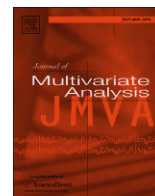




Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: [www.elsevier.com/locate/jmva](http://www.elsevier.com/locate/jmva)

# Functional semiparametric partially linear model with autoregressive errors

Sophie Dabo-Niang<sup>a</sup>, Serge Guillas<sup>b,\*</sup>

<sup>a</sup> Université Charles-De-Gaulle-Lille 3, Lille, France

<sup>b</sup> University College London, London, United Kingdom

## ARTICLE INFO

### Article history:

Available online xxxx

### AMS subject classification:

62G08

### Keywords:

Functional data  
Semiparametric regression  
Autocorrelation

## ABSTRACT

In this paper, we introduce a functional semiparametric model, where a real-valued random variable is explained by the sum of a unknown linear combination of the components of a multivariate random variable and an unknown transformation of a functional random variable. The errors can be autocorrelated. We focus here on the parametric estimation of the coefficients in the linear combination. First, we use a nonparametric kernel method to remove the effect of the functional explanatory variable. Then, we use generalized least squares approach to obtain an estimator of these coefficients. Under some technical assumptions, we prove consistency and asymptotic normality of our estimator. Finally, we present Monte Carlo simulations that illustrate these characteristics.

© 2008 Elsevier Inc. All rights reserved.

## 1. Introduction

Robinson [18] considered the semiparametric regression model  $E(Y|X, Z) = \beta'X + \theta(Z)$ , (a.s.), with independent errors,  $(X, Y, Z)$  in  $\mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^q$ . He established the root- $n$ -consistency of an estimator of  $\beta$ . The technique is based on regressing first  $X$  and  $Y$  on  $Z$  and then estimating  $\beta$  by plugging into its ordinary least squares-based expression the estimates of  $E(X|Z)$  and  $E(Y|Z)$  corresponding to the aforementioned regressions.

Recently, Aneiros and Vieu [2] extended this result to the case where  $Z$  is valued in a semimetric space eventually of infinite dimension. In this context,  $Z$  is called a functional random variable. In a non-functional context, Aneiros and Quintela [1] considered the regression model  $Y_i = \beta'X_i + \theta(Z_i) + \varepsilon_i$ , where  $(X_i, Y_i, Z_i)$  are in  $\mathbb{R}^p \times \mathbb{R} \times [0, 1]$ , and  $\varepsilon_i$  are unobserved dependent errors. They established the root- $n$ -consistency of an estimator of  $\beta$ . In this paper, we extend both of these studies to the case of a random variable  $Z$  that takes its values in a semimetric space  $(E, d)$  which is of infinite dimension [5,9], with autocorrelated errors. Let  $\mu$  be the probability distribution for  $Z$ . Let us denote  $B(z, s)$  the ball whose center is  $z$  in  $(E, d)$ , with radius  $s$ . In this paper,  $\mu(B(z, h_n))$ , for  $z \in (E, d)$ , will play the role of  $h_n^q$  for the kernel-based density estimator when  $E = \mathbb{R}^q$ . Indeed, there is no Lebesgue measure in infinite-dimensional spaces [14] and thus  $\mu(B(z, h_n))$  may depend upon  $z$ . In infinite-dimensional cases, defining densities with respect to other measures than the Lebesgue measure is a strong assumption [15]. We use an extension of the Nadaraya–Watson kernel estimator for the nonparametric part [8, 2]. We must ensure that this estimator is efficient enough to get the optimal rate of convergence (fractal dimension). Our assumptions are therefore slightly more restrictive on the kernel than in the finite-dimensional case.

Regression models for functional data are useful in practice, since observations measured almost continuously are more common, see [9] for a review of methods and examples. In our context, Gervini and Gasser [12] introduced a semiparametric

\* Corresponding address: Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, United Kingdom.  
E-mail address: [serge@stats.ucl.ac.uk](mailto:serge@stats.ucl.ac.uk) (S. Guillas).

model for functional data based on B-splines. They address the root- $m$ -consistency, where  $m$  is the number of observations per individual curve. Here, we examine the root- $n$ -consistency when the number of curves  $n$  goes to infinity. Let  $(\varepsilon_1, \dots, \varepsilon_n)$  be a sequence of real random variable that satisfy

$$\varepsilon_i + \sum_{j=1}^d \rho_j \varepsilon_{i-j} = \zeta_i, \tag{1}$$

where all the roots of the equation  $A(u) = u^d + \sum_{j=1}^d \rho_j u^{d-j} = 0$ , lie inside the unit ball ( $|u| \leq 1$ ), and  $\{\zeta_i\}$  is a centered white noise independent with  $\varepsilon_i$ . The error  $\{\varepsilon_i\}$  is then a weakly stationary AR( $d$ ) process with innovation process  $\{\zeta_i\}$ . Our semiparametric regression model for functional variables is defined as follows:

$$Y_i = X_i' \boldsymbol{\beta} + \theta(Z_i) + \varepsilon_i = \sum_{j=1}^p X_{ij} \beta_j + \theta(Z_i) + \varepsilon_i, \quad i = 1, \dots, n, \tag{2}$$

where  $\theta$  is an unknown function from  $E$  to  $\mathbb{R}$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is an unknown vector in  $\mathbb{R}^p$ ,  $(X_1, Y_1, Z_1), \dots, (X_n, Y_n, Z_n)$  are random variables taking their values in  $\mathbb{R}^p \times \mathbb{R} \times E$  identically distributed as  $(X, Y, Z)$ ,  $(X_i, Z_i)_{i=1, \dots, n}$  is an independent sequence of random variables independent of the sequence  $\{\varepsilon_i\}_{i=1, \dots, n}$ .

