The Annals of Statistics
1999, Vol. 27, No. 6, 2054–2080

LOCAL ASYMPTOTIC NORMALITY FOR REGRESSION MODELS WITH LONG-MEMORY DISTURBANCE

BY MARC HALLIN,[1] MASANOBU TANIGUCHI, [2] ABDESLAM SERROUKH[3] AND KOKYO CHOY

Université Libre de Bruxelles, Osaka University, Université Libre de Bruxelles and Komazawa University, Tokyo

The local asymptotic normality property is established for a regression model with fractional ARIMA(p, d, q) errors. This result allows for solving, in an asymptotically optimal way, a variety of inference problems in the long-memory context: hypothesis testing, discriminant analysis, rank-based testing, locally asymptotically minimax and adaptive estimation, etc. The problem of testing linear constraints on the parameters, the discriminant analysis problem, and the construction of locally asymptotically minimax adaptive estimators are treated in some detail.

1. Introduction. Local asymptotic normality [LAN; see, e.g., Le Cam (1960, 1986); Strasser (1985); Le Cam and Yang (1990)] constitutes a key result in asymptotic inference and allows for the solution of virtually all asymptotic inference problems. The LAN approach recently has been adopted in the study of a variety of time series models. Swensen (1985) established LAN for autoregressive models of finite order with a regression trend and applied it in the derivation of the local power of the Durbin–Watson test. For a stationary ARMA (p, q) process (without trend), Kreiss (1987, 1990) also proved the LAN property, and constructed locally asymptotically minimax LAM adaptive estimators, as well as locally asymptotically maximin tests for AR models. Based on the LAN property, Hallin and Puri (1988, 1994) discussed the problem of testing arbitrary linear constraints on the parameters of an ARMA model with trend. Kreiss (1990) showed the LAN for autoregressive processes of infinite order and discussed adaptive estimation. For multivariate ARMA models with a linear trend, Garel and Hallin (1995) established the LAN property and gave various expressions for the central sequence. Unit root and cointegrated models have been studied by Jeganathan (1995, 1997). Benghabrit

Received March 1997; revised October 1999.

[1]Research supported by the Human Capital contract ERB CT CHRX 940 963 and the Fonds d'Encouragement à la Recherche de l'Université Libre de Bruxelles.

[2]Work completed while visiting the Institut de Statistique et de Recherche Opérationnelle at the Université Libre de Bruxelles under grant of Belgian Fonds National de la Recherche Scientifique.

[3]Research supported by the Fonds d'Encouragement à la Recherche de l'Université Libre de Bruxelles.

AMS 1991 *subject classifications*. Primary 60G10, 62E20; secondary 62A10, 62F05.

Key words and phrases. Long-memory process, FARIMA model, local asymptotic normality, locally asymptotically optimal test, adaptive estimation, discriminant analysis.

and Hallin (1998) investigate the bilinear case in the vicinity of AR dependence. Other nonlinear time series models have been considered, for example, by Hwang and Basawa (1993), Drost, Klaassen and Werker (1997) and Koul and Schick (1997).

All the above results are for *short-memory* time series models. The *long-memory* case, from this point of view, remains largely, if not totally, unexplored. Asymptotic results, though, exist. But they mainly deal with the asymptotic behavior of traditional MLEs or LSEs. For a Gaussian *long-memory* process, Dahlhaus (1989) proved the asymptotic normality of the maximum likelihood estimator of spectral parameters and discussed its asymptotic efficiency. Yajima (1985, 1991) elucidated problems related with the asymptotics of the BLUE and the LSE in a regression model with long-memory stationary errors. Beran (1995) considers the case of the differencing parameter. Sethuraman and Basawa (1997) consider the MLE for a multivariate fractionally differenced AR model and obtain its asymptotic distribution. The monograph by Beran (1994) gives an extensive review of various problems in long-memory time series. However, the systematic approach based on the LAN property apparently has not been considered so far. This paper develops the asymptotic theory based on the LAN approach for a regression model with stationary non-Gaussian FARIMA(p, d, q) long-memory errors. In Section 2, we show that the log-likelihood ratios exhibit the typical LAN behavior, so that the local experiments converge weakly to a Gaussian shift experiment. This result is used, in Section 3, in the solutions of a variety of inference problems. Examples are given, and a general formula for the distribution, under contiguous sequences of alternatives, of a wide class of statistics is established. These results are quite general, and remain valid under a very large family of innovation densities.

Throughout the paper we denote by $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ the sets of (natural) integers.

2. Local asymptotic normality (LAN). Suppose that we observe $\mathbf{Y}^{(n)} = (Y_1, \ldots, Y_n)'$ generated by

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta} + e_t, \qquad t \in \mathbb{Z}, \tag{2.1}$$

where $\{\mathbf{X}_t = (X_{t1}, \ldots, X_{tb})',\ t \in \mathbb{Z}\}$ is a sequence of b-dimensional real-valued nonstochastic regressors, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_b)'$ is a vector of nonserial parameters and $\{e_t,\ t \in \mathbb{Z}\}$ is a stationary long-memory process generated by the *fractional* ARIMA [in short, FARIMA(p, d, q)] model (L, as usual, stands for the lag operator),

$$\sum_{k=0}^{p} \phi_k L^k (1 - L)^d e_t = \sum_{k=0}^{q} \eta_k L^k \varepsilon_t, \qquad t \in \mathbb{Z},\ \phi_0 = \eta_0 = 1, \tag{2.2}$$

with a vector of serial parameters

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+q+1})' = (d, \phi_1, \ldots, \phi_p, \eta_1, \ldots, \eta_q)' = (d, \boldsymbol{\phi}', \boldsymbol{\eta}')'$$

and an i.i.d. *innovation* process $\{\varepsilon_t,\ t \in \mathbb{Z}\}$, with nonvanishing innovation density g. In matrix notation, (2.1) also writes

$$\mathbf{Y}^{(n)} = \mathbf{X}^{(n)}\boldsymbol{\beta} + \mathbf{e}^{(n)}, \qquad t \in \mathbb{Z}, \tag{2.3}$$

where $\mathbf{X}^{(n)} = (\mathbf{X}_1, \dots, \mathbf{X}_n)'$ and $\mathbf{e}^{(n)} = (e_1, \dots, e_n)'$.

The following assumptions are made on $\boldsymbol{\theta}$ and g [(S1)–(S3)], and $\mathbf{X}^{(n)}$ [(G1)–(G6)], respectively.

(S1) The characteristic polynomials $\phi(z) =: \sum_{k=0}^{p} \phi_k z^k$ and $\eta(z) =: \sum_{k=0}^{q} \eta_k z^k$ have no roots within the unit disc $\mathscr{D} = \{z \in \mathbb{C}: |z| \leq 1\}$.
(S2) $0 < d < 1/2$; denote by Θ the set of all $\boldsymbol{\theta}$'s satisfying (S1)–(S2).
(S3) The innovation density g is such that $\int z g(z)\,dz = 0$ and $\int z^2 g(z)\,dz =: \sigma^2 < \infty$; moreover, g is absolutely continuous, with a.e. derivative g' satisfying

$$\begin{aligned} &0 < \mathscr{I}(g) =: \int [g'(z)/g(z)]^2 g(z)\,dz < \infty \quad \text{and} \\ &\int [g'(z)/g(z)]^4 g(z)\,dz < \infty. \end{aligned} \tag{2.4}$$

These assumptions imply the quadratic mean differentiability of $\sqrt{g}$. Also, note that (S3) entails $\mathrm{E}\,[g'(\varepsilon_t)/g(\varepsilon_t)] = 0$ and that the second part of condition (2.4) is not required when inference is restricted to the serial part of the model [i.e., when $\boldsymbol{\beta}$ is specified; see Swensen (1985)].

Under these assumptions, $\psi(z) =: \phi(z)\eta(z)^{-1}(1-z)^d$ admits the absolutely convergent development

$$\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k, \qquad z \in \mathscr{D}, \tag{2.5}$$

which implies that $\{e_t\}$ has the AR(∞) representation

$$\sum_{k=0}^{\infty} \psi_k e_{t-k} = \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{2.6}$$

Similarly, letting $\xi(z) =: \sum_{k=0}^{\infty} \xi_k z^k =: (\psi(z))^{-1}$, $z \in \mathscr{D}$, $\{e_t\}$ has the MA(∞) representation

$$e_t = \sum_{k=0}^{\infty} \xi_k \varepsilon_{t-k}, \qquad t \in \mathbb{Z}, \tag{2.7}$$

where $\{\xi_k\}$ is a square-summable sequence. Of course, the coefficients ψ_k and ξ_k in (2.5)–(2.7) depend on $\boldsymbol{\theta}$, with $\psi_0 = \xi_0 = 1$. The coefficients ψ_ν are differentiable, with gradient $\mathrm{grad}_{\boldsymbol\theta}(\psi_\nu) = (g_{\nu;\,1}, \dots, g_{\nu;\,p+q+1})'$, where

$$g_{\nu;\,j} = \begin{cases} \displaystyle\int_{-\pi}^{\pi} e^{-i\nu\lambda}\phi(e^{i\lambda})\left(\eta(e^{i\lambda})\right)^{-1}\left(1-e^{i\lambda}\right)^{d}\,\log(1-e^{i\lambda})\,d\lambda, & j=1\\ \displaystyle\int_{-\pi}^{\pi} e^{-i(\nu-j)\lambda}\left(\eta(e^{i\lambda})\right)^{-1}\left(1-e^{i\lambda}\right)^{d}\,d\lambda, & j=2,\dots,p+1\\ \displaystyle-\int_{-\pi}^{\pi} e^{-i(\nu-j)\lambda}\phi(e^{i\lambda})\left(\eta(e^{i\lambda})\right)^{-2}\left(1-e^{i\lambda}\right)^{d}\,d\lambda, & \\ & j=p+2,\dots,p+q+1, \end{cases} \tag{2.8}$$

and $\{e_t\}$ has spectral density

$$f_{\boldsymbol\theta}(\lambda) = \frac{\sigma^2\left|\sum_{k=0}^{q}\eta_k e^{ik\lambda}\right|^2}{2\pi\mid 1-e^{i\lambda}\mid^{2d}\left|\sum_{k=0}^{p}\phi_k e^{ik\lambda}\right|^2}\,. \tag{2.9}$$

On the regression constants $\mathbf{X}^{(n)}$, we impose a sort of Grenander's conditions [see, e.g., Hannan (1970)]. Letting

$$a_{kj}^{n}(h) = \begin{cases} \displaystyle\sum_{t=1}^{n-h} X_{t+h,\,k}X_{tj}, & h = \ \ 0,\,1,\dots,\\ \displaystyle\sum_{t=1-h}^{n} X_{t+h,\,k}X_{tj}, & h=-1,-2,\dots, \end{cases} \tag{2.10}$$

the following conditions are assumed to hold:

(G1) $a_{kk}^{n}(0)\to\infty$ as $n\to\infty$, $\quad k=1,\dots,b$.

(G2) $\lim_{n\to\infty}\ X_{n+1,k}^{2}/a_{kk}^{n}(0)=0$, $\quad k=1,\dots,b$.

(G3) $\lim_{n\to\infty}\ a_{kj}^{n}(h)/\{a_{kk}^{n}(0)a_{jj}^{n}(0)\}^{1/2}=r_{kj}(h)$ exists for every $k,\,j=1,\dots,b$ and $h\in\mathbb{Z}$; denote by $\tilde{\mathbf{R}}(h)$ the $b\times b$ matrix $\left(r_{kj}(h)\right)$.

(G4) $\tilde{\mathbf{R}}(0)$ is nonsingular; then there exists a Hermitian matrix function $\mathbf{M}(\lambda)=\left(M_{kj}(\lambda)\right)$ with positive semidefinite increments such that

$$\tilde{\mathbf{R}}(h) = \int_{-\pi}^{\pi} e^{ih\lambda}\,d\mathbf{M}(\lambda). \tag{2.11}$$

Still on $\mathbf{X}^{(n)}$, we also make the following two assumptions:

(G5) $a_{ll}^{n}(0) = O(n^{1+\alpha})$ for some $\alpha\geq 0$, and $\max_{1\leq t\leq n} X_{tl}^{2}/a_{ll}^{n}(0) = O(n^{-\delta})$, $l=1,\dots,b$, for some $\delta>1-2d$.

(G6) There exist $b_1,\,b_2\in\mathbb{N}$, $0\leq b_1\leq b_2\leq b$, such that X_{tk} is of the order of t^{j_k-1} (notation: $X_{tk}\sim t^{j_k-1}$), $k+1,\dots b_1$, for some sequence of integers $1\leq j_1\leq\cdots<j_{k-1}\leq j_k\leq\cdots\leq j_{b_1}$, that is, $0<\liminf_{t\to\infty}X_{tk}t^{-j_k+1}\leq\limsup_{t\to\infty}X_{tk}t^{-j_k+1}<\infty$; for the sake of simplicity, we assume in the sequel that $j_k=k$, $k=1,\dots,b_1$, so that this assumption takes the form

 (i) $X_{tk}\sim t^{k-1}$, $k=1,\dots,b_1$, which implies $M_{kk}(0^+)-M_{kk}(0)=1$, $k=1,\dots,b_1$; moreover,

 (ii) $0<M_{kk}(0^+)-M_{kk}(0)<1$, $k=b_1+1,\dots,b_2$ and

 (iii) $M_{kk}(0^+)-M_{kk}(0)=0$, $k=b_2+1,\dots,b$.

