Electronic Journal of Statistics
Vol. 8 (2014) 2701–2740
ISSN: 1935-7524
DOI: 10.1214/14-EJS968

Estimation of the variance of the quasi-maximum likelihood estimator of weak VARMA models

Yacouba Boubacar Maïnassara*

Laboratoire de Mathématiques de Besançon – UMR 6623 CNRS-UFC, 16 route de Gray, 25030 Besançon Cedex, France

Abstract: This paper considers the problems of computing and estimating the asymptotic variance matrix of the least squares (LS) and/or the quasi-maximum likelihood (QML) estimators of vector autoregressive moving-average (VARMA) models under the assumption that the errors are uncorrelated but not necessarily independent. We firstly give expressions for the derivatives of the VARMA residuals in terms of the parameters of the models. Secondly we give an explicit expression of the asymptotic variance matrix of the QML/LS estimator, in terms of the VAR and MA polynomials, and of the second and fourth-order structure of the noise. We then deduce a consistent estimator of this asymptotic variance matrix. Modified versions of the Wald, Lagrange Multiplier and Likelihood Ratio tests are proposed for testing linear restrictions on the parameters. The theoretical results are illustrated by means Monte Carlo experiments.

MSC 2010 subject classifications: 62H12, 62H15, 62M10.
Keywords and phrases: Covariance matrix estimate, Lagrange multiplier test, likelihood ratio test, QMLE/LSE, residuals derivatives, Wald test, weak VARMA models.

Received February 2014.

Contents

*The work was, partially, supported by a BQR (Bonus Qualité Recherche) of the Université de Franche-Comté.

1. Introduction

The class of vector autoregressive moving-average (VARMA) models and the sub-class of vector autoregressive (VAR) models are used in time series analysis and econometrics to describe not only the properties of the individual time series but also the possible cross-relationships between the time series (see [36, 41]). This paper is devoted to the problems of computing and estimating the asymptotic variance matrix of the least squares (LS) and/or the quasi-maximum likelihood (QML) estimators of VARMA models under the assumption that the errors are uncorrelated but not necessarily independent. These models are called weak VARMA in contrast to the standard VARMA models, also called strong VARMA models, in which the error terms are supposed to be independent and identically distributed (iid). This independence assumption is often considered too restrictive by practitioners. It precludes conditional heteroscedasticity and/or other forms of nonlinearity (see [25] for a review on weak univariate ARMA models).

A process $(X_t)_{t\in\mathbb{Z}}$ is said to be *nonlinear* when the innovation process in the Wold decomposition (see *e.g.* [12], for the univariate case, and Reinsel [41] in the multivariate framework) is uncorrelated but not necessarily independent, and is said to be *linear* in the opposite case (*i.e.* when the innovation process in the Wold decomposition is iid). Relaxing the independence assumption considerably extends the range of applications of the VARMA models, and allows to cover linear representations of general nonlinear processes. Indeed such nonlinearities may arise for instance when the error process follows an autoregressive conditional heteroscedasticity (ARCH) introduced by Engle [18] and extended to the generalized ARCH (GARCH) by [5], all-pass (see [3]) or other models displaying a second order dependence (see [1]). Other situations where the errors are dependent can be found in [25], see also [42]. Leading examples of multivariate linear processes are the VARMA and VAR models with iid error terms. Nonlinear models are becoming more and more employed because numerous real time series exhibit nonlinear dynamics, for instance conditional heteroscedasticity, which can not be generated by autoregressive moving-average (ARMA) models with iid error terms.[1]

[1] To cite few examples of nonlinear processes, let us mention the self-exciting threshold autoregressive (SETAR), the smooth transition autoregressive (STAR), the exponential autoregressive (EXPAR), the bilinear, the random coefficient autoregressive (RCA), the functional autoregressive (FAR) (see [19, 44] for references on these nonlinear time series models). All these nonlinear models have been initially proposed for univariate time series, but have multivariate extensions.

Work on asymptotic results usually focuses on univariate models (see [25] for a review of this topic). For multivariate models, important advances have been obtained by Dufour and Pelletier [16] who study the asymptotic properties of a generalization of the regression-based estimation method proposed by Hannan and Rissanen [29] under weak assumptions on the innovation process, Francq and Raïssi [20] who study portmanteau tests for weak VAR models, Boubacar Maïnassara and Francq [9] study the estimation of weak VARMA models, Boubacar Maïnassara [6, 7] who studies portmanteau tests and the problem of order selection of weak VARMA models, Katayama [33] who also proposes a new portmanteau test statistic for weak VARMA models. In [9], it is shown that the asymptotic variance matrix of the usual estimators has the "sandwich" form $\Omega := J^{-1}IJ^{-1}$, where the two Fisher information matrices J and I depend respectively on second and fourth-order moments of the errors and on the true parameter (denoted, hereafter, θ_0). This proposed asymptotic variance reduces to standard form $\Omega = 2J^{-1}$ in the linear case.

In the framework of (Gaussian) linear processes, the problem of computing the Fisher information matrices and their inverses has been widely studied. Various expressions of these matrices have been given by Whittle [46, 47], Siddiqui [43], Durbin [17] and Box and Jenkins [11]. McLeod [39], Klein and Mélard [34, 35] and Godolphin and Bane [27] have given algorithms for their computation. For few particular cases of weak ARMA models, the matrices I and J have been computed by Boubacar Maïnassara, Carbon and Francq [8], Francq and Zakoian [24, 26] and Francq, Roy and Zakoian [22]. The main goal of the present paper is to complete the available results concerning the statistical analysis of weak VARMA models, by proposing another estimator of Ω, which allows to separate the effects due to the VARMA parameters from those due to the nonlinear structure of the noise.

The paper is organized as follows. Section 2 presents the models that we consider here, and presents the results on the QML/LS estimator asymptotic distribution obtained by Boubacar Maïnassara and Francq [9]. In Section 3 we give expressions for the derivatives of the VARMA residuals in terms of parameters of the models. Section 4 is devoted to find an explicit expression of the asymptotic variance of the QML/LS estimator, in terms of the VAR and MA polynomials, and of the second and fourth-order structure of the noise. In Section 5 we deduce a consistent estimator of this asymptotic variance. We describe, in Section 6, how to obtain numerical evaluations of tolerance for the information matrices J and I up to some tolerance. In Section 7 it is shown how the standard Wald, LM (Lagrange Multiplier) and LR (Likelihood Ratio) tests must be adapted in the weak VARMA case in order to test for general linearity constraints. This section is also of interest in the univariate framework because, to our knowledge, these tests have not been studied for weak ARMA models. Numerical experiments are presented in Section 8. The proofs of the main results are collected in the appendix. We denote by $A \otimes B$ the Kronecker product of two matrices A and B, and by $\mathrm{vec}(A)$ the vector obtained by stacking the columns of A. The reader is refereed to Magnus and Neudecker [37] for the properties of these operators. Let 0_r be the null vector of $\mathbb{R}^r$, and let I_r be the $r \times r$ identity matrix.

2. Model and assumptions

Consider a d-dimensional stationary process (X_t) satisfying a structural VARMA(p,q) representation of the form

$$A_{00}X_t - \sum_{i=1}^{p} A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{i=1}^{q} B_{0i}\epsilon_{t-i}, \quad \forall t \in \mathbb{Z} = \{0, \pm 1, \dots\}, \tag{2.1}$$

where ϵ_t is a white noise, namely a stationary sequence of centered and uncorrelated random variables with a non singular variance Σ_0. The structural forms are mainly used in econometrics to introduce instantaneous relationships between economic variables. Of course, constraints are necessary for the identifiability of these representations. Let $[A_{00} \dots A_{0p} B_{00} \dots B_{0q}]$ be the $d \times (p+q+2)d$ matrix of all the coefficients, without any constraint. The matrix Σ_0 is considered as a nuisance parameter. The parameter of interest, θ_0, belongs to the parameter space $\Theta \subset \mathbb{R}^{k_0}$, where k_0 is the number of unknown parameters, which is typically much smaller that $(p+q+2)d^2$. The matrices $A_{00}, \dots A_{0p}, B_{00}, \dots B_{0q}$ involved in (2.1) are specified by θ_0. More precisely, we write $A_{0i} = A_i(\theta_0)$ and $B_{0j} = B_j(\theta_0)$ for $i = 0, \dots, p$ and $j = 0, \dots, q$, and $\Sigma_0 = \Sigma(\theta_0)$. We need the following assumptions used by Boubacar Maïnassara and Francq [9] to ensure the consistency and the asymptotic normality of the quasi-maximum likelihood estimator (QMLE).

A1: The applications $\theta \mapsto A_i(\theta)$ $i = 0, \dots, p$, $\theta \mapsto B_j(\theta)$ $j = 0, \dots, q$ and $\theta \mapsto \Sigma(\theta)$ admit continuous third order derivatives for all $\theta \in \Theta$.

For simplicity we now write A_i, B_j and Σ instead of $A_i(\theta)$, $B_j(\theta)$ and $\Sigma(\theta)$. Let $A_\theta(z) = A_0 - \sum_{i=1}^{p} A_i z^i$ and $B_\theta(z) = B_0 - \sum_{i=1}^{q} B_i z^i$.

A2: For all $\theta \in \Theta$, we have $\det A_\theta(z) \det B_\theta(z) \neq 0$ for all $|z| \leq 1$.

A3: We have $\theta_0 \in \Theta$, where Θ is compact.

A4: The process (ϵ_t) is stationary and ergodic.

A5: For all $\theta \in \Theta$ such that $\theta \neq \theta_0$, either the transfer functions

$$A_0^{-1} B_0 B_\theta^{-1}(z) A_\theta(z) \neq A_{00}^{-1} B_{00} B_{\theta_0}^{-1}(z) A_{\theta_0}(z)$$

for some $z \in \mathbb{C}$, or

$$A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}.$$

A6: We have $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

We now introduce, as in [23] the strong mixing coefficients of a stationary process $Z = (Z_t)$ denoted by

$$\alpha_Z(h) = \sup_{A \in \sigma(Z_u, u \leq t), B \in \sigma(X_u, u \geq t+h)} |P(A \cap B) - P(A)P(B)|,$$

measuring the temporal dependence of the process Z. Denoting by $\|Z\|$ the Euclidean norm of Z.

A7: We have $\mathbb{E}\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$.

