Statistica Sinica **6**(1996), 147-169

LOCALLY ASYMPTOTICALLY OPTIMAL TESTS FOR AUTOREGRESSIVE AGAINST BILINEAR SERIAL DEPENDENCE

Youssef Benghabrit and Marc Hallin

Ecole Mohammadia d'Ingénieurs and Université Libre de Bruxelles

Abstract. Locally asymptotically most stringent tests for autoregressive against diagonal bilinear time series models are derived. A (restricted) local asymptotic normality property is therefore established for bilinear processes in the vicinity of linear autoregressive ones. The behaviour of the bispectrum under local alternatives of bilinear dependence shows the danger of misinterpreting skewness or kurtosis effects for nonlinearities. The proposed test statistic is a generalization of the Gaussian Lagrange multiplier statistic considered by Saikkonen and Luukkonen (1988), and is expressed as a closed-form function of the estimated residual spectrum and bispectrum. Its local power is explicitly provided. The local power of the Lagrange multiplier test follows as a particular case.

Key words and phrases: Time series, bilinear model, local asymptotic normality, bispectrum.

1. Introduction

1.1. Asymptotic inference for nonlinear time series models

The development of time series analysis has been largely dominated by the theory of linear (ARMA) models. This importance is mainly due to the L^2, hence implicitly Gaussian, approach adopted. Nonlinear models (Granger and Andersen (1978), Priestley (1988), Subba Rao and Gabr (1984), Tong (1990)) were introduced as a reaction against this excessive predominance of ARMA models. Unfortunately, the statistical analysis of nonlinear models also runs into mathematically more complex problems, a number of which remain unsolved. Much attention therefore has been devoted to the problem of testing for linear serial dependence (i.e., ARMA behaviour). The proposed testing procedures either are purely heuristic (Subba Rao and Gabr (1980), Hinich (1982), McLeod and Li (1983), Keenan (1985), Chan and Tong (1990), Petruccelli and Davies (1986), An and Cheng (1991), Subba Rao and da Silva (1992), Hinich and Patterson (1992), Chan and Tran (1992), Skaug and Tjøstheim (1993a, b)), or belong to the class of Gaussian Lagrange multiplier tests (against bilinear alternatives: Saikkonen and Luukkonen (1988, 1991), Guégan and Pham (1992). See Guégan (1990) or Tong (1990) for a review). All of them, of course, are of an asymptotic nature.

If, however, asymptotic inference is to be considered, the most adequate and powerful technique, certainly, is LeCam's (1960, 1986) *local asymptotic normality* (*LAN*) methodology, which has been very successfully applied in the linear context (Akritas and Johnson (1982), Swensen (1985), Hallin, Ingenbleek and Puri (1985), Kreiss (1987, 1990a,b), Garel (1989), Hallin and Puri (1994), Garel and Hallin (1995)). *LAN* indeed allows for deriving asymptotically optimal tests and estimates, for computing asymptotic local powers and asymptotic relative efficiencies, etc. Except however for Guégan and Pham (1992), where a contiguity property is assumed rather than proved, and for Benghabrit and Hallin (1992), where the particular case of testing for white noise against first-order superdiagonal bilinear dependence is treated, *LAN* so far has not been considered in the context of nonlinear time series models. The reason is that nonlinear models, even the simplest bilinear ones, are considerably more complicated than linear ones, due to the presence of infinite products in deconvolution formulae. The same technical difficulties are met in the proof of *LAN* properties as in the study of the asymptotic distribution of maximum likelihood or least square estimates of bilinear coefficients–a question which remains essentially an open problem. Due to these technical difficulties, the elegant method of Swensen (1985), itself based on LeCam (1986, Ch. 10) could not be applied here, and a more traditional Taylor expansion approach had to be adopted.

1.2. Testing for linearity

Optimality or asymptotic optimality is an important and extremely desirable property for a testing procedure − it is not the only one. The three main issues we are emphasizing in this derivation of asymptotic tests are (i) their validity, (ii) their ease of use and interpretability, and (iii) their power and optimality features:

(i) (*asymptotic validity*) Since the L^2 approach is inadequate in the bilinear context, Gaussian assumptions are highly unnatural, and should be avoided. If Gaussian likelihoods are to be considered, they should be treated in a *pseudo-likelihood* perspective, and it is essential that the (asymptotic) distributions of the test statistics to be proposed, hence the (asymptotic) probability levels of the resulting tests, remain valid under arbitrary probability densities.

(ii) (*ease of use and interpretability*) Most of time series practice, and much of the experience and intuitive insight of time series analysts is connected with correlogram inspection. Therefore, it is important that test statistics, in time series analysis, whenever possible, remain correlogram-based. Of course, traditional correlograms cannot be used in the detection of a nonlinear behaviour. In addition to classical autocorrelation coefficients (i.e., standardized versions of $\sum_t Z_t Z_{t-i}$, where Z_t is some residual series), we also consider here the third-order moment structure; what we call the *cubic autocorrelation coefficients* are standardized versions of $\sum_t Z_t Z_{t-i} Z_{t-j}$, constituting the *bispectrum*), and their non-Gaussian counterparts. Both traditional and cubic autocorrelation coefficients enter into the definition of our test statistics; inspecting the corresponding correlograms, however, provides quite an amount of side information (see Gabr (1988)).

(iii) (*optimality*) The tests we are proposing are designed for linear dependence against bilinear alternatives with unspecified "innovation" densities under the null as well as under the alternative. As far as optimality properties are concerned, however, some arbitrary but fixed "innovation" density has to be chosen as a target, since nonadaptive inference procedures cannot be expected to be uniformly (w.r.t. innovation densities) optimal. To be specific, our tests are *locally asymptotically most stringent* against the bilinear alternatives associated with some (arbitrarily) predetermined density type.

In order to keep things as simple and readable as possible, we are restricting the present paper to the case of first-order $BL(1,0;1,1)$ bilinear dependence. Higher-order cases are very similar, but notationally and technically more intricate, and are the subject of a forthcoming paper. All the ideas and conclusions developed here remain valid under the more general $BL(p,q;P,Q)$ setting.

1.3. Outline of the paper

Section 2 presents the notation and main technical assumptions to be considered in the paper, then briefly recalls the definition of local asymptotic normality and its consequences in hypothesis testing context. Section 3 starts with the definition of the statistical tools—viz., generalized versions of the spectrum (*f-autocorrelation coefficients*) and bispectrum (the so-called *cubic autocorrelation coefficients*)—to be used. The asymptotic joint normality of these generalized autocorrelation coefficients is then established (Section 3.3) under very mild assumptions. In Section 3.4, a *correlogram-based* LAN result, i.e. involving a central sequence which is measurable with respect to the residual spectrum and bispectrum, is proved: the estimated residual spectrum and bispectrum accordingly will be locally asymptotically sufficient. Section 4 deals with the asymptotic linearity property needed in order to substitute estimated residuals for exact ones. The optimal tests we are proposing are given in Section 5.1. They are based on an explicit quadratic form in the estimated residual spectrum and bispectrum. The corresponding local power is also provided. A comparison is made in Section 5.2 between the traditional Lagrange multiplier test and ours; an essential difference is that arbitrary estimates (such as M- or R- or adaptive estimates) can be used in the derivation of the residual spectrum and bispectrum, provided that they converge at an appropriate rate. Finally, the local asymptotic powers of our tests are provided explicitly. The local power of the Gaussian Lagrange multipliers method follows as a particular case.

2. Local Asymptotic Normality

2.1. Notation and technical assumptions

Denote by $H_f^{(n)}(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (a,b)'$ or by $H_f^{(n)}(a,b)$ the hypothesis under which an observed series $\mathbf{X}^{(n)} = (X_1^{(n)}, \ldots, X_n^{(n)})$ constitutes a finite realization of some solution of the bilinear stochastic difference equation (first-order diagonal or $BL(1,0;1,1)$ model)

$$X_t - aX_{t-1} - bX_{t-1}\epsilon_{t-1} = \epsilon_t, \quad t \in \mathbb{Z}, \tag{2.1}$$

where $\{\epsilon_t; t \in \mathbb{Z}\}$ is independent white noise, i.e., an i.i.d. process, with density f. The probabilistic properties of this model have been studied in detail by Pham and Tran (1981).

The following assumptions are made on f:

(A1) $f(x) > 0$, $x \in \mathbb{R}$; $\int_{-\infty}^{\infty} xf(x)dx = 0$; $\int_{-\infty}^{\infty} x^4 f(x)dx < \infty$; denoting by σ^2 the variance $\int_{-\infty}^{\infty} x^2 f(x)dx$, let $\sigma^{-1} f_1(\cdot/\sigma) = f(\cdot)$.
(A1$'$) Same as (A1), but $\int_{-\infty}^{\infty} x^6 f(x)dx < \infty$.
(A2) f is absolutely continuous on finite intervals, i.e., there exists $\dot{f}$ such that for all $-\infty < a_1 < a_2 < \infty, f(a_2) - f(a_1) = \int_{a_1}^{a_2} \dot{f}(x)dx$. Letting $\phi_f = -\dot{f}/f$, assume that the Fisher information associated with f is finite: $\int_{-\infty}^{\infty} \phi_f^2(x)f(x)dx = I(f) < \infty$.
(A2$'$) Same as (A2), but $\int_{-\infty}^{\infty} |\phi(f)|^{2+\delta} f(x)dx < \infty$ for some $\delta > 0$.
(A3) ϕ_f is differentiable, with derivative $\dot{\phi}_f$; both ϕ_f and $\dot{\phi}_f$ are Lipschitz, i.e.

$$\max\Big(|\phi_f(x) - \phi_f(y)|, |\dot{\phi}_f(x) - \dot{\phi}_f(y)|\Big) < A_f|x-y|, \tag{2.2}$$

for all $x, y \in \mathbb{R}$.

Denote by $\mathcal{F}$ (resp. $\mathcal{F}'$) the class of all densities satisfying A1-A3 (resp. A1$'$, A2$'$, A3): $\mathcal{F}'$ clearly contains most usual densities, such as the normal, the logistic, the Student with six or more degrees of freedom, … but also skew and possibly contaminated densities. We are interested in testing the null hypothesis

$$H^{(n)} = \bigcup_{g \in \mathcal{F}} H_g^{(n)} \quad \Big(\text{resp.}, \bigcup_{g \in \mathcal{F}'} H_g^{(n)}\Big), \tag{2.3}$$

where

$$H_g^{(n)} = \bigcup_{a\in(-1;1)} H_g^{(n)}(a,0). \tag{2.4}$$

Most of the results below, however, remain valid under much weaker conditions (such as A1 for Proposition 3.1 , etc.).