Assumption (G6) distinguishes three types of regressors, according to their asymptotic behaviors.

Letting $\mathbf{X}^{(m)}_{\cdot k} = (X_{1k}, \dots, X_{nk})^1$, $k = 1, \dots, b$, and

$$\tilde{\mathbf{D}}_n =: \operatorname{diag}\{n^{-d}\|\mathbf{X}_{\cdot 1}\|, \dots, n^{-d}\|\mathbf{X}_{\cdot b_1}\|, \|\mathbf{X}_{\cdot b_1+1}\|, \dots, \|\mathbf{X}_{\cdot b}\|\}$$

(here $\|\cdot\|$ denotes the Euclidean norm), define the *local* sequences

$$\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta} + n^{-1/2}\mathbf{h}, \qquad \boldsymbol{\beta}^{(n)} = \boldsymbol{\beta} + \tilde{\mathbf{D}}_n^{-1}\mathbf{k},$$

where $\mathbf{h}$ is a $(p+q+1)$-dimensional real vector, and $\mathbf{u} = (\mathbf{h}', \mathbf{k}')'$ belongs to some open subset $\mathscr{H}$ of $\mathbb{R}^{p+q+1}$. Also, denote by $\{\psi_k^{(n)}\}$ and $\{\xi_k^{(n)}\}$ the sequences resulting from substituting $\boldsymbol{\theta}^{(n)}$ for $\boldsymbol{\theta}$ in the definitions of ψ and ξ [(2.6) and (2.7)], respectively.

The sequence of statistical experiments under study is

$$\mathscr{E}_n = \left\{\mathbb{R}^{\mathbb{Z}}, \mathscr{B}^{\mathbb{Z}}, \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\beta}} \mid (\boldsymbol{\theta}, \boldsymbol{\beta}) \in \Theta \times \mathbb{R}^b\right\}\right\}, \qquad n \in \mathbb{N},$$

where $\mathscr{B}^{\mathbb{Z}}$ denotes the Borel σ-field on $\mathbb{R}^{\mathbb{Z}}$ and $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\beta}}$ the joint distribution $\mathscr{L}(\varepsilon_s,\ s \le 0;\ Y_1, \dots, Y_n)$ characterized by the parameter value $(\boldsymbol{\theta}, \boldsymbol{\beta})$ and the innovation density g. Denote by $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$ the sequence of simple hypotheses $\left\{\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\beta}}\},\ n \in \mathbb{N}\right\}$. The log-likelihood ratio for $H_g^{(n)}(\boldsymbol{\theta}^{(n)}, \boldsymbol{\beta}^{(n)})$ with respect to $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$ takes the form [under $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$]

$$\begin{aligned}
\Lambda_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) &=: \log \frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)},\boldsymbol{\beta}^{(n)}}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\beta}}} \\
&= \sum_{t=1}^{n} \Bigg\{ \log\ g\Bigg(\varepsilon_t + \sum_{\nu=0}^{t-1} \psi_\nu^{(n)}[Y_{t-\nu} - \mathbf{X}'_{t-\nu}\boldsymbol{\beta}^{(n)}] \\
&\qquad - \sum_{\nu=0}^{t-1} \psi_\nu[Y_{t-\nu} - \mathbf{X}'_{t-\nu}\boldsymbol{\beta}] \\
&\qquad + \sum_{r=0}^{\infty}\sum_{\mu=0}^{r}[\psi^{(n)}_{\mu+t}\xi^{(n)}_{r-\mu} - \psi_{\mu+t}\xi_{r-\mu}]\varepsilon_{-r}\Bigg) - \log\ g(\varepsilon_t)\Bigg\}.
\end{aligned} \tag{2.12}$$

The same log-likelihood ratio would have been obtained from *conditional* likelihoods [associated with the distribution of $(Y_1, \dots, Y_n)$ conditional upon the *starting values* $\{\varepsilon_t,\ t \le 0\}$], since the innovations ε_t have the same distribution under $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$ as under $H_g^{(n)}(\boldsymbol{\theta}^{(n)}, \boldsymbol{\beta}^{(n)})$. As we shall see, these starting values have no influence on the form (2.14) of the subsequent LAN result—hence of the asymptotic form of local experiments. Another way of handling the starting value problem is developed in Koul and Schick (1997), but is hardly applicable here due to the infinite-dimensional nature of initial values.

Define the $b \times b$ matrix

$$\mathbf{W}(\boldsymbol{\theta}) = \frac{1}{2\pi}\begin{pmatrix} \mathbf{W}_1(\boldsymbol{\theta}) & \mathbf{O} \\ \mathbf{O} & \mathbf{W}_2(\boldsymbol{\theta}) \end{pmatrix}, \tag{2.13}$$

where $\mathbf{W}_1(\boldsymbol{\theta})$ is the $b_1 \times b_1$ matrix with (k, j)th entry,

$$\frac{\Gamma(k-d)\Gamma(j-d)\{(2k-1)(2j-1)\}^{-1/2}}{(\sigma^2/2\pi)\,|\eta(1)/\phi(1)|^2\,\Gamma(k-2d)\Gamma(j-2d)(k+j-1-2d)}$$

$[\Gamma(\cdot)$ stands for the classical gamma function$]$ and $\mathbf{W}_2(\boldsymbol{\theta})$ is the $(b-b_1)\times(b-b_1)$ matrix with (k, j)th entry,

$$\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\lambda)^{-1}\, dM_{k+b_1,\, j+b_1}(\lambda).$$

We then have the following LAN result.

PROPOSITION 1 (LAN). *Suppose that* (S1)–(S3) *and* (G1)–(G6) *hold. Then the sequence of experiments* $\mathscr{E}_n$, $n \in \mathbb{N}$, *is locally asymptotically normal* (LAN) *and equicontinuous on compact subsets* $\mathscr{C}$ *of* $\mathscr{H}$. *That is*:

(i) *For all* $\boldsymbol{\theta}, \boldsymbol{\beta}$, *the log-likelihood ratio* (2.12) *admits, under* $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$, *as* $n \to \infty$, *the asymptotic representation*

$$\begin{aligned}\Lambda_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) &= (\mathbf{h}', \mathbf{k}')\boldsymbol{\Delta}_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) \\ &\quad -\tfrac{1}{2}\left[\sigma^2 \mathscr{I}(g)\mathbf{h}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{h} + \mathscr{I}(g)\mathbf{k}'\mathbf{W}(\boldsymbol{\theta})\mathbf{k}\right] + o_P(1),\end{aligned} \tag{2.14}$$

with the $(b+p+q+1)$-*dimensional random vector* (*the* central sequence)

$$\boldsymbol{\Delta}_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) =: n^{-1/2}\begin{pmatrix} \displaystyle\sum_{t=1}^{n} \frac{g'(Z_t)}{g(Z_t)} \sum_{\nu=1}^{t-1} \mathrm{grad}_{\boldsymbol{\theta}}(\psi_\nu)\, e_{t-\nu} \\ \displaystyle -\tilde{\mathbf{D}}_n^{-1} \sum_{t=1}^{n} \frac{g'(Z_t)}{g(Z_t)} \sum_{\nu=0}^{t-1} \psi_\nu \mathbf{X}_{t-\nu} \end{pmatrix}, \tag{2.15}$$

the $(p+q+1) \times (p+q+1)$ *matrix*,

$$\mathbf{Q}(\boldsymbol{\theta}) =: (4\pi)^{-1}\int_{-\pi}^{\pi} \mathrm{grad}_{\boldsymbol{\theta}}(\log f_{\boldsymbol{\theta}}(\lambda))\,\mathrm{grad}'_{\boldsymbol{\theta}}(\log f_{\boldsymbol{\theta}}(\lambda))\, d\lambda, \tag{2.16}$$

and the $b \times b$ *matrix* $\mathbf{W}(\boldsymbol{\theta})$ *defined in* (2.13); Z_t, $t = 1, \dots, n$ *stands for the approximate residual* $Z_t(\boldsymbol{\theta}, \boldsymbol{\beta}) =: \sum_{k=0}^{t-1} \psi_k (Y_{t-k} - \mathbf{X}'_{t-k}\boldsymbol{\beta})$, *and* $\mathrm{grad}'_{\boldsymbol{\theta}}(\psi_\nu)$ *is given in* (2.8).

(ii) *Still under* $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$, *as* $n \to \infty$, $\boldsymbol{\Delta}_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$ *is asymptotically normal, with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{\Gamma}_g(\boldsymbol{\theta}, \boldsymbol{\beta})$, *where*

$$\boldsymbol{\Gamma}_g(\boldsymbol{\theta}, \boldsymbol{\beta}) =: \begin{pmatrix} \sigma^2 \mathscr{I}(g)\mathbf{Q}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathscr{I}(g)\mathbf{W}(\boldsymbol{\theta}) \end{pmatrix}. \tag{2.17}$$

(iii) *For all* $n \in \mathbb{N}$ *and all* $(\boldsymbol{\theta}, \boldsymbol{\beta}) \in \mathscr{H}$, *the mapping* $\mathbf{u} = (\mathbf{h}, \mathbf{k}) \mapsto \mathrm{P}^{(n)}_{\boldsymbol{\theta}^{(n)},\, \boldsymbol{\beta}^{(n)}}$ *is continuous with respect to the variational distance.*

See Section 4 for the proof.

3. Applications.

3.1. *Hypothesis testing.* As mentioned in the introduction, Proposition 1 is the key result for virtually all problems in asymptotic inference connected with the FARIMA model under study. We illustrate this fact with some examples. For a general theory on locally asymptotically optimal testing in LAN families, the reader is referred to Strasser (1985) or Le Cam (1986). The following is borrowed from Strasser (1985).

Let $\mathscr{M}_0$ denote the intersection of $\mathscr{H}$ with a linear subspace of $\mathbb{R}^{p+q+1}$, equipped with the norm $\|\cdot\|_{\boldsymbol{\Gamma}_g}$ associated with the covariance matrix (2.17); namely, $\|\boldsymbol{\xi}\|_{\boldsymbol{\Gamma}_g} =: \boldsymbol{\xi}'\boldsymbol{\Gamma}_g^{-1}\boldsymbol{\xi}$. We consider the problem of testing the null hypothesis under which $(\boldsymbol{\theta}, \boldsymbol{\beta}) \in \mathscr{M}_0$ against the alternative $(\boldsymbol{\theta}, \boldsymbol{\beta}) \notin \mathscr{M}_0$. Denote by $\boldsymbol{\Pi}_{\mathscr{M}_0;\boldsymbol{\Gamma}_g}$ the projection matrix (projection here means the orthogonal projection associated with the norm $\|\cdot\|_{\boldsymbol{\Gamma}_g}$) mapping $\mathbb{R}^{b+p+q+1}$ onto $\mathscr{M}_0$, and by $\mathbf{I}$ the identity map (in $\mathbb{R}^{b+p+q+1}$). Put $\mathscr{B}_c =: \{\mathbf{u} \in \mathscr{H} \mid \|[\mathbf{I} - \boldsymbol{\Pi}_{\mathscr{M}_0;\boldsymbol{\Gamma}_g}]\mathbf{u}\|_{\boldsymbol{\Gamma}_g} = c\}$, $c > 0$. Also, let $\mathrm{E}^{(n)}_{\mathbf{u}}$ stand for the expectation under $H^{(n)}_g((\boldsymbol{\theta}, \boldsymbol{\beta}) + n^{-1/2}\mathbf{u})$. A sequence of tests φ_n, $n \in \mathbb{N}$ is said to be *asymptotically unbiased* at level $\alpha \in [0, 1]$ for the testing problem just described if

$$\limsup_{n\to\infty} \mathrm{E}^{(n)}_{\mathbf{u}}[\varphi_n] \le \alpha, \qquad \mathbf{u} \in \mathscr{M}_0,$$

$$\liminf_{n\to\infty} \mathrm{E}^{(n)}_{\mathbf{u}}[\varphi_n] \ge \alpha, \qquad \mathbf{u} \notin \mathscr{M}_0.$$

PROPOSITION 2. *Suppose that* (S1)–(S3) *and* (G1)–(G6) *hold. Let* $\alpha \in [0, 1]$ *and choose* $k_\alpha \in [0, \infty)$ *such that*

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\beta}}\left[\left\|\left[\mathbf{I} - \boldsymbol{\Pi}_{\mathscr{M}_0;\boldsymbol{\Gamma}_g}\right]\boldsymbol{\Delta}^{(n)}_g(\boldsymbol{\theta}, \boldsymbol{\beta})\right\|_{\boldsymbol{\Gamma}_g} > k_\alpha\right] = \alpha.$$

Then:

(i) *the sequence of tests*

$$\varphi^*_n = \begin{cases} 1, & \textit{if } \left\|[\mathbf{I} - \boldsymbol{\Pi}_{\mathscr{M}_0;\boldsymbol{\Gamma}_g}]\boldsymbol{\Delta}^{(n)}_g(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)})\right\|_{\boldsymbol{\Gamma}_g} > k_\alpha, \\ 0, & \textit{if } \left\|[\mathbf{I} - \boldsymbol{\Pi}_{\mathscr{M}_0;\boldsymbol{\Gamma}_g}]\boldsymbol{\Delta}^{(n)}_g(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)})\right\|_{\boldsymbol{\Gamma}_g} \le k_\alpha, \end{cases} \tag{3.1}$$

where $\hat{\boldsymbol{\theta}}^{(n)}$ *and* $\hat{\boldsymbol{\beta}}^{(n)}$ *are* locally discrete [*see*, *for instance*, *Le Cam and Yang* (1990), *page* 60] *estimators such that* $n^{1/2}\left(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right)$ *and* $\tilde{\mathbf{D}}_n\left(\hat{\boldsymbol{\beta}}^{(n)} - \boldsymbol{\beta}\right)$ *are* $O_P(1)$, *as* $n \to \infty$ *under the null hypothesis* [*note that they also play a role in* *the definition of the matrix* $\mathbf{\Gamma}_g = \mathbf{\Gamma}_g(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)})$] *is asymptotically unbiased at level* α *for the problem under study.*

(ii) $\{\varphi_n^*,\ n \in \mathbb{N}\}$ *is locally asymptotically maximin, in the sense that, denoting by* $\{\varphi_n,\ n \in \mathbb{N}\}$ *any other sequence of asymptotically unbiased tests at level* α *for the same problem,*

$$\limsup_{n\to\infty} \inf_{\mathbf{u}\in\mathscr{B}_c} \mathrm{E}_{\mathbf{u}}^{(n)}[\varphi_n] \le \lim_{n\to\infty} \inf_{\mathbf{u}\in\mathscr{B}_c} \mathrm{E}_{\mathbf{u}}^{(n)}[\varphi_n^*]. \tag{3.2}$$

The result readily follows from Theorem 82.21 in Strasser (1985).

