Asymptotic properties of local polynomial regression with missing data and correlated errors

A. Pérez-González · J. M. Vilar-Fernández · W. González-Manteiga

Abstract The main objective of this work is the nonparametric estimation of the regression function with correlated errors when observations are missing in the response variable. Two nonparametric estimators of the regression function are proposed. The asymptotic properties of these estimators are studied; expresions for the bias and the variance are obtained and the joint asymptotic normality is established. A simulation study is also included.

Keywords Local polynomial regression · Missing response and correlated errors

1 Introduction

The local polynomial fitting is a attractive technique used for estimating the regression function. Many authors have studied the asymptotic properties of the local polynomial fitting in a context of dependence (Masry 1996a,b; Masry and Fan 1997;

A. Pérez-González (✉)
Department of Statistics and Operations Research, Faculty of Business Sciences,
University of Vigo, Campus South, Ourense, Spain
e-mail: anapg@uvigo.es

J. M. Vilar-Fernández
Department of Mathematics, Faculty of Computer Science, University of A Coruña,
Campus de Elviña, A Coruña, Spain
e-mail: eijvilar@udc.es

W. González-Manteiga
Department of Statistics and Operations Research, Faculty of Mathematics,
University of Santiago de Compostela, Campus South, Santiago de Compostela, Spain
e-mail: wenceslao@usc.es

Härdle and Tsybakov 1997; Härdle et al. 1998; Vilar and Vilar 1998, etc.). A broad study of the local polynomial modelling can be found in Fan and Gijbels (1996).

Most of the statistical methods are designed for complete data sets and problems arise when missing observations are present, which is a common situation in biomedical, environmental or socioeconomic studies, for example. Classic examples are found in the field of social sciences with the problem of non-response in sample surveys, in Physics, in Genetics (Meng 2000), etc.

In the regression context, a common method is to impute the incomplete observations and then proceed to carry out the estimation of the conditional or unconditional mean of the response variable with the completed sample. The methods considered include linear regression (Yates 1933), kernel smoothing (Cheng 1994; Chu and Cheng 1995; González-Manteiga and Pérez-González 2004), nearest neighbor imputation (Chen and Shao 2000), semiparametric estimation (Wang et al. 2004), nonparametric multiple imputation (Aerts et al. 2002), empirical likelihood over the imputed values (Wang and Rao 2002), etc.

For dependent data, the problem of missing observations has been studied using various techniques like the likelihood estimation (Peña and Tiao 1991; Jones 1980), least squares (Beveridge 1992; Chow and Lin 1976; etc.) and kernel estimation (Robinson 1984), among others.

The objective of this paper is to introduce a nonparametric estimator of the regression function with correlated errors when observations are missing in the response variable. The observations can be missed for various reasons, for example in a time series: the fault of an equipment of measurement, the inability of to observe the series at some instants, for example on holidays or due to a tempest, etc.

In this paper, we propose two nonparametric estimators when we have missing observations in the response variable based on the local polynomial estimator for complete data studied by Francisco-Fernández and Vilar-Fernández (2001) for fixed design and correlated errors.

The first one is the *simplified local polynomial smoother* (*SLPS*), which only uses complete observations. The second one is based on the techniques of simple imputation already used by Chu and Cheng (1995) or González-Manteiga and Pérez-González (2004). This estimator, that we will refer to as *imputed local polynomial smoother* (*ILPS*), consists in using *SLPS* to estimate the missing observations of the response variable Y; then, the local polynomial estimator for complete data is applied to the completed sample.

Let us consider the fixed regression model where the functional relationship between the design points, $x_{t,n}$, and the responses, $Y_{t,n}$, can be expressed as

$$Y_{t,n} = m(x_{t,n}) + \varepsilon_{t,n}, \quad 1 \le t \le n,$$

where $m(\cdot)$ is a regression function defined in $[0, 1]$, without any loss of generality, and $\varepsilon_{t,n}$, $1 \le t \le n$, are unobserved random variables with zero mean and finite variance, σ_ε^2. We assume, for each n, that $\{\varepsilon_{1,n}, \varepsilon_{2,n}, \ldots, \varepsilon_{n,n}\}$ have the same joint distribution as $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$, where $\{\epsilon_t,\ t \in \mathbb{Z}\}$ is a strictly stationary stochastic process. In this way, it is assumed that the errors of the model can be in general dependent.

The design points $x_{t,n}$, $1 \leq t \leq n$, follow a regular design generated by a density f. So, for each n, the design points are defined by

$$\int_0^{x_{t,n}} f(x)\mathrm{d}(x) = \frac{t-1}{n-1}, \quad 1 \leq t \leq n,$$

where f is a positive function, defined in [0, 1] and its first derivative is continuous.

For simplicity, we are going to avoid the subindex n in the sample data and in the errors notation, that is, we are going to use x_t, Y_t and ε_t.

The response variable Y can have missing data. To check whether an observation is complete ((x_t, Y_t) $\in \mathbb{R}^2$) or not ((x_t, ?)), a new variable δ is introduced into the model as an indicator of the missing observations. Thus, $\delta_t = 1$ if Y_t is observed, and zero if Y_t is missing for $t = 1, \ldots, n$.

Following the patterns in the literature (see Little and Rubin (1987), etc), we need to establish whether the loss of an item of data is independent or not of the value of the observed data and/or the missing data. In this paper we suppose that the data are missing at random (MAR), i.e.:

$$P(\delta_t = 1/Y_t, x_t) = P(\delta_t = 1/x_t) = \mathfrak{p}(x_t), \tag{1}$$

$\mathfrak{p}$ being a positive function, defined on [0, 1] and its first derivative is continuous. We suppose that the variables δ_t are independent.

In the next section we present the regression model with missing data, as well as the nonparametric estimators used. The mean squared error and the asymptotic distribution of the estimators are shown in Sect. 3. In Sect. 4, a simulation study is presented. The conclusions are shown in Sect. 5. And, finally Sect. 6 contains the proofs of the asymptotic results.

2 The regression model and the nonparametric estimators

Our goal is to estimate the unknown regression function $m(x)$ and its derivatives using weighted local polynomial fitting. We assume that the $(p+1)$th derivative of the regression function at point x exist and are continuous. As indicated previously, two nonparametric estimators are studied, the first (Simplified estimator) arises from using only complete observations and is a generalization of the local polynomial estimator. If we assume that the $(p+1)$th derivatives of the regression function at point x exist and are continuous, the parameter vector $\boldsymbol{\beta}(x) = (\beta_0(x), \beta_1(x), \ldots, \beta_p(x))^{\mathrm{T}}$, where $\beta_j(x) = m^{(j)}(x)/(j!)$, with $j = 0, 1, \ldots, p$, can be estimated by minimizing the function

$$\Psi(\boldsymbol{\beta}(x)) = (\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta}(x))^{\mathrm{T}} \mathbf{W}_n^{\delta} (\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta}(x)),$$

where

$$\mathbf{Y}_n = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X}_n = \begin{pmatrix} 1 & (x_1 - x) & \cdots & (x_1 - x)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_n - x) & \cdots & (x_n - x)^p \end{pmatrix},$$

$$\mathbf{W}_n^\delta = \mathrm{diag}\left(n^{-1} K_{h_n}(x_1 - x)\delta_1, .., n^{-1} K_{h_n}(x_n - x)\delta_n\right)$$

with $K_{h_n}(u) = h_n^{-1} K\left(h_n^{-1} u\right)$, K being a kernel function and h_n the bandwidth or smoothing parameter that controls the size of the local neighborhood and so the degree of smoothing.

Assuming the invertibility of $\mathbf{X}_n^t \mathbf{W}_n^\delta \mathbf{X}_n$, the estimator is

$$\hat{\boldsymbol{\beta}}_{S,n}(x) = \left(\mathbf{X}_n^{\mathrm{T}} \mathbf{W}_n^\delta \mathbf{X}_n\right)^{-1} \mathbf{X}_n^{\mathrm{T}} \mathbf{W}_n^\delta \mathbf{Y}_n = \mathbf{S}_n^{-1} \mathbf{T}_n, \tag{2}$$

where $\mathbf{S}_n$ is the array $(p+1) \times (p+1)$ whose (i, j)th element is $s_{i,j,n} = s_{i+j-2,n}$ with

$$s_{k,n} = \frac{1}{n} \sum_{t=1}^{n} (x_t - x)^k K_{h_n}(x_t - x)\delta_t, \quad 0 \le k \le 2p$$

and $\mathbf{T}_n = \left(t_{0,n}, t_{1,n}, \ldots, t_{p,n}\right)^t$, being

$$t_{i,n} = \frac{1}{n} \sum_{t=1}^{n} (x_t - x)^i K_{h_n}(x_t - x) Y_t \delta_t, \quad 0 \le i \le p. \tag{3}$$

A second estimator (imputed estimator) is computed in two steps. In the first step, the Simplified estimator with degree q, kernel L and smoothing parameter g_n, is used to estimate the missing observations. In this way, the sample $\left\{\left(x_t, \hat{Y}_t\right)\right\}_{t=1}^{n}$ is completed, where $\hat{Y}_t = \delta_t Y_t + (1 - \delta_t)\widehat{m}_{S,g_n}(x_t)$, with $\hat{m}_{S,g_n}(x) = \mathbf{e}_1^t \hat{\boldsymbol{\beta}}_{S,n}(x)$ and $\mathbf{e}_j$ is the $(p+1) \times 1$ dimensional vector with 1 at the jth coordinate and zero at the rest. Now, the simplified estimation is applied to the data $\left\{\left(x_t, \hat{Y}_t\right)\right\}_{t=1}^{n}$ with degree $p(p \ge q)$, kernel K and smoothing parameter h_n. The expression of this estimator is

$$\hat{\boldsymbol{\beta}}_{I,n}(x) = \left(\mathbf{X}_n^{\mathrm{T}} \mathbf{W}_n \mathbf{X}_n\right)^{-1} \mathbf{X}_n^{\mathrm{T}} \mathbf{W}_n \hat{\mathbf{Y}}_n = \mathbf{U}_n^{-1} \mathbf{V}_n, \tag{4}$$

where $\widehat{\mathbf{Y}} = \left(\widehat{Y}_1, .., \widehat{Y}_n\right)^{\mathrm{T}}$, $\mathbf{W}_n = \mathrm{diag}\left(n^{-1} K_{h_n}(x_1 - x), .., n^{-1} K_{h_n}(x_n - x)\right)$.

3 Asymptotic properties

In this section asymptotic expressions for the bias and variance/covariance array and the asymptotic normality of the estimate defined in (2) and (4) are obtained. The following assumptions will be needed in our analysis:

A.1. Kernel functions K and L are symmetric, with bounded support, and Lipschitz continuous.

