Sociedad de Estadística e Investigación Operativa
Test (2001) *Vol.* **10**, *No.* 2, pp. 333–355

Asymptotic properties in partial linear models under dependence

Germán Aneiros* and Alejandro Quintela
Departamento de Matemáticas, Facultad de Informática
Universidad de La Coruña, Spain

Abstract

Consider the regression model $y_i = \boldsymbol{\zeta}_i^T \boldsymbol{\beta} + m(t_i) + \varepsilon_i$ for $i = 1, \ldots, n$. Here $(\boldsymbol{\zeta}_i^T, t_i)^T \in \mathbb{R}^p \times [0,1]$ are design points, $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of parameters, m is an unknown smooth function from $[0,1]$ to $\mathbb{R}$ and ε_i are the unobserved errors. We will assume that these errors are not independent. Under suitable assumptions, we obtain expansions for the bias and the variance of a Generalized Least Squares (GLS) type regression estimator, and for an estimator of the nonparametric function $m(\cdot)$. Furthermore, we prove the asymptotic normality of the first estimator. The obtained results are a generalization of those contained in Speckman (1988), who studied a similar model with i.i.d. error variables.

Key Words: Bandwidth selection, kernel smoothing, mixing, partial linear models.
AMS subject classification: 62G05, 62G20, 62M10.

1 Introduction

Partial linear models are more than a modest generalization of a multivariate linear model. These models are used when a response variable can be presumed to be related linearly to one or more variables, and in a non-linear way to one or more different variables. The specific form of the model that we will consider in this paper is

$$y_i = \boldsymbol{\zeta}_i^T \boldsymbol{\beta} + m(t_i) + \varepsilon_i \qquad (i = 1, \ldots, n), \tag{1.1}$$

*Correspondence to: Germán Aneiros, Departamento de Matemáticas, Facultad de Informática, Universidad de La Coruña, Campus de Elviña, s/n, La Coruña, Spain. Email: ganeiros@udc.es

Research supported by the Xunta de Galicia (Spain) and the DGES (Spain) under research projects XUGA 10503A98 and PB98-0182-C02-01, respectively.

Received: May 1999; Accepted: June 2001

where the $\boldsymbol{\zeta}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ and $t_i \in [0,1]$ are design points, $\boldsymbol{\beta}$ $(p \times 1)$ is a vector of unknown parameters, m is a smooth unknown real-valued function defined on $[0,1]$ and the ε_i are a sample of errors identically distributed. Assumptions relating to $\boldsymbol{\zeta}_i$ and t_i will be introduced in Section 2.

Using the notation established in that section and in (1.3) below, if we assume that assumptions (A.1.a), (A.5.b)-(A.7) and (A.9)-(A.12) hold, then $\boldsymbol{\beta}$ can be identified via (see Lemma 3.1 and property (3.12) below)

$$\boldsymbol{\beta} = \boldsymbol{V}^{-1}(\lim_{n\to\infty} n^{-1}\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1} E(\widetilde{\boldsymbol{y}})).$$

The identifiability of $m(\cdot)$ is obtained from the identifiability of $\boldsymbol{\beta}$, together with assumption (A.9) and the continuity of $m(\cdot)$.

Model (1.1) is much more flexible than the standard linear model since it combines both parametric and nonparametric components. It can be used to examine the effect of price changes on the volume of sales. The conventional assumption is that the logarithm of the sales volume is linearly related to price. However, it is natural to expect that weekly and seasonal effects would also be at work, therefore Daniel and Wood (1980) also included dummy variables to indicate the day of the week and the month in which each observation lay. In this situation, an attractive alternative is to model the dependence on time in a nonparametric fashion, where the variable t represents the day of the year.

There are some interesting papers on the estimation of the vector $\boldsymbol{\beta}$ and the function m. One possible method to estimate $\boldsymbol{\beta}$ and m would be by means of a penalized least squares criterion, by minimizing

$$\sum_{i=1}^{n}(y_i - \boldsymbol{\zeta}_i^T\boldsymbol{\beta} - m(t_i))^2 + \lambda \int (m''(t))^2 dt.$$

In this way, we have a spline type estimation, studied by Engle et al. (1986), Denby (1986), Heckman (1986) and Rice (1986), among others. Another method to estimate $\boldsymbol{\beta}$ and m would be by means of estimators based on least squares estimation and kernel type estimation. Thus, if we consider model (1.1) without the linear component,

$$y_i = m(t_i) + \varepsilon_i,$$

a kernel type estimator can be written as

$$m_{n,h}(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i) y_i, \tag{1.2}$$

with $w_{n,h}(\cdot, t_i)$ a weight function (derived from a function $K(\cdot)$, the kernel) that can take different forms, thus providing different estimators: Nadaraya (1964) and Watson (1964), Priestley-Chao (1972), local polynomial estimators (Stone (1977)) or Gasser-Müller (1979). See, for example, the monographs of Härdle (1993), Simonoff (1996) or Fan and Gijbels (1996) for some theoretical results and practical examples in the nonparametric estimation field. It is well known that in all these estimators the selection of an adequate parameter h —the *smoothing parameter* or *bandwidth*— is essential for good behavior of the estimator when we fit a curve to a set of given data. See, for example, Quintela (1996) for a review of smoothing parameter selection methods, and a comparison of the same under dependence assumptions for the errors.

One method developed by Speckman (1988) for the estimation of the vector $\boldsymbol{\beta}$ in (1.1) is based on least squares estimation and kernel type estimation, by means of a regression over the partial residuals of the model of the form

$$\begin{aligned} \widetilde{\boldsymbol{y}} &= (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{y}, \\ \widetilde{\boldsymbol{X}} &= (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{X}, \end{aligned} \tag{1.3}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^T$, $\boldsymbol{X} = \{x_{ij}\}_{i=1,\ldots,n\ j=1,\ldots,p}$ and $\boldsymbol{W}$ is a smoothing matrix with elements $\{w_{ij}\} = \{w_{n,h}(t_i, t_j)\}$. In this way, $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{y}}$ are the matrix $\boldsymbol{X}$ and the vector $\boldsymbol{y}$ after adjustment for dependence on t. Assuming that $\widetilde{\boldsymbol{X}}$ has full rank, he obtains

$$\widehat{\boldsymbol{\beta}}_p = \left(\widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}}\right)^{-1} \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{y}}, \tag{1.4}$$

by minimizing the weighted least squares criterion

$$\|(\boldsymbol{I} - \boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\|_2^2,$$

where $\|\|_2$ denotes the Euclidean norm $\|\boldsymbol{v}\|_2^2 = \sum v_i^2$ for $\boldsymbol{v} = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$, and he uses a kernel type estimator for the function m

$$\widehat{m}_h(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i)(y_i - \boldsymbol{\zeta}_i^T \widehat{\boldsymbol{\beta}}_p). \tag{1.5}$$

Speckman (1988) compared estimator (1.4) with a different estimator of β suggested by Green et al. (1985), with the following expression:

$$\widehat{\beta}_{GJS} = \left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{y}.$$

Depending on the assumptions, the variance of the latter estimator can be dominated by its bias, something that does not happen with $\widehat{\beta}_p$ in (1.4). In his paper, Speckman (1988) supposes that the errors of the model (1.1) are independent. However, the presence of correlation between the errors is something that can often happen in practice (for example, when the observations are recorded through time). In this case, the variance-covariance matrix of the errors has the form $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma_\varepsilon^2\boldsymbol{\Psi}$, where $\sigma_\varepsilon^2 = Var(\varepsilon_i)$ and $\boldsymbol{\Psi}$ is different from the identity matrix.