Note that, when $\mathscr{M}_0$ is the linear space spanned by the columns of a full-rank matrix $\mathbf{B}$, the test statistic in (3.1) takes the form

$$\begin{aligned}
&\left\| [\mathbf{I} - \mathbf{\Pi}_{\mathscr{M}_0;\mathbf{\Gamma}_g}] \mathbf{\Delta}_g^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)}) \right\|_{\mathbf{\Gamma}_g} \\
&\quad = \mathbf{\Delta}_g^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)})' \left[\mathbf{\Gamma}_g^{-1}(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)}) - \mathbf{B}' \left(\mathbf{B}'\mathbf{\Gamma}_g(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)})\mathbf{B} \right)^{-1} \mathbf{B}' \right] \\
&\quad\quad \times \mathbf{\Delta}_g^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \hat{\boldsymbol{\beta}}^{(n)}).
\end{aligned}$$

The local and asymptotic optimality property (ii) of φ_n^* also could be described in terms of local asymptotic *stringency*; see Le Cam [(1986), Chapter 11.9, Corollaries 2 and 3].

Let us give an explicit example. Let the innovation density g be Gaussian; the process $\{e_t\}$ then is completely characterized by its spectral density (2.9). For simplicity, set $\sigma^2 = 1$ (in case σ^2 is unknown, as it usually is in practice, it always can be replaced by its empirical residual version, based on the estimates provided below). The assumptions of Proposition 2 are assumed to hold, with $b_2 = 0$ in (G6). From Theorem 2.4 of Yajima (1991), the BLUE $\hat{\boldsymbol{\beta}}_{\mathrm{BL}}^{(n)} = \left(\mathbf{X}^{(n)\prime}\mathbf{\Sigma}_n^{-1}(\boldsymbol{\theta})\mathbf{X}^{(n)}\right)^{-1}\mathbf{X}^{(n)\prime}\mathbf{\Sigma}_n^{-1}(\boldsymbol{\theta})\mathbf{Y}^{(n)}$, where $\mathbf{\Sigma}_n(\boldsymbol{\theta})$ is the covariance matrix of $\mathbf{e}^{(n)}$, and the LSE $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} = \left(\mathbf{X}^{(n)\prime}\mathbf{X}^{(n)}\right)^{-1}\mathbf{X}^{(n)\prime}\mathbf{Y}^{(n)}$ are asymptotically equivalent (namely, $\tilde{\mathbf{D}}_n(\hat{\boldsymbol{\beta}}_{\mathrm{BL}}^{(n)} - \boldsymbol{\beta}) = \tilde{\mathbf{D}}_n(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}) + o_P(1)$), as $n \to \infty$. The MLE $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)}$ of $\boldsymbol{\theta}$ is obtained by maximizing

$$\ell_n(\boldsymbol{\theta}, \boldsymbol{\beta}) = -\tfrac{1}{2}\log|\mathbf{\Sigma}_n(\boldsymbol{\theta})| - \tfrac{1}{2}\left(\mathbf{Y}_n - \mathbf{X}_n\hat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}\right)' \mathbf{\Sigma}_n^{-1}(\boldsymbol{\theta}) \left(\mathbf{Y}_n - \mathbf{X}_n\hat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)}\right)$$

with respect to $\boldsymbol{\theta}$. As in Dahlhaus (1989),we can show that $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)} \to_P \boldsymbol{\theta}$, and that the mth element of $n^{1/2}\mathbf{Q}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)} - \boldsymbol{\theta})$ satisfies

$$n^{1/2}\left(\mathbf{Q}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)} - \boldsymbol{\theta})\right)_m = n^{-1/2}\left[\mathbf{e}^{(n)\prime}\mathbf{A}^m\mathbf{e}^{(n)} - \mathrm{tr}(\mathbf{\Sigma}_n(\boldsymbol{\theta}))\,\mathbf{A}^m(\boldsymbol{\theta})\right] + o_P(1),$$

where $\mathbf{A}^m = \mathbf{A}^m(\boldsymbol{\theta})$ is the $n \times n$ Toeplitz matrix with entries

$$A_{k,j}^m = \frac{1}{8\pi^2}\int_{-\pi}^{\pi} e^{i(k-j)\lambda}\frac{(\partial/\partial\,\theta_m)f_{\boldsymbol{\theta}}(\lambda)}{f_{\boldsymbol{\theta}}^2(\lambda)}\,d\lambda.$$

Denote by $\mathscr{M}(\mathbf{B})$ the linear space spanned by the columns of a matrix $\mathbf{B}$. The problem consists in testing the null hypothesis $H_0^{(n)}$ under which

$$(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \in \mathscr{M}(\mathbf{B}_1) \quad \text{and} \quad (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \in \mathscr{M}(\mathbf{B}_2)$$

for some given $(p+q+1)\times(p+q+1-\ell_1)$ and $b\times(b-\ell_2)$ matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ of maximal ranks, and given vectors $\boldsymbol{\theta}_0 \in \mathbb{R}^{p+q+1}$ and $\boldsymbol{\beta}_0 \in \mathbb{R}^b$, respectively. Then Proposition 2 implies that the test rejecting $H_0^{(n)}$ whenever the test statistic

$$\begin{aligned} T_n = n(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)} - \boldsymbol{\theta}_0)' &\left\{\mathbf{Q}(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)}) - \mathbf{Q}(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)})\mathbf{B}_1\left[\mathbf{B}_1'\mathbf{Q}(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)})\mathbf{B}_1\right]^{-1}\mathbf{B}_1'\mathbf{Q}(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)})\right\} \\ &\times(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(n)} - \boldsymbol{\theta}_0) \\ &+(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}_0)'\tilde{\mathbf{D}}_n\left\{\frac{\mathbf{W}_2}{2\pi} - \tilde{\mathbf{D}}_n\mathbf{B}_2\left[\mathbf{B}_2'\tilde{\mathbf{D}}_n\frac{\mathbf{W}_2}{2\pi}\tilde{\mathbf{D}}_n\mathbf{B}_2\right]^{-1}\mathbf{B}_2'\tilde{\mathbf{D}}_n\right\} \\ &\times\tilde{\mathbf{D}}_n(\hat{\boldsymbol{\beta}}_{\mathrm{LS}}^{(n)} - \boldsymbol{\beta}_0) \end{aligned} \tag{3.3}$$

exceeds the α-quantile $\chi^2_{\ell;\alpha}$ of a chi-square distribution with $\ell = \ell_1+\ell_2$ degrees of freedom, has asymptotic level α and is locally asymptotically optimal in the sense of (3.2).

3.2. *Discriminant analysis.* Before turning to the discriminant analysis problem, we establish a general result on the asymptotic distribution of a class of statistics which plays a central role in this specific context, but also in a variety of testing and estimation problems. Let $\mathbf{B} = (B_{k,\ell})$ denote the $n \times n$ matrix with elements

$$B_{k,\ell} = \int_{-\pi}^{\pi} e^{i(k-\ell)\lambda} f_{bb}(\lambda)\, d\lambda,$$

where f_{bb} is a real, even integrable function defined on $[-\pi, \pi]$, such that

$$\frac{1}{n}\operatorname{tr}\left([\mathbf{B}\boldsymbol{\Sigma}_n(\boldsymbol{\theta})]^2\right) \longrightarrow \frac{1}{2\pi}\int_{-\pi}^{\pi}[f_{bb}(\lambda)f_{\boldsymbol{\theta}}(\lambda)]^2\, d\lambda < \infty \tag{3.4}$$

as $n \to \infty$. Let $\boldsymbol{\alpha}_n' = (\alpha_1, \dots, \alpha_n)$ be a nonrandom vector satisfying

(T1) $n^{-1/2}\sum_{r=1}^n \alpha_r \longrightarrow 0$;
(T2) There exists a spectral measure $M_\alpha(\lambda)$ such that

$$\lim_{n\to\infty}\sum_{j=1}^n \alpha_j\alpha_{j+s} = \int_{-\pi}^{\pi} e^{is\lambda}\, dM_\alpha(\lambda)$$

and

$$\boldsymbol{\alpha}_n'\boldsymbol{\Sigma}_n(\boldsymbol{\theta})\boldsymbol{\alpha}_n \to \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\lambda)\, dM_\alpha(\lambda) =: c(\boldsymbol{\theta}). \tag{3.5}$$

The class of statistics we are considering here (actually, a class of *sequences* of statistics) is

$$\mathscr{T}^{(n)} = \left\{ T^{(n)} \;\middle|\; T^{(n)} = n^{-1/2} \sum_{k=1}^{n} \sum_{\ell=1}^{n} [e_k e_\ell - R(k-\ell)]\, B_{k,\ell} + \boldsymbol{\alpha}_n' \mathbf{e}_n + o_P(1) \right\}, \tag{3.6}$$

where the $o_P(1)$ term is under some sequence $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$, as $n \to \infty$, and $R(k-\ell) = \mathrm{E}\,[e_k e_\ell]$ [expectations, here and below, are taken under $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$].

The motivation for considering this class $\mathscr{T}^{(n)}$ is as follows. For simplicity, assume that $\boldsymbol{\beta} = \mathbf{0}$ and that the parameter $\boldsymbol{\theta}$ is a scalar θ. Consider the discriminant analysis problem under which the observation $\mathbf{Y}^{(n)}$ is to be assigned either to population Π_1, described by the spectral density $f_{\theta_1}(\lambda)$, or to population Π_2, described by the spectral density $f_{\theta_2}(\lambda)$. The following approximate Gaussian log-likelihood ratio formally can be constructed for this problem [see Dahlhaus (1989)]:

$$\mathrm{GLR} =: \frac{n}{4\pi} \int_{-\pi}^{\pi} \left[\log \frac{f_{\theta_2}(\lambda)}{f_{\theta_1}(\lambda)} + I_n(\lambda) \left(\frac{1}{f_{\theta_2}(\lambda)} - \frac{1}{f_{\theta_1}(\lambda)} \right) \right] d\lambda,$$

where $I_n(\lambda) =: (2\pi n)^{-1} |\sum_{t=1}^{n} Y_t \exp(-it\lambda)|^2$. We can use GLR as a discriminant criterion: namely, select Π_1 or Π_2 according as GLR > 0 or GLR ≤ 0. A measure of performance of GLR as a discriminant criterion is the probability of misclassification when Π_1 and Π_2 are *close* to each other, or the asymptotic behavior of this probability when Π_2 is contiguous to Π_1: for instance, under $\theta_2 = \theta_1 + n^{-1/2} h$. Expanding the corresponding GLR around θ_1 yields a linear term which belongs to $\mathscr{T}^{(n)}$, with $\boldsymbol{\alpha} = \mathbf{0}$.

