Applications of Mathematics

Hira L. Koul; Donatas Surgailis

Testing a sub-hypothesis in linear regression models with long memory covariates and errors

Applications of Mathematics, Vol. 53 (2008), No. 3, 235–248

Persistent URL: `http://dml.cz/dmlcz/140318`

53 (2008) APPLICATIONS OF MATHEMATICS No. 3, 235–248

TESTING A SUB-HYPOTHESIS IN LINEAR REGRESSION MODELS WITH LONG MEMORY COVARIATES AND ERRORS*

Hira L. Koul, East Lansing, Donatas Surgailis, Vilnius

(Invited)

Abstract. This paper considers the problem of testing a sub-hypothesis in homoscedastic linear regression models when the covariate and error processes form independent long memory moving averages. The asymptotic null distribution of the likelihood ratio type test based on Whittle quadratic forms is shown to be a chi-square distribution. Additionally, the estimators of the slope parameters obtained by minimizing the Whittle dispersion is seen to be $n^{1/2}$-consistent for all values of the long memory parameters of the design and error processes.

Keywords: moving averages, linear regression, Whittle quadratic forms, chi-square

MSC 2010: 62M09, 62M10, 62M99

1. Introduction

A classical problem in statistics is to see whether among a given set of predictor variables a subset of variables is significant or not for predicting a response variable. In the regression set up this is done by first stipulating a linear regression model and then testing for the absence of some of these covariates by testing that the corresponding slope parameters are zero. This is the so called problem of testing a sub-hypothesis in linear regression set up.

More precisely, let n, p and k be known positive integers with $k \leqslant p$. Let $\boldsymbol{X}_{t1}$, $1 \leqslant t \leqslant n$ ($\boldsymbol{X}_{t2}$, $1 \leqslant t \leqslant n$) be $k \times 1$ ($(p-k) \times 1$) random vectors and let Y_t, $1 \leqslant t \leqslant n$ denote the response variables. Consider the regression model where for

* Research of the first author was partly supported by the NSF DMS Grant 0701430. Research of the second author was partly supported by the bilateral France-Lithuania scientific project Gilibert and the Lithuanian State Science and Studies Foundation grant T-15/07.

some $\beta_1 \in \mathbb{R}^k$ and $\beta_2 \in \mathbb{R}^{p-k}$,

$$Y_t = \beta_1' \boldsymbol{X}_{t1} + \beta_2' \boldsymbol{X}_{t2} + \varepsilon_t, \quad t = 1, \ldots, n. \tag{1.1}$$

The problem of interest is to test

$$H_0\colon\ \beta_2 = 0, \quad \text{vs.} \quad H_1\colon\ \beta_2 \neq 0.$$

In the case of independent homoscedastic errors $\{\varepsilon_t\}$ and when the design variables are either non-random or random and i.i.d., this problem has been well studied in literature, cf. [12] and [16] and references therein. A classical testing procedure is the likelihood ratio test when errors are Gaussian or the analysis of variance type tests via the least square theory which are asymptotically valid without the Gaussianity assumption.

The focus of this paper is to investigate the large sample behavior of an analog of the likelihood ratio type test for H_0 when both the error and the covariate processes form long memory moving averages and are independent of each other. A discrete time strictly stationary stochastic process with finite second moment is said to have long memory if its auto-covariances tend to zero in a hyperbolic fashion as the lag increases to infinity. Long memory processes arise in numerous physical and social sciences. See [1], [3], [4], [6] and [7] for more on these processes. Regression models with long memory in design and errors are useful when the long memory in design variables may be not enough to explain the long memory in the response process, cf. [13]. Such models are found useful in economics and finance when observing high frequency data where spot returns are regressed on forward premiums. See e.g. [2], [5], [15], among others. See also [10] where some currency exchange data sets are observed to have long memory.

Let $\mathbb{Z} := \{0, \pm 1, \ldots\}$ and let Θ be an open and relatively compact subset of $\mathbb{R}^d$, $d \geqslant 1$. The process $\{\varepsilon_t, t \in \mathbb{Z}\}$ is said to form a long memory moving average if for some functions α from Θ to $(0,1)$ and c from Θ to $\mathbb{R}$,

$$\varepsilon_t = \sum_{s \in \mathbb{Z}} a(t-s; \vartheta) \zeta_s, \quad t \in \mathbb{Z}, \tag{1.2}$$

where $a(t;\vartheta) \sim c(\vartheta) t^{\alpha(\vartheta)/2-1}$ $(t \to \infty)$ and ζ_j, $j \in \mathbb{Z}$, are standardized i.i.d. random variables having finite fourth moment. Let $\hat{a}(u;\vartheta) = (2\pi)^{-1} \sum_{t \in \mathbb{Z}} a(t;\vartheta) \mathrm{e}^{-\mathbf{i}tu}$, $u \in \Pi := [-\pi, \pi]$ be the Fourier transform of a, where $\mathbf{i} = (-1)^{1/2}$. Then the corresponding spectral density is $f(u;\vartheta) = 2\pi |\hat{a}(u;\vartheta)|^2$, $\vartheta \in \Theta$, $u \in \Pi$. Let

$$\hat{b}(u;\vartheta) = \frac{1}{f(u;\vartheta)}, \ \ u \in \Pi; \quad b_t(\vartheta) = \int_\Pi \mathrm{e}^{\mathbf{i}tu} \hat{b}(u;\vartheta)\,\mathrm{d}u, \ \ t \in \mathbb{Z}, \ \ \vartheta \in \Theta.$$

Let $\boldsymbol{X}'_t := (\boldsymbol{X}'_{t1}, \boldsymbol{X}'_{t2})$, $\beta' := (\beta'_1, \beta'_2)$ and define

$$\Lambda_n(\vartheta, \beta) := \sum_{t,s=1}^{n} b_{t-s}(\vartheta)(Y_t - \boldsymbol{X}'_t\beta)(Y_s - \boldsymbol{X}'_s\beta), \quad \beta \in \mathbb{R}^p,$$
$$(\hat{\theta}_n, \hat{\beta}_n) := \underset{(\vartheta,\beta)\in\Theta\times\mathbb{R}^p}{\operatorname{argmin}} \Lambda_n(\vartheta, \beta),$$
$$\Lambda_n(\vartheta, \beta_1) := \sum_{t,s=1}^{n} b_{t-s}(\vartheta)(Y_t - \beta_1 X'_{t1}\beta)(Y_s - \boldsymbol{X}'_{s1}\beta_1), \quad \beta_1 \in \mathbb{R}^k, \ \vartheta \in \Theta,$$
$$(\hat{\theta}_{n1}, \hat{\beta}_{n1}) := \underset{(\vartheta,\beta_1)\in\Theta\times\mathbb{R}^k}{\operatorname{argmin}} \Lambda_{n1}(\vartheta, \beta_1).$$