The reader is referred to [9] for a discussion of these assumptions. Note that (ϵ_t) can be replaced by (X_t) in **A4**, because $X_t = A_{\theta_0}^{-1}(L)B_{\theta_0}(L)\epsilon_t$ and $\epsilon_t = B_{\theta_0}^{-1}(L)A_{\theta_0}(L)X_t$, where L stands for the backward operator. From **A1** the matrices A_0 and B_0 are invertible. Introducing the innovation process $e_t = A_{00}^{-1}B_{00}\epsilon_t$, the structural representation $A_{\theta_0}(L)X_t = B_{\theta_0}(L)\epsilon_t$ can be rewritten as the reduced VARMA representation

$$X_t - \sum_{i=1}^{p} A_{00}^{-1}A_{0i}X_{t-i} = e_t - \sum_{i=1}^{q} A_{00}^{-1}B_{0i}B_{00}^{-1}A_{00}e_{t-i}.$$

We thus recursively define $\tilde{e}_t(\theta)$ for $t = 1, \dots, n$ by

$$\tilde{e}_t(\theta) = X_t - \sum_{i=1}^{p} A_0^{-1}A_iX_{t-i} + \sum_{i=1}^{q} A_0^{-1}B_iB_0^{-1}A_0\tilde{e}_{t-i}(\theta),$$

with initial values $\tilde{e}_0(\theta) = \cdots = \tilde{e}_{1-q}(\theta) = X_0 = \cdots = X_{1-p} = 0$. The gaussian quasi-likelihood is given by

$$\tilde{\mathrm{L}}_n(\theta, \Sigma_e) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma_e}} \exp\left\{-\frac{1}{2}\tilde{e}_t'(\theta)\Sigma_e^{-1}\tilde{e}_t(\theta)\right\},\ \Sigma_e = A_0^{-1}B_0\Sigma B_0'A_0^{-1'}.$$

A QMLE of θ and Σ_e are a measurable solution $(\hat{\theta}_n, \hat{\Sigma}_e)$ of

$$(\hat{\theta}_n, \hat{\Sigma}_e) = \arg\max_{\theta \in \Theta, \Sigma_e} \tilde{\mathrm{L}}_n(\theta, \Sigma_e).$$

We use the matrix M_{θ_0} of the coefficients of the reduced form (8.1), to that made by Boubacar Maïnassara and Francq [9], where

$$M_{\theta_0} = [A_{00}^{-1}A_{01} : \cdots : A_{00}^{-1}A_{0p} : A_{00}^{-1}B_{01}B_{00}^{-1}A_{00} : \cdots : A_{00}^{-1}B_{0q}B_{00}^{-1}A_{00}].$$

Now, we need a local identifiability assumption which completes the global identifiability assumption **A5** (but none is implied by the other) and specifies how this matrix depends on the parameter θ_0. Let $\dot{M}_{\theta_0}$ be the matrix $\partial \mathrm{vec}(M_\theta)/\partial\theta'$ evaluated at θ_0.

A8: The matrix $\dot{M}_{\theta_0}$ is of full rank k_0.

Under Assumptions **A1**–**A8**, Boubacar Maïnassara and Francq [9] have showed the consistency ($\hat{\theta}_n \to \theta_0$ *a.s.* as $n \to \infty$) and the asymptotic normality of the QMLE:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega := J^{-1}IJ^{-1}), \tag{2.2}$$

where $J = J(\theta_0, \Sigma_{e0})$ and $I = I(\theta_0, \Sigma_{e0})$, with

$$J(\theta, \Sigma_e) = \lim_{n\to\infty} \frac{2}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log\tilde{\mathrm{L}}_n(\theta, \Sigma_e) \quad a.s.$$

and

$$I(\theta, \Sigma_e) = \lim_{n\to\infty} \mathrm{Var}\left\{\frac{2}{\sqrt{n}}\frac{\partial}{\partial\theta}\log\tilde{\mathrm{L}}_n(\theta, \Sigma_e)\right\}. \tag{2.3}$$

3. Expression for the derivatives of the VARMA residuals

For a univariate ARMA model, McLeod [38] has defined the noise derivatives by writing

$$\frac{\partial e_t}{\partial \phi_i} = \upsilon_{t-i},\ i=1,\dots,p \quad \text{and} \quad \frac{\partial e_t}{\partial \varphi_j} = u_{t-j},\ j=1,\dots,q,$$

where ϕ_i and φ_j are respectively the univariate AR and MA parameters. Let $\phi_\theta(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\varphi_\theta(L) = 1 - \sum_{i=1}^{q} \varphi_i L^i$. We denote by ϕ_h^* and φ_h^* the coefficients defined by

$$\phi_\theta^{-1}(z) = \sum_{h=0}^{\infty} \phi_h^* z^h, \quad \varphi_\theta^{-1}(z) = \sum_{h=0}^{\infty} \varphi_h^* z^h, \quad |z| \le 1 \quad \text{for} \quad h \ge 0.$$

When p and q are not both equal to 0, let $\theta = (\theta_1, \dots, \theta_p, \theta_{p+1}, \dots, \theta_{p+q})'$. Then, it is easily seen that the univariate noise derivatives can be represented as

$$\frac{\partial e_t(\theta)}{\partial \theta} = (\upsilon_{t-1}(\theta), \dots, \upsilon_{t-p}(\theta), u_{t-1}(\theta), \dots, u_{t-q}(\theta))',$$

where

$$\upsilon_t(\theta) = -\phi_\theta^{-1}(L) e_t(\theta) = -\sum_{h=0}^{\infty} \phi_h^* e_{t-h}(\theta),\ u_t(\theta) = \varphi_\theta^{-1}(L) e_t(\theta) = \sum_{h=0}^{\infty} \varphi_h^* e_{t-h}(\theta)$$

and

$$e_t(\theta) = \varphi_\theta^{-1}(L) \phi_\theta(L) X_t.$$

The subject of this section is to generalize these expansions to the multivariate ARMA case. The reduced VARMA representation can be rewritten as the compact form

$$\mathbf{A}_\theta(L) X_t = \mathbf{B}_\theta(L) e_t(\theta),$$

where $\mathbf{A}_\theta(L) = I_d - \sum_{i=1}^{p} \mathbf{A}_i L^i$ and $\mathbf{B}_\theta(L) = I_d - \sum_{i=1}^{q} \mathbf{B}_i L^i$, with $\mathbf{A}_i = A_0^{-1} A_i$ and $\mathbf{B}_i = A_0^{-1} B_i B_0^{-1} A_0$. For $\ell = 1, \dots, p$ and $\ell' = 1, \dots, q$, let $\mathbf{A}_\ell = (\mathbf{a}_{ij,\ell})$, $\mathbf{B}_{\ell'} = (\mathbf{b}_{ij,\ell'})$, $\mathbf{a}_\ell = \mathrm{vec}[\mathbf{A}_\ell]$ and $\mathbf{b}_{\ell'} = \mathrm{vec}[\mathbf{B}_{\ell'}]$. We denote respectively by

$$\mathbf{a} := (\mathbf{a}_1', \dots, \mathbf{a}_p')' \quad \text{and} \quad \mathbf{b} := (\mathbf{b}_1', \dots, \mathbf{b}_q')',$$

the coefficients of the multivariate AR and MA parts. Thus we can rewrite $\theta = (\mathbf{a}', \mathbf{b}')'$, where $\mathbf{a} \in \mathbb{R}^{k_1}$ depends on $A_0, \dots, A_p$, and where $\mathbf{b} \in \mathbb{R}^{k_2}$ depends on $B_0, \dots, B_q$, with $k_1 + k_2 = k_0$. For $i, j = 1, \dots, d$, let $M_{ij}(L)$ and $N_{ij}(L)$ the $(d \times d)-$matrix operators defined by

$$M_{ij}(L) = \mathbf{B}_\theta^{-1}(L) E_{ij} \mathbf{A}_\theta^{-1}(L) \mathbf{B}_\theta(L) \quad \text{and} \quad N_{ij}(L) = \mathbf{B}_\theta^{-1}(L) E_{ij},$$

where E_{ij} is the $d \times d$ matrix with 1 at position (i, j) and 0 elsewhere. We denote by $\mathbf{A}_{ij,h}^*$ and $\mathbf{B}_{ij,h}^*$ the $(d \times d)$ matrices defined by

$$M_{ij}(z) = \sum_{h=0}^{\infty} \mathbf{A}_{ij,h}^* z^h, \quad N_{ij}(z) = \sum_{h=0}^{\infty} \mathbf{B}_{ij,h}^* z^h, \quad |z| \le 1$$

for $h \geq 0$. Take $\mathbf{A}^*_{ij,h} = \mathbf{B}^*_{ij,h} = 0$ when $h < 0$. Let the $d \times d^3(p+q)$ matrix

$$\lambda_h(\theta) = \left[-\mathbf{A}^*_{h-1} : \cdots : -\mathbf{A}^*_{h-p} : \mathbf{B}^*_{h-1} : \cdots : \mathbf{B}^*_{h-q}\right], \tag{3.1}$$

where

$$\mathbf{A}^*_h = \left[\mathbf{A}^*_{11,h} : \mathbf{A}^*_{21,h} : \cdots : \mathbf{A}^*_{dd,h}\right] \text{ and } \mathbf{B}^*_h = \left[\mathbf{B}^*_{11,h} : \mathbf{B}^*_{21,h} : \cdots : \mathbf{B}^*_{dd,h}\right]$$

are $d \times d^3$ matrices.

The matrix $\lambda_h(\theta)$ is well defined because the coefficients of the series expansions of $\mathbf{A}_\theta^{-1}$ and $\mathbf{B}_\theta^{-1}$ decrease exponentially fast to zero.

We are now able to state the following proposition, which is a generalization of a result given in [38].

Proposition 3.1. *Under the assumptions* **A1–A8***, we have*

$$\frac{\partial e_t(\theta)}{\partial \theta'} = \left[V_{t-1}(\theta) : \cdots : V_{t-p}(\theta) : U_{t-1}(\theta) : \cdots : U_{t-q}(\theta)\right],$$

where

$$V_t(\theta) = -\sum_{h=0}^{\infty} \mathbf{A}^*_h \left(I_{d^2} \otimes e_{t-h}(\theta)\right) \quad \text{and} \quad U_t(\theta) = \sum_{h=0}^{\infty} \mathbf{B}^*_h \left(I_{d^2} \otimes e_{t-h}(\theta)\right)$$

with the $d \times d^3$ matrices

$$\mathbf{A}^*_h = \left[\mathbf{A}^*_{11,h} : \mathbf{A}^*_{21,h} : \cdots : \mathbf{A}^*_{dd,h}\right] \text{ and } \mathbf{B}^*_h = \left[\mathbf{B}^*_{11,h} : \mathbf{B}^*_{21,h} : \cdots : \mathbf{B}^*_{dd,h}\right].$$

Moreover, at $\theta = \theta_0$ we have

$$\frac{\partial e_t}{\partial \theta'} = \sum_{i \geq 1} \lambda_i \left(I_{d^2(p+q)} \otimes e_{t-i}\right),$$

with the λ_i's are defined by (3.1).

4. Explicit expression of I and J

The subject of this section is to give expressions for the information matrices I and J involved in the asymptotic variance Ω of the QMLE. In these expressions, we isolate what is a function of the VARMA parameter θ_0 from what is a function of the distribution of the weak noise e_t.