Letting $S_t = g'(\varepsilon_t)/g(\varepsilon_t)$, also define

$$R_{Se}(\ell) = \mathrm{E}[S_t e_{t+\ell}] \quad \text{and} \quad C_{See}(\ell, m) = \mathrm{cum}\{S_t, e_{t+\ell}, e_{t+m}\}$$

[for the definition of the joint cumulants $\mathrm{cum}\{\cdots\}$, see Brillinger (1981)]. The corresponding cross-spectral ($f_{Se}(\lambda)$) and cumulant spectral ($f_{See}(\lambda, \mu)$) densities are characterized by

$$R_{Se}(\ell) = \int_{-\pi}^{\pi} e^{i\ell\lambda} f_{Se}(\lambda)\, d\lambda$$

and

$$C_{See}(\ell, m) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i[\ell\lambda + m\mu]} f_{See}(\lambda, \mu)\, d\lambda\, d\mu,$$

respectively. Finally, define $\tilde{f}_{\psi X}(\lambda)$ and $\tilde{f}_{\alpha\psi X}(\lambda, \mu, \eta)$ by means of the limits

$$\int_{-\pi}^{\pi} \tilde{f}_{\psi X}(\lambda)\, d\lambda = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} n^{-1/2} \sum_{\nu=1}^{n} \psi_\nu e^{-i\nu\lambda} \mathbf{k}' \tilde{\mathbf{D}}_n^{-1} \sum_{\ell=1}^{n} \mathbf{X}_\ell e^{-i\ell\lambda}\, d\lambda$$

and

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\tilde{f}_{\alpha\psi X}(\lambda,\mu,\eta)\,d\lambda\,d\mu\,d\eta$$
$$=\lim_{n\to\infty}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\sum_{\ell=1}^{n}\alpha_{\ell}e^{i\ell\lambda}\right]$$
$$\times\left[\sum_{j=1}^{n}\sum_{\nu=1}^{j-1}\left(\psi_{\nu}e^{-i\nu\mu}\right)\mathbf{k}'\tilde{\mathbf{D}}_n^{-1}\mathbf{X}_{j-\nu}e^{-i(j-\nu)\eta}\right]d\lambda\,d\mu\,d\eta.$$

The notation $\tilde{f}^{(1)}_{\psi X}$, etc. will be used in an obvious way [instead of $\tilde{f}_{\psi^{(1)}X}$, etc.] when the sequence $\psi^{(1)}_{\nu}=\mathbf{h}'\operatorname{grad}_{\boldsymbol{\theta}}(\psi_{\nu})$ is substituted for ψ_{ν}.

PROPOSITION 3. *Assume that* (S1), (S2), (G1)–(G6), (T1), (T2) *and* (3.4) *hold. Then, under the sequence of alternatives $H^{(n)}_g(\boldsymbol{\theta}^{(n)},\boldsymbol{\beta}^{(n)})$, as $n\to\infty$, the distribution of $T^{(n)}\in\mathscr{T}^{(n)}$ is asymptotically normal $\mathscr{N}(m,\,v^2)$, with*

$$\begin{aligned}m&=8\pi^2\textstyle\int_{-\pi}^{\pi}\left[\sum_{\nu=1}^{\infty}\mathbf{h}'\operatorname{grad}_{\boldsymbol{\theta}}(\psi_{\nu})e^{i\lambda\nu}\right]f_{bb}(\lambda)f_{\boldsymbol{\theta}}(\lambda)f_{Se}(-\lambda)\,d\lambda\\&\quad+4\pi^2\textstyle\int_{-\pi}^{\pi}\sum_{j=1}^{\infty}e^{ij\lambda}\tilde{f}_{\psi X}(\lambda)\,d\lambda\int_{-\pi}^{\pi}f_{bb}(\eta)f_{See}(-\eta,\eta)\,d\eta\\&\quad+2\pi\textstyle\int_{-\pi}^{\pi}\tilde{f}^{(1)}_{\psi\alpha}(\lambda)f_{See}(\lambda,0)\,d\lambda+\int_{-\pi}^{\pi}\tilde{f}_{\alpha\psi X}(\lambda,\lambda,-\lambda)f_{Se}(\lambda)\,d\lambda\\&=m_1+m_2+m_3+m_4\quad\textit{say},\end{aligned}\tag{3.7}$$

and

$$v^2=16\pi^3\int_{-\pi}^{\pi}[f_{bb}(\lambda)f_{\boldsymbol{\theta}}(\lambda)]^2\,d\lambda+\mu_4\left[2\pi\int_{-\pi}^{\pi}f_{bb}(\lambda)f_{\boldsymbol{\theta}}(\lambda)\,d\lambda\right]^2+c(\boldsymbol{\theta}),$$

where μ_4 denotes the fourth cumulant of g, and $c(\boldsymbol{\theta})$ is given in (3.5).

See Section 4 for the proof.

We now describe in some detail the discriminant analysis problem to be solved. Let $b(\lambda)$ denote a real, even, integrable function, defined over $[-\pi,\,\pi]$. Assuming again that g is Gaussian, define

$$T^{(n)}=\frac{n^{1/2}}{4\pi}\int_{-\pi}^{\pi}b(\lambda)\frac{I_n(\lambda)-f_{\boldsymbol{\theta}}(\lambda)}{f_{\boldsymbol{\theta}}(\lambda)}\,d\lambda+\mathbf{k}'\tilde{\mathbf{D}}_n^{-1}\mathbf{X}'\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\theta})\mathbf{e}_n,$$

with $I_n(\lambda)=(2\pi n)^{-1}\left|\sum_{t=1}^{n}e_t\ \exp\{-it\lambda\}\right|^2$. This statistic appears in the discriminant analysis problem for time series. Consider indeed the case when the observed series $\{\mathbf{Y}_t,\ t=1,\dots,n\}$, defined in (2.1), belongs to one of the two types (sequences of models) $\Pi^{(n)}_1$ or $\Pi^{(n)}_2$ described by

$\Pi^{(n)}_1$: residual spectral density $f_1(\lambda)$, regression coefficients $\boldsymbol{\beta}_1$,

$\Pi^{(n)}_2$: residual spectral density $f_2(\lambda)$, regression coefficients $\boldsymbol{\beta}_2$.

Then, $T^{(n)}$ provides an approximation of the log-likelihood ratio under the conditions

$$f_1(\lambda) = f_{\boldsymbol{\theta}}(\lambda),\ \ f_2(\lambda) = f_1(\lambda)(1 + n^{-1/2} b(\lambda)) \text{ and } \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = \tilde{\mathbf{D}}_n^{-1}\mathbf{k}$$

[see Shumway and Unger (1974), Zhang and Taniguchi (1994) and Taniguchi (1998) for the short-memory case]. Obviously, $T^{(n)} \in \mathscr{T}^{(n)}$. Hence, Proposition 3 allows for evaluating the asymptotic probabilities of misclassification associated with the rule assigning the observed series to $\Pi_1^{(n)}$ or $\Pi_2^{(n)}$ according as $T^{(n)}$ is larger or smaller than zero. Indeed, we obtain that $T^{(n)}$ is asymptotically normal, as $n \to \infty$, with mean and variance

$$m = \int_{-\pi}^{\pi} \left[\sum_{\nu=1}^{\infty} \mathbf{h}' \operatorname{grad}'_{\boldsymbol{\theta}}(\psi_\nu) e^{i\lambda\nu} \right] b(\lambda) f_{Se}(-\lambda)\, d\lambda + \int_{-\pi}^{\pi} \tilde{f}_{\alpha\psi Z}(\lambda, \lambda, -\lambda) f_{Se}(\lambda)\, d\lambda$$

and

$$v^2 = \frac{1}{4\pi} \int_{-\pi}^{\pi} [b(\lambda)]^2\, d\lambda + \frac{1}{2\pi}\mathbf{k}' \int_{-\pi}^{\pi} [f_{\boldsymbol{\theta}}(\lambda)]^{-1}\, dM(\lambda)\mathbf{k},$$

respectively, under $\Pi_2^{(n)}$. The asymptotic misclassification probability is thus $\Phi(-m/v)$, where Φ, as usual, stands for the standard normal distribution function.

3.3. *Adaptive estimation.* Another application of the LAN property is adaptive estimation. For the sake of simplicity, we concentrate on adaptive estimation of $\boldsymbol{\theta} = (d,\ \phi)'$ in the FARIMA$(1, d, 0)$ model

$$(1 + \phi L)(1 - L)^d X_t = \varepsilon_t\,, \qquad t \in \mathbb{Z}, \tag{3.8}$$

where $\{\varepsilon_t\,,\ t \in \mathbb{Z}\}$ is an i.i.d. sequence with unspecified probability density g; though, of course, the more general case (2.1) and (2.2) could be considered as well. We provide two distinct versions, according as a symmetry assumption can be made on g or not. The methods we describe are adapted from Koul and Schick (1997), which we refer to for details.

The LAN property for model (3.8) follows from Proposition 2.1, with [same notation as in (2.14)]

$$\Lambda_g^{(n)}(d,\ \phi) = \mathbf{h}'\boldsymbol{\Delta}_g^{(n)}(d,\ \phi) - \tfrac{1}{2}\sigma^2 \mathscr{I}(g)\mathbf{h}'\mathbf{Q}(d,\ \phi)\mathbf{h} + o_P(1),$$

where

$$\psi_\nu(d,\ \phi) = \left(1 - \phi - \frac{1+d}{\nu}\right) \frac{\Gamma(\nu - 1 - d)}{\Gamma(-d)\Gamma(\nu)},$$

$$\operatorname{grad}_{d,\phi}(\psi_\nu) = \begin{cases} (1,\ 1)', & \nu = 1, \\ -\dfrac{\Gamma(\nu-1-d)}{\Gamma(-d)\Gamma(\nu)} \left(1,\ \dfrac{1}{\nu} + 1 - \phi - \dfrac{1+d}{\nu} \displaystyle\sum_{i=0}^{\nu-2} \dfrac{1}{i-d}\right)', & \nu \geq 2, \end{cases}$$

$$\mathbf{Q}(d,\ \phi) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \operatorname{grad}_{d,\phi}(\log f_{d,\phi}(\lambda)) \operatorname{grad}'_{d,\phi}(\log f_{d,\phi}(\lambda))\, d\lambda,$$

with

$$f_{d,\phi}(\lambda) =: \frac{\sigma^2}{2\pi} \frac{|1-e^{i\lambda}|^{-2d}}{|1-\phi e^{i\lambda}|^2} = \frac{\sigma^2}{2\pi} \frac{(2\ \sin(\lambda/2))^{-2d}}{1+\phi^2 - 2\ \phi \cos(\lambda)}$$

and

$$\boldsymbol{\Delta}_g^{(n)}(d,\ \phi) = n^{-1/2} \sum_{t=1}^{n} \frac{g'(Z_t)}{g(Z_t)} \sum_{\nu=1}^{t-1} \mathrm{grad}_{d,\,\phi}(\psi_\nu)\, X_{t-\nu}. \tag{3.9}$$

Putting $\varphi_g =: -g'/g$,

$$\begin{aligned} H_t(d,\ \phi) &=: \sum_{\nu=1}^{t-1} \psi_\nu(d,\ \phi) X_{t-\nu}, \text{ and} \\ \dot{\mathbf{H}}_t(d,\ \phi) &=: \sum_{\nu=1}^{t-1} \mathrm{grad}_{d,\,\phi}(\psi_\nu)\, X_{t-\nu}, \end{aligned} \tag{3.10}$$

the residuals $Z_t^{(n)}$ and the central sequence take the form

$$Z_t^{(n)}(d,\ \phi) = X_t - \sum_{\nu=1}^{t-1} \psi_\nu(d,\ \phi) X_{t-\nu} = X_t - H_t(d,\ \phi) \tag{3.11}$$

and

$$\boldsymbol{\Delta}_g^{(n)}(d,\ \phi) = -n^{-1/2} \sum_{t=1}^{n} \dot{\mathbf{H}}_t(d,\ \phi) \varphi_g(Z_t^{(n)}(d,\ \phi)),$$

respectively.

Before turning to the construction of adaptive estimates, we first show that some of the technical conditions involved in Koul and Schick's results hold under our model.

LEMMA 3.1. *For all local sequence $\boldsymbol{\theta}^{(n)}$ such that $\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta} = O(n^{-1/2})$, the following properties hold, under $H^{(n)}(d, \phi)$, as $n \to \infty$:*

(i) $\sum_{t=1}^{n} \left| H_t(\boldsymbol{\theta}^{(n)}) - H_t(\boldsymbol{\theta}) - (\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta})' \dot{\mathbf{H}}_t(\boldsymbol{\theta}) \right|^2 = o_P(1)$;
(ii) $\max_{1 \le t \le n} \|\dot{\mathbf{H}}_t(\boldsymbol{\theta})\| = o_P(n^{1/2})$;
(iii) $(1/n) \sum_{t=1}^{n} \dot{\mathbf{H}}_t(\boldsymbol{\theta}) = o_P(1)$;
(iv) $(1/n) \sum_{t=1}^{n} \dot{\mathbf{H}}_t(\boldsymbol{\theta}) \left(\dot{\mathbf{H}}_t(\boldsymbol{\theta})\right)' = \mathbf{Q}(\boldsymbol{\theta}) + o_P(1)$;
(v) $(1/n) \sum_{t=1}^{n} \left\| \dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)}) - \dot{\mathbf{H}}_t(\boldsymbol{\theta}) \right\|^2 = o_P(1)$;
(vi) *For any positive sequence $c_n \to \infty$,*

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)}) \right\|^2 I\left[\left\| \dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)}) \right\| > c_n \right] = o_P(1);$$

(vii) *For any positive sequence $a_n \to 0$, there exists a sequence of positive integers $m_n \to \infty$ such that*

$$n^{-1} \sum_{\substack{1 \le t,\ i \le n \\ |t-i| > m_n}} \mathrm{E}_{\boldsymbol{\theta}^{(n)}} \left[\left\| \dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)}) - \mathrm{E}_{\boldsymbol{\theta}^{(n)}} \left[\dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)}) \middle| \mathscr{B}_{n,\,i}(\boldsymbol{\theta}^{(n)}) \right] \right\|^2 \right] = o(a_n^2)$$

as $n \to \infty$, *where* $\mathscr{B}_{n,i}(\boldsymbol{\theta})$ *denotes the* σ*-algebra generated by*

$$((\varepsilon_t,\ t \le 0), Z_1(\boldsymbol{\theta}), \dots, Z_{i-1}(\boldsymbol{\theta}), Z_{i+1}(\boldsymbol{\theta}), \dots, Z_n(\boldsymbol{\theta})).$$

See Section 4 for the proof.

