ELSEVIER

Stochastic Processes and their Applications 82 (1999) 173–193

stochastic processes and their applications

www.elsevier.com/locate/spa

Local linear regression estimation for time series with long-range dependence ☆

Elias Masry[a], Jan Mielniczuk[b,*]

[a] Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093, USA

[b] Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01-237 Warsaw, Poland

Received 21 August 1998; received in revised form 8 February 1999; accepted 12 February 1999

Abstract

Consider the nonparametric estimation of a multivariate regression function and its derivatives for a regression model with long-range dependent errors. We adopt local linear fitting approach and establish the joint asymptotic distributions for the estimators of the regression function and its derivatives. The nature of asymptotic distributions depends on the amount of smoothing resulting in possibly non-Gaussian distributions for large bandwidth and Gaussian distributions for small bandwidth. It turns out that the condition determining this dichotomy is different for the estimates of the regression function than for its derivatives; this leads to a double bandwidth dichotomy whereas the asymptotic distribution for the regression function estimate can be non-Gaussian whereas those of the derivatives estimates are Gaussian. Asymptotic distributions of estimates of derivatives in the case of large bandwidth are the scaled version of that for estimates of the regression function, resembling the situation of estimation of cumulative distribution function and densities under long-range dependence. The borderline case between small and large bandwidths is also examined.

MSC: 62G05; 60F05

Keywords: Local linear regression estimation; Long-range dependent errors; Random-design nonparametric regression; Smoothing dichotomy; Asymptotic distributions

1. Introduction

Let $\{Y_i, \boldsymbol{X}_i\}_{i=1}^{\infty}$ be jointly stationary processes with values in $\mathbb{R}$ and $\mathbb{R}^d$, respectively, and assume that $E|Y_1| < \infty$. Define the multivariate regression function of Y_1 given $\boldsymbol{X}_1 = \boldsymbol{x}_1$ as

$$g(\boldsymbol{x}) := E(Y_1|\boldsymbol{X}_1 = \boldsymbol{x}_1). \tag{1.1}$$

☆ The work of E. Masry was supported by the National Science Foundation under Grant DMS-97-03876. The work of J. Mielniczuk was supported by the Committee of Scientific Research, Poland, Grant 2 P301 001 006.

* Corresponding author.

E-mail address: miel@ipipan.waw.pl (J. Mielniczuk)

PII: S0304-4149(99)00015-0

The estimation of $g(\boldsymbol{x})$ from the observations $\{Y_i, \boldsymbol{X}_i\}_{i=1}^{n}$ is a fundamental problem in statistical data analysis.

There is an extensive literature on the estimation of $g(\boldsymbol{x})$ for weakly dependent processes: we mention in particular Robinson (1983), Roussas (1990), Truong and Stone (1992), Fan and Masry (1992) for estimates of g based on the Nadaraya–Watson approach. We also mention (Masry, 1996a, b) for estimators of g and its derivatives based on local polynomial fitting.

Recently, there has been an increasing interest in the estimation of the regression function g for processes $\{Y_i, \boldsymbol{X}_i\}_{i=1}^{\infty}$ which exhibit long-range dependence (Koul, 1992; Koul and Mukherjee, 1993; Hidalgo, 1997; Csörgő and Mielniczuk, 1998,1999). Unlike the weakly dependent case where no special model is assumed, various models are assumed in the case of long-range dependence. Here we focus on the model

$$Y_i = g(\boldsymbol{X}_i) + \varepsilon_i, \quad \varepsilon_i = G(Z_i, \boldsymbol{X}_i), \tag{1.2}$$

where $E(\varepsilon_1|\boldsymbol{X}_1) = 0$ almost surely, the processes $\{\boldsymbol{X}_i\}_{i=1}^{\infty}$ and $\{Z_i\}_{i=1}^{\infty}$ are independent, $\{Z_i\}_{i=1}^{\infty}$ is a stationary Gaussian process with zero mean, unit variance, such that for some $0 < \alpha < 1$

$$R(i) := E(Z_{i+1}Z_1) = \frac{L(i)}{i^{\alpha}}, \quad i = 1, 2, \ldots, \tag{1.3}$$

where $L(i)$ is slowly varying at infinity and eventually positive function. Model (1.2) was first considered by Cheng and Robinson (1994) who dealt with the estimation of certain moment-type functionals. Csörgő and Mielniczuk (1999) considered the estimation of g for model (1.2)–(1.3), using the Nadaraya–Watson kernel approach. In several papers a qualitatively different behavior for regression estimators in the long-range dependent case was shown under some assumptions on the interplay between amount of smoothing and the strength of dependence. Csörgő and Mielniczuk (1999) proved that for model (1.2)–(1.3) long-range dependence influences the asymptotic behavior of Nadaraya–Watson regression estimators only when the smoothing parameter is sufficiently large in a specified sense. The same phenomenon was established in Csörgő and Mielniczuk (1998) for Nadaraya–Watson estimate when the Gaussian sequence in (1.2) is replaced by long-range dependent linear process.

The purpose of this paper is to study the statistical properties of local linear regression estimators of *g and its derivatives* under the long-range dependent model (1.2)–(1.3). We note that local linear fitting has significant advantages over Nadaraya–Watson regression estimator: it has a smaller bias (see, for example, Fan, 1992,1993), it adapts automatically to the boundary of design points (see Fan and Gijbels, 1992,1996; Ruppert and Wand, 1994) and thus no boundary modification is required. It is superior to the Nadaraya–Watson estimator in the context of estimating the derivatives of the regression function (see Fan and Gijbels, 1992).

Assuming that g has continuous second partial derivatives in a neighbourhood of $\boldsymbol{x}$, our goal is to estimate g and its first-order derivatives

$$\boldsymbol{b}_1(\boldsymbol{x}) = \left(\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_d}\right)^{\mathrm{T}}(\boldsymbol{x}) \tag{1.4}$$

using local linear fitting and to establish the joint asymptotic distributions of their estimates for the long-range dependence model (1.2)–(1.3) ($\boldsymbol{u}^{\mathrm{T}}$ is the transpose of $\boldsymbol{u}$).

We show in this paper that the nature of asymptotic distributions for the estimators of g and its derivatives depends on the interplay between the amount of smoothing and the strength of dependence. In particular, for "small" bandwidths the asymptotic distributions coincide with those for i.i.d. or weakly dependent data. However, the condition determining smallness of bandwidth is different for the estimators of the regression function than for its derivatives (compare (3.11a) and (3.11b) below). On the other hand, when the appropriate conditions are reversed and large bandwidths are considered, asymptotic distributions of the estimates of the regression function and its derivatives are influenced by long-range dependence. Since these conditions do not coincide, it may happen that for certain amount of smoothing the asymptotic distribution of the regression function estimate *is* influenced by long-range dependence, whereas the asymptotic distribution of estimates of the derivatives *is not*. We thus have a multiple bandwidth dichotomy for the asymptotic behavior of the estimates of g and its derivatives. This paper extends the work of Csörgő and Mielniczuk (1999) from Nadaraya–Watson estimators of g to local linear fitting of g *and its derivatives*. Theorem 3 shows that the asymptotic distribution of the estimates of the derivatives of g in the case of large bandwidths is a scaled version of the asymptotic distribution of the estimate of g itself. This is an analogue of the results for estimators of a cumulative distribution function and its derivative, the probability density function, in the case of long-range dependence (Csörgő and Mielniczuk, 1995).

2. Formulation and preliminary results

In this section the processes $\{Y_i, \boldsymbol{X}_i\}$ are jointly stationary, model (1.2)–(1.3) is not assumed here.

Let $b_0(\boldsymbol{x}) := g(\boldsymbol{x})$ and put

$$\boldsymbol{b}(\boldsymbol{x}) = (b_0(\boldsymbol{x}), \boldsymbol{b}_1(\boldsymbol{x})^{\mathrm{T}})^{\mathrm{T}}, \tag{2.1}$$

where $\boldsymbol{b}_1(\boldsymbol{x})$ is defined in (1.4). Let $K(\boldsymbol{u})$ be a bounded integrable weight function on $\mathbb{R}^d$ and h be a bandwidth parameter. Given the observations $\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$, consider the multivariate weighted least squares

$$\sum_{i=1}^n (Y_i - b_0 - (\boldsymbol{X}_i - \boldsymbol{x})^{\mathrm{T}} \boldsymbol{b}_1)^2 K_h(\boldsymbol{X}_i - \boldsymbol{x}). \tag{2.2}$$

All vectors in the paper are column vectors,

$$K_h(\boldsymbol{u}) = h^{-d} K(\boldsymbol{u}/h) \tag{2.3}$$

and

$$h = h_n \to 0 \quad \text{as } n \to \infty. \tag{2.4}$$

Let $\binom{1}{\boldsymbol{u}}$ denote a vector $(1, u_1, \ldots, u_d)^{\mathrm{T}}$, where $\boldsymbol{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$. Eq. (2.2) can be written in the form

$$\sum_{i=1}^n (Y_i - (1, (\boldsymbol{X}_i - \boldsymbol{x})^{\mathrm{T}})\boldsymbol{b})^2 K_h(\boldsymbol{X}_i - \boldsymbol{x}). \tag{2.5}$$

Minimization of (2.5) with respect to $\boldsymbol{b}$ leads to the estimate $\widehat{\boldsymbol{b}} = \widehat{\boldsymbol{b}}_n(\boldsymbol{x})$ as the solution of

$$\sum_{i=1}^{n} Y_i \begin{pmatrix} 1 \\ \boldsymbol{X}_i - \boldsymbol{x} \end{pmatrix} K_h(\boldsymbol{X}_i - \boldsymbol{x}) = \left\{ \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \boldsymbol{X}_i - \boldsymbol{x} \end{pmatrix} (1, (\boldsymbol{X}_i - \boldsymbol{x})^{\mathrm{T}}) K_h(\boldsymbol{X}_i - \boldsymbol{x}) \right\} \widehat{\boldsymbol{b}}. \tag{2.6}$$

Let

$$\boldsymbol{Q}_n = \mathrm{diag}(1, h_n, \ldots, h_n), \tag{2.7}$$

$$\boldsymbol{t}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} Y_i K_h(\boldsymbol{X}_i - \boldsymbol{x}) \begin{pmatrix} 1 \\ h^{-1}(\boldsymbol{X}_i - \boldsymbol{x}) \end{pmatrix} \tag{2.8}$$

and

$$\boldsymbol{S}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ h^{-1}(\boldsymbol{X}_i - \boldsymbol{x}) \end{pmatrix} (1, h^{-1}(\boldsymbol{X}_i - \boldsymbol{x})^{\mathrm{T}}) K_h(\boldsymbol{X}_i - \boldsymbol{x}). \tag{2.9}$$

Then Eq. (2.6) can be written in the form

$$\boldsymbol{t}_n(\boldsymbol{x}) = \boldsymbol{Q}_n \boldsymbol{S}_n(\boldsymbol{x}) \widehat{\boldsymbol{b}}_n(\boldsymbol{x}), \tag{2.10a}$$

so that we have

$$\widehat{\boldsymbol{b}}_n(\boldsymbol{x}) = \boldsymbol{S}_n^{-1}(\boldsymbol{x}) \boldsymbol{Q}_n^{-1} \boldsymbol{t}_n(\boldsymbol{x}) \tag{2.10b}$$

as our estimate of $\boldsymbol{b}(\boldsymbol{x})$. Note that if the linear term is omitted in (2.2), then minimization of (2.2) with respect to b_0 yields the Nadaraya–Watson estimate $\widehat{b}_0 = \widehat{g}(\boldsymbol{x})$. In this section we state an asymptotic centered representation for the estimation error $\widehat{\boldsymbol{b}} - \boldsymbol{b}$ under fairly weak assumptions. This makes establishing asymptotic distributions for $\widehat{\boldsymbol{b}} - \boldsymbol{b}$ considerably simpler.