Let  $X_i = (X_{i1}, \dots, X_{ip})'$ ,  $i = 1, \dots, n$ ,  $X^j = (X_{1j}, \dots, X_{nj})'$ ,  $j = 1, \dots, p$ ,  $\mathbf{X} = [X^1, \dots, X^p]$ ,  $\mathbf{Y} = (Y_1, \dots, Y_n)$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$  ( $A'$  denotes the transpose of a matrix  $A$ ). The variance of  $\boldsymbol{\varepsilon}$  is  $\text{Var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \Phi_n$ , where  $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_i)$  and  $\Phi_n$  is a  $n \times n$  definite positive matrix different to the identity matrix  $\mathbf{I}$ . Since  $\Phi_n$  is positive definite, there exists a  $n \times n$  matrix  $\mathbf{P}$  that satisfies  $\mathbf{P}\Phi_n\mathbf{P}' = \mathbf{I}$ ,  $\mathbf{P}'\mathbf{P} = \Phi_n^{-1}$ . This matrix  $\mathbf{P}$  is not unique [13]. Under (1), the inverse of the covariance matrix  $\Phi_n$  is [22]

$$\Phi_n^{-1} = (\mathbf{I} + \rho_1 \mathbf{U}' + \dots + \rho_d \mathbf{U}^d)(\mathbf{I} + \rho_1 \mathbf{U} + \dots + \rho_d \mathbf{U}^d), \tag{3}$$

where

$$\mathbf{U} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

is the auxiliary  $n \times n$  identity matrix. If the order of the autoregression is  $d = 1$ , then

$$\Phi_n^{-1} = \begin{bmatrix} 1 & \rho_1 & 0 & \dots & 0 & 0 \\ \rho_1 & 1 + \rho_1^2 & \rho_1 & \dots & 0 & 0 \\ 0 & \rho_1 & 1 + \rho_1^2 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & 1 + \rho_1^2 & \rho_1 \\ 0 & 0 & 0 & \dots & \rho_1 & 1 \end{bmatrix}.$$

Let

$$g_j(z) = E[X_{ij}|Z_i = z], \quad \eta_{ij} = X_{ij} - E[X_{ij}|Z_i], \quad i = 1, \dots, n; \quad j = 1, \dots, p \tag{4}$$

and

$$\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{ip})', \quad \boldsymbol{\eta}^j = (\eta_{1j}, \dots, \eta_{nj})',$$

where  $(\eta_{11}, \dots, \eta_{1p})$  are centered independent and identically distributed random variables of variance/covariance  $p \times p$  matrix  $\boldsymbol{\Sigma}_\eta = (\sum_{ij})_{1 \leq i \leq p; 1 \leq j \leq p}$ . Since  $(X_i, Z_i)$  and  $\varepsilon_i$  are independent, so are  $\boldsymbol{\eta}_{ij}$  and  $\varepsilon_i$ . Let:

$$W_{ni}(z) = \frac{K(d(Z_i, z)/h_n)}{\sum_{k=1}^n K(d(Z_k, z)/h_n)}, \quad \mathbf{W} = \{W_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq n},$$

where  $K$  is a function over  $[0, +\infty[$  called kernel,  $h_n > 0$  is the bandwidth parameter and  $W_{ij} = W_{ni}(Z_j)$ . Let

$$\begin{aligned} \tilde{\mathbf{X}} &= (\mathbf{I} - \mathbf{W})\mathbf{X} = (\tilde{X}^1, \dots, \tilde{X}^p), & \tilde{\mathbf{Y}} &= (\mathbf{I} - \mathbf{W})\mathbf{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_n), \\ \tilde{\boldsymbol{\theta}} &= (\tilde{\theta}(Z_1), \dots, \tilde{\theta}(Z_n))', & \tilde{\boldsymbol{\varepsilon}} &= (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n)', \\ \tilde{X}^j &= (\tilde{X}_{1j}, \dots, \tilde{X}_{nj})', & \tilde{X}_i &= (\tilde{X}_{i1}, \dots, \tilde{X}_{ip})', \end{aligned}$$

where

$$\begin{aligned} \tilde{\theta}(Z_i) &= \theta(Z_i) - \sum_{j=1}^n W_{ni}(Z_j)\theta(Z_j), & \tilde{X}_{ij} &= X_{ij} - \sum_{l=1}^n W_{ni}(Z_l)X_{lj}, \\ \tilde{\varepsilon}_i &= \varepsilon_i - \sum_{j=1}^n W_{ni}(Z_j)\varepsilon_j, & \tilde{Y}_i &= Y_i - \sum_{j=1}^n W_{ni}(Z_j)Y_j. \end{aligned}$$

To start, we assume that the covariance matrix  $\Phi_n$  is known and that  $\mathbf{P}\tilde{\mathbf{X}}$  has full rank. Then, a generalized least squares method can be used, see for example [13]. We minimize  $\|\mathbf{P}(\mathbf{I}-\mathbf{W})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\|_2^2$ , ( $\|\mathbf{u}\|_2$  denotes the Euclidean norm for  $\mathbf{u} \in \mathbb{R}^n$ ) to obtain the following estimate of  $\boldsymbol{\beta}$ ,

$$\hat{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\mathbf{Y}}.$$

Since the function  $\theta$  is unknown, we proceed to estimate it by using  $\hat{\boldsymbol{\beta}}$

$$\hat{\theta}_n(z) = \sum_{i=1}^n W_{ni}(z)(Y_i - X_i' \hat{\boldsymbol{\beta}}), \quad z \in E.$$

In what follows, we assume that the noise  $\{\zeta_j\}$  in the autoregressive errors has a density with respect to Lebesgue measure on  $\mathbb{R}$ . So,  $\{\varepsilon_j\}$  is strongly mixing with mixing coefficient  $\alpha(t) = O(\kappa^t)$ , ( $0 < \kappa < 1$ ) (see [16]). We have also

$$c(j) = \text{Cov}(\varepsilon_i, \varepsilon_{i+j}) = O(\kappa^j), \quad 0 < \kappa < 1, \tag{5}$$

(see [3, exercise 3.11]).