Let $a_n \to 0$, $b_n \to 0$, $c_n \to \infty$, and $d_n = o(n)$ denote three sequences of positive real numbers and one sequence of positive integers (all convergences are as $n \to \infty$). Denote by $\boldsymbol{\theta}_0^{(n)}$ a preliminary, root-n consistent estimate of $\boldsymbol{\theta} = (d, \phi)'$, by $\hat{\psi}_\nu^{(n)}$ and $\hat{Z}_t^{(n)} = Z_t^{(n)}(\boldsymbol{\theta}_0^{(n)})$ the corresponding estimated residuals; let $\hat{\mathbf{Z}}^{(n)} =: \left(\hat{Z}_{d_n}^{(n)}, \hat{Z}_{d_n+1}^{(n)}, \dots, \hat{Z}_n^{(n)}\right)$.

First, consider the case under which the unknown g can be assumed to be symmetric. Choosing the sequences a_n, b_n and c_n such that

$$n^{-1}a_n^{-3}b_n^{-1} \to 0 \quad \text{as } n \to \infty, \tag{3.12}$$

define the estimator of the score function φ by

$$\hat{\varphi}^{(n)}(z) =: -\frac{f'_{N_n}\left(z, \hat{\mathbf{Z}}^{(n)}\right) - f'_{N_n}\left(-z, \hat{\mathbf{Z}}^{(n)}\right)}{b_n + f_{N_n}\left(z, \hat{\mathbf{Z}}^{(n)}\right) + f_{N_n}\left(-z, \hat{\mathbf{Z}}^{(n)}\right)}, \tag{3.13}$$

with $N_n = n - d_n + 1$ and

$$\begin{aligned} f_n(z, y_1, \dots, y_n) &=: \frac{1}{na_n}\sum_{t=1}^n k\left(\frac{z - y_t}{a_n}\right), \\ f'_n(z, y_1, \dots, y_n) &=: \frac{1}{na_n^2}\sum_{t=1}^n k'\left(\frac{z - y_t}{a_n}\right), \end{aligned} \tag{3.14}$$

where the kernel k is assumed to satisfy [see Schick (1993)]

(K) The kernel k is three times continuously differentiable, with derivatives $k^{(i)}$ satisfying $|k^{(i)}(z)| \le ck(z)$, $i = 1, 2, 3$ for some positive c, and $\int_{-\infty}^{\infty} z^2 k(z)\,dz < \infty$.

The technical advantages of using this type of kernel were noted first in Bickel and Klaassen (1986).

We then have the following result.

PROPOSITION 4. *Assume that* (S1)–(S3) *hold*, *that* g *is symmetric*, *that* $\boldsymbol{\theta}_0^{(n)}$ *is root-*n *consistent*, *that the kernel* k *is symmetric and satisfies condition* (K), *and that the constants involved in* (3.15) *satisfy* (3.12). *Then*, *the sequence*

$$\hat{\boldsymbol{\vartheta}}^{(n)} =: \boldsymbol{\theta}_0^{(n)} + \left(\hat{\mathcal{J}}^{(n)}\hat{\mathbf{Q}}^{(n)}\right)^{-1}\frac{1}{N_n}\sum_{t=d_n}^n \dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)})\hat{\varphi}^{(n)}(\hat{Z}_t^{(n)}), \tag{3.15}$$

where

$$\begin{aligned} \hat{\mathcal{J}}^{(n)} &=: \frac{1}{N_n}\sum_{t=d_n}^n \left[\hat{\varphi}^{(n)}(\hat{Z}_t^{(n)})\right]^2 \quad \textit{and} \\ \hat{\mathbf{Q}}^{(n)} &=: \frac{1}{N_n}\sum_{t=d_n}^n \dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)})\left(\dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)})\right)', \end{aligned} \tag{3.16}$$

is locally asymptotically minimax (*LAM*).

PROOF. The proposition results from Koul and Schick's (1997) Theorem 5.2, which shows that, under conditions (i)–(vi) of Lemma 1,

$$n^{1/2}\left(\hat{\boldsymbol{\vartheta}}^{(n)} - \boldsymbol{\theta}\right) = \left(\sigma^2 \mathscr{I}(g)\mathbf{Q}(\boldsymbol{\theta})\right)^{-1} \boldsymbol{\Delta}_g^{(n)}(\boldsymbol{\theta}) + o_P(1)$$

under $H_g^{(n)}(\boldsymbol{\theta})$, as $n \to \infty$. □

When the assumption of symmetry cannot be made, the construction of the adaptive estimator is slightly different. Using the same notation as above, we now assume that the constants $a_n,\ b_n,\ c_n$ and the sequence m_n of Lemma 1 are such that

$$n^{-1}a_n^{-4}b_n^{-2}c_n^2 \to 0 \quad \text{and} \quad n^{-1}a_n^{-3}b_n^{-1}c_n^2 m_n \to 0 \tag{3.17}$$

as $n \to \infty$. The estimator of the score function is

$$\hat{\varphi}^{(n)}(z) =: -\frac{f'_{N_n}\left(z, \hat{\mathbf{Z}}^{(n)}\right)}{b_n + f_{N_n}\left(z, \hat{\mathbf{Z}}^{(n)}\right)}, \tag{3.18}$$

still with $N_n = n - d_n + 1$ and f_n given in (3.14).

PROPOSITION 5. *Assume that* (S1)–(S3) *hold*, *that* $\boldsymbol{\theta}_0^{(n)}$ *is root-n consistent*, *that the kernel k satisfies condition* (K), *and that the constants involved in* (3.15) *satisfy* (3.17). *Then*, *the sequence*

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}^{(n)} =: \boldsymbol{\theta}_0^{(n)} + \left(\hat{\mathcal{J}}^{(n)}\hat{\mathbf{Q}}^{(n)}\right)^{-1} \\ \times \tfrac{1}{N_n}\textstyle\sum_{t=d_n}^n \left(\dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)}) - \tfrac{1}{N_n}\sum_{t=d_n}^n \dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)})\right) \\ \times \hat{\varphi}^{(n)}(\hat{Z}_t^{(n)}),\end{aligned} \tag{3.19}$$

with $\hat{\mathcal{J}}^{(n)}$ *given in* (3.16) *and*

$$\begin{aligned}\hat{\mathbf{Q}}^{(n)} =: \frac{1}{N_n}\sum_{t=d_n}^{n}\left(\dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)}) - \frac{1}{N_n}\sum_{t=d_n}^{n}\dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)})\right) \\ \times \left(\dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)}) - \tfrac{1}{N_n}\textstyle\sum_{t=d_n}^n \dot{\mathbf{H}}_t^{(n)}(\boldsymbol{\theta}_0^{(n)})\right)'\end{aligned} \tag{3.20}$$

is locally asymptotically minimax (*LAM*).

The result again follows from Koul and Schick [(1997), Theorem 6.2]; here, conditions (i)–(vii) of Lemma 1 are required.

The adaptive estimation of the lag parameter d in the FARIMA$(0, d, 0)$ model $(1 - L)^d X_t = \varepsilon_t$ has been treated in Hallin and Serroukh (1999).

TABLE 1

Square roots of the mean square errors of the Gaussian maximum likelihood estimator $\phi_0^{(n)}$ and the adaptive estimator $\hat{\phi}^{(n)}$ of ϕ, in model (3.8), with $d = 0.2$ and centered exponential errors ε_t, for $\phi = \pm 0.5$ and ± 0.8, respectively. Series length: $n = 40$. Number of replications: 40

	$\boldsymbol{\phi}_0^{(n)}$	$\hat{\boldsymbol{\phi}}^{(n)}$		$\boldsymbol{\phi}_0^{(n)}$	$\hat{\boldsymbol{\phi}}^{(n)}$
$\phi = -0.8$	0.10575	0.10549	$\phi = 0.8$	0.09663	0.09646
$\phi = -0.5$	0.11061	0.10938	$\phi = 0.5$	0.10419	0.10414

While the theoretical interest of adaptive estimation for long and very long time series is obvious from the LAM property of the resulting estimates, the practical virtues of adaptivity for short series lengths should be investigated by means of an extensive Monte Carlo study. Such an investigation is beyond the scope of this paper. A modest simulation nevertheless has been conducted and yields encouraging results, despite the very short length ($n = 40$) of the (non-Gaussian) series considered.

Forty replications of FARIMA$(1, d, 0)$ processes of length $n = 40$, characterized by (3.8), with centered exponential errors ε_t have been generated for the parameter values $d = 0.2$, $\phi = \pm 0.5$ and ± 0.8, respectively. For each replication, a LAM estimation $\hat{\phi}^{(n)}$ of ϕ has been computed along the lines of Proposition 5. Dahlhaus' (1989) Gaussian MLE was used as the initial estimate $\boldsymbol{\theta}_0^{(n)} = (\phi_0^{(n)}, d_0^{(n)})$. The sequences $a_n = b_n = n^{-1/16}$, $c_n = m_n = n^{-1/8}$, $d_n = n/2$ and $N_n = n - d_n + 1$ were adopted, with the kernel $k(u) = \frac{3}{4}\sqrt{5}\left(1 - u^2/5\right) I\left[|u| \leq 5\right]$. Table 1 presents, for each value of ϕ, the square root of the mean square error computed over the 40 replications, for the Gaussian MLE $\phi_0^{(n)}$ and the adaptive $\hat{\phi}^{(n)}$, respectively.

3.4. *Other inference problems.* As already mentioned, the LAN result in Proposition 1 is the key to most inference problems for the long memory FARIMA model (2.2). Further examples are:

1. Locally asymptotically minimax estimation; the construction of locally asymptotically minimax estimates in the parametric models with specified innovation density g readily follows from Proposition 1 by applying the usual methods described, for example, in Section 5.3 of Le Cam and Yang (1990).
2. Maximum likelihood estimation; the asymptotic behavior of maximum likelihood estimators similarly follows from Proposition 1.
3. Rank-based testing; rank-based tests can be derived, along the same general lines as in Hallin and Puri (1994), from rank-based versions of the central sequences $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$. This approach is adopted in Serroukh (1996).
4. Adaptive rank-based testing methods; adaptive rank-based tests—actually, permutation tests—enjoying the same (conditional) distribution-freeness

properties as rank tests but also yielding the uniformly optimal performance of adaptive tests, can be constructed in the same spirit as in the simpler case of AR models; see Hallin and Werker (1998) for a brief outline in the traditional short memory AR(p) context.

4. Proofs.

4.1. *Proof of Proposition* 1. It is well known that

$$(1-e^{i\lambda})^d = \sum_{k=0}^{\infty} \pi_k e^{i\lambda k} \quad \text{and} \quad (1-e^{i\lambda})^{-d} = \sum_{k=0}^{\infty} \rho_k e^{i\lambda k},$$

where

$$\pi_k = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} \quad \text{and} \quad \rho_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}.$$

It follows that

$$\begin{aligned} \pi_k &= O(k^{-1-d}), & \rho_k &= O(k^{-1+d}),\\ \tfrac{\partial \pi_k}{\partial d} &= O(k^{-1-d}\log k), & \tfrac{\partial \rho_k}{\partial d} &= O(k^{-1+d}\log k),\\ \tfrac{\partial^2 \pi_k}{\partial^2 d} &= O(k^{-1-d}(\log k)^2), & \tfrac{\partial^2 \rho_k}{\partial^2 d} &= O(k^{-1+d}(\log k)^2). \end{aligned} \tag{4.1}$$

Considering the characteristic polynomials $\phi(z)$ and $\eta(z)$ in (S1), $\phi(z)/\eta(z)$ is expanded as the absolutely convergent (for $|z| < 1$) power series,

$$\frac{\phi(z)}{\eta(z)} = \sum_{k=0}^{\infty} \gamma_k z^k,$$

where the γ_k's are such that

$$|\gamma_k| = O(|\gamma(\boldsymbol{\theta})|^k) \tag{4.2}$$

for some bounded, C^2 function $\gamma(\cdot)$ of $\boldsymbol{\theta}$ satisfying $|\gamma(\boldsymbol{\theta})| < 1$; compare Fuller (1996). Similarly, defining the sequence $\{\alpha_k\}$ by

$$\frac{\eta(z)}{\phi(z)} = \sum_{k=0}^{\infty} \alpha_k z^k, \qquad |z| < 1,$$

there exists a C^2 function $\alpha(\boldsymbol{\theta})$ such that $|\alpha(\boldsymbol{\theta})| < 1$ and

$$|\alpha_k| = O(|\alpha(\boldsymbol{\theta})|^k). \tag{4.3}$$

It follows from (4.1) and (4.2) that the sequence $\{\psi_k\}$ defined in (2.5) satisfies

$$\left| \frac{\partial^j \psi_k}{\partial \theta_{i_1} \cdots \partial \theta_{i_j}} \right| = O(k^{-1-d}(\log k)^j), \qquad j = 0, 1, 2, \tag{4.4}$$

as $k \to \infty$. Similarly, for the sequence $\{\xi_k\}$ defined in (2.6), it is easily shown that

$$\left| \frac{\partial^j \xi_k}{\partial\theta_{i_1} \cdots \partial\theta_{i_j}} \right| = O(k^{-1+d}(\log\ k)^j), \qquad j = 0,\ 1,\ 2, \tag{4.5}$$

as $k \to \infty$.