In this paper, we will study asymptotic properties of a Generalized Least Squares (GLS) estimator of the vector β, assuming a dependence structure in the model's errors, i.e. when the sample of $\{\varepsilon_i\}$ is a time series. In Section 2 we describe the assumptions for model (1.1), and the precise form of the estimates of β and m. Next, we obtain the asymptotic results, that prove that the dependence effect between the errors affects, in a very small way, the convergence rate of the estimators. Section 3 is devoted to sketches of proofs.

2 The model estimation

Let us consider model (1.1). We assume that $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, $Var(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma_\varepsilon^2\boldsymbol{\Psi}$, $\boldsymbol{\Psi} \neq \boldsymbol{I}$ and positive definite. Since $\boldsymbol{\Psi}$ is positive definite, there exists a $n \times n$ matrix $\boldsymbol{P}$ such that $\boldsymbol{P}\boldsymbol{\Psi}\boldsymbol{P}^T = \boldsymbol{I}$ hence $\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{\Psi}^{-1}$ and $\boldsymbol{P}$ is not unique. We choose $\boldsymbol{P}$ as in Judge et al. (1985, pp. 26). If we suppose that the correlation matrix $\boldsymbol{\Psi}$ is known and that $\boldsymbol{P}\widetilde{\boldsymbol{X}}$ has full rank (or equivalently, $\widetilde{\boldsymbol{X}}$ has full rank), using the definition of $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{X}}$ given in (1.3), we can estimate β by the generalized least squares method (see, for example, Judge et al. 1985). Through this method, we obtain

$$\widehat{\beta} = \left(\widetilde{\boldsymbol{X}}^T\boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}^T\boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{y}},$$

by minimizing

$$\|\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{y}-\boldsymbol{X}\beta)\|_2^2,$$

and we proceed to estimate the function m by means of estimator (1.5), using $\widehat{\boldsymbol{\beta}}$.

In this paper, we focus on the Gasser and Müller (1979) weights, but our results can be extended to use other types of kernel estimators (for example, kernel estimators such that properties (3.1)-(3.5) —see below— hold with $f_1, f_2 : [0,1] \to \mathbb{R}$ bounded functions and S_ε a constant which depends on the error structure, or kernel estimators such that properties (2.2a) and (2.2b) of Speckman 1988 hold i.e., kernel estimators with "weight function of order v"). Thus, for $t \in [h, 1-h]$, we set

$$w_{n,h}(t, t_i) = h^{-1} \int_{(i-1)/n}^{i/n} K\left(\frac{t-u}{h}\right) du, \tag{2.1}$$

where $h > 0$, $t_i = (i - 1/2)/n$ and $K(\cdot)$ is a function with support on $[-1, 1]$. Because the function to be estimated has bounded support ($[0, 1]$), if $t = t(n) = qh \in [0, h)$ or $t = t(n) = 1 - qh \in (1 - h, 1]$ ("boundary intervals"), the support of the bandwidth-scaled kernel function for estimating in the interval t with bandwidth h is not contained in the support of the function, so that some mass of the scaled kernel is not matched by the data (for example, whenever $t = qh \in [0, h)$ only the interval $[-1, q]$ of the support $[-1, 1]$ of the kernel K is mapped into $[0, 1]$). Therefore, the bias of the nonparametric estimator $m_{n,h}(t)$ (see (1.2)) has different orders in $t \in [0, 1]$. This is known as a "boundary problem" or a "boundary effect". The solution proposed by Gasser and Müller (1984) for the boundary problem is the introduction of modified kernels $K_q(\cdot)$ ($K_q^*(\cdot)$) for estimating in the interval $t = t(n) = qh \in [0, h)$ ($t = t(n) = 1 - qh \in (1 - h, 1]$), defined as follows (see Gasser and Müller 1984).

Definition 2.1. A function $K_q : \mathbb{R} \to \mathbb{R}$ ($K_q^* : \mathbb{R} \to \mathbb{R}$) is called a boundary kernel of order v (for some integer $v \geq 1$) for estimating in the interval $t = t(n) = qh \in [0, h)$ ($t = t(n) = 1 - qh \in (1 - h, 1]$) if:

(a) K_q (K_q^*) has support $[-1, q]$ ($[-q, 1]$) and is Hölder continuous on it.

(b) $\int K_q(u)du = 1$, $\int u^v K_q(u)du \neq 0$ and $\int u^z K_q(u)du = 0$, $z = 1, \ldots, v-1$ ($\int K_q^*(u)du = 1$, $\int u^v K_q^*(u)du \neq 0$ and $\int u^z K_q^*(u)du = 0$, $z = 1, \ldots, v-1$).

(c) $\sup_{q\subset[0,1]} \left|\int u^v K_q(u)du\right| < \infty$ and $\sup_{q\subset[0,1]} \int K_q^2(u)du < \infty$ ($\sup_{q\subset[0,1]} \left|\int u^v K_q^*(u)du\right| < \infty$ and $\sup_{q\subset[0,1]} \int K_q^{*2}(u)du < \infty$).

In this paper, if t is in the boundary region, we will use modified weights (but we maintain the same notation $w_{n,h}(t,t_i)$) obtained by replacing K in (2.1) with a boundary kernel K_q (or K_q^*). We use $t_i = (i-1/2)/n$ for simplicity, but it would suffice that $t_i \leq S_i \leq t_{i+1}$, where $0 = S_0 \leq S_1 \leq \cdots \leq S_n = 1$ and $\max_i |S_i - S_{i-1}| = O(n^{-1})$. In this case, $w_{n,h}(t,t_i) = h^{-1}\int_{S_{i-1}}^{S_i} K((t-u)/h)du$.

As in Speckman (1988), we will assume that there exist smooth functions $g_j(\cdot) : [0,1] \to \mathbb{R}$ such that

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

where $\{\eta_{ij}\}$ is a sequence of real numbers or random variables (in this case, ε_i and η_{ij} are assumed independent, and we must interpret our analysis as being conditional on $\{\eta_{ij}\}$. Therefore, the $O(\cdot)$ terms should be interpreted as bounds holding in probability with respect to the distribution of the $\{\eta_{ij}\}$; see Speckman 1988, pp. 418-419, for more details on this relationship). More assumptions on η_{ij} will be specified later.

Our results for the asymptotic normality will be valid under the following general dependence structure.

Definition 2.2. Let $\mathbb{N}^*$ denote the set of positive integers, and for any i and j in $\mathbb{N}^* \cup \{\infty\}$ $(i \leq j)$ define $\mathcal{F}_i^j$ to be the σ-algebra spanned by the variables $Z_i,\ldots,Z_j$. The sequence $\{Z_i\}$ is said to be *α-mixing* (or *strong mixing)* if there exist mixing coefficients $\alpha(m)$ such that $\lim_{m\to\infty}\alpha(m) = 0$, and for positive integers k and m and for any sets A and B that are, respectively, $\mathcal{F}_1^k$−measurable and $\mathcal{F}_{k+m}^\infty$−measurable,

$$|P(A\cap B) - P(A)P(B)| \leq \alpha(m).$$

We refer to the monograph of Doukhan (1994) for properties of this or more mixing conditions. In this monograph, it can be observed how several regular processes satisfy the strong mixing condition. For example, the stationary ARMA processes are strong mixing, provided the innovations have absolutely continuous distribution with respect to Lebesgue measure

(Mokkadem 1988). For references on kernel estimation with mixing data see, for example, Györfi et al. (1990).