Therefore (see [1, Remark 2.2, p. 341 and (3)])

$$\sum_{i=1}^{\infty} \alpha(i)^{\frac{\delta}{2+\delta}} < \infty, \quad \text{for some } \delta > 0; \quad \sum_{i=1}^{\infty} i|c(i)| < \infty; \tag{6}$$

$$\|\Phi_n\|_2 = O(1); \quad \|\Phi_n^{-1}\|_{\infty} = O(1), \tag{7}$$

where

$$\|\mathbf{A}\|_p = \max_{\|\mathbf{u}\|_p > 0} \frac{\|\mathbf{A}\mathbf{u}\|_p}{\|\mathbf{u}\|_p}, \quad \mathbf{u} \in \mathbb{R}^n,$$

is the  $L_p$  norm of the matrix  $\mathbf{A}$ .

The above conditions on  $\eta$  and (3) entail (see page 341, Remarks 2.2–2.3 of [1]):

$$n^{-1} \eta' \Phi_n^{-1} \eta \rightarrow \mathbf{B} = (B_{ij}) = (\sigma_{\varepsilon}^2 / \sigma_{\zeta}^2) \left(1 + \sum_{j=1}^d \rho_j^2\right) \Sigma_{\eta}, \quad \text{in probability,} \tag{8}$$

where  $\mathbf{B}$  is a positive definite  $p \times p$  matrix and  $\sigma_{\zeta}^2 = \text{Var}(\zeta_i)$ .

## 2. Consistency

Let  $r > 0$  and  $\mathcal{B}(z, h)$  be the open ball centered at  $z \in E$  and of radius  $h > 0$ . Let  $\mathcal{C}$  be a given subset of  $E$  such that  $\mathcal{C} \subset \cup_{k=1}^{d_n} \mathcal{B}(z_k, l_n)$ , where  $d_n l_n^{\gamma} = C$ ,  $d_n \rightarrow \infty$  and  $l_n \rightarrow 0$  as  $n \rightarrow \infty$  ( $\gamma$  and  $C$  are real positive constants,  $z_k \in E$ ,  $l_n > 0$ ). We assume that  $Z$  takes its values in  $\mathcal{C}$ . We also assume that the components  $X_{ij}$  ( $i = 1, \dots, n, j = 1, \dots, p$ ) of  $\mathbf{X}$  are bounded. We consider the following assumptions:

- (H1)  $E|\varepsilon_i|^{2+\delta} < \infty$  for some  $\delta > 0$ .
- (H2)  $K$  is Lipschitzian over  $[0, \infty[$ , of support  $[0, r]$  and there exists a positive constant  $C$  such that:  $\forall u \in [0, r], -K'(u) \geq C$ .
- (H3)  $\forall (u, v) \in \mathcal{C} \times \mathcal{C}, \forall f \in \{\theta, g_1, \dots, g_p\}, |f(u) - f(v)| \leq (d(u, v))^{\alpha}, C > 0, \alpha > 0$ .
- (H4)  $E|\eta_{11}|^s + \dots + E|\eta_{1p}|^s < \infty, s \geq 3$ .

The rates of convergence will be achieved under some assumptions about the small ball probabilities of the variable  $Z$ :

(H5) there exist positive constants  $c_0, c_1$  and a positive valued function  $\psi$  on  $]0, \infty[$  such that

$$\int_0^1 \psi(ht)dt > c_0 \psi(h), \quad \text{and} \quad c_0 \psi(h) \leq P(Z_1 \in \mathcal{B}(z, h)) \leq c_1 \psi(h), \quad \forall z \in E, h > 0.$$

In addition, we assume that the smoothing parameter  $h_n$  satisfies

(H6)

$$\lim_{n \rightarrow +\infty} h_n = 0, \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{n\psi(h_n)}{\log n} = +\infty;$$

**Theorem 1.** Assume (H1)–(H6), (1) and (2) hold. Then

$$\Phi_n \text{ is positive definite} \tag{9}$$

is necessary and sufficient for

$$n^{1/2}(\hat{\beta} - \beta) \rightarrow \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{B}^{-1}) \text{ in distribution.} \tag{10}$$

**Proof.** We have

$$n^{1/2}(\hat{\beta} - \beta) = \left( n^{-1} \tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\mathbf{X}} \right)^{-1} n^{-1/2} \left[ \tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\varepsilon} + \tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\theta} \right].$$

The following lemmas yield the proof. □

**Lemma 2.** Under (H2), (H4), (H5), (H6), we have

$$n^{-1} \tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\mathbf{X}} \rightarrow \mathbf{B}, \text{ almost surely,}$$

if (H3) is satisfied for only  $g_1, \dots, g_p$ .

**Lemma 3.** Under (H1)–(H6), we have

$$n^{-1/2} \tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\varepsilon} \rightarrow \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{B}^{-1}) \text{ in distribution.}$$

**Lemma 4.** Under (H2)–(H6), we have

$$n^{-1/2} \tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\theta} \rightarrow 0, \text{ almost surely.}$$

To avoid interrupting the discussion, the proofs of the lemmas above are given in the appendix.

We assume in the following that the matrix  $\Phi_n$  is unknown. We need to have the exact form of  $\Phi_n$  and an estimate of this matrix. As [1] pointed out,  $\Phi_n$  has  $((n(n+1)/2) - 1)$  different unknown parameters, but it is usual to assume that the elements in  $\Phi_n$  are functions of a  $k \times 1$  vector  $\phi$  ( $k < n$  and remains constant as  $n$  increases). Then [13] the estimation of  $\Phi_n(\phi)$  reduces to the estimation of  $\phi$ .

Let  $\hat{\Phi}_n$  be an estimator of  $\Phi_n$  and

$$\hat{\beta} = \left( \tilde{\mathbf{X}}' \hat{\Phi}_n^{-1} \tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}' \hat{\Phi}_n^{-1} \tilde{\mathbf{Y}}.$$

Let us show in the following theorem that  $\hat{\beta}$  has the same asymptotic distribution as  $\hat{\beta}$ .