A.2. The sequence of smoothing parameters, $\{l_n\}$, with l_n satisfies that $l_n > 0$, $l_n \downarrow 0$, $nl_n \uparrow \infty$, where $l_n = h_n$ or g_n.

A.3. Denote $\mathrm{Cov}\,(\epsilon_i, \epsilon_{i+k}) = \sigma_\varepsilon^2\, v\,(k)\,,\ k = 0, 1, 2 \ldots$ then $\sum_{k=1}^\infty k\,|v(k)| < \infty$.

3.1 Asymptotic properties of the simplified estimator

The following notations will be used. Let $\mu_{K,j} = \int u^j K(u)\,\mathrm{d}u$ and $v_{K,j} = \int u^j K^2(u)\,\mathrm{d}u$ and let us denote $\boldsymbol{\mu}_K = \left(\mu_{K,p+1}, \ldots, \mu_{K,2p+1}\right)^{\mathrm{T}}$ and $\mathbf{S}_K$ and $\tilde{\mathbf{S}}_K$ are the arrays whose (i,j)th elements are $s_{K,i,j} = \mu_{K,i+j-2}$ and $\tilde{s}_{K,i,j} = v_{K,i+j-2}$, respectively.

In the following theorem, expressions for the bias and the variance array of the estimator $\hat{\boldsymbol{\beta}}_{S,n}(x)$ are obtained.

Theorem 1 *If assumptions A1, A2 and A3 are fulfilled, for every $x \in (h_n, 1 - h_n)$, we have*

$$\mathbf{H}_n\left(E\left(\hat{\boldsymbol{\beta}}_{S,n}(x)/\delta\right) - \boldsymbol{\beta}(x)\right) = \frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}\mathbf{S}_K^{-1}\boldsymbol{\mu}_K + o_p(h_n^{p+1}\mathbf{1}), \qquad (5)$$

with $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$ and

$$\mathrm{Var}\left(\mathbf{H}_n\hat{\boldsymbol{\beta}}_{S,n}(x)/\delta\right) = \frac{1}{nh_n}\frac{c^\delta(\varepsilon)}{\mathfrak{p}(x)^2 f(x)}\mathbf{S}_K^{-1}\tilde{\mathbf{S}}_K\mathbf{S}_K^{-1}\left(\mathbf{1} + o_p(\mathbf{1})\right), \qquad (6)$$

where $\mathbf{H}_n = \mathrm{diag}\left(1, h_n, h_n^2, \ldots, h_n^p\right)$, $\delta = (\delta_1, \ldots, \delta_n)^{\mathrm{T}}$ and

$$c^\delta(\varepsilon) = \mathfrak{p}(x)^2 c(\varepsilon) + \mathfrak{p}(x)\mathfrak{q}(x)v(0)\sigma_\varepsilon^2,$$

with $c(\varepsilon) = \sigma_\varepsilon^2\left(v(0) + 2\sum_{k=1}^\infty v(k)\right)$ and $\mathfrak{q}(x) = (1 - \mathfrak{p}(x))$.

Remarks

- From expressions (5) and (6) it is deduced that the AMSE of the estimator $\hat{m}_{S,h_n}^{(j)}(x) = (j!)\,\mathbf{e}_{j+1}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{S,n}(x)$ is

$$\mathrm{AMSE}\left(\hat{m}_{S,h_n}^{(j)}(x)/\delta\right) = \left(\frac{(j!)\,m^{(p+1)}(x)}{(p+1)!}h_n^{p+1-j}\mathbf{e}_{j+1}^{\mathrm{T}}\mathbf{S}_K^{-1}\boldsymbol{\mu}_K\right)^2$$
$$+ \frac{(j!)^2}{nh_n^{2j+1}}\frac{c^\delta(\varepsilon)}{\mathfrak{p}(x)^2 f(x)}\mathbf{e}_{j+1}^{\mathrm{T}}\mathbf{S}_K^{-1}\tilde{\mathbf{S}}_K\mathbf{S}_K^{-1}\mathbf{e}_{j+1},$$
$$j = 0, \ldots, p.$$

- The existence of missing observations has no influence on the bias but does on the variances of the estimators $\hat{m}_{S,h_n}^{(j)}(x)$ through the term

$$\frac{c^{\delta}(\varepsilon)}{\mathfrak{p}(x)^2} = \left(c(\varepsilon) + \frac{\mathfrak{q}(x)}{\mathfrak{p}(x)}\nu(0)\sigma_{\varepsilon}^2\right) \geq c(\varepsilon),$$

which decreases as $\mathfrak{p}(x)$ increases, and therefore, decreases the variance.

- Expressions (5) and (6) generalize those obtained by Francisco-Fernández and Vilar-Fernández (2001) for the case of complete data ($\mathfrak{p}(x) = 1$) under dependence.
- If the observations are independent, considering that $c^{\delta}(\varepsilon) = \mathfrak{p}(x)\nu(0)\sigma_{\varepsilon}^2$, one obtains

$$\text{AMSE}\left(\hat{m}_{S,h_n}^{(j)}(x)/\delta\right) = \left(\frac{(j!)\,m^{(p+1)}(x)}{(p+1)!}h_n^{p+1-j}\mathbf{e}_{j+1}^{\mathrm{T}}\mathbf{S}_K^{-1}\boldsymbol{\mu}_K\right)^2$$
$$+ \frac{(j!)^2}{nh_n^{2j+1}}\frac{\nu(0)\sigma_{\varepsilon}^2}{\mathfrak{p}(x)f(x)}\mathbf{e}_{j+1}^{\mathrm{T}}\mathbf{S}_K^{-1}\tilde{\mathbf{S}}_K\mathbf{S}_K^{-1}\mathbf{e}_{j+1}.$$

To establish the asymptotic normality of $\hat{\boldsymbol{\beta}}_{S,n}(x)$, the following additional assumptions are necessary:

A.4. The process of the random errors $\{\varepsilon_t\}$ has a moving average MA(∞)-type dependence structure, so

$$\varepsilon_t = \sum_{i=0}^{\infty}\phi_i e_{t-i}, \qquad t = 1, 2, \dots$$

where $\{e_t\}$ is a sequence of independent identically distributed random variables with zero mean and variance σ_e^2, and the sequence $\{\phi_i\}$ verifies that $\sum_{i=0}^{\infty}|\phi_i| < \infty$.

A.5. $E|e_t|^{2+\gamma} < \infty$ for some $\gamma > 0$

A.6. $h_n = O(n^{-1/(2p+3)})$

A.7. The sequences of smoothing parameters $\{h_n\}$ and $\{g_n\}$ are the same order, this is, $\lim_{n\to\infty}\dfrac{h_n}{g_n} = \lambda$.

Theorem 2 *If assumptions A1–A6 are fulfilled, for every $x \in (h_n, 1 - h_n)$, we have the asymptotic normality of $\hat{\boldsymbol{\beta}}_{S,n}(x)$ conditional on δ:*

$$\sqrt{nh_n}\left(\mathbf{H}_n\left(\left(\hat{\boldsymbol{\beta}}_{S,n}(x)\right) - \boldsymbol{\beta}(x)\right) - \frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}\mathbf{S}_K^{-1}\boldsymbol{\mu}_K\right) \xrightarrow{\mathcal{L}} N_{(p+1)}(\mathbf{0}, \boldsymbol{\Sigma}_S) \quad (7)$$

where $\boldsymbol{\Sigma}_S = \dfrac{c^{\delta}(\varepsilon)}{\mathfrak{p}(x)^2 f(x)}S_K^{-1}\tilde{\mathbf{S}}_K S_K^{-1}$, and $N_{(p+1)}(\mathbf{0}, \boldsymbol{\Sigma}_S)$ denotes a multivariate normal distribution of dimension $p + 1$.

The asymptotic normality conditional on $\boldsymbol{\delta}$ of the individual components $\hat{m}_{S,h_n}^{(j)}(x) = (j!)\,\hat{\beta}_{S,j}(x)$ is directly derived from Theorem 2. We have, for $j = 0, \ldots, p$,

$$\sqrt{nh_n^{1+2j}}\left(\left(\hat{m}_{S,h_n}^{(j)}(x) - m^{(j)}(x)\right) - h_n^{p+1-j}\frac{m^{(p+1)}(x)}{(p+1)!}(j!)\mathbf{e}_{j+1}^{\mathrm{T}}\mathbf{S}_K^{-1}\boldsymbol{\mu}_K\right) \overset{\mathcal{L}}{\longrightarrow} N$$

$$\left(0, \sigma_j^2\right),$$

where $\sigma_j^2 = (j!)^2\,\dfrac{c^{\delta}(\varepsilon)}{\mathfrak{p}(x)^2 f(x)}\mathbf{e}_{j+1}^t\mathbf{S}_K^{-1}\tilde{\mathbf{S}}_K\mathbf{S}_K^{-1}\mathbf{e}_{j+1}.$

The condition of dependence given in assumption A.4. is very general and a large class of stationary processes have MA(∞) representations (see Sect. 5.7 of Brockwell and Davis (1991)). This condition is taken on to be able to use a Central Limit Theorem for sequences with $m(n)$-dependent main part of Nieuwenhuis (1992). A different strategy can be used to obtain the asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}_{S,n}(x)$. For this, assuming that the process of the errors $\{\varepsilon_t\}$ is strong mixing (α-mixing) and imposing bound conditions on the mixing coefficients, following a similar approach to that employed in Masry and Fan (1997) or Francisco-Fernández and Vilar-Fernández (2001) one can obtain the asymptotic normality of $\hat{\boldsymbol{\beta}}_{S,n}(x)$ using the well known a "small-blocks and large-blocks" method.

3.2 Asymptotic properties of the imputed estimator

Considering that

$$\widehat{Y}_i = \delta_i\left(Y_i - m(x_i)\right) + (1 - \delta_i)\left(\widehat{m}_{S,g_n}(x_i) - m(x_i)\right) + m(x_i)$$

where $\widehat{m}_{S,g_n}(x_i) = \hat{m}_{S,g_n}^{(0)}(x_i)$, the following basic decomposition is obtained

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{I,n}(x) - \boldsymbol{\beta}(x) &= \left(\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n\mathbf{X}_n\right)^{-1}\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n\begin{pmatrix}\delta_1\left(Y_1 - m(x_1)\right)\\ \vdots \\ \delta_n\left(Y_n - m(x_n)\right)\end{pmatrix}\\
&+ \left(\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n\mathbf{X}_n\right)^{-1}\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n\begin{pmatrix}(1-\delta_1)\left(\hat{m}_{S,g_n}(x_1) - m(x_1)\right)\\ \vdots \\ (1-\delta_n)\left(\hat{m}_{S,g_n}(x_n) - m(x_n)\right)\end{pmatrix}\\
&+ \left(\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n\mathbf{X}_n\right)^{-1}\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n\begin{pmatrix}m(x_1)\\ \vdots \\ m(x_n)\end{pmatrix} - \mathbf{X}_n^{\mathrm{T}}\boldsymbol{\beta}(x)\\
&= \boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3. \tag{8}
\end{aligned}$$

Using it as basis, the conditional asymptotic mean square error of $\hat{\beta}_{I,n}(x)$ is obtained as follows. The following notations will be used. Let $L_{g_n,q}^*(v) = \mathbf{e}_1^{\mathrm{T}} \mathbf{S}_L^{-1} (1, \ldots, v^q)^{\mathrm{T}} L_{g_n}(v)$ and $A_{j,q}(v) = g_n \int u^j K(u) L_{g_n,q}^*(h_n(v-u)) \, du$, and let us denote $\mathbf{Z}$ and $\tilde{\mathbf{Z}}$ as the arrays whose (i, j)th elements, $i, j = 1, \ldots, p+1$, are $z_{i,j} = \int A_{i-1,q}(v) A_{j-1,q}(v) \, dv$ and $\tilde{z}_{i,j} = \int v^{i-1} K(v) A_{j-1,q}(v) \, dv$, respectively.