We will write $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n)^T$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p)$, where $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})^T$ and $\boldsymbol{\eta}_j = (\eta_{1j}, \ldots, \eta_{nj})^T$, for each $j = 1, \ldots, p$. Also, for $f : [0,1] \to \mathbb{R}$, we denote $\widetilde{f}(t_i) = f(t_i) - \sum_{j=1}^n w_{n,h}(t_i, t_j) f(t_j)$, $\widetilde{\boldsymbol{f}} = (\widetilde{f}(t_1), \ldots, \widetilde{f}(t_n))^T$ and $\boldsymbol{f} = (f(t_1), \ldots, f(t_n))^T$. Here, $c(j)$ denotes the covariance between ε_i and ε_{i+j} (for the stationary process $\{\varepsilon_i\}$); furthermore, $tr(\boldsymbol{A})$ denotes the trace of the matrix $\boldsymbol{A}$, i.e. $tr(\boldsymbol{A}) = \sum_{i=1}^n A_{ii}$, while $||\boldsymbol{A}||_p$ is the L_p norm of the matrix $\boldsymbol{A}$, i.e.,

$$||\boldsymbol{A}||_p = \max_{||\boldsymbol{v}||_p > 0} ||\boldsymbol{A}\boldsymbol{v}||_p \,/\, ||\boldsymbol{v}||_p \;\text{ for } \boldsymbol{v}(v_1, \ldots, v_n)^T,$$

where

$$||\boldsymbol{v}||_p^p = \sum_{i=1}^n |v_i|^p\,, \qquad 1 \le p < \infty,$$

and

$$||\boldsymbol{v}||_\infty = \max_{1 \le i \le n} |v_i|\,.$$

It can be shown that

$$\begin{aligned} ||\boldsymbol{A}||_1 &= \max_{1 \le j \le n} \sum_{i=1}^n |A_{ij}|\,, \\ ||\boldsymbol{A}||_2 &= \sqrt{\text{maximum eigenvalue of } \boldsymbol{A}^T\boldsymbol{A}} \end{aligned}$$

and that

$$||\boldsymbol{A}||_\infty = \max_{1 \le i \le n} \sum_{j=1}^n |A_{ij}|\,.$$

In what follows, we always consider that $n \to \infty$, $h \to 0$ and $nh^v \to \infty$ (for an integer $v \ge 1$ that will be defined in the assumptions).

We impose the following assumptions:

(A.1) (a) $\{\varepsilon_i\}_{i=1}^n$ are stationary and $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma_\varepsilon^2 < \infty$;
(b) $\{\varepsilon_i\}_{i=1}^n$ are strong mixing.

(A.2) $E\,|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$.

(A.3) $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$, where $\alpha(n)$ are the mixing coefficients of $\{\varepsilon_i\}_i$.

(A.4) $\sum_{k=1}^{\infty} k\,|c(k)| < \infty$.

(A.5) (a) $||\mathbf{\Psi}||_2 = O(1)$;
(b) $||\mathbf{\Psi}^{-1}||_{\infty} = O(1)$.

(A.6) (a) For $t \in [h, 1-h]$ we use the weights $w_{n,h}(t, t_i)$ (see (2.1)), where $K : \mathbb{R} \to \mathbb{R}$ is Hölder continuous, with support $[-1, 1]$. Furthermore $\int K(u)du = 1$, $\int u^v K(u)du \neq 0$ and $\int u^z K(u)du = 0$, $z = 1, \ldots, v-1$, for some integer $v \geq 1$;
(b) For $t \in [0, h) \cup (1-h, 1]$ we use modified weights obtained by replacing K in (2.1) with a boundary kernel of order v (see Definition 2.1).

(A.7) The functions $m(\cdot), g_1(\cdot), \ldots, g_p(\cdot)$ have $v \geq 1$ continuous derivatives on $[0, 1]$.

(A.8) The components of $\boldsymbol{X}$ are uniformly bounded.

(A.9) The design points t_i are $t_i = (i - 0.5)/n$, $i = 1, \ldots, n$.

(A.10) $n^{-1}\boldsymbol{\eta}^T \mathbf{\Psi}^{-1}\boldsymbol{\eta} \longrightarrow \boldsymbol{V}$ where $\boldsymbol{V} = \{V_{ij}\}$ is a positive definite matrix.

(A.11) $||\boldsymbol{W}\boldsymbol{\eta}_j||_2^2 = O(h^{-1}) = ||\boldsymbol{W}^T\boldsymbol{\eta}_j||_2^2$, $1 \leq j \leq p$.

(A.12) $n^{-1}\boldsymbol{\eta}^T \mathbf{\Psi}^{-1}\widetilde{\boldsymbol{m}} = O(n^{-1/2}h^v)$.

(A.13) $||\boldsymbol{W}^T \mathbf{\Psi}^{-1}\boldsymbol{\eta}_j||_2 = O(ne(n))$, $1 \leq j \leq p$, where $e(n) = c_1 h^{2v} + c_2(nh)^{-1}$.

(A.14) $nh^{4v} \to 0$, $nh^2 \to \infty$.

Remark 2.1. Assumptions (A.1.a), (A.2), (A.6), (A.7), (A.9) and (A.14) are frequent conditions in the setting of kernel smoothing. Assumption (A.6b) is sufficient to avoid boundary effects (see Gasser and Müller 1979, 1984). The existence of such boundary kernels for arbitrary $v \in \mathbb{N}$ is established in Gasser et al. (1985). In the setting of partial linear models, the assumption (A.8) is used by Speckman (1988).

Remark 2.2. Suppose that the rows of $\boldsymbol{\eta}$, $(\eta_{i1}, \dots, \eta_{ip})$, $i = 1, \dots, n$, are independent and identically distributed random vectors with mean zero and finite variance-covariance matrix $\Sigma_\eta = (\Sigma_{ij})$, and that $\{\varepsilon_j\}$ is a stationary autoregressive process of order $k \geq 1$ ($AR(k)$ process, i.e., $\varepsilon_j = \phi_1\varepsilon_{j-1} + \phi_2\varepsilon_{j-2} + \cdots + \phi_k\varepsilon_{j-k} + e_j$, where $\{e_j\}$ is a zero mean white noise process independent of $\{\varepsilon_j\}$, and $\phi(z) = 1 - \phi_1 z - \cdots - \phi_k z^k \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$). In this case, the inverse of the correlation matrix can be seen in Wise (1955). Furthermore, let us assume that the innovations e_j have an absolutely continuous distribution with respect to Lebesgue measure. Then $\{\varepsilon_j\}$ is strong mixing with mixing coefficients $\alpha(n) = O(d^n)$, and $c(n) = O(s^n)$ $(0 < d, s < 1)$ (see Mokkadem 1988, and exercise 3.11 of Brockwell and Davis 1991); therefore (A.1.b), (A.3) (A.5a) hold. Utilising the expression for $\boldsymbol{\Psi}^{-1}$, it is easy to see that (A.5b) holds and, together with the above conditions on $\boldsymbol{\eta}$, we see that (A.10) holds in probability, where $\boldsymbol{V} = (\sigma_\varepsilon^2/\sigma_e^2)(1 + \sum_{i=1}^k \phi_i^2)\Sigma_\eta$ (we denote $\sigma_e^2 = Var(e_i)$). We have that $E||\boldsymbol{W}\boldsymbol{\eta}_j||_2^2 = E||\boldsymbol{W}^T\boldsymbol{\eta}_j||_2^2 = \Sigma_{jj}tr(\boldsymbol{W}^T\boldsymbol{W})$, so assumption (A.11) follows from assumption (A.6). Moreover, $E(\boldsymbol{\eta}_j^T\boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}}) = 0$ and $Var(\boldsymbol{\eta}_j^T\boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}}) = \widetilde{\boldsymbol{m}}^T\boldsymbol{\Psi}^{-1}\Sigma_{jj}\boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}}$, so assumption (A.12) follows from (3.2), (3.3) (see Section 3) and assumption (A.5b). By Whittle's inequality (Whittle 1960), and using that under (A.6) it verifies that $\max_{i,j}|w_{n,h}(t, t_i)| = O((nh)^{-1})$, we obtain that

$$E\left(||\boldsymbol{W}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_j||_2^2\right) = O(h^{-2}) = O(\{ne(n)\}^2),$$

and (A.13) holds.