**Theorem 5.** Under the conditions of Theorem 1, we have

$$n^{1/2}(\hat{\beta} - \beta) \rightarrow \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{B}^{-1}) \text{ in distribution.} \tag{11}$$

**Proof.** The proof is similar to the proof of Theorem 2.5 of [1] (which is a modification of [10, Theorem 3]), so we omit some details. Let us replace in  $\hat{\beta}$  the matrix  $\Phi_n^{-1}$  by  $(\sigma_\varepsilon^2/\sigma_e) \Phi_n^{-1}$  and notice that the proof is based on the fact that :

(i) the elements of  $\Phi_n$  are functions of the  $k \times 1$  vector of parameters  $\Phi_n$  such that the elements of the  $k$  matrices

$$\mathbf{M}_{nt}(\phi) = \frac{\partial}{\partial \phi_t} \Phi_n^{-1}(\phi), \quad t = 1, \dots, k,$$

are continuous functions of  $\phi$  in an open sphere  $\mathbf{S}$  of the true value  $\phi^0$  of the parameter  $\phi$  [22].

(ii) The sequences of matrices  $\{\tilde{\mathbf{X}} = \tilde{\mathbf{X}}_n\}$  and  $\{\Phi_n\}$  are such that

$$\lim_{n \rightarrow \infty} n^{-1} \tilde{\mathbf{X}}' \mathbf{M}_{nt}(\phi) \tilde{\mathbf{X}} = \mathbf{N}_t(\phi),$$

where  $\mathbf{N}_t(\phi) = 2\phi_t \sum_{\eta} \eta$  is a  $p \times p$  matrix of continuous functions of  $\phi$ ,  $t = 1, \dots, k$ , and

$$n^{-1} \tilde{\mathbf{X}}' \mathbf{M}_{nt}(\phi) (\tilde{\theta} + \tilde{\varepsilon}) = O_p(n^{-1/2}),$$

because of conditions on  $\eta$ ,  $K$  and  $g_i$ .

(iii) There exist estimators  $\widehat{\phi}$  of  $\phi$  and  $\widehat{\Phi}_n = \Phi_n(\widehat{\phi})$  of  $\Phi_n = \Phi_n(\phi^0)$  such that  $\widehat{\Phi}_n^{-1}(\widehat{\phi})$  exists for all  $n$  and  $\widehat{\phi}$ . The estimator  $\widehat{\phi}$  is based on  $\widehat{\varepsilon}_i = y_i - \mathbf{X}'_i \bar{\beta} - \bar{\theta}(Z_i)$  (see [3, Chap. 8], for estimation methods of  $\phi^0$  using  $\{\varepsilon_i\}$ ), where  $\bar{\beta} = (\widetilde{\mathbf{X}}' \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}' \widetilde{\mathbf{Y}}$  and  $\bar{\theta}(Z_i) = \sum_{j=1}^n W_{ni}(Z_j) \theta(Z_j) (Y_j - \mathbf{X}'_j \bar{\beta})$ .

Using the same arguments as those of Lemmas 2 and 4, we have that  $\|\bar{\beta} - \beta\|_2 = o_p(1)$ ,  $\sup_i |\bar{\theta}(Z_i) - \theta(Z_i)| = o_p(1)$  and  $\sup_i |\mathbf{X}'_i \bar{\beta} + \bar{\theta}(Z_i) - \mathbf{X}'_i \beta - \theta(Z_i)| = o_p(1)$ . This and Theorem 1 of [11] (see also [1, Remark 2.7]), imply that

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

Then the proof of the theorem follows from (i), (ii), (iii) and Theorem 1 because:

$$\begin{aligned} \widehat{\beta} - \beta &= (\widetilde{\mathbf{X}}' \Phi_n^{-1}(\widehat{\phi}) \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}' \Phi_n^{-1}(\widehat{\phi}) (\tilde{\theta} + \tilde{\varepsilon}) \\ &= (n^{-1} \widetilde{\mathbf{X}}' \Phi_n^{-1}(\phi^0) \widetilde{\mathbf{X}})^{-1} (n^{-1} \widetilde{\mathbf{X}}' \Phi_n^{-1}(\phi^0) (\tilde{\theta} + \tilde{\varepsilon})) \\ &\quad + \sum_{t=1}^k \left\{ (n^{-1} \widetilde{\mathbf{X}}' \Phi_n^{-1}(\phi^*) \widetilde{\mathbf{X}})^{-1} (n^{-1} \widetilde{\mathbf{X}}' \mathbf{M}_{nt}(\phi^*) (\tilde{\theta} + \tilde{\varepsilon})) \right\} (\widehat{\phi}_t - \phi_t^0) \\ &\quad - \sum_{t=1}^k \left\{ (n^{-1} \widetilde{\mathbf{X}}' \Phi_n^{-1}(\phi^*) \widetilde{\mathbf{X}})^{-1} (n^{-1} \widetilde{\mathbf{X}}' \mathbf{M}_{nt}(\phi^*) \widetilde{\mathbf{X}}) \right\} \\ &\quad \times \left\{ (n^{-1} \widetilde{\mathbf{X}}' \Phi_n^{-1}(\phi^*) \widetilde{\mathbf{X}})^{-1} (n^{-1} \widetilde{\mathbf{X}}' \Phi_n^{-1}(\phi^*) (\tilde{\theta} + \tilde{\varepsilon})) \right\} \times (\widehat{\phi} - \phi^0), \end{aligned}$$

where  $\phi^*$  is between  $\widehat{\phi}$  and  $\phi^0$ .  $\square$

It is well known that the performance of the kernel estimate depends on the choice of the window parameter  $h$ . The bound in (A.1) is simple and easy to compute. So, this allows us to choose the window parameter  $h$  that minimizes that bound. We can also use different methods like cross-validation. For example, one can consider the bandwidth that minimizes a cross-validation criterion: Rachdi and Vieu [17] proved (under some additional condition about the concentration of  $X$ ) that the cross-validation bandwidth is asymptotically optimal with respect to the average square error or the mean integrated square error. Such result can be extended to our setting but is beyond the scope of this paper and deserves future investigation.

