4n^{-\frac{\Delta^2}{16C}}$. Now,

$$22\left(1+\frac{4CB_n}{\varepsilon_nh_n}\right)^{1/2}q\alpha\left(\left[\frac{n}{2q}\right]\right)=22\left(\frac{B_n}{\varepsilon_n}\right)^{1/2}h^{-1/2}\left(\frac{h_n\varepsilon_n}{B_n}+4C\right)^{1/2}q\alpha\left(\left[\frac{n}{2q}\right]\right)$$

and since $\frac{h_n \varepsilon_n}{B_n} \to 0$ as $n \to \infty$ we have that for n large enough and by Assumption A4, for $\beta > 2$

$$\begin{aligned} 22\left(1+\frac{4CB_n}{\varepsilon_n h_n}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right) &\le C\left(\frac{B_n}{\varepsilon_n}\right)^{1/2} h_n^{-1/2}\frac{n}{2p}[p]^{-\beta} \\ &\le Cnh_n^{-1/2}B_n^{\beta+1.5}\varepsilon_n^{\beta+0.5}. \end{aligned}$$

Thus, $P\left(\max_{1\le k\le l_n}|s_0^B(x_k)-E(s_0^B(x_k))|\ge\varepsilon_n\right) < \frac{Cn^{1/2}}{h_n}\left(4n^{-\frac{\Delta^2}{16C}}+Cnh_n^{-1/2}B_n^{\beta+1.5}\varepsilon_n^{\beta+0.5}\right)$ and if Δ is chosen such that $\frac{\Delta^2}{16C} > 1$ the first term in the summation to the right of the inequality is negligible and we have that $P\left(\max_{1\le k\le l_n}|s_0^B(x_k)-E(s_0^B(x_k))|\ge\varepsilon_n\right) < CB_n^{\beta+1.5}(\ln(n))^{0.25+\beta/2}n^{1.25-\beta/2}h_n^{-1.75-\beta/2}$ and therefore

$$P\left(\max_{1\le k\le l_n}\left|s_0^B(x_k)-E(s_0^B(x_k))\right|\right) = O(B_n^{\beta+1.5}(\ln(n))^{0.25+\beta/2}n^{1.25-\beta/2}h_n^{-1.75-\beta/2}).$$

Lastly, if $B_n \approx n^{1/s+\theta}$ for $s > 2, \theta > 0$ we have that $\sup_{x\in G}|s_0(x)-s_0^B(x)-E(s_0(x)-s_0^B(x))| = o(n^{-1/2})$ and if $n^{(\theta+1/s)(\beta+1.5)+1.25-\beta/2}h_n^{-1.75-\beta/2}(\ln(n))^{0.25+\beta/2}\to 0$ as $n\to\infty$, then

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

which completes the proof. $\square$

Proof of Theorem 2. Note that $m(x) = \frac{1}{nh_n}\sum_{i=1}^n W_n\left(\frac{X_i-x}{h_n},x\right)(m(x)+m^{(1)}(x)(X_i-x))$ and put $S(x) = \begin{pmatrix}\bar f_n(x) & 0\\ 0 & \sigma_K^2\bar f_n(x)\end{pmatrix}$. Then $\check m(x)-m(x) = \frac{1}{nh_n}\sum_{i=1}^n W_n\left(\frac{X_i-x}{h_n},x\right)Y_i^*$, where $Y_i^* = Y_i - m(x) - m^{(1)}(x)(X_i-x)$. Let $A_n(x) = \frac{1}{h_n}\left(e'\left(S_n(x)^{-1}-S(x)^{-1}\right)^2 e\right)^{1/2}$, $D_n(x) = \check m(x)-m(x)-\frac{1}{nh_n\bar f_n(x)}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)Y_i^*$. Then,

$$\begin{aligned} |D_n(x)| &= \frac{1}{nh_n}\left|e'(S_n^{-1}(x)-S^{-1}(x))\begin{pmatrix}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)Y_i^*\\ \sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)Y_i^*\end{pmatrix}\right| \\ &\le h_nA_n(x)\frac{1}{nh_n}\left(\left|\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)Y_i^*\right| + \left|\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)Y_i^*\right|\right) \end{aligned}$$

by Hölder's Inequality. Under the conditions of Theorem 1 $\sup_{x\in G}\left|s_{n,j}(x)-E(s_{n,j}(x))\right| = o_p(h_n)$ for $j=0,1,2$ provided that $\frac{nh_n^3}{\ln(n)}\to\infty$. Now, $\sup_{x\in G}\left|s_{n,2}(x)-\sigma_K^2\bar f_n(x)\right| \le \sup_{x\in G}|s_{n,2}(x)-E(s_{n,2}(x))| + \sup_{x\in G}\left|E(s_{n,2}(x))-\sigma_K^2\bar f_n(x)\right|$, but

$$\sup_{x\in G}\left|E(s_{n,2}(x))-\sigma_K^2\bar f_n(x)\right| \le \frac{1}{n}\sum_{i=1}^n\int\phi^2K(\phi)|f_i(x+h_n\phi)-f_i(x)|\mathrm{d}\phi \le h_nC\sigma_K^2$$

given Assumptions A1 and A2. Therefore, $\sup_{x\in G}\left|s_{n,2}(x)-\sigma_K^2\bar f_n(x)\right| \le o_p(h_n)+O(h_n) = O_p(h_n)$ and similar arguments give $\sup_{x\in G}\left|s_{n,0}(x)-\bar f_n(x)\right| = O_p(h_n)$ and $\sup_{x\in G}|s_{n,1}(x)| = O_p(h_n)$. As a result, $A_n(x) = O_p(1)$ uniformly in G. We now turn our attention to $B_n(x) = \frac{1}{nh_n\bar f_n(x)}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)Y_i^*$. Since $Y_i^* = m(X_i)-m(x)-m^{(1)}(x)(X_i-x)+U_i$ and K has a bounded support, $Y_i^* = \frac{1}{2}m^{(2)}(x)(X_i-x)^2+U_i+o_p(h_n^2)$ and

$$\begin{aligned} B_n(x) &= \frac{h_n^2}{\bar f_n(x)}\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)\frac{1}{2}m^{(2)}(x)\left(\frac{X_i-x}{h_n}\right)^2 + \frac{1}{\bar f_n(x)}\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)U_i \\ &\quad + o(h_n^2)\frac{1}{\bar f_n(x)}\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right) = B_{n,1}(x)+B_{n,2}(x)+B_{n,3}(x). \end{aligned}$$

We examine each $B_{n,j}(x)$ for $j=1,2,3$ separately.

$$B_{n,3}(x) = \frac{1}{\bar f_n(x)}\left(\left(\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)-\bar f_n(x)\right)+\bar f_n(x)\right)o(h_n^2) \quad \text{and}$$

$$|B_{n,3}(x)| \le \frac{1}{\bar f_n(x)}\left(\left|\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{X_i-x}{h_n}\right)-\bar f_n(x)\right|+\bar f_n(x)\right)o(h_n^2).$$

Since $\bar{f}_n(x) \to \bar{f}(x)$ as $n \to \infty$, $|B_{n,3}(x)| \le (O_p(h_n)+1)o(h_n^2) = o_p(h_n^2)$. Furthermore, if $\inf_{x\in G}|\bar{f}_n(x)| > 0$ as $n \to \infty$, $\sup_{x\in G}|B_{n,3}(x)| = o_p(h_n^2)$. $B_{n,1}(x) = \frac{m^{(2)}(x)h_n^2}{2\bar{f}_n(x)}s_{n,2}(x)$ and therefore by Theorem 1, given that $\inf_{x\in G}|\bar{f}_n(x)| > 0$ as $n \to \infty$,

$$\sup_{x\in G}|B_{n,1}(x) - \frac{h_n^2}{2}\sigma_K^2 m^{(2)}(x)| \le C\frac{h_n^2}{2\inf_{x\in G}\bar{f}_n(x)}\sup_{x\in G}|s_{n,2}(x) - \sigma_K^2\bar{f}_n(x)| = O_p(h_n^3).$$

Hence $B_{n,1}(x) = \frac{h_n^2}{2}\sigma_K^2 m^{(2)}(x) + o_p(h_n^2)$ uniformly in G.

Let $Z_i = \frac{1}{h_n}K\left(\frac{X_i-x}{h_n}\right)U_i$; then $B_{n,2}(x) = \frac{1}{\bar{f}_n(x)}\frac{1}{n}\sum_{i=1}^n Z_i$. Since the processes $\{X_i\}_{i=1}^n$ and $\{U_i\}_{i=1}^n$ are independent and $E(U_i)=0$, $E(Z_i)=0$. Now note that $V(Z_i) = \frac{1}{h_n^2}E\left(K^2\left(\frac{X_i-x}{h_n}\right)\right)E(U_i^2) = \frac{1}{h_n}\omega_{ii}(\theta_0)\int K^2(\phi)f_i(x+h_n\phi)\mathrm{d}\phi$. Since $|\omega_{ii}(\theta_0)| < C$ and $f_i(x) < C$ we have that $h_nV(Z_i) \le C\int K^2(\phi)\mathrm{d}\phi$ and $\sup_i h_nV(Z_i) = O(1)$. We now consider

$$\sum_{j=1,i\neq j}^n|\mathrm{cov}(Z_i,Z_j)| = \sum_{j=1,i\neq j}^n|E(Z_i,Z_j)| \le \sum_{j=1}^n|E(Z_i,Z_{i+j})| + \sum_{j=1}^n|E(Z_i,Z_{i-j})|.$$

First write $\sum_{j=1}^n|E(Z_i,Z_{i+j})| = \sum_{j=1}^{d_n-1}|E(Z_i,Z_{i+j})| + \sum_{j=d_n}^n|E(Z_i,Z_{i+j})| = J_{n,1}+J_{n,2}$, where d_n is a sequence of integers such that $d_n \to \infty$ and $d_nh_n \to 0$. Then,

$$\begin{aligned}
J_{n,1} &= \sum_{j=1}^{d_n-1}\frac{1}{h_n^2}\left|EK\left(\frac{X_i-x}{h_n}\right)K\left(\frac{X_{i+j}-x}{h_n}\right)U_iU_{i+j}\right| \\
&= \sum_{j=1}^{d_n-1}|\omega_{i,i+j}(\theta_0)|\int K(\phi_1)K(\phi_2)f_{i,i+j}(x+h_n\phi_1, x+h_n\phi_2)\mathrm{d}\phi_1\mathrm{d}\phi_2 \\
&\le C\sum_{j=1}^{d_n-1}\left(\int K(\phi_1)\,\mathrm{d}\phi_1\right)^2 = C(d_n-1) \le Cd_n.
\end{aligned}$$