2.2. *LAN* and locally asymptotically most stringent tests

The key result, if locally optimal testing procedures are to be constructed, is the locally asymptotically normal (*LAN*) structure of the family of likelihoods under study. Denote by $\Lambda^{(n)}_{\boldsymbol{\tau};f}(\cdot;a)$, with $a \in (-1;1)$ and $\boldsymbol{\tau} = (\alpha,\beta)' \in \mathbb{R}^2$, the logarithm of the likelihood ratio for $H_f^{(n)}(\boldsymbol{\theta}) = H_f^{(n)}(a,0)$ against $H_f^{(n)}(\boldsymbol{\theta} + \boldsymbol{\nu}(n)\boldsymbol{\tau}) = H_f^{(n)}(a+n^{-1/2}\alpha,\, n^{-1/2}\beta/\sigma)$, where

$$\boldsymbol{\nu}(n) = n^{-\frac{1}{2}}\begin{pmatrix} 1 & 0 \\ 0 & \sigma^{-1}\end{pmatrix}.$$

The *restricted LAN* property we need is as follows:
(*LAN1*) *For all* $a \in (-1,1)$, *there exist random vectors* $\boldsymbol{\Delta}_f^{(n)}(a)$ *and nonrandom matrices* $\boldsymbol{W}_f^2(a)$ *such that* $\boldsymbol{W}_f^2(\cdot)$ *is continuous with respect to* a, *and*

$$\Lambda^{(n)}_{\boldsymbol{\tau}^{(n)};f}(\mathbf{X}^{(n)};a) = \boldsymbol{\tau}^{(n)'}\boldsymbol{\Delta}_f^{(n)}(a) - \frac{1}{2}\boldsymbol{\tau}^{(n)'}\boldsymbol{W}_f^2(a)\boldsymbol{\tau}^{(n)} + o_P(1), \tag{2.5}$$

under $H_f^{(n)}(a,0)$, *as* $n\to\infty$, *for all sequence* $\boldsymbol{\tau}^{(n)} = (\alpha^{(n)},\beta^{(n)})'$ *such that* $\sup_n \boldsymbol{\tau}^{(n)'}\boldsymbol{\tau}^{(n)} < \infty$.
(*LAN2*) *For all* $a \in (-1;1)$, $\boldsymbol{\Delta}_f^{(n)}(a)$ *is asymptotically normal under* $H_f^{(n)}(a,0)$, *as* $n\to\infty$, *with mean* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{W}_f^2(a)$.

If $\hat{a}^{(n)}$ denotes a *root n-consistent* estimate (under $H_f^{(n)}(a,0)$) of a (all usual estimates are root n-consistent under $H^{(n)}$), it follows from LeCam (1986, Ch.11) that a *locally asymptotically most stringent* test for $H_f^{(n)}$ against the alternative $\cup_{a\in(-1,1)}\cup_{b\neq 0} H_f^{(n)}\ (a,b)$, at asymptotic probability level α, consists in rejecting $H_f^{(n)}$ whenever $Q_f^{(n)}(\hat{a}^{(n)})$ exceeds the $(1-\alpha)$-quantile of a chi-square variable with one degree of freedom, where

$$Q_f^{(n)}(a) = (\boldsymbol{\Delta}_f^{(n)}(a))'\left[\boldsymbol{W}_f^{-2}(a) - (\boldsymbol{W}_f^2(a)_{11})^{-1}\begin{pmatrix} 1 & 0 \\ 0 & 0\end{pmatrix}\right]\boldsymbol{\Delta}_f^{(n)}(a). \tag{2.6}$$

This, at first sight, settles the problem (provided that *LAN* is proved). However, *LAN1* and *LAN2* are derived under fixed density f, and the asymptotic optimality against $\cup_{a\in(-1,1)}\cup_{b\neq 0} H_f^{(n)}(a,b)$ of a test based on $Q_f^{(n)}(\hat{a}^{(n)})$

holds for the null sequence $\cup_{a\in(-1,1)} H_f^{(n)}(a,0)$ only, not for the more general $\cup_{g\in\mathcal{F}} \cup_{a\in(-1,1)} H_g^{(n)}(a,0)$, as expected. The fact that $Q_f^{(n)}(\hat{a}^{(n)})$ is asymptotically chi-square under $H_f^{(n)}$ does not guarantee indeed that it remains asymptotically chi-square under $H_g^{(n)}$, $g \neq f$. Actually, it is not. The above test accordingly is not valid under $H^{(n)}$ (not even asymptotically) and, therefore, is not very attractive.

Fortunately, central sequences are never uniquely determined, and $\boldsymbol{\Delta}_f^{(n)}(a)$ in (2.5) can be replaced with any alternative sequence $\tilde{\boldsymbol{\Delta}}_f^{(n)}(a)$ such that (for all $a \in (-1;1)$)

$$\tilde{\boldsymbol{\Delta}}_f^{(n)}(a) - \boldsymbol{\Delta}_f^{(n)}(a) = o_P(1), \quad n \to \infty, \tag{2.7}$$

under $H_f^{(n)}(a,0)$. Similarly, any (possibly random) sequence of matrices $\tilde{\boldsymbol{W}}_f^{(n)2}(a)$ can be substituted for $\boldsymbol{W}_f^2(a)$ in (2.7), provided that

$$\tilde{\boldsymbol{W}}_f^{(n)2}(a^{(n)}) - \boldsymbol{W}_f^2(a) = o_P(1), \tag{2.8}$$

still under $H_f^{(n)}(a,0)$, as $n \to \infty$, for any sequence $a^{(n)} \to a$, any $a \in (-1;1)$. Such substitutions clearly do not affect the asymptotic behavior of $Q_f^{(n)}(\hat{a}^{(n)})$, neither under $H_f^{(n)}$ nor under contiguous alternatives, so that the optimality properties of the corresponding test remain unchanged. But (2.7) and (2.8) are not required to hold under $H_g^{(n)}$, $g \neq f$, and this provides quite an amount of flexibility in the definition of an optimal test statistic. The following questions then quite naturally arise: (i) is this flexibility enough to allow for tests that remain valid under $H^{(n)}$? for tests that are measurable with respect to some appropriate, intuitively interpretable, sequence of σ-algebras (e.g., generated by some adequately generalized correlograms)? (ii) does it allow for (asymptotically) invariant (rank-based) tests? etc. The purpose of the present paper is to answer the first of these two questions. Rank-based methods are the subject of another, forthcoming paper (Benghabrit and Hallin (1993)).

3. Generalized Correlograms

3.1. Generalized autocorrelation coefficients

The main tool in the analysis of linear time series models is the correlogram, i.e., (for a series $Z_1, \ldots, Z_n$) the set of autocorrelation coefficients

$$r_i^{(n)} = (n-i)^{-1} \sum_{t=i+1}^{n} Z_t Z_{t-i} / \hat{\sigma}_{(n)}^2, \quad i = 1, \ldots, n-1, \tag{3.1}$$

with $\hat{\sigma}_{(n)}^2 = n^{-1} \sum Z_t^2$ (all series here are assumed to have zero mean). Now, $r_i^{(n)}$ is a typically Gaussian statistic, the non-Gaussian version of which (for density

f) is the so-called *f-autocorrelation coefficient*

$$r_{i;f}^{(n)} = (n-i)^{-1} \sum_{t=i+1}^{n} \phi_{f_1}\left(\frac{Z_t}{\hat{\sigma}_{(n)}}\right) \frac{Z_{t-i}}{\hat{\sigma}_{(n)}} / (\hat{I}^{(n)}(f_1))^{1/2}, \quad i = 1, \dots, n-1, \tag{3.2}$$

with $\hat{I}^{(n)}(f_1) = n^{-1} \sum_t \phi_{f_1}^2(Z_t/\hat{\sigma}_{(n)})$; (cf., e.g., Hallin and Puri (1992, 1994) or Section 2.3 below). This scaling factor is such that $(n-i)^{1/2} r_{i;f}^{(n)}$, just as $(n-i)^{1/2} r_i^{(n)}$, is asymptotically standard normal under the assumptions of Proposition 3.1 below. For notational convenience, we also introduce the nonstandardized quantities

$$c_{i;f}^{(n)} = (\hat{I}^{(n)}(f_1))^{1/2} r_{i;f}^{(n)} = (n-i)^{-1} \sum_{t=i+1}^{n} \phi_{f_1}\left(\frac{Z_t}{\hat{\sigma}_{(n)}}\right) \frac{Z_{t-i}}{\hat{\sigma}_{(n)}}, \tag{3.3}$$

as well as the constants (assuming that the integrals converge)

$$I_g(f, g) = \int_{-\infty}^{\infty} \phi_{f_1}(x) \phi_{g_1}(x) g_1(x) dx \tag{3.4}$$

and

$$I_g(f) = I_g(f, f) = \int_{-\infty}^{\infty} \phi_{f_1}^2(x) g_1(x) dx. \tag{3.5}$$

For particular choices of f, it may happen that $I_g(f)$ does not depend on g and thus reduces to $I(f_1)$: this latter constant can then be used in (3.2) instead of its estimate $\hat{I}^{(n)}(f_1)$; an example is the Gaussian autocorrelations, where $I_g(\mathcal{G}) = -\int_{-\infty}^{\infty} x \phi_{g_1}(x) g_1(x) dx = 1$, so that (3.2) reduces to (3.1).