The analog of the likelihood ratio test for H_0 would be based on

$$\mathcal{Q}_n := -2[\Lambda_n(\hat{\theta}_n, \hat{\beta}_n) - \Lambda_{n1}(\hat{\theta}_{n1}, \hat{\beta}_{n1})].$$

Strictly speaking, in the case of Gaussian errors the exact likelihood ratio test would have the elements of the inverse of the covariance matrix as weights in the quadratic forms Λ_n's instead of $\{b_{t-s}(\vartheta)\}$. The above quadratic forms are their Whittle approximations. For this reason we shall call the test that rejects H_0 in favor of H_1 when $\mathcal{Q}_n$ is large the Whittle test.

In the next section we give minimal sufficient conditions on the spectral density $f(u;\vartheta)$ and the process $\boldsymbol{X}_t$ under which the null distribution of $\mathcal{Q}_n$ is seen to converge to a chi-square distribution. The class of covariate processes satisfying these conditions includes the short and long memory moving averages. Additionally, we show that the consistency rate for the minimizers $\hat{\theta}_n$ and $\hat{\beta}_n$ is $n^{1/2}$. Some proofs appear in the last section.

2. Assumptions and main results

To proceed we need to describe some additional needed assumptions on the spectral density $f(u;\vartheta)$. In the sequel, θ, β_0 denote the true values of ϑ, β, respectively, and Θ_0 denotes an arbitrarily small neighborhood of θ. For any twice differentiable function $h(\vartheta, \beta)$ let

$$\nabla_\beta h(\vartheta, \beta) = (\partial h(\vartheta, \beta)/\partial \beta_j)_{j=1,\dots,p};$$
$$\nabla^2_{\vartheta\beta} h(\vartheta, \beta) = (\partial^2 h(\vartheta, \beta)/\partial \vartheta_i \partial \beta_j)_{i=1,\dots,d;\, j=1,\dots,p},$$
$$\nabla_\vartheta h(\vartheta, \beta) = (\partial h(\vartheta, \beta)/\partial \vartheta_i)_{i=1,\dots,d};$$
$$\nabla^2_{\beta\beta} h(\vartheta, \beta) = (\partial^2 h(\vartheta, \beta)/\partial \beta_i \partial \beta_j)_{i,j=1,\dots,p}.$$

We write $\nabla^2_{\beta\beta}h(\theta,\beta_0)$ for $\nabla^2_{\beta\beta}h(\vartheta,\beta)|_{\vartheta=\theta,\beta=\beta_0}$, $\nabla^2_{\vartheta\beta}h(\vartheta,\beta_0)$ for $\nabla^2_{\vartheta\beta}h(\vartheta,\beta)|_{\beta=\beta_0}$, etc.

We assume the following conditions on the spectral density $f(u;\vartheta)$ where f^{-1} stands for $1/f$, cf. [9] and [14].

(a.1) $f(u;\vartheta)$, $u\in\Pi$, determines ϑ uniquely and $\int_{-\pi}^{\pi}\log f(u;\vartheta)\,\mathrm{d}u=0$, $\vartheta\in\Theta_0$.

(a.2) $\int_\Pi \log f(u;\vartheta)\,\mathrm{d}u$ is twice differentiable under the sign of integral.

Furthermore, there exist a function α from Θ to $(0,1)$ that is continuous at θ and a constant $C<\infty$ satisfying the following conditions (a.3)–(a.6).

(a.3) $f(u,\vartheta)$ is continuous at all (u,ϑ), $u\neq 0$, $\vartheta\in\Theta_0$, f^{-1} is continuous on $\Pi\times\Theta_0$ and

$$f(u;\vartheta)\leqslant C|u|^{-\alpha(\vartheta)},\quad \forall\,(u,\vartheta)\in\Pi\times\Theta_0.$$

(a.4) $\nabla_\vartheta f^{-1}$ and $\nabla^2_{\vartheta\vartheta}f^{-1}$ are continuous on $\Pi\times\Theta_0$ and

$$|\nabla_\vartheta f^{-1}(u;\vartheta)|\leqslant C|u|^{\alpha(\vartheta)},$$
$$|\nabla^2_{\vartheta\vartheta} f^{-1}(u;\vartheta)|\leqslant C|u|^{\alpha(\vartheta)/2},\quad \forall\,(u,\vartheta)\in\Pi\times\Theta_0.$$

(a.5) $|\nabla_\vartheta(\partial f^{-1}(u;\vartheta)/\partial u)|\leqslant C|u|^{\alpha(\vartheta)-1}$, $\forall\,(u,\vartheta)\in\Pi\times\Theta_0$

(a.6) $f^{-1}(u;\vartheta)\leqslant C|u|^{\alpha(\vartheta)/2}$, $\forall\,(u,\vartheta)\in\Pi\times\Theta_0$.

Using arguments as in [8] and [9] one can see that fractional autoregressive moving average processes satisfy these conditions.

About the design covariates $\boldsymbol{X}_t'=(X_{t,1},X_{t,2},\ldots,X_{t,p})$ we shall assume that

$$X_{t,i}=\sum_{j=1}^{p}\sum_{s\in\mathbb{Z}}B_{ij}(t-s)\xi_{s,j},\quad t\in\mathbb{Z},\qquad\qquad (2.1)$$
$$\sum_{s\in\mathbb{Z}}B^2_{ij}(s)<\infty,\quad i,j=1,2,\ldots,p.$$

We further assume that $\xi_s:=(\xi_{s,1},\ldots,\xi_{s,p})'$, $s\in\mathbb{Z}$ are i.i.d. standardized r.v.'s, independent of ζ_s, $s\in\mathbb{Z}$, implying the independence of designs and errors in (1.1). The above condition (2.1) includes both the short memory (in particular, i.i.d.) and the long memory random designs, but it does not allow for an intercept parameter in (1.1).