Since $d_nh_n \to 0$ we have that $h_nJ_{n,1} \le Cd_nh_n = o(1)$ and $J_{n,1} = o(h_n^{-1})$. Given that $K(\cdot)$ is measurable we have that Z_i is $\sigma(X_i,U_i)$ measurable, where $\sigma(X_i,U_i)$ is the σ-algebra generated by (X_i,U_i). By Theorem 3(1) in [6] with $p=q=\delta>2$ we have

$$|E(Z_i,Z_{i+j})| \le 8E(|Z_i|^\delta)E(|Z_{i+j}|^\delta)\alpha(\sigma(X_i,U_i),\sigma(X_{i+j},U_{i+j}))^{1-\frac{2}{\delta}}$$

where $\alpha(\sigma(X_i,U_i),\sigma(X_{i+j},U_{i+j})) = \sup_{A\in\sigma(X_i,U_i),B\in\sigma(X_{i+j},U_{i+j})}|P(A\cap B)-P(A)P(B)|$. Now define $\mathcal{F}_{-\infty}^i = \sigma(\ldots,X_{i-1},U_{i-1},X_i,U_i)$, $\mathcal{F}_{i+j}^\infty = \sigma(X_{i+j},U_{i+j},X_{i+j+1},U_{i+j+1},\ldots)$ and $\alpha(j) = \sup_i\alpha(\mathcal{F}_{-\infty}^i,\mathcal{F}_{i+j}^\infty)$. Then, $\alpha(\sigma(X_i,U_i),\sigma(X_{i+j},U_{i+j})) \le \alpha(j)$. Also,

$$\begin{aligned}
E|Z_i|^\delta &= E(|U_i|^\delta)h_n^{-\delta+1}\frac{1}{h_n}E\left(K^\delta\left(\frac{X_i-x}{h_n}\right)\right) \\
&= E(|U_i|^\delta)h_n^{-\delta+1}\int K^\delta(\phi)f_i(x+h_n\phi)\mathrm{d}\phi \\
&\le CE(|U_i|^\delta)h_n^{-\delta+1}\int K^\delta(\phi)\mathrm{d}\phi \quad \text{by Assumption A1} \\
&\le Ch_n^{-\delta+1}.
\end{aligned}$$

Similarly $E|Z_{i+j}|^\delta \le Ch_n^{-\delta+1}$ and we have $|E(Z_i,Z_{i+j})| \le 8(Ch_n^{-\delta+1})^{2/\delta}\alpha(j)^{1-\frac{2}{\delta}} = Ch_n^{-2+\frac{2}{\delta}}\alpha(j)^{1-\frac{2}{\delta}}$. Hence, $J_{n,2} \le Ch_n^{-2+\frac{2}{\delta}}\sum_{j=d_n}^\infty\alpha(j)^{1-\frac{2}{\delta}}$ and since $j \ge d_n$ we have that for some $a > 1-\frac{2}{\delta} > 0$, $\frac{j^a}{d_n^a} \ge 1$ and $J_{n,2} \le Ch_n^{-2+\frac{2}{\delta}}d_n^{-a}\sum_{j=d_n}^\infty j^a\alpha(j)^{1-\frac{2}{\delta}}$. But, $\sum_{j=d_n}^\infty j^a\alpha(j)^{1-\frac{2}{\delta}} \to 0$ by Assumption A4 as $n \to \infty$. Now, $h_n^{\frac{2}{\delta}-1}d_n^{-a} = \left((h_nd_n^{\frac{a\delta}{\delta-2}})^{1-\frac{2}{\delta}}\right)^{-1}$ and choosing d_n such that $h_n^{1-\frac{2}{\delta}}d_n^a = 1$ the right hand side of the last equality is equal to 1 and we have $J_{n,2} = o(h_n^{-1})$. This is obviously consistent with $d_nh_n \to 0$ in the sense that $\frac{a\delta}{\delta-2} > 1 \Rightarrow a > 1-\frac{2}{\delta}$. Furthermore, it is easily seen from the developments above that $\sup_i|J_{n,1}| + \sup_i|J_{n,2}| = o(h_n^{-1})$ and $h_n\sup_i\sum_{j=1}^n|E(Z_iZ_{i+j})| = o(1)$. Similar arguments show that $\sum_{j=1}^n|E(Z_iZ_{i-j})| = o(h_n^{-1})$ and $h_n\sup_i\sum_{j=1}^n|E(Z_iZ_{i+j})| = o(1)$. Hence, combining results we have $\sum_{j=1,i\neq j}^n|\mathrm{cov}(Z_i,Z_j)| = o(h_n^{-1})$ and $\sup_i\sum_{j=1,i\neq j}^n|\mathrm{cov}(Z_i,Z_j)| = o(h_n^{-1})$. Now, observe that $V\left(\frac{1}{n}\sum_{i=1}^n Z_i\right) = \frac{1}{n^2}\sum_{i=1}^n E(Z_i^2) + \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n E(Z_iZ_j) = V_{n,1}+V_{n,2}$.

$$\begin{aligned}
V_{n,1} &= \frac{1}{n^2}\sum_{i=1}^n\frac{1}{h_n}\omega_{ii}(\theta_0)\int K^2(\phi)(f_i(x+h_n\phi)-f_i(x))\mathrm{d}\phi + \frac{1}{n^2}\sum_{i=1}^n\frac{1}{h_n}\omega_{ii}(\theta_0)\int K^2(\phi)f_i(x)\mathrm{d}\phi \\
&= V_{n,1}^1 + V_{n,1}^2.
\end{aligned}$$

By the Lipschitz condition on $f_i(x)$ and Assumption A2, $|V^1_{n,1}| \leq C\frac{1}{n^2}\sum_{i=1}^n \omega_{ii}(\theta_0)$ and therefore $nh_n|V^1_{n,1}| \leq \frac{Ch_n}{n}\sum_{i=1}^n \omega_{ii}(\theta_0)$ and by Assumption A3 we have $nh_n|V^1_{n,1}| = O(h_n)$. Also,

$$nh_nV^2_{n,1} = \int K^2(\phi)\mathrm{d}\phi \frac{1}{n}\sum_{i=1}^n f_i(x)\omega_{ii}(\theta_0) \to \bar{\omega}_f(x,\theta_0)\int K^2(\phi)\mathrm{d}\phi.$$

Hence, $\frac{h_n}{n}\sum_{i=1}^n E(Z_i^2) = \bar{\omega}_f(x,\theta_0)\int K^2(\phi)\mathrm{d}\phi + O(h_n)$. Now,

$$nh_n\left|\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1,i\neq j}^n E(Z_iZ_j)\right| \leq \frac{1}{n}\sum_{i=1}^n h_n \sup_i \sum_{j=1,i\neq j}^n |E(Z_iZ_j)| = o(1)$$

where the last equality follows from our previous results. Hence, we have that

$$V\left(\sqrt{nh_n}\frac{1}{n}\sum_{i=1}^n Z_i\right) = \bar{\omega}_f(x,\theta_0)\int K(\phi)\mathrm{d}\phi + O(h_n) + o(1). \tag{21}$$

We now consider $B_{n,2}(x)$. Here we adopt the method first proposed by Bernstein [3] and adopted by Masry and Fan [22] to partition the sums into large and small blocks. First, partition the set $\{1,\ldots,n\}$ into $2k_n+1$ subsets with *large* blocks of size r_n and *small* blocks of size s_n and $k_n = \left[\frac{n}{r_n+s_n}\right]$. Let $Z_{n,i} = \sqrt{h_n}Z_{i+1}$ for $i = 0, 1, \ldots, n-1$ so that $B_{n,2}(x) = \frac{1}{\bar{f}_n(x)}\frac{1}{n}\sum_{i=1}^n Z_i$ and $\sqrt{nh_n}\frac{1}{n}\sum_{i=1}^n Z_i = \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} Z_{n,i}$. Now let

$$\eta_j = \sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1} Z_{n,i} \quad \text{for } 0\leq j\leq k_n-1$$
$$\xi_j = \sum_{i=j(r_n+s_n)+r_n}^{(j+1)(r_n+s_n)-1} Z_{n,i} \quad \text{for } 0\leq j\leq k_n-1$$
$$\zeta_j = \sum_{i=k_n(r_n+s_n)}^{n-1} Z_{n,i}$$

and write $\sqrt{nh_n}\frac{1}{n}\sum_{i=1}^n Z_i = \frac{1}{\sqrt{n}}\left(\sum_{j=0}^{k_n-1}\eta_j + \sum_{j=0}^{k_n-1}\xi_j + \zeta_j\right) = \frac{1}{\sqrt{n}}(Q'_n + Q''_n + Q'''_n)$. We show that $E\left(\left(\frac{1}{\sqrt{n}}Q''_n\right)^2\right) \to 0$, $E\left(\left(\frac{1}{\sqrt{n}}Q'''_n\right)^2\right) \to 0$; then the asymptotic distribution of $B_{n,2}(x)$ is determined by $\frac{1}{\sqrt{n}}Q'_n$. Note that $E\left(\left(\frac{1}{\sqrt{n}}Q''_n\right)^2\right) = \frac{1}{n}E\left(\left(\sum_{j=0}^{k_n-1}\xi_j\right)^2\right) = \frac{1}{n}\sum_{j=0}^{k_n-1}E\left(\xi_j^2\right) + \frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{l=0,l\neq j}^{k_n-1}E(\xi_j\xi_l)$ and by Assumption A4 there exists $q_n \to \infty$ such that $q_ns_n = o((nh_n)^{1/2})$, $q_n\left(\frac{n}{h_n}\right)^{1/2}\alpha(s_n) = o(1)$. Then defining $r_n = \left[\frac{(nh_n)^{1/2}}{q_n}\right]$ as $n\to\infty$ we have $\frac{s_n}{r_n} = \frac{o((nh_n)^{1/2})/q_n}{[(nh_n)^{1/2}/q_n]} \to 0$, $\frac{r_n}{n} = \left[\frac{(nh_n)^{1/2}}{q_n}\right]\frac{1}{n} \to 0$, $\frac{r_n}{(nh_n)^{1/2}} = \left[\frac{(nh_n)^{1/2}}{q_n}\right]\frac{1}{(nh_n)^{1/2}} \to 0$, $\frac{n}{r_n}\alpha(s_n) = \frac{n\alpha(s_n)}{\left[\frac{(nh_n)^{1/2}}{q_n}\right]} \approx \left(\frac{n}{h_n}\right)^{1/2}q_n\alpha(s_n) \to 0$. Since $\xi_j = \sum_{i=j(r_n+s_n)+r_n}^{(j+1)(r_n+s_n)-1} Z_{n,i}$ we have

$$\frac{1}{n}\sum_{j=0}^{k_n-1}E(\xi_j^2) = \frac{h_n}{n}\left(\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{s_n}E(Z^2_{j(r_n+s_n)+r_n+\theta}) + \sum_{j=0}^{k_n-1}\sum_{\theta=1}^{s_n}\sum_{\delta=1,\delta\neq\theta}^{s_n}E(Z_{j(r_n+s_n)+r_n+\theta}Z_{j(r_n+s_n)+r_n+\delta})\right).$$

