Bernoulli **10**(1), 2004, 49–78

Estimation and testing in a partial linear regression model under long-memory dependence

GERMÁN ANEIROS-PÉREZ[1], WENCESLAO GONZÁLEZ-MANTEIGA[2] and PHILIPPE VIEU[3]

[1]*Departamento de Matemáticas, Facultad de Informática, Universidad de La Coruña, 15071 La Coruña, Spain. E-mail: ganeiros@udc.es*
[2]*Departamento de Estatística e Investigación Operativa, Facultad de Matemáticas, Universidad de Santiago de Compostela, 15706 Santiago, Spain. E-mail: wences@zmat.usc.es*
[3]*Laboratoire de Statistique et Probabilités, UMR CNRS C55830, Université Paul Sabatier, 31062 Toulouse Cedex, France. E-mail: vieu@cict.fr*

We discuss estimation and testing of hypotheses in a partial linear regression model, that is, a regression model where the regression function is the sum of a linear and a nonparametric component. We focus on the case where the covariables and the random noise do not necessarily have summable autocovariance functions, and the estimators and test statistics are based on kernel smoothing. We obtain the bias, variance and asymptotic distribution of both estimators for the parametric and nonparametric parts, as well as the asymptotic distributions of the statistics used, both under the null hypothesis and local alternatives. We thus generalize the results of Speckman and of Beran and Ghosh to the case of general structures for the autocovariance function and complete the results of González-Manteiga and Vilar-Fernández to the case of a partial linear regression model. Simulations and a real data example provide promising results for our tests.

Keywords: hypothesis testing; kernel smoothing; long-memory process; partial linear models

1. Introduction

Complete agreement on the definition of long-memory processes has not yet been reached. Many definitions in the literature are based on the behaviour at the origin of the spectral density of the process, or even on the asymptotic behaviour of the Allen variance (Beran 1994; Heyde and Yang 1997). In this paper, we adopt one of the more general definitions based on the spectral density. A second-order stationary process $\{W_i\}$ with covariances $r_W(k) \equiv \mathrm{cov}(W_1, W_{1+k})$ and spectral density $f(\lambda) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty} \exp(-\mathrm{i}k\lambda) r_W(k)$ is said to have long memory if $f(0) = \infty$. Thus, we have short memory if $0 < f(0) < \infty$, while the case $f(0) = 0$ is known as antipersistence.

Long-memory processes appear to be of great importance in many different real situations, including agriculture, economics, geophysics and hydrology; see Mandelbrot and van Ness (1968) for a pioneering study on mathematical models with long memory, and Beran (1992) for a survey. Processes with short memory are well known and often used in practice; see Box and Jenkins (1976). As Beran (1994, p. 52) wrote concerning antipersistence, 'in practice, this case is rarely encountered (though it may occur after overdifferencing), mainly because the condition $\sum_{k=-\infty}^{\infty} r_W(k) = 0$ is very unstable'; see Beran and Feng (2002) and the references therein for a discussion of antipersistence and some practical examples.

Several results on statistical inference for these three types of processes have been established and show that many properties that hold under one type do not hold under the others. For example, in the setting of fixed-design nonparametric regression and under suitable conditions, Hall and Hart (1990) showed that the optimal rate of convergence of the Nadaraya–Watson estimator for the trend function under long-memory errors is necessarily slower than under short-memory ones, while Beran and Feng (2002) obtained the result that, in the presence of antipersistent errors, the convergence rate of a local polynomial estimator (or a kernel estimator) is faster than for short-memory errors (thus, if the type of dependence structure is not taken into account, it can completely invalidate statistical inference). The reader will find other examples and references in the general monograph by Beran (1994).

Concerning the conditions imposed on these processes, attention has been paid to two parametric models: fractional Gaussian noise (Mandelbrot and Van Ness 1968) and the fractional autoregressive integrated moving average (FARIMA) process (Granger and Joyeux 1980). Depending on the value of the self-similarity and difference parameters of these two models, respectively, the models can have long memory, short memory or antipersistence. Throughout this paper, most of the conditions on the dependence of the processes will be based on links between finite sums of absolute covariances and smoothing parameters. Therefore, on the one hand, most of our results will not need any parametric assumption on the decay rates of r_W nor any Gaussian condition for the process, and on the other hand, most of them will hold under these three types of dependence structures. Nevertheless, because the case of short memory has been extensively studied and that of antipersistence is less interesting for statistical applications, we will focus on the long-memory case, in the sense that our assumptions will be justified under this structure.

This paper deals with the semiparametric regression model

$$y_i = r(x_{i1}, \ldots, x_{ip}, t_i) + \varepsilon_i = \boldsymbol{\zeta}_i^{\mathrm{T}}\boldsymbol{\beta} + m(t_i) + \varepsilon_i \qquad (i = 1, \ldots, n), \tag{1}$$

where r is the regression function, $\boldsymbol{\beta}$ is a p-vector of unknown parameters, m is an unknown smooth function, and ε_i is the random noise (we have denoted $\boldsymbol{\zeta}_i^{\mathrm{T}} = (x_{i1}, \ldots, x_{ip})$). This model is a special type of partial linear regression model and was introduced by Engle *et al.* (1986) to study the effect of weather on electricity demand. It is clear that it generalizes the linear model and restricts the multivariate nonparametric model; it is more flexible than the linear model and eliminates (or reduces) the 'curse of dimensionality'. See Härdle *et al.* (2000) for several examples concerned with practical problems involving partial linear regression models, and Aneiros-Pérez and Quintela-del-Río (2002) for a recent study of this model under short-memory errors.

As in Speckman (1988), Linton (1995), and Beran and Ghosh (1998), among others, we will assume that the x_{ij} are spread out in some fashion and are correlated with the points $t_i = (i - 0.5)/n$. More specifically, we will suppose that

$$x_{ij} = g_j(t_i) + \eta_{ij} \qquad (1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant p), \tag{2}$$

where the g_j are unknown smooth functions and η_{ij} are random variables with zero mean.

Most studies of model (1) assume short-memory (Aneiros-Pérez and Quintela-del-Río 2001) or particular long-memory conditions on η_{ij} and ε_i (Beran and Ghosh 1998). The objectives of this paper are, under the possible presence of general long-memory dependence, to tackle the estimation of $\boldsymbol{\beta}$ and m in (1) and to consider the problem of testing hypotheses on $\boldsymbol{\beta}$ or m in (1).

This paper is organized as follows. In Section 2 we present the estimators and some asymptotic properties (bias, variance and distribution), together with sufficient assumptions to obtain them. Section 3 is devoted to testing hypotheses on $\boldsymbol{\beta}$ and m in (1). We present the test statistics and their asymptotic distributions, under both the null and alternative hypotheses. In Section 4, the finite-sample behaviour of the proposed tests is illustrated with a simulation study, and a simple example based on real data is given. Section 5 is devoted to the illustration of how our general methodology applies to a typical class of long-dependent processes, including fractional Gaussian noise and FARIMA processes. In Section 6 some useful technical lemmas are given, and in Section 7 proofs of theorems are given.

2. Estimators and asymptotic properties

2.1. Notation

Let us first introduce some notation. In this paper, $[z]$ denotes the integer part of $z \in \mathbb{R}$, and for $f : [0, 1] \to \mathbb{R}$, we use $\mathbf{f}$ to denote $(f(t_1), \ldots, f(t_n))^{\mathrm{T}}$. Furthermore, for $\mathbf{v} = (v_1, \ldots, v_s) \in \mathbb{R}^s$, $|\mathbf{v}|^2$ means $\sum_{i=1}^{s} v_i^2$, and for the matrix $\mathbf{A} = (a_{ij})$, $\|\mathbf{A}\|_2$ denotes the L_2 norm (or spectral norm) of $\mathbf{A}$ (that is, $\|\mathbf{A}\|_2 = \max_{|\mathbf{v}|\neq 0}|\mathbf{A}\mathbf{v}|/|\mathbf{v}|$) and $\|\mathbf{A}\|_1$ and $\|\mathbf{A}\|_\infty$ mean $\max_j \sum_i |a_{ij}|$ and $\max_i \sum_j |a_{ij}|$, respectively. We also will use the notation

$$\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}, \qquad \mathbf{X} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n)^{\mathrm{T}} = (\mathbf{x}_1, \ldots, \mathbf{x}_p) = (x_{ij})_{i=1,\ldots,n;j=1,\ldots,p},$$

$$\boldsymbol{\eta} = (\boldsymbol{\eta}_1^*, \ldots, \boldsymbol{\eta}_n^*)^{\mathrm{T}} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p) = (\eta_{ij})_{i=1,\ldots,n;j=1,\ldots,p}, \qquad \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}.$$

For the autocovariance functions of the stationary processes $\{\varepsilon_i\}_{i=1}^n$ and $\{\eta_{ij}\}_{i=1}^n$ $(j = 1, \ldots, p)$ we will write

$$r_\varepsilon(k) = \mathrm{E}(\varepsilon_1 \varepsilon_{1+k}), \qquad r_{\eta_j}(k) = \mathrm{E}(\eta_{1j}\eta_{1+k,j}),$$

$$S_{\eta_j,n} = \sum_{k=0}^{n} |r_{\eta_j}(k), |(j = 1, \ldots, p), \qquad S_{\varepsilon,n} = \sum_{k=0}^{n} |r_\varepsilon(k)|, \qquad S_{\eta_0,n} = \max_{1\leqslant j\leqslant p} S_{\eta_j,n},$$

$$\mathbf{V}_\eta = \mathrm{E}(\boldsymbol{\eta}_i^* \boldsymbol{\eta}_i^{*\mathrm{T}}), \qquad \mathbf{V}_\varepsilon = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}).$$

Throughout the paper, we will always use the notion of convergence in distribution of a sequence of generic variables $Z_n = G_n(\boldsymbol{\varepsilon}, \mathbf{X})$ conditioned on the design matrix $\mathbf{X}$. We specify that the limit involved in the definition of convergence in distribution of these conditioned variables must be interpreted as holding in probability with respect to the distribution of x_{ij}.

2.2. The estimates

Robinson (1988) and Speckman (1988) proposed to estimate $\boldsymbol{\beta}$ in (1) by means of

$$\hat{\boldsymbol{\beta}}_b = (\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b)^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{y}}_b. \tag{3}$$

Furthermore, Speckman (1988) estimated $m(t)$ in (1) through the kernel estimator

$$\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b) = \sum_{i=1}^{n} w_{n,h}(t, t_i)(y_i - \boldsymbol{\zeta}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_b).$$

In these estimators, b and h are smoothing parameters such that $nb \to \infty$, $nh \to \infty$, $b \to 0$ and $h \to 0$ as $n \to \infty$, these being the usual conditions in the nonparametric literature (note that to estimate $\boldsymbol{\beta}$ in model (1) we must use nonparametric estimation). For any $(n \times q)$ matrix $\mathbf{A}$ $(q \geqslant 1)$, we denote $\tilde{\mathbf{A}}_z = (\mathbf{I} - \mathbf{W}_z)\mathbf{A}$ (for $z = b$ or $z = h$), where $\mathbf{W}_z = (w_{n,z}(t_i, t_j))_{i,j}$ with $w_{n,z}(\cdot, \cdot)$ being a weight function that can take different forms, thus providing different estimators. Speckman (1988) used general weights, while Robinson (1988) worked with Nadaraya–Watson weights. In this paper, we will focus on the Gasser–Müller weights with boundary kernels, that is:

$$w_{n,z}(t, t_j) = \begin{cases} z^{-1}\displaystyle\int_{(j-1)/n}^{j/n} K\left(\frac{t-u}{z}\right)\mathrm{d}u & \text{if } t \in [z, 1-z], \\ z^{-1}\displaystyle\int_{(j-1)/n}^{j/n} K_q\left(\frac{t-u}{z}\right)\mathrm{d}u & \text{if } t = qz \in [0, z), \\ z^{-1}\displaystyle\int_{(j-1)/n}^{j/n} K_q^*\left(\frac{t-u}{z}\right)\mathrm{d}u & \text{if } t = 1 - qz \in (1-z, 1], \end{cases} \tag{4}$$

where we have supposed that $t_j = (j - 0.5)/n$ and $t \in [0, 1]$. Conditions on the kernel functions K, K_q and K_q^* will be specified below. Another option could be to use local polynomial estimators, which were suggested in Stone (1977) and have the property of automatic boundary correction.

Under suitable conditions, including independent errors, Robinson (1988) and Speckman (1988) showed that $\hat{\boldsymbol{\beta}}_b$ is $n^{1/2}$-consistent for $\boldsymbol{\beta}$ and asymptotically normal. Furthermore, Speckman (1988) found that $\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)$ is $n^{v/(2v+1)}$-consistent for $m(t)$, where he assumed that m has $v \geqslant 2$ continuous derivatives on [0, 1]. As usual in many functional estimation problems (see Györfi *et al.* 1989; Bosq 1998), these rates of convergence obtained for independent and identically distributed samples are still the same in the setting of short-memory errors. In this context, Gao (1995), assuming a linear process on the errors, investigated the asymptotic normality of $\hat{\boldsymbol{\beta}}_b$, and the rates of strong convergence of $\hat{\boldsymbol{\beta}}_b$ and $\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)$, and Aneiros-Pérez (2001) proved the asymptotic normality of these estimators under α-mixing errors. In the particular long-memory context where

$r_\varepsilon(k) = ck^{-\alpha_\varepsilon}$ and $r_{\eta_j}(k) = c_j k^{-\alpha_j}$ $(k > 0,\ 0 < \alpha_\varepsilon,\ \alpha_j < 1,\ j = 1, \ldots, p)$, Beran and Ghosh (1998) obtained the $n^{1/2}$-consistency of $\hat{\boldsymbol{\beta}}_b$ by calculating the asymptotic orders for its bias and variance, together with the asymptotic normality of the properly normalized estimator. On the other hand, Gao and Ahn (1999), by approximating m in (1) using a finite series summation, constructed $n^{\alpha_\varepsilon/2}$- and $n^{(v+1)\alpha_\varepsilon/(2(v+1)+\alpha_\varepsilon)}$-consistent estimators for $\boldsymbol{\beta}$ and m, respectively. They assumed that η_{ij} and g_j in (2) are equal to zero and known, respectively. On the errors ε_i, they supposed either $r_\varepsilon(k) = ck^{-\alpha_\varepsilon}$ $(k > 0,\ 0 < \alpha_\varepsilon < 1)$ or summability conditions which involve x_{ij}, r_ε and the family of functions used to approximate m. As far as we know, partial linear regression models have not been studied under general long-memory conditions on the errors.