Let

$$V(\vartheta):=\int_\Pi \mathbf{g}(u)f^{-1}(u;\vartheta)\,\mathrm{d}u,\quad W(\vartheta):=\int_\Pi f(u;\vartheta)\nabla^2_{\vartheta\vartheta}f^{-1}(u;\vartheta)\,\mathrm{d}u,$$

where $\mathbf{g}(u)$ is the matrix-valued spectral density of $\boldsymbol{X}_t$ of order $p\times p$. Write

$$\mathbf{g}(u)=\begin{pmatrix}\mathbf{g}_{11}(u) & \mathbf{g}_{12}(u)\\ \mathbf{g}_{12}(u)' & \mathbf{g}_{22}(u)\end{pmatrix},\qquad V(\vartheta)=\begin{pmatrix}V_{11}(\vartheta) & V_{12}(\vartheta)\\ V_{12}(\vartheta)' & V_{22}(\vartheta)\end{pmatrix},$$

where $\mathbf{g}_{12}(u)$ and $\mathbf{g}_{22}(u)$ are matrix-valued (cross-)spectral densities of $\boldsymbol{X}_{t1}$ and $\boldsymbol{X}_{t2}$ having dimensions $k \times p$ and $(p-k)\times(p-k)$, respectively. We are now ready to state the following theorem.

Theorem 2.1. *Assume the model* (1.1), *where* (a.1)–(a.6), (2.1) *hold and* $V(\theta)$ *is positive definite. Then, under* H_0, $\mathcal{Q}_n \Longrightarrow 8\pi^2\chi^2_{p-k}$.

We remark here that a similar chi-square limit distribution seems to hold also in the case when the intercept parameter μ (unknown mean) is included in the model, provided one does not test for $\mu = 0$ but only for absence of some random zero mean covariates. However, the consistency rate of the Whittle estimate of μ seems to be $n^{(1-\alpha(\theta))/2}$, i.e., much slower than the rate $n^{1/2}$ for the remaining parameters. This case will be dealt with elsewhere when dealing with general deterministic designs.

The proof of the above theorem is facilitated by the following lemma. For any positive integer q, let $\mathcal{N}_q(\mu,\Sigma)$ denote the q-dimensional normal distribution with mean vector μ and the covariance matrix Σ. Let Z_θ and Z_β denote two independent random vectors with Z_θ (Z_β) having $\mathcal{N}_d(0,16\pi^3W(\theta))$ ($\mathcal{N}_p(0,8\pi^3V(\theta))$) distribution. Then $(Z'_\theta, Z'_\beta)'$ has $\mathcal{N}_{d+p}(0,\Gamma)$ distribution where

$$\Gamma := \begin{pmatrix} 16\pi^3 W(\theta) & 0 \\ 0 & 8\pi^3 V(\theta) \end{pmatrix}.$$

We also need to define

$$\begin{aligned}
\Lambda_{n0}(\vartheta) &:= \Lambda_n(\vartheta,\beta_0) := \sum_{t,s=1}^n b_{t-s}(\vartheta)\varepsilon_t\varepsilon_s, & Z_{n\theta} &:= n^{-1/2}\nabla_\vartheta \Lambda_{n0}(\theta), \\
T_n(\vartheta) &:= \sum_{t,s=1}^n b_{t-s}(\vartheta)\varepsilon_t \boldsymbol{X}_s, & Z_{n\beta} &:= n^{-1/2}T_n(\theta), \\
T_{n1}(\vartheta) &:= \sum_{t,s=1}^n b_{t-s}(\vartheta)\varepsilon_t \boldsymbol{X}_{s1}, & Z_{n\beta 1} &:= n^{-1/2}T_{n1}(\theta), \\
A_n(\vartheta) &:= \sum_{t,s=1}^n b_{t-s}(\vartheta)\boldsymbol{X}_t \boldsymbol{X}'_s, & \beta'_0 &:= (\beta'_{01},\beta'_{02}).
\end{aligned} \tag{2.2}$$

We are now ready to state

Lemma 2.1. *Under the conditions of Theorem* 2.1,

$$\Lambda_n(\hat\theta_n,\hat\beta_n) = \Lambda_n(\theta,\beta_0) - \frac{1}{2}Z'_{n\theta}(2\pi W(\theta))^{-1}Z_{n\theta} - Z'_{n\beta}(2\pi V(\theta))^{-1}Z_{n\beta} + o_p(1) \tag{2.3}$$

$$(Z'_{n\theta}, Z'_{n\beta}) \Longrightarrow (Z'_\theta, Z'_\beta). \tag{2.4}$$

We remark here that (2.3) also holds under H_0. The proof of Lemma 2.1 is postponed till later. We use it to yield the following proof.

P r o o f of Theorem 2.1. Apply Lemma 2.1 under H_0 to obtain

$$\Lambda_n(\hat\theta_n, \hat\beta_{n1}) = \Lambda_n(\theta, \beta_{01}) - \frac{1}{2} Z'_{n\theta}(2\pi W(\theta))^{-1} Z_{n\theta} - Z'_{n\beta 1}(2\pi V_{11}(\theta))^{-1} Z_{n\beta 1} + o_p(1). \tag{2.5}$$

Under H_0, $\Lambda_n(\theta, \beta_0) = \Lambda_n(\theta, \beta_{01})$, (2.3) and (2.5) yield that

$$\mathcal{Q}_n = 2[Z'_{n\beta}(2\pi V(\theta))^{-1} Z_{n\beta} - Z'_{n\beta 1}(2\pi V_{11}(\theta))^{-1} Z_{n\beta 1}] + o_p(1). \tag{2.6}$$

Hence (2.4) implies that $\mathcal{Q}_n \Longrightarrow \mathcal{Q} := 2[Z'_\beta(2\pi V(\theta))^{-1} Z_\beta - Z'_{\beta 1}(2\pi V_{11}(\theta))^{-1} Z_{\beta 1}]$, where $Z_{\beta 1}$ is the vector of the first k components of Z_β. Now the claim follows from the following fact. Let $Z \sim \mathcal{N}_p(0, \Sigma)$ where Σ is a positive definite covariance matrix, and let Z_1 denote its first k components with covariance matrix Σ_{11}. Then the distribution of $Z'\Sigma^{-1}Z - Z'_1\Sigma_{11}^{-1}Z_1$ is χ^2_{p-k}. See, e.g., [16]. □