Assumption 1. (a) The kernel K is bounded with a compact support: $K(\boldsymbol{u}) = 0$ for $||\boldsymbol{u}|| > 1$.

(b) Let $f(\boldsymbol{u}, \boldsymbol{v}; l)$ be the joint probability density of $(\boldsymbol{X}_1, \boldsymbol{X}_{l+1})$ which is assumed to exist and $f(\boldsymbol{u})$ be the density of $\boldsymbol{X}_1$. Assume that

$$\rho(n) := \sup_{\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d} \frac{1}{n} \sum_{l=1}^{n} |f(\boldsymbol{u}, \boldsymbol{v}; l) - f(\boldsymbol{u}) f(\boldsymbol{v})| = \mathrm{o}\left(\frac{1}{n h_n^d} \right). \tag{2.11}$$

Remark 1. Observe that if $\{X_i\}$ is zero mean, unit variance univariate stationary Gaussian sequence with $E[X_i X_j] = R(i-j)$ such that $\sum_{i=1}^{\infty} |R(i)| < \infty$ then (2.11) is satisfied. Moreover, if instead of absolute summability of covariances we assume that (1.3) is satisfied for $\{X_i\}$ then (2.11) is fulfilled provided $L(n) n^{1-\alpha} h(n) = \mathrm{o}(1)$. This follows from the observation that in this case $|f(\boldsymbol{u}, \boldsymbol{v}; l) - f(\boldsymbol{u}) f(\boldsymbol{v})| \leqslant C |r(l)|$ (cf. Castellana and Leadbetter, 1986, p. 180).

Denote $\int_{\mathbb{R}^d}$ as $\int$ and define the moment matrices

$$\boldsymbol{M} = \int \begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix} (1, \boldsymbol{u}^{\mathrm{T}}) K(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u} = \begin{pmatrix} m_0 & \boldsymbol{m}_1 \\ \boldsymbol{m}_1^{\mathrm{T}} & \boldsymbol{M}_2 \end{pmatrix}, \tag{2.12}$$

$$\boldsymbol{B} = \int \begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix} \mathrm{vech}^{\mathrm{T}}(\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}) K(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u}, \tag{2.13}$$

where vech($\boldsymbol{A}$) of a $d\times d$ symmetric matrix $\boldsymbol{A}$ is the $d(d+1)/2$-dimensional column vector consisting of the concatenated column vectors of $\boldsymbol{A}$ which lie on and below the diagonal. Note that when $K(\boldsymbol{u})$ is a symmetric probability density then $\boldsymbol{M}$ is a diagonal matrix and its inverse exists. We assume throughout the paper that $\boldsymbol{M}$ is invertible. Now center the vector $\boldsymbol{t}_n$ by defining $\boldsymbol{t}_n^*$,

$$\boldsymbol{t}_n^*(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^{n}(Y_i-g(\boldsymbol{X}_i))K_h(\boldsymbol{X}_i-\boldsymbol{x})\begin{pmatrix}1\\ h^{-1}(\boldsymbol{X}_i-\boldsymbol{x})\end{pmatrix}. \tag{2.14}$$

Theorem 1. *Let Assumption* 1 *hold, and let* g *be twice continuously differentiable in the neighbourhood of* $\boldsymbol{x}$. *If* $nh_n^d\to\infty$ *and* $f(\boldsymbol{x})>0$ *we have*

$$\boldsymbol{Q}_n(\widehat{\boldsymbol{b}}-\boldsymbol{b})+\frac{1}{2}h_n^2\boldsymbol{M}^{-1}\boldsymbol{B}\,\mathrm{vech}(2\boldsymbol{V}-\mathrm{diag}(\boldsymbol{V}))=\frac{\boldsymbol{M}^{-1}\boldsymbol{t}_n^*(\boldsymbol{x})}{f(\boldsymbol{x})}(1+\mathrm{o_P}(1))+\mathrm{o_P}(h_n^2), \tag{2.15}$$

where the $d\times d$ *matrix* $\boldsymbol{V}=\boldsymbol{V}(\boldsymbol{x})$ *is given by*

$$\boldsymbol{V}(\boldsymbol{x})=\left(\frac{\partial^2 g(\boldsymbol{x})}{\partial x_i\partial x_j}\right)_{i,j=1}^{d}. \tag{2.16}$$

Remark 2. Since the kernel K has a compact support, the proof of Theorem 1 shows that the condition on the dependence index $\rho(n)$ can be weakened to a local supremum over a neighbourhood of $(\boldsymbol{x},\boldsymbol{x})$:

$$\tilde{\rho}_\varepsilon(n):=\sup_{||\boldsymbol{u}-\boldsymbol{x}||\leqslant\varepsilon,||\boldsymbol{v}-\boldsymbol{x}||\leqslant\varepsilon}\frac{1}{n}\sum_{l=1}^{n}|f(\boldsymbol{u},\boldsymbol{v};l)-f(\boldsymbol{u})f(\boldsymbol{v})|=\mathrm{o}\left(\frac{1}{nh_n^d}\right)$$

for some $\varepsilon>0$.

3. Main results

In this section we adopt the regression model

$$Y_i=g(\boldsymbol{X}_i)+\varepsilon_i,\quad \varepsilon_i=G(Z_i,\boldsymbol{X}_i), \tag{3.1}$$

where the processes $\{Z_i\}$ and $\{\boldsymbol{X}_i\}$ are independent, $\{\boldsymbol{X}_i\}$ are i.i.d. The d-dimensional random variables, and $\{Z_i\}$ is a stationary Gaussian zero mean, unit variance process such that for some $0<\alpha<1$

$$R(i):=E(Z_{i+1}Z_1)=\frac{L(i)}{i^\alpha},\quad i=1,2,\ldots, \tag{3.2}$$

where $L(i)$ is slowly varying and eventually positive function. We assume throughout that $EG(Z_1,\boldsymbol{x})=0$ for any $\boldsymbol{x}\in\mathbb{R}^d$. Let $H_j(z)=(-1)^j\mathrm{e}^{z^2/2}D^j(\mathrm{e}^{-z^2/2})$, $j\in\mathbb{N}$, denote the jth Hermite polynomial and ϕ the standard normal density. If $EG^2(Z_1,\boldsymbol{x})<\infty$ for fixed $\boldsymbol{x}\in\mathbb{R}^d$, then $G(\cdot,\boldsymbol{x})$ admits the Fourier–Hermite decomposition

$$G(z,\boldsymbol{x})=\sum_{j=r(\boldsymbol{x})}^{\infty}\frac{c_j(\boldsymbol{x})}{j!}H_j(z), \tag{3.3}$$

in $\mathscr{L}^2(\mathbb{R}, \phi)$, where

$$c_j(\boldsymbol{x}) = \int_{-\infty}^{\infty} G(z, \boldsymbol{x}) H_j(z) \phi(z)\, \mathrm{d}z$$

and

$$r(\boldsymbol{x}) = \min\{j \in \mathbb{N}\colon\ c_j(\boldsymbol{x}) \neq 0\} \tag{3.4}$$

is the Hermite rank of $G(\cdot, \boldsymbol{x})$. Note that $r(\boldsymbol{x}) > 0$ in view of the assumption $EG(Z_1, \boldsymbol{x}) = 0$. We assume throughout this section that $r\alpha < 1$. The main reason for imposing this condition is that it entails dichotomous behavior for the asymptotic distributions of the estimates of the regression and its derivatives. We conjecture that when $\{G(Z_i, \boldsymbol{x})\}$ is short-range dependent the dichotomy discussed in this paper does not occur. Under $r\alpha < 1$ we have (Dobrushin and Major, 1979; Taqqu, 1979) as $n \to \infty$, with $\xrightarrow{\mathscr{D}}$ denoting convergence in distribution

$$\left(\frac{n^\alpha}{L(n)}\right)^{r/2} \frac{1}{n} \sum_{i=1}^{n} G(Z_i, \boldsymbol{x}) \xrightarrow{\mathscr{D}} c_r(\boldsymbol{x}) \left(\frac{2}{r!(1 - r\alpha)(2 - r\alpha)}\right)^{1/2} \eta_r, \tag{3.5}$$

where η_r denotes the value at $t = 1$ of a Hermite process of rank r, given as a standardized multiple Wiener–Itô integral with respect to Brownian motion on $[0, 1]$ (cf. Taqqu, 1979, Theorem 5.6). The random variable η_1 is normal but $\eta_2, \eta_3, \ldots$ are not normally distributed.

It would be of interest to allow for long-range dependence of the explanatory random variables $\{\boldsymbol{X}_i\}$ in model (1.3) but since our approach cannot be easily generalized to incorporate this, it is beyond the scope of this paper. Note, however, that Koul (1992) treats such a case for a linear model and Hidalgo (1997) allows for long-range dependent $\{\boldsymbol{X}_i\}$ if (3.1) assumes the special form $\varepsilon_i = G(Z_i)$.

Put

$$|G'(z, \boldsymbol{x})| = \sum_{k=1}^{d} \left| \frac{\partial G(z, \boldsymbol{x})}{\partial x_k} \right|.$$

We impose the following conditions:

$\mathbb{C}_1$: K is bounded, $\int K(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = 1$ and $K(\boldsymbol{y}) = 0$ for $||\boldsymbol{y}|| > 1$;

$\mathbb{C}_2$: $f(\cdot)$ is continuous in a neighbourhood of $\boldsymbol{x}$ and $f(\boldsymbol{x}) > 0$;

$\mathbb{C}_3$: $E(G(Z_1, \boldsymbol{u}) - G(Z_1, \boldsymbol{x}))^2 \to 0$ as $\boldsymbol{u} \to \boldsymbol{x}$;

$\mathbb{C}_4$: $\min\{r(\boldsymbol{u})\colon\ ||\boldsymbol{u} - \boldsymbol{x}|| \leqslant \varepsilon\} = r(\boldsymbol{x})$ for some $\varepsilon > 0$;

$\mathbb{C}_5$: g is twice continuously differentiable in the neighbourhood of $\boldsymbol{x}$;

$\mathbb{C}_6$: f is continuously differentiable in a neighbourhood of $\boldsymbol{x}$;

$\mathbb{C}_7$: For each $z \in \mathbb{R}$ outside of a set of Lebesgue measure zero, the function $G(z, \cdot)$ is continuously differentiable in a neighbourhood of $\boldsymbol{x}$ and such that $E(\sup\{|G'(Z, \boldsymbol{y})|^2\colon ||\boldsymbol{y} - \boldsymbol{x}|| \leqslant \delta\}) < \infty$.