McLeod [38] gave a nice expression for J, for the univariate ARMA model, as the variance of a VAR model involving only the ARMA parameter θ_0 (see (8.8.3) in [12]). Francq, Roy and Zakoïan [22] obtained an expression of I involving the ARMA parameter θ_0 and the fourth-order moments of the weak noise (ϵ_t) (with their notations, $J = \mathbf{\Lambda}'_\infty \mathbf{\Lambda}_\infty$ and $I = \mathbf{\Lambda}'_\infty \mathbf{\Gamma}_{\infty,\infty} \mathbf{\Lambda}_\infty$ where $\mathbf{\Lambda}_\infty$ depends on θ_0 and $\mathbf{\Gamma}_{\infty,\infty}$ depends on moments of (ϵ_t)). For certain statistical applications, recently, Boubacar Maïnassara, Carbon and Francq [8] give similar expressions for $I(\theta)$ and $J(\theta)$ when $\theta \neq \theta_0$. Let us define the matrix

$$\mathcal{M} := \mathbb{E}\left\{\left(I_{d^2(p+q)} \otimes e_t'\right)^{\otimes 2}\right\}$$

involving the second-order moments of (e_t). We are now able to state the following proposition, which provides a form for $J = J(\theta_0, \Sigma_{e0})$, in which the terms depending on θ_0 (through the matrices λ_i) are distinguished from the terms depending on the second-order moments of (e_t) (through the matrix $\mathcal{M}$) and the terms of the noise variance of the multivariate innovation process (through the matrix Σ_{e0}).

Proposition 4.1. *Under Assumptions* **A1**–**A8***, we have*

$$\operatorname{vec} J = 2\sum_{i\geq 1}\mathcal{M}\left\{\lambda_i' \otimes \lambda_i'\right\}\operatorname{vec}\Sigma_{e0}^{-1},$$

where the λ_i*'s are defined by (3.1).*

We now search similar tractable expressions for I. In view of (2.3), we have

$$I = \lim_{n\to\infty}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Upsilon_t\right) = \sum_{h=-\infty}^{+\infty}\operatorname{Cov}(\Upsilon_t, \Upsilon_{t-h}), \tag{4.1}$$

where

$$\Upsilon_t = \frac{\partial}{\partial\theta}\left\{\log\det\Sigma_{e0} + e_t'(\theta)\Sigma_{e0}^{-1}e_t(\theta)\right\}_{\theta=\theta_0}. \tag{4.2}$$

Note that, the existence of the sum of the right-hand side of (4.1) is a consequence of **A7** and of Davydov's inequality [13] (see *e.g.* Lemma 11 in [9]). Let the matrices

$$\mathcal{M}_{ij,h} := \mathbb{E}\left(\left\{e_{t-h}' \otimes \left(I_{d^2(p+q)} \otimes e_{t-j-h}'\right)\right\} \otimes \left\{e_t' \otimes \left(I_{d^2(p+q)} \otimes e_{t-i}'\right)\right\}\right).$$

The terms depending on the VARMA parameter are the matrices λ_i defined in (3.1) and let the matrices

$$\boldsymbol{\Gamma}(i,j) = \sum_{h=-\infty}^{+\infty}\mathcal{M}_{ij,h}$$

involving the fourth-order moments of the innovations e_t. The terms depending on the noise variance of the multivariate innovation process are in Σ_{e0}. We now state an analog of Proposition 4.1 for $I = I(\theta_0, \Sigma_{e0})$.

Proposition 4.2. *Under Assumptions* **A1**–**A8***, we have*

$$\operatorname{vec} I = 4\sum_{i,j=1}^{+\infty}\boldsymbol{\Gamma}(i,j)\left(\left\{I_d \otimes \lambda_j'\right\} \otimes \left\{I_d \otimes \lambda_i'\right\}\right)\operatorname{vec}\left(\operatorname{vec}\Sigma_{e0}^{-1}\left\{\operatorname{vec}\Sigma_{e0}^{-1}\right\}'\right),$$

where the λ_i*'s are defined by (3.1).*

Remark 4.1. Consider the univariate case $d = 1$. We obtain

$$\mathrm{vec}\, J = 2\sum_{i\geq 1}\{\lambda_i \otimes \lambda_i\}' \ \text{ and } \ \mathrm{vec}\, I = \frac{4}{\sigma^4}\sum_{i,j=1}^{+\infty}\gamma(i,j)\{\lambda_j \otimes \lambda_i\}',$$

where $\gamma(i,j) = \sum_{h=-\infty}^{+\infty}\mathbb{E}(e_t e_{t-i} e_{t-h} e_{t-j-h})$ and $\lambda_i' \in \mathbb{R}^{p+q}$ are defined by (3.1).

Remark 4.2. Francq, Roy and Zakoïan [22] considered the univariate case $d = 1$. In their paper, they used the LS estimator and they obtained

$$\mathbb{E}\left(\frac{\partial^2 e_t(\theta_0)}{\partial\theta\partial\theta'}\right) = 2\sum_{i\geq 1}\sigma^2\lambda_i\lambda_i' \ \text{ and } \ \mathrm{Var}\left\{2e_t(\theta_0)\frac{\partial e_t(\theta_0)}{\partial\theta}\right\} = 4\sum_{i,j\geq 1}\gamma(i,j)\lambda_i\lambda_j'$$

where σ^2 is the variance of the univariate process e_t and the vectors $\lambda_i = (-\phi_{i-1}^*, \dots, -\phi_{i-p}^*, \varphi_{i-1}^*, \dots, \varphi_{i-q}^*)' \in \mathbb{R}^{p+q}$, with the convention $\phi_i^* = \varphi_i^* = 0$ when $i < 0$. Let $\tilde{\ell}_n(\theta, \Sigma_e) = -2n^{-1}\log\tilde{\mathrm{L}}_n(\theta, \Sigma_e)$. In [9], it is shown that $\ell_n(\theta, \Sigma_e) = \tilde{\ell}_n(\theta, \Sigma_e) + o(1)$ a.s., where

$$\ell_n(\theta, \Sigma_e) = \frac{1}{n}\sum_{t=1}^{n}\left\{d\log(2\pi) + \log\det\Sigma_e + e_t'(\theta)\Sigma_e^{-1}e_t(\theta)\right\}.$$

Using the vec operator and the elementary relation $\mathrm{vec}(aa') = a \otimes a'$, their result writes

$$\mathrm{vec}\, J = \mathrm{vec}\left\{\mathbb{E}\left(\frac{\partial^2\ell_n(\theta_0, \Sigma_{e0})}{\partial\theta\partial\theta'}\right)\right\} = \frac{1}{\sigma^2}\,\mathrm{vec}\left\{\mathbb{E}\left(\frac{\partial^2}{\partial\theta\partial\theta'}e_t^2(\theta_0)\right)\right\} = 2\sum_{i\geq 1}\lambda_i \otimes \lambda_i$$

$$\begin{aligned}\text{and } \ \mathrm{vec}\, I &= \mathrm{vec}\left\{\mathrm{Var}\left(\frac{\partial\ell_n(\theta_0, \Sigma_{e0})}{\partial\theta}\right)\right\} = \frac{1}{\sigma^4}\,\mathrm{vec}\left\{\mathrm{Var}\left(2e_t\frac{\partial e_t(\theta_0)}{\partial\theta}\right)\right\} \\ &= \frac{4}{\sigma^4}\sum_{i,j\geq 1}\gamma(i,j)\lambda_i \otimes \lambda_j,\end{aligned}$$

which are the expressions given in Remark 4.1.

The following example illustrates that how the matrices I and J are depend on the terms θ_0 and terms involving the distribution of the innovations ϵ_t.

Example 1. Consider for instance a weak, univariate, ARMA(1,1) of the form

$$X_t = aX_{t-1} - b\epsilon_{t-1} + \epsilon_t,$$

with variance σ^2. Then, with our notations, we have $\theta = (a, b)'$, $\theta_0 = (a_0, b_0)'$, $\mathbf{A}_i^* = a^i$, $\mathbf{B}_i^* = b^i$, $\lambda_i = (-\mathbf{A}_{i-1}^*, \mathbf{B}_{i-1}^*) = (-a^{i-1}, b^{i-1})$, $\mathbb{E}(\epsilon_t^2) = \sigma_0^2$,

$$\mathbf{\Gamma}(i,j) = \sum_{h=-\infty}^{+\infty}\mathbb{E}\left(\epsilon_t\epsilon_{t-i}\epsilon_{t-h}\epsilon_{t-j-h}\right)I_4 = \gamma(i,j)I_4$$

and

$$\lambda_j \otimes \lambda_i = \left(a^{j+i-2}, -a^{j-1}b^{i-1}, -a^{j-1}b^{i-1}, b^{j+i-2}\right).$$

Thus, we have

$$\begin{aligned}
\operatorname{vec} J &= 2\sum_{i\geq 1}\left(a_0^{2(i-1)}, -(a_0b_0)^{i-1}, -(a_0b_0)^{i-1}, b_0^{2(i-1)}\right)' \text{ and} \\
\operatorname{vec} I &= \frac{4}{\sigma_0^4}\sum_{i,j=1}^{+\infty}\gamma(i,j)\left(a_0^{j+i-2}, -a_0^{j-1}b_0^{i-1}, -a_0^{j-1}b_0^{i-1}, b_0^{j+i-2}\right)',
\end{aligned}$$

where σ_0 is the true value of σ. We then deduce that

$$J = 2\begin{bmatrix} \frac{1}{1-a_0^2} & \frac{-1}{1-a_0b_0} \\ \frac{-1}{1-a_0b_0} & \frac{1}{1-b_0^2} \end{bmatrix} \text{ and } I = \frac{4}{\sigma_0^4}\sum_{i,j=1}^{+\infty}\gamma(i,j)\begin{bmatrix} a_0^{j+i-2} & -a_0^{j-1}b_0^{i-1} \\ -a_0^{j-1}b_0^{i-1} & b_0^{j+i-2} \end{bmatrix}.$$

Thus, in the standard strong ARMA case, *i.e.* when **A4** is replaced by the assumption that (ϵ_t) is iid, it is easily seen that $\gamma(i,j) = [\mathbb{E}(\epsilon_t^2)]^2 = \sigma_0^4$ when $i = j$ and 0 if $i \neq j$, so that $I = 2J$. In the general case we have $I \neq 2J$. □

5. Estimating the asymptotic variance matrix

In section 4, we obtained explicit expressions for I and J. We now turn to the estimation of these matrices. Let $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ be the QMLE residuals when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$ and

$$\hat{e}_t = X_t - \sum_{i=1}^{p} A_0^{-1}(\hat{\theta}_n)A_i(\hat{\theta}_n)\hat{X}_{t-i} + \sum_{i=1}^{q} A_0^{-1}(\hat{\theta}_n)B_i(\hat{\theta}_n)B_0^{-1}(\hat{\theta}_n)A_0(\hat{\theta}_n)\hat{e}_{t-i},$$

for $t = 1,\dots,n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$. Let $\hat{\Sigma}_{e0} = n^{-1}\sum_{t=1}^{n}\hat{e}_t\hat{e}_t'$ be an estimator of Σ_{e0}. The matrix $\mathcal{M}$ involved in the expression of J can easily be estimated by its empirical counterpart

$$\hat{\mathcal{M}}_n := \frac{1}{n}\sum_{t=1}^{n}\left\{\left(I_{d^2(p+q)} \otimes \hat{e}_t'\right)^{\otimes 2}\right\}.$$

In view of Proposition 4.1, we define an estimator $\hat{J}_n$ of J by

$$\operatorname{vec}\hat{J}_n = \sum_{i\geq 1}\hat{\mathcal{M}}_n\left\{\hat{\lambda}_i' \otimes \hat{\lambda}_i'\right\}\operatorname{vec}\hat{\Sigma}_{e0}^{-1}.$$

We are now able to state the following theorem, which shows the strong consistency of $\hat{J}_n$.