P r o o f of (2.3) of Lemma 2.1. Assume that (2.4) holds. By using an argument as in [11], one can verify that under the assumed set up $(\hat\theta_n, \hat\beta_n)$ is consistent for (θ, β_0). In other words, for all aufficiently large n, $\hat\theta_n \in \Theta_0$, $\hat\beta_n \in B_0$ with probability arbitrarily close to 1, where $\Theta_0 \ni \theta$, $B_0 \ni \beta_0$ are arbitrarily small neighborhoods of the true parameters. The Taylor expansion around (θ, β_0) yields

$$\begin{aligned}\Lambda_n(\hat\theta_n, \hat\beta_n) = \Lambda_n(\theta, \beta_0) &+ (\hat\theta_n - \theta)'\nabla_\vartheta \Lambda_n(\theta, \beta_0) + (\hat\beta_n - \beta_0)'\nabla_\beta \Lambda_n(\theta, \beta_0) \\ &+ \frac{1}{2}(\hat\theta_n - \theta)'\nabla^2_{\vartheta\vartheta}\Lambda_n(\theta^*_n, \beta^*_n)(\hat\theta_n - \theta) \\ &+ (\hat\beta_n - \beta_0)'\nabla^2_{\vartheta\beta}\Lambda_n(\theta^*_n, \beta^*_n)(\hat\theta_n - \theta) \\ &+ \frac{1}{2}(\hat\beta_n - \beta_0)'\nabla^2_{\beta\beta}\Lambda_n(\theta^*_n, \beta^*_n)(\hat\beta_n - \beta_0),\end{aligned} \tag{2.7}$$

where $(\theta^*_n, \beta^*_n) \in \Theta \times \mathbb{R}^p$ are some random vectors such that

$$\|\theta^*_n - \theta\| \leqslant \|\hat\theta_n - \theta\| = o_p(1), \quad \|\beta^*_n - \beta_0\| \leqslant \|\hat\beta_n - \beta_0\| = o_p(1). \tag{2.8}$$

We shall prove below that the following asymptotic relations hold:

$$n^{1/2} g(\hat\theta_n - \theta) = -(2\pi W(\theta))^{-1} Z_{n\theta} + o_p(1), \tag{2.9}$$

$$n^{1/2}(\hat\beta_n - \beta_0) = (2\pi V(\theta))^{-1} Z_{n\beta} + o_p(1). \tag{2.10}$$

Now, the Taylor expansion yields

$$\Lambda_n(\vartheta, \beta) = \Lambda_{n0}(\vartheta) - 2(\beta - \beta_0)'T_n(\vartheta) + (\beta - \beta_0)'A_n(\vartheta)(\beta - \beta_0). \tag{2.11}$$

□

The following lemma will also be proved later.

Lemma 2.2. *Under the assumptions of Theorem* 2.1 *the following hold:*

$$n^{-1}\nabla^2_{\vartheta\vartheta}\Lambda_{n0}(\vartheta) \to 2\pi W(\vartheta), \tag{2.12}$$

$$n^{-1}\nabla_{\vartheta}T_n(\vartheta) \to 0, \tag{2.13}$$

$$n^{-1}A_n(\vartheta) \to 2\pi V(\vartheta), \tag{2.14}$$

uniformly in $\vartheta \in \Theta_0$ *a.s. Moreover,*

$$n^{-1}\sup_{\vartheta\in\Theta_0}\left(|\nabla_\vartheta A_n(\vartheta)| + |\nabla_{\vartheta\vartheta}A_n(\vartheta)| + |\nabla_\vartheta T_n(\vartheta)| + |\nabla_{\vartheta\vartheta}T_n(\vartheta)|\right) = O_p(1). \tag{2.15}$$

Note that (2.11), (2.12)–(2.14), (2.15) and (2.8) imply the convergence of the second derivatives in (2.7) (in probability):

$$n^{-1}\nabla^2_{\vartheta\vartheta}\Lambda_n(\theta^*_n, \beta^*_n) \to 2\pi W(\theta), \tag{2.16}$$

$$n^{-1}\nabla^2_{\vartheta\beta}\Lambda_n(\theta^*_n, \beta^*_n) \to 0, \tag{2.17}$$

$$n^{-1}\nabla^2_{\beta\beta}\Lambda_n(\theta^*_n, \beta^*_n) \to 4\pi V(\theta). \tag{2.18}$$

The representation (2.3) now follows from (2.7) and (2.9), (2.10), (2.4) and (2.16)–(2.18). Relations (2.4) and (2.9), (2.10) are proved in Sections 2 and 3 below. □

3. Proof of the CLT in (2.4)

Recall the definitions of ε_t and $\boldsymbol{X}_t$ from (1.2) and (2.1). In the sequel, we shall suppress θ in $a(t;\theta)$ in (1.2) and some other notation. The following proof of (2.4) has its roots in [9]. For notational simplicity, we shall assume that $d = p = 1$, in particular, that X_t is a scalar process. The idea of the proof is to approximate the quadratic forms $Z_{n\theta}$, $Z_{n\beta}$ by the corresponding "diagonal" forms $\tilde{Z}_{n\theta}$, $\tilde{Z}_{n\beta}$ defined as follows. Let $\hat{a}_1(u) := (2\pi)^{1/2}\hat{a}(u)|\nabla_\vartheta \hat{b}(u)|^{1/2}$, $\hat{a}_2(u) := (2\pi)^{1/2}\hat{a}(u)|\nabla_\vartheta \hat{b}(u)|^{1/2}\operatorname{sgn}(\nabla_\vartheta \hat{b}(u))$, $a_j(t) := \int_\Pi \mathrm{e}^{\mathrm{i}tu}\hat{a}_j(u)\,\mathrm{d}u$, $j = 1, 2$, $G(t) := \sum_{s\in\mathbb{Z}} b_{t-s}B(s)$, $t \in \mathbb{Z}$, and

$$\tilde{\varepsilon}_{jt} := \sum_{s\in\mathbb{Z}} a_j(t-s)\zeta_s, \quad j = 1, 2, \quad \tilde{X}_t := \sum_{s\in\mathbb{Z}} G(t-s)\xi_s, \tag{3.1}$$

where $\xi_s \equiv \xi_{s,1}$, $B(s) \equiv B_{11}(s)$. Then the approximating quadratic forms are

$$\tilde{Z}_{n\theta} := n^{-1/2}\sum_{t=1}^{n}(\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t} - E\tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t}), \quad \tilde{Z}_{n\beta} := n^{-1/2}\sum_{t=1}^{n}\varepsilon_t\tilde{X}_t. \tag{3.2}$$

The processes $\{\tilde{\varepsilon}_{jt}\}$, $j = 1, 2$ follow the definition in [9, (1.9)] with weights $a_j(t)$, $j = 1, 2$ defined through their Fourier transforms as in [9, (2.1), (2.2)].