Let

$$m_i = \int u_i K(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}, \quad i = 0, 1, \ldots, d, \quad (u_0 = 1),$$

$$\gamma_{ij} = \int u_i u_j K^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}, \quad i, j = 0, 1, \ldots, d \tag{3.6}$$

and $\boldsymbol{m} = (m_0, \ldots, m_d)^{\mathrm{T}}$, $\Gamma = (\gamma_{ij})_{i,j=0}^{d}$. Moreover, put $\nabla v(\boldsymbol{x}) = (\partial v/\partial x_1, \ldots, \partial v/\partial x_d)^{\mathrm{T}}(\boldsymbol{x})$.

The following theorem gives the asymptotic covariances of $(t^*_{n,i})^d_{i=0}$, the components of the vector $\boldsymbol{t}^*_n$ defined in (2.14). It will determine the distinct norming factors for the cases of "large" and "small" bandwidth leading subsequently to two different asymptotic distributions (Theorems 3 and 4).

Theorem 2. (a) *Under conditions* $\mathbb{C}_1$–$\mathbb{C}_4$ *we have for* $i,j=0,1,\dots,d$

$$\operatorname{Cov}\{t^*_{n,i},t^*_{n,j}\} =\gamma_{ij}E[G^2(Z_1,\boldsymbol{x})]f(\boldsymbol{x})\frac{(1+\mathrm{o}(1))}{nh_n^d}+\frac{2m_im_jc_r^2(\boldsymbol{x})f^2(\boldsymbol{x})}{r!(1-r\alpha)(2-r\alpha)}\left(\frac{L(n)}{n^\alpha}\right)^r(1+\mathrm{o}(1)). \tag{3.7a}$$

(b) *If in addition* $\mathbb{C}_6$–$\mathbb{C}_7$ *are satisfied, K is symmetric and* $c_rf(\cdot)$ *is continuously differentiable in the neighbourhood of* $\boldsymbol{x}$, *then for* $i,j=1,\dots,d$

$$\operatorname{Cov}\{t^*_{n,i},t^*_{n,j}\} =\gamma_{ij}E[G^2(Z_1,\boldsymbol{x})]f(\boldsymbol{x})\frac{(1+\mathrm{o}(1))}{nh_n^d} +h_n^2w_{ij}(\boldsymbol{x},r)\left(\frac{2}{r!(1-r\alpha)(2-r\alpha)}\right)\left(\frac{L(n)}{n^\alpha}\right)^r(1+\mathrm{o}(1)), \tag{3.7b}$$

where

$$w_{ij}(\boldsymbol{x},r)=\left(\frac{\partial(fc_r)}{\partial x_i}(\boldsymbol{x})\right)\left(\frac{\partial(fc_r)}{\partial x_j}(\boldsymbol{x})\right)m_{ii}m_{jj}+\mathrm{o}(1)$$

and $m_{ij}=\int u_iu_jK(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$ *for* $i,j=1,\dots,d$.

Remark. (a) When the first term on the right-hand side of (3.7) is dominant we have

$$nh_n^d\operatorname{Var}(t^*_{n,i})\to\gamma_{ii}E[G^2(Z_1,\boldsymbol{x})]f(\boldsymbol{x}),\quad i=0,1,\dots,d.$$

This happens when

$$nh_n^d=\mathrm{o}((n^\alpha/L(n))^r) \tag{3.10a}$$

for $i=0$, and when

$$nh_n^{d+2}=\mathrm{o}((n^\alpha/L(n))^r) \tag{3.10b}$$

for $i=1,\dots,d$. Conditions (3.10a) and (3.10b) will be called small bandwidth condition for estimates of g and ∇g, respectively. In this case, under respectively, (3.10a) and (3.10b), the asymptotic distributions of the estimates of g and its derivatives are normal (Theorem 4). When

$$(n^\alpha/L(n))^r=\mathrm{o}(nh_n^d), \tag{3.11a}$$

the second term in (3.7a) is dominant for $i=0$, in which case

$$\left(\frac{n^\alpha}{L(n)}\right)^r\operatorname{Var}(t^*_{n,0})\to\frac{2m_0^2c_r^2(\boldsymbol{x})f^2(\boldsymbol{x})}{r!(1-r\alpha)(2-r\alpha)}.$$

On the other hand, if the kernel is symmetric then $m_i=0$ for $i=1,\dots,d$ and if a more stringent condition holds, namely

$$(n^\alpha/L(n))^r=\mathrm{o}(nh_n^{d+2}) \tag{3.11b}$$

then for $i = 1,\dots,d$ the second term in (3.7b) is dominant and

$$\frac{1}{h_n^2}\left(\frac{n^\alpha}{L(n)}\right)^r \mathrm{Var}(t^*_{n,i}) \to \left(\frac{\partial(c_r f)}{\partial x_i}\right)^2(\boldsymbol{x})\left(\frac{2}{r!(1-r\alpha)(2-r\alpha)}\right)m_{ii}^2.$$

In this case the asymptotic distributions of the estimates of g and its derivatives are scaled distributions of a fixed random variable η_r defined below (3.5). Conditions (3.11a) and (3.11b) will be called large bandwidth conditions for estimates of g and ∇g, respectively. Observe also that an equivalent form of condition (3.11b) is $(nh_n^{d+2})^{-1} = \mathrm{o}(\mathrm{Var}(\sum_{i=1}^n G(Z_i,\boldsymbol{x})/n)$. Here $(nh_n^{d+2})^{-1}$ is asymptotically proportional to the variance of any component of $\widehat{\boldsymbol{b}}_1(\boldsymbol{x})$ when the errors $\varepsilon_1,\varepsilon_2,\dots$ in (1.2) are independent and identically distributed. Hence under (3.11b) such a variance will be dominated by the variance of the sample mean of the errors in the present model (1.2).

(b) It is seen from the proof of Theorem 2 that (3.7b) does not hold if i or j is equal to 0. Moreover, note that the condition of differentiability of $c_r(\cdot)$ which implies the assumed differentiability of $c_r f(\cdot)$ in view of $\mathbb{C}_5$ is not always satisfied. For example, if $G(z,\boldsymbol{x}) = G_1(z)G_2(\boldsymbol{x})$ it holds provided $G_2(\cdot)$ is differentiable and positive in a neighbourhood of $\boldsymbol{x}$ but it fails at points for which $G_2(\cdot)$ is zero. However, in such cases $\mathbb{C}_4$ is violated also.

Theorem 3. (a) *Suppose $r\alpha < 1$ and conditions $\mathbb{C}_1$–$\mathbb{C}_5$ are satisfied. If* (3.11a) *holds and the sequence $\{nh_n^{d+4}\}$ is bounded then as $n \to \infty$*

$$\left(\frac{n^\alpha}{L(n)}\right)^{r/2} \boldsymbol{Q}_n(\widehat{\boldsymbol{b}} - \boldsymbol{b}) \xrightarrow{\mathcal{D}} \left(\frac{2}{r!(1-r\alpha)(2-r\alpha)}\right)^{1/2} c_r(\boldsymbol{x})\boldsymbol{M}^{-1}\boldsymbol{m}\eta_r,$$

where $\boldsymbol{m} = (m_0, m_1, \dots, m_d)^{\mathrm{T}}$ and the moment matrix $\boldsymbol{M}$ is given in (2.12).

(b) *Assume* (3.11b) *instead of* (3.11a), *K is symmetric, $\mathbb{C}_6$–$\mathbb{C}_7$ are satisfied and $c_r f(\cdot)$ is continuously differentiable in the neighbourhood of $\boldsymbol{x}$. Then*

$$\left(\frac{n^\alpha}{L(n)}\right)^{r/2} (\widehat{\boldsymbol{b}}_1 - \boldsymbol{b}_1) \xrightarrow{\mathcal{D}} \left(\frac{2}{r!(1-r\alpha)(2-r\alpha)}\right)^{1/2} \frac{\nabla(c_r f)(\boldsymbol{x})}{f(\boldsymbol{x})}\eta_r.$$

Remark. Since $\boldsymbol{m}$ is the first column of $\boldsymbol{M}$, $\boldsymbol{M}^{-1}\boldsymbol{m} = (1,\boldsymbol{0})^{\mathrm{T}}$. Consequently, Theorem 3(a) implies that for large bandwidth h_n, in the sense of (3.11a), we have

$$\left(\frac{n^\alpha}{L(n)}\right)^{r/2} (\widehat{g}(\boldsymbol{x}) - g(\boldsymbol{x})) \xrightarrow{\mathcal{D}} \left(\frac{2}{r!(1-r\alpha)(2-r\alpha)}\right)^{1/2} c_r(\boldsymbol{x})\eta_r,$$

whereas for the estimates of the derivatives of g we have under (3.11b)

$$\left(\frac{n^\alpha}{L(n)}\right)^{r/2} (\widehat{\boldsymbol{b}}_1(\boldsymbol{x}) - \boldsymbol{b}_1(\boldsymbol{x})) \xrightarrow{\mathcal{D}} \left(\frac{2}{r!(1-r\alpha)(2-r\alpha)}\right)^{1/2} \frac{\nabla(c_r f)(\boldsymbol{x})}{f(\boldsymbol{x})}\eta_r.$$

This behavior is in sharp contrast to local polynomial fitting under a weak dependence assumption (Masry, 1996b), where the norming factors for g and its first-order derivatives, leading to nondegenerate asymptotic distributions, differ by a multiplicative factor h_n.

It will be shown in Theorem 4 that for small bandwidth h_n, in the sense specified in (3.10a), the local linear estimators of g and its derivatives have a (nondegenerate) normal asymptotic distributions.

Let

$$\boldsymbol{\Sigma} := \frac{\sigma^2(\boldsymbol{x})}{f(\boldsymbol{x})}\boldsymbol{M}^{-1}\Gamma\boldsymbol{M}^{-1} =: \begin{pmatrix} \sigma_{00} & \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{01}^{\mathrm{T}} & \boldsymbol{\Sigma}_{11} \end{pmatrix},$$

where $\sigma^2(\boldsymbol{x}) = E(G^2(Z, \boldsymbol{x}))$.

Theorem 4. (a) *Suppose that* $r\alpha < 1$ *and conditions* $\mathbb{C}_1$–$\mathbb{C}_7$ *hold. If as* $n \to \infty$, $nh_n^d \to \infty$, $nh_n^{d+4} \to 0$, *and condition* (3.10a) *holds, then*

$$(nh_n^d)^{1/2}\boldsymbol{Q}_n(\widehat{\boldsymbol{b}} - \boldsymbol{b}) \xrightarrow{\mathscr{D}} \mathrm{N}(0, \boldsymbol{\Sigma}).$$

(b) *Assume* (3.10b) *instead of* (3.10a), K *is symmetric and impose the remaining conditions of* (a). *Then*

$$(nh_n^{d+2})^{1/2}(\widehat{\boldsymbol{b}}_1 - \boldsymbol{b}_1) \xrightarrow{\mathscr{D}} \mathrm{N}(0, \boldsymbol{\Sigma}_{11}).$$

Remark. Theorem 4(a) gives the joint asymptotic normality of the estimates of g and its first-order partial derivatives when the bandwidth is small in the sense of (3.10a). Note that when K is a symmetric kernel then with $u_0 = 1$, $m_{ij} = \int u_i u_j K(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 0$, $\gamma_{ij} = \int u_i u_j K^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 0$ for $i \neq j$, $i, j = 0, 1, \ldots, d$. Thus

$$\boldsymbol{\Sigma} = \mathrm{diag}\left(\frac{\gamma_{00}}{m_{00}^2}, \ldots, \frac{\gamma_{dd}}{m_{dd}^2}\right),$$

i.e. the scaled estimates of $g(\boldsymbol{x})$ and its derivatives are asymptotically *independent* normal variates. Moreover, in the setup of Theorem 4(b), $\boldsymbol{\Sigma}_{11} = \mathrm{diag}(\gamma_{11}/m_{11}^2, \ldots, \gamma_{dd}/m_{dd}^2)$. Note also that we need $nh_n^{d+2} \to \infty$ in order to ensure weak consistency of the estimators of derivatives.