Theorem 5.1. *Under Assumptions* **A1**–**A8**, *we have*

$$\hat{J}_n \to J \text{ a.s. as } n \to \infty.$$

In the standard strong VARMA case $\hat{\Omega} = 2\hat{J}^{-1}$ is a strongly consistent estimator of Ω. In the general weak VARMA case this estimator is not consistent when $I \neq 2J$. So we now need a consistent estimator of I. The estimation of the long-run variance I is more complicated. In the literature, two types of estimators are generally employed: HAC estimators (see [2, 40] for general references, and [26] for an application to testing strong linearity in weak ARMA models) and spectral density estimators (see [4] and also den [14] for a general reference; see also [9] for an application to a weak VARMA model). Let

$$\mathcal{M}_{n\ ij,h} := \frac{1}{n} \sum_{t=1}^{n-|h|} \left(\left\{ e'_{t-h} \otimes \left(I_{d^2(p+q)} \otimes e'_{t-j-h} \right) \right\} \otimes \left\{ e'_t \otimes \left(I_{d^2(p+q)} \otimes e'_{t-i} \right) \right\} \right).$$

To estimate $\mathbf{\Gamma}(i,j)$ consider a sequence of real numbers $(b_n)_{n \in \mathbb{N}^*}$ such that

$$b_n \to 0 \quad \text{and} \quad n b_n^{\frac{10+4\nu}{\nu}} \to \infty \quad \text{as} \quad n \to \infty, \tag{5.1}$$

and a weight function $f : \mathbb{R} \to \mathbb{R}$ which is bounded, with compact support $[-a, a]$ and continuous at the origin with $f(0) = 1$. Note that under the above assumptions, we have

$$b_n \sum_{|h|<n} |f(hb_n)| = O(1). \tag{5.2}$$

Consider the matrix

$$\hat{\mathbf{\Gamma}}_n(i,j) := \sum_{h=-T_n}^{+T_n} f(hb_n) \hat{\mathcal{M}}_{n\ ij,h} \quad \text{and} \quad T_n = \left[\frac{a}{b_n} \right],$$

where $[x]$ denotes the integer part of x and where

$$\hat{\mathcal{M}}_{n\ ij,h} := \frac{1}{n} \sum_{t=1}^{n-|h|} \left(\left\{ \hat{e}'_{t-h} \otimes \left(I_{d^2(p+q)} \otimes \hat{e}'_{t-j-h} \right) \right\} \otimes \left\{ \hat{e}'_t \otimes \left(I_{d^2(p+q)} \otimes \hat{e}'_{t-i} \right) \right\} \right).$$

In view of Proposition 4.2, we define an estimator $\hat{I}_n$ of I by

$$\operatorname{vec} \hat{I}_n = 4 \sum_{i,j=1}^{+\infty} \hat{\mathbf{\Gamma}}_n(i,j) \left(\left\{ I_d \otimes \hat{\lambda}'_i \right\} \otimes \left\{ I_d \otimes \hat{\lambda}'_j \right\} \right) \operatorname{vec} \left(\operatorname{vec} \hat{\Sigma}_{e0}^{-1} \left\{ \operatorname{vec} \hat{\Sigma}_{e0}^{-1} \right\}' \right).$$

We are now able to state the following theorem, which shows the weak consistency of an empirical estimator of $\hat{I}_n$.

Theorem 5.2. *Under Assumptions* **A1**–**A8***, we have*

$$\hat{I}_n \to I \text{ in probability as } n \to \infty.$$

Therefore Theorems 5.1 and 5.2 show that

$$\hat{\Omega}_n := \hat{J}_n^{-1} \hat{I}_n \hat{J}_n^{-1}$$

is a weakly estimator of the asymptotic covariance matrix $\Omega = J^{-1} I J^{-1}$, which allows to separate the effects due to the VARMA parameters from those due to the nonlinear structure of the noise.

6. Approximation of the information matrices by finite sums

In practice the infinite sums involved in J and I are truncated. This section concentrates on the choice of the truncation parameter for J and I. Matrix J is truncated by the matrix J^M and defined by

$$\operatorname{vec} J^M = 2\sum_{i=1}^{M} \mathcal{M}\left\{\lambda_i' \otimes \lambda_i'\right\} \operatorname{vec} \Sigma_{e0}^{-1}.$$

The following proposition defines a value of M such that J^M be equal to J up to an arbitrarily small tolerance number ε. Let the matrix norm defined by $\|A\| = \sum_{i,j} |A(i,j)|$ with obvious notations.

Proposition 6.1. *Let $\overline{\rho}$ be the inverse of the largest modulus of the zeroes of the polynomials $\det A_\theta(z)$ and $\det B_\theta(z)$ and let*

$$K_1 = d^6(p+q)^3\left(\frac{-2d(p+q)}{\log\overline{\rho}}\right)^{d(p+q)} \overline{\rho}^{\,-0.5-d(p+q)/\log\overline{\rho}}.$$

For all $\varepsilon > 0$, we can therefore choose an integer

$$M \geq M_\epsilon := \frac{\log\left(\sqrt{\varepsilon/2\overline{\pi}\overline{\Gamma}}(1-\sqrt{\overline{\rho}})/K_1\right)}{\log\overline{\rho}}$$

such that $\|\operatorname{vec} J - \operatorname{vec} J^M\| \leq \varepsilon$, where $\overline{\Gamma} = \|\mathcal{M}\|$ and $\overline{\pi} = \|\operatorname{vec}\Sigma_{e0}^{-1}\|$.

Similarly to J, the matrix I is truncated by the matrix I^M of M^2 terms, defined by

$$\operatorname{vec} I^M = 4\sum_{i,j=1}^{M} \mathbf{\Gamma}(i,j)\left(\left\{I_d\otimes\lambda_j'\right\}\otimes\left\{I_d\otimes\lambda_i'\right\}\right)\operatorname{vec}\left(\operatorname{vec}\Sigma_{e0}^{-1}\left\{\operatorname{vec}\Sigma_{e0}^{-1}\right\}'\right).$$

We now state an analog of Proposition 6.1 for I.

Proposition 6.2. *Let*

$$\overline{\mathbf{\Gamma}} = \max_{i,j\geq 0}\|\mathbf{\Gamma}(i,j)\|, \qquad \overline{\pi}_1 = \left\|\operatorname{vec}\left(\operatorname{vec}\Sigma_{e0}^{-1}\left\{\operatorname{vec}\Sigma_{e0}^{-1}\right\}'\right)\right\|.$$

For all $\varepsilon > 0$, we can therefore choose an integer

$$M \geq M_\epsilon := \frac{\log\left(\sqrt{\varepsilon/4\overline{\pi}_1 d^2\overline{\mathbf{\Gamma}}}(1-\sqrt{\overline{\rho}})/K_1\right)}{\log\overline{\rho}}$$

such that $\|\operatorname{vec} I - \operatorname{vec} I^M\| \leq \varepsilon$.

7. Testing linear restrictions on the parameter

It may be of interest to test s_0 linear constraints on the elements of θ_0 (in particular $A_{0p} = 0$ or $B_{0q} = 0$). We thus consider a null hypothesis of the form

$$H_0 : R_0\theta_0 = \mathfrak{r}_0$$

where R_0 is a known $s_0 \times k_0$ matrix of rank s_0 and $\mathfrak{r}_0$ is a known s_0-dimensional vector. The Wald, LM and LR principles are employed frequently for testing H_0. The LM test is also called the score or Rao-score test. We now examine if these principles remain valid in the non standard framework of weak VARMA models.

Let $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$, where $\hat{J}$ and $\hat{I}$ are consistent estimator of J and I, as defined in Section 5. Under Assumptions **A1**–**A8**, and the assumption that I is invertible, the Wald statistic

$$\mathbf{W}_n = n(R_0\hat{\theta}_n - \mathfrak{r}_0)'(R_0\hat{\Omega}R_0')^{-1}(R_0\hat{\theta}_n - \mathfrak{r}_0)$$

asymptotically follows a $\chi^2_{s_0}$ distribution under H_0. Therefore, the standard formulation of the Wald test remains valid. More precisely, at the asymptotic level α, the Wald test consists in rejecting H_0 when $\mathbf{W}_n > \chi^2_{s_0}(1-\alpha)$. It is however important to note that a consistent estimator of the form $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$ is required. The estimator $\hat{\Omega} = 2\hat{J}^{-1}$, which is routinely used in the time series softwares, is only valid in the strong VARMA case.