Remark 2.3. Justification of the assumptions (A.5b), (A.10), (A.12) and (A.13) is not possible under a general strong mixing condition, because we need the structure of $\boldsymbol{\Psi}^{-1}$. Thus, in the above remark we have focused on the $AR(k)$ process. This condition is not more restrictive than that given in the related literature. In a model like (1.1), Schick (1996, 1998) assumes $AR(1)$ errors, Gao (1995) supposes that $\{\varepsilon_i\}$ is a class of linear processes and Schick (1999) works with $ARMA(1, 1)$ errors. Furthermore, these authors do not use $\boldsymbol{\Psi}^{-1}$.

In Section 3 we obtain the following results

Theorem 2.1.

(a) *Under assumptions (A.1a), (A.5b)-(A.7) and (A.9)-(A.12) we have:*

$$E(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = O\left(h^{2v}\right) + O(h^{v}(nh)^{-1/2}).$$

(b) *Under assumptions (A.1a), (A.5)-(A.7), (A.9)-(A.11), (A.13) and (A.14) we have:*

$$Var(\widehat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2 n^{-1}\boldsymbol{V}^{-1} + o(n^{-1}).$$

In the next theorem, let $\widehat{m}_{h,0}(t)$ denote the estimator of $m(t)$ that would be obtained by kernel smoothing if $\boldsymbol{\beta}$ were known precisely, i.e.

$$\widehat{m}_{h,0}(t) = \sum_{i=1}^{n} w_{n,h}(t,t_i)(y_i - \boldsymbol{\zeta}_i^T\boldsymbol{\beta}).$$

Theorem 2.2.

(a) *Under the assumptions in part (a) of the Theorem 2.1, we have:*

$$Bias(\widehat{m}_h(t)) = Bias(\widehat{m}_{h,0}(t))(1+o(1)) = O(h^v).$$

(b) *Under assumptions in part (b) of the Theorem 2.1, together with (A.4), we have:*

$$Var(\widehat{m}_h(t)) = Var(\widehat{m}_{h,0}(t))(1+o(1)) = O((nh)^{-1}).$$

Theorem 2.3. *Under assumptions (A.1)-(A.3), (A.5)-(A.11), (A.13) and (A.14), we have:*

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - E(\widehat{\boldsymbol{\beta}})) \xrightarrow{D} \mathcal{N}(\boldsymbol{0}, \sigma_\varepsilon^2\boldsymbol{V}^{-1}).$$

Corollary 2.1. *Under assumption (A.12) and the conditions of Theorem 2.3, we have:*

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathcal{N}(\boldsymbol{0}, \sigma_\varepsilon^2\boldsymbol{V}^{-1}).$$

Remark 2.4. Note that in Theorem 2.3 we need a mixing condition on the errors. Essentially, this condition is neccesary to apply Lemma 1.1 in Volkonskii and Rozanov (1959).

Obviously, in a practical case, when we use a set of data, it is difficult to know the exact form of the matrix $\boldsymbol{\Psi}$ and, consequently, of $\boldsymbol{\Psi}^{-1}$. In this case, we need to obtain an estimate of this matrix. In the most general case $\boldsymbol{\Psi}$ will have $[(n(n+1)/2)-1]$ different unknown parameters, but it is customary to make some further assumptions on the structure of this matrix. A usual condition in the econometric literature entails that the elements in $\boldsymbol{\Psi}$ are functions of a $(k\times 1)$ vector $\boldsymbol{\phi}$, where $k<n$ and k remains constant as n increases. Then the problem of estimating $\boldsymbol{\Psi}=\boldsymbol{\Psi}(\boldsymbol{\phi})$ reduces to one of estimating $\boldsymbol{\phi}$ (see Judge et al. 1985).

Let $\widehat{\boldsymbol{\Psi}}$ be an estimator of $\boldsymbol{\Psi}$. Let

$$\widehat{\widehat{\boldsymbol{\beta}}}=\left(\widetilde{\boldsymbol{X}}^T\widehat{\boldsymbol{\Psi}}^{-1}\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}^T\widehat{\boldsymbol{\Psi}}^{-1}\widetilde{\boldsymbol{y}}.$$

Evaluation of the finite sample properties of $\widehat{\widehat{\boldsymbol{\beta}}}$ is, in general, a difficult problem, because $\widehat{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{y}}$ will be correlated. Consequently, inferences about $\boldsymbol{\beta}$ need to be based on the asymptotic distribution of $\widehat{\widehat{\boldsymbol{\beta}}}$. For the asymptotic properties of $\widehat{\widehat{\boldsymbol{\beta}}}$ we first investigate the asymptotic properties of $\widehat{\boldsymbol{\beta}}$ (Theorems 2.1 and 2.3), and then we give sufficient conditions to show that $\widehat{\widehat{\boldsymbol{\beta}}}$ and $\widehat{\boldsymbol{\beta}}$ have the same asymptotic distribution.

We suppose that

(A.15) $n^{-1}\widetilde{\boldsymbol{X}}^T(\widehat{\boldsymbol{\Psi}}^{-1}-\boldsymbol{\Psi}^{-1})\widetilde{\boldsymbol{X}}\longrightarrow\boldsymbol{0}$ in probability.

(A.16) $n^{-1/2}\widetilde{\boldsymbol{X}}^T(\widehat{\boldsymbol{\Psi}}^{-1}-\boldsymbol{\Psi}^{-1})(\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{m}+\boldsymbol{\varepsilon})\longrightarrow\boldsymbol{0}$ in probability.

Theorem 2.4. *Under conditions of Corollary 2.1 and assumptions (A.15) and (A.16), we have:*

$$n^{1/2}(\widehat{\widehat{\boldsymbol{\beta}}}-\boldsymbol{\beta})\xrightarrow{D}\mathcal{N}(\boldsymbol{0},\sigma_\varepsilon^2\boldsymbol{V}^{-1}).$$

Remark 2.5. Assumptions (A.15) and (A.16) have the form of the conditions used in the estimation of $\boldsymbol{\beta}$ in a classical linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (see, for example, Judge et al. 1985). Nevertheless, if we do not assume a parametric structure for the errors, it is difficult to obtain an estimator $\widehat{\boldsymbol{\Psi}}$ which satisfies both. A possible procedure to estimate $c(k) = Cov(\varepsilon_j, \varepsilon_{j+k})$, and therefore $\boldsymbol{\Psi}$, might be based on second order differences defined as $\widehat{\varepsilon}_{i,k,\mu} = y_i - k(k+\mu)^{-1}y_{i+\mu} - \mu(k+\mu)^{-1}y_{i-k}$ (see Herrmann et al. 1992, for the expression of $\widehat{c}(k)$). Nevertheless, (A.15) and (A.16) are not necessarily verified.