Our proof of the LAN result in Proposition 1 is based on a slight modification [see, e.g., Garel and Hallin (1995)] of Swensen's (1985) Lemma 1 [itself based on McLeish (1974)], applicable to a wide class of models with dependent observations. Note that, from (2.12),

$$\Lambda_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{t=1}^{n} \log\left((\Phi_t^{(n)}(\boldsymbol{\theta}^{(n)}, \boldsymbol{\beta}^{(n)}; \boldsymbol{\theta}, \boldsymbol{\beta}))^2 \right),$$

with

$$\begin{aligned}
&\left[\Phi_t^{(n)}(\boldsymbol{\theta}^{(n)}, \boldsymbol{\beta}^{(n)}; \boldsymbol{\theta}, \boldsymbol{\beta})\right]^2 \\
&\quad =: \Bigg\{ g\Bigg(\varepsilon_t + \sum_{\nu=0}^{t-1} \psi_\nu^{(n)}[Y_{t-\nu} - \mathbf{Z}'_{t-\nu}\boldsymbol{\beta}^{(n)}] - \sum_{\nu=0}^{t-1} \psi_\nu[Y_{t-\nu} - \mathbf{Z}'_{t-\nu}\boldsymbol{\beta}] \\
&\qquad\qquad + \sum_{r=0}^{\infty}\sum_{\mu=0}^{r} [\psi_{t+\mu}^{(n)}\xi_{r-\mu}^{(n)} - \psi_{t+\mu}\xi_{r-\mu}]\varepsilon_{-r} \Bigg) \Bigg\} \\
&\quad \times \{g(\varepsilon_t)\}^{-1}.
\end{aligned}$$

Define

$$U_t^{(n)} = \Phi_t^{(n)}\left(\boldsymbol{\theta}^{(n)}, \boldsymbol{\beta}^{(n)}; \boldsymbol{\theta}, \boldsymbol{\beta}\right) - 1,$$

and

$$W_t^{(n)} = \frac{1}{2}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)} \left\{ n^{-1/2} \sum_{\nu=1}^{t-1} [\mathbf{h}'\ \mathrm{grad}_{\boldsymbol{\theta}}(\psi_\nu)] e_{t-\nu} - \sum_{\nu=0}^{t-1} \psi_\nu \mathbf{Z}'_{t-\nu} \tilde{\mathbf{D}}_n^{-1}\mathbf{k} \right\},$$

and denote by $\mathscr{F}_t$ the σ-field generated by $\{\varepsilon_k,\ k \le 0;\ Y_1, \ldots, Y_t\}$. In order to prove parts (i) and (ii) of Proposition 1, it is sufficient to check for the following six conditions: see Swensen's Lemma 1 [note that (L1)–(L5) are Swensen's conditions (1.2)–(1.6), whereas (L6) is assumption (iii) in his Theorem 1]. All expectations and convergences are under $H_g^{(n)}(\boldsymbol{\theta}, \boldsymbol{\beta})$, as $n \to \infty$.

(L1) $\mathrm{E}\left[W_t^{(n)} | \mathscr{F}_{t-1}\right] = 0$ a.s.;

(L2) $\lim_{n\to\infty}\ \mathrm{E}\left[\sum_{t=1}^n (U_t^{(n)} - W_t^{(n)})^2\right] = 0$;

(L3) $\sup_n\ \mathrm{E}\left[\sum_{t=1}^n (W_t^{(n)})^2\right] < \infty$;

(L4) $\max_{1\le t\le n} |W_t^{(n)}| = o_P(1)$;

(L5) $\sum_{t=1}^n (W_t^{(n)})^2 - \tau^2/4 = o_P(1)$ for some positive constant τ;

(L6) $\sum_{t=1}^{n} \mathrm{E}\left[(W_t^{(n)})^2 I[|W_t^{(n)}| > \delta] | \mathscr{F}_{t-1}\right] = o_P(1)$ for some $\delta > 0$ ($I[\cdot]$ stands for the indicator function).

We successively prove that (L1)–(L6) hold here. Condition (L1) immediately follows from the definition of $W_t^{(n)}$. In order to prove (L2), let

$$\begin{aligned}
Q_t^{(n)} &= \sum_{\nu=0}^{t-1} \psi_\nu^{(n)} \left[Y_{t-\nu} - \mathbf{Z}'_{t-\nu}\boldsymbol{\beta}^{(n)}\right] - \sum_{\nu=0}^{t-1} \psi_\nu \left[Y_{t-\nu} - \mathbf{Z}'_{t-\nu}\boldsymbol{\beta} \quad\right] \\
&\quad + \sum_{r=0}^{\infty} \sum_{\mu=0}^{r} \left[\psi_{\mu+t}^{(n)} \xi_{r-\mu}^{(n)} - \psi_{\mu+t} \xi_{r-\mu}\right] \varepsilon_{-r} \\
&= Q_{t;1}^{(n)} + Q_{t;2}^{(n)} + Q_{t;3}^{(n)} \quad \text{say.}
\end{aligned}$$

Expanding at $\boldsymbol{\theta}$ yields

$$\begin{aligned}
&\sum_{\mu=0}^{r} \left[\psi_{\mu+t}^{(n)} \xi_{r-\mu}^{(n)} - \psi_{\mu+t} \xi_{r-\mu}\right] \\
&\quad = \left[O(n^{-1/2}) \sum_{\mu=0}^{r} \|\mathrm{grad}_{\boldsymbol{\theta}}\ \psi_{\mu+t}\| |\xi_{r-\mu}| + O(n^{-1/2}) \sum_{\mu=0}^{r} |\psi_{\mu+t}| \|\mathrm{grad}_{\boldsymbol{\theta}}\ \xi_{r-\mu}\|\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \\
&\quad = O(n^{-1/2}) A_1 + O(n^{-1/2}) A_2 \quad \text{say,}
\end{aligned}$$

where $\boldsymbol{\theta} \le \boldsymbol{\theta}^* \le \boldsymbol{\theta}^{(n)}$. In view of (4.4) and (4.5),

$$\begin{aligned}
A_1 &= \sum_{\mu=0}^{[r/2]} O\left((\mu+t)^{-1-d} \log(\mu+t)\right) O\left((r-\mu)^{-1+d} \log(r-\mu)\right) \\
&\quad + \sum_{\mu=[r/2]+1}^{r} O\left((\mu+t)^{-1-d} \log(\mu+t)\right) O\left((r-\mu)^{-1+d} \log(r-\mu)\right) \\
&\quad = O(r^{-1+d}) O(t^{-d/2}) O(\log r) \sum_{\mu=0}^{[r/2]} O\left((\mu+t)^{-1-d/2}\right) O(\log \mu + \log t) \\
&\qquad + O(r^{-1-d/2}) O(t^{-d/2}) O(\log t) O(\log r) \sum_{\mu=[r/2]+1}^{r} O\left((r-\mu)^{-1+d}\right).
\end{aligned}$$

Noting that $\sum_{\mu=0}^{r} \mu^\alpha = O(r^{\alpha+1})$ for $\alpha > -1$, we have that

$$A_1 = O(t^{-d/2} \log t) O(r^{-1+d} \log r).$$

Similarly, it can be shown that A_2 is $O(t^{-d/2} \log t) O(r^{-1+d} \log r)$, which implies

$$\sum_{\mu=0}^{r} \left[\psi_{\mu+t}^{(n)} \xi_{r-\mu}^{(n)} - \psi_{\mu+t} \xi_{r-\mu}\right] = O(n^{-1/2}) O(t^{-d/2} \log t) O(r^{-1+d} \log r). \tag{4.6}$$

Hence, $Q_{t;1}^{(n)} + Q_{t;2}^{(n)}$ can be written as

$$\sum_{\nu=0}^{t-1} \left[\psi_\nu^{(n)} - \psi_\nu\right] e_{t-\nu} - \sum_{\nu=0}^{t-1} \psi_\nu \mathbf{Z}'_{t-\nu}\tilde{\mathbf{D}}_n^{-1}\mathbf{k} - \sum_{\nu=0}^{t-1} \left[\psi_\nu^{(n)} - \psi_\nu\right] \mathbf{Z}'_{t-\nu}\tilde{\mathbf{D}}_n^{-1}\mathbf{k}$$

$$(4.7)\qquad = n^{-1/2} \sum_{\nu=0}^{t-1} \left[\mathbf{h}' \operatorname{grad}_{\boldsymbol{\theta}}(\psi_\nu)\right] e_{t-\nu} + \frac{1}{2n} \sum_{\nu=1}^{t-1} \left[\mathbf{h}' \left(\frac{\partial^2}{\partial\theta_i\theta_j}\psi_\nu\right)_{\boldsymbol{\theta}^*} \mathbf{h}\right] e_{t-\nu}$$

$$- \sum_{\nu=0}^{t-1} \psi_\nu \mathbf{Z}'_{t-\nu}\tilde{\mathbf{D}}_n^{-1}\mathbf{k} - \sum_{\nu=0}^{t-1} \left[\psi_\nu^{(n)} - \psi_\nu\right] \mathbf{Z}'_{t-\nu}\tilde{\mathbf{D}}_n^{-1}\mathbf{k},$$

where $\boldsymbol{\theta}^*$ is some intermediate point between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{(n)}$. On account of (4.4), (4.6) and (4.8), and by Theorem 2.3 in Yajima (1991), we obtain that

$$(4.8)\qquad \lim_{n\to\infty} \sum_{t=1}^{n} \mathrm{E}\left[(Q_t^{(n)})^2\right] < \infty.$$

Next, let us show that

$$(4.9)\qquad E_1^{(n)} =: \sum_{t=1}^{n} \mathrm{E}\left[\left(U_t^{(n)} - \frac{1}{2}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)} Q_t^{(n)}\right)^2\right] = o(1).$$

For every $c_1 > 0$, we have

$$E_1^{(n)} = \sum_{t=1}^{n} \mathrm{E}\left[I\left[n^{1/2}|Q_t^{(n)}| \le c_1\right] \left(U_t^{(n)} - \frac{1}{2}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)} Q_t^{(n)}\right)^2\right]$$
$$+ \sum_{t=1}^{n} \mathrm{E}\left[I\left[n^{1/2}|Q_t^{(n)}| > c_1\right] \left(U_t^{(n)} - \frac{1}{2}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)} Q_t^{(n)}\right)^2\right]$$
$$= E_{1;1}^{(n)} + E_{1;2}^{(n)} \quad \text{say.}$$

It follows from Lemma 2.2(ii) in Garel and Hallin (1995) that

$$E_{1;1}^{(n)} \le \left\{\sum_{t=1}^{n} \mathrm{E}\left[\left(Q_t^{(n)}\right)^2\right]\right\} o_{c_1}(1),$$

where $\lim_{n\to\infty} o_{c_1}(1) = 0$ for any given $c_1 > 0$. Hence, $E_{1;1}^{(n)}$ also is $o(1)$. Part (i) of the same lemma yields

$$(4.10)\qquad E_{1;2}^{(n)} \le \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[I\left[n^{1/2}|Q_t^{(n)}| > c_1\right] n \left(Q_t^{(n)}\right)^2 \mathscr{I}(g)\right].$$

From Ash [(1972), page 297], $n\left(Q_t^{(n)}\right)^2$ is uniformly integrable, which implies that the right-hand side in (4.10) converges to zero as $c_1 \to \infty$; (4.9) follows.