We now turn to the LM test. Let $\hat{\theta}^c_n$ be the restricted QMLE of the parameter under H_0. Define the Lagrangean

$$\mathcal{L}(\theta, \lambda) = \tilde{\ell}_n(\theta) - \lambda'(R_0\theta - \mathfrak{r}_0),$$

where λ denotes a s_0-dimensional vector of Lagrange multipliers. The first-order conditions yield

$$\frac{\partial\tilde{\ell}_n}{\partial\theta}(\hat{\theta}^c_n) = R_0'\hat{\lambda}, \quad R_0\hat{\theta}^c_n = \mathfrak{r}_0.$$

It will be convenient to write $a \stackrel{c}{=} b$ to signify $a = b + c$. A Taylor expansion gives under H_0

$$0 \;=\; \sqrt{n}\frac{\partial\tilde{\ell}_n(\hat{\theta}_n)}{\partial\theta} \stackrel{o_P(1)}{=} \sqrt{n}\frac{\partial\tilde{\ell}_n(\hat{\theta}^c_n)}{\partial\theta} - J\sqrt{n}\left(\hat{\theta}_n - \hat{\theta}^c_n\right).$$

We deduce that

$$\sqrt{n}(R_0\hat{\theta}_n - \mathfrak{r}_0) \;=\; R_0\sqrt{n}(\hat{\theta}_n - \hat{\theta}^c_n) \stackrel{o_P(1)}{=} R_0J^{-1}\sqrt{n}\frac{\partial\tilde{\ell}_n(\hat{\theta}^c_n)}{\partial\theta} = R_0J^{-1}R_0'\sqrt{n}\hat{\lambda}.$$

Thus under H_0 and the previous assumptions,

$$\sqrt{n}\hat{\lambda} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, (R_0J^{-1}R_0')^{-1}R_0\Omega R_0'(R_0J^{-1}R_0')^{-1}\right\}, \tag{7.1}$$

so that the LM statistic is defined by

$$\begin{aligned}\mathbf{LM}_n &= n\hat{\lambda}'\left\{(R_0\hat{J}^{-1}R_0')^{-1}R_0\hat{\Omega}R_0'(R_0\hat{J}^{-1}R_0')^{-1}\right\}^{-1}\hat{\lambda}\\ &= n\frac{\partial\tilde{\ell}_n}{\partial\theta'}(\hat{\theta}_n^c)\hat{J}^{-1}R_0'\left(R_0\hat{\Omega}R_0'\right)^{-1}R_0\hat{J}^{-1}\frac{\partial\tilde{\ell}_n}{\partial\theta}(\hat{\theta}_n^c).\end{aligned}$$

Note that in the strong VARMA case, $\hat{\Omega} = 2\hat{J}^{-1}$ and the LM statistic takes the more conventional form $\mathbf{LM}_n^* = (n/2)\hat{\lambda}'R_0\hat{J}^{-1}R_0'\hat{\lambda}$. In the general case, strong and weak as well, the convergence (7.1) implies that the asymptotic distribution of the $\mathbf{LM}_n$ statistic is $\chi^2_{s_0}$ under H_0. The null is therefore rejected when $\mathbf{LM}_n > \chi^2_{s_0}(1-\alpha)$. Of course the conventional LM test with rejection region $\mathbf{LM}_n^* > \chi^2_{s_0}(1-\alpha)$ is not asymptotically valid for general weak VARMA models. Standard Taylor expansions show that

$$\sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^c) \stackrel{o_P(1)}{=} -\sqrt{n}J^{-1}R_0'\hat{\lambda},$$

and that the LR statistic satisfies

$$\begin{aligned}\mathbf{LR}_n &:= 2\left\{\log\tilde{\mathrm{L}}_n(\hat{\theta}_n) - \log\tilde{\mathrm{L}}_n(\hat{\theta}_n^c)\right\} \stackrel{o_P(1)}{=} \frac{n}{2}(\hat{\theta}_n - \hat{\theta}_n^c)'J(\hat{\theta}_n - \hat{\theta}_n^c)\\ &\stackrel{o_P(1)}{=} \frac{n}{2}\hat{\lambda}'R_0\hat{J}^{-1}R_0'\hat{\lambda} \stackrel{o_P(1)}{=} \mathbf{LM}_n^*.\end{aligned}$$

Using the previous computations and standard results on quadratic forms of normal vectors (see *e.g.* Lemma 17.1 in [45]), we find that the $\mathbf{LR}_n$ statistic is asymptotically distributed as $\sum_{i=1}^{s_0}\lambda_i Z_i^2$ where the Z_i's are iid $\mathcal{N}(0,1)$ and $\lambda_1,\dots,\lambda_{s_0}$ are the eigenvalues of

$$\Sigma_{\mathbf{LR}} = J^{-1/2}S_{\mathbf{LR}}J^{-1/2}, \quad S_{\mathbf{LR}} = \frac{1}{2}R_0'(R_0J^{-1}R_0')^{-1}R_0\Omega R_0'(R_0J^{-1}R_0')^{-1}R_0.$$

Note that when $\Omega = 2J^{-1}$, the matrix $\Sigma_{\mathbf{LR}} = J^{-1/2}R_0'(R_0J^{-1}R_0')^{-1}R_0J^{-1/2}$ is a projection matrix. Its eigenvalues are therefore equal to 0 and 1, and the number of eigenvalues equal to 1 is $\mathrm{Tr}\, J^{-1/2}R_0'(R_0J^{-1}R_0')^{-1}R_0J^{-1/2} = \mathrm{Tr}\, I_{s_0} = s_0$. Therefore we retrieve the well-known result that $\mathbf{LR}_n \sim \chi^2_{s_0}$ under H_0 in the strong VARMA case. In the weak VARMA case, the asymptotic null distribution of $\mathbf{LR}_n$ is complicated. It is possible to evaluate the distribution of a quadratic form of a Gaussian vector by means of the Imhof algorithm (see [31]). An alternative is to use the transformed statistic

$$\mathbf{LR}_n^- := \frac{n}{2}(\hat{\theta}_n - \hat{\theta}_n^c)'\hat{J}\hat{S}_{\mathbf{LR}}^-\hat{J}(\hat{\theta}_n - \hat{\theta}_n^c) \tag{7.2}$$

which follows a $\chi^2_{s_0}$ under H_0, when $\hat{J}$ and $\hat{S}_{\mathbf{LR}}^-$ are weakly consistent estimators of J and of a generalized inverse of $S_{\mathbf{LR}}$. The estimator $\hat{S}_{\mathbf{LR}}^-$ can be obtained from the singular value decomposition of any weakly consistent estimator $\hat{S}_{\mathbf{LR}}$ of $S_{\mathbf{LR}}$. More precisely, defining the diagonal matrix $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1,\dots,\hat{\lambda}_{k_0})$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{k_0}$ denote the eigenvalues of the symmetric matrix $\hat{S}_{\mathbf{LR}}$, and denoting by $\hat{P}$ an orthonormal matrix such that $\hat{S}_{\mathbf{LR}} = \hat{P}\hat{\Lambda}\hat{P}'$, one can set

$$\hat{S}_{\mathbf{LR}}^{-} = \hat{P}\hat{\Lambda}^{-}\hat{P}', \quad \hat{\Lambda}^{-} = \mathrm{diag}\left(\hat{\lambda}_1^{-1}, \dots, \hat{\lambda}_{s_0}^{-1}, 0, \dots, 0\right).$$

The matrix $\hat{S}_{\mathbf{LR}}^{-}$ then converges weakly to a matrix $S_{\mathbf{LR}}^{-}$ satisfying $S_{\mathbf{LR}}S_{\mathbf{LR}}^{-}S_{\mathbf{LR}} = S_{\mathbf{LR}}$, because $S_{\mathbf{LR}}$ has full rank s_0.

The obvious problem with this modified version of the LR statistic is that $\Sigma_{\mathbf{LR}}$ is not invertible in the general situation where $s_0 < k_0$, which invalidates the asymptotic $\chi^2_{s_0}$ distribution and also entails numerical problems in the computation of (7.2).

8. Numerical illustrations

In this section, by means of Monte Carlo experiments, we illustrate the finite sample behavior of our estimators of the information matrices I and J (hereafter denoted respectively I_c and J_c, where the subscript on I and J denoting computed) involved in the asymptotic matrix variance Ω, for strong and weak VARMA models. We used, the three kernels described in Table 1 below in the calculation of the estimator of the matrix I_c. Let,

$$\hat{J} = \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial}{\partial\theta}\tilde{e}_t'(\hat{\theta}_n)\right\}\hat{\Sigma}_e^{-1}\left\{\frac{\partial}{\partial\theta'}\tilde{e}_t(\hat{\theta}_n)\right\}.$$

We compare our estimator, of Ω, with the standard and the spectral density estimators $\hat{\Omega}^{SP} := \hat{J}^{-1}\hat{I}^{SP}\hat{J}^{-1}$ ($\hat{I}^{SP}$ is defined in Theorem 3 of [9]) considered by Boubacar Maïnassara and Francq [9].

8.1. Simulating models

The structural VARMA(p,q) representation (2.1) can be rewritten in a standard reduced VARMA(p,q) form if the matrices A_{00} and B_{00} are non singular. Indeed, premultiplying (2.1) by A_{00}^{-1} and introducing the innovation process $e_t = A_{00}^{-1}B_{00}\epsilon_t$, with non singular variance $\Sigma_{e0} = A_{00}^{-1}B_{00}\Sigma_0 B_{00}' A_{00}^{-1'}$, we obtain the reduced VARMA representation

$$X_t - \sum_{i=1}^{p} A_{00}^{-1}A_{0i}X_{t-i} = e_t - \sum_{i=1}^{q} A_{00}^{-1}B_{0i}B_{00}^{-1}A_{00}e_{t-i}. \tag{8.1}$$

The structural form (2.1) allows to handle seasonal models, instantaneous economic relationships, VARMA in the so-called echelon form representation, and many other constrained VARMA representations (see [36], chap. 12). The reduced form (8.1) is more practical from a statistical viewpoint, because it gives the forecasts of each component of (X_t) according to the past values of the set of the components.

The above discussion shows that VARMA representations are not unique, that is, a given process (X_t) can be written in reduced form or in structural form by premultiplying by any non singular $(d \times d)$ matrix. Of course, in order to ensure the uniqueness of a VARMA representation, constraints are necessary for the identifiability of the $(p+q+2)d^2$ elements of the matrices involved in the VARMA equation (2.1). In contrast, the echelon form guarantees uniqueness of the VARMA representation (see also [36]). The echelon form is the most widely identified VARMA representation employed in the literature. The identifiability of VARMA processes has been studied in particular by Hannan [28] who gave several procedures ensuring identifiability.

To generate the strong and the weak VARMA models, we consider the following bivariate VARMA(1,1) model in echelon form considered in Lütkepohl ([36], chap. 12, eq. 12.1.19)

$$\begin{aligned}\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & a_1(2,2) \end{pmatrix}\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \\ &\quad - \begin{pmatrix} 0 & 0 \\ b_1(2,1) & b_1(2,2) \end{pmatrix}\begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix},\end{aligned} \tag{8.2}$$

where $\theta_0 = (a_1(2,2), b_1(2,1), b_1(2,2)) = (0.95, 2, 0)$ and $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ follows a strong or weak white noise.