2.3. Some asymptotic properties

In this section we will need some of the following assumptions. Since we wish to make our study quite general, we keep these assumptions in quite a complicated form, but in Section 5 we will show how they are satisfied for common long-dependent processes.

Assumption 1.

(a) $\{\varepsilon_i\}_{i=1}^n$ *is a stationary process with* $\mathrm{E}(\varepsilon_1) = 0$ *and* $\mathrm{E}(\varepsilon_1^2) = \sigma_\varepsilon^2 < \infty$.

(b) $\{\eta_{ij}\}_{i=1}^n$ *is a stationary process with* $\mathrm{E}(\eta_{1j}) = 0$, $\mathrm{E}(\eta_{1j}^2) < \infty$ *and autocovariance function* $r_{\eta_j}(k)$ *monotone and convergent to* 0 *as* $k \to \infty$ $(j = 1, \ldots, p)$.

(c) $\{\varepsilon_i\}_{i=1}^n$ *is independent of* $\{\eta_{ij}\}_{i=1}^n$ $(j = 1, \ldots, p)$.

Assumption 2.

(a) $n^{-1}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta} \xrightarrow{P} \mathbf{V}_\eta$, *where* $\mathbf{V}_\eta$ *is a positive definite matrix.*

(b) $n(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta}(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1} \xrightarrow{P} \mathbf{A}$, *where* $\mathbf{A}$ *is a positive definite matrix.*

(c) $n^{1/2}(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\varepsilon} \xrightarrow{d} N(\mathbf{0}, \mathbf{A})$.

Assumption 3.

(a) *The function* m *has two continuous derivatives on* $[0, 1]$.

(b) *The functions* $g_1, \ldots, g_p$ *have two continuous derivatives on* $[0, 1]$.

Assumption 4. *The design points* t_i *are* $t_i = (i - 0.5)/n$ $(i = 1, \ldots, n)$.

Assumption 5.

(a) $K\colon \mathbb{R} \to \mathbb{R}$ *is a Hölder continuous and non-negative function with support* $[-1, 1]$. *Furthermore,* $\int K(u)\mathrm{d}u = 1$ *and* $\int uK(u)\mathrm{d}u = 0$.

(b) *For* $q \in [0, 1]$, $K_q(\cdot)$ $(K_q^*(\cdot))$ *has support* $[-1, q]$ $([-q, 1])$ *and is Hölder continuous,* $\int K_q(u)\mathrm{d}u = \int K_q^*(u)\mathrm{d}u = 1$ *and* $\int uK_q(u)\mathrm{d}u = \int uK_q^*(u)\mathrm{d}u = 0$.

(c) $\sup_{q\in[0,1]}\max\{|\int u^2K_q(u)\mathrm{d}u|,\ |\int u^2K_q^*(u)\mathrm{d}u|,\ \int K_q^2(u)\mathrm{d}u,\ \int K_q^{*2}(u)\mathrm{d}u\} < \infty$.

Assumption 6.
(a) $nb^8 \to 0$ *as* $n \to \infty$.
(b) $nh^2 \to \infty$ *as* $n \to \infty$.
(c) $b^2 = o(h)$.

Assumption 1 is quite usual in nonparametric time series (see Hall *et al.* 1995), while Assumptions 3, 4 and 6 are common in nonparametric kernel smoothing. Note that Assumption 5 is sufficient to avoid boundary effects, and the existence of such modified kernels was proven by Gasser *et al.* (1985). It is clear that Assumption 5(c) is satisfied if the kernels K_q and K_q^* are uniformly bounded in q. Assumption 2 is related to the long-memory setting and will be discussed specifically in Section 5.

Theorem 1. (a) *Under Assumptions* 1(b), 2(a) *and* 3–5, *if* $\mathrm{E}(\varepsilon_i) = 0$ *then we have that*

$$\mathrm{E}(\hat{\boldsymbol{\beta}}_b|\mathbf{X}) - \boldsymbol{\beta} = O(b^4 + n^{-2}) + O_p((b^4 + n^{-2})^{1/2}((nb)^{-1} S_{\eta 0,[nb]})^{1/2}).$$

(b) *Under Assumptions* 1, 2(a), 2(b), 3(b), 4 *and* 5 *and if, in addition,* $(b^2 + n^{-1}) S_{\varepsilon,n} \to 0$ *and* $(nb)^{-1} S_{\eta 0,[nb]} S^2_{\varepsilon,n} \to 0$ *as* $n \to \infty$, *then we have that*

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_b|\mathbf{X}) = n^{-1}\mathbf{A} + o_p(n^{-1}).$$

(c) *Under Assumptions* 1, 2(a), 2(c), 3–5 *and* 6(a), *and if, in addition,* $b^3 S_{\eta 0,[nb]} \to 0$, $(b^4 + n^{-2}) S_{\varepsilon,n} \to 0$ *and* $(nb)^{-1} S_{\eta 0,[nb]} S_{\varepsilon,n} \to 0$ *as* $n \to \infty$, *then, conditionally on* $\mathbf{X}$,

$$n^{1/2}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}).$$

Remark 1. Theorem 1 generalizes Speckman's (1988) Theorems 2 and 4 and Beran and Ghosh's (1998) Theorem 1 to the case of a general structure of autocovariances for η_{ij} and ε_i. The autocovariance structure of $\{\varepsilon_i\}$ has no effect on the conditional bias of $\hat{\boldsymbol{\beta}}_b$, but we can observe the effect of the autocovariance of $\{\eta_{ij}\}$. Furthermore, it must be noted that, in the case of Theorem 1(b, c), the conditions on the bandwidth depend on the autocovariance functions of η_{ij} and ε_i. We observe that, under assumptions of Theorem 1(a, b) together with $nb^8 \to 0$ and $b^3 S_{\eta 0,[nb]} \to 0$ as $n \to \infty$, the bias of $\hat{\boldsymbol{\beta}}_b$ is of smaller order than its variance.

In the next theorem we reveal some asymptotic properties of $\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)$. To obtain the asymptotic normality of the properly normalized nonparametric estimator, we need to assume additional conditions on $\{\varepsilon_i\}_{i=1}^n$. Specifically, one of the following assumptions must be satisfied:

Assumption 7. $\{\varepsilon_i\}_{i=1}^n$ *is a stationary Gaussian process with* $\mathrm{E}(\varepsilon_1) = 0$, $\mathrm{var}(\varepsilon_1) = \sigma_\varepsilon^2 < \infty$ *and autocovariance function* $r_\varepsilon(k) = \sigma_\varepsilon^2 L(k) k^{-\alpha_\varepsilon}$ $(k > 0)$, *where* $0 < \alpha_\varepsilon < 1$ *is a fixed constant and* L *is a function defined on* $[0, \infty)$, *slowly varying and positive in some neighbourhood of infinity. Furthermore,* $C_L^* = \sup\{|L(x)|/|L(nh)| : 1 \leqslant x \leqslant nh, n \in \mathbb{N}\} < \infty$, $nh^{1+4/\alpha_\varepsilon}/L^{1/\alpha_\varepsilon}(nh) \to 0$ *and* $n^{\alpha_\varepsilon - 1} h^{\alpha_\varepsilon}/L(nh) \to 0$ *as* $n \to \infty$.

Assumption 8. *$\{\varepsilon_i\}_{i=1}^n$ is a mean-zero stationary process such that $\varepsilon_i = \sum_{j=0}^{\infty} c_j e_{j-i}$, where $\sum_{j=0}^{\infty} c_j^2 < \infty$, and the $\{e_i\}$ are independent and identically distributed with $\mathrm{E}(e_i) = 0$ and $\mathrm{E}(e_i^4) < \infty$. The spectral density of the process $\{\varepsilon_i\}$ is of the form $f_\varepsilon(\lambda) = |1 - \mathrm{e}^{\mathrm{i}\lambda}|^{\alpha_\varepsilon - 1} g(\lambda)$, where g is a continuous spectral density on $[-\pi, \pi]$ which is bounded both away from zero and infinity, and $0 < \alpha_\varepsilon < 1$.*

Let us introduce the following notation. We denote $C_{n,h,\alpha_\varepsilon,L} = ((nh)^{\alpha_\varepsilon}/L(nh))^{1/2}$ (with $L = 1$ if Assumption 8 holds) and $\sigma^2_{\alpha_\varepsilon,g,K} = s(\alpha_\varepsilon) g(0) \int_{-1}^{1}\int_{-1}^{1} K(u)K(v)|u-v|^{-\alpha_\varepsilon} \mathrm{d}u\, \mathrm{d}v$ (with $s(\alpha_\varepsilon) = 1$ and $g(0) = \sigma_\varepsilon^2$ if Assumption 7 holds, and $s(\alpha_\varepsilon) = 2\pi\Gamma(\alpha_\varepsilon)$ $(\Gamma((1-\alpha_\varepsilon)/2)\Gamma((1+\alpha_\varepsilon)/2))^{-1}$ if Assumption 8 holds).

We also need the following usual condition on the kernel:

Assumption 9. *The kernel K is symmetric and differentiable with bounded derivative.*

Theorem 2. (a) *Under the assumptions of Theorem* 1(a)*, together with Assumptions* 6(b) *and* 6(c)*, if $h^{-4} b^3 n^{-1} S_{\eta_0,[nb]} \to 0$ as $n \to \infty$, we have*

$$\mathrm{E}(\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)|\mathbf{X}) - m(t) = \mathrm{bias}(\hat{m}_h(t, \boldsymbol{\beta}))(1 + o_p(1)) = O_p(h^2).$$

(b) *Under the assumptions of Theorem* 1(b)*, and if r_ε is monotone and $r_\varepsilon(k) \to 0$ as $k \to \infty$, we have*

$$\mathrm{var}(\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)|\mathbf{X}) = \mathrm{var}(\hat{m}_h(t, \boldsymbol{\beta}))(1 + o_p(1)) = O_p((nh)^{-1} S_{\varepsilon,[nh]}).$$

(c) *Under the assumptions of Theorem* 1(a, b)*, together with Assumptions* 6(a)*,* 7 *or* 8*, and* 9*, and if, in addition, $h \log(nh) \to 0$ and $b^3 S_{\eta_0,[nb]} \to 0$ as $n \to \infty$, then, conditionally on $\mathbf{X}$,*

$$C_{n,h,\alpha_\varepsilon,L}(\hat{m}_h(u_1, \hat{\boldsymbol{\beta}}_b) - m(u_1), \ldots, \hat{m}_h(u_k, \hat{\boldsymbol{\beta}}_b) - m(u_k)) \xrightarrow{d} \sigma_{\alpha_\varepsilon,g,K}(N_1, \ldots, N_k),$$

where, for each fixed $k \in \mathbb{N}$, $0 < u_1 < \cdots < u_k < 1$ are arbitrary fixed points and $N_1, \ldots, N_k$ are independent standard normal variables.

Remark 2. As in Theorem 1 above, the conditions on the bandwidths depend on the autocovariance functions of the processes η_{ij} and ε_i. Theorem 2(a, b) generalizes Speckman's (1988) Theorem 3 to the case of a general structure of autocovariances for η_{ij} and ε_i. As in Speckman (1988), and using the results of Csörgő and Mielniczuk (1995) and Deo (1997) on normality, we observe similar asymptotic behaviour for the nonparametric estimators $\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)$ and $\hat{m}_h(t, \boldsymbol{\beta})$. From Theorem 2(c), we have that the finite-dimensional distributions of the properly normalized estimator $\hat{m}_h(\cdot, \hat{\boldsymbol{\beta}}_b)$ converge to those of a Gaussian white noise process. This is a common fact in fixed-design nonparametric regression (see Csörgő and Mielniczuk 1995; Deo 1997) but, as Csörgő and Mielniczuk (1999) showed, in the case of random-design nonparametric regression with long-memory errors, depending on the amount of smoothing employed, these finite-dimensional distributions can converge to a degenerate process with completely dependent marginals. It is natural to expect that this fact will remain in a partial linear regression model with random explanatory variable t.