In practice, for regression models with autoregressive errors, the unknown matrix  $\Phi_n$  (or its inverse) is not directly estimated, as it can be numerically unstable or difficult to compute. In a non-functional context, a more stable iterative procedure has been introduced [4,21]. We adapt the algorithm presented in [21, section 2.12] to the functional data context. It consists of repeatedly applying the lag transformation inferred from the structure of the residuals of a regression with no autocorrelation to the original regression equation. This eventually yields an uncorrelated regression equation. It proceeds as follows in the AR(1) case:

- (i) Run the semiparametric regression of model (2) under the assumption that the autocorrelation in the error term is zero. Retain the residuals  $\widehat{\varepsilon}_i = Y_i - \widehat{\beta} X_i - \widehat{\theta}(Z_i)$ .
- (ii) Fit an AR(1) model to the residuals, say:

$$\widehat{\rho}(B) \widehat{\varepsilon}_i = \zeta_i \tag{12}$$

where  $B$  is the lag operator, with  $\widehat{\rho}(B) \widehat{\varepsilon}_i = \widehat{\varepsilon}_i + \widehat{\rho} \widehat{\varepsilon}_{i-1}$ .

- (iii) Apply the transformation to both sides of (2). Let  $u_i = \widehat{\rho}(B) Y_i$ ,  $v_i = \widehat{\rho}(B) X_i$  and  $w_i = \widehat{\rho}(B) Z_i$ .
- (iv) Build the transformed regression model

$$u_i = v_i' \beta + \theta(w_i) + \eta_i, \tag{13}$$

where the noise  $\eta_i$  is assumed uncorrelated. Obtain the estimate of  $\beta$ . Retain the residuals. Go to step (ii).

- (v) Stop if convergence is reached (say by small value of autocorrelation) or the number of iterations is large.

Sargan [20] discussed the convergence of this type of algorithm in the parametric case. The algorithm will converge to a local maximum in many cases, but may converge to a saddle point. The study of convergence of the functional semiparametric algorithm above is beyond the scope of this paper, and deserves investigation.

### 3. Simulation results

We simulate  $n$  random curves as a Wiener noise [6]. We smooth these curves with a standard Tukey's smoother to get more regular curves  $Z_1, \dots, Z_n$ . For each curve, we keep track of 32 equally spaced values over  $[0, 1]$ . We define the operators

$$\theta_1(z) = 300 \int_0^1 \sin(2\pi x) z(x) dx$$

**Table 1**  
Estimates of parameter  $\beta = (\beta_1, \beta_2, \beta_3) = (1, 2, -3)$ , case  $\theta_1$

n	$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
	Quartile	Median	Quartile	Quartile	Median	Quartile	Quartile	Median	Quartile
Autocorrelation 0.2									
20	0.09	1.09	2.18	1.18	2.02	2.77	-4.13	-3.00	-1.85
50	0.46	0.91	1.63	1.60	2.02	2.50	-3.63	-3.01	-2.46
200	0.82	1.04	1.25	1.87	2.03	2.15	-3.22	-3.03	-2.81
Autocorrelation 0.6									
20	0.13	1.20	1.82	1.34	1.98	2.62	-3.95	-3.04	-2.11
50	0.32	0.97	1.45	1.50	1.83	2.26	-3.48	-2.88	-2.37
200	0.80	1.03	1.22	1.86	2.00	2.14	-3.19	-3.05	-2.85
Autocorrelation 0.95									
20	0.18	0.97	1.78	1.38	2.04	2.78	-3.90	-3.10	-1.92
50	0.28	1.02	1.58	1.62	1.96	2.38	-3.48	-2.97	-2.35
200	0.73	1.02	1.24	1.81	1.95	2.12	-3.24	-2.97	-2.71

Monte Carlo study with 100 seeds and sample sizes  $n = 20, 50, 200$ . The empirical lower quartile, median and upper quartile are reported.

**Table 2**  
Estimates of parameter  $\beta = (\beta_1, \beta_2, \beta_3) = (1, 2, -3)$ , case  $\theta_2$

n	$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
	Quartile	Median	Quartile	Quartile	Median	Quartile	Quartile	Median	Quartile
Autocorrelation 0.2									
20	-8.77	0.84	7.57	-3.69	0.46	8.03	-10.56	-3.34	7.25
50	-2.70	0.73	5.48	-0.79	1.96	5.23	-6.78	-3.65	-0.03
200	-1.11	0.96	2.56	0.28	1.95	3.25	-4.52	-2.72	-0.47
Autocorrelation 0.6									
20	-5.64	-1.01	6.36	-1.44	1.61	5.96	-7.98	-2.10	5.05
50	-2.78	2.24	5.39	-0.80	1.93	4.51	-7.63	-3.33	1.11
200	-1.19	0.78	2.79	0.40	1.95	3.32	-5.10	-2.83	-0.72
Autocorrelation 0.95									
20	-7.40	-1.43	7.66	-5.27	0.23	5.37	-7.27	1.80	6.43
50	-4.58	-0.51	4.50	-1.18	1.32	3.92	-6.10	-2.65	1.65
200	-0.74	1.08	3.58	0.88	2.05	3.77	-5.68	-3.36	-1.82

Monte Carlo study with 100 seeds and sample sizes  $n = 20, 50, 200$ . The empirical lower quartile, median and upper quartile are reported.

and

$$\theta_2(z) = 300 \int_0^1 x^3 z(x) dx.$$

We simulate an AR(1) sequence  $\varepsilon_1, \dots, \varepsilon_n$ , with autocorrelation  $\rho_1$  taking values 0.2, 0.6 and 0.95 respectively and a corresponding Gaussian centered noise with variance 1. Our regression model is  $Y_i = \beta'X_i + \theta(Z_i) + \varepsilon_i$  with signal-to-noise ratio from  $\theta_1(Z_i)$  and  $\beta'X_i$  each around 2, when compared to  $\varepsilon_i$ , but larger signal ratio for  $\theta_2(Z_i)$ , when compared to  $\varepsilon_i$ . The covariance matrix of  $X$  is

$$\begin{pmatrix} 8 & -2 & 6 \\ -2 & 10 & 4 \\ 6 & 4 & 9 \end{pmatrix}.$$

We carry out a Monte Carlo study with 100 different seeds, and sample size 20, 50 and 200. The bandwidth was selected to give significant ranges in the kernel matrix based on the semimetric.