Now we define the following conditions, which are an extension of regularity conditions (1) (3) from Fuller and Battese (1973).

(A.17) The elements of $\boldsymbol{\Psi} = \boldsymbol{\Psi}_n$ are functions of a $(k \times 1)$ vector of parameters $\boldsymbol{\phi}$, such that the elements of the matrices

$$\boldsymbol{S}_{nr}(\boldsymbol{\phi}) = \frac{\partial}{\partial \phi_r} \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\phi}) \quad (r = 1, 2, \ldots, k)$$

are continuous functions of $\boldsymbol{\phi}$ in an open sphere $\boldsymbol{B}$ of $\boldsymbol{\phi}^0$, the true value of the parameter vector $\boldsymbol{\phi}$.

(A.18) The sequences of matrices $\{\widetilde{\boldsymbol{X}}_n\}$ and $\{\boldsymbol{\Psi}_n\}$ are such that

$$\lim_{n\to\infty} n^{-1} \widetilde{\boldsymbol{X}}_n^T \boldsymbol{S}_{nr}(\boldsymbol{\phi}) \widetilde{\boldsymbol{X}}_n = \boldsymbol{H}_r(\boldsymbol{\phi}),$$

where $\boldsymbol{H}_r(\boldsymbol{\phi})$ is a matrix whose elements are continuous functions of $\boldsymbol{\phi}$, $r = 1, \ldots, k$, and

$$n^{-1} \widetilde{\boldsymbol{X}}_n^T \boldsymbol{S}_{nr}(\boldsymbol{\phi}) (\widetilde{\boldsymbol{m}}_n + \widetilde{\boldsymbol{\varepsilon}}_n) = O_p(n^{-1/2}).$$

(A.19) An estimator, $\widehat{\boldsymbol{\Psi}} = \boldsymbol{\Psi}(\widehat{\boldsymbol{\phi}})$, for $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\phi}^0)$ is available such that $\boldsymbol{\Psi}^{-1}(\widehat{\boldsymbol{\phi}})$ exists for all n, and $\widehat{\boldsymbol{\phi}}$ satisfies the condition

$$\widehat{\boldsymbol{\phi}} = \boldsymbol{\phi}^0 + o_p(1).$$

Theorem 2.5. *Under the conditions of Corollary 2.1 and assumptions (A.17), (A.18) and (A.19), we have:*

$$n^{1/2}(\widehat{\widehat{\boldsymbol{\beta}}} - \boldsymbol{\beta}\ ) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{V}^{-1}).$$

Remark 2.6. This theorem is an extension of Theorem 3 in Fuller and Battese (1973).

Remark 2.7. Assumption (A.18) is probably not the most natural, and is certainly not the most general. For autoregressive errors of order k, we can change $\boldsymbol{\Psi}^{-1}$ to $(\sigma_\varepsilon^2/\sigma_e^2)^{-1}\boldsymbol{\Psi}^{-1}$ in $\widehat{\boldsymbol{\beta}}$ (see Wise 1955). In this case, (A.17) holds (see Wise 1955). In addition, if the rows of $\boldsymbol{\eta}$ are i.i.d. random vectors with mean zero and finite variance-covariance matrix Σ_η, then (A.18) follows from (A.6), (A.7) and (A.11), with $\boldsymbol{H}_r(\boldsymbol{\phi}) = 2\phi_r\Sigma_\eta$, $r = 1, \ldots, k$. Furthermore, if we denote $\overline{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{y}}$ and $\overline{m}_h(t_i) = \sum_{j=1}^n w_{n,h}(t_i, t_j)(y_j - \boldsymbol{\zeta}_j^T\overline{\boldsymbol{\beta}})$, it can be shown using the methods of Section 3 that $\|\overline{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 = O_p(n^{-1/2}) = o_p(1)$ and $\sup_i |\overline{m}_h(t_i) - m(t_i)| = O_p(h^v + (nh)^{-1/2}) = o_p(1)$. Together with (A.8) (this assumption can be changed to the less restrictive assumption: $\sup_{i,j} E|\eta_{ij}|^{2+\delta'} < C < \infty$, for some $\delta' > 0$), we have that $\sup_i |\boldsymbol{\zeta}_i^T\overline{\boldsymbol{\beta}} + \overline{m}_h(t_i) - \boldsymbol{\zeta}_i^T\boldsymbol{\beta} - m(t_i)| = o_p(1)$. Let $\widehat{\varepsilon}_i = y_i - \boldsymbol{\zeta}_i^T\overline{\boldsymbol{\beta}} - \overline{m}_h(t_i)$. From Theorem 1 in Cao et al. (1995) we obtain that $\widehat{\boldsymbol{\phi}}$ (an estimator that uses $\widehat{\varepsilon}_i$) is consistent for $\boldsymbol{\phi}^0$, where the estimation is carried out using any mechanism that would be consistent if the estimation were made using the unobserved series $\{\varepsilon_i\}$. For details about several consistent methods for estimating $\boldsymbol{\phi}^0$ using $\{\varepsilon_i\}$, see Brockwell and Davis (1991, Ch. 8).

3 Proofs

It is easy to see that, under assumptions (A.6) and (A.9), we have

$$||\boldsymbol{W}||_\infty = O(1), \qquad ||\boldsymbol{W}||_1 = O(1). \tag{3.1}$$

Furthermore, let us denote $f_{n,h}(t) = \sum_{i=1}^n w_{n,h}(t, t_i)(f(t_i) + \varepsilon_i)$, where $f : [0, 1] \to \mathbb{R}$ has $v \geq 1$ continuous derivatives and $nh^v \to \infty$. Gasser and

Müller (1984) show that, under assumptions (A.1a), (A.6) and (A.9) we have that

$$Bias(f_{n,h}(t)) = h^v f_1(t) m^{(v)}(t) + o(h^v), \tag{3.2}$$

uniformly in $t \in [h, 1-h]$, and

$$Bias(f_{n,h}(t)) = O(h^v) \tag{3.3}$$

uniformly in $t \in [0,1]$. In addition, if we assume (A.4), then (see Hart 1991)

$$Var(f_{n,h}(t)) = (nh)^{-1} f_2(t) S_\varepsilon (1 + o(1)) \tag{3.4}$$

uniformly in $t \in [h, 1-h]$, and

$$Var(f_{n,h}(t)) = O((nh)^{-1}) \tag{3.5}$$

uniformly in $t \in [0,1]$. We have denoted $f_1(t) = (-1)^v (v!)^{-1} \int_{-1}^{1} u^v K(u) du$, $f_2(t) = \int_{-1}^{1} K^2(u) du$ and $S_\varepsilon = c(0) + 2\sum_{k=1}^{\infty} c(k)$.

Now, we demonstrate the following lemma, which is required later.