But $\frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{s_n}E(Z^2_{j(r_n+s_n)+r_n+\theta}) \leq \frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{s_n}h_n\sup_i E(Z_i^2) \leq C\frac{1}{n}k_ns_n \leq C\frac{s_n}{r_n+s_n} = o(1)$. Also, since $\sup_i\sum_{j=1,i\neq j}^n|\mathrm{cov}(Z_i,Z_j)| = o(h_n^{-1})$,

$$\begin{aligned}
&\left|\frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{s_n}\sum_{\delta=1,\delta\neq\theta}^{s_n}E(Z_{j(r_n+s_n)+r_n+\theta}Z_{j(r_n+s_n)+r_n+\delta})\right| \\
&\quad\leq \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{s_n}\sum_{\delta=1,\delta\neq\theta}^{s_n}|\mathrm{cov}(Z_{j(r_n+s_n)+r_n+\theta}, Z_{j(r_n+s_n)+r_n+\delta})| \\
&\quad\leq \frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{s_n}h_n \sup_{j(r_n+s_n)+r_n+\theta}\sum_{l=1,l\neq j(r_n+s_n)+r_n+\theta}^{n}|\mathrm{cov}(Z_{j(r_n+s_n)+r_n+\theta}, Z_l)| \\
&\quad= o(1)\frac{k_n}{s_n}s_n \leq o(1)\frac{s_n}{r_n+s_n} = o(1)
\end{aligned}$$

and therefore $\frac{1}{n}E\left(\left(\sum_{j=0}^{k_n-1}\xi_j\right)\right)^2 = o(1)$. Now, $\xi_j\xi_l = h_n\sum_{\theta=1}^{s_n}\sum_{\delta=1}^{s_n} Z_{j(r_n+s_n)+r_n+\delta}Z_{l(r_n+s_n)+r_n+\theta}$ and consequently

$$\left|\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{l=0,l\neq j}^{k_n-1}E(\xi_j\xi_l)\right| \le \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{l=0,l\neq j}^{k_n-1}\sum_{\delta=1}^{s_n}\sum_{\theta=1}^{s_n}|E(Z_{j(r_n+s_n)+r_n+\delta}Z_{l(r_n+s_n)+r_n+\theta})|$$

and since $j\neq l$ the distance between the indexes must be greater than r_n as $|j(r_n+s_n)+r_n+\delta-(l(r_n+s_n)+r_n+\theta)| \ge r_n+1 > r_n$. Thus,

$$\begin{aligned}\left|\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{l=0,l\neq j}^{k_n-1}E(\xi_j\xi_l)\right| &\le 2\frac{h_n}{n}\sum_{i=1}^{n-r_n}\sum_{j=i+r_n}^{n}|E(Z_iZ_j)| \le 2\frac{h_n}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}|E(Z_iZ_j)|\\ &= \frac{h_n}{n}\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}|E(Z_iZ_j)| \le \frac{1}{n}\sum_{i=1}^{n}h_n\sup_i\sum_{j=1,j\neq i}^{n}|\mathrm{cov}(Z_i,Z_j)| = o(1).\end{aligned}$$

Combining the results above we have that $E\left(\left(\frac{1}{\sqrt{n}}Q_n''\right)^2\right) = o(1)$. We now turn our attention to the Q_n''' term.

$$\begin{aligned}E\left(\left(\frac{1}{\sqrt{n}}Q_n'''\right)^2\right) &= \frac{1}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}E(Z_{n,i}^2) + \frac{1}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}\sum_{j=k_n(r_n+s_n),i\neq j}^{n-1}E(Z_{n,i}Z_{n,j})\\ &= \frac{h_n}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}E(Z_{i+1}^2) + \frac{h_n}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}\sum_{j=k_n(r_n+s_n),i\neq j}^{n-1}E(Z_{i+1}Z_{j+1}).\end{aligned}$$

Given $\sup_i h_nE(Z_i^2) \le C$ we have that $\frac{h_n}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}E(Z_{i+1}^2) \le \frac{1}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}\sup_i h_nE(Z_i^2) = Cn^{-1}(n-k_n(r_n+s_n)) = o(1)$, since by construction $n-k_n(r_n+s_n) \le r_n+s_n$ and therefore $n^{-1}(n-(r_n+s_n)) \le n^{-1}(r_n+s_n) = o(1)$. Now,

$$\begin{aligned}\frac{h_n}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}\sum_{j=k_n(r_n+s_n),i\neq j}^{n-1}E(Z_{i+1}Z_{j+1}) &\le \frac{1}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}h_n\sum_{j=k_n(r_n+s_n),i\neq j}^{n-1}|\mathrm{cov}(Z_{i+1},Z_{j+1})|\\ &\le \frac{1}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}\sup_i h_n\sum_{j=1,i\neq j}^{n}|\mathrm{cov}(Z_i,Z_j)|\\ &\le o(1)\frac{1}{n}(n-k_n(r_n+s_n)) = o(1)\end{aligned}$$

and by combining the results above we have $E\left(\left(\frac{1}{\sqrt{n}}Q_n'''\right)^2\right) = o(1)$. We now turn our attention to the Q_n' term. $\eta_j = \sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}Z_{n,i}$ for $0\le j\le k_n-1$ and by construction $\eta_j = h_n^{1/2}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}Z_{i+1}$. Now let $\mathcal{F}_i^j$ be the σ-algebra generated by the random variables $\{X_t,U_t : i\le t\le j\}$, i.e., $\mathcal{F}_i^j = \sigma(X_i,U_i,\ldots,X_j,U_j)$ so that η_j is $\mathcal{F}_{j(r_n+s_n)+1}^{j(r_n+s_n)+r_n}$ measurable. Note that $j(r_n+s_n)+1-((j-1)(r_n+s_n)+r_n) = s_n+1$ and if we define $V_j = \exp(it\eta_j)$, by Lemma 1.1 in [33] we have

$$\left|E\left(\prod_{j=0}^{k_n-1}V_j\right)-\prod_{j=0}^{k_n-1}E(V_j)\right| = \left|E\left(\exp(it\sum_{j=0}^{k_n-1}\eta_j)\right)-\prod_{j=0}^{k_n-1}E(\exp(it\eta_j))\right| \le 16(k_n-1)\alpha(s_n+1). \tag{22}$$

$(k_n-1)\alpha(s_n+1) \le \frac{n}{r_n+s_n}\alpha(s_n+1) = \frac{n}{r_n(1+\frac{s_n}{r_n})}\alpha(s_n+1)$ and since by construction $\frac{s_n}{r_n}\to 0$, $\frac{n}{r_n}\alpha(s_n)\to 0$ we have that $16(k_n-1)\alpha(s_n+1)\to 0$. Thus, by Corollary 14.1 in [17], $\{\eta_j\}_{0\le j\le k_n-1}$ forms a sequence which is independent as $n\to\infty$. Now, $\eta_j = h_n^{1/2}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}Z_{i+1}$ and

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{k_n-1}E(\eta_j^2) &= \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\sum_{l=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}E(Z_{i+1}Z_{l+1})\\ &= \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}E(Z_{i+1}^2) + \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\sum_{l=j(r_n+s_n),i\neq l}^{j(r_n+s_n)+r_n-1}E(Z_{i+1}Z_{l+1})\\ &= I_{n,1}+I_{n,2}.\end{aligned}$$

Also,

$$\begin{aligned}|I_{n,2}| &= \left|\frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{r_n}\sum_{\delta=1,\delta\neq\theta}^{r_n}E(Z_{j(r_n+s_n)+\theta}Z_{j(r_n+s_n)+\delta})\right|\\ &\le \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{r_n}\sum_{\delta=1,\delta\neq\theta}^{r_n}|\mathrm{cov}(Z_{j(r_n+s_n)+\theta},Z_{j(r_n+s_n)+\delta})|\end{aligned}$$

$$\le \frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{\theta=1}^{r_n} h_n \sup_{j(r_n+s_n)+\theta} \sum_{l=1, l\neq j(r_n+s_n)+\theta}^{n} |\mathrm{cov}(Z_{j(r_n+s_n)+\theta}, Z_l)|$$
$$= o(1)\frac{k_n r_n}{n} \le o(1)\frac{r_n}{r_n+s_n} = o(1).$$

For the term $I_{n,1}$ note that $E(Z_i^2) = \frac{1}{h_n}\omega_{ii}(\theta_0)\int K^2(\phi)f_i(x+h_n\phi)\mathrm{d}\phi$ and from Taylor's expansion $|f_i(x+h_n\phi) - f_i(x)| \le O(h_n)$. Therefore,

$$I_{n,1} = \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\left(\frac{1}{h_n}\omega_{i+1,i+1}(\theta_0)\int K^2(\phi)(f_{i+1}(x+h_n\phi) - f_{i+1}(x))\mathrm{d}\phi\right.$$
$$\left. + \frac{1}{h_n}\omega_{i+1,i+1}(\theta_0)f_{i+1}(x)\int K^2(\phi)\mathrm{d}\phi\right) = I_{n,11} + I_{n,12}$$

and looking at the last two terms separately we have

$$|I_{n,11}| \le \frac{h_n}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\frac{1}{h_n}\omega_{i+1,i+1}(\theta_0)\int K^2(\phi)|f_{i+1}(x+h_n\phi) - f_{i+1}(x)|\mathrm{d}\phi$$
$$\le O(h_n)\int K^2(\phi)\mathrm{d}\phi\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\omega_{i+1,i+1}(\theta_0)$$

and since $\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\omega_{i+1,i+1}(\theta_0) \le n^{-1}\sum_{i=1}^{n}\omega_{ii}(\theta_0) \to \bar{\omega}(\theta_0)$ as $n\to\infty$ we have that $|I_{n,11}| = O(h_n)$.

$$I_{n,12} = \int K^2(\phi)\mathrm{d}\phi\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\omega_{i+1,i+1}(\theta_0)f_{i+1}(x) = \int K^2(\phi)\mathrm{d}\phi\frac{1}{n}\sum_{i=1}^{n}\omega_{ii}(\theta_0)f_i(x)$$
$$-\left(\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)+r_n}^{(j+1)(r_n+s_n)-1}\omega_{i+1,i+1}(\theta_0)f_{i+1}(x) + \frac{1}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}\omega_{i+1,i+1}(\theta_0)f_{i+1}(x)\right)\int K^2(\phi)\mathrm{d}\phi.$$

Now, $n^{-1}\sum_{i=1}^{n}\omega_{ii}(\theta_0)f_i(x) \to \bar{\omega}_f(x,\theta_0) < \infty$ by Assumption A3 and since $|\omega_{ii}(\theta_0)|, f_i(x) < C$,

$$\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)+r_n}^{(j+1)(r_n+s_n)-1}\omega_{i+1,i+1}(\theta_0)f_{i+1}(x) \le C\frac{s_n}{r_n+s_n} \to 0.$$

Similarly, $\frac{1}{n}\sum_{i=k_n(r_n+s_n)}^{n-1}\omega_{i+1,i+1}(\theta_0)f_{i+1}(x) \to 0$. Combining the above results we have that $I_{n,1} = \bar{\omega}_f(x,\theta_0)\int K^2(\phi)\mathrm{d}\phi + o(1) + O(h_n)$, and given that $I_{n,2} = o(1)$ we conclude that

$$\frac{1}{n}\sum_{j=0}^{k_n-1}E(\eta_j^2) = \bar{\omega}_f(x,\theta_0)\int K^2(\phi)\mathrm{d}\phi + o(1) + O(h_n).$$