8.2. Implementation of the estimating of I_c and J_c

Let $X_1, \dots, X_n$, be observations of the bivariate VARMA$(1,1)$ process (8.2). With ours notations, the VARMA representation (8.2) can be rewritten as the compact form

$$\mathbf{A}_\theta(L)X_t = \mathbf{B}_\theta(L)\epsilon_t(\theta),$$

where $\mathbf{A}_\theta(L) = I_2 - \mathbf{A}L$ and $\mathbf{B}_\theta(L) = I_2 - \mathbf{B}L$, with

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & a_1(2,2) \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 0 & 0 \\ b_1(2,1) & b_1(2,2) \end{pmatrix}.$$

For simplicity, we now write a_1, b_1 and b_2 instead of $a_1(2,2)$, $b_1(2,1)$ and $b_1(2,2)$. For estimating the asymptotic matrix variance Ω, introduced in this paper, we implement the information matrices I_c and J_c involved in Ω, using the following steps:

1. Compute the estimates $\hat{\mathbf{A}}, \hat{\mathbf{B}}$ by QMLE.
2. Compute the QMLE residuals $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$ and
$$\hat{e}_t = X_t - \mathbf{A}(\hat{\theta}_n)\hat{X}_{t-1} + \mathbf{B}(\hat{\theta}_n)\hat{e}_{t-1},$$
for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$.
3. Compute the polynomials inverses

$$\mathbf{A}_\theta^{-1}(L) = \sum_{i=0}^{\infty} \mathbf{A}^i L^i = I_2 + \sum_{i=1}^{\infty} \begin{pmatrix} 0 & 0 \\ 0 & a_1^i \end{pmatrix} \text{ and}$$

$$\mathbf{B}_\theta^{-1}(L) = \sum_{i=0}^{\infty} \mathbf{B}^i L^i = I_2 + \sum_{i=1}^{\infty} \begin{pmatrix} 0 & 0 \\ b_1 b_2^{i-1} & b_2^i \end{pmatrix}.$$

4. Compute the $(2 \times 2)-$matrix operators, introduced in this paper, defined by

$$\begin{aligned} M_{22}(L) &= \mathbf{B}_\theta^{-1}(L) E_{22} \mathbf{A}_\theta^{-1}(L) \mathbf{B}_\theta(L) \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -b_1 & a_1 \end{pmatrix} L + \sum_{i=2}^{\infty} \left[\begin{pmatrix} 0 & 0 \\ -b_1 a_1^{i-1} & b_2^i - b_2 a_1^{i-1} \end{pmatrix} \right. \\ &\quad \left. + \sum_{k=1}^{i-1} \begin{pmatrix} 0 & 0 \\ b_1 b_2^k a_1^{i-k-1} & b_2^k \left(a_1^{i-k} - b_2 a_1^{i-k-1} \right) \end{pmatrix} \right] L^i \quad \text{and} \\ N_{2j}(L) &= \mathbf{B}_\theta^{-1}(L) E_{2j}, \text{ for } j = 1, 2 \\ N_{21}(L) &= \sum_{i=0}^{\infty} \begin{pmatrix} 0 & 0 \\ b_2^i & 0 \end{pmatrix} L^i \text{ and } N_{22}(L) = \sum_{i=0}^{\infty} \begin{pmatrix} 0 & 0 \\ 0 & b_2^i \end{pmatrix} L^i \end{aligned}$$

where E_{2j} is the 2×2 matrix with 1 at position $(2, j)$ and 0 elsewhere. We are now able to compute the matrices $\lambda_i(\theta)$'s given in (3.1).

5. We then compute the 2×6 matrix

$$\lambda_h(\theta) = \left[-\mathbf{A}_{h-1}^* : \mathbf{B}_{h-1}^* \right],$$

with

$$\mathbf{A}_h^* = \left[\mathbf{A}_{22,h}^* \right] \text{ and } \mathbf{B}_h^* = \left[\mathbf{B}_{21,h}^* : \mathbf{B}_{22,h}^* \right]$$

and where the 2×2 matrices $\mathbf{A}_{22,h}^*$, $\mathbf{B}_{21,h}^*$ and $\mathbf{B}_{22,h}^*$ are respectively given, in the precedent step, by

$$\begin{aligned} M_{22}(z) &= \sum_{h=0}^{\infty} \mathbf{A}_{22,h}^* z^h, \ N_{21}(z) = \sum_{h=0}^{\infty} \mathbf{B}_{21,h}^* z^h, \\ N_{22}(z) &= \sum_{h=0}^{\infty} \mathbf{B}_{22,h}^* z^h, \ |z| \le 1 \end{aligned}$$

for $h \ge 0$. Take $\mathbf{A}_{22,h}^* = \mathbf{B}_{21,h}^* = \mathbf{B}_{22,h}^* = 0$ when $h < 0$. Thus the matrices $\lambda_i(\theta)$'s can be estimated by plugging, using

$$\hat{\lambda}_h := \lambda_h(\hat{\theta}_n) = \left[-\hat{\mathbf{A}}_{h-1}^* : \hat{\mathbf{B}}_{h-1}^* \right].$$

6. Compute the 9×1 matrix: $\operatorname{vec} \hat{J}_n = \sum_{i=1}^{M} \hat{\mathcal{M}}_n \{ \hat{\lambda}_i' \otimes \hat{\lambda}_i' \} \operatorname{vec} \hat{\Sigma}_{e0}^{-1}$, where $\hat{\Sigma}_{e0} = n^{-1} \sum_{t=1}^{n} \hat{e}_t \hat{e}_t'$ is (an estimator of Σ_{e0}) the empirical variance of $\hat{e}_1, \dots, \hat{e}_n$ and where $\hat{\mathcal{M}}_n = n^{-1} \sum_{t=1}^{n} \{ (I_3 \otimes \hat{e}_t')^{\otimes 2} \}$.

Table 1
Kernels used in the calculation of the matrices $\hat{\mathbf{\Gamma}}_n(i,j)$

Truncated uniform or Rectangular (REC):	$f(x)=\begin{cases}1, & \text{if } \lvert x\rvert\le 1\\ 0, & \text{otherwise}\end{cases}$
Bartlett (BAR):	$f(x)=\begin{cases}1-\lvert x\rvert, & \text{if } \lvert x\rvert\le 1\\ 0, & \text{otherwise}\end{cases}$
Parzen (PAR):	$f(x)=\begin{cases}1-6x^2+6\lvert x\rvert^3, & \text{if } \lvert x\rvert\le 1/2\\ 2(1-\lvert x\rvert)^3, & \text{if } 1/2\le\lvert x\rvert\le 1\\ 0, & \text{otherwise}\end{cases}$

7. Define the 9×144 matrices estimators

$$\hat{\mathbf{\Gamma}}_n(i,j)=\sum_{h=-T_n}^{+T_n} f(hb_n)\hat{\mathcal{M}}_{n\ ij,h} \quad \text{and} \quad T_n=[a/b_n]\,,$$

where $[x]$ denotes the integer part of x and where

$$\hat{\mathcal{M}}_{n\ ij,h}:=\frac{1}{n}\sum_{t=1}^{n-|h|}\left(\left\{\hat{e}'_{t-h}\otimes\left(I_3\otimes\hat{e}'_{t-j-h}\right)\right\}\otimes\left\{\hat{e}'_t\otimes\left(I_3\otimes\hat{e}'_{t-i}\right)\right\}\right).$$

The real numbers $(b_n)_{n\in\mathbb{N}^*}$ satisfy (5.1) and a weight function $f:\mathbb{R}\to\mathbb{R}$ is supposed bounded, with compact support $[-a,a]$ and continuous at the origin with $f(0)=1$. With regard to the choice of (b_n), a few heuristic remarks can be made (see, for instance, [10, 15, 26, 30]). When h is small relative to n, a weight $f(hb_n)$ close to one is required. Therefore, it is supposed that (b_n) decreases to zero as n tends to ∞ (see condition (5.1)). On the contrary, when h is large relative to n, one wants a weight $f(hb_n)$ close to zero. Therefore, it is supposed that (b_n) does not decrease to zero too quickly. We used, in this paper, the three kernels described in Table 1 in the calculation of the matrices estimators $\hat{\mathbf{\Gamma}}_n(i,j)$. For each kernel, we have taken b_n equal to $1/\ln(n)$ which corresponds to the truncation point $T_n=[\ln(n)]$.

8. Compute the 9×1 matrix:

$$\operatorname{vec}\hat{I}_n=4\sum_{i,j=1}^{M}\hat{\mathbf{\Gamma}}_n(i,j)\Big(\left\{I_3\otimes\hat{\lambda}'_i\right\}\otimes\left\{I_3\otimes\hat{\lambda}'_j\right\}\Big)\operatorname{vec}\Big(\operatorname{vec}\hat{\Sigma}_{e0}^{-1}\left\{\operatorname{vec}\hat{\Sigma}_{e0}^{-1}\right\}'\Big).$$

8.3. Empirical size

The numerical illustrations of this section are made with the free statistical software R (see http://cran.r-project.org/). We simulated N independent trajectories of different sizes n of Model (8.2), first with the strong Gaussian noise (8.3), second with the two weak noises (8.4) and (8.5).

8.3.1. Strong VARMA model case

We first consider the strong VARMA case. To generate this model, we assume that in (8.2) the innovation process (ϵ_t) is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IID}\,\mathcal{N}(0, I_2). \tag{8.3}$$

We simulated N independent trajectories of size $n = 1,000$ of Model (8.2) with the strong gaussian noise (8.3). For each of these N replications we estimated the coefficients $(a_1(2,2), b_1(2,1), b_1(2,2))$, using the Gaussian maximum likelihood estimation method, and we compared estimates of the asymptotic variance Ω, of the QMLE, of standard and modified (sandwich) estimators of Ω.

8.3.2. Weak VARMA model case

The GARCH(p,q) models constitute important examples of weak white noises in the univariate case. These models have numerous extensions to the multivariate framework. Jeantheau [32] has proposed a simple extension of the multivariate GARCH(p,q) with conditional constant correlation. For simplicity, we consider the bivariate ARCH(1) model. In which, the process (ϵ_t) verifies the following relation $\epsilon_t = H_t \eta_t$ where $\{\eta_t = (\eta_{1,t}, \eta_{2,t})'\}_t$ is an iid centered process with $\text{Var}\{\eta_{i,t}\} = 1$, for $i = 1, 2$ and H_t is a diagonal matrix whose elements $h_{ii,t}$ verify

$$\begin{pmatrix} h_{11,t}^2 \\ h_{22,t}^2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{pmatrix}.$$

The elements of the matrix A, as well as the vector $(c_1, c_2)'$, are supposed to be positive. In addition, suppose that the stationarity conditions hold. Then, we have

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} h_{11,t} & 0 \\ 0 & h_{22,t} \end{pmatrix} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix}. \tag{8.4}$$

We now repeat the same experiment on two weak VARMA$(1,1)$ models. We first assume that in (8.2) the innovation process (ϵ_t) is an ARCH(1) model defined in equation (8.4) with $c_1 = 0.3$, $c_2 = 0.2$, $a_{11} = 0.45$, $a_{21} = 0.4$ and $a_{22} = 0.25$. In second set of experiments, we assume that in (8.2) the innovation process (ϵ_t) is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}(|\eta_{1,t-1}| + 1)^{-1} \\ \eta_{2,t}(|\eta_{2,t-1}| + 1)^{-1} \end{pmatrix}, \quad \text{with} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID}\,\mathcal{N}(0, I_2). \tag{8.5}$$

This noise is a direct extension of a weak noise defined by Romano and Thombs [42] in the univariate case.