3. Testing of hypotheses

In any regression analysis, it is very important to reduce the dimension of the vector of explanatory variables or, in general, to simplify the model. In the particular case of a partial linear regression model like (1), one can try to do this by testing the hypotheses

$$H_{0,\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{and} \quad H_{0,m} : m = m_0.$$

Under the assumption that the errors are independent and identically distributed, Gao (1997) designed a method for testing the hypothesis $H_{0,\boldsymbol{\beta}}$, where the statistic used was based on m approximated by a B-spline function. Actually, in many cases the observations are sequentially gathered in time, and then there is correlation among the errors. Taking this fact into account, González-Manteiga and Aneiros-Pérez (2003) assumed that $\{\boldsymbol{\eta}_i^*\}$ are independent and identically distributed and $\{\varepsilon_i\}$ is an MA(∞) short-memory process, and they tested $H_{0,\boldsymbol{\beta}}$ and $H_{0,m}$ by using statistics based on kernel smoothing (see (5) and (6) below). As we have noted in Section 1, there exist settings where the assumption of short-memory processes is restrictive. In this section we focus on the problem of testing the hypotheses $H_{0,\boldsymbol{\beta}}$ and $H_{0,m}$ under the possible presence of long-memory dependence in η_{ij} and/or ε_i. For this, we will use the same statistics as González-Manteiga and Aneiros-Pérez (2003), that is,

$$d^2_{\boldsymbol{\beta}}(\hat{r}_n, H_{0,\boldsymbol{\beta}}) = n^{-1}\sum_{i=1}^{n}(\hat{r}_n(\boldsymbol{\zeta}_i^{\mathrm{T}}, t_i) - \hat{r}_{n,\beta_0}(\boldsymbol{\zeta}_i^{\mathrm{T}}, t_i))^2 = (\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0)^{\mathrm{T}}(n^{-1}\tilde{\mathbf{X}}_h^{\mathrm{T}}\tilde{\mathbf{X}}_h)(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0) \qquad (5)$$

and

$$d^2_{m}(\hat{r}_n, H_{0,m}) = n^{-1}\sum_{i=1}^{n}(\hat{r}_n(\boldsymbol{\zeta}_i^{\mathrm{T}}, t_i) - \hat{r}_{n,m_0}(\boldsymbol{\zeta}_i^{\mathrm{T}}, t_i))^2 = n^{-1}\sum_{i=1}^{n}(\hat{m}_h(t_i, \hat{\boldsymbol{\beta}}_b) - m_0(t_i))^2, \qquad (6)$$

where $\hat{r}_n(\boldsymbol{\zeta}^{\mathrm{T}}, t) = \boldsymbol{\zeta}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_b + \hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)$, $\hat{r}_{n,\beta_0}(\boldsymbol{\zeta}^{\mathrm{T}}, t) = \boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\beta}_0 + \hat{m}_h(t, \boldsymbol{\beta}_0)$ and $\hat{r}_{n,m_0}(\boldsymbol{\zeta}^{\mathrm{T}}, t) = \boldsymbol{\zeta}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_b + m_0(t)$. The motivation for (5) and (6) is clear. Taking into account that under general conditions (see Theorems 1 and 2) $\hat{r}_n(\boldsymbol{\zeta}^{\mathrm{T}}, t)$ is a consistent estimator for the regression function $r(\boldsymbol{\zeta}^{\mathrm{T}}, t) = \boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\beta} + m(t)$, together with the fact that, under the null hypotheses $H_{0,\boldsymbol{\beta}}$ and $H_{0,m}$, $r(\boldsymbol{\zeta}^{\mathrm{T}}, t)$ also is consistently estimated by $\hat{r}_{n,\beta_0}(\boldsymbol{\zeta}^{\mathrm{T}}, t)$ and $\hat{r}_{n,m_0}(\boldsymbol{\zeta}^{\mathrm{T}}, t)$, respectively, we have that (5) and (6) are natural statistics for testing the parametric hypothesis $H_{0,\boldsymbol{\beta}}$ and the nonparametric hypothesis $H_{0,m}$, respectively. Observe that under $H_{0,m}$ we could use the least-squares estimator of $\boldsymbol{\beta}$ in the model $y_i - m_0(t_i) = \boldsymbol{\zeta}_i^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i$, say $\hat{\boldsymbol{\beta}}$, which would be introduced in the expression $\hat{r}_{n,m_0}(\boldsymbol{\zeta}_i^{\mathrm{T}}, t_i)$ in (6) instead of $\hat{\boldsymbol{\beta}}_b$. As a result, the summand $\boldsymbol{\zeta}_i^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_b - \hat{\boldsymbol{\beta}})$ appears, the study of d_m^2 is more complicated and perhaps the final distance is less natural. Because of this, we have preferred to use $\hat{\boldsymbol{\beta}}_b$.

3.1. Parametric test

By means of the asymptotic distribution under $H_{0,\boldsymbol{\beta}}$ of the statistic $d^2_{\boldsymbol{\beta}}(\hat{r}_n, H_{0,\boldsymbol{\beta}})$ (see (5)), we will obtain the approximate Type I error of the test. To study the power of this test, and to obtain information about the Type II error, we introduce the following alternative hypotheses:

$$H_{1,\boldsymbol{\beta}}^{\mathbf{c},n} : \boldsymbol{\beta} = \boldsymbol{\beta}_1^n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-1/2}, \text{ where } \mathbf{c} \neq \mathbf{0} \ (p \times 1) \text{ is an arbitrary fixed vector;}$$

$$H_{1,\boldsymbol{\beta}} : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

To obtain the asymptotic behaviour of the statistic $d_{\boldsymbol{\beta}}^2(\hat{r}_n, H_{0,\boldsymbol{\beta}})$ under $H_{0,\boldsymbol{\beta}}$, $H_{1,\boldsymbol{\beta}}^{\mathbf{c},n}$ and $H_{1,\boldsymbol{\beta}}$, we will need the following additional assumption, which will be discussed in Section 5:

Assumption 10. $n^{-1}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta} \xrightarrow{P} \sigma_\varepsilon^2\mathbf{V}_\eta$, *where* $\mathbf{V}_\eta$ *was defined in Assumption* 2(a) *above.*

Theorem 3. *Suppose that the assumptions of Theorem* 1(c)*, together with Assumptions* 2(b) *and* 10*, hold. Then, conditionally on* $\mathbf{X}$*, we have that:*

(a) *Under the null hypothesis* $H_{0,\boldsymbol{\beta}}$,

$$F(b, h) \equiv \frac{nd_{\boldsymbol{\beta}}^2(\hat{r}_n, H_{0,\boldsymbol{\beta}})}{\sigma_\varepsilon^2} = \frac{(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0)^{\mathrm{T}}(\tilde{\mathbf{X}}_h^{\mathrm{T}}\tilde{\mathbf{X}}_h)(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}_0)}{\sigma_\varepsilon^2} \xrightarrow{d} \chi_p^2,$$

where χ_p^2 *denotes the chi-square distribution with p degrees of freedom.*

(b) *Under* $H_{1,\boldsymbol{\beta}}^{\mathbf{c},n}$,

$$F(b, h) \xrightarrow{d} \chi_p^2(\theta), \tag{7}$$

$\chi_p^2(\theta)$ *being a chi-square distribution with p degress of freedom and non-centrality parameter* $\theta = \sigma_\varepsilon^{-2}\mathbf{c}^{\mathrm{T}}\mathbf{V}_\eta\mathbf{c}$.

(c) Under $H_{1,\boldsymbol{\beta}}$,

$$F(b, h) \to \infty \qquad \text{as } n \to \infty.$$

Remark 3. Note that the local alternative hypothesis $H_{1,\boldsymbol{\beta}}^{\mathbf{c},n}$ depends on *n*, and therefore the notion of convergence in distribution should be specified. Indeed, expression (7) should be understood as

$$\forall\delta > 0, \quad P_{\beta_1^n}(F(b, h)|\mathbf{X} < \delta) - \Psi_\theta(\delta) \xrightarrow{P} 0, \qquad \text{as } n \to \infty,$$

where Ψ_θ is the $\chi_p^2(\theta)$ distribution function.

Remark 4. It is clear that σ_ε^2 in $F(b, h)$ can be replaced by any consistent estimator $\hat{\sigma}_\varepsilon^2$ without affecting Theorem 3. Therefore, in practice, in testing $H_{0,\boldsymbol{\beta}}$ at a given significance level α by means of the asymptotic distribution proven in Theorem 3(a), $H_{0,\boldsymbol{\beta}}$ is rejected if

$$d_{\boldsymbol{\beta}}^2(\hat{r}_n, H_{0,\boldsymbol{\beta}}) > n^{-1}\hat{\sigma}_\varepsilon^2\Psi^{-1}(1-\alpha),$$

where Ψ is the χ_p^2 distribution function. A natural estimate of σ_ε^2 is $\hat{\sigma}_\varepsilon^2 = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^2$, where $\hat{\varepsilon}_i = y_i - \hat{r}_n(\boldsymbol{\zeta}_i^{\mathrm{T}}, t_i)$ are residuals from the semi-parametric fit. Observe that Theorem 3 complements Theorem 2.1 of González-Manteiga and Aneiros-Pérez (2003), who obtained a similar result but unconditionally on the design matrix and assuming that $\boldsymbol{\eta}_i^*$ and ε_i are independent and identically distributed and short-memory processes, respectively.

3.2. Nonparametric test

In the nonparametric test, acceptance of the null hypothesis yields a p-variate linear regression model, and the estimation errors can be large if we wrongly accept this hypothesis. So there really is a need for information about the power of the test, and for this purpose we will look at local alternatives of the form

$$H_{1,m}^{c_n} : m(t_i) = m_1^n(t_i) = m_0(t_i) + c_n m^*(t_i) \quad (i = 1, \ldots, n).$$

To obtain the asymptotic behaviour of the statistic $d_m^2(\hat{r}_n, H_{0,m})$ (see (6)) under $H_{0,m}$ and $H_{1,m}^{c_n}$ we will need the following additional assumptions:

Assumption 11. *The errors ε_i are generated by a stationary causal process $\varepsilon_i = \sum_{j=0}^{\infty} b_j e_{i-j}$, where $\{e_i\}$ is a sequence of independent and identically distributed random variables with zero mean, zero kurtosis and $\mathrm{E}(|e_1|^{4+2\lambda}) < \infty$ (for some $\lambda > 0$). Furthermore, $r_\varepsilon(k) = \mathrm{var}(e_1)\sum_{j=0}^{\infty} b_j b_{j+k}$ $(k \geqslant 0)$ is absolutely summable and $\sum_{k=1}^{\infty} k|r_\varepsilon(k)| < \infty$.*

Assumption 12. $\mathrm{E}(|\eta_{1j}|^{4\lambda/(2+\lambda)}) < \infty$, $j = 1, \ldots, p$.

Assumption 13. $nh^{3/2} \to \infty$ *and* $hn^{(2+\lambda)/(2\lambda+2)} \to 0$ *as* $n \to \infty$.

Assumption 14. m^* *is twice continuously differentiable and non-identically null on* $[0, 1]$.

Assumptions 11–13 are quite usual in such problems (see González-Manteiga and Vilar-Fernández 1995; Biedermann and Dette 2000), while Assumption 14 is obviously necessary to make possible the application of the results of Section 2. Assumption 11 will be discussed in Section 5.

Theorem 4. *Suppose that the assumptions of Theorem* 1(a, b)*, together with Assumptions* 6(a) *and* 11–14*, hold. Then, if $b^3 S_{\eta_0,[nb]} \to 0$ as $n \to \infty$, we have that, conditionally on* $\mathbf{X}$

(a) *Under the null hypothesis $H_{0,m}$,*

$$(n^2 h)^{1/2}\left(d_m^2(\hat{r}_n, H_{0,m}) - \frac{\sum_{s=-\infty}^{\infty} r_\varepsilon(s) \int K^2}{nh} \right) \xrightarrow{d} N(0, \sigma_d^2),$$

where $\sigma_d^2 = 2(\sum_{k=-\infty}^{\infty} r_\varepsilon(k))^2 \int (K * K)^2$ and $K * K$ denotes the convolution of K with itself.

(b) *Under* $H^{c_n}_{1,m}$, *if* $c_n = (n^2 h)^{-1/4}$ *then*

$$(n^2 h)^{1/2}\left(d^2_m(\hat{r}_n,\, H_{0,m}) - \frac{\sum_{s=-\infty}^{\infty} r_\varepsilon(s)\int K^2}{nh}\right) \xrightarrow{d} N\left(\int m^*(u)^2 \mathrm{d}u,\, \sigma^2_d\right),$$

and, if $(n^2 h)^{1/4} c_n \to \infty$ *as* $n \to \infty$, *then*

$$(n^2 h)^{1/2}\left(d^2_m(\hat{r}_n,\, H_{0,m}) - \frac{\sum_{s=-\infty}^{\infty} r_\varepsilon(s)\int K^2}{nh}\right) \to \infty.$$

Remark 5. From Theorem 4, we observe that the use of the statistic $d^2_m(\hat{r}_n,\, H_{0,m})$ allows the detection of alternatives at a distance of $(n^2 h)^{-1/4}$. In practice, in testing $H_{0,m}$ at a given significance level α by means of the asymptotic distribution proven in Theorem 4(a), $H_{0,m}$ is rejected if

$$d^2_m(\hat{r}_n,\, H_{0,m}) > n^{-1}h^{-1/2}\hat{\sigma}_d\Phi(1-\alpha) + n^{-1}h^{-1}\hat{S}\int K^2,$$

where $\hat{S}$ is an estimate of $S = \sum_{k=-\infty}^{\infty} r_\varepsilon(k)$, $\hat{\sigma}^2_d = 2\hat{S}^2 \int (K * K)^2$ and Φ is the standard normal distribution. As for $\hat{S}$, under the assumption that the errors ε_i follow an AR(k) model, Aneiros-Pérez and Quintela-del-Río (2002) proposed a consistent estimator for S, which can be generalized to ARMA(k, z) models. This was based on consistent estimators of the parameters of the AR(k) model, and it used the residuals of the semi-parametric fit. On the other hand, González-Manteiga and Aneiros-Pérez (2003) constructed a consistent estimator for S which does not assume a parametric model on the errors. This was based on second-order differences $\hat{\hat{\varepsilon}}_{i,k,s,b} = \hat{y}^*_{i,b} - k(k+s)^{-1}\hat{y}^*_{i+s,b} - s(k+s)^{-1}\hat{y}^*_{i-k,b}$ (where $\hat{y}^*_{i,b} = y_i - \zeta_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_b$), and is an extension of the Herrmann *et al.* (1992) estimator to the setting of a partial linear regression model.

Remark 6. As seen in the proof of Theorem 4, its results remain valid if we replace 'assumptions of Theorem 1(a, b)' in its statement by 'assumptions of Theorem 1(a, c)'. Note that Theorem 4 extends Theorem 2.2 of González-Manteiga and Aneiros-Pérez (2003), who obtained a similar result but unconditionally on the design matrix and assuming that $\boldsymbol{\eta}^*_i$ are independent and identically distributed.