Now let $\frac{1}{\sqrt{n}}Q_n' = \sum_{j=0}^{k_n-1}Z_{jn}$ where $Z_{jn} = \frac{1}{(nh_n)^{1/2}}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}K\left(\frac{X_{i+1}-x}{h_n}\right)U_{i+1}$ and $S_n^2 = \sum_{j=0}^{k_n-1}E(Z_{jn} - E(Z_{jn}))^2$, where $S_n^2 = \sum_{j=0}^{k_n-1}\frac{1}{n}E(\eta_j^2) \to \bar{\omega}_f(x,\theta_0)\int K^2(\phi)\mathrm{d}\phi$ as $n\to\infty$. We first observe that if we define $W_n = \frac{1}{S_n}\frac{1}{\sqrt{n}}Q_n'$ and let $\psi_{W_n}(\lambda) = E(\exp(i\lambda W_n))$ be the characteristic function of W_n we have

$$|\psi_{W_n}(\lambda) - \exp(-\lambda^2/2)| \le \left|E\left(\exp\left(\mathrm{i}\lambda\sum_{j=0}^{k_n-1}\frac{1}{n^{1/2}S_n}\eta_j\right)\right) - \prod_{j=0}^{k_n-1}E\left(\exp\left(\mathrm{i}\lambda\frac{1}{n^{1/2}S_n}\eta_j\right)\right)\right|$$
$$+ \left|\prod_{j=0}^{k_n-1}E\left(\exp\left(i\lambda\frac{1}{n^{1/2}S_n}\eta_j\right) - \exp(-\lambda^2/2)\right)\right| = A_1 + A_2.$$

But $A_1 = o(1)$ by the result on Eq. (22) and $A_2 = o(1)$ by Lindeberg's CLT (Theorem 23.6 in [5]), which is implied by Lyapunov's condition. Hence,

$$\sum_{j=0}^{k_n-1}\frac{Z_{jn}}{S_n} \xrightarrow{d} N(0,1) \quad \text{as } n\to\infty \text{ provided that } \lim_{n\to\infty}\sum_{j=0}^{k_n-1}E\left|\frac{Z_{jn}}{S_n}\right|^{2+\delta} = 0 \text{ for some } \delta > 0.$$

$$\sum_{j=0}^{k_n-1}E\left|\frac{Z_{jn}}{S_n}\right|^{2+\delta} = (S_n^2)^{-1-\delta/2}(nh_n)^{-\delta/2}\frac{1}{nh_n}\sum_{j=0}^{k_n-1}E\left|\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}K\left(\frac{X_{i+1}-x}{h_n}\right)U_{i+1}\right|^{2+\delta}$$
$$\le (S_n^2)^{-1-\delta/2}(nh_n)^{-\delta/2}2^{1+\delta}\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\frac{1}{h_n}E\left|K\left(\frac{X_{i+1}-x}{h_n}\right)U_{i+1}\right|^{2+\delta}$$

by the c_r inequality. Furthermore, $\frac{1}{h_n}E\left|K\left(\frac{X_{i+1}-x}{h_n}\right)U_{i+1}\right|^{2+\delta} = \frac{1}{h_n}E\left(K^{2+\delta}\left(\frac{X_{i+1}-x}{h_n}\right)\right)E\left|U_{i+1}\right|^{2+\delta}$ and given that $E\left|U_{i+1}\right|^{2+\delta} < C$ we have that

$$\frac{1}{h_n}E\left|K\left(\frac{X_{i+1}-x}{h_n}\right)U_{i+1}\right|^{2+\delta} \leq C\int K^{2+\delta}(\phi)f_{i+1}(x+h_n\phi)\mathrm{d}\phi < C$$

by Assumption A2. Therefore,

$$\frac{1}{n}\sum_{j=0}^{k_n-1}\sum_{i=j(r_n+s_n)}^{j(r_n+s_n)+r_n-1}\frac{1}{h_n}E\left|K\left(\frac{X_{i+1}-x}{h_n}\right)U_{i+1}\right|^{2+\delta} \leq C\frac{r_n}{r_n+s_n} \to C$$

and since $S_n^2 \to \bar{\omega}_f(\theta_0, x)\int K^2(\phi)d\phi$ as $nh_n \to \infty$ we have $\lim_{n\to\infty}\sum_{j=0}^{k_n-1}E\left|\frac{Z_{jn}}{S_n}\right|^{2+\delta} = 0$.

Finally, combining the results of $\frac{Q_n'}{\sqrt{n}}$, $\frac{Q_n''}{\sqrt{n}}$ and $\frac{Q_n'''}{\sqrt{n}}$ we conclude that $(nh_n)^{1/2}B_{n,2}(x) \xrightarrow{d} N(0, \frac{\bar{\omega}_f(x,\theta_0)}{\bar{f}(x)^2}\int K^2(\phi)\mathrm{d}\phi)$ as $n\to\infty$. Combining with $B_{n,1}(x) = \frac{h_n^2}{2}\sigma_K^2 m^{(2)}(x) + o_p(h_n^2)$ gives

$$\left(\frac{1}{(nh_n)^{1/2}\bar{f}_n(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)Y_i^* - B_{n,1}(x)\right) \xrightarrow{d} N\left(0, \frac{\bar{\omega}_f(x,\theta_0)}{\bar{f}(x)^2}\int K^2(\phi)\mathrm{d}\phi\right) \quad \text{as } n\to\infty.$$

Now, we note from our previous results on $B_{n,1}(x)$, $B_{n,3}(x)$ and by applying Theorem 1 to $\bar{f}_n(x)B_{n,2}(x)$ with $g(U_i) = U_i$, $j = 0$ and $v_i = 1$ for all i, that we have $\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)Y_i^* = O_p(h_n^2) + O_p\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}\right)$ and $\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h_n}\right)\left(\frac{X_i-x}{h_n}\right)Y_i^* = O_p(h_n^2) + O_p\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}\right)$ uniformly in G. Hence,

$$(nh_n)^{1/2}|D_n(x)| \leq (nh_n)^{1/2}O_p(h_n^3) + (nh_n)^{1/2}O_p\left(\left(\frac{h_n\ln(n)}{n}\right)^{1/2}\right).$$

Now, provided that $h_n^2\ln(n) = o(1)$ the right hand side of the inequality is $o(1)$ and we have

$$(nh_n)^{1/2}\left(\check{m}(x) - m(x) - B_{n,1}(x)\right) \xrightarrow{d} N\left(0, \frac{\bar{\omega}_f(x,\theta_0)}{\bar{f}(x)^2}\int K^2(\phi)\mathrm{d}\phi\right) \quad \text{as } n\to\infty. \quad \square$$

Proof of Theorem 3. Let $\check{Z}_i$ be the ith component of the vector $\check{Z}$. Note that $\hat{m}(x) - m(x) = \frac{1}{ng_n}\sum_{i=1}^{n}W_n\left(\frac{X_i-x}{g_n}, x\right)\check{Z}_i^*$, where $\check{Z}_i^* = \check{Z}_i - m(x) - m^{(1)}(x)(X_i - x)$. Let $A_n(x) = \frac{1}{g_n}\left(e'\left(S_n(x)^{-1} - S(x)^{-1}\right)^2 e\right)^{1/2}$, $D_n(x) = \hat{m}(x) - m(x) - \frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\check{Z}_i^*$. As in Theorem 1

$$\begin{aligned}|D_n(x)| &= \frac{1}{nh_n}\left|e'(S_n^{-1}(x) - S^{-1}(x))\begin{pmatrix}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\check{Z}_i^* \\ \sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\left(\frac{X_i-x}{g_n}\right)\check{Z}_i^*\end{pmatrix}\right| \\ &\leq g_nA_n(x)\frac{1}{ng_n}\left(\left|\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\check{Z}_i^*\right| + \left|\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\left(\frac{X_i-x}{g_n}\right)\check{Z}_i^*\right|\right)\end{aligned}$$

and $A_n(x) = O_p(1)$ uniformly in G. We now turn our attention to $B_n(x) = \frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\check{Z}_i^*$. Since $\check{Z}_i = m(X_i) - \sum_{j=1,j\neq i}^{n}\frac{v_{ij}}{v_{ii}}(\check{m}(X_j) - m(X_j)) + \gamma_i$ we have

$$\begin{aligned}B_n(x) &= \frac{1}{\bar{f}_n(x)}\frac{1}{ng_n}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\frac{m^{(2)}(x)}{2}(X_i-x)^2 + \frac{1}{\bar{f}_n(x)}\frac{1}{ng_n}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\gamma_i \\ &\quad + o(g_n^2)\frac{1}{\bar{f}_n(x)}\frac{1}{ng_n}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right) - \frac{1}{\bar{f}_n(x)}\frac{1}{ng_n}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{v_{ij}}{v_{ii}}(\check{m}(X_j) - m(X_j)) \\ &= B_{n,1}(x) + B_{n,2}(x) + B_{n,3}(x) - B_{n,4}(x).\end{aligned}$$

We examine each $B_{n,j}(x)$ for $j = 1, 2, 3, 4$ separately. From Theorem 2 $B_{n,1}(x) = \frac{g_n^2}{2}\sigma_K^2m^{(2)}(x) + o_p(g_n^2)$, $B_{n,3}(x) = o_p(g_n^2)$ uniformly in G. Also, from Theorem 2, $(ng_n)^{1/2}B_{n,2}(x) \to N\left(0, \frac{\bar{\omega}_f(x,\theta_0)}{\bar{f}(x)^2}\int K^2(\phi)\mathrm{d}\phi\right)$ where $\bar{\omega}_f(x,\theta_0) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f_i(x)v_{ii}^{-2}$. We now examine $B_{n,4}(x)$. From the definition of Y_i^* and Theorem 2

$$\begin{aligned}\check{m}(X_j)-m(X_j) &= \frac{1}{nh_n\bar{f}_n(X_j)}\sum_{l=1}^{n}K\left(\frac{X_l-X_j}{h_n}\right)(m(X_l)-m(X_j)-m^{(1)}(X_j)(X_l-X_j))\\ &\quad+\frac{1}{nh_n\bar{f}_n(X_j)}\sum_{l=1}^{n}K\left(\frac{X_l-X_j}{h_n}\right)U_l+O_p(h_n^3)+O_p\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}h_n\right)\end{aligned}$$

and therefore we can write $B_{n,4}(x)=B_{n,41}(x)+B_{n,42}(x)+B_{n,43}(x)$ where

$$B_{n,41}(x)=\frac{1}{n^2g_nh_n\bar{f}_n(x)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{l=1}^{n}\frac{v_{ij}}{v_{ii}}\frac{1}{\bar{f}_n(X_j)}K\left(\frac{X_i-x}{g_n}\right)K\left(\frac{X_l-X_j}{h_n}\right)(m(X_l)-m(X_j)-m^{(1)}(X_j)(X_l-X_j))$$

$$B_{n,42}(x)=\frac{1}{n^2g_nh_n\bar{f}_n(x)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{l=1}^{n}\frac{v_{ij}}{v_{ii}}\frac{1}{\bar{f}_n(X_j)}K\left(\frac{X_i-x}{g_n}\right)K\left(\frac{X_l-X_j}{h_n}\right)U_l$$

$$B_{n,43}(x)=\frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{v_{ij}}{v_{ii}}K\left(\frac{X_i-x}{g_n}\right)\left(O_p(h_n^3)+O_p\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}h_n\right)\right).$$