For the estimation of the coefficients, we used the quasi-maximum likelihood estimation method and we compared estimates of the asymptotic variance Ω, of the QMLE, of standard and modified (sandwich) estimators of Ω.

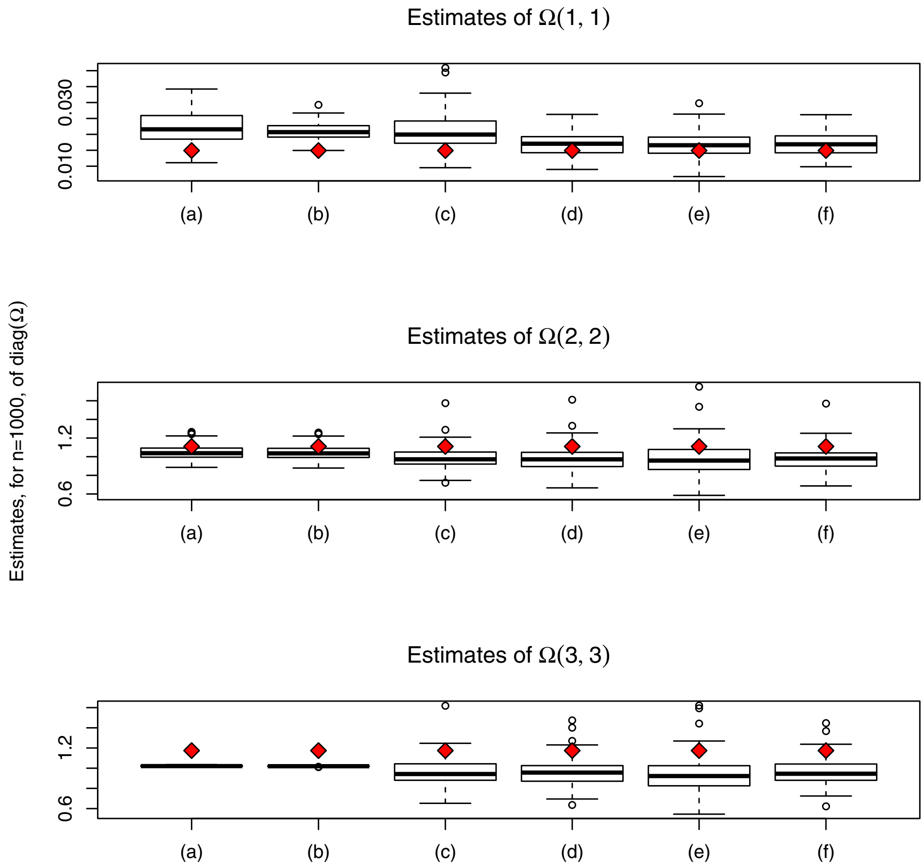

Fig 1. *Comparison of standard and modified estimates of the asymptotic variance* Ω *of the QMLE, on the simulated model (8.2)–(8.3). The diamond symbols represent the mean, over the* $N = 100$ *replications, of the standardized squared errors* $n\{\hat{a}_1(2,2) - 0.95\}^2 = 0.01492$ *for* $\Omega(1,1)$, $n\{\hat{b}_1(2,1) - 2\}^2 = 1.1097$ *for* $\Omega(2,2)$ *and* $n\{\hat{b}_1(2,2)\}^2 = 1.1743$ *for* $\Omega(3,3)$.

8.3.3. Comments

Figures 1, 2 and 3 compare the standard and the sandwich estimators of the QMLE asymptotic variance Ω. In these Figures, (a), (b), (c), (d), (e) and (f) correspond respectively to the box-plots of the estimation of $\Omega^s = 2J^{-1}$, $\Omega^{sc} = 2J_c^{-1}$, $\Omega^{SP} = J^{-1}I^{SP}J^{-1}$, $\Omega^{BAR} = J_c^{-1}I^{BAR}J_c^{-1}$, $\Omega^{REC} = J_c^{-1}I^{REC}J_c^{-1}$ and $\Omega^{PAR} = J_c^{-1}I^{PAR}J_c^{-1}$.

In the strong VARMA case we know that the two estimators, standard and sandwich, are consistent. In view of the three top panels of Figure 1, it seems that the sandwich estimators is less accurate in the strong case. This is not surprising because the sandwich estimators are more robust, in the sense that these estimators continue to be consistent in the weak VARMA case, contrary to

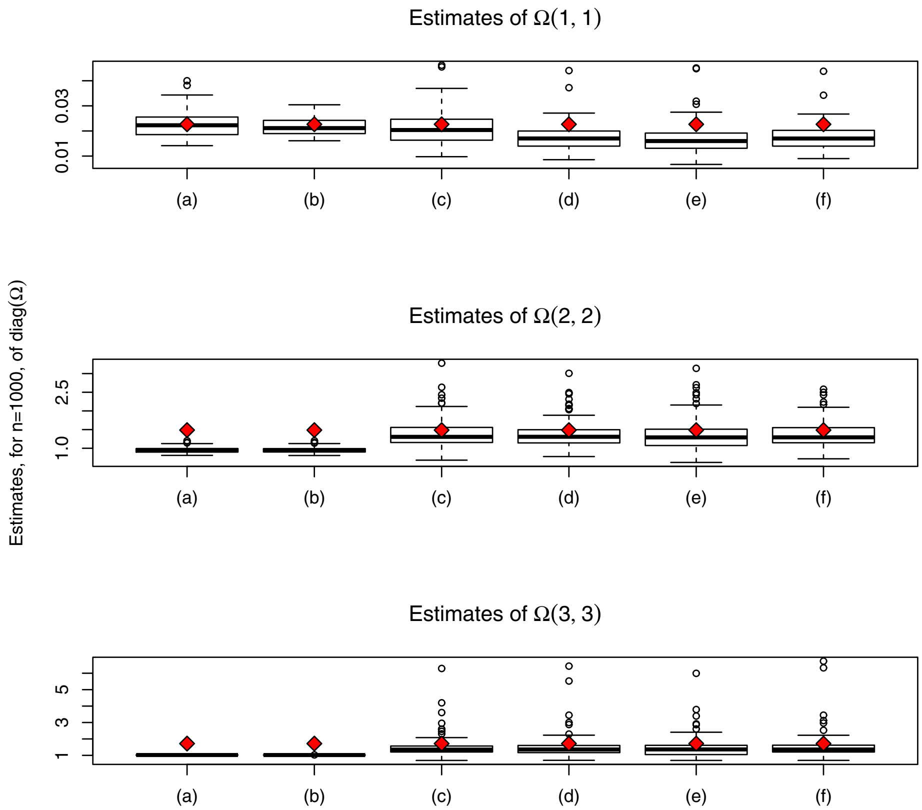

FIG 2. *Comparison of standard and modified estimates of the asymptotic variance* Ω *of the QMLE, on the simulated model (8.2)–(8.4). The diamond symbols represent the mean, over the* $N = 100$ *replications, of the standardized squared errors* $n\{\hat{a}_1(2,2) - 0.95\}^2 = 0.0227$ *for* $\Omega(1,1)$, $n\{\hat{b}_1(2,1) - 2\}^2 = 1.4865$ *for* $\Omega(2,2)$ *and* $n\{\hat{b}_1(2,2)\}^2 = 1.7172$ *for* $\Omega(3,3)$.

the standard estimator (see Figures 2 and 3). It is clear that in the weak Model (8.2)–(8.4) case $n\mathrm{Var}\{\hat{b}_1(2,1)-b_1(2,1)\}^2$ and $n\mathrm{Var}\{\hat{b}_1(2,2)-b_1(2,2)\}^2$ are better estimated by the sandwich estimators than by the standard ones (see Figure 2). We draw the same conclusion for the second weak Model (8.2)–(8.5) case for $n\mathrm{Var}\{\hat{b}_1(2,2)-b_1(2,2)\}^2$ (see Figure 3). We draw the conclusion that, the failure of the standard estimators of Ω in the weak VARMA framework may have important consequences in terms of identification (see [7]) or hypothesis testing.

Table 2 displays the empirical sizes of the standard Wald, LM and LR tests, and that of the modified versions proposed in Section 7. We use the **CompQuadForm** R package to evaluate the p-values using the Imohf algorithm, 1961). For the nominal level $\alpha = 5\%$, the empirical size over the $N = 1,000$ independent replications should vary between the significant limits 3.6% and 6.4% with probability 95%. For the nominal level $\alpha = 1\%$, the significant limits are 0.3% and 1.7%, and for the nominal level $\alpha = 10\%$, they

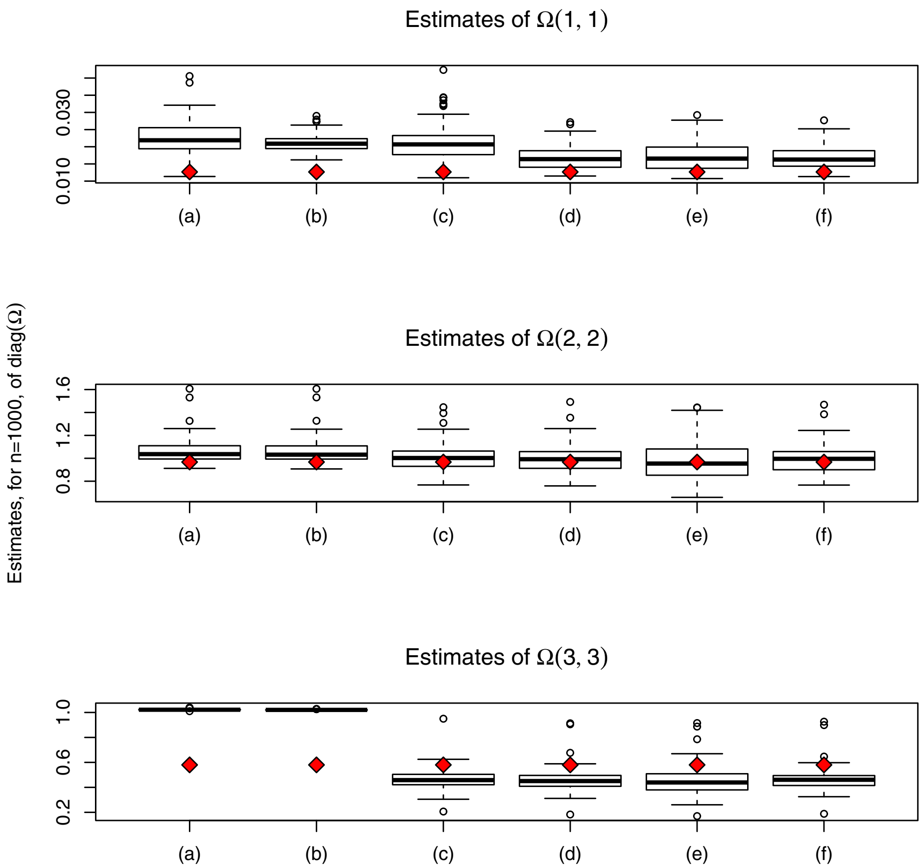

Fig 3. *Comparison of standard and modified estimates of the asymptotic variance* Ω *of the QMLE, on the simulated model (8.2)–(8.5). The diamond symbols represent the mean, over the* $N = 100$ *replications, of the standardized squared errors* $n\{\hat{a}_1(2,2) - 0.95\}^2 = 0.0127$ *for* $\Omega(1,1)$, $n\{\hat{b}_1(2,1) - 2\}^2 = 0.9673$ *for* $\Omega(2,2)$ *and* $n\{\hat{b}_1(2,2)\}^2 = 0.5802$ *for* $\Omega(3,3)$.