Note that $\sum_{t\in\mathbb{Z}} G(t)^2 < \infty$ since the Fourier transform $\hat{G}(u) = (2\pi)^{-1} \sum_{s\in\mathbb{Z}} G(s)\mathrm{e}^{\mathbf{i}su}$, $u \in \Pi$ satisfies $\hat{G}(u) = 2\pi\hat{b}(u)\hat{B}(u) = 2\pi f^{-1}(u)\hat{B}(u)$ where $\hat{B}$ is the Fourier transform of $(B(s))$, and so $\hat{G} \in L^2(\Pi)$ by the assumption of boundedness of f^{-1}. In particular, the process $\{\tilde{X}_t\}$ in (3.1) is a well-defined moving average process independent of the process $\{\varepsilon_t\}$. Clearly, for $d = p = 1$, (2.4) follows from

$$(\tilde{Z}_{n\theta}, \tilde{Z}_{n\beta}) \Longrightarrow \mathcal{N}_2(0, \Gamma) \tag{3.3}$$

and

$$E(Z_{n\theta} - \tilde{Z}_{n\theta})^2 = o(1), \quad E(Z_{n\beta} - \tilde{Z}_{n\beta})^2 = o(1). \tag{3.4}$$

The first relation in (3.4) follows from [9, Lemmas 1 and 4]. Then

$$\begin{aligned} E(Z_{n\beta} - \tilde{Z}_{n\beta})^2 &= n^{-1} \sum_{t_2,t_2=1}^{n} \sum_{s_1,s_2\notin[1,n]} b_{t_1-s_1} b_{t_2-s_2} E\varepsilon_{t_1}\varepsilon_{t_2} EX_{s_1}X_{s_2} \\ &= \int_{\Pi^2} f(x)g(z)\,\mathrm{d}x\,\mathrm{d}z \Big(n^{-1/2}\int_{\Pi} D_n(x+y)D_n(y-z)(f^{-1}(y) - f^{-1}(z))\,\mathrm{d}y\Big)^2, \end{aligned} \tag{3.5}$$

where $D_n(x) := \sum_{s=1}^{n} \mathrm{e}^{\mathbf{i}sx}\mathrm{e}^{-\mathbf{i}(n+1)x/2}$ is the Dirichlet kernel. Recall that

$$\begin{aligned} |D_n(x)| &\leqslant \pi n^c |x|^{c-1}, \quad \forall\, 0 < c < 1,\ x \in \Pi, \\ &\leqslant 2\pi n(1 + n|x|)^{-1}, \quad \forall\, x \in \Pi. \end{aligned}$$

Using these bounds and the boundedness and continuity of f^{-1}, it follows that the bracketed expression in (3.5) is bounded in x, z, n, and tends to zero as $n \to \infty$. This proves (3.4).

Next, we turn to the proof of (3.3). We use an approach similar to that in [9, proof of Thm. 3]. Accordingly, we define below "$2M$-memory" ($M < \infty$) approximations

$$\tilde{\varepsilon}_{it}^M := \sum_{|t-s|\leqslant M} a_i^M(t-s)\zeta_s, \tag{3.6}$$

$$\varepsilon_i^M := \sum_{|t-s|\leqslant M} a^M(t-s)\zeta_s, \tag{3.7}$$

$$\tilde{X}_t^M := \sum_{|t-s|\leqslant M} G^M(t-s)\xi_s, \tag{3.8}$$

to (3.1) and (1.2), and the corresponding quadratic forms:

$$\tilde{Z}^M_{n\theta} := n^{-1/2} \sum_{t=1}^{n} (\tilde{\varepsilon}^M_{1t} \tilde{\varepsilon}^M_{2t} - E\tilde{\varepsilon}^M_{1t}\tilde{\varepsilon}^M_{2t}), \quad \tilde{Z}^M_{n\beta} := n^{-1/2} \sum_{t=1}^{n} \varepsilon^M_t \tilde{X}^M_t. \tag{3.9}$$

Relation (3.3) follows from

$$\tilde{Z}^M_{n\theta}, \tilde{Z}^M_{n\beta}) \Longrightarrow \mathcal{N}_2(0, \Gamma^M) \tag{3.10}$$

for any $M < \infty$, where Γ^M is a covariance matrix, and

$$\lim_{M\to\infty} \|\Gamma^M - \Gamma\| = 0, \tag{3.11}$$

$$\lim_{M\to\infty} \limsup_{n\to\infty} E(\tilde{Z}^M_{n\theta} - \tilde{Z}_{n\theta})^2 = 0, \tag{3.12}$$

$$\lim_{M\to\infty} \limsup_{n\to\infty} E(\tilde{Z}^M_{n\beta} - \tilde{Z}_{n\beta})^2 = 0, \tag{3.13}$$

see [9, p. 97] for details.

The approximating processes (i.e., the weight functions $a^M_i(t)$, $a^M(t)$, $G^M(t)$) are defined through their Fourier transforms. In particular, the sequences $(a^M_i(t))$ $(i = 1, 2)$ are defined in [9, p. 96]. Relation (3.12) also follows from [9, (3.8)]. The CLT in (3.10) is a consequence of the fact that the approximating processes in (3.6)–(3.8) are all stationary and $2M$-dependent.