We look at each of these terms separately. Note that

$$B_{n,41}(x)=\frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{v_{ij}}{v_{ii}}K\left(\frac{X_i-x}{g_n}\right)\left\{\frac{1}{nh_n\bar{f}_n(X_j)}\sum_{l=1}^{n}K\left(\frac{X_l-X_j}{h_n}\right)(m(X_l)-m(X_j)-m^{(1)}(X_j)(X_l-X_j))\right\}$$

and the term inside the curly brackets $\{\cdot\}$ is $O_p(h_n^2)$ uniformly in G from Theorem 2. Hence,

$$\begin{aligned}|B_{n,41}(x)| &\leq O_p(h_n^2)\frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{|v_{ij}|}{|v_{ii}|}K\left(\frac{X_i-x}{g_n}\right)\\ &\leq O_p(h_n^2)\frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\sup_i\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{|v_{ij}|}{|v_{ii}|}\\ &\leq O_p(h_n^2)O(1)\frac{1}{ng_n\bar{f}_n(x)}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)\end{aligned}$$

where $\sup_i\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{|v_{ij}|}{|v_{ii}|}=O(1)$ by assumption. Furthermore, from Theorem 1 $\frac{1}{ng_n}\sum_{i=1}^{n}K\left(\frac{X_i-x}{g_n}\right)=O_p(1)$ and by Assumption A1 $\bar{f}_n(x)\to\bar{f}(x)$. Hence, $\sup_{x\in G}|B_{n,41}(x)|=O_p(h_n^2)$. Using similar arguments and Theorem 2 we have $\sup_{x\in G}|B_{n,43}(x)|=O_p(h_n^3)+O_p\left(\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}h_n\right)$.

$$\begin{aligned}B_{n,42}(x) &= \frac{1}{n\bar{f}_n(x)}\sum_{l=1}^{n}U_l\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{ng_nh_n\bar{f}_n(X_j)}\sum_{i=1}^{n}\frac{v_{ij}}{v_{ii}}K\left(\frac{X_i-x}{g_n}\right)K\left(\frac{X_l-X_j}{h_n}\right)\\ &= \frac{1}{n\bar{f}_n(x)}\sum_{l=1}^{n}U_l\lambda_{ln}(x).\end{aligned}$$

Note that $E(B_{n,42}(x))=0$ and

$$\begin{aligned}V((ng_n)^{1/2}B_{n,42}(x)) &= \frac{g_n}{n\bar{f}_n(x)^2}\sum_{l=1}^{n}\sum_{k=1}^{n}E(U_lU_k\lambda_{ln}(x)\lambda_{kn}(x))\\ &\leq \frac{g_n}{n\bar{f}_n(x)^2}\sum_{l=1}^{n}\sum_{k=1}^{n}|\omega_{lk}(\theta_0)||E(\lambda_{ln}(x)\lambda_{kn}(x))|.\end{aligned}$$

We define $a_{ij}=\frac{v_{ij}}{v_{ii}}$, $K_i=K\left(\frac{X_i-x}{g_n}\right)$, $K_{lj}=K\left(\frac{X_l-X_j}{h_n}\right)$ and examine

$$\begin{aligned}|E(\lambda_{ln}(x)\lambda_{kn}(x))| &= E\left|\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{m=1}^{n}\sum_{\substack{o=1\\ o\neq m}}^{n}\frac{1}{n^2g_n^2h_n^2\bar{f}_n(X_j)\bar{f}_n(X_o)}a_{ij}a_{mo}K_iK_mK_{lj}K_{ko}\right|\\ &\leq \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{m=1}^{n}\sum_{\substack{o=1\\ o\neq m}}^{n}\frac{1}{n^2g_n^2h_n^2}|a_{ij}||a_{mo}|E\left(\frac{K_iK_mK_{lj}K_{ko}}{\bar{f}_n(X_j)\bar{f}_n(X_o)}\right).\end{aligned}$$

Since $\inf_{x\in G}|\bar{f}_n(x)| > 0$ we have

$$
\begin{aligned}
V((ng_n)^{1/2}B_{n,42}(x)) &\le \frac{Cg_n}{n\bar{f}_n(x)^2}\sum_{l=1}^{n}\sum_{k=1}^{n}|\omega_{lk}(\theta_0)|\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{m=1}^{n}\sum_{\substack{o=1\\ o\neq m}}^{n}\frac{|a_{ij}||a_{mo}|}{n^2g_n^2h_n^2}E\left(K_iK_mK_{lj}K_{ko}\right)\\
&= \frac{Cg_n}{n\bar{f}_n(x)^2}\sum_{l=1}^{n}|\omega_{ll}(\theta_0)|\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{m=1}^{n}\sum_{\substack{o=1\\ o\neq m}}^{n}\frac{|a_{ij}||a_{mo}|}{n^2g_n^2h_n^2}E\left(K_iK_mK_{lj}K_{lo}\right)\\
&\quad+\frac{Cg_n}{n\bar{f}_n(x)^2}\sum_{l=1}^{n}\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{m=1}^{n}\sum_{\substack{o=1\\ o\neq m}}^{n}\frac{|a_{ij}||a_{mo}|}{n^2g_n^2h_n^2}E\left(K_iK_mK_{lj}K_{ko}\right)\\
&= T_{1n}+T_{2n}.
\end{aligned}
$$

We need to show that $T_{1n}, T_{2n} = o(1)$. The strategy that we use is to establish the order of the partial sums that emerge from considering all possible combinations of the indexes l, k, i, j, m, o in T_{1n}, T_{2n}.[4] Each of these partial sums is shown to be $o_p(1)$ by first establishing the order of $\pi_n = \frac{1}{h_n^2g_n^2}E(K_iK_mK_{lj}K_{lo})$ and $\rho_n = \frac{1}{h_n^2g_n^2}E(K_iK_mK_{lj}K_{ko})$. Here we show the cases in which l and k are distinct from the indexes in the four inner sums, i.e., i, j, m, o.[5] We need to consider seven cases, and given Assumption A1 we have from calculating the expectations the following bounds: Case 1 ($i = m$ and $j = o$): $\pi_n \le \frac{C}{g_nh_n}$, $\rho_n \le \frac{C}{g_n}$; Case 2 ($i = o$ and $j = m$) $\pi_n \le \frac{C}{h_n}$, $\rho_n \le C$; Case 3 ($i = m$): $\pi_n \le \frac{C}{g_n}$, $\rho_n \le \frac{C}{g_n}$; Case 4 ($i = o$), Case 5 ($j = m$), Case 7 ($i \neq j \neq m \neq o$): $\pi_n \le C$, $\rho_n \le C$; Case 6 ($j = o$): $\pi_n \le \frac{C}{h_n}$, $\rho_n \le C$. We now denote the partial sums associated with $V((ng_n)^{1/2}B_{n,42}(x))$ in each of these cases by $s_i, i = 1, \dots, 7$. Hence, we have the following inequalities, where the first term refers to the partial sums in T_{1n} and the second term refers to the partial sums in T_{2n} for each case:

$$
\begin{aligned}
s_1 &\le \frac{C\bar{\omega}_n}{h_n\bar{f}_n^2(x)}\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|^2\right)+\frac{C}{ng_n\bar{f}_n^2(x)}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|^2\right)\\
s_2 &\le \frac{C\bar{\omega}_n}{h_n\bar{f}_n^2(x)}g_n\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ji}||a_{ij}|\right)+\frac{C}{n\bar{f}_n^2(x)}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ji}||a_{ij}|\right)\\
s_3 &\le \frac{C\bar{\omega}_n}{\bar{f}_n^2(x)}\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{o=1\\ o\neq i\neq j}}^{n}|a_{io}|\right)+\frac{C}{ng_n\bar{f}_n^2(x)}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{o=1\\ o\neq i\neq j}}^{n}|a_{io}|\right)\\
s_4 &\le \frac{C\bar{\omega}_ng_n}{\bar{f}_n^2(x)}\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{m=1\\ m\neq i\neq j}}^{n}|a_{mi}|\right)+\frac{C}{n\bar{f}_n^2(x)}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{m=1\\ m\neq i\neq j}}^{n}|a_{mi}|\right)\\
s_5 &\le \frac{C\bar{\omega}_ng_n}{\bar{f}_n^2(x)}\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{o=1\\ o\neq i\neq j}}^{n}|a_{jo}|\right)+\frac{C}{n\bar{f}_n^2(x)}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{o=1\\ o\neq i\neq j}}^{n}|a_{jo}|\right)\\
s_6 &\le \frac{C\bar{\omega}_ng_n}{nh_n\bar{f}_n^2(x)}\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{m=1\\ m\neq i\neq j}}^{n}|a_{mj}|\right)+\frac{C}{n\bar{f}_n^2(x)}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\left(n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\sum_{\substack{m=1\\ m\neq i\neq j}}^{n}|a_{mj}|\right)\\
s_7 &\le \frac{C\bar{\omega}_ng_n}{\bar{f}_n^2(x)}\left(n^{-2}\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}|a_{ij}|\right)^2\right)+\frac{C}{n\bar{f}_n^2(x)}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\neq l}}^{n}|\omega_{lk}(\theta_0)|\left(n^{-2}\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}|a_{ij}|\right)^2\right).
\end{aligned}
$$

By Assumptions A1.6 and A3 we have that $\frac{1}{n}\sum_{l=1}^{n}\omega_{ll}(\theta_0) \to \bar{\omega}(\theta_0)$ and $\inf_{x\in G}|\bar{f}_n(x)| > 0$. Furthermore, we note that from Theorem 1 $g_n\sum_{k=1,l\neq k}^{n}|\omega_{lk}(\theta_0)| = o(1)$ and consequently, provided that $\sup_i\sum_{j=1,j\neq i}\frac{|v_{ij}|}{|v_{ii}|} = O(1)$ and $\sup_i\sum_{j=1,j\neq i}\frac{|v_{ji}|}{|v_{jj}|} = O(1)$ the first term and second term in each case are $o(1)$.

Therefore, $B_{n,42}(x) = o_p((ng_n)^{-1/2})$ and $B_{n,4}(x) = O_p(h_n^2) + o_p((ng_n)^{-1/2}) + O_p\left(\left(\frac{h_n}{n}\ln(n)\right)^{1/2}\right)$. Now, provided that $\frac{h_n}{g_n} \to 0$ and $\frac{ng_n^3}{\ln(n)} \to \infty$ we have that the last term is $o(g_n^2)$ and we obtain $B_{n,4}(x) = O_p(h_n^2)+o_p((ng_n)^{-1/2})+o_p(g_n^2)$. Now, if $g_n = O(n^{-1/5})$ then $(ng_n)^{1/2}B_{n,3} = o_p(1)$ and consequently we have

[4] See the note on indexes at the end of this Appendix A.