3.2. Cubic autocorrelation coefficients

The f-autocorrelation coefficients are locally inefficient in the problem of detecting departures from linearity, i.e., their asymptotic distributions are exactly the same under linearity as under local alternatives of bilinear dependence. We therefore need another class of statistics, the *cubic autocorrelation coefficients*

$$q_{i,j;f}^{(n)} = \begin{cases} (n-j)^{-1} \sum_{t=j+1}^{n} \phi_{f_1}\left(\frac{Z_t}{\hat{\sigma}_{(n)}}\right)\left(\frac{Z_{t-i}}{\hat{\sigma}_{(n)}}\right)\left(\frac{Z_{t-j}}{\hat{\sigma}_{(n)}}\right) / [\hat{I}^{(n)}(f_1)]^{1/2}, & \text{if } 0 < i < j < n, \\ (n-i)^{-1} \sum_{t=i+1}^{n} \phi_{f_1}\left(\frac{Z_t}{\hat{\sigma}_{(n)}}\right) \frac{Z_{t-i}^2}{\hat{\sigma}_{(n)}^2} / [\hat{I}^{(n)}(f_1) \hat{m}_4^{(n)}]^{1/2}, & \text{if } 0 < i = j < n, \end{cases} \tag{3.6}$$

where $\hat{m}_4^{(n)} = n^{-1} \sum_t (Z_t^{(n)})^4 / \hat{\sigma}_{(n)}^4$. For Gaussian f, (3.6) takes the much simpler form

$$q_{i,j}^{(n)} = \begin{cases} (n-j)^{-1} \sum_{t=j+1}^{n} Z_t Z_{t-i} Z_{t-j} / \hat{\sigma}_{(n)}^3, & \text{if } 0 < i < j < n, \\ (n-i)^{-1} \sum_{t=i+1}^{n} Z_t Z_{t-i}^2 / \hat{\sigma}_{(n)}^3 (\hat{m}_4^{(n)})^{1/2}, & \text{if } 0 < i = j < n, \end{cases} \tag{3.7}$$

which (under the terminology *bispectrum*) constitutes the main tool in bilinear model identification; (see Gabr (1988)). For notational convenience, we also introduce the non standardized quantities

$$d_{i,j;f}^{(n)} = (n-j)^{-1} \sum_{t=j+1}^{n} \phi_{f_1}\Big(\frac{Z_t}{\hat{\sigma}_{(n)}}\Big) \frac{Z_{t-i}}{\hat{\sigma}_{(n)}} \frac{Z_{t-j}}{\hat{\sigma}_{(n)}}, \qquad 0 < i \leq j < n. \tag{3.8}$$

3.3. Asymptotic normality

In the sequel, we only need cubic autocorrelation coefficients of the form $q_{1,\ell;f}^{(n)}, \ell \geq 1$, which we henceforth denote by $q_{\ell;f}^{(n)}$. For simplicity, the following central-limit theorem is restricted to that particular case. A similar result could be stated in the general case.

Proposition 3.1. *Let f satisfy Assumptions* A1 *and* A2. *Assume that $Z_1, \ldots, Z_n$ are i.i.d., with density g, where g satisfies* A1 *and is such that $\int_{-\infty}^{\infty} \phi_f(x) g(x) dx = 0$ and $I_g(f) < \infty$. Then , for all k and ℓ,*

$$\Big((n-1)^{1/2} r_{1;f}^{(n)}, \ldots, (n-k)^{1/2} r_{k;f}^{(n)}, (n-1)^{1/2} q_{1;f}^{(n)}, \ldots, (n-\ell)^{1/2} q_{\ell;f}^{(n)}\Big) \tag{3.9}$$

is asymptotically multinormal, with mean $\mathbf{0}$ *and covariance matrix*

$$\begin{pmatrix} \mathbf{I}_{k\times k} & \mathbf{C}_{f;g} \\ \mathbf{C}'_{f;g} & \mathbf{I}_{\ell\times\ell} \end{pmatrix}, \tag{3.10}$$

where $(\mathbf{C}_{f;g})_{ij} = (\frac{m_3}{\sigma^3})_g (\frac{m_4}{\sigma^4})_g^{-1/2} I[i = j = 1]$, *with* $(m_i/\sigma^i)_g = \int_{-\infty}^{\infty} x^i g_1(x) dx$, $i = 3, 4$.

Proof. The proof is straightforward, and we only briefly sketch it here. Consider arbitrary linear combinations of the components of (3.9): they are sums of $\max(k, \ell)$-dependent variables, and their asymptotic normality results from the classical Robbins-Hoeffding central-limit theorem for m-dependent summands (see, e.g., Anderson (1971) p.427). Application of the usual Cramér-Wold argument completes the proof.

Note that the assumptions of the proposition are satisfied in the following two cases:

- f is Gaussian and g satisfies A1: then $\int_{-\infty}^{\infty} \phi_f g dx$ reduces to $\int_{-\infty}^{\infty} x g(x) dx = 0$, and $\int_{-\infty}^{\infty} \phi_f^2(x) g(x) dx$ to $\int_{-\infty}^{\infty} x^2 g(x) dx = \sigma_g^2$;
- f and g are symmetric with respect to the origin, g satisfies A1 and $I_g(f) < \infty$.

Asymptotic normality still holds under densities g such that $\int \phi_f(x) g(x) dx \neq 0$, but the centering terms and asymptotic covariance matrices are then considerably more complicated.

3.4. A correlogram-based LAN result

Assume that X_0 and ϵ_0 are observable. These starting values have no influence upon asymptotic results (their influence on central sequences is $o_P(1)$), but they allow for a considerable simplification of the form of likelihood functions. Thus the likelihood functions and log-likelihood ratios $\Lambda^{(n)}_{\boldsymbol{\tau};f}(\cdot;a)$ considered here (see Section 2.2) are those associated with $(\epsilon_0, X_0, \mathbf{X}^{(n)})$.

Proposition 3.2. *Assume that f satisfies* A1′, A2′ *and* A3. *Let* $Z_t = Z_t^{(n)}(a) = X_t^{(n)} - aX_{t-1}^{(n)}$, $t = 1, \ldots, n$. *Then*

(i) *LAN1 and LAN2 hold, with*

$$\boldsymbol{\Delta}_f^{(n)}(a) = \sum_{i=1}^{n-1} a^{i-1}(n-i)^{1/2} \begin{pmatrix} c_{i;f}^{(n)} \\ d_{i;f}^{(n)} \end{pmatrix} \tag{3.11}$$
$$= [\hat{I}^{(n)}(f_1)]^{1/2} \left\{ (n-1)^{1/2} \begin{pmatrix} r_{1;f}^{(n)} \\ (\hat{m}_4^{(n)})^{1/2} q_{1;f}^{(n)} \end{pmatrix} + \sum_{i=2}^{n-1} a^{i-1}(n-i)^{1/2} \begin{pmatrix} r_{i;f}^{(n)} \\ q_{i;f}^{(n)} \end{pmatrix} \right\}$$

and

$$\mathbf{W}_f^2(a) = I(f_1) \begin{pmatrix} (1-a^2)^{-1} & (\frac{m_3}{\sigma^3})_f \\ (\frac{m_3}{\sigma^3})_f & (\frac{m_4}{\sigma^4})_f + \frac{\sigma^2}{1-a^2} \end{pmatrix}. \tag{3.12}$$

(ii) *under* $H_g^{(n)}(a,0)$, *where g satisfies* A1, $\int_{-\infty}^{\infty} \phi_f g dx = 0$ *and* $I_g(f) < \infty$, *as* $n \to \infty$, $\boldsymbol{\Delta}_f^{(n)}(a)$ *is asymptotically normal, with mean* $\mathbf{0}$ *and covariance matrix*

$$\mathbf{W}_{f;g}^2(a) = I_g(f) \begin{pmatrix} (1-a^2)^{-1} & (\frac{m_3}{\sigma^3})_g \\ (\frac{m_3}{\sigma^3})_g & (\frac{m_4}{\sigma^4})_g + \frac{a^2}{1-a^2} \end{pmatrix}. \tag{3.13}$$

Proof. See Appendix A.

Note that, for Gaussian f, (3.12) yields

$$\mathbf{W}_{\mathcal{G}}^2(a) = \begin{pmatrix} (1-a^2)^{-1} & 0 \\ 0 & 3 + \frac{a^2}{1-a^2} \end{pmatrix}, \tag{3.14}$$

whereas under an arbitrary innovation density g, (3.13) (still for Gaussian f) in general is not diagonal, with

$$\mathbf{W}_{\mathcal{G};g}^2(a) = \begin{pmatrix} (1-a^2)^{-1} & (\frac{m_3}{\sigma^3})_g \\ (\frac{m_3}{\sigma^3})_g & (\frac{m_4}{\sigma^4})_g + \frac{a^2}{1-a^2} \end{pmatrix}; \tag{3.15}$$

this latter matrix clearly strongly depends on the skewness and kurtosis of the underlying (in practice, unknowm) density g.

4. Local Asymptotic Linearity

4.1. Asymptotic distributions under local alternatives

The *LAN* result of Proposition 3.2 and LeCam's so-called third lemma also provide for asymptotic distributions under local alternatives of bilinear dependence. Inspecting the form ((4.3) below) of $\mathbf{\Delta}_f^{(n)}(a)$'s asymptotic mean immediately reveals the main problem we are faced with: the skewness of the underlying density g induces the same type of asymptotic behavior as does the presence of a bilinear term in the model.

Proposition 4.1. *Assume that f satisfies* A1 *and* A2. *Let g satisfy* A1′, A2′, A3, *and be such that $\int_{-\infty}^{\infty} \phi_f g dx = 0$ and $I_g(f) < \infty$. Denote by $(\alpha^{(n)}, \beta^{(n)})$ a bounded sequence*: $\sup_n(|\alpha^{(n)}| + |\beta^{(n)}|) < \infty$. *Then, under $H_g^{(n)}(a + n^{-1/2}\alpha^{(n)}, n^{-1/2}\beta^{(n)}/\sigma)$, where σ^2 denotes the variance associated with g, as $n \to \infty$,*

(i) *the joint distribution of* (3.9) *is asymptotically multinormal, still with covariance matrix* (3.10), *but with centering terms*

$$\begin{cases} [\alpha^{(n)} + \beta^{(n)}(\frac{m_3}{\sigma^3})_g](I_g(f))^{-1/2} I_g(f,g), & i = 1, \\ \alpha^{(n)} a^{i-1}(I_g(f))^{-1/2} I_g(f,g), & i > 1, \end{cases} \tag{4.1}$$

for $(n-i)^{1/2} r_{i;f}^{(n)}$, and

$$\begin{cases} [\alpha^{(n)}(\frac{m_3}{\sigma^3})_g + \beta^{(n)}(\frac{m_4}{\sigma^4})_g](\frac{m_4}{\sigma^4})_g^{-1/2}(I_g(f))^{-1/2} I_g(f,g), & i = 1, \\ \beta^{(n)} a^{i-1}(I_g(f))^{-1/2} I_g(f,g), & i > 1, \end{cases} \tag{4.2}$$

for $(n-i)^{1/2} q_{i;f}^{(n)}$.

(ii) *$\mathbf{\Delta}_f^{(n)}(a)$ is asymptotically bivariate normal, with mean*

$$\left(\alpha^{(n)}(1-a^2)^{-1} + \beta^{(n)}(\frac{m_3}{\sigma^3})_g, \alpha^{(n)}(\frac{m_3}{\sigma^3})_g + \beta^{(n)}\left[(\frac{m_4}{\sigma^4})_g + \frac{a^2}{1-a^2}\right]\right)' I_g(f,g), \tag{4.3}$$

and covariance matrix (3.12).