It remains to show the existence of $a^M(t)$, $G^M(t)$ satisfying (3.13). Consider the Fourier transforms $\hat{a}$, $\hat{G}$ (the transfer functions of filters $\{\varepsilon_t\}$, $\{\tilde{X}_t\}$), then $V(\theta) = \int_\Pi g(x) f^{-1}(x)\,\mathrm{d}x = \int_\Pi |\hat{a}(x)\hat{G}(x)|^2\,\mathrm{d}x < \infty$. Note that

$$E\tilde{Z}^2_{n\beta} = 4\pi^2 n^{-1} \int_{\Pi^2} D^2_n(x+y) |\hat{a}(x)\hat{G}(z)|^2 \,\mathrm{d}x\,\mathrm{d}z.$$

We shall prove below that

$$n^{-1} \int_{\Pi^2} D^2_n(x+y) |\hat{a}(x)\hat{G}(z)|^2 \,\mathrm{d}x\,\mathrm{d}z \to 2\pi \int_\Pi |\hat{a}(x)\hat{G}(x)|^2\,\mathrm{d}x. \tag{3.14}$$

Put

$$\begin{aligned} a_K(j) &:= \int_\Pi \mathrm{e}^{-\mathbf{i}jx} \hat{a}(x) \mathbf{1}(|\hat{a}(x)| \leqslant K)\,\mathrm{d}x, \\ G_K(j) &:= \int_\Pi \mathrm{e}^{-\mathbf{i}jx} \hat{G}(x) \mathbf{1}(|\hat{G}(x)| \leqslant K)\,\mathrm{d}x, \end{aligned} \tag{3.15}$$

and

$$\varepsilon_{t,K} := \sum_{j\in\mathbb{Z}} a_K(j)\zeta_{t-j}, \quad \tilde{X}_{t,K} := \sum_{j\in\mathbb{Z}} G_K(j)\xi_{t-j}. \tag{3.16}$$

Then

$$(3.17)\quad n^{-1}E\Big(\sum_{t=1}^{n}(\varepsilon_t\tilde{X}_t-\varepsilon_{t,K}\tilde{X}_{t,K})\Big)^2$$
$$=4\pi^2\int_{\Pi^2}D_n^2(x+z)|\hat{a}(x)\hat{G}(z)|^2(1-\mathbf{1}(|\hat{a}(x)|\leqslant K,\,|\hat{G}(x)|\leqslant K))\,\mathrm{d}x\,\mathrm{d}z$$
$$\to 8\pi^3\int_{\Pi}|\hat{a}(z)\hat{G}(z)|^2\mathbf{1}(|\hat{a}(z)|>K\text{ or }|\hat{G}(z)|>K)\,\mathrm{d}z,\quad n\to\infty,$$

where we have used (3.14) and the well-known fact that a similar convergence holds for any bounded measurable functions instead of $\hat{a}$ and $\hat{G}$. Next, using boundedness of truncated (by K) transfer functions in (3.15), one can approximate them in $L^4(\Pi)$ by trigonometric polynomials of degree $2M$ similarly to [9, p. 96], and the corresponding moving averages by $2M$-dependent moving averages. More precisely, for any $K,\varepsilon>0$ there exist $M>0$ and trigonometric polynomials $\hat{a}^M$, $\hat{G}^M$ such that the Fourier coefficients $a^M(j)=\int_\Pi \mathrm{e}^{\mathbf{i}jx}\hat{a}^M(x)\,\mathrm{d}x$, $G^M(j)=\int_\Pi \mathrm{e}^{\mathbf{i}jx}\hat{G}^M(x)\,\mathrm{d}x$ vanish for $|j|>M$ and

$$(3.18)\quad \int_\Pi(|\hat{a}(x)\mathbf{1}(|\hat{a}(x)|\leqslant K)-\hat{a}^M(x)|^4+|\hat{G}(x)\mathbf{1}(|\hat{G}(x)|\leqslant K)-\hat{G}^M(x)|^4)\,\mathrm{d}x<\varepsilon.$$

Similarly to (3.17),

$$(3.19)\quad n^{-1}E\Big(\sum_{t=1}^{n}(\varepsilon_t^M\tilde{X}_t^M-\varepsilon_{t,K}\tilde{X}_{t,K})\Big)^2$$
$$\to 8\pi^3\int_\Pi|\hat{a}(x)\hat{G}(x)\mathbf{1}(|\hat{a}(x)|\leqslant K,\,|\hat{G}(x)|\leqslant K)-\hat{a}^M(x)\hat{G}^M(x)|^2\,\mathrm{d}x.$$

Clearly, by (3.18), the right-hand side of (3.19) does not exceed $C_K\varepsilon^{1/2}$, i.e. it can be made arbitrarily small by an appropriate choice of M and the approximating polynomials. Together with (3.17) and the fact that the right-hand side of (3.17) vanishes as $K\to\infty$, this concludes the proof of (3.13). Relation (3.11) easily follows from relations (3.14), (3.17), (3.19).

It remains to show the limit (3.14). Using $n^{-1}\int_\Pi D_n^2(x+y)\,\mathrm{d}y=2\pi$, the difference between the left- and right-hand sides in (3.14) can be rewritten as

$$u_n:=n^{-1}\int_\Pi|\hat{G}(z)|^2\,\mathrm{d}z\int_\Pi D_n^2(x-z)(|\hat{a}(x)|^2-|\hat{a}(z)|^2)\,\mathrm{d}x$$
$$=\int_\Pi g(z)f^{-1}(z)\,\mathrm{d}z\Big\{n^{-1}\int_\Pi D_n^2(x-z)\Big(\frac{f(x)}{f(z)}-1\Big)\,\mathrm{d}x\Big\}.$$