Theorem 5. (a) *Assume that conditions* $\mathbb{C}_1$–$\mathbb{C}_7$ *hold. If* $nh_n^d/(n^\alpha/L(n))^r \to C_b^2$ *for some constant* $0 < C_b < \infty$ *and* $nh_n^{d+4} \to 0$, *then*

$$(nh_n^d)^{1/2}\boldsymbol{Q}_n(\widehat{\boldsymbol{b}} - \boldsymbol{b}) \xrightarrow{\mathscr{D}} \zeta_1 + \zeta_2,$$

where

$$\zeta_1 = C_b\left(\frac{2}{r!(1 - r\alpha)(2 - r\alpha)}\right)^{1/2} c_r(\boldsymbol{x})\boldsymbol{M}^{-1}\boldsymbol{m}\eta_r$$

and ζ_2 *has* $\mathrm{N}(0, \boldsymbol{\Sigma})$ *distribution. Moreover,* ζ_1 *and* ζ_2 *are independent.*

(b) *Assume that* $nh_n^{d+2}/(n^\alpha/L(n))^r \to \bar{\mathbb{C}}_b^2$ *for some constant* $0 < \bar{C}_b < \infty$, $nh_n^{d+4} \to 0$, K *is symmetric and the conditions* $\mathbb{C}_1 - \mathbb{C}_7$ *hold. Then*

$$(nh_n^{d+2})^{1/2}(\widehat{\boldsymbol{b}}_1 - \boldsymbol{b}_1) \xrightarrow{\mathscr{D}} \bar{\zeta}_1 + \bar{\zeta}_2,$$

where

$$\bar{\zeta}_1 = \bar{C}_b\left(\frac{2}{r!(1 - r\alpha)(2 - r\alpha)}\right)^{1/2} \frac{\nabla(c_r f)}{f(\boldsymbol{x})}\eta_r,$$

and $\bar{\zeta}_2$ *has* $\mathrm{N}(0, \boldsymbol{\Sigma}_{11})$ *distribution. Moreover,* $\bar{\zeta}_1$ *and* $\bar{\zeta}_2$ *are independent.*

4. Discussion

Theorems 3 and 4 show that the fundamental dichotomy of behavior for the Nadaraya–Watson regression estimator, under LRD, carries over to locally linear estimator of regression. Dichotomy of behavior also occurs for estimates of derivatives; however the borderline condition distinguishing between large and small bandwidths is different in this case than for estimating regression. In the case of derivatives, larger bandwidths than for estimating regression are necessary in order for long-range dependence to influence the asymptotic distribution of estimators. Thus it may happen that for a certain sequence of bandwidths the asymptotic distribution of regression estimates is influenced by long-range dependence, whereas the asymptotic distributions for estimates of derivatives are not. Moreover, for estimating derivatives, Theorem 4 shows that in the case of small bandwidths asymptotic behavior of locally linear estimator of $\boldsymbol{b}_1$ is the same as in the i.i.d. or weakly dependent case.

We discuss the interplay between the parameters d, r and α and the bandwidth in the case when $b_n = Cn^{-\delta}$. The assumptions of Theorem 3(a) imply $1/(d+4) \leqslant \delta \leqslant (1-r\alpha)/d$, whereas those of Theorem 4(a) entail $\max\{(1-r\alpha)/d, 1/(d+4)\} < \delta < 1/d$. The lower bound on δ, namely $1/(d+4) < \delta$ may be easily weakened to $1/(d+2(p+1)) < \delta$ when local polynomial fitting of order p is used. Note that the conditions of Theorem 3(b) impose more stringent condition on δ than those of Theorem 3(a): namely $1/(d+4) \leqslant \delta \leqslant (1-r\alpha)/(d+2)$.

Observe also that in the restrictive case of $G(z, \boldsymbol{x}) = G(z)$ considered e.g. in Hidalgo (1997) assumptions $\mathbb{C}_3, \mathbb{C}_4, \mathbb{C}_7$ are automatically satisfied.

The results of this paper can be easily generalized to joint asymptotic distributions for $(\widehat{\boldsymbol{b}}(\boldsymbol{x}_1) - \boldsymbol{b}(\boldsymbol{x}_1), \dots, \widehat{\boldsymbol{b}}(\boldsymbol{x}_k) - \boldsymbol{b}(\boldsymbol{x}_k))^{\mathrm{T}}$ for distinct points $\{\boldsymbol{x}_i\}_{i=1}^k$, where $k \geqslant 1$. Under obvious modifications of assumptions, the asymptotic law for estimates of $(b_0(\boldsymbol{x}_1), \dots, b_0(\boldsymbol{x}_k))^{\mathrm{T}}$ in the case of large bandwidth (Theorem 3) is proportional to

$$(c_r(\boldsymbol{x}_1), \dots, c_r(\boldsymbol{x}_k))^{\mathrm{T}} \eta_r,$$

where r is the minimal Hermite rank of the functions $G(\cdot, \boldsymbol{x}_1), \dots, G(\cdot, \boldsymbol{x}_k)$. Analogously, the asymptotic law for estimates of $\boldsymbol{b}_1$ is proportional to

$$\left(\frac{\nabla(c_r f)(\boldsymbol{x}_1)}{f(\boldsymbol{x}_1)}, \dots, \frac{\nabla(c_r f)(\boldsymbol{x}_k)}{f(\boldsymbol{x}_k)} \right)^{\mathrm{T}} \eta_r.$$

In the case of small bandwidth (Theorem 4) the scaled variates $\{\widehat{\boldsymbol{b}}(\boldsymbol{x}_k) - \boldsymbol{b}(\boldsymbol{x}_k)\}$ are asymptotically normal variables, each of them with covariance structure specified in Theorem 4.

5. Derivations

Define the $(d+1) \times d(d+1)/2$ matrix $\boldsymbol{B}_n(\boldsymbol{x})$ by

$$\boldsymbol{B}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ h^{-1}(\boldsymbol{X}_i - \boldsymbol{x}) \end{pmatrix} \mathrm{vech}^{\mathrm{T}} \left[\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h} \right) \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h} \right)^{\mathrm{T}} \right] K_h(\boldsymbol{X}_i - \boldsymbol{x}). \tag{5.1}$$

In the proof of Theorem 1 we need the following lemma. Mean-square convergence is denoted by $\stackrel{\text{m.s.}}{\to}$.

Lemma 1. *Under Assumption* 1 *and* $nh_n^d \to \infty$ *we have*

$$S_n(\boldsymbol{x}) \stackrel{\text{m.s.}}{\to} f(\boldsymbol{x})\boldsymbol{M}$$

and

$$B_n(\boldsymbol{x}) \stackrel{\text{m.s.}}{\to} f(\boldsymbol{x})\boldsymbol{B}$$

as $n \to \infty$ *at continuity points* $\boldsymbol{x}$ *of* $f(\boldsymbol{x})$, *where the moment matrices* $\boldsymbol{M}$ *and* $\boldsymbol{B}$ *are given in* (2.12)–(2.13).

Proof of the Lemma. We focus our attention on the convergence of $\boldsymbol{S}_n(\boldsymbol{x})$. By (2.9) and stationarity

$$E(\boldsymbol{S}_n(\boldsymbol{x})) = \int \begin{pmatrix} 1 \\ h^{-1}(\boldsymbol{u}-\boldsymbol{x}) \end{pmatrix} \left(1, \left(\frac{\boldsymbol{u}-\boldsymbol{x}}{h}\right)^{\mathrm{T}}\right) K_h(\boldsymbol{u}-\boldsymbol{x}) f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

and by Bochner's lemma

$$E(\boldsymbol{S}_n(\boldsymbol{x})) \to f(\boldsymbol{x}) \int \begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix} (1, \boldsymbol{u}^{\mathrm{T}}) K(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = f(\boldsymbol{x})\boldsymbol{M}$$

as $n \to \infty$ at continuity points $\boldsymbol{x}$ of $f(\cdot)$. Write $\boldsymbol{S}_n(\boldsymbol{x}) = (s_n^{(i,j)}(\boldsymbol{x}))_{i,j=0}^d$. Then for $i, j \geqslant 1$

$$\operatorname{Var}(s_n^{(i,j)}(\boldsymbol{x})) = \frac{1}{n}\operatorname{Var}(U_1^{(i,j)}) + \frac{2}{n}\sum_{l=1}^{n-1}\left(1 - \frac{l}{n}\right)\operatorname{Cov}(U_1^{(i,j)}, U_{l+1}^{(i,j)}) =: I_1 + I_2, \quad (5.2)$$

where

$$U_l^{(i,j)} = \left(\frac{X_{l,i} - x_i}{h}\right)\left(\frac{X_{l,j} - x_j}{h}\right) K_h(\boldsymbol{X}_l - \boldsymbol{x}).$$

Now,

$$\begin{aligned} nh_n^d I_1 &= h_n^d E[(U_1^{(i,j)})^2] + \mathcal{O}(h_n^d) \\ &= \int \left(\frac{u_i - x_i}{h}\right)^2 \left(\frac{u_j - x_j}{h}\right)^2 \left[\frac{1}{h^d} K^2\left(\frac{\boldsymbol{u}-\boldsymbol{x}}{h}\right)\right] f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \mathcal{O}(h_n^d) \end{aligned}$$

and under Assumption 1 and Bochner's lemma

$$nh_n^d I_1 \to f(\boldsymbol{x}) \int u_i^2 u_j^2 K^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \quad (5.3)$$

at continuity points $\boldsymbol{x}$ of $f(\cdot)$. Next,

$$\begin{aligned} |I_2| &\leqslant \frac{2}{n}\left|\sum_{l=1}^{n-1} \operatorname{cov}(U_1^{(i,j)}, U_{l+1}^{(i,j)})\right| \\ &\leqslant 2 \int\int \left| \left(\frac{u_i - x_i}{h}\right)\left(\frac{u_j - x_j}{h}\right)\left(\frac{v_i - x_i}{h}\right) \right. \end{aligned}$$

$$\times\left(\frac{v_j-x_j}{h}\right)\rho(n)K_h(\boldsymbol{u}-\boldsymbol{x})K_h(\boldsymbol{v}-\boldsymbol{x})\bigg|\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}$$

$$\leqslant 2\rho(n)\left(\int\left|\frac{u_i-x_i}{h}\right|\left|\frac{u_j-x_j}{h}\right||K_h(\boldsymbol{u}-\boldsymbol{x})|\,\mathrm{d}\boldsymbol{u}\right)^2 = 2\rho(n)\left(\int|u_iu_j||K(\boldsymbol{u})|\,\mathrm{d}\boldsymbol{u}\right)^2 \tag{5.4}$$