Proof. Here, again, the proof is lengthy but easy. Denoting by

$$\mathbf{\Delta}_{g;\kappa}^{(n)}(a) = \sum_{i=1}^{\kappa} a^{i-1}(n-i)^{1/2} \begin{pmatrix} c_{i;g}^{(n)} \\ d_{i;g}^{(n)} \end{pmatrix}$$

a truncated version of $\mathbf{\Delta}_g^{(n)}(a)$, the asymptotic joint normality of (3.9) and $\mathbf{\Delta}_{g;\kappa}^{(n)}(a)$ follows along the same lines as in the proof of Proposition 3.1. Now, the difference $\mathbf{\Delta}_g^{(n)}(a) - \mathbf{\Delta}_{g;\kappa}^{(n)}(a)$ converges to zero in probability, as $\kappa \to \infty$, uniformly in n, i.e., for all $\epsilon > 0$ and $\eta > 0$, there exists K such that $P[\|\mathbf{\Delta}_g^{(n)}(a) - \mathbf{\Delta}_{g;\kappa}^{(n)}(a)\| > \epsilon] < \eta$ for all $\kappa > K$ and all n. Applying Anderson's (1971) Theorem 7.7.1

to arbitrary linear combinations of (3.9) and $\mathbf{\Delta}_g^{(n)}(a)$, along with the classical Cramér-Wold argument then provides the joint asymptotic normality of (3.9) and $\mathbf{\Delta}_g^{(n)}$, hence that of (3.9) and $\Lambda^{(n)}_{\boldsymbol{\tau}^{(n)};g}$. LeCam's third Lemma completes the proof of Part(i) of the proposition, from which Part(ii) is an immediate consequence. Details are left to the reader.

4.2. Local asymptotic linearity

Asymptotic linearity results play an essential role when estimated residuals are to be substituted for the exact ones in test statistics. Such results have been obtained in Kreiss (1987, (4.5)) and Hallin and Puri (1994), Proposition 5.1 (iii) for f-autocorrelation coefficients; their analogues for cubic autocorrelation coefficients are derived here.

Proposition 4.2. *Assume that f satisfies* A1 *and* A2. *Let g satisfy* A1′, A2′, A3, *and be such that $\int_{-\infty}^{\infty}\phi_f g dx = 0$ and $I_g(f) < \infty$. Then, under $H_g^{(n)}(a)$, as $n \to \infty$, for all bounded sequences $\alpha^{(n)}$ (such that* $\sup_n[\alpha^{(n)}]^2 < \infty$*),*

$$(n-i)^{1/2}(r_{i;f}^{(n)}(a+n^{-1/2}\alpha^{(n)}) - r_{i;f}^{(n)}(a)) + \alpha^{(n)}a^{i-1}(I_g(f))^{-1/2}I_g(f,g) = o_P(1),\quad i = 1,2,\ldots, \tag{4.4}$$

$$(n-1)^{1/2}(q_{1;f}^{(n)}(a+n^{-1/2}\alpha^{(n)}) - q_{1;f}^{(n)}(a)) + \alpha^{(n)}(\frac{m_3}{\sigma^3})_g(\frac{m_4}{\sigma^4})_g^{-1/2}(I_g(f))^{-1/2}I_g(f,g) = o_P(1), \tag{4.5}$$

and, for $i = 2,3,\ldots$,

$$(n-i)^{1/2}(q_{i;f}^{(n)}(a+n^{-1/2}\alpha^{(n)}) - q_{i;f}^{(n)}(a)) = o_P(1). \tag{4.6}$$

Proof. See Appendix B.

Denote by $\hat{a}^{(n)}$ an estimate of a satisfying the following assumptions.

(A4) (i) $\hat{a}^{(n)}$ is *root n-consistent*, i.e., for all $f \in \mathcal{F}$, all $a \in (-1,1)$, all $\epsilon > 0$, there exist $\eta = \eta(f;a;\epsilon)$ and $N = N(f;a;\epsilon)$ such that, under $H_f^{(n)}(a)$, $P[n^{1/2}|\hat{a}^{(n)} - a| > \eta] < \epsilon$, all $n \geq N$. (ii) $\hat{a}^{(n)}$ is *locally discrete*, i.e. for all $f \in \mathcal{F}$, all $a \in (-1,1)$, all fixed $c > 0$, the number of possible values of $\hat{a}^{(n)}$ in intervals of the form $[a \pm cn^{-1/2}]$ remains bounded, under $H_f^{(n)}(a)$, as $n \to \infty$.

Assumption A4(i) is satisfied by all usual estimates (e.g., by $a^{(n)} = \sum X_t X_{t-1}/\sum X_t^2$; but "approximate" or robust estimates can be considered as well). As for Assumption A4(ii), though discretization techniques can be used (see LeCam (1960), Hallin and Puri (1994), Section 5.3), it has no practical implications and safely can be ignored (except for the purpose of proving convergence theorems). The number of digits in numerical applications indeed is always strictly bounded. We then have the following Corollary to Proposition 4.2.

Corollary 4.2. *Let* $\hat{a}^{(n)}$ *satisfy* A4(i) *and* (ii). *Denote by* $\hat{r}^{(n)}_{i;f}$, $\hat{q}^{(n)}_{i;f}$ *and* $\hat{\mathbf{\Delta}}^{(n)}_f$, *respectively, the correlation coefficients and central sequence* $r^{(n)}_{i;f}(\hat{a}^{(n)})$, $q^{(n)}_{i;f}(\hat{a}^{(n)})$ *and* $\mathbf{\Delta}^{(n)}_f(\hat{a}^{(n)})$ *computed from the resulting estimated residuals* $Z^{(n)}_t(\hat{a}^{(n)})$. *Then, under the assumptions of Proposition* 4.2,

$$(n-i)^{1/2}(\hat{r}^{(n)}_{i;f}-r^{(n)}_{i;f}) = -a^{i-1}(I_g(f))^{-1/2}I_g(f,g)n^{1/2}(\hat{a}^{(n)}-a)+o_P(1), i=1,2,\ldots, \tag{4.7}$$

$$(n-1)^{1/2}(\hat{q}^{(n)}_{1;f}-q^{(n)}_{1;f}) = -(\frac{m_3}{\sigma^3})_g(\frac{m_4}{\sigma^4})_g^{-1/2}(I_g(f))^{-1/2}I_g(f,g)n^{1/2}(\hat{a}^{(n)}-a)+o_P(1), \tag{4.8}$$

$$(n-i)^{1/2}(\hat{q}^{(n)}_{i;f}-q^{(n)}_{i;f}) = o_P(1), \quad i=2,3,\ldots \tag{4.9}$$

and

$$\hat{\mathbf{\Delta}}^{(n)}_f(\hat{a}^{(n)}) - \mathbf{\Delta}^{(n)}_f(a) = -\begin{pmatrix} (1-a^2)^{-1} \\ (\frac{m_3}{\sigma^3})_g \end{pmatrix} I_g(f,g)n^{1/2}(\hat{a}^{(n)}-a)+o_P(1).$$

The proof readily follows from an argument which goes back to LeCam (see Kreiss (1987), Lemma 4.4 for a formal proof).

5. Locally Asymptotically Optimal Tests

5.1. General case

We now can state the main result of the paper. Let

$$\hat{m}^{(n)}_3(\hat{a}^{(n)}) = \frac{n^{-1}\sum_{t=1}^n[Z^{(n)}_t(\hat{a}^{(n)})]^3}{n^{-1}[\sum_{t=1}^n[Z^{(n)}_t(\hat{a}^{(n)})]^2]^{3/2}} \tag{5.1}$$

and

$$\hat{m}^{(n)}_4(\hat{a}^{(n)}) = \frac{n^{-1}\sum_{t=1}^n[Z^{(n)}_t(\hat{a}^{(n)})]^4}{[n^{-1}\sum_{t=1}^n[Z^{(n)}_t(\hat{a}^{(n)})]^2]^2}. \tag{5.2}$$

Proposition 5.1. *Assume that* f *satisfies* A1 *and* A2. *Let* g *satisfy* A1′, A2′, A3, *and be such that* $\int_{-\infty}^{\infty}\phi_f g dx = 0$ *and* $I_g(f) < \infty$. *Let* $\hat{a}^{(n)}$ *satisfy* A4. *Then, the test rejecting* $H^{(n)}$ *whenever*

$$\hat{Q}^{(n)}_f = \frac{1-(\hat{a}^{(n)})^2}{\hat{I}^{(n)}(f_1)[\hat{m}^{(n)}_4(1-(\hat{a}^{(n)})^2)+(\hat{a}^{(n)})^2-(\hat{m}^{(n)}_3)^2(1-(\hat{a}^{(n)})^2)^2]} \tag{5.3}$$
$$\times\, \mathbf{\Delta}^{(n)}_f(\hat{a}^{(n)})'\begin{pmatrix} (\hat{m}^{(n)}_3)^2(1-(\hat{a}^{(n)})^2) & -\hat{m}^{(n)}_3(1-(\hat{a}^{(n)})^2) \\ -\hat{m}^{(n)}_3(1-(\hat{a}^{(n)})^2) & 1 \end{pmatrix}\mathbf{\Delta}^{(n)}_f(\hat{a}^{(n)})$$

exceeds the $(1-\alpha)$*-quantile* $\chi^2_{1-\alpha}$ *of a chi-square variable with one degree of freedom*