Write $u_n = \int_\Pi f^{-1}(z)g(z)j_n(z)\,\mathrm{d}z$, where $j_n(z)$ denotes the quantity in the curly brackets. Further, split $u_n = \int_{|z|<\delta} \cdots + \int_{z\in\Pi,|z|>\delta} \cdots =: u_{n1}(\delta) + u_{n2}(\delta)$. Note that $u_{n2}(\delta) = o(1)$ for any fixed $\delta > 0$, which follows from the fact that $j_n(z) = o(1)$ uniformly in $|z| > \delta$, $z \in \Pi$, which fact in turn follows from joint continuity of $f(x)/f(z)$ on $x, z \in [-\pi, -\delta] \cup [\delta, \pi]$ for any $\delta > 0$ fixed. It remains to show that $u_{n1}(\delta) \to 0$ ($\delta \to 0$) uniformly in n. As g is integrable, it suffices to show that $f^{-1}(z)j_n(z)$ is bounded on $(-\delta, \delta)$ uniformly in n. Using (a.6) and the bound $D_n^2(x) \leqslant Cn^2/(1+n^2x^2)$, for $0 < z < \delta$ one obtains

$$
\begin{aligned}
f^{-1}(z)|j_n(z)| &\leqslant f^{-2}(z)n^{-1}\int_\Pi D_n^2(x-z)f(x)\,\mathrm{d}x + Cf^{-1}(z) \\
&\leqslant C\int_0^\infty \frac{nz^\alpha\,\mathrm{d}x}{(1+n^2(x-z)^2)x^\alpha} + O(1) \\
&= C\int_0^\infty \frac{nz\,\mathrm{d}u}{(1+n^2z^2(1-u)^2)u^\alpha} + O(1) =: CJ_n(z) + O(1).
\end{aligned}
$$

Let $nz \geqslant 1$, then

$$
J_n(z) \leqslant nz\int_{|1-u|\leqslant 1/nz} u^{-\alpha}\,\mathrm{d}u + (nz)^{-1}\int_{|1-u|\geqslant 1/nz} (1-u)^{-2}u^{-\alpha}du \leqslant C.
$$

Next, let $nz \leqslant 1$, then

$$
J_n(z) \leqslant nz\int_0^{1+1/nz} u^{-\alpha}\,\mathrm{d}u + (nz)^{-1}\int_{1+1/nz} (u-1)^{-2}u^{-\alpha}\,\mathrm{d}u = O((nz)^\alpha) = O(1).
$$

This proves the boundedness of $f^{-1}(z)j_n(z)$ as well as $u_n = o(1)$ and the limit in (3.14), thereby completing the proof of the CLT in (2.4).

4. Proof of (2.9)–(2.10)

Next, we prove the $n^{1/2}$-consistency of the estimator $(\hat\theta_n, \hat\beta_n)$. Under some regularity conditions, $(\hat\theta_n, \hat\beta_n)$ is a unique solution in $\Theta_0 \times B_0$ of the equation

$$
\sum_{t,s=1}^n \nabla_\vartheta b_{t-s}(\hat\theta_n)(Y_t - X_t'\hat\beta_n)(Y_s - X_s'\hat\beta_n) = 0, \tag{4.1}
$$

$$
\sum_{t,s=1}^n b_{t-s}(\hat\theta_n)\nabla_\beta(Y_t - X_t'\hat\beta_n)(Y_s - X_s'\hat\beta_n) = 0. \tag{4.2}
$$

Relation (4.2) yields

$$
\hat\beta_n = \beta_0 + \hat A_n^{-1}\hat T_n, \tag{4.3}
$$

where $\hat{A}_n := A_n(\hat{\theta}_n)$, $\hat{T}_n := T_n(\hat{\theta}_n)$ and where $A_n(\vartheta)$, $T_n(\vartheta)$ are defined in (2.2) above. Note the equation in (4.1) for $\hat{\theta}_n$ is the same as in [14, p. 134]. Therefore we can use the results in [14] to estimate $|\hat{\theta}_n - \theta|$ given a rate of $|\hat{\beta}_n - \beta_0|$ and vice versa. Namely, since $\hat{\theta}_n = \operatorname{argmin}_\vartheta \tilde{\Lambda}_n(\vartheta)$ where $\tilde{\Lambda}_n(\vartheta) := \Lambda_n(\vartheta, \hat{\beta}_n)$, so expanding (4.1) as in [14] yields

$$0 = \nabla_\vartheta \tilde{\Lambda}_n(\hat{\theta}_n) = \nabla_\vartheta \tilde{\Lambda}_n(\theta) + \nabla^2_{\vartheta\vartheta}(\vartheta^*_n)(\hat{\theta}_n - \theta). \tag{4.4}$$

From (2.12) and consistency of $(\hat{\theta}_n, \hat{\beta}_n)$ it follows that $n^{-1}\nabla^2_{\vartheta\vartheta}\tilde{\Lambda}_n(\vartheta^*_n) = 2\pi W(\theta) + o_p(1)$ a.s. and therefore

$$|\hat{\theta}_n - \theta| = O_p(|n^{-1}\nabla_\vartheta \tilde{\Lambda}_n(\theta)|) \tag{4.5}$$

according to (4.4). However,

$$\nabla_\vartheta \tilde{\Lambda}_n(\theta) = n^{1/2} Z_{n\theta} + \nabla_\vartheta \Lambda_{n2}(\theta)(\hat{\beta}_n - \beta_0) + (\hat{\beta}_n - \beta_0)'\nabla_\vartheta \Lambda_{n3}(\theta)(\hat{\beta}_n - \beta_0), \tag{4.6}$$

where $Z_{n\theta} = O_p(1)$ (see (2.4)) and

$$\nabla_\vartheta \Lambda_{n3}(\theta) := \sum_{t,s=1}^{n} \nabla_\vartheta b_{t-s}(\theta) X_t X_s' = O_p(n), \tag{4.7}$$

$$\nabla_\vartheta \Lambda_{n2}(\theta) := \sum_{t,s=1}^{n} \nabla_\vartheta b_{t-s}(\theta) \varepsilon_t X_s' = O_p(n^{3/4}), \tag{4.8}$$

see [14], (3.2), (3.4). From (4.5) and (4.6)–(4.8) we conclude

$$\begin{aligned} |\hat{\theta}_n - \theta| &= O_p(n^{-1/2}) + O_p(n^{-1/4}|\hat{\beta}_n - \beta_0|) + O_p(|\hat{\beta}_n - \beta_0|^2) \\ &= O_p(n^{-1/2}) + O_p(|\hat{\beta}_n - \beta_0|^2) \\ &= O_p(\max(|\hat{\beta}_n - \beta_0|^2, n^{-1/2})). \end{aligned} \tag{4.9}$$