4. Simulation study and real data example

In this section we will extend Beran and Ghosh's (1998) simulation study and real data example. In their simulation study, they compared the least-squares estimate (LSE) of $\boldsymbol{\beta}$ applied to x and y directly with the semi-parametric estimate (3). In one of their two models the LSE of $\boldsymbol{\beta}$ was unbiased, and in the other it was biased. In the two cases, the semi-parametric estimate was better from the mean-square error point of view. The objective of their real data example was estimation of $\boldsymbol{\beta}$. In this section we will test hypotheses on $\boldsymbol{\beta}$ and m. The Epanechnikov kernel (modified at the boundaries) was used.

4.1. A simulation study

A modest simulation study was carried out to observe the finite-sample behaviour of the tests proposed (Theorems 3 and 4), together with the effect of the dependence and bandwidths. Models (28) and (29) in Beran and Ghosh (1998) were used for the parametric and nonparametric tests, respectively. Like these authors, we have used bandwidths of the type $b = n^{-\gamma}$ and $h = n^{-\theta}$, satisfying the assumptions of the theorems. For each possible combination of the sample sizes ($n = 200$, 500), parameters, bandwidths and functions considered, $M = 5000$ samples were generated. The null hypotheses $H_{0,\beta} : \beta = \beta_0$ and $H_{0,m}$: $m = m_0$ were tested at the significance level $\alpha = 0.1$. Tables 1 and 2 show the relative frequency of acceptance of $H_{0,\beta}$ and $H_{0,m}$, respectively, both assuming unknown and known σ_ε^2 and $\sum_{s=-\infty}^{\infty} r_\varepsilon(s)$.

Model (28) in Beran and Ghosh (1998) was $y_i = \zeta_i\beta + m(t_i) + \varepsilon_i$, where the $\zeta_i = x_{i1}$ were independent and identically distributed with $P(\zeta_t = 1) = P(\zeta_t = -1) = 0.5$ ($p = 1$, $g_1 = 0$), $m(t) = 2 + 20/(1 + t)$ and ε_i were generated by a FARIMA(0, d_ε, 0) process with unit variance. Values of $d_\varepsilon = 0$ and $d_\varepsilon = 0.3$ were considered (recall that $d_\varepsilon = 0$ means independent and identically distributed standard normal, and $0 < d_\varepsilon < 0.5$ means a long-memory process, with $r_\varepsilon(k) \sim c_{d_\varepsilon}|k|^{2d_\varepsilon - 1}$ as $|k| \to \infty$). The null hypothesis $H_{0,\beta} : \beta = 0$

Table 1. Relative frequency of acceptance of $H_{0,\beta} : \beta = 0$ $(1 - \alpha = 0.90)$

		$\beta = 0$		$\beta = 0.05$		$\beta = 0.1$	
d_ε	$b = h$	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
0	0.15	0.8906[a]	0.8936	0.8112	0.6960	0.5906	0.2798
		0.8954[b]	0.8946	0.8164	0.6990	0.5960	0.2830
	0.20	0.8910	0.8936	0.8104	0.6948	0.5892	0.2796
		0.8964	0.8948	0.8182	0.6966	0.5954	0.2798
	0.25	0.8920	0.8948	0.8106	0.6942	0.5880	0.2788
		0.8976	0.8948	0.8156	0.6948	0.5954	0.2810
0.3	0.15	0.9022	0.9142	0.8052	0.6714	0.5342	0.2200
		0.9370	0.9326	0.8534	0.7170	0.6108	0.2594
	0.20	0.9008	0.9072	0.7966	0.6666	0.5324	0.2206
		0.9332	0.9296	0.8464	0.7138	0.6098	0.2592
	0.25	0.8974	0.9038	0.7940	0.6642	0.5342	0.2208
		0.9314	0.9254	0.8464	0.7104	0.6074	0.2568

[a]Using the estimated variance.
[b]Using the true variance.

was tested under $\beta = 0$, $\beta = 0.05$ and $\beta = 0.1$. Because we have observed similar behaviours for $b = h$ and $b \neq h$ (for reasonable values of b and h), we only present the former case. The estimate $\hat{\sigma}_\varepsilon^2$ with pilot bandwidths $b_0 = h_0 = 0.25$ was used (see Remark 4).

Remark 7. We observe that the frequencies in the $\beta = 0$ column in Table 1 are close to $1 - \alpha = 0.90$, and they decrease as β increases. Therefore, the test may be judged as exhibiting desirable behaviour. Furthermore, we see that the test turns out to be quite insensitive both to bandwidth selection (in the case of β estimation, this was observed by Beran and Ghosh 1998) and to long memory, the level being similar with both $n = 200$ and $n = 500$ (convergence is achieved with moderate sample size) and the power increasing as the sample size increases.

In model (29) in Beran and Ghosh (1998), $\beta = 5$, $\zeta_i = x_{i1} = \log(1 + t_i) + \eta_{i1}$ and $m(t_i) = 2 + 20t_i$ were considered. Furthermore, η_{i1} were generated by a FARIMA(0, d_η, 0) process with unit variance and values of $d_\eta = 0$ and $d_\eta = 0.3$. These authors considered these same types of processes for ε_i. Nevertheless, if we assume that ε_i follows a FARIMA(0, d_ε, 0) process with $d_\varepsilon > 0$, then the assumptions of our Theorem 4 are not satisfied. For this reason, we have supposed an AR(1) process for ε_i, with parameter ρ_ε ($\rho_\varepsilon = 0$ (similar to the case of $d_\varepsilon = 0$) and $\rho_\varepsilon = 0.6$ were considered). The null hypothesis $H_{0,m} : m(t) = 2 + 20t$ was tested under $m(t) = 2 + 20t$, $m(t) = 2 + 20t^{1.05}$ and $m(t) = 2 + 20t^{1.1}$. $S = \sum_{k=-\infty}^{\infty} r_\varepsilon(k) = \sigma_\varepsilon^2(1 + \rho_\varepsilon)(1 - \rho_\varepsilon)^{-1}$ was estimated by means of $\hat{S} = \hat{\sigma}_\varepsilon^2(1 + \widehat{\rho}_\varepsilon)(1 - \widehat{\rho}_\varepsilon)^{-1}$, where $\widehat{\rho}_\varepsilon = (\sum_{i=1}^{n-1} \hat{\varepsilon}_i \hat{\varepsilon}_{i+1})(\sum_{i=1}^{n} \hat{\varepsilon}_i^2)^{-1}$. The pilot bandwidths $b_0 = h_0 = 0.25$ were used.

Remark 8. As in the parametric test, we observe that the nonparametric test may be judged as exhibiting desirable behaviour. Nevertheless, important differences are seen. On the one hand, the bandwidth h has greater influence than in the parametric case, and on the other hand, the level and the power are affected by the dependence in ε (not in η).

As noted above, the assumptions of our Theorem 4 include a short-memory structure for ε_i. To illustrate the possible asymptotic behaviour of the nonparametric test $d_m^2(\hat{r}_n, H_{0,m})$ under $H_{0,m}$, in the case of long-memory conditions on ε_i, $M = 100\,000$ values of $d_m^2(\hat{r}_n, H_{0,m})$ (with $b = 0.15$ and $h = 0.04$) were obtained from M samples of size $n = 200$. The density of $d_m^2(\hat{r}_n, H_{0,m})$ was then estimated by means of the Rosenblatt–Parzen estimator, with the plug-in bandwidth selector recommended by Jones *et al.* (1996). The same null hypothesis and structures for η as in Table 2 were considered, while ε_i was taken to be: (1) an AR(1) process with parameter $\rho_\varepsilon = 0.6$; (2) a FARIMA(0, 0.2, 0) process; and (3) a FARIMA(0, 0.4, 0) process. Each process had unit variance. Because we have observed that the corresponding estimated density is quite insensitive to the dependence structure in η, in Figure 1 we only show the combination of a FARIMA(0, 0.3, 0) process in η with cases (1)–(3) on ε. Furthermore, we show the approximation of the density of the test given by our Theorem 4, for case (1).

Table 2. Relative frequency of acceptance of $H_{0,m}: m(t) = 2 + 20t$ $(1 - \alpha = 0.90)$

			$m(t) = 2 + 20t$		$m(t) = 2 + 20t^{1.05}$		$m(t) = 2 + 20t^{1.1}$	
ρ_ε	d_η	(b, h)	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
0	0	(0.15, 0.03)	0.8222[a]	0.8388	0.1716	0.0030	0.0002	0.0000
			0.8426[b]	0.8460	0.1968	0.0038	0.0002	0.0000
		(0.15, 0.04)	0.8042	0.8278	0.1348	0.0018	0.0000	0.0000
			0.8274	0.8398	0.1492	0.0014	0.0002	0.0000
		(0.20, 0.03)	0.8224	0.8400	0.1716	0.0030	0.0002	0.0000
			0.8432	0.8464	0.1968	0.0038	0.0002	0.0000
		(0.20, 0.04)	0.8054	0.8294	0.1358	0.0018	0.0000	0.0000
			0.8278	0.8396	0.1498	0.0014	0.0002	0.0000
	0.3	(0.15, 0.03)	0.8180	0.8366	0.1744	0.0060	0.0004	0.0000
			0.8420	0.8426	0.1988	0.0058	0.0004	0.0000
		(0.15, 0.04)	0.7952	0.8224	0.1366	0.0034	0.0002	0.0000
			0.8214	0.8302	0.1550	0.0032	0.0002	0.0000
		(0.20, 0.03)	0.8198	0.8376	0.1744	0.0058	0.0004	0.0000
			0.8466	0.8442	0.1998	0.0062	0.0004	0.0000
		(0.20, 0.04)	0.8020	0.8238	0.1380	0.0034	0.0002	0.0000
			0.8256	0.8340	0.1576	0.0032	0.0004	0.0000
0.6	0	(0.15, 0.03)	0.8558	0.8426	0.6034	0.3118	0.1482	0.0030
			0.9618	0.8996	0.8348	0.4466	0.3532	0.0062
		(0.15, 0.04)	0.7860	0.8118	0.5030	0.2662	0.0948	0.0018
			0.9236	0.8754	0.7490	0.3572	0.2372	0.0042
		(0.20, 0.03)	0.8564	0.8434	0.6054	0.3124	0.1496	0.0030
			0.9630	0.8994	0.8342	0.4470	0.3530	0.0064
		(0.20, 0.04)	0.7868	0.8120	0.5048	0.2660	0.0958	0.0018
			0.9246	0.8756	0.7492	0.3582	0.2388	0.0042
	0.3	(0.15, 0.03)	0.8474	0.8418	0.5922	0.3132	0.1484	0.0038
			0.9626	0.8982	0.8392	0.4438	0.3592	0.0070
		(0.15, 0.04)	0.7658	0.8078	0.4978	0.2644	0.0932	0.0028
			0.9230	0.8752	0.7492	0.3664	0.2446	0.0034
		(0.20, 0.03)	0.8492	0.8434	0.5946	0.3144	0.1512	0.0038
			0.9638	0.9002	0.8414	0.4462	0.3600	0.0076
		(0.20, 0.04)	0.7680	0.8090	0.5006	0.2674	0.0942	0.0026
			0.9252	0.8772	0.7530	0.3684	0.2466	0.0034

[a]Using the estimated sum of covariances.
[b]Using the true sum of covariances.

Remark 9. Two interesting facts can be seen in Figure 1. On the one hand, sample size n and bandwidth h possibly cause the difference between the approximated density from Theorem 4 (solid line) and the theoretical density obtained from the Monte Carlo study (dotted line). On the other hand, bandwidth h and the long memory in the errors ε_i, in some cases, appear to

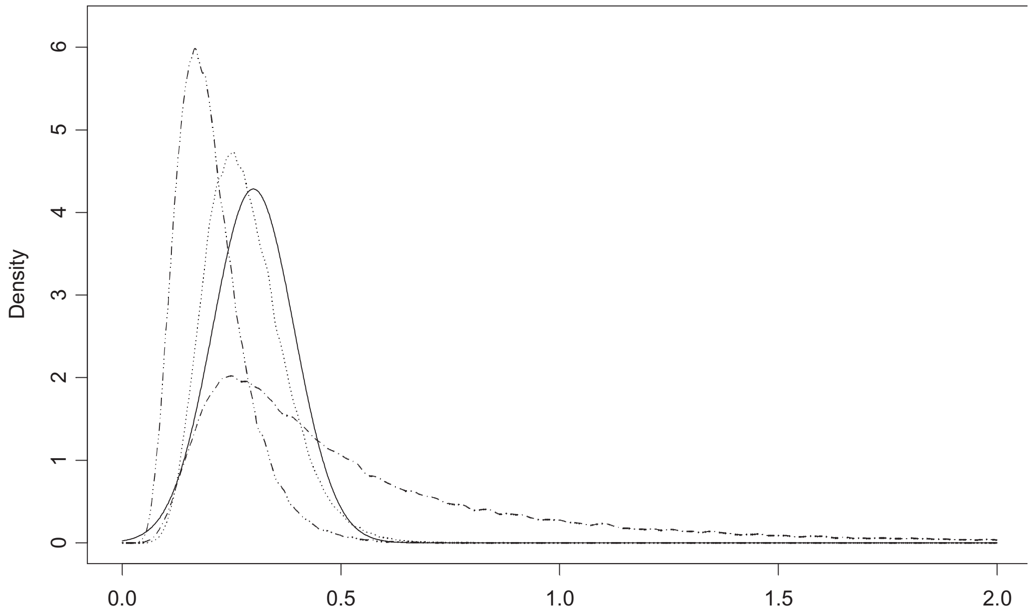

Figure 1. Distribution from Theorem 4 —— (approximation of ⋯), $\rho_\varepsilon = 0.6$ ⋯, $d_\varepsilon = 0.2$ -⋯-, $d_\varepsilon = 0.4$ - · -

destroy the asymptotic normality of the nonparametric test. This empirical fact has been proven in other settings; see Ho (1996) for the setting of density estimation.