Thus $nh_n^d|I_2| = \mathrm{o}(1)$ by Assumption 1(b). Hence by (5.2)–(5.4),

$$nh_n^d\,\mathrm{Var}(s_n^{(i,j)}(\boldsymbol{x})) \to f(\boldsymbol{x})\int u_i^2u_j^2K^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}, \quad i,j\geqslant 1$$

at continuity points $\boldsymbol{x}$ of f. The proof for i,j such that $ij=0$ as well as for $B_n(\boldsymbol{x})$ is similar. □

Proof of Theorem 1. By (2.8) and (2.15)

$$\boldsymbol{t}_n(\boldsymbol{x}) - \boldsymbol{t}_n^*(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ h^{-1}(\boldsymbol{X}_i-\boldsymbol{x})\end{pmatrix}g(\boldsymbol{X}_i)K_h(\boldsymbol{X}_i-\boldsymbol{x}). \tag{5.5}$$

Expanding $g(\boldsymbol{X}_i)$ in a Taylor series around $\boldsymbol{x}$ for $||\boldsymbol{X}_i-\boldsymbol{x}||\leqslant h$ we have

$$g(\boldsymbol{X}_i) = (1,(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}})\boldsymbol{b} + \tfrac{1}{2}(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{V}(\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x}) + \mathrm{o_P}(h_n^2).$$

Substituting in (5.5)

$$\boldsymbol{t}_n(\boldsymbol{x}) - \boldsymbol{t}_n^*(\boldsymbol{x}) = \frac{1}{n}\left\{\sum_{i=1}^{n}\begin{pmatrix}1\\ h^{-1}(\boldsymbol{X}_i-\boldsymbol{x})\end{pmatrix}(1,(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}})K_h(\boldsymbol{X}_i-\boldsymbol{x})\right\}\boldsymbol{b}$$
$$+\frac{1}{2n}\sum_{i=1}^{n}\begin{pmatrix}1\\ h^{-1}(\boldsymbol{X}_i-\boldsymbol{x})\end{pmatrix}(\boldsymbol{X}_i-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{V}(\boldsymbol{x})(\boldsymbol{X}_i-\boldsymbol{x})K_h(\boldsymbol{X}_i-\boldsymbol{x})$$
$$+\mathrm{o_p}(h_n^2)\frac{1}{n}\sum_{i=1}^{n}\left|\begin{pmatrix}1\\ h^{-1}(\boldsymbol{X}_i-\boldsymbol{x})\end{pmatrix}K_h(\boldsymbol{X}_i-\boldsymbol{x})\right|. \tag{5.6}$$

Note that the first term on the right-hand side of (5.6) is $\boldsymbol{Q}_n\boldsymbol{S}_n(\boldsymbol{x})\boldsymbol{b}$. For the second term, since $\boldsymbol{V}(\boldsymbol{x})$ is symmetric, we have for any vector $\boldsymbol{a}$

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{a} = \mathrm{vech}^{\mathrm{T}}(\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}})\mathrm{vech}(2\boldsymbol{V}-\mathrm{diag}(\boldsymbol{V})).$$

Now with the matrix $\boldsymbol{B}_n(\boldsymbol{x})$ defined in (5.1), it is seen that that the second term on the right-hand side of (5.6) is equal to $\frac{1}{2}h_n^2\boldsymbol{B}_n(\boldsymbol{x})\mathrm{vech}(2\boldsymbol{V}-\mathrm{diag}(\boldsymbol{V}))$. Finally, if we define

$$\begin{pmatrix}\bar{s}_{n,0}\\ \bar{\boldsymbol{s}}_{n,1}\end{pmatrix} := \frac{1}{n}\sum_{i=1}^{n}\left|\begin{pmatrix}1\\ h^{-1}(\boldsymbol{X}_i-\boldsymbol{x})\end{pmatrix}K_h(\boldsymbol{X}_i-\boldsymbol{x})\right|,$$

then the third term on the right-hand side of (5.6) is equal to $\mathrm{o_p}(h_n^2)(\bar{s}_{n,0},\bar{\boldsymbol{s}}_{n,1}^{\mathrm{T}})^{\mathrm{T}}$. Thus

$$\boldsymbol{t}_n(\boldsymbol{x}) - \boldsymbol{t}_n^*(\boldsymbol{x}) = \boldsymbol{Q}_n\boldsymbol{S}_n(\boldsymbol{x})\boldsymbol{b} + \tfrac{1}{2}h_n^2\boldsymbol{B}_n(\boldsymbol{x})\mathrm{vech}(2\boldsymbol{V}-\mathrm{diag}(\boldsymbol{V})) + \mathrm{o_P}(h_n^2)(\bar{s}_{n,0},\bar{\boldsymbol{s}}_{n,1}^{\mathrm{T}})^{\mathrm{T}}. \tag{5.7}$$

It now follows from (2.10a) and (5.7) that

$$\boldsymbol{t}_n^*(\boldsymbol{x}) = \boldsymbol{Q}_n\boldsymbol{S}_n(\boldsymbol{x})(\widehat{\boldsymbol{b}} - \boldsymbol{b}) + \tfrac{1}{2}h_n^2\boldsymbol{B}_n(\boldsymbol{x})\mathrm{vech}(2\boldsymbol{V} - \mathrm{diag}(\boldsymbol{V})) + (\bar{s}_{n,0}, \bar{\boldsymbol{s}}_{n,1}^{\mathrm{T}})^{\mathrm{T}} o_{\mathrm{P}}(h_n^2).$$

The result follows from the above relation with both sides multiplied by $\boldsymbol{S}_n^{-1}$, the lemma, the fact that $(\bar{s}_{n,0}, \bar{\boldsymbol{s}}_{n,1}^{\mathrm{T}}) = \mathcal{O}_{\mathrm{P}}(1)$ by an obvious adaptation of the proof of Lemma 1, and the invertability of $\boldsymbol{M}$. □

Proof of Theorem 2. (*Part* (a)). For arbitrary $\boldsymbol{a} \in \mathbb{R}^{d+1}$ with $||\boldsymbol{a}|| > 0$, put

$$W_n := \boldsymbol{a}^{\mathrm{T}}\boldsymbol{t}_n^* = \frac{1}{n}\sum_{i=1}^{n} G(Z_i, \boldsymbol{X}_i)\tilde{K}_h(\boldsymbol{X}_i - \boldsymbol{x}), \tag{5.8}$$

where

$$\tilde{K}_h(\boldsymbol{u}) = \frac{1}{h^d}\tilde{K}(\boldsymbol{u}/h), \qquad \tilde{K}(\boldsymbol{u}) = \boldsymbol{a}^{\mathrm{T}}\begin{pmatrix} 1 \\ \boldsymbol{u} \end{pmatrix} K(\boldsymbol{u}). \tag{5.9}$$

Then

$$\begin{aligned}\mathrm{Var}(W_n) &= \frac{1}{n}E[G^2(Z_1, \boldsymbol{X}_1)\tilde{K}_h^2(\boldsymbol{X}_1 - \boldsymbol{x})] \\ &\quad + \frac{1}{n^2}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} E[G(Z_i, \boldsymbol{X}_i)G(Z_j, \boldsymbol{X}_j)\tilde{K}_h(\boldsymbol{X}_i - \boldsymbol{x})\tilde{K}_h(\boldsymbol{X}_j - \boldsymbol{x})] \\ &=: J_1(\boldsymbol{x}) + J_2(\boldsymbol{x}). \end{aligned} \tag{5.10}$$

We have

$$\begin{aligned} nh_n^d J_1(\boldsymbol{x}) &= \int E[G^2(Z, \boldsymbol{u})]f(\boldsymbol{u})\left[\frac{1}{h_n^d}\tilde{K}^2\left(\frac{\boldsymbol{u} - \boldsymbol{x}}{h}\right)\right]\mathrm{d}\boldsymbol{u} \\ &\to E[G^2(Z, \boldsymbol{x})]f(\boldsymbol{x})\int \tilde{K}^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}, \end{aligned} \tag{5.11}$$

since $EG^2(Z, \cdot)f(\cdot)$ is continuous at $\boldsymbol{x}$ in view of $\mathbb{C}_2$–$\mathbb{C}_3$. For $J_2(\boldsymbol{x})$ we have by (3.3) and $E(H_{l_1}(Z_i)H_{l_2}(Z_j)) = \delta_{l_1,l_2}R^{l_1}(|i-j|)l_1!$ that

$$E(G(Z_i, \boldsymbol{u})G(Z_j, \boldsymbol{v})) = \sum_{l=1}^{\infty}\frac{c_l(\boldsymbol{u})c_l(\boldsymbol{v})}{l!}R^l(|i-j|).$$

Consequently,

$$J_2(\boldsymbol{x}) = \frac{1}{n^2}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{\infty}\frac{R^l(|i-j|)}{l!}v_{n,l}^2(\boldsymbol{x}), \tag{5.12}$$

where

$$v_{n,l}(\boldsymbol{x}) = \int c_l(\boldsymbol{u})\tilde{K}_h(\boldsymbol{u} - \boldsymbol{x})f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}. \tag{5.13}$$

Since $\tilde{K}$ has compact support by $\mathbb{C}_1$, (5.9), and condition $\mathbb{C}_4$ implies that $r(\boldsymbol{u}) \geqslant r(\boldsymbol{x})$ for $\boldsymbol{u}$ close to $\boldsymbol{x}$, so that for large n in view of $h_n \to 0$,

$$J_2(\boldsymbol{x}) = \frac{1}{n^2} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \sum_{l=r}^{\infty} \frac{R^l(|i-j|)}{l!} v_{n,l}^2(\boldsymbol{x}) = J_{2,1}(\boldsymbol{x}) + J_{2,2}(\boldsymbol{x})$$

$$=: \frac{v_{n,r}^2(\boldsymbol{x})}{r!} \frac{1}{n^2} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} R^r(|i-j|) + \frac{1}{n^2} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \sum_{l=r+1}^{\infty} \frac{R^l(|i-j|)}{l!} v_{n,l}^2(\boldsymbol{x}). \tag{5.14}$$

Now by dominated convergence

$$v_{n,r}(\boldsymbol{x}) \to c_r(\boldsymbol{x}) f(\boldsymbol{x}) \int \tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \tag{5.15}$$

since $\tilde{K}$ has compact support and $c_r(\cdot) f(\cdot)$ is continuous at $\boldsymbol{x}$. Observe that continuity of $c_r(\cdot)$ at $\boldsymbol{x}$ follows from $\mathbb{C}_3$, since by Parseval's equality

$$E[G(Z_1,\boldsymbol{u}) - G(Z_1,\boldsymbol{x})]^2 = \sum_{l=1}^{\infty} \frac{[c_l(\boldsymbol{u}) - c_l(\boldsymbol{x})]^2}{l!} \geqslant \frac{[c_r(\boldsymbol{u}) - c_r(\boldsymbol{x})]^2}{r!}.$$

Moreover,

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} R^r(|i-j|) \sim \frac{2}{(1-r\alpha)(2-r\alpha)} \left(\frac{L(n)}{n^\alpha}\right)^r,$$

thus

$$J_{2,1}(\boldsymbol{x}) \sim \frac{2c_r^2(\boldsymbol{x}) f^2(\boldsymbol{x})}{r!(1-r\alpha)(2-r\alpha)} \left(\int \tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\right)^2 \left(\frac{L(n)}{n^\alpha}\right)^r.$$