Theorem 3 *Let us suppose that assumptions A1, A2, A3 are verified. Then, for every* $x \in (r_n, 1 - r_n)$*, with* $r_n = \max\{h_n, g_n\}$*, we have*

$$\mathbf{H}_n \left(E\left(\hat{\beta}_{I,n}(x)/\delta\right) - \beta(x) \right) = \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} \mathbf{S}_K^{-1} \boldsymbol{\mu}_K$$
$$+ \mathfrak{q}(x) \frac{m^{(q+1)}(x)}{(q+1)!} g_n^{q+1} \mathbf{S}_K^{-1} \tilde{\boldsymbol{\mu}}_K \, \mathbf{e}_1 \mathbf{S}_L^{-1} \boldsymbol{\mu}_L$$
$$+ o_p \left(\mathbf{1}(h_n^{p+1} + g_n^{q+1}) \right) \tag{9}$$

and, if h_n *and* g_n *verify A7, then*

$$\mathrm{Var}\left(\mathbf{H}_n \hat{\beta}_{I,n}(x)/\delta\right) = \frac{1}{nh_n} \frac{c^\delta(\varepsilon)}{f(x)} \mathbf{S}_K^{-1}$$
$$\times \left(\tilde{\mathbf{S}}_K + \lambda^2 \frac{\mathfrak{q}(x)^2}{\mathfrak{p}(x)^2} \mathbf{Z} + 2\lambda \frac{\mathfrak{q}(x)}{\mathfrak{p}(x)} \tilde{\mathbf{Z}} \right) \mathbf{S}_K^{-1} (1 + o_p(1)), \tag{10}$$

where $\tilde{\boldsymbol{\mu}}_K = \left(\mu_{K,0}, \ldots, \mu_{K,p}\right)^t$.

Remarks

- From expressions (9) and (10) it is deduced that the two smoothing parameters used to calculate the estimator $\hat{\beta}_{I,n}(x)$ have influence in the expressions of the bias and asymptotic variance. The conditional AMSE of the estimator $\hat{m}_{I,h_n}^{(j)}(x) = (j!) \, \mathbf{e}_{j+1}^{\mathrm{T}} \hat{\beta}_{I,n}(x)$ is the following:

$$\mathrm{AMSE}\left(\hat{m}_{I,h_n}^{(j)}(x)/\delta\right) = \left(\begin{array}{c} \dfrac{(j!) \, m^{(p+1)}(x)}{(p+1)!} h_n^{p+1-j} \mathbf{e}_{j+1}^{\mathrm{T}} \mathbf{S}_K^{-1} \boldsymbol{\mu}_K + \\ \dfrac{(j!) \, m^{(q+1)}(x)}{(q+1)!} \mathfrak{q}(x) \dfrac{g_n^{q+1}}{h_n^j} \mathbf{e}_{j+1}^{\mathrm{T}} \mathbf{S}_K^{-1} \tilde{\boldsymbol{\mu}}_K \, \mathbf{e}_1 \mathbf{S}_L^{-1} \boldsymbol{\mu}_L \end{array} \right)^2$$
$$+ \frac{(j!)^2}{nh_n^{2j+1}} \frac{c^\delta(\varepsilon)}{f(x)} \mathbf{e}_{j+1}^{\mathrm{T}} \mathbf{S}_K^{-1}$$
$$\times \left(\tilde{\mathbf{S}}_K + \frac{h_n^2}{g_n^2} \frac{\mathfrak{q}(x)^2}{\mathfrak{p}(x)^2} \mathbf{Z} + 2 \frac{h_n}{g_n} \frac{\mathfrak{q}(x)}{\mathfrak{p}(x)} \tilde{\mathbf{Z}} \right) \mathbf{S}_K^{-1} \mathbf{e}_{j+1},$$
$$j = 0, \ldots, p.$$

- The existence of missing observations has influence in the bias and the variances of the estimator $\hat{m}_{I,h_n}^{(j)}(x)$. The expressions (9) and (10) generalize those obtained

by Francisco-Fernández and Vilar-Fernández (2001) for the case of complete data ($\mathfrak{p}(x) = 1$) under dependence.

- The dependence of the errors influences the variance of the estimator $\hat{m}_{I,h_n}^{(j)}(x)$ through the term $c^\delta(\varepsilon)$. If the observations are independent one obtains that

$$
\text{AMSE}\left(\hat{m}_{I,h_n}^{(j)}(x)/\delta\right) = \left(
\begin{array}{c}
\dfrac{(j!)\, m^{(p+1)}(x)}{(p+1)!} h_n^{p+1-j} \mathbf{e}_{j+1}^{\mathrm{T}} \mathbf{S}_K^{-1} \boldsymbol{\mu}_K + \\[2mm]
\dfrac{(j!)\, m^{(q+1)}(x)}{(q+1)!} \mathfrak{q}(x)\, \dfrac{g_n^{q+1}}{h_n^j} \mathbf{e}_{j+1}^{\mathrm{T}} \mathbf{S}_K^{-1} \tilde{\boldsymbol{\mu}}_K\, \mathbf{e}_1 \mathbf{S}_L^{-1} \boldsymbol{\mu}_L
\end{array}
\right)^2
$$
$$
+ \frac{(j!)^2}{n h_n^{2j+1}} \frac{c(0)\,\mathfrak{p}(x)\sigma_\varepsilon^2}{f(x)} \mathbf{e}_{j+1}^{\mathrm{T}} \mathbf{S}_K^{-1}
$$
$$
\times \left(\tilde{\mathbf{S}}_K + \frac{h_n^2}{g_n^2}\frac{\mathfrak{q}(x)^2}{\mathfrak{p}(x)^2}\mathbf{Z} + 2\frac{h_n}{g_n}\frac{\mathfrak{q}(x)}{\mathfrak{p}(x)}\tilde{\mathbf{Z}}\right)\mathbf{S}_K^{-1}\mathbf{e}_{j+1}.
$$

This expression generalizes those obtained by González-Manteiga and Pérez-González (2004) who used local linear regression ($p = q = 1$) for the case of incomplete data under independence.

- In the case $\dfrac{g_n}{h_n} \to 0$, the expression of the bias is that given in (9) although note that if $q = p$, the second summand is asymptotically null with respect to the first since $g_n = o(h_n)$. With respect to the asymptotic variance, its expression is the following

$$
\text{Var}\left(\mathbf{H}_n \hat{\boldsymbol{\beta}}_{I,n}(x)/\delta\right) = \frac{1}{n h_n}\frac{c^\delta(\varepsilon)}{\mathfrak{p}(x)^2 f(x)}\mathbf{S}_K^{-1}\tilde{\mathbf{S}}_K \mathbf{S}_K^{-1}\left(1 + o_p(1)\right).
$$

This expression coinciding with that obtained for the variance of the Simplified estimator $\hat{\boldsymbol{\beta}}_{S,n}(x)$ (see 6), hence, if $q = p$, the estimators, Simplified $\hat{m}_{S,h_n}^{(j)}(x)$ and Imputed $\hat{m}_{I,h}^{(j)}(x)$, have the same asymptotic mean squared error. But if $q < p$, the second term of the bias of the Imputed estimator can be dominant, and there, the bias of estimator $\hat{m}_{I,h_n}^{(j)}(x)$ is greater than that of $\hat{m}_{S,h}^{(j)}(x)$ and also has greater AMSE.

- In the case of $\dfrac{h_n}{g_n} \to 0$, again, the expression of the asymptotic bias is that given in (9), but in this case because $q \le p$, the second summand on the right side of Eq. (9) is the dominant term of the bias. To obtain the expression of the variance of the estimator, let us denote $\mathbf{R}$ and $\tilde{\mathbf{R}}$ as the arrays whose (i, j)th elements, $i, j = 1, \ldots, p+1$, are

$$
r_{i,j} = \left(\int v^{i+j-2} L^2(v)\mathrm{d}v\right)\left(\int v^{i-1} K(v)\mathrm{d}v\right)\left(\int v^{j-1} K(v)\mathrm{d}v\right)
$$

and

$$\tilde{r}_{i,j} = \left(\int v^{i-1} K(v) dv \right) \left(\int v^{j-1} K(v) dv \right),$$

respectively. Then, the expression of the variance, in this case, is the following

$$\text{Var}\left(\mathbf{H}_n \hat{\boldsymbol{\beta}}_{I,n}(x)/\boldsymbol{\delta} \right) = \frac{1}{nh_n} \frac{c^\delta(\varepsilon)}{f(x)} \mathbf{S}_K^{-1}$$

$$\times \left(\tilde{\mathbf{S}}_K + \frac{h_n}{g_n} \frac{\mathsf{q}(x)^2}{\mathsf{p}(x)^2} \mathbf{R} + 2 \frac{h_n}{g_n} \frac{\mathsf{q}(x)}{\mathsf{p}(x)} \tilde{\mathbf{R}} L_q^*(0) \right) \mathbf{S}_K^{-1} \left(1 + o_p(1) \right)$$

$$\approx \frac{1}{nh_n} \frac{c^\delta(\varepsilon)}{f(x)} \mathbf{S}_K^{-1} \tilde{\mathbf{S}}_K \mathbf{S}_K^{-1} \left(1 + o_p(1) \right).$$

In this case, the imputed estimator provides better asymptotic variance than the simplified estimator due to the oversmoothing of the bandwidth parameter g_n. But the bias for the imputed estimator is bigger than that for the simplified one.