Remark 10. As we suspected, the normal approximation of the statistics is worse for the nonparametric test than for the parametric (the relative frequency of acceptance of H_0 under H_0 is closest to $1 - \alpha$ in the parametric case; see Tables 1 and 2), this fact being motivated by the very slow speed of convergence of the corresponding statistic to the normal distribution (very large sample sizes for the nonparametric test are then necessary). To solve this problem, a bootstrap algorithm could be used to approximate the distribution of $d_m^2(\hat{r}_n, H_{0,m})$. In the setting of nonparametric regression ($\boldsymbol{\beta} = \mathbf{0}$), Vilar-Fernández and González-Manteiga (2000) suggested bootstrap mechanisms ('naive bootstrap' and 'wild bootstrap') for this goodness-of-fit test, assuming a stationary and reversible ARMA process for the errors. Their simulation study showed that bootstrap tests gave better behaviour with respect to the normal test, and it seems natural to expect a similar performance in our setting of partial linear regression models. In the context of long-memory errors, and without assuming a parametric structure on the errors, the 'block bootstrap' (Hall *et al.* 1995) could be one option. In any case, the extension of these and other resampling mechanisms both to the case of partial linear regression models and to the setting of long-memory dependence is an open problem.

4.2. A real example

Our Theorems 3 and 4 are illustrated by analysis of the average annual temperature for the southern hemisphere against both the average annual temperature for the northern hemisphere and the year (t_i). The data set was extracted from the database held at the Climate Research Unit of the University of East Anglia (http://www.cru.uea.ac.uk). We define $x_i = \sum_{j=1}^{12}(x_{ij} - n_j)/12$ and $y_i = \sum_{j=1}^{12}(y_{ij} - s_j)/12$, where n_j and s_j are the average monthly temperatures in the period 1950–1979 ($j = 1, \ldots, 12$) for the northern and southern hemispheres, respectively. x_{ij} and y_{ij} denote the average monthly temperatures in the month j of year i for the northern and southern hemispheres, respectively. The two series cover the period between 1856 and 2001. Thus the sample size is $n = 146$.

Beran and Ghosh (1998) studied the series x_i and y_i, corresponding to the period 1854–1989, from the point of view of β estimation. They modelled the residuals $\hat{\eta}_i$ and $\hat{\varepsilon}_i$ by a FARIMA(0, d, 0) process, with $d = 0.2$ and 0.1, respectively, and they obtained approximate confidence intervals for β and the parameters d. They showed that long memory is clearly present in η, but this fact is not clear in ε.

We analysed our residuals $\hat{\eta}_i$ and $\hat{\varepsilon}_i$ with the same type of bandwidth as these authors, and we have essentially obtained similar conclusions for the process η_i. Nevertheless, Box–Jenkins analysis of the residuals $\hat{\varepsilon}_i$ suggests that an AR(1) process could be a reasonable model for the errors ε_i (see Figure 2). Then, assuming an AR(1) process for ε_i and a FARIMA(0, 0.2, 0) process for η_i, we tested two important hypotheses: (1) $H_{0,\beta} : \beta = 0$ and (2) $H_{0,m} : m = 0$. We calculated the statistics $d_\beta^2(\hat{r}_n, H_{0,\beta})$ and $d_m^2(\hat{r}_n, H_{0,m})$ for a grid of values $(b, h) \in \{0.05 + 0.05i, i = 0, 1, \ldots, 6\} \times \{0.04 + 0.01j, j = 0, 1, 2, 3\}$ ($b = h \in \{0.05 + 0.05i, i = 0, 1, \ldots, 6\}$ were also used for $d_\beta^2(\hat{r}_n, H_{0,\beta})$). In all cases, the p-value was 0.0000.

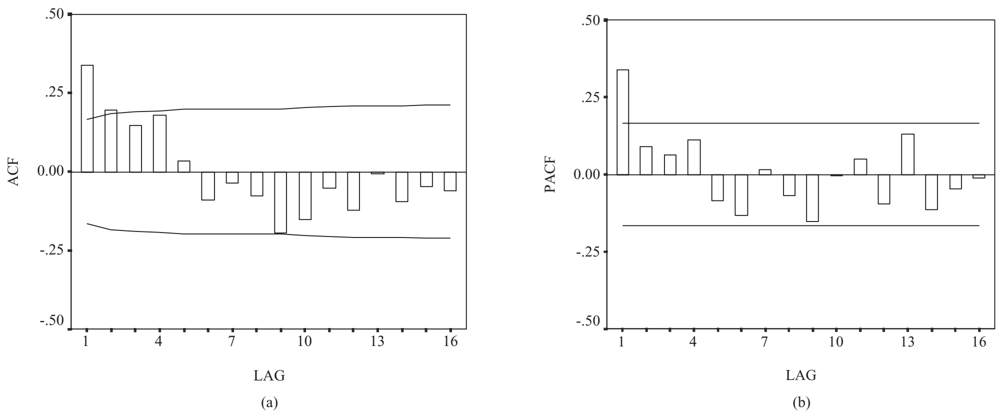

Figure 2. (a) Sample and (b) sample partial autocorrelation functions

5. Concluding remarks

For the purposes of this study, we have made our conditions on the dependence structures of the errors as general as we could. To highlight our approach, this section will include a discussion of some typical examples of long-dependent processes. More precisely, we will look deeply at fractional Gaussian noise and FARIMA processes.

First of all, let us note that most of the conditions we have introduced in this paper are quite usual in long-memory estimation. For instance, Assumptions 2(b) and 2(c) are quite common (see Beran and Ghosh 1998), and they are known to be satisfied, under weak conditions, both by fractional Gaussian noise and by FARIMA processes, as proven by Yajima (1991) and Künsch *et al.* (1993). Furthermore, concerning Assumption 2(a), it should be noted that this is a quite unrestrictive condition, satisfied for instance if $\boldsymbol{\eta}_i^*$ is ergodic; see Yajima (1988) and Beran and Ghosh (1998) for discussion of the use of ergodicity for noise regression modelling. Assumptions 7 and 8 are also quite usual in a nonparametric setting (see Csörgő and Mielniczuk 1995; Deo 1997), and it should be noted that they play a minor role in our work since they are only needed to prove Theorem 2(c). Finally, note that Assumption 10 is also easily satisfied in many cases. For instance, it is clear that if the ε_i are uncorrelated, then Assumption 10 reduces to Assumption 2(a), while on the other hand if the $\boldsymbol{\eta}_i^*$ are uncorrelated and $\sum_{k=1}^n r_\varepsilon^2(k) = o(n)$ (so that, ε_i may have long memory), then Assumption 10 holds. Our Assumption 11 is the only restrictive assumption of this paper since it implies that $\{\varepsilon_i\}$ is a short-memory process. The reason why we introduced it is that throughout the proof of Theorem 4 we need probabilistic results, which are still unknown for general memory structures, on the limit distribution of a properly normalized quadratic form in ε_i (specifically, in the study of $A_{1,n}$ below). At the time of writing, the only existing results are on special quadratic forms in special types of long-memory processes (Fox and Taqqu 1987), and they need to be extended to allow general coefficients depending on n and general types of long-memory processes (obtaining, possibly, central and non-central limit theorems). When ε_i satisfies Assumption 11, these results can be obtained by means of Nieuwenhuis's (1992) Theorem 2.3. Note, however, that Assumption 11 is only needed for Theorem 4, and observe that we still allow long-memory dependence for η_{ij}.

The only assumptions which are really specific to our work are those linking the autocovariance sums $S_{\varepsilon,n}$ and $S_{\eta_0,n}$ with the bandwidths h and b. Because they are new, these conditions have been emphasized by being presented within the statement of each theorem. It should be noted that these new conditions are easily satisfied by most of the usual long-dependent processes. For instance (see Beran 1994), many classes of long-memory processes have autocovariance of the form

$$|k|^{-D}L(|k|) \qquad (0 < D < 1), \tag{8}$$

L being some slowly varying function. It is easy to see that, if ε_i or η_{ij} is a process of this kind, we can write $S_{\varepsilon,n}$ and $S_{\eta_0,n}$ in the form

$$J(n)n^{-D+1}, \tag{9}$$

for some other slowly varying function J. Therefore, it is easy to check, one by one, that all the conditions linking $S_{\varepsilon,n}$ and $S_{\eta_0,n}$ with the bandwidths h and b hold for reasonable choices of bandwidth. For instance, looking at Theorem 1(b), it is clear that a reasonable bandwidth choice should make the trade-off between both components b^2 and n^{-1} of the rate of convergence, and so a good candidate is $b \sim Cn^{-1/2}$. For this value of b, it is easy to check by using (9), that if ε_i is long-dependent of the form of (8) then the condition $(b^2 + n^{-1})S_{\varepsilon,n} \to 0$ is always satisfied. Similarly, all the conditions of this type appearing in our work could be checked for processes ε_i or η_{ij} satisfying (8). To conclude this point, note that both fractional Gaussian noise and FARIMA processes are of type (8) (see Beran 1994).

Finally, we note that in most results the assumptions used are general, without restriction to a particular type of dependence. Therefore, Theorems 1, 2(a, b) and 3 hold when ε_i and/or η_{ij} are either antipersistent, short memory or long memory (observe that Theorems 2(c) and 4 use long- and short-memory conditions on the errors, respectively).

6. Some technical lemmas

The following lemmas will be used in the proof of the theorems.

Lemma 1. *Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a monotone function such that $f(k) \to 0$ as $k \to \infty$. Then, if nh is large enough,*

$$C_1(nh)^{-1}\sum_{k=0}^{[nh]} f(k) \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} w_{n,h}(t, t_i)w_{n,h}(t, t_j)f(|i-j|) \leqslant C_2(nh)^{-1}\sum_{k=0}^{[nh]} f(k),$$

where C_1 and C_2 are positive constants, $t \in [0, 1]$, $t_i = (i - 0.5)/n$ and $w_{n,h}(t, t_i)$ are the Gasser–Müller weights (see (4)) with kernel function satisfying Assumption 5.

Proof. Using the fact that $w_{n,h}(t, t_i) \neq 0 \Rightarrow t_i \in (t - h - 0.5n^{-1}, t + h + 0.5n^{-1})$, the proof of the lemma is a direct consequence of the assumptions. □

Remark 11. As a direct consequence of Lemma 1, it is clear that if $\{u_i\}_{i=1}^n$ is a zero-mean stationary process with monotone autocovariance function $r_u(k) \to 0$ as $k \to \infty$, then, if $nh \to \infty$ as $n \to \infty$, we have that

$$C_1(nh)^{-1}\sum_{k=0}^{[nh]} r_u(k) \leqslant \mathrm{var}(\hat{g}_h(t)) \leqslant C_2(nh)^{-1}\sum_{k=0}^{[nh]} r_u(k),$$

where we have denoted $\hat{g}_h(t) = \sum_{i=1}^n w_{n,h}(t, t_i)(g(t_i) + u_i)$, with $g : [0, 1] \to \mathbb{R}$.

Lemma 2. *Assume that $\{u_i\}_{i=1}^n$ is a zero-mean stationary process with autocovariance function r_u, and $\{\theta_i\}$ is a sequence of real numbers. Then*

$$\mathrm{E}\left[\left(\sum_{i=1}^{n}\theta_i u_i\right)^2\right] \leqslant 2\left(\max_{i\in\{1,\dots,n\}}|\theta_i|\right)^2 n\sum_{k=0}^{n-1}|r_u(k)|.$$

Proof. The proof is trivial and therefore omitted. □

Lemma 3. *Let $\{u_i\}_{i=1}^n$ and $\{v_i\}_{i=1}^n$ be stationary processes with $\mathrm{E}(u_i) = \mathrm{E}(v_i) = 0$ and autocovariance functions r_u and r_v, respectively. Suppose that $\{u_i\}_{i=1}^n$ is independent of $\{v_i\}_{i=1}^n$. Let $\{a_{ij}\}_{ij}$ and k be real numbers such that $a_{ij} = 0$ if $|i-j| > k > 0$. Then*

$$\mathrm{E}\left(\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}u_i v_j\right)^2\right) \leqslant 4\left(\max_{i,j}|a_{ij}|\right)^2 n(2[k]+1)\sum_{i=0}^{n-1}|r_u(i)|\sum_{j=0}^{n-1}|r_v(j)|.$$

If $|r_v(j)|$ decreases, then $\sum_{j=0}^{n-1}|r_v(j)|$ above can be changed to $\sum_{j=0}^{2[k]}|r_v(j)|$.

Proof. As a direct consequence of the assumptions, we have that

$$\begin{aligned}\mathrm{E}\left(\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}u_i v_j\right)^2\right) &= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}r_u(i_1-i_2)\times\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}a_{i_1j_1}a_{i_2j_2}r_v(j_1-j_2)\\ &\leqslant \left(\max_{i,j}|a_{ij}|\right)^2\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}|r_u(i_1-i_2)|\times\sum_{j_1=i_1-[k]}^{i_1+[k]}\sum_{j_2=i_2-[k]}^{i_2+[k]}|r_v(j_1-j_2)|\\ &\leqslant \left(\max_{i,j}|a_{ij}|\right)^2 2n\sum_{i=0}^{n-1}|r_u(i)|\times 2(2[k]+1)\sum_{j=0}^{n-1}|r_v(j)|.\end{aligned}$$

□

Lemma 4. *Let $\{u_i\}_{i=1}^n$ be a stationary process such that $\mathrm{E}(|u_1|^k) < \infty$ $(k > 0)$. If $nha_n^{-k} \to 0$ as $n \to \infty$, then*

$$\sup_{t\in[0,1]}\left|\sum_{i=1}^{n}w_{n,h}(t,\, t_i)u_i\right| = o_p(a_n),$$

where $t_i = (i-0.5)/n$ and $w_{n,h}(\cdot,\cdot)$ are Gasser–Müller weights with bounded kernel function K with support $[-1, 1]$ (see (4)).