For $J_{2,2}(\boldsymbol{x})$, using $R(i) \leqslant 1$ we have

$$J_{2,2}(\boldsymbol{x}) \leqslant \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} R^{r+1}(|i-j|) \sum_{l=r+1}^{\infty} \frac{v_{n,l}^2(\boldsymbol{x})}{l!}.$$

Note that $v_{n,l}(\boldsymbol{x})$ is the Fourier–Hermite coefficient in the expansion of $\int G(z,\boldsymbol{u}) \tilde{K}_h(\boldsymbol{u} - \boldsymbol{x}) f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$, so that by Parseval's theorem

$$\sum_{l=0}^{\infty} \frac{v_{n,l}^2(\boldsymbol{x})}{l!} = E\left(\int G(Z_1,\boldsymbol{u}) \tilde{K}_h(\boldsymbol{u} - \boldsymbol{x}) f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\right)^2$$

$$\leqslant \left(\int E^{1/2}[G^2(Z_1, \boldsymbol{x} + h_n \boldsymbol{u})] f(\boldsymbol{x} + h_n \boldsymbol{u}) |\tilde{K}(\boldsymbol{u})|\,\mathrm{d}\boldsymbol{u}\right)^2 < \infty \tag{5.16}$$

since $\tilde{K}$ has a compact support and $E[G^2(Z,\cdot)]$ and $f(\cdot)$ are bounded in the neighbourhood of $\boldsymbol{x}$ by $\mathbb{C}_1$–$\mathbb{C}_3$ (the second inequality in (5.16) follows by expanding the square and using the Cauchy–Schwarz inequality for the expectation of the product with respect to Z). Finally, using Karamata's theorem (cf. e.g. Resnick, 1987, p. 17)

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} R^{r+1}(|i-j|) = \mathrm{o}\left(\left(\frac{L(n)}{n^\alpha}\right)^r\right). \tag{5.17}$$

Thus

$$J_{2,2} = \mathrm{o}\left(\left(\frac{L(n)}{n}\right)^r\right). \tag{5.18}$$

The result for $\mathrm{Var}(W_n)$ now follows from (5.10), (5.11), (5.14)–(5.16). The theorem follows. □

Proof of part (b). Note that for symmetric K, $m_i = 0$ for $i = 1, \dots, d$, thus the second main term in (3.7a) disappears for $i, j \geqslant 1$. Consider $\tilde{K}$ for $\boldsymbol{a} = (a_0, \dots, a_d)^{\mathrm{T}}$ such that $a_0 = 0$. The only difference in the proof of (b) is taking into account higher order terms in $J_{2,1}(\boldsymbol{x})$ and $J_{2,2}(\boldsymbol{x})$ of (5.14). Writing $(c_r f)(\boldsymbol{x}+h_n\boldsymbol{u}) = c_r f(\boldsymbol{x}) + h_n \int_0^1 \boldsymbol{u}^{\mathrm{T}} \nabla(c_r f)(\boldsymbol{x}+h_n t\boldsymbol{u})\,\mathrm{d}t$ when $\nabla c_r f(\cdot)$ is continuous in the neighbourhood of $\boldsymbol{x}$, we have

$$v_{n,r}(\boldsymbol{x}) = c_r(\boldsymbol{x}) f(\boldsymbol{x}) \int \tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + h_n \int \boldsymbol{u}^{\mathrm{T}} \tilde{K}(\boldsymbol{u}) \int_0^1 \nabla(c_r f)(\boldsymbol{x} + t h_n \boldsymbol{u}))\,\mathrm{d}t\,\mathrm{d}\boldsymbol{u}$$

and dominated convergence implies

$$v_{n,r}(\boldsymbol{x}) = c_r(\boldsymbol{x}) f(\boldsymbol{x}) \int \tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + h_n(\nabla(c_r f)(\boldsymbol{x}))^{\mathrm{T}} \int \boldsymbol{u}\tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \mathrm{o}(h_n). \tag{5.19}$$

Observe that for our choice of $\boldsymbol{a}$, $\int \tilde{K} = 0$ and the first term in the expansion of $v_{n,r}(\boldsymbol{x})$ is 0.

Moreover, reasoning as in part (a) we have by (5.14), (5.16), and (5.17)

$$J_{2,2}(\boldsymbol{x}) = \mathrm{o}\left(\left(\frac{L(n)}{n^\alpha}\right)^r\right) E\left(\int G(Z_1, \boldsymbol{u}) \tilde{K}_h(\boldsymbol{u} - \boldsymbol{x}) f(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\right)^2.$$

Now using $\mathbb{C}_6$–$\mathbb{C}_7$ and expanding $f(\boldsymbol{x} + h_n\boldsymbol{u}) G(Z_1, \boldsymbol{x} + h_n\boldsymbol{u})$ in a first-order Taylor series around $\boldsymbol{x}$, the expected value above may be written as

$$h_n^2 E\left(\int \boldsymbol{u}^{\mathrm{T}} \nabla(Gf)(\boldsymbol{y}_{\boldsymbol{u}}) \tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\right)^2,$$

where $\boldsymbol{y}_{\boldsymbol{u}}$ is an intermediate point satisfying $\boldsymbol{x} - \boldsymbol{u}h_n \leqslant \boldsymbol{y}_{\boldsymbol{u}} \leqslant \boldsymbol{x} + \boldsymbol{u}h_n$. Using the Cauchy–Schwarz inequality as in (5.16), and using $\mathbb{C}_6$ and $\mathbb{C}_7$ yield that it is of the order $\mathcal{O}(h_n^2)$ and thus

$$J_{2,2}(\boldsymbol{x}) = h_n^2 \mathrm{o}\left(\left(\frac{L(n)}{n^\alpha}\right)^r\right). \quad \square \tag{5.20}$$

Proof of Theorem 3. (*Part* (a)) Observe that is enough to show that

$$\left(\frac{n^\alpha}{L(n)}\right)^{r/2} \boldsymbol{t}_n^* \xrightarrow{\mathscr{D}} \left(\frac{2}{r!(1 - r\alpha)(2 - r\alpha)}\right)^{1/2} c_r(\boldsymbol{x}) f(\boldsymbol{x}) \boldsymbol{m}\, \eta_r, \tag{5.21}$$

since in view of Theorem 1

$$\begin{aligned} &\boldsymbol{Q}_n(\widehat{\boldsymbol{b}} - \boldsymbol{b}) + \tfrac{1}{2} h_n^2 \boldsymbol{M}^{-1} \boldsymbol{B}\,\mathrm{vech}\,(2\boldsymbol{V} - \mathrm{diag}(\boldsymbol{V})) \\ &\quad = \frac{\boldsymbol{M}^{-1} \boldsymbol{t}_n^*(\boldsymbol{x})}{f(\boldsymbol{x})} (1 + \mathrm{o}_{\mathrm{P}}(1)) + \mathrm{o}_{\mathrm{P}}(h_n^2). \end{aligned} \tag{5.22}$$

This together with (5.21) and conditions $(n^\alpha/L(n))^r = \mathrm{o}(nh_n^d)$ and nh_n^{d+4} is bounded imply the result. In order to prove (5.21), we use the familiar Cramér–Wold device and consider the variable W_n given in (5.8). Write

$$W_n = (1/n)\sum_{i=1}^{n} U_{n,i}, \qquad U_{n,i} = G(Z_i, \boldsymbol{X}_i)\tilde{K}_h(\boldsymbol{X}_i - \boldsymbol{x}), \tag{5.23}$$

where $\tilde{K}(\boldsymbol{u}) = \boldsymbol{a}^{\mathrm{T}}\binom{1}{\boldsymbol{u}}K(\boldsymbol{u})$. Define

$$\xi_{n,i} = E(U_{n,i}|Z_i) = \int G(Z_i, \boldsymbol{u})f(\boldsymbol{u})\tilde{K}(\boldsymbol{u} - \boldsymbol{x})\,\mathrm{d}\boldsymbol{u},$$

which in turn is approximated by $\zeta_{n,i} = p_n(\boldsymbol{x})G(Z_i, \boldsymbol{x})$, where $p_n(\boldsymbol{x}) = \int f(\boldsymbol{u})\tilde{K}_h(\boldsymbol{u} - \boldsymbol{x})\,\mathrm{d}\boldsymbol{u}$. If we show that

$$J_1(\boldsymbol{x}) := E\left(W_n - \frac{1}{n}\sum_{i=1}^{n}\xi_{n,i}\right)^2 = \mathcal{O}\left(\frac{1}{nh_n^d}\right) \tag{5.24}$$

and

$$J_2(\boldsymbol{x}) := E\left(\frac{1}{n}\sum_{i=1}^{n}\xi_{n,i} - \frac{1}{n}\sum_{i=1}^{n}\zeta_{n,i}\right)^2 = \mathrm{o}\left(\left(\frac{L(n)}{n^{\alpha}}\right)^r\right), \tag{5.25}$$

then

$$W_n = \frac{p_n(\boldsymbol{x})}{n}\sum_{i=1}^{n} G(Z_i, \boldsymbol{x}) + \mathcal{O}_{\mathrm{P}}\left(\left(\frac{1}{nh_n^d}\right)^{1/2}\right) + \mathrm{o}_{\mathrm{P}}\left(\left(\frac{L(n)}{n^{\alpha}}\right)^{r/2}\right). \tag{5.26}$$

By dominated convergence $p_n(\boldsymbol{x}) \to f(\boldsymbol{x})\int\tilde{K}$ at continuity points of $f(\cdot)$. Now using (3.5) and the condition $(n^{\alpha}/L(n))^r = \mathrm{o}(nh_n^d)$ we have by (5.26)

$$\left(\frac{n^{\alpha}}{L(n)}\right)^r W_n \xrightarrow{\mathcal{D}} \left(\frac{2}{r!(1 - r\alpha)(2 - r\alpha)}\right)^{1/2} c_r(\boldsymbol{x})f(\boldsymbol{x})\int\tilde{K}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\eta_r.$$

Thus it remains to show (5.24) and (5.25). This can be proved analogously to proof of Theorem 1 in Csörgő and Mielniczuk (1999) observing that their result actually holds for an arbitrary kernel K satisfying $\mathbb{C}_1$ without the restriction that K is positive or is a a product of univariate kernels. More specifically, the kernel $\tilde{K}(\boldsymbol{u})$ replaces the kernel $K(\boldsymbol{u})$ in Csörgő and Mielniczuk (1999) and condition $\mathbb{C}_3$ replaces the second part of condition $\mathbb{C}_5$ in that paper. In particular, for the term $J_2(\boldsymbol{x})$ we have (using condition $\mathbb{C}_4$ and the compactness of the support of K)

$$J_2(\boldsymbol{x}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=r}^{\infty}\frac{R^l(|i-j|)}{l!}v_{n,l}^2(\boldsymbol{x})$$

with $v_{n,l}(\boldsymbol{x}) = \int[c_l(\boldsymbol{u}) - c_l(\boldsymbol{x})]f(\boldsymbol{u})\tilde{K}_h(\boldsymbol{u} - \boldsymbol{x})\,\mathrm{d}\boldsymbol{u}$. We have

$$J_2(\boldsymbol{x}) \leqslant \frac{1}{n^2}\sum_{l=0}^{\infty}\frac{v_{n,l}^2(\boldsymbol{x})}{l!}\sum_{i=1}^{n}\sum_{j=1}^{n}R^r(|i-j|) = \frac{1}{n^2}\sum_{l=0}^{\infty}\frac{v_{n,l}^2(\boldsymbol{x})}{l!}\mathcal{O}(n^{2-r\alpha}L^r(n)).$$