(i) *has asymptotic level* α *under* $H_g^{(n)}$,
(ii) *is locally asymptotically most stringent against* $\cup_{a\in(-1,1)}\cup_{b\neq 0} H_f^{(n)}(a,b)$ (*at asymptotic probability level* α),
(iii) *has asymptotic power*

$$1-\Phi\left(z_{\alpha/2}+\beta((\frac{m_4}{\sigma^4})_g+\frac{a^2}{1-a^2}-(\frac{m_3}{\sigma^3})_g^2(1-a^2))^{1/2}I_g(f,g)(I_g(f))^{-1/2}\right)$$
$$+\Phi\left(-z_{\alpha/2}+\beta((\frac{m_4}{\sigma^4})_g+\frac{a^2}{1-a^2}-(\frac{m_3}{\sigma^3})_g^2(1-a^2))^{1/2}I_g(f,g)(I_g(f))^{-1/2}\right) \quad (5.4)$$

against $H_g^{(n)}(a,n^{-1/2}\beta/\sigma)$, *where* z_α *and* Φ *stand for the* $(1-\alpha)$*-quantile and distribution function of the standard normal variable, respectively.*
(iv) *If moreover* $\hat{a}^{(n)}$ *is the (pseudo-)maximum likelihood estimator associated with* f, *then* $\mathbf{\Delta}_f^{(n)}(\hat{a}^{(n)})_1=o_P(1)$, *and* $\hat{Q}_f^{(n)}$ *reduces to*

$$\hat{Q}_f^{(n)}=\frac{[1-(\hat{a}^{(n)})^2][\mathbf{\Delta}_f^{(n)}(\hat{a}^{(n)})_2]^2}{\hat{I}^{(n)}(f_1)[\hat{m}_4^{(n)}(1-(\hat{a}^{(n)})^2)+(\hat{a}^{(n)})^2-(\hat{m}_3^{(n)})^2(1-(\hat{a}^{(n)})^2)^2]}.$$

Proof. The form of the test statistic (5.3) follows from (3.13) and the general result (2.6) by an explicit computation of

$$\mathbf{W}_f^{-2}(\hat{a}^{(n)})-\left[(\mathbf{W}_f^2(\hat{a}^{(n)}))_{11}\right]^{-1}\begin{pmatrix}1&0\\0&0\end{pmatrix}.$$

Define

$$\begin{aligned}Q^{(n)}(a)&=(\mathbf{\Delta}_f^{(n)}(a))'(\mathbf{W}_f^2(\hat{a}^{(n)}))^{-1/2}\\&\times\{\mathbf{I}-(\mathbf{W}_f^2(\hat{a}^{(n)}))^{1/2}(\hat{I}^{(n)}(f_1))^{-1}\begin{pmatrix}1-(\hat{a}^{(n)})^2&0\\0&0\end{pmatrix}(\mathbf{W}_f^2(\hat{a}^{(n)}))^{1/2}\}\\&\times(\mathbf{W}_f^2(\hat{a}^{(n)}))^{-1/2}\mathbf{\Delta}_f^{(n)}(a).\end{aligned}\quad (5.5)$$

The matrix between braces is symmetric and idempotent, of rank one, and, since $\mathbf{W}_f^2(\cdot)$ has been assumed continuous, $(\mathbf{W}_f^2(\hat{a}^{(n)}))^{-1/2}\mathbf{\Delta}_f^{(n)}(a)$ under $H_f^{(n)}(a,0)$ is asymptotically $N(\mathbf{0},\mathbf{I})$; $Q^{(n)}(a)$ therefore is asymptotically chi square under $H_f^{(n)}(a,0)$, with one degree of freedom. Substituting $\hat{a}^{(n)}$, $\hat{m}_3^{(n)}$ and $\hat{m}_4^{(n)}$ for a, m_3/σ^3 and m_4/σ^4, respectively, in (5.5) yields $Q^{(n)}(\hat{a}^{(n)})=\hat{Q}_f^{(n)}$. Now,

$$\begin{aligned}&Q^{(n)}(\hat{a}^{(n)})-Q^{(n)}(a)\\&=(\mathbf{\Delta}_f^{(n)}(\hat{a}^{(n)})+\mathbf{\Delta}_f^{(n)}(a))'\left\{(\mathbf{W}_f^2(\hat{a}^{(n)}))^{-1}-(\hat{I}^{(n)}(f_1))^{-1}\begin{pmatrix}1-(\hat{a}^{(n)})^2&0\\0&0\end{pmatrix}\right\}\\&\quad\times(\mathbf{\Delta}_f^{(n)}(\hat{a}^{(n)})-\mathbf{\Delta}_f^{(n)}(a)),\end{aligned}$$

where, from Corollary 4.2,

$$\mathbf{\Delta}_f^{(n)}(\hat{a}^{(n)})-\mathbf{\Delta}_f^{(n)}(a)=\xi^{(n)}((1-(\hat{a}^{(n)})^2)^{-1},\hat{m}_3^{(n)})'+o_P(1)$$

under $H_g^{(n)}(a)$, as $n \to \infty$, and $\xi^{(n)}$ itself is $O_P(1)$. And, since $\hat{I}^{(n)}(f_1)((1 - (\hat{a}^{(n)})^2)^{-1}, \hat{m}_3^{(n)})$ coincides with $\mathbf{W}_f^2(\hat{a}^{(n)})$'s first row,

$$\left\{(\mathbf{W}_f^2(\hat{a}^{(n)}))^{-1} - (\hat{I}^{(n)}(f_1))^{-1}\begin{pmatrix}1-(\hat{a}^{(n)})^2 & 0\\ 0 & 0\end{pmatrix}\right\}\begin{pmatrix}(1-(\hat{a}^{(n)})^2)^{-1}\\ \hat{m}_3^{(n)}\end{pmatrix}$$
$$= (\hat{I}^{(n)}(f_1))^{-1}\left(\begin{pmatrix}1\\0\end{pmatrix} - \begin{pmatrix}1\\0\end{pmatrix}\right) = \mathbf{0}.$$

This entails the asymptotic equivalence, under $H_g^{(n)}(a)$ (hence, also under the contiguous $H_g^{(n)}(a, n^{-1/2}b)$), of $Q^{(n)}(a)$ and the test statistic $Q_f^{(n)}(\hat{a}^{(n)})$, and completes the proof of part (i) of the proposition.

Part (ii) now is obvious, since (5.3) under $H_f^{(n)}(a)$ is of the form $Q_f^{(n)}(\hat{a}^{(n)})$, with $Q_f^{(n)}(\cdot)$ given in (2.6). Part (iii) follows from the same asymptotic equivalence (under $H_g^{(n)}(a, n^{-1/2}b)$) as above, and the asymptotic distribution of $\mathbf{\Delta}_f^{(n)}(a)$ given in Proposition 4.1 (ii): $\hat{Q}_f^{(n)}$ then is asymptotically non central chi-square, still with one degree of freedom, and with noncentrality parameter

$$\beta^2[(\frac{m_4}{\sigma^4})_g + \frac{a^2}{1-a^2} - (\frac{m_3}{\sigma^3})_g^2(1-a^2)]I_g^2(f,g)(I_g(f))^{-1}.$$

The explicit local power finally results from expressing the distribution function of the noncentral chi-square variable with one degree of freedom in terms of the standard normal one Φ.

Note that the local power (5.4) is a complicated function of $a^2/(1-a^2)$, $(m_3/\sigma^3)_g^2$, $(m_4/\sigma^4)_g$ and $I_g(f,g)(I_g(f))^{-1/2}$. For given β^2 and g, bilinearity is more easily detected as a^2 gets closer to one. Under Gaussian f, and under the assumption that $(m_3/\sigma^3)_g = 0$ and $(m_4/\sigma^4)_g = 3$, we retrieve Guégan and Pham's (1992) noncentrality parameter $\beta^2[3 + a^2/(1-a^2)]$.

5.2. The Gaussian (Lagrange multiplier) test

A popular method for deriving locally optimal tests is the Lagrange multiplier method applied to Gaussian likelihood functions. This approach has been considered (in the present context of testing against bilinear dependence) by Saikkonen and Luukkonen (1988, 1991) and Guégan and Pham (1992). The resulting test asymptotically coincides with the optimal test (associated with Gaussian densities f) described here when the stronger assumption is made that $\hat{a}^{(n)}$ is a Gaussian maximum likelihood or least square estimate (see Part (iv) of Proposition 5.1).

The traditional derivation of Lagrange multiplier tests relies on the analytical expression of Gaussian likelihoods, so that special attention has to be given to the matrices involved in the quadratic forms: if (5.3) and (2.7) are based on the asymptotic covariance matrix (3.14), the much simpler test statistic indeed is obtained:

$$\begin{aligned} Q_{\mathcal{G}}^{(n)}(\hat{a}^{(n)}) &= (1-(\hat{a}^{(n)})^2)/(3-2(\hat{a}^{(n)})^2)(\boldsymbol{\Delta}_{\mathcal{G}}^{(n)}(\hat{a}^{(n)}))_2^2 \\ &= \left[(1-(\hat{a}^{(n)})^2)/(3-2(\hat{a}^{(n)})^2)(\hat{\sigma}^{(n)})^3\right] \\ &\quad \times \left(\sum_{i=1}^{n-1}(\hat{a}^{(n)})^{i-1}(n-i)^{-1/2}\sum_{t=i+1}^{n}\hat{Z}_t\hat{Z}_{t-1}\hat{Z}_{t-i}\right)^2, \end{aligned}$$

with $\hat{Z}_t = X_t - \hat{a}^{(n)}X_{t-1}$. Clearly, $Q_{\mathcal{G}}^{(n)}(\hat{a}^{(n)})$ is asymptotically chi-square under AR(1) dependence with Gaussian innovation process, but not (in view of (3.15)) under arbitrary innovation densities g (as soon as g either is skew, or has kurtosis $(m_4/\sigma^4)_g \neq 3$). The resulting Lagrange multiplier test is not (even asymptotically) valid under arbitrary densities. In order to avoid this, Guégan and Pham (1992) make the rather restrictive assumptions that $(\frac{m_3}{\sigma^3})_g = 0$ and $(\frac{m_4}{\sigma^4})_g = 3$. The test statistic proposed by Saikkonen and Luukkonen (1988), however, avoids this defect and, after simple algebra, easily reduces to the particular form given in part (iv) of Proposition 5.1.

Acknowledgement

The research of the second author was 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.