Proof. Let $\varepsilon > 0$. We have that

$$\begin{aligned}P\left(\sup_{t\in[0,1]}\left|\sum_{i=1}^{n}w_{n,h}(t,\, t_i)u_i\right| > \varepsilon a_n\right) &\leqslant C_3 nhP(C_4(nh)^{-1}|u_1| > \varepsilon C_5(nh)^{-1}a_n)\\ &\leqslant C_6 nh\mathrm{E}(|u_1|^k)a_n^{-k} \to 0,\end{aligned}$$

where C_i $(i = 3, 4, 5, 6)$ are positive constants. □

Lemma 5. *Under Assumptions* 1(b), 2(a), 3(b), 4 *and* 5, *we have that*

$$n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b \xrightarrow{P} \mathbf{V}_\eta.$$

Proof. The (i, j)th element of $n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b$ is

$$n^{-1}\tilde{\mathbf{x}}_{ib}^{\mathrm{T}}\tilde{\mathbf{x}}_{jb} = n^{-1}(\tilde{\mathbf{g}}_{ib}^{\mathrm{T}}\tilde{\mathbf{g}}_{jb} + \tilde{\mathbf{g}}_{ib}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_{jb} + \tilde{\mathbf{g}}_{jb}^{\mathrm{T}}\tilde{\boldsymbol{\eta}}_{ib} + \boldsymbol{\eta}_i^{\mathrm{T}}\boldsymbol{\eta}_j - \boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{W}_b\boldsymbol{\eta}_j - \boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{W}_b^{\mathrm{T}}\boldsymbol{\eta}_j + \boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{W}_b^{\mathrm{T}}\mathbf{W}_b\boldsymbol{\eta}_j). \tag{10}$$

Because $\tilde{\mathbf{g}}_i$ is a vector of biases, it satisfies $|\tilde{\mathbf{g}}_{ib}|^2 = O(nb^4 + n^{-1})$ (see Gasser and Müller 1984). Furthermore, from Assumption 2(a) we have that $|\boldsymbol{\eta}_i| = O_p(n^{1/2})$, and by Lemma 1 (with $f = r_{\eta_{ij}}$) $\mathrm{E}(|\mathbf{W}_b\boldsymbol{\eta}_j|^2) = O(b^{-1}S_{\eta_0,[nb]})$, and then $|\mathbf{W}_b\boldsymbol{\eta}_j| = O_p(b^{-1/2}S_{\eta_0,[nb]}^{1/2})$. From the last two results, and using the fact that $(nb)^{-1}S_{\eta_0,[nb]} \to 0$ (because of Assumption 1(b) and the fact that $nb \to \infty$), it is easy to obtain that $|\tilde{\boldsymbol{\eta}}_{ib}| = O_p(n^{1/2})$. These results, together with Assumption 2(a) and expression (10), give the result of the lemma. □

7. Proofs of main results

Proof of Theorem 1. (a) We have that

$$\mathrm{E}(\hat{\boldsymbol{\beta}}_b|\mathbf{X}) - \boldsymbol{\beta} = (n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b)^{-1}n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{m}}_b. \tag{11}$$

From Lemma 5, it is sufficient to study the p-vector $n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{m}}_b$. Its jth element is

$$n^{-1}\tilde{\mathbf{x}}_{jb}^{\mathrm{T}}\tilde{\mathbf{m}}_b = n^{-1}\left(\tilde{\mathbf{g}}_{jb}^{\mathrm{T}}\tilde{\mathbf{m}}_b + \boldsymbol{\eta}_j^{\mathrm{T}}\tilde{\mathbf{m}}_b - (\mathbf{W}_b\boldsymbol{\eta}_j)^{\mathrm{T}}\tilde{\mathbf{m}}_b\right). \tag{12}$$

Using results obtained in the proof of Lemma 5, it is easy to see that

$$n^{-1}\tilde{\mathbf{g}}_{jb}^{\mathrm{T}}\tilde{\mathbf{m}}_b = O(b^4 + n^{-2}), \qquad n^{-1}(\mathbf{W}_b\boldsymbol{\eta}_j)^{\mathrm{T}}\tilde{\mathbf{m}}_b = O_p\left((b^4 + n^{-2})^{1/2}((nb)^{-1}S_{\eta_0,[nb]})^{1/2}\right). \tag{13}$$

As for the second summand in (12), from Lemma 2 we obtain that

$$n^{-1}\boldsymbol{\eta}_j^{\mathrm{T}}\tilde{\mathbf{m}}_b = O_p\left((b^4 + n^{-2})^{1/2}(n^{-1}S_{\eta_0,n})^{1/2}\right). \tag{14}$$

Now, Lemma 5 and expressions (11)–(14), together with the fact that, for large enough n, $n^{-1}S_{\eta_0,n} \leqslant (nb)^{-1}S_{\eta_0,[nb]}$, give the result.

(b) We have that

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_b|\mathbf{X}) = n^{-1}(n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b)^{-1}n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}(\mathbf{I} - \mathbf{W}_b)\mathbf{V}_\varepsilon(\mathbf{I} - \mathbf{W}_b^{\mathrm{T}})\tilde{\mathbf{X}}_b(n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b)^{-1}. \tag{15}$$

First of all, we consider the asymptotic behaviour of the $(p \times p)$ matrix $n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{X}}_b$. Its (i, j)th element can be written as

$$n^{-1}\tilde{\mathbf{x}}_{ib}^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{x}}_{jb} = n^{-1}\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta}_j + n^{-1}(\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{g}}_{jb} - \boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{W}_b\boldsymbol{\eta}_j + \tilde{\mathbf{g}}_{ib}^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta}_j + \tilde{\mathbf{g}}_{ib}^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{g}}_{jb} - \tilde{\mathbf{g}}_{ib}^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{W}_b\boldsymbol{\eta}_j - (\mathbf{W}_b\boldsymbol{\eta}_i)^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta}_j - (\mathbf{W}_b\boldsymbol{\eta}_i)^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{g}}_{jb} + (\mathbf{W}_b\boldsymbol{\eta}_i)^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{W}_b\boldsymbol{\eta}_j). \tag{16}$$

Let us study the summands in (16). For this purpose, we will use the fact that

$$\|\mathbf{V}_\varepsilon\|_2 \leqslant (\|\mathbf{V}_\varepsilon\|_1\|\mathbf{V}_\varepsilon\|_\infty)^{1/2} = \|\mathbf{V}_\varepsilon\|_1 = O(S_{\varepsilon,n}). \tag{17}$$

From (17) and results obtained in the proof of Lemma 5, together with the Cauchy–Schwarz inequality, it is easy to obtain that

$$\begin{aligned}
&n^{-1}\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{g}}_{jb} = O_p((b^4+n^{-2})^{1/2}S_{\varepsilon,n}), \qquad n^{-1}\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{W}_b\boldsymbol{\eta}_j = O_p((nb)^{-1/2}S_{\eta_0,[nb]}^{1/2}S_{\varepsilon,n}),\\
&n^{-1}\tilde{\mathbf{g}}_{ib}^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{g}}_{jb} = O((b^4+n^{-2})S_{\varepsilon,n}), \quad n^{-1}(\mathbf{W}_b\boldsymbol{\eta}_i)^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta}_j = O_p((nb)^{-1/2}S_{\eta_0,[nb]}^{1/2}S_{\varepsilon,n}),\\
&n^{-1}\tilde{\mathbf{g}}_{ib}^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{W}_b\boldsymbol{\eta}_j = O_p((b^4+n^{-2})^{1/2}(nb)^{-1/2}S_{\eta_0,[nb]}^{1/2}S_{\varepsilon,n}),\\
&n^{-1}(\mathbf{W}_b\boldsymbol{\eta}_i)^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{g}}_{jb} = O_p((b^4+n^{-2})^{1/2}(nb)^{-1/2}S_{\eta_0,[nb]}^{1/2}S_{\varepsilon,n}),\\
&n^{-1}(\mathbf{W}_b\boldsymbol{\eta}_i)^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{W}_b\boldsymbol{\eta}_j = O_p((nb)^{-1}S_{\eta_0,[nb]}S_{\varepsilon,n}).
\end{aligned} \tag{18}$$

Now, from (16), (18) and the fact that $(b^2+n^{-1})S_{\varepsilon,n} \to 0$ and $((nb)^{-1}S_{\eta_0,[nb]})^{1/2}S_{\varepsilon,n} \to 0$, we obtain that

$$n^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\mathbf{V}_\varepsilon\tilde{\mathbf{X}}_b = n^{-1}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta} + o_p(1). \tag{19}$$

Using Lemma 5, (15) and (19), it is easy to see that

$$\operatorname{var}(\hat{\boldsymbol{\beta}}_b|\mathbf{X}) = (\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{V}_\varepsilon\boldsymbol{\eta}(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1} + o_p(n^{-1}) + n^{-1}\mathbf{V}_\eta^{-1}\mathbf{R}_n\mathbf{V}_\eta^{-1}, \tag{20}$$

where

$$\mathbf{R}_n = n^{-1}(-\tilde{\mathbf{X}}_b^{\mathrm{T}}\mathbf{W}_b\mathbf{V}_\varepsilon\tilde{\mathbf{X}}_b - \tilde{\mathbf{X}}_b^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{W}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b + \tilde{\mathbf{X}}_b^{\mathrm{T}}\mathbf{W}_b\mathbf{V}_\varepsilon\mathbf{W}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b).$$

By Assumption 2(b) and (20), the proof concludes if we demonstrate that $\mathbf{R}_n = o_p(1)$. Taking into account results obtained in the proof of Lemma 5, together with the fact that $\|\mathbf{W}_b\|_2 = \|\mathbf{W}_b^{\mathrm{T}}\|_2 = O(1)$, we obtain that

$$|\mathbf{W}_b^{\mathrm{T}}\tilde{\mathbf{x}}_{jb}| = |\mathbf{W}_b^{\mathrm{T}}(\boldsymbol{\eta}_j - \mathbf{W}_b\boldsymbol{\eta}_j + \tilde{\mathbf{g}}_{jb})| = O_p\Big(b^{-1/2}S_{\eta_0,[nb]}^{1/2} + (nb^4+n^{-1})^{1/2}\Big),$$

and therefore

$$\|\mathbf{W}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b\|_2 = \|\tilde{\mathbf{X}}_b^{\mathrm{T}}\mathbf{W}_b\|_2 = O_p\Big(b^{-1/2}S_{\eta_0,[nb]}^{1/2} + (nb^4+n^{-1})^{1/2}\Big).$$

Furthermore, Lemma 5 gives

$$\|\tilde{\mathbf{X}}_b\|_2 = \|\tilde{\mathbf{X}}_b^{\mathrm{T}}\|_2 = O_p\Big(n^{1/2}\Big).$$

Now, using these last two results, together with (17) and the fact that $(b^2+n^{-1})S_{\varepsilon,n}$ and $((nb)^{-1}S_{\eta_0,[nb]})^{1/2}S_{\varepsilon,n}$ converge to zero, it is easy to see that

$$\mathbf{R}_n = O_p\Big(((nb)^{-1}S_{\eta_0,[nb]})^{1/2}S_{\varepsilon,n} + (b^2+n^{-1})S_{\varepsilon,n}\Big) = o_p(1).$$

(c) From Theorem 1(a), we obtain that

$$n^{1/2}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) = n^{1/2}(\hat{\boldsymbol{\beta}}_b - \mathrm{E}(\hat{\boldsymbol{\beta}}_b|\mathbf{X})) + o_p(1).$$

Furthermore, using Lemma 5 and Assumption 2(a), it is easy to see that

$$n^{1/2}(\hat{\boldsymbol{\beta}}_b - \mathrm{E}(\hat{\boldsymbol{\beta}}_b|\mathbf{X})) = n^{1/2}(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\varepsilon} + O_p(|\mathbf{B}_n|_2),$$

where

$$\mathbf{B}_n = -n^{1/2}(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{W}_b\boldsymbol{\varepsilon} + n^{1/2}(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1}(\tilde{\mathbf{g}}_b - \mathbf{W}_b\boldsymbol{\eta})^{\mathrm{T}}(\boldsymbol{\varepsilon} - \mathbf{W}_b\boldsymbol{\varepsilon}).$$

Demonstrate that $\mathbf{B}_n = o_p(1)$ proves this part of the theorem (see Assumption 2(c)). By Lemma 5, to obtain $\mathbf{B}_n = o_p(1)$ it is sufficient to show that

$$-n^{-1/2}\boldsymbol{\eta}^{\mathrm{T}}\mathbf{W}_b\boldsymbol{\varepsilon} + n^{-1/2}(\tilde{\mathbf{g}}_b - \mathbf{W}_b\boldsymbol{\eta})^{\mathrm{T}}(\boldsymbol{\varepsilon} - \mathbf{W}_b\boldsymbol{\varepsilon}) = o_p(1). \tag{21}$$

Expression (21) can be written as $\mathbf{Q}_{n,b} + \mathbf{L}_{n,b} = o_p(1)$, where $\mathbf{Q}_{n,b} = \sum_{i=1}^n \sum_{j=1}^n \rho_{ij}\boldsymbol{\eta}_i^* \varepsilon_j$ and $\mathbf{L}_{n,b} = n^{-1/2}\tilde{\mathbf{G}}_b^{\mathrm{T}}\tilde{\boldsymbol{\varepsilon}}_b$, with $\rho_{ij} = n^{-1/2}(\sum_{k=1}^n w_{n,b}(t_k, t_i)w_{n,b}(t_k, t_j) - w_{n,b}(t_i, t_j) - w_{n,b}(t_j, t_i))$. We now bound $\mathbf{Q}_{n,b}$ and $\mathbf{L}_{n,b}$. Using the fact that we can write $\mathbf{L}_{n,b} = n^{-1/2}\sum_{i=1}^n \mathbf{c}_{i,b}\varepsilon_i$, $\mathbf{c}_{i,b}$ being non-random p-vectors such that $\mathbf{c}_{i,b}\mathbf{c}_{j,b}^{\mathrm{T}} = O(b^4 + n^{-2})$ if $i, j \in B_b$ or $\mathbf{c}_{i,b}\mathbf{c}_{j,b}^{\mathrm{T}} = o(b^4) + O(n^{-2})$ otherwise, where B_b is a set with $\#B_b = O(nb)$, it is easy to obtain that $\|\mathrm{var}(\mathbf{L}_{n,b})\|_2 = o(b^4 S_{\varepsilon,n}) + O(n^{-2}S_{\varepsilon,n})$. Therefore,

$$\mathbf{L}_{n,b} = o_p((b^4 S_{\varepsilon,n})^{1/2}) + O_p((n^{-2}S_{\varepsilon,n})^{1/2}). \tag{22}$$