By Parseval's equality

$$\sum_{l=0}^{\infty}\frac{v_{n,l}^2(\boldsymbol{x})}{l!} = E\left\{\left[\int(G(Z,\boldsymbol{u}) - G(Z,\boldsymbol{x}))f(\boldsymbol{u})\tilde{K}_h(\boldsymbol{u} - \boldsymbol{x})\,\mathrm{d}\boldsymbol{u}\right]^2\right\}$$

and in view of Cauchy–Schwarz inequality (as in (5.16)), the last quantity is bounded by

$$\left\{\int (E(G(Z,\boldsymbol{u})-G(Z,\boldsymbol{x}))^2)^{1/2} f(\boldsymbol{u})|\tilde{K}_h(\boldsymbol{u}-\boldsymbol{x})|\,\mathrm{d}u\right\}^2$$

$$=\left\{\int (E(G(Z,\boldsymbol{x}+h_n\boldsymbol{u})-G(Z,\boldsymbol{x}))^2)^{1/2} f(\boldsymbol{x}+h_n\boldsymbol{u})|\tilde{K}(\boldsymbol{u})|\,\mathrm{d}u\right\}^2 \tag{5.27}$$

which in view of compactness of the support of $\tilde{K}$, $\mathbb{C}_2$ and $\mathbb{C}_3$ tends to 0, so that $J_2(\boldsymbol{x})$ satisfies (5.25).

Proof of part (b). Let $\boldsymbol{t}_{n1}^*$ denote the subvector of $\boldsymbol{t}_n^*$ consisting of its last d components. Observe that since K is symmetric, (5.22) implies that

$$\widehat{\boldsymbol{b}}_1-\boldsymbol{b}_1=\frac{\boldsymbol{M}_2^{-1}}{h_n}\frac{\boldsymbol{t}_{n1}^*}{f(\boldsymbol{x})}(1+\mathrm{o_P}(1))+\mathcal{O}(h_n)+\mathrm{o_P}(h_n).$$

Moreover, $(n^\alpha/L(n))^{r/2}h_n=\mathrm{o}(1)$ in view of condition (3.11b) and the fact that nh_n^{d+4} is bounded. Thus it suffices to show that

$$\frac{1}{h_n}\left(\frac{n^\alpha}{L(n)}\right)^{r/2}\boldsymbol{t}_{n1}^* \xrightarrow{\mathscr{D}} \left(\frac{2}{r!(1-r\alpha)(2-r\alpha)}\right)^{1/2}\boldsymbol{\Lambda}\nabla(c_rf)(\boldsymbol{x}).$$

where $\boldsymbol{\Lambda}=\mathrm{diag}(m_{11},\dots,m_{dd})$, since for symmetric K, $\boldsymbol{\Lambda}=\boldsymbol{M}_2$. As in part (a) we use the Cramér–Wold device to define W_n of (5.23) with $\boldsymbol{a}\in\mathbb{R}^{d+1}$ such that $a_0=0$. Observe that $\int\tilde{K}=0$ for such $\boldsymbol{a}$. We get from (5.24) in view of the condition (3.11b)

$$\left(\frac{n^\alpha}{L(n)}\right)^r\frac{1}{h_n^2}J_1(\boldsymbol{x})=\mathcal{O}\left(\left(\frac{n^\alpha}{L(n)}\right)^r\right)\frac{1}{nh_n^{d+2}}=\mathrm{o}(1).$$

Thus it suffices to find the limit of

$$\left(\frac{n^\alpha}{L(n)}\right)^{r/2}\frac{1}{nh_n}\sum_{i=1}^n\xi_{ni},$$

where

$$\xi_{ni}=E(U_{ni}|Z_i)=\int G(Z_i,\boldsymbol{u})f(\boldsymbol{u})\tilde{K}_h(\boldsymbol{u}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{u}.$$

Consider

$$\bar{\xi}_{ni}=\frac{H_r(Z_i)}{r!}\int c_r(\boldsymbol{u})f(\boldsymbol{u})\tilde{K}_h(\boldsymbol{u}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{u}.$$

Using a first-order series expansion of $(c_rf)(\boldsymbol{x}+h_n\boldsymbol{u})$ around $\boldsymbol{x}$, and $\int\tilde{K}=0$ it follows as in the proof of Theorem 2b ((5.19)) that

$$\frac{1}{h_n}\int c_r(\boldsymbol{u})f(\boldsymbol{u})\tilde{K}_h(\boldsymbol{u}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{u}\;\to\;(\nabla(c_rf)(\boldsymbol{x}))^{\mathrm{T}}\int\boldsymbol{u}\tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}.$$

Thus part (b) follows from (3.5) once we prove

$$E\left(\frac{1}{n}\sum_{i=1}^n\xi_{ni}-\frac{1}{n}\sum_{i=1}^n\bar{\xi}_{ni}\right)^2=\mathrm{o}\left(h_n^2\left(\frac{L(n)}{n^\alpha}\right)^r\right).$$

However, it is easy to see that the term on the left-hand side of the equation above is equal to $J_{2,2}(\boldsymbol{x})$ of (5.14) which was shown in the proof of Theorem 2b to be $\mathrm{o}(h_n^2(L(n)/n^\alpha)^r)$ (see (5.20)) provided $\int\tilde{K}=0$. □

Proof of Theorem 4(a). We apply the Cramér–Wold device once again. With W_n of (5.8) we have

$$W_n = \frac{1}{n}\sum_{i=1}^{n} U_{n,i}, \qquad U_{n,i} = G(Z_i, \boldsymbol{X}_i)\tilde{K}_h(\boldsymbol{X}_i - \boldsymbol{x}).$$

$EW_n = 0$ and by proof of Theorem 2, $nh_n^d \operatorname{Var}(W_n) \to \theta^2(\boldsymbol{x})$, where

$$\theta^2(\boldsymbol{x}) = E(G^2(Z,\boldsymbol{x}))f(\boldsymbol{x})\int \tilde{K}^2(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}.$$

Suppose we show that

$$(nh_n^d)^{1/2} W_n \xrightarrow{\mathscr{D}} \mathrm{N}(0, \theta^2(\boldsymbol{x})). \tag{5.28}$$

Then since $W_n = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{t}_n^*$ for arbitrary $\boldsymbol{a} \in \mathbb{R}^{d+1}$

$$(nh_n^d)^{1/2}\boldsymbol{t}_n^* \xrightarrow{\mathscr{D}} \mathrm{N}(0, \boldsymbol{\Sigma}(\boldsymbol{x})), \tag{5.29}$$

where $\boldsymbol{\Sigma}(\boldsymbol{x}) = E(G^2(Z,\boldsymbol{x}))f(\boldsymbol{x})\Gamma$. Using Theorem 1 together with the condition $nh_n^{d+4} \to 0$ we obtain that $(nh_n^d)^{1/2}\boldsymbol{Q}_n(\widehat{\boldsymbol{b}} - \boldsymbol{b})$ has the same asymptotic distribution as that of $(nh_n^d)^{1/2}\boldsymbol{M}^{-1}\boldsymbol{t}_n^*/f(\boldsymbol{x})$. Thus the result follows from (5.29). It remains to establish (5.28). It is seen that

$$V_n := (nh_n^d)^{1/2} W_n = \frac{(nh_n^d)^{1/2}}{n}\sum_{i=1}^{n} G(Z_i, \boldsymbol{X}_i)\tilde{K}_h(\boldsymbol{X}_i - \boldsymbol{x}) \tag{5.30}$$

is, up to the scaling factor, equal to S_n in the proof of Theorem 2 in Csörgő and Mielniczuk (1999) except that here $\tilde{K}(\boldsymbol{u}) = \boldsymbol{a}^{\mathrm{T}}\binom{1}{\boldsymbol{u}}K(\boldsymbol{u})$. Thus, if we define

$$\mu_n = E(V_n | Z_1, \ldots, Z_n) \tag{5.31}$$

and $\sigma_n^2(\boldsymbol{x}) = \operatorname{var}(V_n | Z_1, \ldots, Z_n)$, then it suffices to show that

$$\mu_n \xrightarrow{\mathscr{P}} 0, \tag{5.32}$$

$$\sigma_n^2(\boldsymbol{x}) \xrightarrow{\mathscr{P}} \theta^2(\boldsymbol{x}) \tag{5.33}$$

and

$$\frac{V_n - \mu_n}{\sigma_n} \xrightarrow{\mathscr{D}} \mathrm{N}(0,1). \tag{5.34}$$

The proof of (5.32) in Csörgő and Mielniczuk (1999) is not applicable here because it would require that $\int \boldsymbol{u}\tilde{K}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \boldsymbol{0}$ for every $a \in \mathbb{R}^{d+1}$, which in turn requires $\int \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}K(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \boldsymbol{0}$ which cannot be satisfied for nonnegative kernel K. We proceed here by using the decomposition

$$\mu_n =: \mu_n' + \mu_n'' = \frac{(nh_n^d)^{1/2}}{n}\left\{\int \tilde{K}_h(\boldsymbol{s} - \boldsymbol{x})f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\sum_{i=1}^{n} G(Z_i, \boldsymbol{x}) + \sum_{i=1}^{n}\int (G(Z_i, \boldsymbol{s}) - G(Z_i, \boldsymbol{x}))f(\boldsymbol{s})\tilde{K}_h(\boldsymbol{s} - \boldsymbol{x})\,\mathrm{d}\boldsymbol{s}\right\}. \tag{5.35}$$

Observe that

$$\mu_n' = \frac{(nh_n^d)^{1/2}}{(n^\alpha/L(n))^{r/2}} \left(\frac{n^\alpha}{L(n)}\right)^{r/2} \frac{1}{n}\sum_{i=1}^n G(Z_i,\boldsymbol{x}) \int \tilde{K}_h(\boldsymbol{s}-\boldsymbol{x}) f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} \xrightarrow{\mathrm{P}} 0 \tag{5.36}$$

in view of the fact that by (3.5) $(n^\alpha/L(n))^{r/2}\frac{1}{n}\sum_{i=1}^n G(Z_i,\boldsymbol{x}) = \mathcal{O}_{\mathrm{P}}(1)$ and applying Bochner's lemma to $\int f(\boldsymbol{s})\tilde{K}_h(\boldsymbol{s}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{s}$. Now observe that

$$E(\mu_n'')^2 = \frac{nh_n^d}{n^2} E\left(\sum_{i=1}^n \int \{G(Z_i,\boldsymbol{s}) - G(Z_i,\boldsymbol{x})\} f(\boldsymbol{s})\tilde{K}_h(\boldsymbol{s}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{s}\right)^2 = nh_n^d J_2(\boldsymbol{x}), \tag{5.37a}$$

where $J_2(\boldsymbol{x})$ is defined by (5.25) in the proof of Theorem 3(a). Since from (5.25) $J_2(\boldsymbol{x}) = \mathrm{o}((L(n)/n^\alpha)^r)$ then

$$E(\mu_n'')^2 = (nh_n^d)\mathcal{O}((L(n)/n^\alpha)^r) = \mathrm{o}(1). \tag{5.37b}$$