Next, we shall estimate $|\hat{\beta}_n - \beta_0|$ given a rate of $|\hat{\theta}_n - \theta|$. From (4.3) we obtain

$$\begin{aligned} |\hat{\beta}_n - \beta_0| &= \Big|\hat{A}_n^{-1} \sum_{t,s=1}^{n} (b_{t-s}(\hat{\theta}_n) - b_{t-s}(\theta)) X_t \varepsilon_s + \hat{A}_n^{-1} \sum_{t,s=1}^{n} b_{t-s}(\theta) X_t \varepsilon_s| \\ &\leqslant |\hat{\theta}_n - \theta| \sup_{\vartheta' \in \Theta_0} |n A_n^{-1}(\vartheta')| \sup_{\vartheta'' \in \Theta_0} |n^{-1}\nabla_\vartheta T_n(\vartheta'')| \\ &\quad + n^{-1/2} \sup_{\vartheta \in \Theta_0} |n A_n^{-1}(\vartheta)||Z_{n\beta}|. \end{aligned} \tag{4.10}$$

Here $Z_{n\beta} = O_p(1)$ (see (2.4)), $\sup_{\vartheta\in\Theta_0} |nA_n^{-1}(\vartheta)| = O_p(1)$ (see (2.14)), and

$$\sup_{\vartheta''\in\Theta_0} |n^{-1}\nabla_\vartheta T_n(\vartheta'')| = O_p(1)$$

(see (2.15)). Therefore from (4.10) we obtain

$$|\hat\beta_n - \beta_0| = O_p(\max(|\hat\theta_n - \theta|, n^{-1/2})). \tag{4.11}$$

Substituting (4.9) into (4.11) and using $|\hat\beta_n - \beta_0| = o_p(1)$ yields

$$|\hat\beta_n - \beta_0| = O_p(n^{-1/2}), \tag{4.12}$$

which in turn together with (4.9) yields

$$|\hat\theta_n - \theta| = O_p(n^{-1/2}). \tag{4.13}$$

Next, we prove (2.9) and (2.10). Using $n\hat A_n^{-1} \to (2\pi V(\theta))^{-1}$ a.s. (which follows from (2.14) and the consistency of $\hat\theta_n$), from (4.3) we obtain

$$n^{1/2}(\hat\beta_n - \beta_0) = (2\pi V(\theta))^{-1} Z_{n\beta} + O_p(n^{-1/2}|T_n(\hat\theta_n) - T_n(\theta)|) + o_p(1),$$

where

$$\begin{aligned} |T_n(\hat\theta_n) - T_n(\theta)| &= \Big|(\hat\theta_n - \theta)'\nabla_\vartheta T_n(\theta) + (1/2)(\hat\theta_n - \theta)'\nabla^2_{\vartheta\vartheta} T_n(\vartheta_n^*)(\hat\theta_n - \theta)| \\ &\leqslant n|\hat\theta_n - \theta||n^{-1}\nabla_\vartheta T_n(\theta)| + n|\hat\theta_n - \theta|^2 \sup_{\vartheta\in\Theta_0} |n^{-1}\nabla^2_{\vartheta\vartheta} T_n(\vartheta)| \\ &= o_p(n^{1/2}) \end{aligned}$$

according to (4.13), (2.13) and (2.15). This proves (2.10). Relation (2.9) follows from [14, Thm. 1.2] and (4.12).

Relations (2.12)–(2.14) follow from a generalization of the results in [11]. Finally, the details of the proof of (2.15) are completely analogous to those appearing in [14, p. 146]. This concludes the proof of Lemma 2.1.

References

[1] *R. Baillie*: Long memory processes and fractional integration in econometrics. J. Econom. *73* (1996), 5–59.

[2] *G. G. Booth, R. R. Kaen, P. E. Koves*: R/S analysis of foreign exchange rates under two international money regimes. J. Monet. Econom. *10* (1982), 4070–415.

[3] *J. Beran*: Statistical methods for data with long-range dependence (with discussion). Stat. Sci. *7* (1992), 404–427.

[4] *J. Beran*: Statistics for Long Memory Processes. Monographs on Statistics and Applied Probability, Vol. 61. Chapman and Hall, London, 1994.

[5] *Y.-W. Cheung*: Long memory in foreign exchange rates. Journal of Business and Economic Statistics *11* (1993), 93–101.

[6] Empirical Process Techniques for Dependent Data (H. Dehling, T. Mikosch, M. Sørensen, eds.). Birkhäuser, Boston, 2002.

[7] Theory and Applications of Long-Range Dependence (P. Doukhan, G. Oppenheim, M. S. Taqqu, eds.). Birkhäuser, Boston, 2003.

[8] *R. Fox, M. S. Taqqu*: Large-sample properties of parameter estimates for strongly dependent Gaussian time series. Ann. Stat. *14* (1986), 517–532.

[9] *L. Giraitis, D. Surgailis*: A central limit theorem for quadratic forms in strongly dependent linear variables and its applications to asymptotical normality of Whittle's estimate. Prob. Theory Relat. Fields *86* (1990), 87–104.

[10] *H. Guo, H. L. Koul*: Nonparametric regression with heteroscedastic long memory errors. J. Stat. Plann. Inference *137* (2007), 379–404.

[11] *E. J. Hannan*: The asymptotic theory of linear time-series models. J. Appl. Probab. *10* (1973), 130–145.

[12] *J. D. Hart*: Nonparametric Smoothing and Lack-of-Fit Tests. Springer Series in Statistics. Springer, New York, 1997.

[13] *J. Hidalgo, P. M. Robinson*: Adapting to unknown disturbance autocorrelation in regression with long memory. Econometrica *70* (2002), 1545–1581.

[14] *H. L. Koul, D. Surgailis*: Asymptotic normality of the Whittle estimator in linear regression models with long memory errors. Stat. Inference Stoch. Process. *3* (2000), 129–147.

[15] *A. W. Lo*: Long term memory in stock market prices. Econometrica *59* (1991), 1279–1313.

[16] *C. R. Rao*: Linear Statistical Inference and Its Applications. Wiley, New York, 1965.

Authors' addresses: *H. L. Koul*, Department of Statistics and Probability, A435 Wells Hall, Michigan State University, East Lansing, MI 48824, U.S.A., e-mail: `koul@stt.msu.edu`; *D. Surgailis*, Vilnius Institute of Mathematics and Informatics, 2600 Vilnius, Lithuania, e-mail: `sdonatas@ktl.mii.lt`.