- From the above, it is deduced that the imputed estimator gives good results if $h_n = \xi g_n$ is chosen. A selection method of these parameters consists of to find h_n and ξ that minimize the AMSE $\left(\hat{m}_{I,h_n}^{(j)}(x)/\boldsymbol{\delta} \right)$. For example, using $p = q$ and considering $h_n = \varphi n^{-1/(2p+3)}$ as the usual selection one obtains

$$\text{AMSE}\left(\hat{m}_{I,h_n}^{(j)}(x)/\boldsymbol{\delta} \right) = \text{AMSE}(\varphi, \xi)$$

$$= \left(\frac{\frac{(j!)\, m^{(p+1)}(x)}{(p+1)!} \varphi^{p+1-j} n^{\frac{j-p-1}{2p+3}}}{\left(\mathbf{e}_{j+1}^{\mathsf{T}} \mathbf{S}_K^{-1} \boldsymbol{\mu}_K + \mathsf{q}(x) \xi^{q+1} \mathbf{e}_{j+1}^{\mathsf{T}} \mathbf{S}_K^{-1} \tilde{\boldsymbol{\mu}}_K \, \mathbf{e}_1 \mathbf{S}_L^{-1} \boldsymbol{\mu}_L \right)} \right)^2$$

$$+ (j!)^2 \frac{c^\delta(\varepsilon)}{f(x)} \varphi^{2j+1} n^{\frac{2(p+1-j)}{2p+3}} \mathbf{e}_{j+1}^{\mathsf{T}} \mathbf{S}_K^{-1}$$

$$\times \left(\tilde{\mathbf{S}}_K + \xi^2 \frac{\mathsf{q}(x)^2}{\mathsf{p}(x)^2} \mathbf{Z} + 2\xi \frac{\mathsf{q}(x)}{\mathsf{p}(x)} \tilde{\mathbf{Z}} \right) \mathbf{S}_K^{-1} \mathbf{e}_{j+1}. \qquad (11)$$

- From expression (11), plug-in selections of φ and ξ can be obtained as values that minimize one estimation of the previous function, $\widehat{\text{AMSE}}(\varphi, \xi)$. This estimation is obtained when substituting in (11) the values $m^{(p+1)}(x)$, $c^\delta(\varepsilon)$, $\mathsf{p}(x)$ and $f(x)$ for estimations of these.

  Taking into account $\mathsf{p}(x) = P(\delta = 1/x) = E(\delta/x)$, this function can be estimated from sample $\left\{ (x_t, \delta_t)_{t=1}^n \right\}$ by a nonparametric regression method. The estimator of $c(\varepsilon) = \sum_{k=-\infty}^{\infty} E(\varepsilon_t, \varepsilon_{t+k}) = \sum_{k=-\infty}^{\infty} \gamma(k)$, and therefore, the estimation of $c^\delta(\varepsilon)$ can be obtained in several ways. In any case, we will use the completed sample $\left\{ \left(x_t, \tilde{Y}_t \right)_{t=1}^n \right\}$, where $\tilde{Y}_t = \delta_t Y_t + (1-\delta_t) \hat{m}_{S,gpilot}(x_t)$, and $\hat{m}_{S,gpilot}(x_t)$ was computed with pilot bandwidth $gpilot$. Using this sample, several procedures can be used to estimate $c(\varepsilon)$.

Müller and Stadtmüller (1988) suggested an estimator for $\gamma(k)$ based on first order differences of sequence $\xi_{t,r} = \tilde{Y}_t - \tilde{Y}_{t-r}$, $t = 1, \ldots, n$, with lag r $(r \geq 1)$. Supposing that ε_t is an m-dependent process, Müller and Stadtmüller (1988) proved the asymptotic properties of the proposed estimator, $\hat{\gamma}_{MS}(k)$. Now an estimator $\hat{c}_{MS}(\varepsilon)$ for $c(\varepsilon)$ is defined as follows:

$$\hat{c}_{MS}(\varepsilon) = \sum_{k=r(n)}^{r(n)} \hat{\gamma}_{MS}(k),$$

where $r(n)$ is a sequence of positive integers, with $r(n) \to \infty$ as $n \to \infty$. Under more general conditions of dependence, ε_t is $\alpha-$mixing. Herrmann et al. (1992) used second order differences of $\tilde{Y}_t$, defined as

$$\eta_{t,\alpha,\beta} = \tilde{Y}_t - \frac{\alpha}{\alpha+\beta}\tilde{Y}_{t+\beta} - \frac{\beta}{\alpha+\beta}\tilde{Y}_{t-\alpha},$$

and they proposed estimators $\hat{\gamma}_{HGK}(k)$ for the covariances of the process ε_t based on these differences $\eta_{t,\alpha,\beta}$. Again, an estimator $\hat{c}_{HGK}(\varepsilon)$ for $c(\varepsilon)$ is defined as follows:

$$\hat{c}_{HGK}(\varepsilon) = \sum_{k=r(n)}^{r(n)} \hat{\gamma}_{HGK}(k).$$

The consistency of this estimator is proven in Herrmann et al. (1992). Also, Hall and Van Keilegom (2003) suggested a new difference-based method for estimating error autocovariance with time series errors.

The asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}_{I,n}(x)$ is established in the following theorem.

Theorem 4 *Assume A1–A7. Then, for every* $x \in (r_n, 1 - r_n)$, *we have the asymptotic normality of* $\hat{\boldsymbol{\beta}}_{I,n}(x)$ *conditional on* $\boldsymbol{\delta}$:

$$\sqrt{nh_n}\left(\mathbf{H}_n\left(\left(\hat{\boldsymbol{\beta}}_{I,n}(x)\right) - \boldsymbol{\beta}(x)\right) - \mathbf{B}_I\right) \xrightarrow{\mathcal{L}} N_{(p+1)}\left(\mathbf{0}, \boldsymbol{\Sigma}_I\right),$$

where

$$\mathbf{B}_I = \frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}\mathbf{S}_K^{-1}\boldsymbol{\mu}_K + \mathfrak{q}(x)\frac{m^{(q+1)}(x)}{(q+1)!}g_n^{q+1}\mathbf{S}_K^{-1}\tilde{\boldsymbol{\mu}}_K\,\mathbf{e}_1\mathbf{S}_L^{-1}\boldsymbol{\mu}_L,$$

and

$$\boldsymbol{\Sigma}_I = \frac{c^{\delta}(\varepsilon)}{f(x)}\mathbf{S}_K^{-1}\left(\tilde{\mathbf{S}}_K + \lambda^2\frac{\mathfrak{q}(x)^2}{\mathfrak{p}(x)^2}\mathbf{Z} + 2\lambda\frac{\mathfrak{q}(x)}{\mathfrak{p}(x)}\tilde{\mathbf{Z}}\right)\mathbf{S}_K^{-1}.$$

The asymptotic normality of the individual components $\hat{m}_{I,h_n}^{(j)}(x) = (j!)\,\hat{\beta}_{I,j}(x)$ is directly derived from Theorem 2. Moreover, the considerations made in Theorem 2 with respect to the dependence structure of the errors process are also valid for this estimator.

4 A simulation study

In this section, we compare the performance of the simplified estimator (2) and the imputed estimator (4). For this purpose, we use the complete data estimator as reference. The simulation study was carried out using a local linear smoother ($p = 1$), considering $p = q = 1$ for the Imputed estimator.

We consider a fixed design model in the interval $[0, 1]$, with equispaced data and with random errors following an AR(1) process

$$\varepsilon_t = \rho\varepsilon_{t-1} + e_t$$

with $N(0, \sigma = 0.3)$ distribution. The regression function considered is $m(x) = 5(x - 0.5)^3$, and the missing data model (1) is $\mathfrak{p}(x) = 0.8\exp(-x^2)$. The kernel functions used (K and L) were the Epanechnikov kernel.

To study the influence of the dependence of the observations, different degrees of dependence were considered, specifically, the following correlation coefficient values were considered for $\rho : \rho = -0.8, -0.5, -0.25, 0, 0.25, 0.5$, and 0.8.

In the first part of the study the global smoothing parameters needed for the three estimators were estimated. For this, the mean integrated square error, MISE, was considered as error criterion. Three hundred samples, of size 100, of the previous model were generated, and the MISE value was approached by Monte Carlo for each smoothing parameter value taken over a grid of size 100 of interval $(0, 1)$. For the Imputed estimator the minimization process was carried out using a double grid for the smoothing parameters used (h_n and g_n). Table 1 shows the values obtained for the optimal global smoothing parameters for each correlation coefficient (ρ) value.

Observing the values obtained for the case of complete data and the simplified estimator, it is apparent how the missing data imply an increase in the smoothing parameter. For another hand, for the three estimators, we can observe that when the dependence of the observations increases, that is, the value of ρ increases, then, the variability of the data increases, and therefore, the optimal smoothing parameter also increases.

Table 1 Optimal global bandwidth

	ρ						
	−0.8	−0.5	−0.25	0.0	0.25	0.5	0.8
Complete	0.17	0.19	0.20	0.23	0.26	0.35	1.00
Simplified	0.38	0.32	0.31	0.32	0.36	0.47	1.00
Imputed							
g	1.00	0.51	0.49	0.51	1.00	1.00	0.49
h	0.21	0.17	0.17	0.19	0.22	0.28	1.00

Table 2 Approximated MISE for 500 samples

	ρ						
	-0.8	-0.5	-0.25	0.0	0.25	0.5	0.8
Complete	0.0018	0.0024	0.0031	0.0045	0.0070	0.0133	0.0503
Simplified	0.0127	0.0084	0.0079	0.0086	0.0110	0.0177	0.0576
Imputed	0.0103	0.0078	0.0076	0.0084	0.0107	0.0164	0.0555
Eff. (%)	18.765	6.987	4.306	2.389	2.583	7.439	3.583
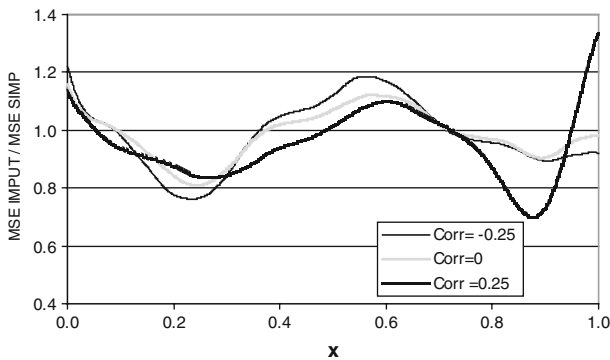
Fig. 1 Quotient between mean squared errors for imputed and simplified estimators with $\rho = -0.25$, 0 and 0.25

Once the optimal bandwidths are obtained, the estimation of the regression function was carried out on another 500 different samples. For these samples, the mean squared error and the MISE were estimated. To compare the simplified and the imputed estimators we computed the efficiency of the latter in the following way:

$$\text{Eff.}(\%) = \frac{\text{MISE}_{\text{SIMP}} - \text{MISE}_{\text{IMPUT}}}{\text{MISE}_{\text{SIMP}}} \times 100,$$

obtaining the values observed in the last row of Table 2.