[5] Bounds for all other cases described in Appendix A are available from the authors upon request.

$$\sqrt{ng_n}\left(B_n(x)-\left(\sigma_K^2\frac{m^{(2)}(x)}{2}g_n^2+o_p(g_n^2)\right)\right)\xrightarrow{d}N\left(0,\frac{\bar{\omega}_f(x,\theta_0)}{\bar{f}^2(x)}\int K^2(\phi)\mathrm{d}\phi\right). \tag{23}$$

Lastly, it follows from arguments similar to those in the proof of Theorem 2 that

$$\sqrt{ng_n}\left(\hat{m}(x)-m(x)-\left(\sigma_K^2\frac{m^{(2)}(x)}{2}g_n^2+o_p(g_n^2)\right)\right)\xrightarrow{d}N\left(0,\frac{\bar{\omega}_f(x,\theta_0)}{\bar{f}^2(x)}\int K^2(\phi)\mathrm{d}\phi\right) \tag{24}$$

which proves the theorem. □

Proof of Theorem 4. $\sqrt{ng_n}(\hat{m}(x)-\dot{m}(x))=e'S_n^{-1}\begin{pmatrix}\frac{1}{\sqrt{ngn}}\sum_{i=1}^n K\left(\frac{X_i-x}{gn}\right)q_i\\ \frac{1}{\sqrt{ngn}}\sum_{i=1}^n K\left(\frac{X_i-x}{gn}\right)\left(\frac{X_i-x}{gn}\right)q_i\end{pmatrix}$ where $q_i=\sum_{j=1,j\neq i}^n(a_{ij}(\dot{\theta})-a_{ij}(\theta_0))(\check{m}(X_j)-m(X_j)-U_j)$ and since $S_n^{-1}(x)=O_p(1)$ and K has compact support, it suffices to show that $\frac{1}{\sqrt{ng_n}}\sum_{i=1}^n K\left(\frac{X_i-x}{g_n}\right)q_i=o_p(1)$. Hence, we must show that

$$\alpha_n=\frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K\left(\frac{X_i-x}{g_n}\right)(a_{ij}(\dot{\theta})-a_{ij}(\theta_0))U_j=o_p(1) \tag{25}$$

and

$$\beta_n=\frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n K\left(\frac{X_i-x}{g_n}\right)(a_{ij}(\dot{\theta})-a_{ij}(\theta_0))(\check{m}(X_j)-m(X_j))=o_p(1). \tag{26}$$

Let $g_0(\theta)=0$ and $I_{iwn}=\{j=1,2,\ldots,n:a_{ij}(\theta)=g_{wn}(\theta)\}$. Then,

$$\begin{aligned}\alpha_n &= \frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\left(\sum_{w=1}^W\sum_{\substack{j\in I_{iwn}\\ j\neq i}}^n K\left(\frac{X_i-x}{g_n}\right)(a_{ij}(\dot{\theta})-a_{ij}(\theta_0))U_j\right.\\ &\quad\left.+\sum_{j\notin\cup_{w=1}^W I_{iwn},j\neq i}^n K\left(\frac{X_i-x}{g_n}\right)(a_{ij}(\dot{\theta})-a_{ij}(\theta_0))U_j\right)\\ &= \frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\sum_{w=1}^W\sum_{\substack{j\in I_{iwn}\\ j\neq i}}^n K\left(\frac{X_i-x}{g_n}\right)(g_{wn}(\dot{\theta})-g_{wn}(\theta_0))U_j\\ &\quad+\frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\sum_{j\notin\cup_{w=1}^W I_{iwn},j\neq i}^n K\left(\frac{X_i-x}{g_n}\right)(g_0(\dot{\theta})-g_0(\theta_0))U_j\\ &= \frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\sum_{w=1}^W\sum_{\substack{j\in I_{iwn}\\ j\neq i}}^n K\left(\frac{X_i-x}{g_n}\right)(g_{wn}(\dot{\theta})-g_{wn}(\theta_0))U_j\\ &= \sum_{w=1}^W(g_{wn}(\dot{\theta})-g_{wn}(\theta_0))\frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\sum_{\substack{j\in I_{iwn}\\ j\neq i}}^n K\left(\frac{X_i-x}{g_n}\right)U_j.\end{aligned}$$

But given TA 4.1, the consistency of $\dot{\theta}$ and the fact that W is finite and does not depend on n, it suffices to show that $\alpha_{n1}=\frac{1}{\sqrt{ng_n}}\sum_{i=1}^n\sum_{j\in I_{iwn},j\neq i}^n K\left(\frac{X_i-x}{g_n}\right)U_j=O_p(1)$ for arbitrary w. Given the independence of $\{X_i\}$ and $\{U_i\}$ and taking expectation of the square yields

$$\begin{aligned}E(\alpha_{n1}^2) &= \frac{1}{ng_n}\sum_{i=1}^n E\left(K^2\left(\frac{X_i-x}{g_n}\right)\right)E\left(\left(\sum_{\substack{\tau\in I_{iwn}\\ \tau\neq i}}U_\tau\right)^2\right)\\ &\quad+\frac{1}{ng_n}\sum_{i=1}^n\sum_{\substack{\tau\in I_{iwn}\\ \tau\neq i}}\sum_{\substack{j=1\\ j\neq i}}^n\sum_{\substack{t\in I_{jwn}\\ t\neq j}}E\left(K\left(\frac{X_i-x}{g_n}\right)K\left(\frac{X_j-x}{g_n}\right)\right)E(U_tU_\tau)\\ &\le \frac{C}{n}\sum_{i=1}^n E\left(\left(\sum_{\substack{\tau\in I_{iwn}\\ \tau\neq i}}U_\tau\right)^2\right)+\frac{Cg_n}{n}\sum_{i=1}^n\sum_{\substack{\tau\in I_{iwn}\\ \tau\neq i}}\sum_{\substack{j=1\\ j\neq i}}^n\sum_{\substack{t\in I_{jwn}\\ t\neq j}}E(U_tU_\tau)\\ &\le \frac{C}{n}\sum_{i=1}^n\sum_{\substack{\tau\in I_{iwn}\\ \tau\neq i}}\sum_{\substack{t\in I_{iwn}\\ t\neq i}}|\omega_{t\tau}|+\frac{Cg_n}{n}\sum_{i=1}^n\sum_{\substack{\tau\in I_{iwn}\\ \tau\neq i}}\sum_{\substack{j=1\\ j\neq i}}^n\sum_{\substack{t\in I_{jwn}\\ t\neq j}}|\omega_{t\tau}|.\end{aligned}$$

By TA 4.2 τ belongs to at most $\aleph$ different index sets I_{iwn} (the same for t); hence given that $|\omega_{t\tau}|$ is bounded the first term on the right hand side of the last inequality is bounded by $C\aleph^2$. For the second term, note that $\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{t\in I_{jwn}\\t\neq j}}|\omega_{t\tau}| \le \aleph\sum_{t=1}^{n}|\omega_{t\tau}| \le C\aleph$ by assumptions TA 4.3; hence

$$\frac{Cg_n}{n}\sum_{i=1}^{n}\sum_{\substack{\tau\in I_{iwn}\\ \tau\neq i}}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{t\in I_{jwn}\\ t\neq j}}|\omega_{t\tau}| \le g_n C\aleph^2 = o(1).$$

The same manipulations as were used above show that

$$\beta_n = \sum_{w=1}^{W}(g_{wn}(\dot{\theta}) - g_{wn}(\theta_0))\frac{1}{\sqrt{ng_n}}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}K\left(\frac{X_i-x}{g_n}\right)(\check{m}(X_j)-m(X_j))$$

and therefore we need only show that $\frac{1}{\sqrt{ng_n}}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}K\left(\frac{X_i-x}{g_n}\right)(\check{m}(X_j)-m(X_j)) = O_p(1)$. Let K_i and K_{lj} be as defined in the proof of Theorem 3; then we can write

$$\frac{1}{\sqrt{ng_n}}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}K\left(\frac{X_i-x}{g_n}\right)(\check{m}(X_j)-m(X_j)) = \beta_{1n}(x)+\beta_{2n}(x)+\beta_{3n}(x),$$

where

$$\beta_{1n}(x) = \frac{1}{\sqrt{ng_n}}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}\sum_{l=1}^{n}\frac{K_iK_{lj}}{nh_n\bar{f}_n(X_j)}(m(X_l)-m(X_j)-m^{(1)}(X_j)(X_l-X_j)),$$

$$\beta_{2n}(x) = \frac{1}{\sqrt{ng_n}}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}\sum_{l=1}^{n}\frac{K_iK_{lj}}{nh_n\bar{f}_n(X_j)}U_l,$$

$$\beta_{3n}(x) = \frac{1}{\sqrt{ng_n}}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}K_i\left(O_p(h_n^3)+O_p\left(h_n\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}\right)\right).$$

We show that $\beta_{in}(x) = O_p(1)$ for $i = 1, 2, 3$. From Theorem 2,

$$\begin{aligned}
|\beta_{1n}(x)| &\le h_n^2 O_p(1)\frac{1}{\sqrt{ng_n}}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}K_i\\
&\le \aleph h_n^2 O_p(1)(ng_n)^{1/2}\frac{1}{ng_n}\sum_{i=1}^{n}K_i \le \aleph(ng_n)^{1/2}h_n^2O_p(1) \quad \text{since } \frac{1}{ng_n}\sum_{i=1}^{n}K_i = O_p(1)\\
&= O_p(1) \quad \text{provided } g_n = O(n^{-1/5}),\ h_n = O(n^{-1/5}).
\end{aligned}$$

Also,

$$\begin{aligned}
|\beta_{3n}(x)| &\le \aleph h_n^3(ng_n)^{1/2}O_p(1)\frac{1}{ng_n}\sum_{i=1}^{n}K_i + \aleph\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}h_n(ng_n)^{1/2}O_p(1)\frac{1}{ng_n}\sum_{i=1}^{n}K_i\\
&\le \aleph h_n^3(ng_n)^{1/2}O_p(1) + \aleph\left(\frac{nh_n}{\ln(n)}\right)^{-1/2}h_n(ng_n)^{1/2}O_p(1)\\
&= \left((nh_n^6g_n)^{1/2} + (g_nh_n\ln(n))^{1/2}\right)\aleph O_p(1)\\
&= O_p(1) \quad \text{provided } g_n = O(n^{-1/5}),\ h_n = O(n^{-1/5}).
\end{aligned}$$

We now examine $\beta_{2n}(x)$. We write

$$\begin{aligned}
\beta_{2n}(x) &= \sqrt{ng_n}\frac{1}{n}\sum_{l=1}^{n}U_l\frac{1}{nh_ng_n}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}\frac{K_iK_{lj}}{\bar{f}_n(X_j)}\\
&= \sqrt{ng_n}\frac{1}{n}\sum_{l=1}^{n}U_lc_{nl} \quad \text{where } c_{nl} = \frac{1}{nh_ng_n}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\neq i}}\frac{K_iK_{lj}}{\bar{f}_n(X_j)}.
\end{aligned}$$