As for $\mathbf{Q}_{n,b}$, applying Lemma 3 over each component and taking into account that $\max_{i,j}|\rho_{ij}| = O(n^{-3/2}b^{-1})$ and $\rho_{ij} = 0$ if $|i - j| > k = 3nb$ for large enough n (see Aneiros-Pérez and Quintela-del-Río 2002), it is easy to obtain that

$$\mathbf{Q}_{n,b} = O_p\Big(((nb)^{-1}S_{\eta_0,[nb]}S_{\varepsilon,n})^{1/2}\Big). \tag{23}$$

Taking into account that $(b^4 + n^{-2})S_{\varepsilon,n} \to 0$ and $(nb)^{-1}S_{\eta_0,[nb]}S_{\varepsilon,n} \to 0$, (22) and (23) give (21). This concludes the proof. □

Proof of Theorem 2. (a) We have that

$$\mathrm{E}(\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)|\mathbf{X}) - m(t) = \mathrm{bias}(\hat{m}_h(t, \boldsymbol{\beta})) - \mathbf{w}_h^{\mathrm{T}}(t)\mathbf{X}\,\mathrm{bias}(\hat{\boldsymbol{\beta}}_b|\mathbf{X}), \tag{24}$$

where $\mathbf{w}_h^{\mathrm{T}}(t) = (w_{n,h}(t, t_1), \ldots, w_{n,h}(t, t_n))$. Taking into account relation (2), Lemma 1 and the fact that the bias of the nonparametric estimator tends to zero, we obtain that

$$\mathbf{w}_h^{\mathrm{T}}(t)\mathbf{X} \xrightarrow{P} (g_1(t), \ldots, g_p(t)). \tag{25}$$

Now, from (25), Theorem 1(a) and the fact that

$$\mathrm{bias}(\hat{m}_h(t, \boldsymbol{\beta})) = 0.5h^2 m''(t)\int u^2 K(u)\mathrm{d}u + o(h^2)$$

(we suppose that $nh^2 \to \infty$), together with $b^2 = o(h)$ and $h^{-4}b^3n^{-1}S_{\eta_0,[nb]} \to 0$, we obtain that

$$\left| \frac{\mathbf{w}_h^{\mathrm{T}}(t)\mathbf{X}\,\mathrm{bias}(\hat{\boldsymbol{\beta}}_b|\mathbf{X})}{\mathrm{bias}(\hat{m}_h(t, \boldsymbol{\beta}))} \right| = o_p(1). \tag{26}$$

Now (24) and (26) give the result.

(b) We have that

$$\begin{aligned}\mathrm{var}(\hat{m}_h(t, \hat{\boldsymbol{\beta}}_b)|\mathbf{X}) = \mathrm{var}(\hat{m}_h(t, \boldsymbol{\beta})) + \mathbf{w}_h^{\mathrm{T}}(t)\mathbf{X}\,\mathrm{var}(\hat{\boldsymbol{\beta}}_b|\mathbf{X})\mathbf{X}^{\mathrm{T}}\mathbf{w}_h(t) \\ - 2\mathbf{w}_h^{\mathrm{T}}(t)\mathbf{X}\,\mathrm{cov}(\hat{\boldsymbol{\beta}}_b,\, \hat{m}_h(t, \boldsymbol{\beta})|\mathbf{X}).\end{aligned} \tag{27}$$

From Theorem 1(b) and Lemma 1 (with $f = r_\varepsilon$) we obtain that

$$\frac{\|\mathrm{var}(\hat{\boldsymbol{\beta}}_b|\mathbf{X})\|_2}{\mathrm{var}(\hat{m}_h(t, \boldsymbol{\beta}))} = O_p((h^{-1}S_{\varepsilon,[nh]})^{-1}) = o_p(1) \tag{28}$$

and, together with (25),

$$\frac{\|\mathbf{w}_h^{\mathrm{T}}(t)\mathbf{X}\,\mathrm{var}(\hat{\boldsymbol{\beta}}_b|\mathbf{X})\mathbf{X}^{\mathrm{T}}\mathbf{w}_h(t)\|_2}{\mathrm{var}(\hat{m}_h(t, \boldsymbol{\beta}))} = o_p(1). \tag{29}$$

Now, using (28) together with (25), it is easy to see that, for $k = 1, \ldots, p$,

$$\frac{|\mathbf{w}_h^{\mathrm{T}}(t)\mathbf{X}\,\mathrm{cov}(\hat{\boldsymbol{\beta}}_{b,k},\, \hat{m}_h(t, \boldsymbol{\beta})|\mathbf{X})|}{\mathrm{var}(\hat{m}_h(t, \boldsymbol{\beta}))} \leqslant C\left(\frac{\mathrm{var}(\hat{\boldsymbol{\beta}}_{b,k}|\mathbf{X})}{\mathrm{var}(\hat{m}_h(t, \boldsymbol{\beta}))}\right)^{1/2} = o_p(1), \tag{30}$$

where $\hat{\boldsymbol{\beta}}_{b,k}$ denotes the kth component of $\widehat{\boldsymbol{\beta}}_b$ and C is a positive constant. Now (27), (29) and (30) give the result.

(c) We have that $\hat{m}_h(u_j, \hat{\boldsymbol{\beta}}_b) = \hat{m}_h(u_j, \boldsymbol{\beta}) - \mathbf{w}_h^{\mathrm{T}}(u_j)\mathbf{X}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})$. Therefore, using (25) and the $n^{1/2}$-consistency of $\hat{\boldsymbol{\beta}}_b$, it is easy to see that $\hat{m}_h(u_j, \hat{\boldsymbol{\beta}}_b) = \hat{m}_h(u_j, \boldsymbol{\beta}) + O_p(n^{-1/2})$, $j = 1, \ldots, k$. Under Assumption 7 we can apply Csörgő and Mielniczuk's (1995) Theorem 1 to obtain that

$$C_{n,h,\alpha_\varepsilon}(\hat{m}_h(u_1, \boldsymbol{\beta}) - m(u_1), \ldots, \hat{m}_h(u_k, \boldsymbol{\beta}) - m(u_k)) \xrightarrow{d} \sigma_{\alpha_\varepsilon,g,K}(N_1, \ldots, N_k),$$

while under Assumption 8 we can apply Deo's (1997) Theorem 3 to obtain the same result. Now, taking into account that, from Assumption 7 (or 8), $C_{n,h,\alpha_\varepsilon} = o_p(n^{1/2})$, it is clear that the result holds. □

Proof of Theorem 3. (a) From Theorem 1(c), we obtain that

$$n(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{A}^{-1}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) \xrightarrow{d} \chi_p^2. \tag{31}$$

Furthermore, Assumptions 2(a), 2(c) and 10 give $\mathbf{A} = \sigma_\varepsilon^2\mathbf{V}_\eta^{-1}$. Now, using this result, Lemma 5 and the $n^{1/2}$-consistency of $\hat{\boldsymbol{\beta}}_b$, it easy to obtain that

$$|n(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{A}^{-1}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) - \sigma_\varepsilon^{-2}n(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^{\mathrm{T}}(n^{-1}\tilde{\mathbf{X}}_h^{\mathrm{T}}\tilde{\mathbf{X}}_h)(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})| = o_p(1).$$

This last result, together with (31), gives

$$\frac{(\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta})^{\mathrm{T}}(\tilde{\mathbf{X}}_h^{\mathrm{T}}\tilde{\mathbf{X}}_h)(\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta})}{\sigma_\varepsilon^2}\xrightarrow{d}\chi_p^2. \tag{32}$$

(b) Because the difference between $\hat{\boldsymbol{\beta}}_b$ and the true value of the vector of parameters $\boldsymbol{\beta}$ does not depend on $\boldsymbol{\beta}$, from Theorem 1(c) we have that, under $H_{1,\boldsymbol{\beta}}^{\mathbf{c},n}$, $n^{1/2}(\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta}_0)\xrightarrow{d}N(\mathbf{c},\mathbf{A})$. Therefore, arguments similar to those of Theorem 3(a) can be used to obtain the result.

(c) We will use the decomposition

$$F(b,h)=\frac{(\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta})^{\mathrm{T}}(\tilde{\mathbf{X}}_h^{\mathrm{T}}\tilde{\mathbf{X}}_h)(\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta})}{\sigma_\varepsilon^2}$$
$$+\,n\left(2\frac{(\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta}_0)^{\mathrm{T}}(n^{-1}\tilde{\mathbf{X}}_h^{\mathrm{T}}\tilde{\mathbf{X}}_h)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{\sigma_\varepsilon^2}-\frac{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}(n^{-1}\tilde{\mathbf{X}}_h^{\mathrm{T}}\tilde{\mathbf{X}}_h)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{\sigma_\varepsilon^2}\right). \tag{33}$$

From the consistency of $\hat{\boldsymbol{\beta}}_b$ and Lemma 5, we obtain that under $H_{1,\boldsymbol{\beta}}$ the difference in large parentheses in (33) tends to $\sigma_\varepsilon^{-2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{V}_\eta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)>0$ ($\mathbf{V}_\eta$ is a positive definite matrix) as $n\to\infty$. This result, together with (32), concludes the proof. □

Proof of Theorem 4. We first give some general results which will be used throughout the proof of the theorem. When these results depend on c_n, they must be interpreted as valid under $H_{1,m}^{c_n}$, with c_n verifying the conditions of the theorem (part (b)) or $c_n=0$. Observe that in this last case, $H_{1,m}^{c_n}=H_{0,m}$.

We have that $\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta}=(\hat{\boldsymbol{\beta}}_b^{(0)}-\boldsymbol{\beta})+\mathbf{B}_b$, where $\mathbf{B}_b=c_n(\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{X}}_b)^{-1}\tilde{\mathbf{X}}_b^{\mathrm{T}}\tilde{\mathbf{m}}_b^*$ and $\hat{\boldsymbol{\beta}}_b^{(0)}$ denotes the estimator (3) in the regression model $y_i=\boldsymbol{\zeta}_i^{\mathrm{T}}\boldsymbol{\beta}+m_0(t_i)+\varepsilon_i$. Furthermore, as we can see in the proof of Theorem 1(a) above, it can be verified that $\mathbf{B}_b=O_p(c_nd_n)$, where $d_n=b^4+n^{-2}+((b^4+n^{-2})(nb)^{-1}S_{\eta_0,[nb]})^{1/2}$. Now from this result, together with Theorem 1(a,b) and our assumptions on b and $S_{\eta_0,[nb]}$, it is easy to obtain that

$$|\hat{\boldsymbol{\beta}}_b-\boldsymbol{\beta}|=O_p(n^{-1/2}+c_nd_n). \tag{34}$$

Furthermore, by means of Lemma 4 and our conditions on the weights and g_j, we have that

$$\sup_{t\in[0,1]}\left|\sum_{i=1}^n w_{n,h}(t,t_i)\boldsymbol{\zeta}_i^{\mathrm{T}}\right|=o_p(h^{-1/4}),$$

and then, together with (34), we have that

$$\sup_{1\leqslant i\leqslant n}|a_i|=o_p(h^{-1/4}(n^{-1/2}+c_nd_n)), \tag{35}$$

where we have denoted

$$a_i=\left(\sum_{j=1}^n w_{n,h}(t_i,t_j)\boldsymbol{\zeta}_j^{\mathrm{T}}\right)(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_b). \tag{36}$$

A result that we will also use is that

$$B_n = O_p(n^{-1/2}), \tag{37}$$

where

$$B_n \equiv n^{-1} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} w_{n,h}(t_i,\, t_j)\varepsilon_j \right) \left(\sum_{k=1}^{n} w_{n,h}(t_i,\, t_k)\zeta_k^{\mathrm{T}} \right). \tag{38}$$

To show (37), it is sufficient to obtain an order $O_p(n^{-1/2})$ for each of the p components $B_{n,s}$ ($s = 1, \ldots, p$). We have that

$$B_{n,s} \equiv n^{-1} \sum_{j=1}^{n} c_{j,s}\varepsilon_j + n^{-1} \sum_{j=1}^{n} \sum_{k=1}^{n} d_{j,k,s}\varepsilon_j \eta_{ks} \equiv B_{n,s}^{(1)} + B_{n,s}^{(2)},$$

where

$$c_{j,s} = \sum_{i=1}^{n} w_{n,h}(t_i,\, t_j) \left(\sum_{k=1}^{n} w_{n,h}(t_i,\, t_k) g_s(t_k) \right), \qquad d_{j,k,s} = \sum_{i=1}^{n} w_{n,h}(t_i,\, t_j) w_{n,h}(t_i,\, t_k).$$

Using our conditions on the weights and g_j, we obtain that

$$\sup_{1\leqslant j\leqslant n} |c_{j,s}| = O(1), \qquad \sup_{1\leqslant j,k\leqslant n} |d_{j,k,s}| = O((nh)^{-1}), \qquad d_{j,k,s} = 0 \quad \text{if } |j-k| > 2nh+1. \tag{39}$$

From Lemma 2 above, and taking into account (39), we obtain that

$$\mathrm{E}((B_{n,s}^{(1)})^2) = n^{-2}\mathrm{E}\left(\left(\sum_{j=1}^{n} c_{j,s}\varepsilon_j \right)^2 \right) = O(n^{-1}).$$

Therefore, $B_{n,s}^{(1)} = O_p(n^{-1/2})$. As for $B_{n,s}^{(2)}$, from Lemma 3 above, and using (39) and the facts that $S_{\varepsilon,n} = O(1)$ and $(nh)^{-1} S_{\eta_0,[nh]} \to 0$, we obtain that

$$\mathrm{E}((B_{n,s}^{(2)})^2) = n^{-2}\mathrm{E}\left(\sum_{j=1}^{n} \sum_{k=1}^{n} d_{j,k,s}\varepsilon_j \eta_{ks} \right)^2 = O((n^2 h)^{-1} S_{\eta_0,[nh]} S_{\varepsilon,n}) = o(n^{-1}),$$

and then $B_{n,s}^{(2)} = o_p(n^{-1/2})$. Now it is clear that (37) follows from the results above.