The results show better behavior for the imputed estimator than for the simplified, with a benefit above 2.5%. Just as in the case of complete data, we see that as the correlation coefficient increases, the value of the MISE increases drastically.

Figure 1 shows the quotient between the mean squared errors, MSE, for imputed and for simplified estimators for three correlation coefficient values ($\rho = -0.25, 0$ and 0.25). It is apparent that at certain points the simplified estimator is better than the imputed. This finding, along with the fact that when using global measures such as the MISE, the imputed estimator is better, justifies that selection of a local bandwidth (for each point) would substantially improve the results.

Figure 2 shows the boxplots of the MSE for the three estimators with $\rho = -0.25$, 0 and 0.25.
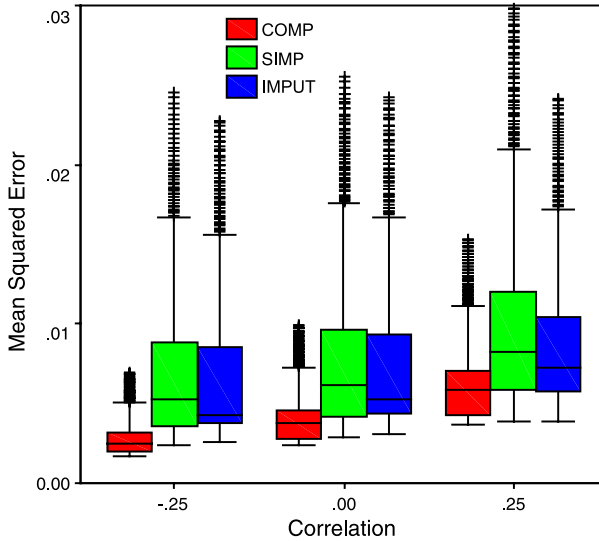
Fig. 2 Boxplots of MSE for the complete (*red*), simplified (*green*) and imputed (*blue*) estimators with $\rho = -0.25$, 0 and 0.25

It is observed that as the correlation coefficient increases, the MSE also increase. Moreover, the good behavior of the Imputed estimator is apparent compared with the Simplified in the three cases.

4.1 The effect of strong dependence

We were interested in studying the behavior of the estimators under strong dependence for different sample sizes. For this reason, we performed more simulations using larger sample size for this value ($\rho = 0.8$). Tables 3 and 4 show the optimal bandwidth and the MISE obtained for several sample sizes.

We can see that as the sample size increases, the MISE decreases and the behavior of the imputed estimator is better. The bandwidth for the imputation of the imputed estimator (g_n) is very big because when the correlation is big, the variance is big also.

Table 3 Optimal bandwidth with correlation $\rho = 0.8$

	n				
	100	200	300	400	500
Complete	1	1	0.55	0.46	0.37
Simplified	1	1	0.56	0.47	0.39
Imputed					
g	0.49	1	1	1	1
h	1	0.44	0.34	0.34	0.3

Table 4 Mean integrated squared error with correlation $\rho = 0.8$

	n				
	100	200	300	400	500
Complete	0.0503	0.0303	0.0235	0.0188	0.0161
Simplified	0.0576	0.0334	0.0258	0.0205	0.0174
Imputed	0.0555	0.0325	0.0244	0.0193	0.0163
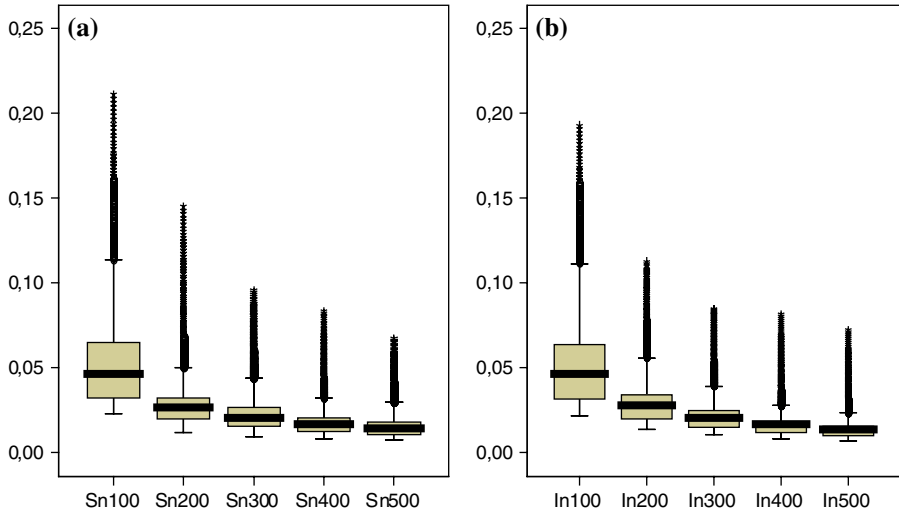
Fig. 3 Boxplots of MSE for simplified (**a**) and imputed (**b**) estimators with $\rho = 0.8$ and $n = 100$, 200, 300, 400 and 500

In the asymptotic results we can see that if we choose $\frac{h_n}{g_n} \to 0$ the variance for the imputed estimator is lower than that for the simplified estimator. The oversmoothing tends to decrease the variance.

Figures 3 and 4 show the boxplots of the MSE for the estimators for incomplete data, and for five sample sizes.

The two estimators have similar behavior with respect to sample size, as long as n grows the MSE decreases.

If we compare the three estimators (complete data case, simplified and imputed estimators) for various sample sizes, we see the following graphs.

The complete data case has the best behavior, and the imputed estimator performs better than the simplified estimator.

5 Conclusions

In this paper we have proposed two nonparametric estimators of the regression functions with correlated errors and missing observations in the response variable.
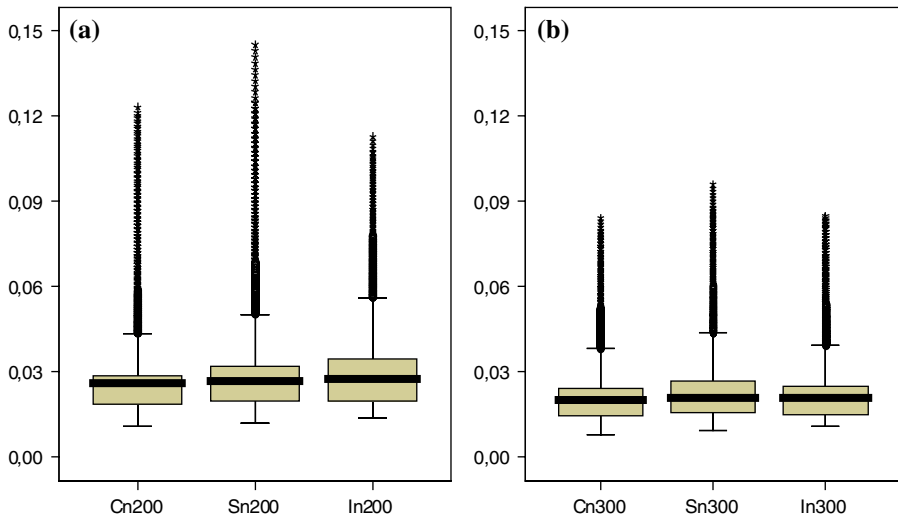
Fig. 4 Boxplots of MSE for the complete (boxplot on the *left*), simplified (*middle*) and imputed (*right*) estimators with $\rho = 0.8$, $n = 200$ (**a**) and $n = 300$ (**b**)

The MSE and the asymptotic normality of the estimators have been studied. We observed that the performance of the estimators depends on the bandwidth parameters. The imputed estimator needs two bandwidth parameters, and when a suitable choice of these parameters is made, the behavior of this one is better than that of the simplified one. In the case of the imputation and the estimation bandwidth verify $\frac{g_n}{h_n} \to 0$, the simplified estimator is better than the imputed estimator. In the case of $\frac{h_n}{g_n} \to 0$, the imputed estimator has a smaller variance but the bias can be bigger. Nevertheless, if the bandwidth parameters are of similar order ($h_n = \xi g_n$ with $\xi > 0$), the imputed estimator can be better than the simplified estimator.

6 Proofs

In this section, we sketch proofs of the results presented in Sect. 3.

First, the convergence for arrays $\mathbf{S}_n$ and $\mathbf{T}_n$ are established.

Proposition 1 *Under assumptions A1 and A3, for every $x \in (h_n, 1 - h_n)$, we have*

$$\lim_{n \to \infty} \mathbf{H}_n^{-1} \mathbf{S}_n \mathbf{H}_n^{-1} = \mathfrak{p}(x) f(x) \mathbf{S}_K \left(1 + o_p(1)\right). \tag{12}$$

Proposition 2 *Under assumptions A1, A2 and A3, for every $x \in (h_n, 1 - h_n)$, we have*

$$\lim_{n \to \infty} n h_n \mathrm{Var} \left(\mathbf{H}_n^{-1} \mathbf{T}_n / \boldsymbol{\delta}\right) = \mathfrak{p}(x) f(x) \tilde{\mathbf{S}}_K \left(1 + o_p(1)\right).$$

Using Taylor's expansion and following a similar approach to that employed in Francisco-Fernández and Vilar-Fernández (2001) we can deduce Propositions 1 and 2.

Proof of Theorem 1. Let $\mathbf{M}_n = (m(x_1), \ldots, m(x_n))^t$. Performing a $(p+1)$th-order Taylor expansion of function m in a neighborhood of x, we obtain

$$E\left(\hat{\boldsymbol{\beta}}_{S,n}(x)/\boldsymbol{\delta}\right) - \boldsymbol{\beta}(x) = \left(\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n^{\delta}\mathbf{X}_n\right)^{-1}\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n^{\delta}\left(\mathbf{M}_n - \mathbf{X}_n^t\boldsymbol{\beta}\right)$$

$$= \frac{m^{(p+1)}(x)}{(p+1)!}\left(\mathbf{X}_n^{\mathrm{T}}\mathbf{W}_n^{\delta}\mathbf{X}_n\right)^{-1}\left(\begin{pmatrix} s_{p+1,n} \\ \cdots \\ s_{2p+1,n} \end{pmatrix} + o_p\left(h_n^{p+1}\mathbf{1}\right)\right).$$

Using Proposition 1 we obtain the bias of $\hat{\boldsymbol{\beta}}_{S,n}(x)$ given in (5).
Again, using (12) we have

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_{S,n}(x)/\boldsymbol{\delta}\right) = \frac{1}{\mathfrak{p}(x)^2 f(x)^1}\mathbf{H}_n^{-1}\mathbf{S}_K\mathbf{H}_n^{-1}Var\left(\mathbf{T}_n/\boldsymbol{\delta}\right)\mathbf{H}_n^{-1}\mathbf{S}_K\mathbf{H}_n^{-1}\left(1 + o_p(1)\right).$$

$$(13)$$

From (13) and Proposition 2, we deduce the expression of the conditional variance of $\hat{\boldsymbol{\beta}}_{S,n}(x)$ (6).