The empirical lower and upper quartiles, and the median of the estimates of each component of  $\beta = (1, 2, -3)$  are reported in Tables 1 and 2. Our estimates show consistence over these sequence of sample sizes, with good results obtained for a sample size of 200. It appears that the level of noise autocorrelation has little effect on the estimation. Only 3 iterations of the algorithm in the last section were enough to ensure convergence. However, the case where  $\theta_2$  is used shows estimation results that are deteriorated compared to the case where  $\theta_1$  is used. It is due to a larger signal-to-noise ratio for the nonparametric part since the function  $\sin(2\pi x)$  does some averaging over  $[0, 1]$ , whereas  $x^3$  does not.

To assess asymptotic normality, we computed estimates of  $\beta$  for 100 different seeds, with the autocorrelation level 0.95 and sample size 200. Fig. 1 shows the Q-Q plots for these estimates. These plots demonstrate that our estimates are close to normal, for finite but large sample size.

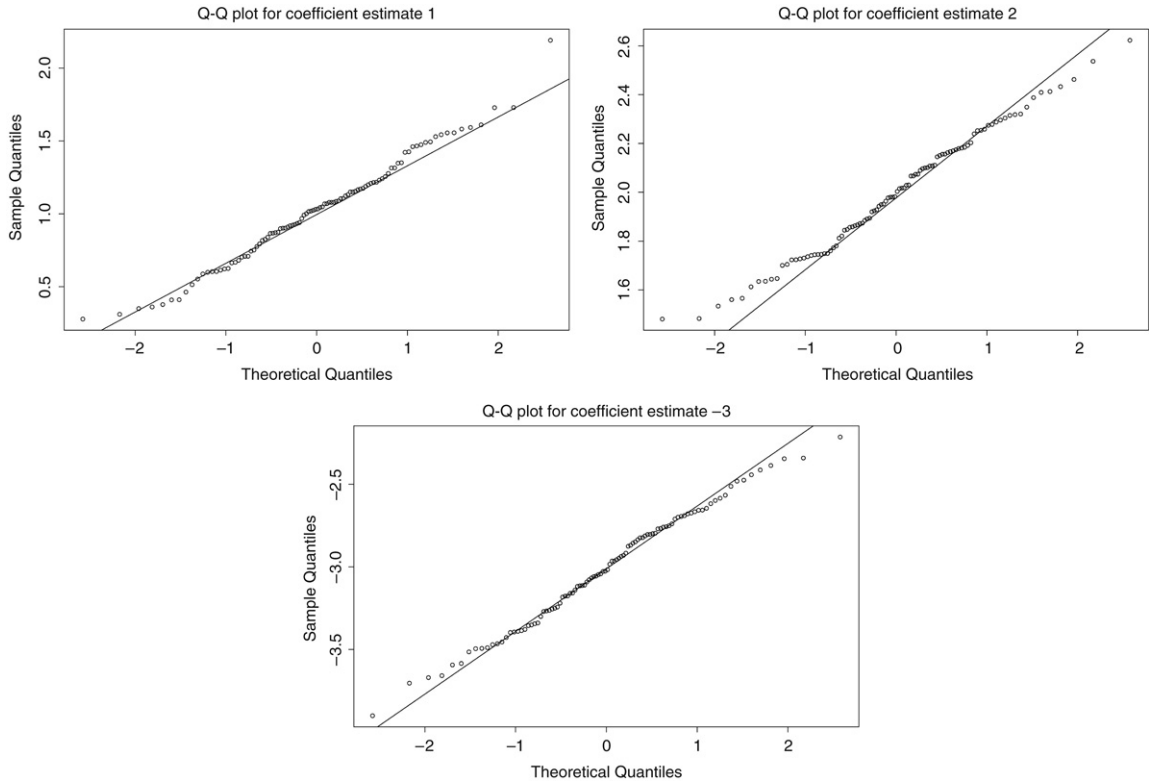


Fig. 1. Q-Q plots for sampled estimates of parameter  $\beta = (\beta_1, \beta_2, \beta_3) = (1, 2, -3)$ , over 100 Monte Carlo experiments, case  $\theta_1$ . Sample size 200 and autocorrelation level 0.95.

**Acknowledgments**

The authors thanks the referees for clarifications, simplifications and references that helped improve the quality of the paper.

**Appendix. Technical derivations**

**Proof of Lemma 2.** Let

$$\bar{g}_j(t) = g_j(t) - \hat{g}_j(t), \quad \bar{g}_j = (\bar{g}_j(Z_1), \dots, \bar{g}_j(Z_n))',$$

where  $\hat{g}_j(t) = \sum_{i=1}^n W_{ni}(t)X_{ij}$ . Then we have :  $\tilde{X}_{ij} = \tilde{\eta}_{ij} + \tilde{g}_j(Z_i) = \bar{g}_j(Z_i) + \eta_{ij}$ , and the  $(i, j)$  element of  $\tilde{\mathbf{X}}' \Phi_n^{-1} \tilde{\mathbf{X}}$  is:

$$\tilde{X}_i' \Phi_n^{-1} \tilde{X}_j = \bar{g}_i' \Phi_n^{-1} \bar{g}_j + (\eta^i)' \Phi_n^{-1} \bar{g}_j + \bar{g}_i' \Phi_n^{-1} \eta^j + (\eta^i)' \Phi_n^{-1} \eta^j.$$

We have that  $n^{-1}(\eta^i)' \Phi_n^{-1} \eta^j \rightarrow \mathbf{B}_{ij}$ , in probability, by (8). Under the hypotheses (H2)–(H6), [9] proved that

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq n} |\bar{g}_j(Z_i)| = O(h_n^\alpha) + o\left(\sqrt{\frac{\log n}{n\psi(h_n)}}\right), \text{ almost surely.} \tag{A.1}$$