Since $\{X_i\}$ and $\{U_i\}$ are independent it is easy to verify $E(\beta_{2n}(x)) = 0$ and

$$
\begin{aligned}
V(\beta_{2n}(x)) &= ng_n\frac{1}{n^2}\sum_{l=1}^{n}\sum_{k=1}^{n}E(U_lU_k)E(c_{nl}c_{nk})\\
&\le \frac{g_n}{n}\sum_{l=1}^{n}\sum_{k=1}^{n}|\omega_{lk}(\theta_0)||E(c_{nl}c_{nk})| \quad \text{and since } \inf_{x\in G}|\bar{f}_n(x)|>0,\\
&\le C\frac{g_n}{n}\sum_{l=1}^{n}\sum_{k=1}^{n}|\omega_{lk}(\theta_0)|\frac{1}{n^2h_n^2g_n^2}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\ne i}}\sum_{m=1}^{n}\sum_{\substack{o\in I_{mwn}\\ o\ne m}}E(K_iK_{lj}K_mK_{ko})\\
&= C\frac{g_n}{n}\sum_{l=1}^{n}\omega_{ll}(\theta_0)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\ne i}}\sum_{m=1}^{n}\sum_{\substack{o\in I_{mwn}\\ o\ne m}}E(K_iK_{lj}K_mK_{lo})\frac{1}{h_n^2g_n^2}\\
&\quad + C\frac{g_n}{n}\sum_{l=1}^{n}\sum_{\substack{k=1\\ k\ne l}}^{n}|\omega_{lk}(\theta_0)|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j\in I_{iwn}\\ j\ne i}}\sum_{m=1}^{n}\sum_{\substack{o\in I_{mwn}\\ o\ne m}}E(K_iK_{lj}K_mK_{ko})\frac{1}{h_n^2g_n^2}\\
&= T_{1n}+T_{2n}.
\end{aligned}
$$

We need to show that $T_{1n}, T_{2n} = O(1)$. We adopt the same strategy as was used in Theorem 3, i.e., establish the order of partial sums that emerge from considering all possible combinations of the indexes l, k, i, j, m, o in $T_{n1}, T_{n,2}$. Each of these partial sums is bounded by establishing the order $\pi_n = E(K_iK_{lj}K_mK_{lo})\frac{1}{h_n^2g_n^2}$ and $\rho_n = E(K_iK_{lj}K_mK_{ko})\frac{1}{h_n^2g_n^2}$.

We need to show that $T_{1n}, T_{2n} = o(1)$. The strategy that we use is to establish the order of the partial sums that emerge from considering all possible combinations of the indexes l, k, i, j, m, o in T_{1n}, T_{2n}.[6] Each of these partial is are shown to be $o_p(1)$ by first establishing the order of $\pi_n = \frac{1}{h_n^2g_n^2}E(K_iK_mK_{lj}K_{lo})$ and $\rho_n = \frac{1}{h_n^2g_n^2}E(K_iK_mK_{lj}K_{ko})$. Here we show the cases in which l and k are distinct from the indexes in the four inner sums, i.e., i, j, m, o.[7] We need to consider seven cases, and given Assumption A1 we have from calculating the expectations the following bounds: Case 1 ($i = m$ and $j = o$): $\pi_n \le \frac{C}{g_nh_n}$, $\rho_n \le \frac{C}{g_n}$; Case 2 ($i = o$ and $j = m$) $\pi_n \le \frac{C}{h_n}$, $\rho_n \le C$; Case 3 ($i = m$): $\pi_n \le \frac{C}{g_n}$, $\rho_n \le \frac{C}{g_n}$; Case 4 ($i = o$), Case 5 ($j = m$), Case 7 ($i \ne j \ne m \ne o$): $\pi_n \le C$, $\rho_n \le C$; Case 6 ($j = o$): $\pi_n \le \frac{C}{h_n}$, $\rho_n \le C$. We now denote the partial sums associated with $V((ng_n)^{1/2}B_{n,42}(x))$ in each of these cases by s_i, $i = 1, \dots, 7$. Hence, we have the following inequalities, where the first term refers to the partial sums in T_{1n} and the second term refers to the partial sums in T_{2n} for each case:

$$
s_1 \le \frac{C}{nh_n}\bar{\omega}_n + \frac{\aleph C}{n^2g_n}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\ne l}}^{n}|\omega_{lk}(\theta_0)|, \qquad s_2 \le \frac{Cg_n}{nh_n}\bar{\omega}_n + \frac{C}{n^2}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\ne l}}^{n}|\omega_{lk}(\theta_0)|
$$

$$
s_3 \le \frac{C\aleph^2}{n}\bar{\omega}_n + \frac{C\aleph^2}{n^2g_n}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\ne l}}^{n}|\omega_{lk}(\theta_0)|, \qquad s_4 \le \frac{C\aleph^2g_n}{n}\bar{\omega}_n + \frac{C\aleph^2}{n^2}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\ne l}}^{n}|\omega_{lk}(\theta_0)|.
$$

Case 5 is identical to Case 4 and

$$
s_6 \le \frac{C\aleph^2g_n}{nh_n}\bar{\omega}_n + \frac{C\aleph^2}{n^2}\sum_{l=1}^{n}g_n\sum_{\substack{k=1\\ k\ne l}}^{n}|\omega_{lk}(\theta_0)|, \qquad s_7 \le C\bar{\omega}_ng_n\aleph^2C + \frac{Cg_n}{n}\sum_{l=1}^{n}\sum_{\substack{k=1\\ k\ne l}}^{n}|\omega_{lk}(\theta_0)|\aleph^2.
$$

Hence, given Assumption A1 and the fact that from Theorem 1 $g_n\sum_{k=1,l\ne k}^{n}|\omega_{lk}(\theta_0)| = o(1)$ we conclude that in each case the first and second terms are O(1). □

Note on indexes: To construct the set of all index combinations for the sixfold sums we first note that for the four inner sums we need to consider seven different possible cases for i, j, m, o: Case 1 ($i = m$ and $j = o, i \ne j$); Case 2 ($i = o$ and $j = m, i \ne j$); Case 3 ($i = m$, but i, j, o distinct); Case 4 ($i = o$, but i, j, m distinct); Case 5 ($j = m$, but i, m, o distinct); Case 6 ($j = o$, but i, j, m distinct); Case 7 ($i \ne j \ne m \ne o$). In each of these cases we must then investigate all possible subcases where l and k are equal or distinct from the indexes considered in T_{1n} and T_{2n}.

Case 1: For the term T_{1n} there are 3 subcases: (1.1) l, i, j distinct; (1.2) $l = i$ and i, j distinct; (1.3) $l = j$ and i, j distinct. For the term T_{2n} there are 7 subcases: (1.1) l, k, i, j distinct; (1.2) $k = i, l, k, j$ distinct; (1.3) $k = j, l, k, i$ distinct; (1.4) $l = i, l, k, j$ distinct; (1.5) $l = j, l, k, i$ distinct; (1.6) $l = i, k = j, l, k$ distinct; (1.7) $l = j, k = i, l, k$ distinct.

Case 2: The subcases are identical to those in Case 1.

Case 3: For the term T_{1n} there are 4 subcases: (3.1) l, i, j, o distinct; (3.2) $l = i$ and i, j, o distinct; (3.3) $l = j$ and i, j, o distinct; (3.4) $l = o$ and i, j, o distinct. For the term T_{2n} there are 13 subcases: (3.1) l, k, i, j, o distinct; (3.2) $k = i, l, k, j, o$ distinct; (3.3) $l = i, i, k, j, o$ distinct; (3.4) $k = j, i, l, j, o$ distinct; (3.5) $l = j, l, k, i, o$ distinct; (3.6) $l = o, l, k, i, j$ distinct; (3.7) $k = o, l, i, j, k$ distinct; (3.8) $l = i, k = j, l, k, o$ distinct; (3.9) $l = j, i = k, l, k, o$ distinct; (3.10) $l = i, k = o, l, k, j$ distinct; (3.11) $l = o, i = k$, l, k, j distinct; (3.12) $l = j, k = o, i, l, k$ distinct; (3.13) $l = o, k = j, l, k, i$ distinct.

6 See the note on indexes in the end of Appendix A.

7 Bounds for all other cases described in Appendix A are available from the authors upon request.

Case 4: For the term T_{1n} there are 4 subcases: (4.1) l, i, j, m distinct; (4.2) $l = m$ and i, j, l distinct; (4.3) $l = i$ and i, j, m distinct; (4.4) $l = j$ and i, j, m distinct. For the term T_{2n} there are 13 subcases: (4.1) l, k, i, j, m distinct; (4.2) $k = m, l, k, j, i$ distinct; (4.3) $l = m, l, i, k, j$ distinct; (4.4) $k = i, l, k, j, m$ distinct; (4.5) $l = i, l, k, j, m$ distinct; (4.6) $k = j, l, k, i, m$ distinct; (4.7) $l = j, m, i, j, k$ distinct; (4.8) $l = m, k = i, l, k, j$ distinct; (4.9) $l = i, m = k, l, k, j$ distinct; (4.10) $l = m, k = j, l, k, i$ distinct; (4.11) $l = j, m = k, l, k, i$ distinct; (4.12) $l = i, k = j, m, l, k$ distinct; (4.13) $l = j, k = i, l, k, m$ distinct.

Case 5: identical to Case 4 due to symmetry.

Case 6: For the term T_{1n} there are 4 subcases: (6.1) l, i, j, m distinct; (6.2) $l = i$ and l, m, j distinct; (6.3) $l = m$ and i, j, l distinct; (6.4) $l = j$ and i, l, m distinct. For the term T_{2n} there are 13 subcases: (6.1) l, k, i, j, m distinct; (6.2) $k = i, l, k, j, m$ distinct; (6.3) $l = i, l, k, m, j$ distinct; (6.4) $k = m, l, k, j, i$ distinct; (6.5) $l = m, l, k, i, j$ distinct; (6.6) $k = j, l, k, i, m$ distinct; (6.7) $l = j, m, i, l, k$ distinct; (6.8) $l = i, k = m, l, k, j$ distinct; (6.9) $k = i, m = l, l, k, j$ distinct; (6.10) $l = i, k = j, l, k, m$ distinct; (6.11) $l = j, i = k, l, k, m$ distinct; (6.12) $l = m, k = j, i, l, k$ distinct; (6.13) $l = j, k = m, l, k, i$ distinct.

Case 7: For the term T_{1n} there are 5 subcases: (7.1) $l \neq i \neq j \neq m \neq o$; (7.2) $l = i$ and l, j, m, o are distinct; (7.3) $l = j$ and l, i, m, o are distinct; (7.4) $l = m$ and i, j, l, o are distinct; (7.5) $l = o$ and i, j, m, l are distinct. For the term T_{2n} there are 21 subcases: (7.1) $l \neq k \neq i \neq j \neq m \neq o$; (7.2) $l = i, j = k$ and l, j, m, o are distinct; (7.3) $l = k, j = l$ and i, j, m, o are distinct; (7.4) $l = i, k = m$ and i, j, m, o are distinct; (7.5) $i = k, l = m$ and i, j, m, o are distinct; (7.6) $l = i, k = o$ and i, j, m, o are distinct; (7.7) $i = k, l = o$ and i, j, m, o are distinct; (7.8) $l = j, k = m$ and i, j, m, o are distinct; (7.9) $j = k, l = m$ and i, j, m, o are distinct; (7.10) $l = j, k = o$ and i, j, m, o are distinct; (7.11) $j = k, l = o$ and i, j, m, o are distinct; (7.12) $l = m, k = o$ and i, j, m, o are distinct; (7.13) $m = k, l = o$ and i, j, m, o are distinct; (7.14) $i = k, l, k, j, m, o$ are distinct; (7.15) $i = l, l, k, j, m, o$ are distinct; (7.16) $j = k, l, k, i, m, o$ are distinct; (7.17) $l = j, l, k, i, m, o$ are distinct; (7.18) $m = k, l, k, i, j, o$ are distinct; (7.19) $m = l, l, k, i, j, o$ are distinct; (7.20) $o = k, l, k, i, j, m$ are distinct; (7.21) $l = o, l, k, i, j, m$ are distinct.