Lemma 3.1. *Under assumptions (A.5b) (A.7) and (A.9) (A.11) we have*

$$n^{-1}(\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{X}}) \longrightarrow \boldsymbol{V}.$$

Proof. The $(i,j)-$th element of $n^{-1}(\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{X}})$ is

$$\begin{aligned} n^{-1} \widetilde{\boldsymbol{x}}_i^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{x}}_j &= n^{-1}(\widetilde{\boldsymbol{g}}_i^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{g}}_j + \widetilde{\boldsymbol{g}}_i^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{\eta}}_j + \widetilde{\boldsymbol{g}}_j^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{\eta}}_i + \boldsymbol{\eta}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}_j \\ &\quad - \boldsymbol{\eta}_i^T \boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}_j - \boldsymbol{\eta}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{W} \boldsymbol{\eta}_j + \boldsymbol{\eta}_i^T \boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \boldsymbol{W} \boldsymbol{\eta}_j), \end{aligned} \tag{3.6}$$

since $\widetilde{\boldsymbol{x}}_j = \widetilde{\boldsymbol{g}}_j + \widetilde{\boldsymbol{\eta}}_j = \widetilde{\boldsymbol{g}}_j + (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{\eta}_j$.

Using (A.10), we only have to prove that all the terms of (3.6), except $n^{-1} \boldsymbol{\eta}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}_j$, tend to zero. We also have, by assumption (A.10) (remember that $\boldsymbol{\Psi}^{-1} = \boldsymbol{P}^T \boldsymbol{P}$)

$$||\boldsymbol{P}\boldsymbol{\eta}_j||_2 = O(n^{1/2}), \tag{3.7}$$

hence

$$\begin{aligned} ||\boldsymbol{P}\widetilde{\boldsymbol{\eta}}_j||_2 &= ||\boldsymbol{P}(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{\eta}_j||_2 \le ||\boldsymbol{P}\boldsymbol{\eta}_j||_2 + ||\boldsymbol{P}||_2 \cdot ||\boldsymbol{W}\boldsymbol{\eta}_j||_2 \\ &= O(n^{1/2}) + O(h^{-1/2}) = O(n^{1/2}), \end{aligned} \tag{3.8}$$

using (A.5b), (A.11) and the fact that $nh \to \infty$.

Now, by assumption (A.5.b) and properties (3.2) and (3.3), we have

$$||\boldsymbol{P}\widetilde{\boldsymbol{g}}_j||_2 = O(n^{1/2}h^v). \tag{3.9}$$

(3.7), (3.8) and (3.9), together with (A.5b), (A.10) and (A.11), complete the proof of the lemma. □

Proof of Theorem 2.1(a). We have that the bias of $\widehat{\boldsymbol{\beta}}$ is

$$E(\widehat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = (\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}}. \tag{3.10}$$

In view of Lemma 3.1, it suffices to consider

$$\begin{aligned} n^{-1}\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}} &= n^{-1}\widetilde{\boldsymbol{g}}_i^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}} + n^{-1}\boldsymbol{\eta}_i^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}} \\ &\quad - n^{-1}\boldsymbol{\eta}_i^T \boldsymbol{W}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}}. \end{aligned} \tag{3.11}$$

The first term on the right hand side of (3.11) is $O(h^{2v})$ by (A.5b), (3.2) and (3.3). The second term is $O(n^{-1/2}h^v)$, by (A.12). The third term is $O((nh)^{-1/2}h^v)$, by (A.5b), (A.11), (3.2) and (3.3). Then,

$$n^{-1}\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{m}} = O(h^{2v}) + O((nh)^{-1/2}h^v). \tag{3.12}$$

Finally, using Lemma 3.1 and (3.10), we have the result. □

Proof of Theorem 2.1(b). We have that

$$Var(\widehat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2(\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}})^{-1} + \sigma_\varepsilon^2(\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}})^{-1}\boldsymbol{M}(\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}})^{-1},$$

where the matrix $\boldsymbol{M}$ is

$$\boldsymbol{M} = -\widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}\boldsymbol{W}\widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{X}}^T \boldsymbol{W}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}} + \widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{W}^T \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}}.$$

Using Lemma 3.1, it suffices to prove that $\boldsymbol{M} = o(n)$.

Using (3.1), we have

$$||\boldsymbol{W}^T||_2 = ||\boldsymbol{W}||_2 \leq (||\boldsymbol{W}||_1 ||\boldsymbol{W}||_\infty)^{1/2} = O(1). \tag{3.13}$$

Using (A.11) and (3.13) we have

$$\begin{aligned} ||\boldsymbol{W}\widetilde{\boldsymbol{x}}_i||_2 &\leq ||\boldsymbol{W}\boldsymbol{\eta}_i||_2 + ||\boldsymbol{W}||_2 \cdot ||\boldsymbol{g}_i - \boldsymbol{W}\boldsymbol{x}_i||_2 \\ &= O(h^{-1/2}) + O((ne(n))^{1/2}) = O((ne(n))^{1/2}). \end{aligned} \tag{3.14}$$

In this last equation we use the fact that (from assumption (A.11) and properties (3.2) and (3.3))

$$n^{-1}||\boldsymbol{g}_i - \boldsymbol{W}\boldsymbol{x}_i||_2^2 \leq e(n), \tag{3.15}$$

uniformly in i, where $e(n) = c_1 h^{2v} + c_2(nh)^{-1}$, and $ne(n) \to \infty$.

From (3.14) and (A.5b) we obtain

$$||\boldsymbol{P}\boldsymbol{W}\widetilde{\boldsymbol{x}}_j||_2 = O((ne(n))^{1/2}). \tag{3.16}$$

From Lemma 3.1 we have

$$||\boldsymbol{P}\widetilde{\boldsymbol{x}}_i||_2 = O(n^{1/2}). \tag{3.17}$$

Combining (3.16) and (3.17) it follows that, for the elements of the matrix $\boldsymbol{M}$,

$$|\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{W} \widetilde{\boldsymbol{x}}_j| = O(n(e(n))^{1/2}). \tag{3.18}$$

Also,

$$|\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{W} \boldsymbol{\Psi} \boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{x}}_j| \leq ||\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{W}||_2 ||\boldsymbol{\Psi}||_2 ||\boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{x}}_j||_2. \tag{3.19}$$

For the last term on right hand side of this inequality we have

$$||\boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{x}}_j||_2 \leq ||\boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}_j||_2 + ||\boldsymbol{W}^T||_2 ||\boldsymbol{\Psi}^{-1}||_2 ||\boldsymbol{g}_j - \boldsymbol{W}\boldsymbol{x}_j||_2. \tag{3.20}$$

Using (A.5b) and (A.13), together with (3.13), (3.15) and (3.20), we have

$$||\boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{x}}_j||_2 = O(ne(n)). \tag{3.21}$$

Then, it follows, by assumption (A.5a), (3.19) and (3.21), that

$$|\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{W} \boldsymbol{\Psi} \boldsymbol{W}^T \boldsymbol{\Psi}^{-1} \widetilde{\boldsymbol{x}}_j| = O(n^2 e^2(n)). \tag{3.22}$$

Due to (3.18), (3.22) and (A.14), we obtain

$$\boldsymbol{M} = O(n(e(n))^{1/2}) + O(n^2 e^2(n)) = o(n). \tag{3.23}$$

□

Proof of Theorem 2.2(a). From the expression for the estimate $\widehat{m}_h(t)$ (1.5) we obtain

$$Bias(\widehat{m}_h(t)) = Bias(\widehat{m}_{h,0}(t)) + \boldsymbol{w}^T(t)\boldsymbol{X}Bias(\widehat{\boldsymbol{\beta}}), \tag{3.24}$$

with $\boldsymbol{w}^T(t) = (w_{n,h}(t,t_1), \ldots, w_{n,h}(t,t_n))$.