A known property (Gasser and Müller 1984), which we will use below, is that, if f has two continuous derivatives on [0, 1], then

$$\sup_{t\in[0,1]} \left| \sum_{i=1}^{n} w_{n,h}(t,\, t_i) f(t_i) - f(t) \right| = O(h^2 + n^{-1}). \tag{40}$$

We now decompose $d_m^2(\hat{r}_n,\, H_{0,m})$ into five summands, which will facilitate the asymptotic study of this statistic. It is easy to see that

$$d_m^2(\hat{r}_n,\ H_{0,m}) = \sum_{i=1}^{5} A_{i,n}, \tag{41}$$

where

$$A_{1,n} = n^{-1}\sum_{i=1}^{n}(\hat{m}_h(t_i, \boldsymbol{\beta}) - m(t_i))^2, \qquad A_{2,n} = n^{-1}\sum_{i=1}^{n}(m_0(t_i)^2 - m(t_i)^2),$$

$$A_{3,n} = 2n^{-1}\sum_{i=1}^{n}\hat{m}_h(t_i, \boldsymbol{\beta})(m(t_i) - m_0(t_i)), \qquad A_{4,n} = n^{-1}\sum_{i=1}^{n} a_i^2,$$

$$A_{5,n} = 2n^{-1}\sum_{i=1}^{n}(\hat{m}_h(t_i, \boldsymbol{\beta}) - m_0(t_i))a_i.$$

We will show that $A_{1,n}$ properly standardized is asymptotically normal, while $A_{2,n} + A_{3,n}$ gives the mean or the convergence to ∞ of $(n^2h)^{1/2}(d_m^2(\hat{r}_n,\ H_{0,m}) - (nh)^{-1}\sum_{s=-\infty}^{\infty} r_\varepsilon(s)$ $\int K^2)$, depending on the conditions on c_n. The rest of the terms are asymptotically negligible.

(a) Under $H_{0,m}$, it is clear that $A_{2,n} = A_{3,n} = 0$. Using (35) we have that $A_{4,n} = o_p((n^2h)^{-1/2})$. Furthermore,

$$A_{5,n} = 2B_n(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_b) + 2n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{n,h}(t_i,\ t_j)m_0(t_j) - m_0(t_i)\right)a_i \equiv A_{5,n}^{(1)} + A_{5,n}^{(2)}.$$

From (34) and (37) we obtain that $A_{5,n}^{(1)} = O_p(n^{-1})$, and by means of (35) and (40) we have that $A_{5,n}^{(2)} = o_p((n^{-1/2}h^{-1/4})(h^2 + n^{-1})) = o_p((n^2h)^{-1/2})$ (as a consequence of Assumption 13, $n^2h^9 \to 0$). Therefore, $A_{5,n} = o_p((n^2h)^{-1/2})$. The term $A_{1,n}$ can be treated essentially in the same way as the term Δ_1 in González-Manteiga and Vilar-Fernández (1995), taking into account the correction according to Biedermann and Dette (2000). Therefore,

$$\sqrt{n^2h}\left(A_{1,n} - \frac{\sum_{s=-\infty}^{\infty} r_\varepsilon(s)\int K^2}{nh}\right) \xrightarrow{d} N(0,\ \sigma_d^2).$$

These results give this part of the theorem.

(b) Under $H_{1,m}^{c_n}$, from Assumption 4 and the Hölder continuity of m_0 and m^*, it is easy to obtain that

$$A_{2,n} = -2c_n\int m_0(u)m^*(u)\mathrm{d}u - c_n^2\int m^*(u)^2\,\mathrm{d}u + O(n^{-1}(c_n + c_n^2)). \tag{42}$$

$A_{3,n}$ can be broken down as

$$A_{3,n} = 2n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{n,h}(t_i,\, t_j)m(t_j)\right)(m(t_i) - m_0(t_i))$$

$$+ 2n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{n,h}(t_i,\, t_j)\varepsilon_j\right)(m(t_i) - m_0(t_i)) \equiv A_{3,n}^{(1)} + A_{3,n}^{(2)}. \quad (43)$$

It follows from (40), $H_{1,m}^{c_n}$, Assumptions 4 and 13, and the Hölder continuity of m_0 and m^* that

$$A_{3,n}^{(1)} = 2c_n\int m_0(u)m^*(u)\mathrm{d}u + 2c_n^2\int m^*(u)^2\mathrm{d}u + O((h^2 + n^{-1})(c_n + c_n^2)). \quad (44)$$

Furthermore, considering $A_{3,n}^{(2)} = \sum_{j=1}^{n}\theta_j\varepsilon_j$, where $\theta_j = 2n^{-1}\sum_{i=1}^{n} w_{n,h}(t_i,\, t_j)(m(t_i) - m_0(t_i))$, and using Lemma 2 above, we obtain that, under $H_{1,m}^{c_n}$, $\mathrm{E}((A_{3,n}^{(2)})^2) = O(n^{-1}c_n^2)$ (we have used the fact that, under $H_{1,m}^{c_n}$, $\sup_j|\theta_j| = O(n^{-1}c_n)$). Therefore,

$$A_{3,n}^{(2)} = O_p((n^{-1}c_n^2)^{1/2}). \quad (45)$$

Using (35), we have that

$$A_{4,n} = o_p\Big((h^{-1/4}(n^{-1/2} + c_n d_n))^2\Big). \quad (46)$$

We can break $A_{5,n}$ down as

$$A_{5,n} = 2B_n(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_b) + 2n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{n,h}(t_i,\, t_j)m_0(t_j) - m_0(t_i)\right)a_i$$

$$+ 2n^{-1}c_n\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{n,h}(t_i,\, t_j)m^*(t_j)\right)a_i$$

(see (36) and (38) for expressions for a_i and B_n, respectively). Therefore, using (34), (35), (37) and (40), together with Assumption 13, we have that, under $H_{1,m}^{c_n}$,

$$A_{5,n} = O_p(n^{-1} + n^{-1/2}c_n d_n) + o_p(h^{-1/4}(n^{-1/2} + c_n d_n)(h^2 + n^{-1})) + o_p(h^{-1/4}c_n(n^{-1/2} + c_n d_n)). \quad (47)$$

As for $A_{1,n}$, we have that

$$A_{1,n} = n^{-1}\sum_{i=1}^{n}(\hat{m}_{0h}(t_i) - m_0(t_i))^2 + c_n^2 n^{-1}\sum_{i=1}^{n}\mathrm{bias}(\hat{m}_{*h}(t_i))^2$$

$$+ 2c_n n^{-1}\sum_{i=1}^{n}(\hat{m}_{0h}(t_i) - m_0(t_i))\mathrm{bias}(\hat{m}_{*h}(t_i)) \equiv A_{1,n}^{(1)} + A_{1,n}^{(2)} + A_{1,n}^{(3)}, \quad (48)$$

where

$$\hat{m}_{0h}(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i)(m_0(t_i) + \varepsilon_i), \qquad \hat{m}_{*h}(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i)(m^*(t_i) + \varepsilon_i).$$

Furthermore, as in (a), we have that

$$(n^2 h)^{1/2} \left(A_{1,n}^{(1)} - \frac{\sum_{s=-\infty}^{\infty} r_\varepsilon(s) \int K^2}{nh} \right) \xrightarrow{d} N(0, \sigma_d^2). \tag{49}$$

From (40), it is easy to obtain that

$$A_{1,n}^{(2)} = O(c_n^2(h^4 + n^{-2})). \tag{50}$$

By the Cauchy–Schwarz inequality, together with (49) and (50), we have that

$$A_{1,n}^{(3)} = O_p((nh)^{-1/2} c_n (h^2 + n^{-1})). \tag{51}$$

Now (41)–(51), together with our assumptions on c_n, give the result. □

Acknowledgements

This research was supported in part by DGESIC, Xunta de Galicia and MCyT (European FEDER support included) grants PB98-0182-C02-01, PB98-0182-C02-02, PGIDT01P-XI10504PR, PGIDT00PXI20704PN, BFM2002-00265 and BFM2002-03213. P. Vieu wishes to thank all the participants of the working group STAPH on Functional Statistics in Toulouse for their numerous and interesting comments. We would like to thank the Climate Research Unit of the University of East Anglia for providing us with the data, via the website www.cru.uea.ac.uk. We are also grateful to the associate editor and a referee for their constructive remarks, which helped to improve the quality of the paper.

References

Aneiros-Pérez, G. (2001) Estimación de modelos parcialmente lineales con errores dependientes. Doctoral thesis, Department of Statistics and Operations Research, University of Santiago de Compostela, Spain.

Aneiros-Pérez, G. and Quintela-del-Río, A. (2001) Asymptotic properties in partial linear models under dependence. *Test*, **10**, 333–355.

Aneiros-Pérez, G. and Quintela-del-Río, A. (2002) Plug-in bandwidth choice in partial linear models with autoregressive errors. *J. Statist. Plann. Inference*, **100**, 23–48.

Beran, J. (1992) Statistical methods for data with long-range dependence. *Statist. Sci.*, **7**, 404–427.

Beran, J. (1994) *Statistics for Long-Memory Processes*. New York: Chapman & Hall.

Beran, J. and Feng, Y. (2002) Local polynomial fitting with long-memory, short-memory and antipersistent errors. *Ann. Inst. Statist. Math.*, **54**, 291–311.

Beran, J. and Ghosh, S. (1998) Root-*n*-consistent estimation in partial linear models with long-memory errors. *Scand. J. Statist.*, **25**, 345–357.

Biedermann, S. and Dette, H. (2000) Testing linearity of regression models with dependent errors by kernel based methods. *Test*, **9**, 417–438.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Lecture Notes in Statist., 110. New York: Springer-Verlag.

Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Csörgő, S. and Mielniczuk, J. (1995) Nonparametric regression under long-range dependent normal errors. *Ann. Statist.*, **23**, 1000–1014.

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

Deo, R.S. (1997) Nonparametric regression with long-memory errors. *Statist. Probab. Lett.*, **33**, 89–94.

Engle, R., Granger, C., Rice, J. and Weiss, A. (1986) Nonparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.*, **81**, 310–320.

Fox, R. and Taqqu, M.S. (1987) Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields*, **74**, 213–240.

Gao, J.T. (1995) Asymptotic theory for partly linear models. *Comm. Statist. Theory Methods*, **24**, 1985–2009.

Gao, J.T. (1997) Adaptive parametric test in a semiparametric regression model. *Comm. Statist. Theory Methods*, **26**, 787–800.

Gao, J.T. and Anh, V.V. (1999) Semiparametric regression under long-range dependent errors. *J. Statist. Plann. Inference*, **80**, 37–57.

Gasser, T. and Müller, H.G. (1984) Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171–185.

Gasser, T., Müller, H.G. and Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B*, **47**, 238–252.

González-Manteiga, W. and Aneiros-Pérez, G. (2003) Testing in partial linear regression models with dependent errors. *J. Nonparametr. Statist*., **15**, 93–111.

González-Manteiga, W. and Vilar-Fernández, J. M. (1995) Testing linear regression models using non-parametric regression estimators when errors are non-independent. *Comput. Statist. Data Anal.*, **20**, 521–541.

Granger, C.W.J. and Joyeux, R. (1980) An introduction to long-range time series models and fractional differencing. *J. Time Ser. Anal.*, **1**, 15–30.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989) *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statist. 60. Berlin: Springer-Verlag.

Hall, P. and Hart, J.D. (1990) Nonparametric regression with long-range dependence. *Stochastic Process. Appl.*, **36**, 339–351.

Hall, P., Lahiri, S.N. and Polzehl, J. (1995) On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.*, **23**, 1921–1936.

Härdle, W., Liang, H. and Gao, J.T. (2000) *Partially Linear Models*. Heidelberg: Physica-Verlag.

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

Heyde, C.C. and Yang, Y. (1997) On defining long-range dependence. *J. Appl. Probab.*, **34**, 939–944.

Ho, H. (1996) On central and non-central limit theorems in density estimation for sequences of long-range dependence. *Stochastic Process. Appl.*, **63**, 153–174.

Jones, M.C., Marron, J.S. and Sheather, S.J. (1996) A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.*, **91**, 401–407.

Künsch, H., Beran, J. and Hampel, F. (1993) Contrast under long-range correlations. *Ann. Statist.*, **21**, 943–964.

Linton, O. (1995) Second order approximation in the partially linear regression model. *Econometrica*, **63**, 1079–1112.

Mandelbrot, B.B. and van Ness, J.W. (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, **10**, 422–437.

Nieuwenhuis, G. (1992) Central limit theorems for sequences with $m(n)$-dependent main part. *J. Statist. Plann. Inference*, **32**, 229–241.

Robinson, P. (1988) Root-n-consistent semiparametric regression. *Econometrica*, **56**, 931–954.

Speckman, P. (1988) Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B*, **50**, 413–436.

Stone, C. (1977) Consistent nonparametric regression. *Ann. Statist.* **5**, 595–645.

Vilar-Fernández, J.M. and González-Manteiga, W. (2000) Resampling for checking linear regression models via non-parametric regression estimation. *Comput. Statist. Data Anal.*, **35**, 211–231.

Yajima, Y. (1988) On estimation of a regression model with long-memory stationary errors. *Ann. Statist.*, **16**, 791–807.

Yajima, Y. (1991) Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.*, **19**, 158–177.

Received April 2002 and revised August 2003