Proof of Theorem 2. First, we study the asymptotic distribution conditional on $\boldsymbol{\delta}$ of the vector $\sqrt{nh_n}\left(\mathbf{H}_n^{-1}\left(\mathbf{T}_n^*\right)\right)$, where $\left(\mathbf{T}_n^*\right) = (\mathbf{T}_n) - E\left(\mathbf{T}_n/\boldsymbol{\delta}\right)$. For it, let $Q_{S,n}$ be an arbitrary linear combination of

$$\mathbf{H}_n^{-1}\left(\mathbf{T}_n^*\right), Q_{S,n} = \tilde{\mathbf{a}}^{\mathrm{T}}\left(\mathbf{H}_n^{-1}\left(\mathbf{T}_n^*\right)\right) = \sum_{i=0}^{p}\alpha_i h_n^{-i}\left((t_{i,n}) - E\left(t_{i,n}/\boldsymbol{\delta}\right)\right)$$

with $\tilde{\mathbf{a}} = (a_0, a_1, \ldots, a_p) \in \mathbb{R}^{p+1}$.
So

$$\sqrt{nh_n}Q_{S,n} = \sum_{t=1}^{n}\xi_{t,n},$$

where

$$\xi_{t,n} = \sqrt{\frac{h_n}{n}}\Psi_{h_n}(x_t - x)\delta_t\varepsilon_t,$$

with

$$\Psi(u) = K(u)\sum_{i=0}^{p}\alpha_i u^j \quad \text{and} \quad \Psi_{h_n}(u) = \frac{1}{h_n}\Psi\left(\frac{u}{h_n}\right).$$

Using Proposition 2, the variance of $Q_{S,n}$ is obtained

$$\sigma_{QS}^2 = \lim_{n \to \infty} \text{Var}(Q_{S,n}/\delta) = \lim_{n \to \infty} nh_n \tilde{\mathbf{a}}^{\mathsf{T}} \text{Var}\left(\mathbf{H}_n^{-1}\left(\mathbf{T}_n^*/\delta\right)\right) \mathbf{a} \tilde{} = \mathfrak{p}(x) f(x) \tilde{\mathbf{S}}_K < \infty. \tag{14}$$

It remains to prove the asymptotic normal distribution conditional on δ of $Q_{S,n} = \sum_{t=1}^n \xi_{t,n}$. For it, we will use the following central limit theorem for sequences with $m(n)$-dependent main part of Nieuwenhuis (1992).

Theorem A.5. *Suppose that the array* $\{X_{i,n}, 1 \leq i \leq q(n)\}$ *has an* $m(n)$-*dependent main part* $\{\tilde{X}_{i,n}\}$ *and a residual part* $\{\bar{X}_{i,n}\}$. *Set* $b_n^2 = \text{Var}\left(\sum_{i=1}^{q(n)} X_{i,n}\right)$. *Assume that the arrays* $(X_{i,n}/b_n)$ *and* $(\bar{X}_{i,n}/b_n)$ *satisfy the variance conditions* C_1 *and* C_1^*, *respectively, and that both arrays satisfy the* $(2 + \delta)$-*moment condition* C_2 *for some* $\delta > 0$ *with* $m(n)^{2+2/\delta}/q(n) \to 0$. *Then*

$$\frac{1}{b_n} \sum_{t=1}^{q(n)} (X_{t,n} - E(X_{t,n})) \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as } n \to \infty.$$

The above conditions are:

Condition C_1. $\displaystyle\max_{i < j \leq q(n)} \frac{1}{j-i} \text{Var}\left(\sum_{t=i+1}^{j} \frac{X_{t,n}}{b_n}\right) = O\left(\frac{1}{q(n)}\right)$ *as* $n \to \infty$.

Condition C_1^*. $\displaystyle\max_{i < j \leq q(n)} \frac{1}{j-i} \text{Var}\left(\sum_{k=i+1}^{j} \frac{\bar{X}_{t,n}}{b_n}\right) = o\left(\frac{1}{q(n)}\right)$ *as* $n \to \infty$.

Condition C_2. $\displaystyle\max_{1 < t \leq q(n)} E(|Z_{t,n}|^{2+\gamma}) = O(q(n)^{-1-(\gamma/2)})$ *as* $n \to \infty$,

$$\text{with } Z_{t,n} = \frac{X_{t,n}}{b_n} \text{ or } \frac{\bar{X}_{t,n}}{b_n}.$$

Here an array $Z_{t,n}$ is called $m(n)$-dependent if for all $n \in \mathbb{N}$ and $k \in \{2, \ldots, q(n) - m(n)\}$ the random vectors $(Z_{t,n} : 1 \leq t \leq k-1)$ and $(Z_{i,n} : k + m(n) \leq i \leq q(n))$ are independent. The sequence $m(n)$ verifies $m(n)/n \to 0$ as $n \to \infty$.

In our case, using assumption A.4. we have

$$\varepsilon_t = \sum_{i=0}^{\infty} \phi_i e_{t-i} = \sum_{i=0}^{m(n)} \phi_i e_{t-i} + \sum_{i=m(n)}^{\infty} \phi_i e_{t-i} = \tilde{\varepsilon}_t + \bar{\varepsilon}_t.$$

Then, taking into account that the kernel K has bounded support, changing the indices and without loss of generality, it can be written as

$$\sqrt{nh_n} Q_{S,n} = \sum_{t=-|h_n n|}^{|h_n n|} \xi_{t,n} = \sum_{t=-|h_n n|}^{|h_n n|} (\hat{\xi}_{t,n} + \bar{\xi}_{t,n}),$$

where $q(n) = 2|h_n n|$ ($|\bullet|$ denotes the integer part), $\xi_{t,n}$ is a triangular array with $m(n)$- dependent main part $\hat{\xi}_{t,n}$ and residual part $\bar{\xi}_{t,n}$, obtained when substituting in $\xi_{t,n}$, ε_t for $\tilde{\varepsilon}_t$ and $\bar{\varepsilon}_t$, respectively.

In the following, we shall prove that our array verifies the conditions of Theorem A.5.

Condition C_1. Under A.1, A.3 and using (14) we obtain

$$\frac{1}{j-i}\operatorname{Var}\left(\sum_{t=i+1}^{j}\frac{\xi_{t,n}}{b_n}\right) \leq \frac{1}{j-i}\frac{1}{b_n^2}\sum_{t=i+1}^{j}\sum_{k=i+1}^{j}E\left(\xi_{t,n}\xi_{k,n}\right)$$

$$\leq \frac{C}{j-i}\frac{1}{nh_n}\sum_{s=-(j-i-1)}^{j-(i+1)}(j-i-|s|)\,|\nu(s)|$$

$$= O\left(\frac{1}{nh_n}\right) = O\left(\frac{1}{q(n)}\right),$$

where C is a positive constant (notation that will be used from here on). Therefore, $\{\xi_{t,n}/b_n\}$ verifies C_1.

Condition C_1^*. Reasoning in a similar way, it is easy to show that

$$\frac{1}{j-i}\operatorname{Var}\left(\sum_{t=i+1}^{j}\frac{\bar{\xi}_{t,n}}{b_n}\right) \leq \frac{1}{j-i}\frac{1}{b_n^2}\sum_{t=i+1}^{j}\sum_{k=i+1}^{j}E\left(\bar{\xi}_{t,n}\bar{\xi}_{k,n}\right)$$

$$\leq \frac{C}{nh_n}\sum_{s=-(j-i-1)}^{j-(i+1)}E\left(\bar{\varepsilon}_{t,n}\bar{\varepsilon}_{t+s,n}\right)$$

$$= \frac{C}{nh_n}\sum_{s=-(j-i-1)}^{j-(i+1)}\sum_{j>m(n)}\phi_j\sum_{k>m(n)}\phi_k E(e_{t-j}e_{k-j})$$

$$\leq \frac{C}{nh_n}\sum_{s}\left(\sum_{j>m(n)}\phi_j\phi_{j+s}\right)\sigma_e^2 \leq \frac{C}{nh_n}\left(\sum_{j>m(n)}\phi_j\right)^2$$

$$= o\left(\frac{1}{nh_n}\right) = o\left(\frac{1}{q(n)}\right).$$

Condition C_2. We want to prove $(2+\gamma)$-moment condition for the arrays $\xi_{t,n}/b_n$ and $\bar{\xi}_{t,n}/b_n$. This is done for the first one, but for the second, an similar approach is used. Taking into account the form of the function $\Psi_{h_n}(u)$, assumptions A.1. and A.5. and using (14) we obtain

$$E\left|\frac{\xi_{t,n}}{b_n}\right|^{2+\gamma} = C\,E\left|\frac{1}{\sqrt{nh_n}}\varepsilon_t\right|^{2+\gamma} = O\left(\frac{1}{\sqrt{nh_n}}\right)^{2+\gamma} = O(q(n)^{-1-(\gamma/2)}).$$

Therefore, we have proven the asymptotic normality conditional on δ of

$$Q_{S,n} = \sqrt{nh_n}\tilde{\mathbf{a}}^{\mathrm{T}}\left(\mathbf{H}_n^{-1}\left(\mathbf{T}_n^*\right)\right) \xrightarrow{\mathcal{L}} N(0, \sigma_{QS}^2).$$

Now, using the Cramer-Wold Theorem we obtain the asymptotic normality conditional on $\boldsymbol{\delta}$ of

$$\sqrt{nh_n}\left(\mathbf{H}_n^{-1}\left(\mathbf{T}_n^*\right)\right) \xrightarrow{\mathcal{L}} N_{(p+1)}\left(0, \boldsymbol{\Sigma}_S\right). \tag{16}$$

Finally, taking into account

$$\sqrt{nh_n}\mathbf{H}_n\left(\left(\hat{\boldsymbol{\beta}}_{S,n}\left(x\right)\right) - \boldsymbol{\beta}\left(x\right)\right)$$
$$= \sqrt{nh_n}\left(\mathbf{H}_n^{-1}\mathbf{S}_n\mathbf{H}_n^{-1}\right)^{-1}\mathbf{H}_n^{-1}\left(\mathbf{T}_n^*\right) + \sqrt{nh_n}\mathbf{H}_n\left(E\left(\hat{\boldsymbol{\beta}}_{S,n}\left(x\right)/\boldsymbol{\delta}\right) - \boldsymbol{\beta}\left(x\right)\right)$$

from (5) and (16) and using Proposition 1, the asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}_{S,n}\left(x\right)$ conditional on $\boldsymbol{\delta}$ (given in (7)) is established.