Appendix A: Proof of Proposition 3.2

All limits and expectations are taken under $H_f^{(n)}(a,0)$, as $n \to \infty$. For simplicity, we omit the superscript n in $X_t^{(n)}$ and $Z_t^{(n)}$. It follows from the invertibility results of Pham and Tran (1981) that, for n sufficiently large, the log-likelihood $\Lambda_{\boldsymbol{\tau}^{(n)}}^{(n)}$ (with $\boldsymbol{\tau}^{(n)} = (\alpha^{(n)}, \beta^{(n)})'$) decomposes into

$$\Lambda_{\boldsymbol{\tau}^{(n)};f}^{(n)} = \sum_{t=1}^{n}\left[\log f(Z_t - Y_t^{(n)}) - \log f(Z_t)\right] + o_P(1),$$

with

$$\begin{aligned} Y_t^{(n)} &= n^{-\frac{1}{2}}(\alpha^{(n)} + \sigma^{-1}\beta^{(n)}X_{t-1})X_{t-1} \\ &\quad + \sum_{j=2}^{t-1}(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j-1}(a + n^{-\frac{1}{2}}(\alpha^{(n)} + \sigma^{-1}\beta^{(n)}X_{t-j}))\prod_{k=1}^{j}X_{t-k} \\ &\quad - (-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{t-1}\left(\prod_{k=1}^{t-1}X_{t-k}\right)\left[(a + n^{-\frac{1}{2}}\alpha^{(n)})X_0 + n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)}X_0\epsilon_0\right]. \end{aligned}$$

Since ϕ_f is differentiable, for some sequence $\delta_t^{(n)} \in (0,1)$,

$$\begin{aligned}\Lambda^{(n)}_{\boldsymbol{\tau}^{(n)};f} &= \sum_{t=2}^{n} \phi_f(Z_t) Y_t^{(n)} - \frac{1}{2}\sum_{t=2}^{n} (Y_t^{(n)})^2 \dot{\phi}(Z_t - \delta_t^{(n)} Y_t^{(n)}) + o_P(1) \\ &= \Sigma_1^{(n)} - \frac{1}{2}\Sigma_2^{(n)} + o_P(1).\end{aligned}$$

The first term in this latter expression in turn decomposes into

$$\begin{aligned}\Sigma_1^{(n)} &= n^{-\frac{1}{2}} \sum_{t=2}^{n} \phi_f(Z_t) \left[(\alpha^{(n)} + \sigma^{-1}\beta^{(n)} X_{t-1}) X_{t-1} - \sigma^{-1} a\beta^{(n)} X_{t-1} X_{t-2}\right] \\ &\quad - (n\sigma)^{-1} \sum_{t=2}^{n} \phi_f(Z_t) \left[(\alpha^{(n)} + \sigma^{-1}\beta^{(n)} X_{t-2}) X_{t-1} X_{t-2}\right] \\ &\quad + \sum_{t=2}^{n} \phi_f(Z_t) \left\{ \sum_{j=3}^{t-1} (-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j-1} (a + n^{-\frac{1}{2}} (\alpha^{(n)} + \sigma^{-1}\beta^{(n)} X_{t-1})) \prod_{k=1}^{j} X_{t-k} \right\} \\ &\quad + \sum_{t=2}^{n} \phi_f(Z_t) (-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{t-1} \prod_{k=1}^{t-1} X_{t-k} \left[(a + n^{-\frac{1}{2}}\alpha^{(n)}) X_0 + n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)} X_0 \epsilon_0\right] \\ &= \Sigma_{11}^{(n)} + \Sigma_{12}^{(n)} + \Sigma_{13}^{(n)} + \Sigma_{14}^{(n)}, \text{ say.}\end{aligned}$$

The identity $X_t = \sum_{i=0}^{t-1} a^i Z_{t-i} + a^t X_0$ and some algebra yield

$$\begin{aligned}\Sigma_{11}^{(n)} &= n^{-\frac{1}{2}} \sum_{t=2}^{n} \phi_f(Z_t) \left\{ \alpha^{(n)} \sum_{i=0}^{t-2} a^i Z_{t-i-1} \right. \\ &\quad \left. + \sigma^{-1}\beta^{(n)} \left[\left(\sum_{i=0}^{t-2} a^i Z_{t-i-1} \right)^2 - \left(\sum_{i=0}^{t-2} a^i Z_{t-i-1} \right) \left(\sum_{i=0}^{t-3} a^{i+1} Z_{t-i-2} \right) \right] \right\} + o_P(1) \\ &= n^{-\frac{1}{2}} \sum_{t=2}^{n} \phi_f(Z_t) \left\{ \alpha^{(n)} \sum_{i=0}^{t-2} a^i Z_{t-i-1} + \sigma^{-1}\beta^{(n)} \sum_{i=0}^{t-2} a^i Z_{t-1} Z_{t-i-1} \right\} + o_P(1) \\ &= \alpha^{(n)} \sum_{i=1}^{n-1} a^{i-1} (n-i)^{-\frac{1}{2}} \sum_{t=i+1}^{n} \phi_{f_1}\left(\frac{Z_t}{\hat{\sigma}^{(n)}}\right) \frac{Z_{t-i}}{\hat{\sigma}^{(n)}} \\ &\quad + \beta^{(n)} \sum_{i=1}^{n-1} a^{i-1} (n-i)^{-\frac{1}{2}} \sum_{t=i+1}^{n} \phi_{f_1}\left(\frac{Z_t}{\hat{\sigma}^{(n)}}\right) \frac{Z_{t-1}}{\hat{\sigma}^{(n)}} \frac{Z_{t-i}}{\hat{\sigma}^{(n)}} + o_P(1) \\ &= (\boldsymbol{\tau}^{(n)})' \boldsymbol{\Delta}_f^{(n)}(a) + o_P(1).\end{aligned}$$

The remaining terms $\Sigma_{12}^{(n)}$, $\Sigma_{13}^{(n)}$ and $\Sigma_{14}^{(n)}$ all are $o_P(1)$'s. Considering for instance $\Sigma_{13}^{(n)}$,

$$\begin{aligned}\Sigma_{13}^{(n)} &= a(\beta^{(n)})^2(n\sigma^2)^{-1}\sum_{t=2}^{n}\phi_f(Z_t)X_{t-1}X_{t-2}X_{t-3}\\ &\quad + a\sum_{t=2}^{n}\sum_{j=4}^{t-1}(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j-1}\phi_f(Z_t)\prod_{k=1}^{j}X_{t-k}\\ &\quad + \alpha n^{-\frac{1}{2}}\sum_{t=2}^{n}\sum_{j=3}^{n}(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j-1}\phi_f(Z_t)\prod_{k=1}^{j}X_{t-k}\\ &\quad - \sum_{t=2}^{n}\sum_{j=3}^{n}(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j}\phi_f(Z_t)\prod_{k=1}^{j}X_{t-k}\\ &= \Sigma_{131}^{(n)} + \Sigma_{132}^{(n)} + \Sigma_{133}^{(n)} - \Sigma_{134}^{(n)}.\end{aligned}$$

The first term $\Sigma_{131}^{(n)}$ clearly converges in probability to $E[\phi_f(Z_4)X_3X_2X_1] = 0$. Since $|\prod_{k=1}^{j}X_{t-k}|^{-j}$ converges a.s. to $\exp[E\log|X_1|]$, $|\Sigma_{132}^{(n)}|$ is a.s. bounded by

$$|a|\sum_{t=2}^{n}\sum_{j\geq 4}(n^{-\frac{1}{2}}\sigma^{-1}|\beta^{(n)}|)^{j-1}|\phi_f(Z_t)|[\exp(E\log|X_1|+C)]^j$$

which in turn converges to zero almost surely. The third and fourth terms are handled similarly, as well as $\Sigma_{12}^{(n)}$, $\Sigma_{13}^{(n)}$ and $\Sigma_{14}^{(n)}$. Details are left to the reader. Turning to $\Sigma_{2}^{(n)}$, it follows from assumption A3 that

$$\begin{aligned}&\Big|\sum_{t=2}^{n}(Y_t^{(n)})^2\dot{\phi}(Z_t-\delta_t^{(n)}Y_t^{(n)}) - \sum_{t=2}^{n}(Y_t^{(n)})^2\dot{\phi}(Z_t)\Big| \leq A_f\sum_{t=2}^{n}|Y_t^{(n)}|^3\\ &\leq 16\sum_{t=2}^{n}(n^{-\frac{1}{2}}(|\alpha^{(n)}| + \sigma^{-1}|\beta^{(n)}X_{t-1}|)^3\\ &\quad + 16\sum_{t=2}^{n}\Big\{\sum_{j=2}^{t-1}(n^{-\frac{1}{2}}\sigma^{-1}|\beta^{(n)}|)^{j-i}(|a| + n^{-\frac{1}{2}}(|\alpha^{(n)}| + |\beta^{(n)}X_{t-1}|)\prod_{k=1}^{j}|X_{t-k}|\Big\}^3\\ &\quad + 16\sum_{t=2}^{n}\Big\{(n^{-\frac{1}{2}}\sigma^{-1}|\beta^{(n)}|)^{t-1}\prod_{k=1}^{t-1}|X_{t-k}|((|a|+n^{-\frac{1}{2}}|\alpha^{(n)}|)|X_0|+n^{-\frac{1}{2}}\sigma^{-1}|\beta^{(n)}||X_0\epsilon_0|)\Big\}^3\\ &= 16\left(\Sigma_{21}^{(n)} + \Sigma_{22}^{(n)} + \Sigma_{23}^{(n)}\right).\end{aligned}$$

Clearly, $\Sigma_{21}^{(n)}$ is $o_P(1)$; $\Sigma_{22}^{(n)}$ and $\Sigma_{23}^{(n)}$ can be treated along the same lines as $\Sigma_{13}^{(n)}$, and also are $o_P(1)$'s. Up to $o_P(1)$ terms, $\Sigma_2^{(n)}$ thus reduces to

$$
\begin{aligned}
&\sum_{t=2}^{n}(Y_t^{(n)})^2\dot{\phi}_f(Z_t)\\
&=\sum_{t=2}^{n}\dot{\phi}_f(Z_t)\{n^{-\frac{1}{2}}(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-1})X_{t-1}\\
&\quad+\sum_{j=2}^{t-1}(-n^{-1/2}\sigma^{-1}\beta^{(n)})(a+n^{-1/2}(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-j}))\prod_{k=1}^{j}X_{t-k}\}^2\\
&\quad+\sum_{t=2}^{n}\dot{\phi}_f(Z_t)\Big\{(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{t-1}((a+n^{-\frac{1}{2}}\alpha^{(n)})X_0+n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)}X_0\epsilon_0\prod_{k=1}^{j}X_{t-k}\Big\}^2\\
&\quad+2\sum_{t=2}^{n}\dot{\phi}_f(Z_t)\Big\{n^{-\frac{1}{2}}(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-1})X_{t-1}\\
&\quad+\sum_{j=2}^{t-1}(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j-1}(a+n^{-\frac{1}{2}}(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-j})\prod_{k=1}^{j}X_{t-k})\\
&\quad\times(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{t-1}\prod_{k=1}^{t-1}X_{t-k}((a+n^{-\frac{1}{2}}\alpha^{(n)})X_0+n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)}X_0\epsilon_0)\Big\}\\
&=\Sigma_{21}^{(n)}+\Sigma_{22}^{(n)}+\Sigma_{23}^{(n)}.
\end{aligned}
$$