Thus (5.32) follows from (5.35)–(5.37). Consider now (5.33). In view of the independence of the sequence $(X_i)_{i=1}^n$ we have

$$\sigma_n^2(\boldsymbol{x}) = \frac{h_n^d}{n}\sum_{k=1}^n \int G^2(Z_k,\boldsymbol{s})\tilde{K}_h^2(\boldsymbol{s}-\boldsymbol{x}) f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}$$
$$-\frac{h_n^d}{n}\sum_{k=1}^n \left(\int G(Z_k,\boldsymbol{s})\tilde{K}_h(\boldsymbol{s}-\boldsymbol{x}) f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right)^2 = I_1(\boldsymbol{x}) + I_2(\boldsymbol{x}). \tag{5.38}$$

Writing $(\int G(Z_1,s)\tilde{K}_h(\boldsymbol{s}-\boldsymbol{x}) f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s})^2$ as a double integral, bringing the expectation inside and applying the Cauchy–Schwarz inequality to the expectation we obtain

$$E\left(\int G(Z_1,\boldsymbol{s})\tilde{K}_h(\boldsymbol{s}-\boldsymbol{x}) f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right)^2 \leqslant \left(\int (EG^2(Z_1,\boldsymbol{s}))^{1/2}|\tilde{K}_h(\boldsymbol{s}-\boldsymbol{x})| f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right)^2 < \infty$$

in view of $\mathbb{C}_1$–$\mathbb{C}_3$. Thus $I_2(\boldsymbol{x}) = \mathcal{O}(h_n^d) = \mathrm{o}_{\mathrm{P}}(1)$. Csörgő and Mielniczuk (1999) prove that $I_1(\boldsymbol{x})$ converges to $\theta^2(\boldsymbol{x})$ with probability 1 under conditions $\mathbb{C}_6 - \mathbb{C}_7$. The proof of (5.34) remains the same as in Csörgő and Mielniczuk (1999) with a change of one bound, namely the term W_n^* in Csörgő and Mielniczuk (1999) is identical to $-I_2(\boldsymbol{x})$ which is $\mathrm{o}_{\mathrm{P}}(1)$ by the argument given above.

Proof of (b). Consider $\tilde{K}$ for $\boldsymbol{a}$ such that $a_0 = 0$. Observe that in the proof of Part (a) condition (3.10a) was used only to verify that $\mu_n \xrightarrow{\mathrm{P}} 0$. Thus it suffices to verify that for our choice of $\tilde{K}$ with K symmetric $\mu_n \xrightarrow{\mathrm{P}} 0$ under the weaker condition (3.10b). This readily follows from (5.36) and (5.25). Namely, under $\mathbb{C}_6$ and $\int \tilde{K} = 0$, $\int \tilde{K}_h(\boldsymbol{s} - \boldsymbol{x}) f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = \mathcal{O}(h_n)$ and thus

$$\mu_n' = \mathcal{O}\left(\frac{(nh_n^d)^{1/2}h_n}{(n^\alpha/L(n))^{r/2}}\right) = \mathrm{o}(1).$$

For μ_n'' of (5.37a), we bound $J_2(\boldsymbol{x})$ as in the proof of Theorem 3(b) except that in (5.27) we expand $G(Z_1,\boldsymbol{x}+h_n\boldsymbol{u}) - G(Z_1,\boldsymbol{x})$ in a first-order Taylor series and using $\mathbb{C}_7$ it can be seen that (5.27) is of order $\mathcal{O}(h_n^2)$ (rather than o(1)) and thus

$$E[\mu_n''(\boldsymbol{x})]^2 = \mathcal{O}(nh_n^{d+2})\mathcal{O}\left(\left(\frac{L(n)}{n}\right)^r\right) = \mathrm{o}(1). \qquad \square$$

Proof of Theorem 5(a). With V_n and μ_n given by (5.30) and (5.31), respectively, let $I_n(\boldsymbol{x}) := V_n - \mu_n(\boldsymbol{x})$. Decomposing $\mu_n(\boldsymbol{x}) = \mu_n'(\boldsymbol{x}) + \mu_n''(\boldsymbol{x})$ as in (5.35) and noting that under condition $\mathbb{C}_3$ we have $E[G(Z,\boldsymbol{s}) - G(Z,\boldsymbol{x})]^2 \to 0$ as $\boldsymbol{s} \to \boldsymbol{x}$ which together with $nh_n^d \sim (n^\alpha/L(n))^r$ implies that $\mu_n''(\boldsymbol{x}) \xrightarrow{\text{P}} 0$ (see the proof of Theorem 4 for $\mu_n''(\boldsymbol{x})$). Hence

$$V_n = (nb_n^d)^{1/2} \boldsymbol{a}^{\text{T}} \boldsymbol{t}_n^* = I_n(\boldsymbol{x}) + \mu_n'(\boldsymbol{x}) + o_{\text{P}}(1).$$

Also, under the borderline condition we have by (3.5) that

$$\mu_n'(\boldsymbol{x}) \xrightarrow{\mathscr{D}} C_b \left(\frac{2}{r!(1 - r\alpha)(2 - r\alpha)} \right)^{1/2} c_r(\boldsymbol{x}) f(\boldsymbol{x}) \int \tilde{K}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \eta_r \tag{5.39}$$

noting that by dominated convergence $\int \tilde{K}_h(\boldsymbol{x}-\boldsymbol{u}) f \boldsymbol{u}\, \mathrm{d}\boldsymbol{u} \to f(\boldsymbol{x}) \int \tilde{K}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}$. Denote the characteristic function of the limit of $\mu_n'(\cdot)$ by $\psi(\cdot, \boldsymbol{x})$ and observe that μ_n' is measurable with respect to $\sigma(Z_1, \ldots, Z_n)$. Now by the proof of Theorem 2 in Csörgő and Mielniczuk (1999)

$$\phi_n(s, \boldsymbol{x}) := E[\mathrm{e}^{\mathrm{i}sI_n(\boldsymbol{x})} | Z_1, \ldots, Z_n]] \xrightarrow{\text{P}} \phi(s, \boldsymbol{x}), \tag{5.40}$$

where $\phi(s, \boldsymbol{x})$ is the characteristic function of $\text{N}(0, \theta^2(\boldsymbol{x}))$. We then have

$$E[\mathrm{e}^{\mathrm{i}(sI_n(\boldsymbol{x}) + t\mu_n'(\boldsymbol{x}))}] = E[\mathrm{e}^{\mathrm{i}t\mu_n'(\boldsymbol{x})} E[\mathrm{e}^{\mathrm{i}sI_n(\boldsymbol{x})} | Z_1, \ldots, Z_n]] = E[\phi_n(s, \boldsymbol{x}) \mathrm{e}^{\mathrm{i}t\mu_n'(\boldsymbol{x})}].$$

Hence

$$\begin{aligned} &|E[\phi_n(s, \boldsymbol{x}) \mathrm{e}^{\mathrm{i}t\mu_n'(\boldsymbol{x})}] - \phi(s, \boldsymbol{x}) \psi(t, \boldsymbol{x})| \\ &\quad \leqslant |E[(\phi_n(s, \boldsymbol{x}) - \phi(s, \boldsymbol{x})) \mathrm{e}^{\mathrm{i}t\mu_n'(\boldsymbol{x})}]| + |\phi(s, \boldsymbol{x}) E[\mathrm{e}^{\mathrm{i}t\mu_n'(\boldsymbol{x})} - \psi(t, \boldsymbol{x})]| \\ &\quad \leqslant E|\phi_n(s, \boldsymbol{x}) - \phi(s, \boldsymbol{x})| + |E[\mathrm{e}^{\mathrm{i}t\mu_n'(\boldsymbol{x})}] - \psi(t, \boldsymbol{x})| \to 0 \end{aligned}$$

since $\phi_n(s, \boldsymbol{x}) \xrightarrow{\text{P}} \phi(s, \boldsymbol{x})$, $E[\mathrm{e}^{\mathrm{i}t\mu_n(\boldsymbol{x})}] \to \psi(t, \boldsymbol{x})$ and $\phi_n(\boldsymbol{s}, \boldsymbol{x})$ is uniformly bounded in n. Thus the limiting law has independent marginals and the result follows by (5.39) and (5.40) from Theorem 1.

Proof of Theorem 5(b) is analogous. □

References

Castellana, J.V., Leadbetter, M.R., 1986. On smoothed probability density estimation. Stochastic Process. Appl. 21, 179–193.

Cheng, B., Robinson, P.M., 1994. Semiparametric estimation from time series with long-range dependence. J. Econometrics 64, 335–353.

Csörgő, S., Mielniczuk, J., 1995. Density estimation under long-range dependence. Ann. Statist. 23, 990–999.

Csörgő, S., Mielniczuk, J., 1998. The smoothing dichotomy in random-design regression with long-memory errors based on moving averages. Technical Report #316. Dept. Stat., University of Michigan.

Csörgő, S., Mielniczuk, J., 1999. Random-design regression under long-range dependent errors. Bernoulli 5, 209–224.

Dobrushin, R.L., Major, P., 1979. Non-central limit theorems for non-linear functionals of Gaussian fields. Z. Wahrscheinlichkeitstheorie verw. Gebiete 50, 27–52.

Fan, J., 1992. Design-adaptive nonparametric regression. J. Amer. Statist. Assoc. 87, 998–1004.

Fan, J., 1993. Local linear regression smoothers and their minimax efficiency. Ann. Statist. 21, 196–216.

Fan, J., Gijbels, I., 1992. Variable bandwidth and local linear regression smoothers. Ann. Statist. 20, 2008–2036.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman and Hall, London.

Fan, J., Masry, E., 1992. Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. J. Multivariate Anal. 43, 237–271.

Hidalgo, J., 1997. Nonparametric estimation with strongly dependent multivariate time series. J. Time Ser. Anal. 18, 95–122.

Koul, H.L., 1992. M-estimators in linear models with long range dependent errors. Statist. Probab. Lett. 14, 153–164.

Koul, H.L., Mukherjee, K., 1993. Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors. Probab. Theory Related Fields 95, 535–553.

Masry, E., 1996a. Multivariate local polynomial regression for time series: uniform consistency and rates. J. Time Ser. Anal. 17, 571–599.

Masry, E., 1996b. Multivariate regression estimation: local polynomial fitting for time series. J. Stochastic Processes Appl. 65, 81–101.

Resnick, S., 1987. Extreme Values, Regular Variation, and Point Processes. Springer, New York.

Robinson, P.M., 1983. Nonparametric estimators for time series. J. Time Ser. Anal. 4, 81–101.

Roussas, G.G., 1990. Nonparametric regression under mixing conditions. Stochastic Processes Appl. 36, 107–116.

Ruppert, D., Wand, M.P., 1994. Multivariate weighted least squares regression. Ann. Statist. 22, 107–116.

Taqqu, M.S., 1979. Convergence of integrated processes of arbitrary Hermite rank. Z. Wahrscheinlichkeitstheorie verw. Gebiete 50, 53–83.

Truong, Y.K., Stone, C.J., 1992. Nonparametric estimation involving time series. Ann. Statist. 20, 77–97.