Proof of Theorem 3. To obtain the bias of $\hat{\boldsymbol{\beta}}_{I,n}\left(x\right)$, from (8) it follows that

$$\mathbf{H}_n E\left(\hat{\boldsymbol{\beta}}_{I,n}\left(x\right)/\boldsymbol{\delta}\right) - \boldsymbol{\beta}\left(x\right) = \mathbf{H}_n E(\boldsymbol{\Gamma}_2/\boldsymbol{\delta}) + \mathbf{H}_n\boldsymbol{\Gamma}_3, \tag{17}$$

because $\boldsymbol{\Gamma}_3$ is no random term.

From (5) and using Proposition 1 it is easy to obtain that

$$\mathbf{H}_n E(\boldsymbol{\Gamma}_2/\boldsymbol{\delta}) = \mathfrak{q}\left(x\right)\frac{m^{(q+1)}\left(x\right)}{(q+1)!}g_n^{q+1}\mathbf{S}_K^{-1}\tilde{\boldsymbol{\mu}}_K\,e_1\mathbf{S}_L^{-1}\boldsymbol{\mu}_L + \mathbf{1}\,g_n^{q+1}. \tag{18}$$

Again, using Proposition 1 we have

$$\mathbf{H}_n E(\boldsymbol{\Gamma}_3/\boldsymbol{\delta}) = \frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}\mathbf{S}_K^{-1}\boldsymbol{\mu}_K + \mathbf{1}\,h_n^{p+1} \tag{19}$$

From (17), (18) and (19) it follows (9).
With respect to the variance of $\hat{\boldsymbol{\beta}}_{I,n}\left(x\right)$, we have

$$\text{Var}\left(\hat{\boldsymbol{\beta}}_{I,n}\left(x\right)/\boldsymbol{\delta}\right) = \text{Var}\left(\boldsymbol{\Gamma}_1/\boldsymbol{\delta}\right) + \text{Var}\left(\boldsymbol{\Gamma}_2/\boldsymbol{\delta}\right) + 2\,\text{Cov}(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2/\boldsymbol{\delta}).$$

Using the same kind of arguments as those used in the proof of (6) we obtain

$$\text{Var}\left(\boldsymbol{\Gamma}_1/\boldsymbol{\delta}\right) = \frac{1}{nh_n}\frac{c^\delta\left(\varepsilon\right)}{f\left(x\right)}\mathbf{H}_n^{-1}\mathbf{S}_K^{-1}\widetilde{\mathbf{S}}_K\mathbf{S}_K^{-1}\mathbf{H}_n^{-1}\left(1 + o_p\left(1\right)\right). \tag{20}$$

With respect to the variance of $\boldsymbol{\Gamma}_2$, we have

$$\text{Var}\left(\boldsymbol{\Gamma}_2/\boldsymbol{\delta}\right) = \left(\mathbf{X}_n^{\text{T}}\mathbf{W}_n\mathbf{X}_n\right)^{-1}\text{Var}(\hat{\mathbf{T}}_n/\boldsymbol{\delta})\left(\mathbf{X}_n^{\text{T}}\mathbf{W}_n\mathbf{X}_n\right)^{-1},$$

where $\hat{\mathbf{T}}_n = \left(\hat{t}_{0,n}, \hat{t}_{1,n}, \ldots, \hat{t}_{p,n}\right)^{\mathrm{T}}$, being

$$\hat{t}_{j,n} = \frac{1}{n} \sum_{i=1}^{n} (x_i - x)^j (1 - \delta_i) K_{h_n} (x_i - x) \left(\hat{m}_{S,g_n} (x_i) - m(x_i)\right), \quad 0 \le j \le p.$$
(21)

Using the approximation of the local polynomial estimator by equivalent kernels (see Sect. 3.2.2 of Fan and Gijbels 1996) $\hat{m}_{S,g_n}(x_i)$ can be written as

$$\hat{m}_{S,g_n}(x_i) = \sum_{s=1}^{n} \Psi^\delta \left(\frac{x_s - x_i}{g_n}\right) Y_s = \frac{1}{nf(x_i)\mathfrak{p}(x_i)} \sum_{s=1}^{n} L_{g_n,q}^* (x_s - x_i)\delta_s Y_s \left(1 + o_p(1)\right).$$
(22)

From (21) and (22) it follows that

$$\begin{aligned}
\mathrm{Cov}\left(h_n^{-j}\hat{t}_{j,n}, h_n^{-k}\hat{t}_{k,n}/\delta\right) &= \frac{1}{n^4 h_n^{j+k}} \sum_i K_{h_n}(x_i - x)(1 - \delta_i)(x_i - x)^j \\
&\quad \cdot \sum_t K_{h_n}(x_t - x)(1 - \delta_t)(x_t - x)^k \\
&\quad \times \frac{1}{f(x_i)\mathfrak{p}(x_i)} \frac{1}{f(x_t)\mathfrak{p}(x_t)} \\
&\quad \cdot \sum_s L_{g_n,q}^*(x_s - x_i)\delta_s \sum_r L_{g_n,q}^*(x_r - x_t)\delta_r c(|s - r|) \\
&= \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4,
\end{aligned}$$
(23)

where we have split $\mathrm{Cov}\left(h_n^{-j}\hat{t}_{j,n}, h_n^{-k}\hat{t}_{k,n}\right)$ into four terms: in Δ_1 we have considered the case $s = r$ and $i = t$; in Δ_2, $s = r i \ne t$; in Δ_3, $s \ne r i = t$; and, in Δ_4, $s \ne r$ and $i \ne t$.

Developing each of these four terms, we obtain that

$$\Delta_1 = O\left(\frac{1}{n^2 g_n^2}\right),$$

$$\Delta_2 = \sigma_\varepsilon^2 c(0) \frac{f(x)\mathfrak{q}(x)^2}{n\mathfrak{p}(x)} \frac{h_n}{g_n^2} \int A_{j,q}(v) A_{k,q}(v) \, dv \left(1 + o_p(1)\right),$$

$$\Delta_3 = O\left(\frac{1}{n^2 g_n^2}\right),$$

$$\Delta_4 = \frac{1}{n}\left(c(\varepsilon) - \sigma_\varepsilon^2 c(0)\right) \mathfrak{q}(x)^2 \frac{h_n}{g_n^2} f(x) \int A_{j,q}(v) A_{k,q}(v) \, dv \left(1 + o_p(1)\right).$$

Therefore,

$$\mathrm{Var}(\hat{\mathbf{T}}_n/\delta) = \frac{h_n}{n g_n^2} f(x) \frac{\mathfrak{q}(x)^2}{\mathfrak{p}(x)^2} c^\delta(\varepsilon) \mathbf{Z} \left(1 + o_p(1)\right).$$
(24)

From (24) and Proposition 1, we deduce that

$$\text{Var}\left(\boldsymbol{\Gamma}_2/\boldsymbol{\delta}\right) = \frac{h_n}{ng_n^2}\frac{1}{f(x)}\frac{\mathfrak{q}(x)^2}{\mathfrak{p}(x)^2}c^{\delta}(\varepsilon)\,\mathbf{H}_n^{-1}\mathbf{S}_K^{-1}\mathbf{Z}\mathbf{S}_K^{-1}\mathbf{H}_n^{-1}\left(1+o_p(1)\right). \qquad (25)$$

Finally, we study the term $\text{Cov}(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2/\boldsymbol{\delta})$,

$$\text{Cov}(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2/\boldsymbol{\delta}) = \left(\mathbf{X}_n^t\mathbf{W}_n\mathbf{X}_n\right)^{-1}\text{Cov}(\mathbf{T}_n\tilde{\mathbf{T}}_n^t/\boldsymbol{\delta})\left(\mathbf{X}_n^t\mathbf{W}_n\mathbf{X}_n\right)^{-1}, \qquad (26)$$

where $\mathbf{T}_n$ is given in (3) and $\tilde{\mathbf{T}}_n = \left(\tilde{t}_{0,n}, \tilde{t}_{1,n}, \ldots, \tilde{t}_{p,n}\right)^t$, being

$$\tilde{t}_{i,n} = \frac{1}{n}\sum_{t=1}^{n}(x_t - x)^i\,K_{h_n}(x_t - x)\left(\hat{m}_{S,g_n}(x_t) - m(x_t)\right)(1 - \delta_t), \quad 0 \le i \le p.$$

Using (22) we can expand the terms of matrix $\text{Cov}(\mathbf{T}_n\tilde{\mathbf{T}}_n^t/\boldsymbol{\delta})$, and we have

$$\text{Cov}\left(h_n^{-j}\hat{t}_{j,n}, h_n^{-k}\hat{t}_{k,n}/\boldsymbol{\delta}\right) = \frac{1}{n^2 h_n^{-j-k}}\sum_i K_{h_n}(x_i - x)(x_i - x)^j\,\delta_i\sum_t K_{h_n}(x_t - x)$$

$$\cdot (x_t - x)^k(1 - \delta_t)\sum_s \Psi^{\delta}\left(\frac{x_s - x_t}{g_n}\right)\text{Cov}\left[\varepsilon_i, \varepsilon_s\right]$$

$$= \Lambda_1 + \Lambda_2,$$

where in Δ_1 we assume that $i = s$, and in Δ_1 we consider $i \ne s$.

Simple algebraic expansions allow us to obtain that

$$\Lambda_1 = \frac{\sigma_\varepsilon^2 c(0)\mathfrak{q}(x)}{ng_n}f(x)\int K(v)\,v^j A_{k,q}(v)\,dv\left(1 + o_p(1)\right).$$

Using Taylor expansions we have

$$\Lambda_2 = \frac{\left(c(\varepsilon) - \sigma_\varepsilon^2 c(0)\right)}{n}\mathfrak{p}(x)\mathfrak{q}(x)f(x)\frac{1}{g_n}\left(\int K(v)(v)^j A_{k,q}(v)\,dv\right)\left(1 + o_p(1)\right),$$

Hence,

$$\text{Cov}\left(h_n^{-j}\hat{t}_{j,n}, h_n^{-k}\hat{t}_{k,n}/\boldsymbol{\delta}\right) = \frac{1}{ng_n}f(x)\frac{\mathfrak{q}(x)}{\mathfrak{p}(x)}c^{\delta}(\varepsilon)\int K(v)\,v^j A_{k,q}(v)\,dv\left(1 + o_p(1)\right),$$

and

$$\text{Cov}(\mathbf{H}_n^{-1}\mathbf{T}_n\mathbf{H}_n^{-1}\tilde{\mathbf{T}}_n^t/\boldsymbol{\delta}) = \frac{1}{ng_n}f(x)\frac{\mathfrak{q}(x)}{\mathfrak{p}(x)}c^{\delta}(\varepsilon)\,\tilde{\mathbf{Z}}\left(1 + o_p(1)\right). \qquad (27)$$

From (26), (27) and using again Proposition 1, we conclude that

$$\text{Cov}(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2/\boldsymbol{\delta}) = \frac{1}{ng_n}f(x)\frac{\mathfrak{q}(x)}{\mathfrak{p}(x)}c^{\delta}(\varepsilon)\,\mathbf{H}_n^{-1}\mathbf{S}_K^{-1}\tilde{\mathbf{Z}}\mathbf{S}_K^{-1}H_n^{-1}\left(1 + o_p(1)\right).$$

By substituting (20), (25) and (27) in it follows (10).