The first term $\Sigma_{21}^{(n)}$ decomposes into

$$
\begin{aligned}
\Sigma_{21}^{(n)}&=n^{-1}\sum_{t=2}^{n}\dot{\phi}_f(Z_t)\{\alpha^{(n)}X_{t-1}+\sigma^{-1}\beta^{(n)}X_{t-1}^2-a\sigma^{-1}\beta^{(n)}X_{t-1}X_{t-2}\}^2\\
&\quad+\sum_{t=2}^{n}\dot{\phi}_f(Z_t)\{-n^{-1}\sigma^{-1}\beta^{(n)}(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-2})X_{t-2}X_{t-1}\\
&\quad+\sum_{j=3}^{t-1}(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j-1}(a+n^{-\frac{1}{2}}\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-j}))\prod_{k=1}^{j}X_{t-k}\}^2\\
&\quad+2n^{-\frac{1}{2}}\sum_{t=2}^{n}\dot{\phi}_f(Z_t)\Big\{[(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-1})X_{t-1}-a\sigma^{-1}\beta^{(n)}X_{t-1}X_{t-2}]\\
&\quad\times[-n^{-1}\sigma^{-1}\beta^{(n)}(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-2})X_{t-1}X_{t-2}\\
&\quad+\sum_{j=3}^{t-1}(-n^{-\frac{1}{2}}\sigma^{-1}\beta^{(n)})^{j-1}(a+n^{-\frac{1}{2}}(\alpha^{(n)}+\sigma^{-1}\beta^{(n)}X_{t-j}))\prod_{k=1}^{j}X_{t-k}]\Big\}\\
&=\Sigma_{211}^{(n)}+\Sigma_{212}^{(n)}+2\Sigma_{213}^{(n)},
\end{aligned}
$$

with

$$\Sigma_{211}^{(n)} = n^{-1} \sum_{t=k+1}^{n} \dot{\phi}_f(Z_t) \Big\{ \alpha^{(n)} \sum_{i=0}^{k-1} a^i Z_{t-i-1} + \sigma^{-1}\beta^{(n)} (\sum_{i=0}^{k-1} a^i Z_{t-i-1})^2$$
$$- a\sigma^{-1}\beta^{(n)} \sum_{i=0}^{k-1} a^i Z_{t-i-1} \sum_{i=0}^{k-2} a^i Z_{t-i-2} \Big\}^2 + R_k^{(n)} + n^{-1} L_k^{(n)}$$
$$= n^{-1} \sum_{t=k+1}^{n} \zeta_t^{(k;n)} + R_k^{(n)} + n^{-1} L_k^{(n)}.$$

It is easily seen that $R_k^{(n)}$ converges to zero (in L_1 norm), as $k \to \infty$, uniformly in n, that $n^{-1} L_k^{(n)}$ (k fixed) converges to zero (in L_1 norm) as $n \to \infty$, and that, the $\zeta_t^{(k;n)}$'s being k-dependent (fixed k), $n^{-1} \sum_{t=k+1}^{n} \zeta_t^{(k;n)} - E(\zeta_t^{(k;n)}) = o_P(1)$ as $n \to \infty$. Now,

$$E(\zeta_t^{(k;n)}) = I(f_1) \Big\{ (\alpha^{(n)})^2 (1 + \frac{a^2}{1-a^2}(1 - a^{2k-3}))$$
$$+ 2\alpha^{(n)}\beta^{(n)} (\frac{m_3}{\sigma^3})_f + (\beta^{(n)})^2 ((\frac{m_4}{\sigma^4})_f + \frac{a^2}{1-a^2}(1 - a^{2k-3})) \Big\},$$

which in turn is asymptotically equivalent, as $k \to \infty$, to

$$I(f_1) \left\{ (\alpha^{(n)})^2 (1-a^2)^{-1} + 2\alpha^{(n)}\beta^{(n)} (\frac{m_3}{\sigma^3})_f + (\beta^{(n)})^2 ((\frac{m_4}{\sigma^4})_f + \frac{a^2}{1-a^2}) \right\}$$
$$= (\boldsymbol{\tau}^{(n)})' \mathbf{W}_f^2 \boldsymbol{\tau}^{(n)}.$$

It follows that $\Sigma_{211}^{(n)} = (\boldsymbol{\tau}^{(n)})' \mathbf{W}_f^2 \boldsymbol{\tau}^{(n)} + o_P(1)$. Consider now $\Sigma_{212}^{(n)}$:

$$|\Sigma_{212}^{(n)}| \leq 2\sigma^{-2} (\beta^{(n)})^2 n^{-2} \sum_{t=2}^{n} |\dot{\phi}_f(Z_t)| |(\alpha^{(n)} + \sigma^{-1}\beta^{(n)} X_{t-2}) X_{t-1} X_{t-2}|^2$$
$$+ 2 \sum_{t=2}^{n} |\dot{\phi}_f(Z_t)| \Big| \sum_{j=3}^{t-1} (-n^{-1/2}\sigma^{-1}\beta^{(n)})^{j-1}$$
$$\times (a + n^{-1/2}(\alpha^{(n)} + \sigma^{-1}\beta^{(n)} X_{t-j})) \prod_{k=1}^{j} X_{t-k} \Big|^2.$$

The two terms in this upper bound can be treated in the same way as $\Sigma_{12}^{(n)}$ and $\Sigma_{13}^{(n)}$, respectively. The case of $\Sigma_{213}^{(n)}$ is entirely similar to that of $\Sigma_{212}^{(n)}$. As for $\Sigma_{22}^{(n)}$ and $\Sigma_{23}^{(n)}$, they can be handled along the same lines as $\Sigma_{21}^{(n)}$; details are left to the reader.

This completes the proof of Part (i) of Proposition 3.2. Part (ii) readily follows from Proposition 3.1 and the definition of the central sequence $\boldsymbol{\Delta}_f^{(n)}$ in terms of autocorrelation and cubic autocorrelation coefficients.

Appendix B: Proof of Proposition 4.2

We concentrate on the proof of (4.6); the derivation of (4.5) is very similar to that of Part (iii) of Proposition 5.1 in Hallin and Puri (1994), of which (4.4) is a particular case.

Under $H_g^{(n)}(a)$, the process $Z_t^{(n)} = Z_t(a+n^{-1/2}\alpha^{(n)}) = X_t - (a+n^{-1/2}\alpha^{(n)})X_{t-1}$, $t \in \mathbb{Z}$, is ARMA (1,1), so that Lemmas 5.2, 5.3 and 5.4 of Hallin and Puri (1994) hold. For simplicity, put $Z_t^0 = Z_t(a) = X_t - aX_{t-1}$, and write $F_1 \circ F_2(\cdot)$ instead of $F_1(F_2(\cdot))$. Also, define

$$F_{[m]}^{-1}(u) = \begin{cases} F_1^{-1}(u_m^-), & \text{if } u < u_m^-, \\ F_1^{-1}(u), & \text{if } u_m^- \leq u < 1 - u_m^+, \\ F_1^{-1}(1-u_m^+), & \text{if } 1 - u_m^+ \leq u, \end{cases} \qquad u \in (0,1),$$

where $u_m^- + u_m^+ = 2/(n+1)$ are such that

$$\int_0^{u_m^-} F_1^{-1}(u)du + \int_{1-u_m^+}^1 F_1^{-1}(u)du = 0,$$

which implies $\int_0^1 F_{[m]}^{-1}(u)du = \int_0^1 F_1^{-1}(u)du = 0$. Then (all limits and expectations are taken under $H_g^{(n)}(a)$)

$$\begin{aligned} & n^{\frac{1}{2}}(I_g(f))^{\frac{1}{2}}(q_{i;f}^{(n)}(a+n^{-1/2}\alpha^{(n)}) - q_{i;f}^{(n)}(a)) \\ &= n^{-\frac{1}{2}} \sum_{t=i+1}^{n} \{[\phi_{f_1} \circ F_{[m]}^{-1} \circ F(Z_t^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-1}^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-i}^{(n)})] \\ &\qquad - [\phi_{f_1} \circ F_{[m]}^{-1} \circ F(Z_t^0)][F_{[m]}^{-1} \circ F(Z_{t-1}^0)][F_{[m]}^{-1} \circ F(Z_{t-i}^0)]\} \\ &\quad - E\left\{n^{-\frac{1}{2}} \sum_{t=i+1}^{n} [\phi_{f_1} \circ F_{[m]}^{-1} \circ F(Z_t^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-1}^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-i}^{(n)})]\right\} \\ &\quad + R_{[m]}^{(n)} + o_P(1) \\ &= \Sigma_{[m]}^{(n)} + R_{[m]}^{(n)} + o_P(1), \end{aligned} \tag{B.1}$$

where

$$\begin{aligned} R_{[m]}^{(n)} &= n^{-\frac{1}{2}} \sum_{t=i+1}^{n} \Big\{[\phi_{f_1} \circ F_1^{-1} \circ F(Z_t^{(n)})][F_1^{-1} \circ F(Z_{t-1}^{(n)})][F_1^{-1} \circ F(Z_{t-i}^{(n)})] \\ &\qquad - [\phi_{f_1} \circ F_1^{-1} \circ F(Z_t^0)][F_1^{-1} \circ F(Z_{t-1}^0)][F_1^{-1} \circ F(Z_{t-i}^0)]\Big\} \\ &\quad - n^{-\frac{1}{2}} \sum_{t=i+1}^{n} \Big\{[\phi_{f_1} \circ F_{[m]}^{-1} \circ F(Z_t^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-1}^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-i}^{(n)})] \\ &\qquad - [\phi_{f_1} \circ F_{[m]}^{-1} \circ F(Z_t^0)][F_{[m]}^{-1} \circ F(Z_{t-1}^0)][F_{[m]}^{-1} \circ F(Z_{t-i}^0)]\Big\} \\ &\quad + E\Big\{n^{-\frac{1}{2}} \sum_{t=i+1}^{n} [\phi_{f_1} \circ F_{[m]}^{-1} \circ F(Z_t^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-1}^{(n)})][F_{[m]}^{-1} \circ F(Z_{t-i}^{(n)})]\Big\} \\ &= R_{[m;1]}^{(n)} - R_{[m;2]}^{(n)} + E_{[m]}^{(n)}. \end{aligned}$$