Finally, one easily checks that

$$\sum_{t=1}^{n} \mathrm{E}\left[\left(\frac{1}{2}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)}Q_t^{(n)} - W_t^{(n)}\right)^2\right]$$

$$= \sum_{t=1}^{n}\mathrm{E}\left[\left(\frac{1}{4n}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)}\sum_{\nu=1}^{t-1}\left[\mathbf{h}'\left(\frac{\partial^2}{\partial\theta_i\theta_j}\psi_\nu\right)_{\boldsymbol{\theta}^*}\mathbf{h}\right]e_{t-\nu}\right.\right.$$

$$\left.\left. - \sum_{\nu=0}^{t-1}\left[\psi_\nu^{(n)} - \psi_\nu\right]\mathbf{Z}'_{t-\nu}\tilde{\mathbf{D}}_n^{-1}\mathbf{k}\right)^2\right]$$

also converges to zero as $n \to \infty$. This completes the proof of (L2).
(L3). Note that

$$\sum_{t=1}^{n}\mathrm{E}\left[\left(W_t^{(n)}\right)^2\right] = \frac{1}{4n}\mathscr{I}(g)\sum_{t=1}^{n}\sum_{\nu_1=1}^{t-1}\sum_{\nu_2=1}^{t-1}\left[\mathbf{h}'\,\mathrm{grad}_{\boldsymbol{\theta}}(\psi_{\nu_1})\right]R_e(\nu_1-\nu_2)\left[\mathrm{grad}'_{\boldsymbol{\theta}}(\psi_{\nu_2})\mathbf{h}\right]$$

$$+\frac{1}{4}\mathscr{I}(g)\left(\tilde{\mathbf{D}}_n^{-1}\mathbf{k}\right)'\left[\sum_{t=1}^{n}\sum_{\nu_1=0}^{t-1}\sum_{\nu_2=0}^{t-1}\psi_{\nu_1}\psi_{\nu_2}\mathbf{Z}_{t-\nu_1}\mathbf{Z}'_{t-\nu_2}\right]\tilde{\mathbf{D}}_n^{-1}\mathbf{k},$$

where $R_e(\nu_1-\nu_2) =: \mathrm{E}[e_{t-\nu_1}e_{t-\nu_2}]$. Lemma 4.4 of Garel and Hallin (1995) and Theorem 2.3 of Yajima (1991), along with (4.4), imply that $\sum_{t=1}^{n}\mathrm{E}[\left(W_t^{(n)}\right)^2] \to \tau^2/4$ as $n \to \infty$, where

$$\tau^2 =: \sigma^2\mathscr{I}(g)\mathbf{h}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{h} + \mathscr{I}(g)\mathbf{k}'\mathbf{W}(\boldsymbol{\theta})\mathbf{k};$$

hence, (L3) is satisfied.
(L4). Turning to (L4), decompose $W_t^{(n)}$ into $W_{t;1}^{(n)} + W_{t;2}^{(n)}$, with

$$W_{t;1}^{(n)} =: \frac{1}{2\sqrt{n}}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)}\sum_{\nu_1=0}^{t-1}\left[\mathbf{h}'\,\mathrm{grad}_{\boldsymbol{\theta}}(\psi_\nu)\right]e_{t-\nu} \tag{4.11}$$

and

$$W_{t;2}^{(n)} =: -\frac{1}{2}\frac{g'(\varepsilon_t)}{g(\varepsilon_t)}\sum_{\nu_1=0}^{t-1}\psi_\nu\mathbf{Z}'_{t-\nu}\tilde{\mathbf{D}}_n^{-1}\mathbf{k}. \tag{4.12}$$

For every $\varepsilon > 0$,

$$P\left[\max_{1\le t\le n}|W_t^{(n)}| > 2\varepsilon\right] \le P\left[\max_{1\le t\le n}|W_{t;1}^{(n)}| > \varepsilon\right] + P\left[\max_{1\le t\le n}|W_{t;2}^{(n)}| > \varepsilon\right].$$

The Markov inequality implies that

$$\begin{aligned} P\left[\max\nolimits_{1\le t\le n}|W_{t;1}^{(n)}| > \varepsilon\right] &\le P\left[\sum_{t=1}^{n}\left(W_{t;1}^{(n)}\right)^2 I\left[|W_{t;1}^{(n)}| > \varepsilon\right] > \varepsilon^2\right] \\ &\le \varepsilon^{-2}n^{-1}\sum_{t=1}^{n}\mathrm{E}\left[n\left(W_{t;1}^{(n)}\right)^2 I[n^{1/2}|W_{t;1}^{(n)}| > \varepsilon n^{1/2}]\right]. \end{aligned} \tag{4.13}$$

Just as $n\left(Q_t^{(n)}\right)^2$, $n\left(W_{t;1}^{(n)}\right)^2$ is uniformly integrable, so that (4.13) is $o(1)$ as $n \to \infty$. In view of (G5), the case of $W_{t;2}^{(n)}$ is entirely similar; (L4) follows.

(L5). For the sake of simplicity, write $\dot{\psi}_\nu = \dot{\psi}_\nu(\mathbf{h})$ instead of $\mathbf{h}'\ \mathrm{grad}_{\boldsymbol{\theta}}(\psi_\nu)$. Then, $W_{t;1}^{(n)} = (1/2\sqrt{n})(g'(\varepsilon_t)/g(\varepsilon_t))\sum_{\nu=1}^{t-1}\dot{\psi}_\nu e_{t-\nu}$, with an ℓ^1-summable sequence $\{\dot{\psi}_\nu\}$. For any fixed $\varepsilon > 0$, we can choose $M > 0$ such that $\sum_{\nu=M}^{\infty}|\dot{\psi}_\nu|\mathrm{E}\left[|e_{t-\nu}|\right] < \varepsilon$. Note that the process $(g'(\varepsilon_t)/g(\varepsilon_t))\sum_{\nu=1}^{\infty}\dot{\psi}_\nu e_{t-\nu}$ is ergodic. Theorem 2 of Hannan [(1970), page 203] thus entails

$$\sum_{t=1}^{n}\left(W_{t;1}^{(n)}\right)^2 \longrightarrow \frac{\sigma^2}{4}\mathscr{I}(g)\mathbf{h}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{h} \quad \text{a.s.}$$

Similarly, showing that

$$\mathrm{E}\left[\sum_{t=1}^{n}\left(W_{t;2}^{(n)}\right)^2\right] \longrightarrow \tfrac{1}{4}\mathscr{I}(g)\mathbf{k}'\mathbf{W}\mathbf{k}$$

and that

$$\mathrm{Var}\left[\sum_{t=1}^{n}\left(W_{t;2}^{(n)}\right)^2\right] \longrightarrow 0,$$

we obtain

$$\sum_{t=1}^{n}\left(W_{t;2}^{(n)}\right)^2 \xrightarrow{P} \frac{1}{4}\mathscr{I}(g)\mathbf{k}'\mathbf{W}\mathbf{k} \quad \text{and} \quad \sum_{t=1}^{n}W_{t;1}^{(n)}W_{t;2}^{(n)} \xrightarrow{P} 0.$$

(L6). Since $\sum_{t=1}^{n}\mathrm{E}\left[\left(W_t^{(n)}\right)^2 I[|W_t^{(n)}| > \delta]\middle|\mathscr{F}_{t-1}\right]$ is a nonnegative variable, it is sufficient, in order to prove (L6), to show that

$$\sum_{t=1}^{n}\mathrm{E}\left[\left(W_t^{(n)}\right)^2 I[|W_t^{(n)}| > \delta]\right] = o(1) \tag{4.14}$$

as $n \to \infty$. This, however, already has been shown in the proof of (L4).

Finally, part (iii) of Proposition 1 readily follows from Scheffé's lemma and the continuity of g. $\square$

4.2. *Proof of Proposition* 3. The proof relies on Proposition 1 and Le Cam's so-called *third lemma*. We just briefly outline it here. Giraitis and Surgailis (1990) established the asymptotic normality, under (3.4), as $n \to \infty$, of

$$T_1^{(n)} = n^{-1/2}\sum_{t=1}^{n}\sum_{s=1}^{n}\left[e_t e_s - R(t-s)\right]B(t-s).$$

In their proof, they approximate the sums $T_{1;1}^{(n)} = n^{-1/2}\sum_{t=1}^{n}\sum_{s=1}^{n}e_t e_s$ with sequences of the form $U_1^{(n)} = n^{-1/2}\sum_{t=1}^{n}e_1(t)e_2(t)$, where $e_i(t)$, $i = 1, 2$, are finite linear combinations of the ε_t's, involving $2M$ terms. For M sufficiently large, the approximation of $T_{1;1}^{(n)}$ by $U_1^{(n)}$ can be made arbitrarily good. On the other hand, we can approximate $\boldsymbol{\alpha}'_n \mathbf{e}_n$ by $V_1^{(n)} = \sum_{t=1}^n \alpha_t \sum_{j=0}^t \beta_j \varepsilon_{t-j}$, where $\{\beta_j\}$ is ℓ^2-summable. The joint asymptotic normality of $(\Lambda^{(n)}, T^{(n)})$ then follows from that of $(\Lambda^{(n)}, U_1^{(n)} + V_1^{(n)})$, which in turn follows from checking for McLeish's classical conditions. Next, we need an evaluation of the variance of $T^{(n)}$, which is equal to

$$\lim_{n\to\infty} \Big[\mathrm{Var}(T_{1;1}^{(n)}) + 2\ \mathrm{Cov}(T_{1;1}^{(n)},\ \boldsymbol{\alpha}'_n \mathbf{e}_n)\Big] + c(\boldsymbol{\theta}).$$

Giraitis and Surgailis (1990) give

$$\lim_{n\to\infty} \mathrm{Var}(T_{1;1}^{(n)}) = 16\pi^3 \int_{-\pi}^{\pi} [f_{\boldsymbol{\theta}}(\lambda) f_{bb}(\lambda)]^2\ d\lambda + \mu_4 \left[2\pi \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\lambda) f_{bb}(\lambda)\, d\lambda\right]^2.$$

On the other hand, we have

$$\begin{aligned}
\mathrm{Cov}(T_{1;1}^{(n)},\ \boldsymbol{\alpha}'_n \mathbf{e}_n) &= n^{-1/2} \sum_{r=1}^n \sum_{k=1}^n \sum_{\ell=1}^n \alpha_r B(k-\ell) \mathrm{cum}(e_r, e_k, e_\ell) \\
&= n^{-1/2} \sum_{r=1}^n \alpha_r \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} D_n(\eta+\lambda) D_n(-\eta+\mu) e^{-ir\lambda - ir\mu} \qquad (4.15)\\
&\qquad \times f_{bb}(\lambda) f_{eee}(\lambda, \mu)\, d\lambda\, d\mu\, d\eta,
\end{aligned}$$

where $D_n(\lambda) = \sum_{k=1}^n e^{ik\lambda}$. Now, (4.15) reduces to

$$n^{-1/2} \sum_{r=1}^n \alpha_r \left[(2\pi)^2 \int_{-\pi}^{\pi} f_{bb}(\eta) f_{eee}(-\eta, \eta)\, d\eta + o(1)\right],$$

a quantity which, by assumption (T1), tends to zero as $n \to \infty$. Finally, we evaluate the asymptotic covariance

$$\begin{aligned}
&\mathrm{Cov}(\Lambda^{(n)}, T^{(n)}) \\
&\quad = \mathrm{Cov}\Bigg(n^{-1/2} \sum_{j=1}^n S_j \sum_{\nu=1}^{j-1} \big[\mathbf{h}'\ \mathrm{grad}_{\boldsymbol{\theta}}(\psi_\nu)\big] e_{j-\nu} - \mathbf{k}'\tilde{\mathbf{D}}_n^{-1} \sum_{j=1}^n S_j \sum_{\nu=0}^{j-1} \psi_\nu \mathbf{Z}_{j-\nu}, \\
&\qquad\qquad n^{-1/2} \sum_{k=1}^n \sum_{\ell=1}^n e_k e_\ell B(k-\ell) + \boldsymbol{\alpha}'_n \mathbf{e}_n\Bigg) \\
&\quad = \mathrm{Cov}\,[(1)+(2),\ (3)+(4)] \quad \text{say.}
\end{aligned}$$

The resulting four terms yield $m_1, \ldots, m_4$ in (3.7). We provide some details on the derivation of m_1; the three other terms are obtained along the same lines. Putting $\dot{\psi}_\nu(\mathbf{h}) =: \mathbf{h}'\ \mathrm{grad}_{\boldsymbol{\theta}}(\psi_\nu)$, we obtain

$$\begin{aligned}
\mathrm{Cov}((1),\ (3)) &= \mathrm{Cov}\Bigg(n^{-1/2} \sum_{j=1}^n S_j \sum_{\nu=1}^{j-1} \dot{\psi}_\nu(\mathbf{h}) e_{j-\nu},\ n^{-1/2} \sum_{k=1}^n \sum_{\ell=1}^n e_k e_\ell B(k-\ell)\Bigg) \\
&= n^{-1} \sum_{j=1}^n \sum_{\nu=1}^{j-1} \sum_{k=1}^n \sum_{\ell=1}^n \dot{\psi}_\nu(\mathbf{h}) B(k-\ell) \big[\mathrm{Cov}(S_j, e_k) \mathrm{Cov}(e_{j-\nu}, e_\ell)
\end{aligned}$$

$$+\mathrm{Cov}(S_j, e_\ell)\mathrm{Cov}(e_{j-\nu}, e_k)\big]$$

$$= C_1^{(n)} + C_2^{(n)} \quad \text{say.}$$

The first term yields

$$C_1^{(n)} = \frac{2\pi}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i\lambda(k-j)} \left[\sum_{\nu=1}^{j-1} \dot{\psi}_\nu(\mathbf{h}) e^{i\lambda\nu} \right] \times f_{bb}(\lambda) f_{\boldsymbol{\theta}}(\lambda)\, d\lambda\, e^{i(k-j)\mu} f_{Se}(\mu)\, d\mu$$

$$= (2\pi)^2 \int_{-\pi}^{\pi} \left[\sum_{\nu=1}^{\infty} \dot{\psi}_\nu(\mathbf{h}) e^{i\lambda\nu} \right] f_{bb}(\lambda) f_{\boldsymbol{\theta}}(\lambda) f_{Se}(-\lambda)\, d\lambda + o(1).$$

The case of $C_2^{(n)}$ is quite similar, and it is easily shown that $C_2^{(n)}$ converges to the same limit as $C_1^{(n)}$. The proposition follows. $\square$

4.3. *Proof of Lemma 1.* We successively establish the seven statements of the lemma.