In view of (3.2), (3.3) and (A.11) it is easy to see that

$$\boldsymbol{w}^T(t)\boldsymbol{X} \longrightarrow (g_1(t), \ldots, g_p(t)) \qquad \text{as } n \to \infty. \tag{3.25}$$

Now, using (3.24), to finish the proof we only have to check that

$$\frac{\boldsymbol{w}^T(t)\boldsymbol{X}Bias(\widehat{\boldsymbol{\beta}})}{Bias(\widehat{m}_{h,0}(t))} = o(1). \tag{3.26}$$

Using (3.25), Theorem 2.1 part (a) and the asymptotic expression for the bias of $\widehat{m}_{h,0}(t)$ ((3.2) and (3.3)), we get an asymptotic order of $O(h^v) + O((nh)^{-1/2}) = o(1)$ for the quotient of (3.26).

It follows that $Bias(\widehat{m}_h(t)) = Bias(\widehat{m}_{h,0}(t))(1 + o(1)) = O(h^v)$, where this last order is a result of (3.2) and (3.3). □

Proof of Theorem 2.2(b). From the expression of the estimate $\widehat{m}_h(t)$ (1.5) we have

$$\begin{aligned} Var(\widehat{m}_h(t)) &= Var(\widehat{m}_{h,0}(t)) + \boldsymbol{w}^T(t)\boldsymbol{X}Var(\widehat{\boldsymbol{\beta}})\boldsymbol{X}^T\boldsymbol{w}(t) \\ &\quad -2\boldsymbol{w}^T(t)(t)\boldsymbol{X}Cov(\widehat{\boldsymbol{\beta}}, \widehat{m}_{h,0}(t)). \end{aligned}$$

Using the convergence of $\boldsymbol{w}^T(t)\boldsymbol{X}$ (3.25) together with $Var(\widehat{\boldsymbol{\beta}}) = O(n^{-1}) = o(Var(\widehat{m}_{h,0}(t)))$ and $|Cov(\widehat{\beta}_k, \widehat{m}_{h,0}(t))| = o(Var(\widehat{m}_{h,0}(t)))$, $k = 1, \ldots, p$, where $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T$ (from Theorem 2.1(b), (3.4) and (3.5)), the result of the theorem holds. □

Proof of Theorem 2.3. Notice that

$$\widehat{\boldsymbol{\beta}} - E(\widehat{\boldsymbol{\beta}}) = (\widetilde{\boldsymbol{X}}^T\boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{\varepsilon}.$$

Let us define $\boldsymbol{c}_n^T = (c_{1n}, \ldots, c_{nn}) = \boldsymbol{a}^T \widetilde{\boldsymbol{X}}^T \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{W})$ (where $\boldsymbol{a} \neq 0$ is a fixed arbitrary $p \times 1$ vector), $\sigma_n = (Var(\boldsymbol{c}_n^T \boldsymbol{\varepsilon}))^{1/2} = (Var(\sum_{i=1}^n c_{in}\varepsilon_i))^{1/2}$, and

$$S_n = \sum_{i=1}^n (c_{in}/\sigma_n)\varepsilon_i.$$

Because of Lemma 3.1 and (3.23), to prove the theorem, we only have to check that

$$S_n \xrightarrow{D} \mathcal{N}(0,1). \tag{3.27}$$

To see this, we first check that

$$\max_{1 \le i \le n} \frac{c_{in}^2}{\sigma_n^2} = O(n^{-1}). \tag{3.28}$$

Using (A.5.b), (A.8) and (3.1), we easily obtain that

$$||\boldsymbol{c}_n||_\infty = O(1). \tag{3.29}$$

From Lemma 3.1 and the proof of Theorem 2.1(b) we have

$$n^{-1}\sigma_n^2 \longrightarrow \sigma_\varepsilon^2 \boldsymbol{a}^T \boldsymbol{V} \boldsymbol{a}. \tag{3.30}$$

Now, (3.28) follows from (3.29) and (3.30).

If we denote $Z_{in} = (c_{in}/\sigma_n)\varepsilon_i$, then, from (A.1a) and (3.28), we obtain

$$\max_{1 \le i \le n} Var(Z_{in}) = O(n^{-1}) \tag{3.31}$$

and, using Davydov's inequality (1968) ((A.1) and (A.2)),

$$|Cov(Z_{in}, Z_{jn})| \le Cn^{-1}\alpha(|i-j|)^{\frac{\delta}{2+\delta}} \quad . \tag{3.32}$$

Furthermore, using (A.3) and the fact that the sequence $\{\alpha(n)\}$ decreases to zero, it is easy to see that $n^{(1-b)(2+\delta)/\delta}\alpha(n) \to 0$ as $n \to \infty$, for all $b > 0$. Therefore, if we consider $a = (1-b_0)(2+\delta)/\delta - 1$, where $0 < b_0 < 1 - \delta/(2+\delta)$ (observe that $a > 0$), we have that

$$n^{1+a}\alpha(n) \longrightarrow 0. \tag{3.33}$$

Now, if we define $p_n = [n^{1-c(1+a)}]$ and $q_n = [n^c]$ with $(2+2a)^{-1} < c < (2+a)^{-1}$ ($[s]$ denotes the integer part of s), it is easy to see that

$$p_n \to \infty, \quad q_n \to \infty, \quad p_n^{-1} q_n \to 0 \quad \text{and} \quad n^{-1} p_n^2 \to 0 \tag{3.34}$$

and, together with (3.33),

$$n p_n^{-1} \alpha(q_n) \longrightarrow 0. \tag{3.35}$$

The asymptotic normality of S_n is proved using a classical argument which consists in decomposing the sum of dependent random variables into a sum of large and small blocks, where the contribution of the small ones is negligible and where the large ones are approximately independent. After this, Lindeberg-Feller's central limit theorem is used. Because of (3.31), (3.32), (3.34), (3.35) and (A.3), the proof of the asymptotic normality of S_n is similar to the one presented in Roussas et al. (1992). For this reason, we omit the proof. □

Proof of Corollary 2.1. Follows from Theorem 2.1(a) and Theorem 2.3. □

Proof of Theorem 2.4. We have that

$$\begin{aligned} n^{1/2}(\widehat{\widehat{\boldsymbol\beta}} - \widehat{\boldsymbol\beta}) &= \left(n^{-1} \widetilde{\boldsymbol X}^T \widehat{\boldsymbol\Psi}^{-1} \widetilde{\boldsymbol X}\right)^{-1} n^{-1/2} \widetilde{\boldsymbol X}^T \widehat{\boldsymbol\Psi}^{-1} (\boldsymbol I - \boldsymbol W)(\boldsymbol m + \boldsymbol\varepsilon) \\ &\quad - \left(n^{-1} \widetilde{\boldsymbol X}^T \boldsymbol\Psi^{-1} \widetilde{\boldsymbol X}\right)^{-1} n^{-1/2} \widetilde{\boldsymbol X}^T \boldsymbol\Psi^{-1} (\boldsymbol I - \boldsymbol W)(\boldsymbol m + \boldsymbol\varepsilon). \end{aligned}$$