In other hand, we get by Cauchy–Schwartz inequality and  $\|\Phi_n^{-1}\|_\infty = O(1)$ :

$$n^{-1} \bar{g}_i' \Phi_n^{-1} \eta^j \leq C n^{-1} \sum_{k=1}^n \bar{g}_i(Z_k) \eta_{kj} \leq C \left( n^{-1} \sum_{k=1}^n \bar{g}_i^2(Z_k) \right)^{1/2} \left( n^{-1} \sum_{k=1}^n \eta_{kj}^2 \right)^{1/2},$$

$$n^{-1} \bar{g}_i' \Phi_n^{-1} \bar{g}_j \leq C \left( n^{-1} \sum_{k=1}^n \bar{g}_i^2(Z_k) \right)^{1/2} \left( n^{-1} \sum_{k=1}^n \bar{g}_j^2(Z_k) \right)^{1/2}.$$

We deduce from (A.1) that

$$\left( n^{-1} \sum_{k=1}^n \bar{g}_i^2(Z_k) \right)^{1/2} = O(h_n^\alpha) + o\left(\sqrt{\frac{\log n}{n\psi(h_n)}}\right), \text{ almost surely.}$$

The strong law of large numbers and the hypotheses on  $\eta$  allow to have that

$$\left( n^{-1} \sum_{k=1}^n \eta_{kj}^2 \right) \rightarrow E(\eta_{11}^2) < \infty, \quad \text{almost surely.}$$

Then we have that  $n^{-1} \bar{g}_i' \Phi_n^{-1} \eta^j \rightarrow 0$  almost surely. We prove in the same way that  $n^{-1} (\eta^i)' \Phi_n^{-1} \bar{g}_j$ ,  $(\eta^i)' \Phi_n^{-1} \eta^j$  and  $n^{-1} \bar{g}_i' \Phi_n^{-1} \bar{g}_j$  converge to zero almost surely. This finishes the proof.  $\square$

**Proof of Lemma 3.** Let  $C'_n = (C_{1n}, \dots, C_{nn}) = \mathbf{a}' \tilde{\mathbf{X}} \Phi_n^{-1} (\mathbf{I} - \mathbf{W})$ , where  $\mathbf{a}$  is a  $p \times 1$  vector. We have  $\|C_n\|_\infty < \infty$  because of the fact that  $\|\mathbf{W}\|_\infty < \infty$ ,  $\|\Phi_n^{-1}\|_\infty < \infty$  and  $\|\mathbf{X}\|_\infty < \infty$ . We have :

$$n^{-1/2} \mathbf{a}' \tilde{\mathbf{X}} \Phi_n^{-1} \tilde{\varepsilon} = n^{-1/2} C'_n \mathbf{e} = \sum_{i=1}^n \frac{C_{in} \varepsilon_i}{n^{1/2}}.$$

Let  $[x]$  denotes the integer part of a real  $x$ , and  $\sigma_n^2 = \text{Var}(\sum_{i=1}^n C_{in} \varepsilon_i)$ . To prove the lemma, we use the well-known technique of big and small blocks. We have to check that:  $S_n^1 \rightarrow \mathcal{N}(0, 1)$ ,  $S_n^2 \rightarrow 0$  and  $S_n^3 \rightarrow 0$  in probability, because  $S_n = \sum_{i=1}^n \left( \frac{C_{in}}{\sigma_n} \right) \varepsilon_i = \sum_{j=0}^{k-1} A_j + \sum_{j=0}^{k-1} \Gamma_j + \Upsilon_k = S_n^1 + S_n^2 + S_n^3$ , where

$$A_j = \sum_{i=(p_n+q_n)j+1}^{(p_n+q_n)j+p_n} Z_{in}, \quad \Gamma_j = \sum_{i=(p_n+q_n)j+p_n+1}^{(p_n+q_n)(j+1)} Z_{in}, \quad \Upsilon_k = \sum_{i=k(p_n+q_n)+1}^n Z_{in},$$

$$Z_{in} = \left( \frac{C_{in}}{\sigma_n} \right) \varepsilon_i, \quad k = k(n) = \left\lfloor \frac{n}{p_n + q_n} \right\rfloor, \quad p_n + q_n \leq n, \quad \text{for } n \text{ great enough.}$$

Let us first check the convergence of  $S_n^2$ . We have by Lemma 2 that

$$n^{-1} \sigma_n^2 \rightarrow \sigma_\varepsilon^2 \mathbf{a}' \mathbf{B} \mathbf{a}, \quad \text{then } \|C_n\|_\infty < \infty \text{ implies that } \max_{1 \leq i \leq n} \frac{C_{in}^2}{\sigma_n^2} = O(n^{-1})$$

and  $\max_{1 \leq i \leq n} \text{Var}(Z_{in}) = O(n^{-1})$ . By [7] inequality, we get

$$\sum_{i=1}^{q_n} \sum_{l=1}^{q_n} \text{Cov}(Z_{in}, Z_{ln}) \leq C n^{-1} \sum_{i=0}^{\infty} \alpha (|i|)^{\frac{\delta}{2+\delta}}.$$

Then we have

$$\text{Var}(\Gamma_j) = q_n \text{Var}(Z_{1n}) + \sum_{i=1}^{q_n} \sum_{l=1}^{q_n} \text{Cov}(Z_{in}, Z_{ln}) = O(n^{-1} q_n).$$

If we consider,  $a = (1 - b)(2 + \delta)/\delta - 1 > 0$  where  $0 < b < 1 - \delta/(2 + \delta)$ , we deduce from (6), the fact that  $\alpha(n)$  decreases to zero, that

$$n^{1+a} \alpha(n) \rightarrow 0. \tag{A.2}$$

Then let  $p_n = [n^{1-c(1+a)}]$  and  $q_n = [n^c]$  with  $(2 + 2a)^{-1} < c < (2 + a)^{-1}$ . It is easy to see that  $p_n \rightarrow \infty$ ,  $q_n \rightarrow \infty$ ,  $p_n^{-1} q_n \rightarrow 0$ ,  $n^{-1} p_n^2 \rightarrow 0$ ,

$$\sum_{j=0}^{k-1} \text{Var}(\Gamma_j) = O\left(\left[\frac{q_n}{p_n + q_n}\right]\right) \rightarrow 0, \quad \left| \sum_{i \neq j} \text{Cov}(\Gamma_i, \Gamma_j) \right| \leq n^{-1} p_n^2 \sum_{l \geq p_n} \alpha (|l|)^{\frac{\delta}{2+\delta}}.$$