Proof of Theorem 4. The same method as that used in the demonstration of Theorem 2 is followed,

$$\sqrt{nh_n}\mathbf{H}_n\left(\left(\hat{\boldsymbol{\beta}}_{I,n}(x)\right)-\boldsymbol{\beta}(x)\right)=\sqrt{nh_n}\left(\mathbf{H}_n^{-1}\mathbf{U}_n\mathbf{H}_n^{-1}\right)^{-1}\mathbf{H}_n^{-1}\boldsymbol{\chi}_n,$$

where $\tilde{\boldsymbol{\chi}}=(\boldsymbol{\chi}_{0,n},\boldsymbol{\chi}_{1,n},\ldots,\boldsymbol{\chi}_{p,n})$, with

$$\begin{aligned}\boldsymbol{\chi}_{j,n}=&\frac{1}{n}\sum_{t=1}^{n}(x_t-x)^j\,K_{h_n}(x_t-x)\\&\times\left(\delta_t\varepsilon_t+(1-\delta_t)\left(\widehat{m}_{S,g_n}(x_t)-E\left(\widehat{m}_{S,g_n}(x_t)/\delta\right)\right)\right)\quad 0\le j\le p.\end{aligned}$$

Using Proposition 1 it is sufficient to prove the asymptotic normality conditional on δ of term $\sqrt{nh_n}\mathbf{H}_n^{-1}\tilde{\boldsymbol{\chi}}$. For it, let $Q_{I,n}$ be an arbitrary linear combination of $\mathbf{H}_n^{-1}\tilde{\boldsymbol{\chi}}_n$,

$$Q_{I,n}=\tilde{\mathbf{a}}^{\mathrm{T}}\mathbf{H}_n^{-1}\tilde{\boldsymbol{\chi}}_n=\sum_{i=0}^{p}a_i h_n^{-i}\tilde{\boldsymbol{\chi}}_{i,n},\quad\text{with}\quad\tilde{\mathbf{a}}=(a_0,a_1,\ldots,a_p)\in\mathbb{R}^{p+1}.$$

Using the approximation given in (22) we have

$$\sqrt{nh_n}Q_{I,n}=\sqrt{nh_n}\sum_{i=0}^{p}a_i(u_{i,n}+v_{i,n})=\sqrt{nh_n}\tilde{\mathbf{a}}^{\mathrm{T}}(\tilde{\mathbf{u}}_n+\tilde{\mathbf{v}}_n),$$

where $\tilde{\mathbf{u}}_n=(u_{0,n},\ldots,u_{p,n})^{\mathrm{T}}$ and $\tilde{\mathbf{v}}_n=(v_{0,n},\ldots,v_{p,n})^{\mathrm{T}}$ with

$$u_{i,n}=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{x_t-x}{h_n}\right)^i K_{h_n}(x_t-x)\varepsilon_t,\quad 0\le i\le p,$$

$$\begin{aligned}v_{i,n}=&\frac{1}{n}\sum_{t=1}^{n}\left(\frac{x_t-x}{h_n}\right)^i K_{h_n}(x_t-x)\\&\cdot\left(\frac{1-\delta_t}{nf(x_t)\mathfrak{p}(x_t)}\sum_{s=1}^{n}L_{g_n,q}^*(x_s-x_t)\delta_s\varepsilon_s\right),\quad 0\le i\le p.\end{aligned}$$

First, we compute the variance of $Q_{I,n}$. By (20), (25) and (26) and using the assumption A.7. we obtain

$$\begin{aligned}\sigma_{IS}^2&=\lim_{n\to\infty}\mathrm{Var}(Q_{I,n}/\delta)=\lim_{n\to\infty}nh_n\tilde{\mathbf{a}}^t\mathrm{Var}\left((\tilde{\mathbf{u}}+\tilde{\mathbf{v}})(\tilde{\mathbf{u}}+\tilde{\mathbf{v}})^t/\delta\right)\tilde{\mathbf{a}}\\&=\frac{c^\delta(\varepsilon)}{f(x)}\tilde{\mathbf{a}}^t\left(\tilde{\mathbf{S}}_K+\lambda^2\frac{\mathfrak{q}(x)^2}{\mathfrak{p}(x)^2}\mathbf{Z}+2\lambda\frac{\mathfrak{q}(x)}{\mathfrak{p}(x)}\tilde{\mathbf{Z}}\right)\tilde{\mathbf{a}}.\end{aligned}\tag{28}$$

Expanding the term $Q_{I,n}$ we obtain

$$
\sqrt{nh_n} Q_{I,n} = \sum_{t=1}^{n} \eta_{t,n}
$$
$$
= \sqrt{\frac{h_n}{n}} \, \Psi_{h_n}(x_t - x) \left(\delta_t \varepsilon_t + \frac{1 - \delta_t}{n f(x_t) \mathfrak{p}(x_t)} \sum_{s=1}^{n} L_{g_n,q}^* (x_s - x_t) \delta_s \varepsilon_s \right).
$$

Again, taking into account the form of functions $\Psi_{h_n}(u)$ and $L_{g_n,q}^*(u)$, that the kernels K and L have bounded support, reordering the sums and if $\lambda = 1$, we have

$$
\sqrt{nh_n} Q_{I,n} = \sum_{t=-|h_n n|}^{|h_n n|} \zeta_{t,n} = \sum_{t=-|h_n n|}^{|h_n n|} (\hat{\zeta}_{t,n} + \bar{\zeta}_{t,n}),
$$

with

$$
\zeta_{t,n} = \sqrt{\frac{h_n}{n}} \, \Psi_{h_n}(x_t - x) \, \delta_t \left(1 + \frac{1 - \delta_t}{n f(x_t) \mathfrak{p}(x_t)} \sum_{j=-|h_n n|+t}^{|h_n n|-t} L_{g_n,q}^* (x_t - x_j) \right) \varepsilon_t,
$$

where $q(n) = 2 |h_n n|$, $\zeta_{t,n}$ is a triangular array with $m(n)$-dependent main part $\hat{\zeta}_{t,n}$ and residual part $\bar{\zeta}_{t,n}$ obtained when substituting in $\zeta_{t,n}$, ε_t by $\tilde{\varepsilon}_t$ and $\bar{\varepsilon}_t$, respectively.

Under assumptions A.1. and A.7. we have

$$
\zeta_{t,n} = \sqrt{\frac{h_n}{n}} \, \Psi_{h_n}(x_t - x) \, \delta_t (1 + C) \, \varepsilon_t = \xi_{t,n}(1 + C). \tag{29}
$$

Using (28) and (29) and reasoning in a similar way as that in the proof of Theorem 2, it is easy to prove that the arrays $\zeta_{t,n}$ and $\bar{\zeta}_{t,n}$ satisfy the conditions of Theorem A.5. Now the proof of Theorem 4 is complete.

References

Aerts, M., Claeskens, G., Hens, N., Molenberghs, G. (2002). Local multiple imputation. *Biometrika, 89(2)*, 375–388.

Beveridge, S. (1992). Least squares estimation of missing values in time series. *Communications in Statistics. Theory and Methods, 21(12)*, 3479–3496.

Brockwell, P. J., Davis, R. A. (1991). *Time series. Theory and methods*, 2 edn. Heidelberg: Springer.

Chen, J. H., Shao, J. (2000). Nearest neighbor imputation for survey data. *Journal of Official Statistics, 16*, 113–131.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association, 89(425)*, 81–87.

Chow, G. C., Lin, A. L. (1976). Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association, 71(355)*, 719–721.

Chu, C. K., Cheng, P. E. (1995). Nonparametric regression estimation with missing data. *Journal of Statistical Planning and Inference, 48*, 85–99.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.

Francisco-Fernández, M., Vilar-Fernández, J. M. (2001). Local polynomial regression estimation with correlated errors. *Communications in Statistics. Theory and Methods, 30(7)*, 1271–1293.

Glasser, M. (1964). Linear regression analysis with missing observations among the independent variables. *Journal of the American Statistical Association, 59*, 834–844.

González-Manteiga, W., Pérez-González, A. (2004). Nonparametric mean estimation with missing data. *Communications in Statistics. Theory and Methods, 33(2)*, 277–303.

Hall, P., Van Keilegom, I. (2003). Using differences based methods for inference in nonparametric regression with time-series errors. *Journal of the Royal Statistical Society: Series B, 65*, 443–456.

Härdle, W., Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics, 81(1)*, 223–242.

Härdle, W., Tsybakov, A., Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference, 68(2)*, 221–245.

Herrmann, E., Gasser, T., Kneip, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika, 79(4)*, 783–795.

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association, 85(411)*, 765–769.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics, 22(3)*, 389–395.

Little, R. J. A. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association, 87*, 1227–1237.

Little, R. J. A., Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Masry, E. (1996a). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis, 17(6)*, 571–599.

Masry, E. (1996b). Multivariate regression estimation—local polynomial fitting for time series. *Stochastic Processes and their Applications, 65(1)*, 81–101.

Masry, E., Fan, J. (1997). Local polynomial estimation of regression function for mixing processes. Scandinavian Journal of Statistics. *Theory and Applications, 24*, 165–179.

Meng, X. L. (2000). Missing data: Dial M for ???. *Journal of the American Statistical Association, 95(452)*, 1325–1330.

Müller, H., Stadtmüller, U. (1988). Detecting dependencies in smooth regression models. *Biometrika, 75(4)*, 639–650.

Nieuwenhuis, G. (1992). Central limits theorems for sequences with m(n)-dependent main part. *Journal of Statistical Planning and Inference, 32*, 229–241.

Peña, D., Tiao, G. C. (1991). A note on likelihood estimation of missing values in time series. *The American Statistician 45(3)*, 212–213.

Robinson, P. M. (1984). Kernel estimation and interpolation for time series containing missing observations. *Annals of the Institute of Statistical Mathematics, 36(1)*, 403–417.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. In *Wiley series in probability and mathematical statistics: applied probability and statistics*. New York: Wiley.

Vilar, J., Vilar, J. (1998). Recursive estimation of regression function by local polynomial fitting. *Annals of the Institute of Statistical Mathematics, 50(4)*, 729–754.

Wang, Q., Linton, O., Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association, 99(466)*, 334–345.

Wang, W., Rao, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *Annals of Statistics, 30(3)*, 896–924.

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture, 1*, 129–142.