Hence,

$$\begin{aligned}\Sigma^{(n)}_{[m]} + E^{(n)}_{[m]} &= n^{-\frac{1}{2}} \sum_{t=i+1}^{n} [\phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^0_t)] \Big\{ [F_{[m]} \circ F(Z^{(n)}_{t-1})][F_{[m]} \circ F(Z^{(n)}_{t-i})] \\ &\quad - [F^{-1}_{[m]} \circ F(Z^0_{t-1})][F_{[m]} \circ F(Z^0_{t-i})] \Big\} \\ &\quad + n^{-\frac{1}{2}} \sum_{t=i+1}^{n} \Big[\phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^{(n)}_t) - \phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^0_t) \Big] \\ &\quad \times [F^{-1}_{[m]} \circ F(Z^{(n)}_{t-1})][F^{-1}_{[m]} \circ F(Z^{(n)}_{t-i})] - E^{(n)}_{[m]} + E^{(n)}_{[m]} \\ &= D^{(n)}_{[m;1]} + D^{(n)}_{[m;2]},\end{aligned}$$

so that (B.1) reduces to $D^{(n)}_{[m;1]} + D^{(n)}_{[m;2]} + R^{(n)}_{[m;1]} - R^{(n)}_{[m;2]} + o_P(1)$. Lemma 5.4 in Hallin and Puri (1994) and the fact that

$$\max_{i+1 \le t \le n} [F^{-1}_{[m]} \circ F(Z^{(n)}_t)][F^{-1}_{[m]} \circ F(Z^{(n)}_{t-i+1})] - [F^{-1}_{[m]} \circ F(Z^0_t)][F^{-1}_{[m]} \circ F(Z^0_{t-i+1})]$$

is $o_P(1)$ as $n \to \infty$, and remains uniformly bounded by $(F^{-1}(1 - u^+_m))^2 - (F^{-1}(u^-_m))^2$, imply that $D^{(n)}_{[m;1]}$ is $o_P(1)$ (as $n \to \infty$, for fixed m). Turning to $D^{(n)}_{[m;2]}$, let

$$\begin{aligned}\bar{D}^{(n)}_{[m;2]} =& n^{-\frac{1}{2}} \sum_{t=i+1}^{n} \Big\{ \phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^{(n)}_t) - \phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^0_t) \Big\} \\ &\times [F^{-1}_{[m]} \circ F(Z^0_{t-1})][F^{-1}_{[m]} \circ F(Z^0_{t-i})].\end{aligned}$$

Clearly, $D^{(n)}_{[m;2]}$ is $o_P(1)$ under $H^{(n)}_g(a)$ iff $\bar{D}^{(n)}_{[m;2]}$ is $o_P(1)$ under $H^{(n)}_g(a + n^{-1/2}\alpha^{(n)})$ (still, as $n \to \infty$). Examining the variance of $\bar{D}^{(n)}_{[m;2]}$, we obtain

$$\begin{aligned}&\mathrm{Cov}(k) \\ &= \mathrm{Cov}\Big(\Big\{ [\phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^{(n)}_t)] - [\phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^0_t)] \Big\} [F^{-1}_{[m]} \circ F(Z^0_{t-1})][F^{-1}_{[m]} \circ F(Z^0_{t-i})], \\ &\quad \Big\{ [\phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^{(n)}_{t-k})] - [\phi_{f_1} \circ F^{-1}_{[m]} \circ F(Z^0_{t-k}] \Big\} [F^{-1}_{[m]} \circ F(Z^0_{t-1})][F^{-1}_{[m]} \circ F(Z^0_{t-i})] \Big).\end{aligned}$$

It then follows from Lemma 5.3 in Hallin and Puri (1994) that

$$\begin{aligned}\mathrm{Var}(\bar{D}^{(n)}_{[m;2]}) \le\ & 4A^2_f (F^{-1}(1 - u^+_m))^4 E\{[F^{-1}_{[m]} \circ F(Z^{(n)}_t)] - [F^{-1}_{[m]} \circ F(Z^0_t)]\}^2 \\ & + 4A^2_f (F^{-1}(1 + u^+_m))^4 E\{[F^{-1}_{[m]} \circ F(Z^{(n)}_t)] - [F^{-1}_{[m]} \circ F(Z^0_t)]\}^2 \sum_{k=1}^{\infty} (\beta(k))^{1/2},\end{aligned}$$

where A_f denotes the Lipschitz constant in Assumption A3, and $\beta(k)$ is the mixing rate associated with $Z^{(n)}_t$ as an absolutely regular process. It follows from Lemma 5.3 in Hallin and Puri (1994) that $\beta(k)$ is decreasing exponentially in k, which ensures the convergence of $\Sigma_k \beta(k)$. Again, the continuity of $F^{-1}_{[m]} \circ F$ and the fact that $\max_{1 \le t \le n} |Z^{(n)}_t - Z^0_t|$ is $o_P(1)$ imply that $\mathrm{Var}(D^{(n)}_{[m;2]})$ converges to zero as $n \to \infty$.

Finally, in view of Lemmas 5.6 and 5.7 of Hallin and Puri (1994), $R_{[m;1]}^{(n)}$ and $R_{[m;2]}^{(n)}$ also are $o_P(1)$ as $m \to \infty$, uniformly in n. Application of Theorem 7.7.1 of Anderson (1971) completes the proof.

References

Akritas, M. G. and Johnson, R. A. (1982). Efficiencies of tests and estimators for p-ORDER autoregressive processes when the error distribution is nonnormal. Ann. Inst. Statist. Math. **34**, 579-589.

An, H. Z. and Cheng, B. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. Internat. Statist. Rev. **59**, 287-307.

Anderson, T. W. (1971). The Statistical Analysis of Time Series. John Wiley, New York.

Benghabrit, Y. and Hallin, M. (1992). Optimal rank-based tests against first-order superdiagonal bilinear dependence. J. Statist. Plann. Inference **32**, 45-61.

Benghabrit, Y. and Hallin, M. (1993). Rank-based tests against nonlinearity in time series. J. Nonparametr. Statist., to appear.

Brockwell, P. J. and Davis, R. A. (1987). Time Series: Theory and Methods. Springer-Verlag, New York.

Chan, W. S. and Tong, H. (1986). On tests for non-linearity in time series analysis. J. Forecast. **5**, 217-228.

Chan, N. H. and Tran, L. T. (1992). Nonparametric tests for serial dependence. J. Time Ser. Anal. **13**, 19-28.

Gabr, M. M. (1988). On the third-order moment structure and bispectral analysis of some bilinear time series. J. Time Ser. Anal. **9**, 11-20.

Garel, B. (1989). The asymptotic distribution of the likelihood ratio for MA. processes with a regression trend. Statist. Decisions **7**, 167-184.

Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. Ann. Inst. Statist. Math. **47**, 551-579.

Granger, C. W. J. and Andersen, A. P. (1978). An Introduction to Bilinear Time Series Models. Vandenhoeck and Ruprecht, Göttingen.

Guégan, D. (1990). Detecting non-linearity: A review. Statistique et Analyse des Données **15**, 1-17.

Guégan, D. and Pham, T. D. (1992). Power of the score test against bilinear time series models. Statist. Sinica **2**, 157-169.

Hallin, M., Ingenbleek, J.-F. and Puri, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. Ann. Statist. **13**, 1156-1181.

Hallin, M. and Puri, M. L. (1992). Rank tests for time series analysis: A survey. In New Directions in Time Series Analysis, (Edited by D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu), 112-153, Springer-Verlag, New York.

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

Hinich, M. J. (1982). Testing for Gaussianity and linearity of a stationary time series. J. Time Ser. Anal. **3**, 169-176.

Hinich, M. J. and Patterson, D. M. (1992). A new diagnostic test of model inadequacy which uses the martingale difference criterion. J. Time Ser. Anal. **13**, 233-252.

Keenan, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity. Biometrika **72**, 39-44.

Kreiss, J.-P. (1987). On adaptive estimation in stationary ARMA processes. Ann. Statist. **15**, 112-133.

Kreiss, J.-P. (1990a). Testing linear hypotheses in autoregressions. Ann. Statist. **18**, 1470-1482.

Kreiss, J.-P. (1990b). Local asymptotic normality for autoregression with infinite order. J. Statist. Plann. Inference **26**, 185-219.

LeCam, L. (1960). Locally asymptotically normal families of distributions. Univ. Calif. Publ. in Statist. **3**, 37-98.

LeCam, L. (1986). Asymptotic Methods in Statistical Decision Theory. Springer-Verlag, New York.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations. J. Time Ser. Anal. **4**, 269-273.

Petruccelli, J. D. and Davies, N. (1986). A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. Biometrika **73**, 687-694.

Pham, T. D. and Tran, L. T. (1981). On the first-order bilinear time series model. J. Appl. Probab. **18**, 617-627.

Priestley, M. B. (1988). Non-linear and Non stationary Time Series Analysis. Academic Press, London.

Saikkonen, P. and Luukkonen, R. (1988). Lagrange multiplier tests for testing non-linearities in time series models. Scand. J. Statist. **15**, 55-68.

Saikkonen, P. and Luukkonen, R. (1991). Power properties of a time series linearity test against some simple bilinear alternatives. Statist. Sinica **1**, 453-464.

Skaug, H. J. and Tjøstheim, D. (1993a). Nonparametric tests of serial independence. In Developments in Time Series Analysis (Edited by T. Subba Rao), 207-229, Volume in Honour of M.B. Priestley, Chapman and Hall, London.

Skaug, H. J. and Tjøstheim, D. (1993b). A nonparametric test of serial independence based on the empirical distribution function. Biometrika **80**, 591-602.

Subba Rao, T. and Da Silva, M. E. (1992). Identification of bilinear time series models BL(p, 0, p, 1). Statist. Sinica **2**, 465-478.

Subba Rao, T. and Gabr, M. M. (1980). A test for linearity of stationary time series. J. Time Ser. Anal. **1**, 145-158.

Subba Rao, T. and Gabr, M. M. (1984). An introduction to Bispectral Analysis and Bilinear Time Series Models. Springer-Verlag, 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.

Tong, H. (1983). Threshold Models in Non-linear Time Series Analysis. Springer-Verlag, Berlin.

Tong, H. (1990). Non-linear Time Series: A dynamical system approach. Clarendon Press, Oxford.

Ecole Mohammadia d'Ingénieurs, Avenue Ibn Sina, B.P. 765, Agdal, Rabat, Morocco.

Département de Mathématique and Institut de Statistique, C.P. 210, Université Libre de Bruxelles, B 1050 Bruxelles, Belgium.

(Received January 1994; accepted December 1994)