This last term tends to zero because  $n^{-1} p_n^2 \rightarrow 0$  and  $\sum_{l \geq p_n} \alpha (|l|)^{\frac{\delta}{2+\delta}} < \infty$  (see (6)). Thus  $S_n^2$  tends to zero in probability. Let us focus on  $S_n^3$ . Because of  $\text{Var}(Z_{in}) = O(n^{-1})$ ,  $\text{Cov}(Z_{in}, Z_{jn}) = O(n^{-1} \alpha (|i - j|)^{\frac{\delta}{2+\delta}})$ ,  $\sum_{i > 0} \alpha (|i|)^{\frac{\delta}{2+\delta}} < \infty$ , we have

$$E\left((S_n^3)^2\right) \leq C \left(1 - \frac{1}{n} \left\lfloor \frac{n}{p_n + q_n} \right\rfloor (p_n + q_n)\right),$$

since  $k = \left\lfloor \frac{n}{p_n + q_n} \right\rfloor$ . The fact that  $\frac{1}{n} \left\lfloor \frac{n}{p_n + q_n} \right\rfloor (p_n + q_n) \rightarrow 1$  yields the proof for  $S_n^3$ .

The proof of  $S_n^1 \rightarrow \mathcal{N}(0, 1)$  is similar to that of [19] and will be omitted. It suffices to apply Lindeberg–Feller’s central limit theorem.  $\square$

**Proof of Lemma 4.** [2] proved that under (H2)–(H6):

$$\begin{aligned} \tilde{\mathbf{X}} \tilde{\theta} &= O(nh_n^\alpha + \psi(h_n)^{-1}(\log n)^2) + O(n^{1/2}h_n^\alpha \log n + \psi(h_n)^{-1/2}(\log n)^2) \\ &\quad + O(n^{1/2}h_n^\alpha \psi(h_n)^{-1/2} \log n + \psi(h_n)^{-1}(\log n)^2) = o(n^{1/2}), \quad a.s; \end{aligned}$$

The fact that  $\|\Phi_n^{-1}\|_\infty < \infty$  yields the proof.  $\square$

## References

- [1] G. Aneiros, A. Quintela, Asymptotic properties in partial linear models under dependence, *Test* 10 (2) (2001) 333–355.
- [2] G. Aneiros, P. Vieu, Semi-functional partial linear regression, *Statist. Probab. Lett.* 76 (11) (2006) 1102–1110.
- [3] P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, in: Springer Series in Statistics, second ed., Springer-Verlag, New York, 1991.
- [4] D. Cochran, G.H. Orcutt, Application of least squares regression to relationships containing auto-correlated error terms, *J. Amer. Statist. Assoc.* 44 (245) (1949) 32–61.
- [5] S. Dabo-Niang, N. Rhomari, Kernel regression estimation when the regressor takes its values in a metric space, *C. R. Math.* 336 (2003) 75–80.
- [6] J. Damon, S. Guillas, Estimation and simulation of autoregressive hilbertian processes with exogenous variables, *Statist. Inference Stoch. Process.* 8 (2005) 185–204.
- [7] J.A. Davydov, The convergence of distributions which are generated by stationary random processes, *Teor. Veroyatnost. i Primenen.* 13 (1968) 730–737.
- [8] F. Ferraty, P. Vieu, The functional nonparametric model and application to spectrometric data, *Comput. Stat.* 17 (2002) 545–564.
- [9] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, in: Springer Series in Statistics, Springer-Verlag, London, 2006.
- [10] W.A. Fuller, G.E. Battese, Transformations for estimation of linear models with nested-error structure, *J. Amer. Statist. Assoc.* 68 (1973) 626–632.
- [11] I. García-Jurado, W. González Manteiga, J.M. Prada Sánchez, M. Febrero-Bande, R. Cao, Predicting using Box-Jenkins, nonparametric, and bootstrap techniques, *Technometrics* 37 (3) (1995) 303–310.
- [12] D. Gervini, T. Gasser, Self-modelling warping functions, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 (4) (2004) 959–971.
- [13] G.G. Judge, R.C. Hill, W.E. Griffiths, H. Lütkepohl, T.C. Lee, *Introduction to the Theory and Practice of Econometrics*, second ed., John Wiley & Sons Inc., New York, 1988.
- [14] H.H. Kuo, *Gaussian Measures in Banach Spaces*, Springer-Verlag, Berlin, 1975.
- [15] R.S. Liptser, A.N. Shiryaev, *Statistics of Random Processes. I*, in: *Applications of Mathematics (New York)*, vol. 5, Springer-Verlag, Berlin, 2001.
- [16] A. Mokkadem, Mixing properties of ARMA processes, *Stochastic Process. Appl.* 29 (2) (1988) 309–315.
- [17] M. Rachdi, P. Vieu, Nonparametric regression for functional data: Automatic smoothing parameter selection, *J. Statist. Plann. Inference* 137 (9) (2007) 2784–2801.
- [18] P.M. Robinson, Root-n-consistent semiparametric regression, *Econometrica* 56 (1988) 931–954.
- [19] G.G. Roussas, L.T. Tran, D.A. Ioannides, Fixed design regression for time series: Asymptotic normality, *J. Multivariate Anal.* 40 (2) (1992) 262–291.
- [20] J. Sargan, Wages and prices in the united kingdom: A study in econometric methodology, in: K.F. Wallis, D.F. Hendry (Eds.), *Quantitative Economics and Econometric Analysis*, Basil Blackwell, Oxford, 1984, pp. 275–314.
- [21] R.H. Shumway, D.S. Stoffer, *Time series analysis and its applications*, in: Springer Texts in Statistics, Springer-Verlag, New York, 2000.
- [22] J. Wise, The autocorrelation function and the spectral density function, *Biometrika* 42 (1955) 151–159.