(i) Letting $(\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta}) = n^{-1/2}(h_1^{(n)}, h_2^{(n)})'$, we have

$$\sum_{t=1}^{n} \left| H_t(\boldsymbol{\theta}^{(n)}) - H_t(\boldsymbol{\theta}) - (\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta})' \dot{\mathbf{H}}_t(\boldsymbol{\theta}) \right|^2$$

$$= \frac{1}{4n^2} \sum_{t=1}^{n} \left| \sum_{\nu=1}^{t-1} \sum_{i,j=1}^{2} h_i^{(n)} h_j^{(n)} \left(\partial_{ij}^2 \psi_\nu\right)_{\boldsymbol{\theta}+c(\boldsymbol{\theta}^{(n)}-\boldsymbol{\theta})} X_{t-\nu} \right|^2,$$

where $c = c_t(X_1, \ldots, X_n) \in [0,1]$. Taking expectations and bounding $\mathrm{E}[X_{t-\nu_1} X_{t-\nu_2}]$ with $\mathrm{E}[X_t^2]$ yields

$$\mathrm{E}\left[\sum_{t=1}^{n} \left| H_t(\boldsymbol{\theta}^{(n)}) - H_t(\boldsymbol{\theta}) - (\boldsymbol{\theta}^{(n)} - \boldsymbol{\theta})' \dot{\mathbf{H}}_t(\boldsymbol{\theta}) \right|^2 \right]$$

$$\leq \frac{1}{4n^2} \sum_{t=1}^{n} \left| \sum_{\nu=1}^{t-1} \sum_{i,j=1}^{2} h_i^{(n)} h_j^{(n)} \sup_{c\in[0,1]} \left(\partial_{ij}^2 \psi_\nu\right)_{\boldsymbol{\theta}+c(\boldsymbol{\theta}^{(n)}-\boldsymbol{\theta})} \right|^2 \mathrm{E}\left[X_t^2\right],$$

a quantity which, for n sufficiently large, is less than

$$\frac{K}{4n} \left| \sum_{\nu=1}^{\infty} \sum_{i,j=1}^{2} \left| \partial_{ij}^2 \psi_\nu(\boldsymbol{\theta}) \right| \right|^2,$$

where K is an appropriate constant. The result then follows from the fact that, in view of (4.4), $\sum_{\nu=1}^{\infty} \sum_{i,j=1}^{2} \left| \partial_{ij}^2 \psi_\nu(\boldsymbol{\theta}) \right| < \infty$.

(ii) This is an immediate consequence of (L4); see the proof of Proposition 1.

(iii) and (iv) For any $\mathbf{h} \in \mathbb{R}^2$, $\mathbf{h}'\dot{\mathbf{H}}_t$ has (up to the choice of the score function, which here is $g'(x)/g(x) =: x$) the same form as defined as $2n^{1/2} W_{t;1}^{(n)}$

appearing in (4.11) [proof of conditions (L4) and (L5) of Proposition 1]. The same ergodicity argument used there yields

$$\frac{1}{n}\sum_{t=1}^{n}\mathbf{h}'\dot{\mathbf{H}}_t \xrightarrow{P} \mathrm{E}[\mathbf{h}'\dot{\mathbf{H}}_t]=0 \quad \text{and} \quad \frac{1}{4n}\sum_{t=1}^{n}\mathbf{h}'\dot{\mathbf{H}}_t\dot{\mathbf{H}}_t'\mathbf{h} \xrightarrow{\text{a.s.}} \frac{\sigma^2}{4}\mathbf{h}'\mathbf{Q}(\boldsymbol{\theta})\mathbf{h}.$$

The statement follows.

(v) A Taylor expansion of $\psi_\nu^{(n)} =: \psi_\nu(\boldsymbol{\theta}^{(n)})$ yields, for some $c \in [0,1]$,

$$\begin{aligned}
\|\dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)})-\dot{\mathbf{H}}_t(\boldsymbol{\theta})\|^2 &= \sum_{i=1}^{2}\left|\sum_{\nu=1}^{t-1}\left(\frac{\partial}{\partial\theta_i}\psi_\nu(\boldsymbol{\theta}^{(n)})-\frac{\partial}{\partial\theta_i}\psi_\nu(\boldsymbol{\theta})\right)X_{t-\nu}\right|^2\\
&\le \sum_{i=1}^{2}\left\{\sum_{j=1}^{2}\sum_{\nu=1}^{t-1} n^{-1/2}\left|h_j^{(n)}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\psi_\nu(\boldsymbol{\theta}+cn^{-1/2}\mathbf{h}^{(n)})X_{t-\nu}\right|\right\}^2\\
&\le \frac{4}{n}\|\mathbf{h}^{(n)}\|^2\sum_{i,j=1}^{2}\left\{\sum_{\nu=1}^{t-1}\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\psi_\nu(\boldsymbol{\theta}+cn^{-1/2}\mathbf{h}^{(n)})X_{t-\nu}\right|\right\}^2.
\end{aligned}$$

Summing over t and taking expectation yields, as in the proof of (i), an $O(1)$ quantity. The result follows.

(vi) This statement is an immediate consequence of (4.13).

(vii) From (3.10), we have

$$\dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)})=\sum_{k=0}^{\infty}\sum_{j=1}^{t-1}\mathrm{grad}(\psi_j^{(n)})\xi_k^{(n)}Z_{t-j-k}(\boldsymbol{\theta}^{(n)}).$$

Under $H^{(n)}(\boldsymbol{\theta}^{(n)})$, $Z_i(\boldsymbol{\theta}^{(n)})$ coincides with

$$\varepsilon_i+\sum_{j=t}^{\infty}\psi_j^{(n)}X_{t-j}.$$

Since $\sum_{j=t}^{\infty}\psi_j^{(n)}X_{t-j}$ is measurable with respect to $(\varepsilon_0,\varepsilon_{-1},\dots)$,

$$\begin{aligned}
&\dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)})-\mathrm{E}_{\boldsymbol{\theta}^{(n)}}\left[\dot{\mathbf{H}}_t(\boldsymbol{\theta}^{(n)})\middle|\mathscr{B}_{n,i}(\boldsymbol{\theta}^{(n)})\right]\\
&\qquad=\begin{cases}\displaystyle\sum_{k=0}^{t-i-1}\mathrm{grad}(\psi_{t-i-k}^{(n)})\xi_k^{(n)}\varepsilon_i, & 1\le i\le t,\\ 0, & t<i.\end{cases}
\end{aligned}$$

Since moreover $\sum_{k=0}^{\infty}\left(\xi_k^{(n)}\right)^2<\infty$, and since, in view of (4.4), $\|\mathrm{grad}(\psi_j^{(n)})\|^2 = O((j^{-1-d}\log j)^2)$, it follows that

$$n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{t-m_n}\mathrm{E}_{\boldsymbol{\theta}^{(n)}}\left[\left\|\sum_{k=0}^{t-i-1}\mathrm{grad}(\psi_{t-i-k}^{(n)})\xi_k^{(n)}\varepsilon_i\right\|^2\right]$$

$$= Kn^{-1} \sum_{t=m_n}^{n} \sum_{i=1}^{t-m_n} \left\| \text{grad}(\psi_{t-i-k}^{(n)}) \right\|^2$$

$$< K' n^{-1} \sum_{t=m_n}^{n} (t - m_n) O(((t - m_n)^{-1-d} \log (t - m_n))^2) \leq K'' \frac{n - m_n}{n},$$

where K, K', and K'' are adequate constants. This latter quantity is $o(a_n^2)$ provided that m_n is such that the ratio $(n - m_n)/n$ goes to zero sufficiently fast, letting, for example, $m_n = n - o(na_n^2)$. □

Acknowledgments. The authors gratefully acknowledge the careful reading and very pertinent remarks by an anonymous referee, which lead to a substantial improvement of the initial version of this paper.

REFERENCES

ASH, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York.

BENGHABRIT, Y. and HALLIN, M. (1998). Locally asymptotically optimal tests for *AR*(p) against diagonal bilinear dependence. *J. Statist. Plann. Inference* **68** 47–63.

BERAN, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

BERAN, J. (1995). Maximum likelihood of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B* **57** 659–672.

BICKEL, P. J. and KLAASSEN, C. A. J. (1986). Empirical Bayes estimation in functional and structural models, and uniformly adaptive estimation of location. *Adv. in Appl. Math.* **7** 55–69.

BRILLINGER, D. R. (1981). *Time Series. Data Analysis and Theory*, 2nd ed. Holden-Day, San Francisco.

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.

DROST, F. C., KLAASSEN, C. A. J. and WERKER, B. J. M. (1997). Adaptive estimation in time series models. *Ann. Statist.* **25** 786–818.

FULLER, W. A. (1996). *Introduction to Statistical Time Series*, 2nd ed. Wiley, New York.

GAREL, B. and HALLIN, M. (1995). Local asymptotic normality of multivariate ARMA processes with linear trend. *Ann. Inst. Statist. Math.* **47** 551–579.

GIRAITIS, L. and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104.

HALLIN, M. and PURI, M. L. (1988). Optimal rank-based procedures for time series analysis: testing an ARMA model against other ARMA models. *Ann. Statist.* **16** 402–432.

HALLIN, M. and PURI, M. L. (1994). Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.* **50** 175–237.

HALLIN, M. and SERROUKH, A. (1999). Adaptive estimation of the lag of a long-memory process. *Statist. Inference Stochastic Process.* **2** 1–19.

HALLIN, M. and WERKER, B. (1998). Optimal testing for semi-parametric AR models: from Lagrange multipliers to autoregression rank scores and adaptive tests. In *Asymptotics, Nonparametrics and Time Series* (S. Ghosh, ed.) 295–358 M. Dekker, New York.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HWANG, S.Y. and BASAWA, I. V. (1993). Asymptotic optimal inference for a class of nonlinear time series models. *Stochastic Process. Appl.* **46** 91–113.

JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time-series models. *Econometric Theory* **11** 818–887.

JEGANATHAN, P. (1997). On asymptotic inference in linear cointegrated time-series systems. *Econometric Theory* **13** 692–745.
KOUL, H. L. and SCHICK, A. (1997). Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* **3** 247–277.
KREISS, J. P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.
KREISS, J. P. (1990a). Testing linear hypotheses in autoregression. *Ann. of Statist.* **18** 1470–1482.
KREISS, J. P. (1990b). Local asymptotic normality for autoregression with infinite order. *J. Statist. Plann. Inference* **26** 185–219.
LE CAM, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics* **3** 37–98.
LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics*. Springer, New York.
MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **4** 620–633.
SCHICK, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21** 1486–1521.
SETHURAMAN, S. and BASAWA, I. V. (1997). The asymptotic distribution of the maximum likelihood estimator for a vector time series model with long memory dependence. *Statist. Probab. Letters* **31** 285–293.
SERROUKH, A. (1996). *Inférence asymptotique paramétrique et non paramétrique pour les modèles ARMA fractionnaires*. Ph.D. thesis, Institut Statist., Univ. Libre Bruxelles.
SHUMWAY, R. H. and UNGER, A. N. (1974). Linear discriminant functions for stationary time series. *J. Amer. Statist. Assoc.* **69** 948–956.
STRASSER, H. (1985). *Mathematical Theory of Statistics*. W. de Gruyter, Berlin.
SWENSEN, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16** 54–70.
TANIGUCHI, M. (1998). Statistical analysis based on functionals of nonparametric spectral density estimators. In *Asymptotics, Nonparametrics and Time Series* (S. Ghosh, ed.) 351–394 M. Dekker, New York.
YAJIMA, Y. (1985). On estimation of long memory time series models. *Austral. J. Statist.* **27** 303–320.
YAJIMA, Y. (1991). Asymptotic properties of the LSE in a regression model with long memory stationary errors. *Ann. Statist.* **19** 158–177.
ZHANG, G. and TANIGUCHI, M. (1994). Discriminant analysis for stationary vector time series. *J. Time Ser. Anal.* **15** 117–126.

M. HALLIN
I.S.R.O., E.C.A.R.E.S.
AND DÉPARTEMENT DE MATHÉMATIQUE
UNIVERSITÉ LIBRE DE BRUXELLES
BRUSSELS, B 1050
BELGIUM
E-mail: mhallin@ulb.ac.be

M. TANIGUCHI
DEPARTMENT OF MATHEMATICAL SCIENCE
OSAKA UNIVERSITY
TOYONAKA, OSAKA 500
JAPAN
E-mail: taniguti@sigmath.es.osaka-u.ac.jp

A. SERROUKH
DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE OF SCIENCE
TECHNOLOGY AND MEDICINE
LONDON
UNITED KINGDOM
E-mail: a.serroukh@ic.ac.uk

K. CHOY
KOMAZAWA UNIVERSITY
TOKYO
JAPAN