Appendix B

See Tables 1 and 2.

Table 1

Average bias ($\times 10^{-2}$) (B), standard deviation (S) and root mean squared error (R) with panel data models and $J = 2$

Estimators	$m_1(x)$			$m_2(x)$			$m_3(x)$		
	B	S	R	B	S	R	B	S	R
$N = 100$									
LLE	.335	.336	.336	.392	.333	.335	1.078	.349	.356
HU1	−.709	.472	.474	.721	.467	.477	−10.294	.519	.569
HU2	.315	.338	.338	.175	.333	.335	.420	.352	.358
RWC	.322	.284	.285	.318	.281	.285	1.449	.294	.306
2SLL	.278	.277	.278	.268	.275	.278	1.042	.289	.298
FHU1	−.707	.463	.465	.755	.460	.470	−9.999	.506	.551
FHU2	.329	.337	.337	.163	.333	.335	.431	.351	.357
FRWC	.327	.285	.286	.320	.282	.286	1.451	.296	.308
F2SLL	.289	.280	.280	.271	.277	.280	1.056	.291	.300
$N = 150$									
LLE	−.020	.271	.272	−.118	.270	.274	1.371	.285	.295
HU1	−.496	.373	.375	−.416	.374	.385	−9.795	.423	.479
HU2	.093	.271	.272	.304	.273	.276	.906	.289	.297
RWC	−.047	.228	.230	−.121	.229	.236	1.694	.242	.257
2SLL	−.051	.223	.224	−.162	.225	.230	1.364	.238	.249
FHU1	−.502	.368	.370	−.409	.370	.381	−9.597	.419	.471
FHU2	.102	.271	.271	.297	.272	.276	.931	.289	.297
FRWC	−.048	.229	.231	−.120	.230	.236	1.689	.243	.257
F2SLL	−.054	.224	.225	−.158	.226	.231	1.365	.239	.250
$N = 200$									
LLE	−.348	.237	.237	−.638	.237	.240	.120	.249	.256
HU1	−.397	.330	.335	.203	.334	.348	−10.232	.376	.451
HU2	−.604	.237	.237	−.955	.239	.241	.062	.247	.253
RWC	−.372	.198	.199	−.705	.201	.207	.364	.209	.221
2SLL	−.387	.194	.194	−.652	.197	.201	.125	.204	.213
FHU1	−.393	.327	.331	.210	.331	.345	−10.050	.373	.443
FHU2	−.602	.236	.237	−.953	.238	.241	.061	.247	.253
FRWC	−.371	.199	.200	−.706	.202	.207	.365	.210	.221
F2SLL	−.383	.194	.195	−.652	.197	.201	.129	.205	.214

Table 2
Average bias ($\times 10^{-2}$) (B), standard deviation (S) and root mean squared error (R) with AR(2) model

Estimators	$m_1(x)$			$m_2(x)$			$m_3(x)$		
	B	S	R	B	S	R	B	S	R
$n = 100$									
LLE	.081	.227	.227	.149	.225	.229	.510	.245	.252
HU1	−.285	.207	.208	.415	.210	.213	−.623	.236	.241
HU2	.214	.221	.221	.419	.220	.223	.648	.239	.246
VFF	.071	.202	.203	.243	.203	.207	.567	.220	.228
2SLL	.089	.203	.203	.228	.203	.208	.554	.221	.228
FHU1	−.284	.212	.212	.357	.213	.216	−.838	.243	.248
FHU2	.198	.221	.222	.419	.220	.223	.662	.239	.246
FVFF	.069	.203	.204	.225	.204	.209	.576	.222	.230
F2SLL	.085	.204	.205	.212	.205	.209	.561	.222	.230
$n = 200$									
LLE	.384	.156	.157	.011	.162	.166	.452	.171	.179
HU1	.214	.146	.147	.273	.151	.155	−.649	.166	.171
HU2	.335	.153	.154	.038	.158	.162	.418	.170	.177
VFF	.420	.141	.142	−.018	.145	.149	.422	.154	.162
2SLL	.424	.142	.142	−.017	.146	.150	.419	.154	.162
FHU1	.230	.147	.148	.264	.152	.155	−.633	.169	.174
FHU2	.347	.153	.154	.015	.158	.162	.419	.170	.176
FVFF	.412	.141	.142	−.029	.145	.150	.435	.154	.162
F2SLL	.415	.142	.142	−.023	.146	.150	.435	.154	.163
$n = 400$									
LLE	−.174	.111	.112	−.102	.114	.119	.332	.128	.135
HU1	−.484	.103	.104	.089	.108	.113	−.513	.125	.128
HU2	−.181	.108	.109	.000	.112	.117	.332	.126	.132
VFF	−.184	.099	.101	−.113	.102	.109	.297	.114	.121
2SLL	−.188	.099	.101	−.115	.102	.109	.290	.114	.121
FHU1	−.488	.104	.105	.063	.109	.113	−.515	.127	.130
FHU2	−.193	.108	.109	−.009	.112	.117	.327	.126	.132
FVFF	−.182	.099	.101	−.113	.103	.109	.295	.114	.122
F2SLL	−.188	.099	.101	−.114	.103	.109	.289	.114	.122

References

[1] D. Aigner, C.A.K. Lovell, P. Schmidt, Formulation and estimation of stochastic frontiers production function models, Journal of Econometrics 6 (1977) 21–37.
[2] B. Baltagi, Econometric Analysis of Panel Data, Wiley, New York, 1995.
[3] S. Bernstein, Sur l'extension du théorème du calcul des probabilités aux sommes de quantités dependantes, Mathematische Annalen 97 (1927) 1–59.
[4] D. Bosq, Nonparametric Statistics for Stochastic Processes, Springer-Verlag, New York, 1996.
[5] J. Davidson, Stochastic Limit Theory, Oxford University Press, New York, 1994.
[6] P. Doukhan, Mixing, Springer-Verlag, New York, 1994.
[7] J. Fan, Design adaptive nonparametric regression, Journal of the American Statistical Association 87 (1992) 998–1004.
[8] J. Fan, Local linear regression smoothers and their minimax efficiencies, Annals of Statistics 21 (1993) 196–216.
[9] J. Fan, I. Gijbels, Data driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation, Journal of the Royal Statistical Society B 57 (1995) 371–394.
[10] J. Fan, Q. Yao, Nonlinear Time Series, Springer-Verlag, New York, 2003.
[11] Y. Fan, Q. Li, A. Weersink, Semiparametric estimation of stochastic frontier models, Journal of Business and Economic Statistics 14 (1996) 460–468.
[12] M. Francisco-Fernandez, J.M. Vilar-Fernandez, Local polynomial regression estimation with correlated errors, Communications in Statistics — Theory and Methods 30 (2001) 1271–1293.
[13] A.R. Gallant, Nonlinear Statistical Models, John Wiley, New York, 1987.
[14] F. Graybill, Matrices with Applications in Statistics, Wadsworth, Belmont, CA, 1983.
[15] L. Györfi, et al., A Distribution-free Theory of Nonparametric Regression, Springer-Verlag, New York, 2002.
[16] D.J. Henderson, A. Ullah, A nonparametric random effects estimator, Economics Letters 88 (2005) 403–407.
[17] J. Jacod, P. Protter, Probability Essentials, Springer, Berlin, 2002.
[18] X. Lin, R.J. Carroll, Nonparametric function estimation for clustered data when the predictor is measured without/with error, Journal of the American Statistical Association 95 (2000) 520–534.
[19] Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Zeitschrift Wahrscheinlichkeitstheorie und Verwandte Gebiete 61 (1982) 405–415.
[20] D. Mandy, C. Martins-Filho, A unified approach to the asymptotic equivalence of Aitken and feasible Aitken instrumental variables estimator, International Economic Review 35 (1994) 957–979.
[21] C. Martins-Filho, F. Yao, Estimation of value-at-risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory, Studies in Nonlinear Dynamics and Econometrics 10 (2) (2006) Article 4.
[22] E. Masry, J. Fan, Local polynomial estimation of regression functions for mixing processes, Scandinavian Journal of Statistics 24 (1997) 165–179.
[23] A. Pagan, A. Ullah, Nonparametric Econometrics, Cambridge University Press, Cambridge, UK, 1999.
[24] T. Pham, L. Tran, Some mixing properties of time series models, Stochastic Processes and their Applications 19 (1985) 297–303.
[25] A.F. Ruckstuhl, A.H. Welsh, R.J. Carroll, Nonparametric function estimation of the relationship between two repeatedly measured variables, Statistica Sinica 10 (2000) 51–71.

[26] D. Ruppert, S.J. Sheather, M.P. Wand, An effective bandwidth selector for local least squares regression, Journal of the American Statistical Association 90 (1995) 1257–1270.
[27] T.A. Severini, J.G. Staniswalis, Quasi-likelihood estimation in semiparametric models, Journal of the American Statistical Association 89 (1994) 501–511.
[28] M. Smith, R. Kohn, Nonparametric seemingly unrelated regression, Journal of Econometrics 98 (2000) 257–281.
[29] L. Su, A. Ullah, More efficient estimation in nonparametric regression with nonparametric autocorrelated errors, Econometric Theory 22 (2006) 98–126.
[30] L. Su, A. Ullah, More efficient estimation in nonparametric regression with nonparametric autocorrelated errors, Economic Letters 96 (2007) 375–380.
[31] A. Ullah, N. Roy, Nonparametric and Semiparametric Econometrics of Panel Data, in: A. Ullah, D.E.A. Giles (Eds.), Handbook of Applied Economic Statistics, vol. 1, Marcel Dekker, New York, 1998, pp. 579–604.
[32] J.M. Vilar-Fernández, M. Francisco-Fernández, Local polynomial regression smoothers with AR-error structure, Test 11 (2002) 439–464.
[33] V.A. Volkonskii, Y.A. Rozanov, Some limit theorems for random functions I, Theory of Probability and its Applications 4 (1959) 178–197.
[34] N. Wang, Marginal nonparametric kernel regression accounting for within subject correlation, Biometrika 90 (2003) 43–52.
[35] T.J. Wansbeek, A. Kapetyn, A note on spectral decomposition and maximum likelihood estimation of ANOVA models with balanced data, Statistics and Probability Letters 1 (1983) 213–215.
[36] A.H. Welsh, T.W. Yee, Local regression for vector responses, Journal of Statistical Planning and Inference 136 (2006) 3007–3031.
[37] H. White, Asymptotic Theory for Econometricians, Academic Press, Orlando, FL, 2001.
[38] Z. Xiao, O.B. Linton, R.J. Carroll, E. Mammen, More efficient local polynomial estimation in nonparametric regression with autocorrelated errors, Journal of the American Statistical Association 98 (2003) 980–992.
[39] A. Zellner, An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias, Journal of the American Statistical Association 57 (1962) 500–509.