Now, Theorem 2.4 follows from (A.15), (A.16) and Corollary 2.1 (having taken into account that, by Lemma 3.1, (3.12) and (3.27), we have that $n^{-1} \widetilde{\boldsymbol X}^T \boldsymbol\Psi^{-1} \widetilde{\boldsymbol X} \to \boldsymbol V$ and $n^{-1/2} \widetilde{\boldsymbol X}^T \boldsymbol\Psi^{-1} (\boldsymbol I - \boldsymbol W)(\boldsymbol m + \boldsymbol\varepsilon) \xrightarrow{D} \mathcal{N}(\boldsymbol 0, \sigma_\varepsilon^2 \boldsymbol V)$). □

Proof of Theorem 2.5. This proof is a modification of the proof of Theorem 3 in Fuller and Battese (1973). Following their indications, we have

$$\widehat{\widehat{\boldsymbol\beta}} = \boldsymbol\beta + \left(\widetilde{\boldsymbol X}_n^T \boldsymbol\Psi_n^{-1}(\widehat{\boldsymbol\phi}) \widetilde{\boldsymbol X}_n\right)^{-1} \widetilde{\boldsymbol X}_n^T \boldsymbol\Psi_n^{-1}(\widehat{\boldsymbol\phi})(\widetilde{\boldsymbol m}_n + \widetilde{\boldsymbol\varepsilon}_n).$$

By a Taylor's expansion we obtain

$$\begin{aligned}
&\left(\widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\widehat{\boldsymbol{\phi}})\widetilde{\boldsymbol{X}}_n\right)^{-1} \widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\widehat{\boldsymbol{\phi}})(\widetilde{\boldsymbol{m}}_n + \widetilde{\boldsymbol{\varepsilon}}_n) \\
&\quad = \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\phi}^0)\widetilde{\boldsymbol{X}}_n\right)^{-1} \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\phi}^0)(\widetilde{\boldsymbol{m}}_n + \widetilde{\boldsymbol{\varepsilon}}_n)\right) \\
&\qquad + \sum_{r=1}^{k} \Bigg\{ \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\phi}^*)\widetilde{\boldsymbol{X}}_n\right)^{-1} \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{S}_{nr}(\boldsymbol{\phi}^*)(\widetilde{\boldsymbol{m}}_n + \widetilde{\boldsymbol{\varepsilon}}_n)\right) \\
&\qquad - \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\phi}^*)\widetilde{\boldsymbol{X}}_n\right)^{-1} \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{S}_{nr}(\boldsymbol{\phi}^*)\widetilde{\boldsymbol{X}}_n\right) \\
&\qquad \times \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\phi}^*)\widetilde{\boldsymbol{X}}_n\right)^{-1} \left(n^{-1}\widetilde{\boldsymbol{X}}_n^T \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\phi}^*)(\widetilde{\boldsymbol{m}}_n + \widetilde{\boldsymbol{\varepsilon}}_n)\right) \Bigg\} \left(\widehat{\phi}_r - \phi_r^0\right),
\end{aligned}$$

where $\boldsymbol{\phi}^*$ is between $\boldsymbol{\phi}^0$ and $\widehat{\boldsymbol{\phi}}$. By Lemma (3.1), (3.12), (3.27), (A.18) and (A.19), it follows that

$$\widehat{\widehat{\boldsymbol{\beta}}} - \boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} + o_p(n^{-1/2}).$$

By Corollary 2.1, the result of the theorem holds. □

Acknowledgments

The authors are grateful to the referees for their constructive suggestions which have greatly improved this paper.

References

Brockwell, P. and R.A. Davis (1991). *Time Series: Theory and Methods.* Springer, New York.

Cao, R., I. García, W. González, J.M. Prada and M. Febrero (1995). Predicting using Box-Jenkins, nonparametric, and bootstrap techniques. *Technometrics*, **37**, 303–310.

Daniel, C. and F.S. Wood (1980). *Fitting Equations to Data: Computer Analysis of Multifactor Data for Scientists and Engineers.* John Wiley and Sons, New York.

Davydov, Yu.A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and its Applications*, **13**, 691–696.

Denby, L. (1986). Smooth regression functions. Statistical Research Report , 26. AT&T Bell Laboratories, Murray Hill.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics, **85.** Springer, New York.

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

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications.* Chapman and Hall.

Fuller, W.A. and G.E. Battese (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association*, **68**, 626–632.

Gao, J. (1995). Asymptotic theory for partly linear models. *Communications in Statistics-Theory and Methods*, **24**, 1985–2009.

Gasser, T. and H.G. Müller (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation.* Lecture Notes in Mathematics, **757** (T. Gasser and M. Rosenblatt, eds.) Springer, New York.

Gasser, T. and H.G. Müller (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, **11**, 171–185.

Gasser, T., H.G. Müller and V. Mammitzsch (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society, B*, **47**, 238–252.

Green, P., C. Jennison and A. Seheult (1985). Analysis of fields experiments by least squares smoothing. *Journal of the Royal Statistical Society, B*, **47**, 299–315.

Györfi, L., W. Härdle, P. Sarda and P. Vieu (1990). *Nonparametric Curve Estimation from Time Series.* Lecture Notes in Statistics, **60.** Springer, New York.

Härdle, W. (1993). *Applied Nonparametric Regression.* Cambridge University Press.

Hart, J. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society, B*, **53**, 173–187.

Heckman, N. (1986). Spline smoothing in partly linear model. *Journal of the Royal Statistical Society, B*, **48**, 244–248.

Herrmann, E., T. Gasser and A. Kneip (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, **79**, 783 95.

Judge, G., W. Griffiths, R. Carter, H. Lütkepohl and T.C. Lee (1985). *The Theory and Practice of Econometrics.* John Wiley and Sons, New York.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and their Applications*, **29**, 309–315.

Nadaraya, E.A. (1964). Remarks on nonparametric estimates for density functions and regression curves. *Theory of Probability and Applications*, **15**, 134–137.

Priestley, M.B. and M.T. Chao (1972). Nonparametric function fitting. *Journal of the Royal Statistical Society, B*, **34**, 385–392.

Quintela-del-Río, A. (1996). Comparison of bandwidth selectors in nonparametric regression under dependence. *Computational Statistics and Data Analysis*, **21**, 563 580.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics and Probability Letters*, **4**, 203 208.

Roussas, G., L.T. Tran and D.A. Ioannides (1992). Fixed design regression for time series: Asymptotic normality. *Journal of Multivariate Analysis*, **40**, 262–291.

Schick, A. (1996). Efficient estimation in a semiparametric additive regression model with autoregressive errors. *Stochastic Processes and their Applications*, **61**, 339 361.

Schick, A. (1998). An adaptive estimator of the autocorrelation coefficient in regression models with autoregressive errors. *Journal of the Time Series Analysis*, **19**, 575–589.

Schick, A. (1999). Efficient estimation in a semiparametric additive regression model with ARMA errors. In *Asymptotics, Nonparametrics and Time Series*, 395-425. Marcel Dekker, New York.

Simonoff, J.S. (1996). *Smoothing Methods in Statistics.* Springer, New York.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, B*, **50**, 413–436.

Stone, C. (1977). Consistent nonparametric regression. *Annals of Statistics*, **5**, 595 645.

Volkonskii, V.A. and Y.A. Rozanov (1959). Some limit theorems for random functions I. *Theory of Probability and Applications*, 4 , 178 197.

Watson, G.S. (1964). Smooth regression analysis. *Sankhya, A*, **26**, 359–372.

Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and Applications*, **5**, 302–305.

Wise, J. (1955). The autocorrelation function and the spectral density function. *Biometrika*, **42**, 151–159.