are 8.1% and 11.9%. When the relative rejection frequencies are outside the significant limits, they are displayed in bold type in Table 2. For the strong VARMA Model I, all the relative rejection frequencies are inside the significant limits. As expected, for the weak VARMA Models II and III cases, the relative rejection frequencies of the standard tests are definitely outside the significant limits. Thus the error of first kind is well controlled by all the tests in the strong case, but only by modified versions of the tests in the weak case. Table 3 shows that the powers of all the tests are very similar in the strong VARMA Model IV. The same is also true for the modified tests in the weak VARMA Models V and VI cases. The empirical powers of the standard tests are hardly interpretable for these weak VARMA Models V and VI cases, because we have already seen in Table 2 that the standard versions of the tests do not well control the error of first kind in the weak VARMA framework.

Table 2
Empirical size of standard and modified tests: relative frequencies (in %) of rejection of $H_0 : b_1(2,2) = 0$. The number of replications is $N = 1000$

Model	Length n	Level	Standard Test			Modified Test			
			Wald	LM	LR	Wald	LM	LR$^-$	LR
		$\alpha = 1\%$	1.3	0.8	1.0	1.6	1.0	1.6	1.6
I	$n = 500$	$\alpha = 5\%$	5.0	4.3	4.8	6.1	4.9	6.1	6.1
		$\alpha = 10\%$	11.3	10.2	11.0	**12.1**	11.0	**12.2**	**12.2**
		$\alpha = 1\%$	1.5	1.1	1.2	1.3	1.0	1.4	1.3
I	$n = 2,000$	$\alpha = 5\%$	5.2	4.9	4.5	5.5	5.2	5.5	5.5
		$\alpha = 10\%$	9.9	9.7	9.7	10.3	9.8	10.3	10.3
		$\alpha = 1\%$	1.3	1.0	1.2	1.3	1.2	1.3	1.3
I	$n = 5,000$	$\alpha = 5\%$	4.4	4.4	4.4	4.4	4.1	4.3	4.4
		$\alpha = 10\%$	8.7	8.5	8.7	8.5	8.6	8.5	8.5
		$\alpha = 1\%$	**4.3**	**3.7**	**4.0**	**2.7**	1.7	**2.7**	1.6
II	$n = 500$	$\alpha = 5\%$	**11.2**	**11.3**	**11.0**	**7.0**	6.0	**7.0**	6.1
		$\alpha = 10\%$	**21.1**	**19.3**	**20.2**	**12.0**	11.2	11.8	**12.2**
		$\alpha = 1\%$	**6.3**	**5.8**	**6.0**	1.5	1.3	1.5	1.3
II	$n = 2,000$	$\alpha = 5\%$	**15.5**	**15.1**	**15.2**	**7.2**	6.3	**7.2**	5.5
		$\alpha = 10\%$	**23.8**	**23.7**	**24.1**	**12.7**	11.6	**12.7**	10.3
		$\alpha = 1\%$	**6.4**	**6.1**	**6.5**	1.0	1.2	1.0	1.3
II	$n = 5,000$	$\alpha = 5\%$	**14.6**	**14.1**	**14.6**	5.3	5.1	5.3	4.4
		$\alpha = 10\%$	**21.0**	**21.4**	**21.0**	10.6	10.4	10.6	8.5
		$\alpha = 1\%$	**0.1**	**0.0**	**0.0**	**2.2**	1.4	**2.2**	1.6
III	$n = 500$	$\alpha = 5\%$	**1.2**	**0.9**	**0.8**	**7.3**	5.9	**7.3**	6.1
		$\alpha = 10\%$	**3.1**	**2.3**	**2.7**	**12.0**	11.5	**12.0**	**12.2**
		$\alpha = 1\%$	**0.0**	**0.0**	**0.0**	1.7	1.3	1.7	1.3
III	$n = 2,000$	$\alpha = 5\%$	**0.7**	**0.5**	**0.5**	**6.7**	**6.8**	**6.7**	5.5
		$\alpha = 10\%$	**2.7**	**2.4**	**2.7**	11.2	10.9	11.1	10.3
		$\alpha = 1\%$	**0.0**	**0.0**	**0.0**	0.8	0.4	0.8	1.3
III	$n = 5,000$	$\alpha = 5\%$	**0.1**	**0.2**	**0.1**	5.3	4.7	5.2	4.4
		$\alpha = 10\%$	**1.4**	**1.3**	**1.3**	10.2	9.9	10.1	8.5

I: Strong VARMA(1,1) model (8.2)–(8.3) with $\theta_0 = (0.95, 2, 0)$
II: Weak VARMA(1,1) model (8.2)–(8.4) with $\theta_0 = (0.95, 2, 0)$
III: Weak VARMA(1,1) model (8.2)–(8.5) with $\theta_0 = (0.95, 2, 0)$

From these simulation experiments, we demonstrated that the validity of the different steps of the traditional methodology of Box and Jenkins, identification, estimation and validation, depends on the noise properties. This standard methodology needs however to be adapted to take into account the possible lack of independence of the errors terms. Under nonindependent errors, it appears that the standard tests are generally unreliable while the proposed tests offer satisfactory levels in most cases. Moreover, the error of first kind is well controlled by the modified versions of the tests. We draw the conclusion that the modified versions are preferable to the standard ones. Therefore the modified tests of the present article can be considered as complementary to the above-mentioned available results concerning the statistical analysis of weak VARMA models.

Table 3
Empirical power of standard and modified tests: relative frequencies (in %) of rejection of $H_0 : b_1(2,2) = 0$. The number of replications is $N = 1000$

Model	Length n	Level	Standard Test			Modified Test			
			Wald	LM	LR	Wald	LM	LR$^-$	LR
		$\alpha = 1\%$	7.8	6.3	6.7	8.7	7.7	8.6	9.0
IV	$n = 500$	$\alpha = 5\%$	18.7	17.2	18.3	20.2	18.8	20.2	20.3
		$\alpha = 10\%$	28.9	26.8	27.8	30.2	27.9	30.3	30.3
		$\alpha = 1\%$	33.7	33.6	33.8	34.4	33.7	34.2	34.1
IV	$n = 2,000$	$\alpha = 5\%$	59.7	60.0	59.8	58.7	59.4	58.8	58.7
		$\alpha = 10\%$	72.0	71.5	71.8	71.6	71.1	71.7	71.6
		$\alpha = 1\%$	81.5	81.0	81.6	81.4	81.4	81.3	81.4
IV	$n = 5,000$	$\alpha = 5\%$	94.2	94.2	94.3	94.0	94.3	94.0	94.0
		$\alpha = 10\%$	97.4	97.3	97.3	97.0	96.8	97.0	97.0
		$\alpha = 1\%$	11.0	9.7	10.4	5.3	4.3	5.1	9.0
V	$n = 500$	$\alpha = 5\%$	24.2	22.8	23.4	15.4	14.1	15.5	20.3
		$\alpha = 10\%$	35.4	33.0	33.4	24.3	23.1	24.3	30.3
		$\alpha = 1\%$	39.8	37.6	39.1	21.4	20.9	21.2	34.1
V	$n = 2,000$	$\alpha = 5\%$	56.6	55.8	56.6	42.2	40.8	42.2	58.7
		$\alpha = 10\%$	65.7	64.6	64.9	51.9	51.6	51.7	71.6
		$\alpha = 1\%$	78.3	77.8	77.9	55.2	54.1	55.1	81.4
V	$n = 5,000$	$\alpha = 5\%$	88.3	88.2	88.2	76.9	76.9	76.8	94.0
		$\alpha = 10\%$	92.3	92.2	92.2	85.2	85.6	85.2	97.0
		$\alpha = 1\%$	2.6	1.1	1.9	15.6	14.2	15.5	9.0
VI	$n = 500$	$\alpha = 5\%$	10.5	8.5	9.5	33.7	32.4	33.6	20.3
		$\alpha = 10\%$	19.2	17.3	18.7	45.3	44.9	45.3	30.3
		$\alpha = 1\%$	30.9	29.5	30.6	76.2	74.9	76.0	34.1
VI	$n = 2,000$	$\alpha = 5\%$	66.4	65.5	66.2	90.3	90.1	90.3	58.7
		$\alpha = 10\%$	81.4	80.0	81.2	93.6	93.5	93.6	71.6
		$\alpha = 1\%$	90.4	90.5	90.6	99.5	99.2	99.5	81.4
VI	$n = 5,000$	$\alpha = 5\%$	98.5	98.3	98.5	100.0	100.0	100.0	94.0
		$\alpha = 10\%$	99.7	99.7	99.7	100.0	100.0	100.0	97.0

IV: Strong VARMA(1,1) model (8.2)–(8.3) with $\theta_0 = (0.95, 2, 0.05)$
V: Weak VARMA(1,1) model (8.2)–(8.4) with $\theta_0 = (0.95, 2, 0.05)$
VI: Weak VARMA(1,1) model (8.2)–(8.5) with $\theta_0 = (0.95, 2, 0.05)$

Appendix: Technical proofs

Proof of Proposition 3.1. Because $\theta' = (\mathbf{a}', \mathbf{b}')$, Lemmas A.1 and A.2 below show that

$$\begin{aligned}\frac{\partial e_t(\theta)}{\partial \theta'} &= [V_{t-1}(\theta) : \cdots : V_{t-p}(\theta) : U_{t-1}(\theta) : \cdots : U_{t-q}(\theta)] \\ &= \sum_{h=0}^{\infty} \Big[-\mathbf{A}^*_{h-1}\left(I_{d^2} \otimes e_{t-h}(\theta)\right) : \cdots : -\mathbf{A}^*_{h-p}\left(I_{d^2} \otimes e_{t-h}(\theta)\right) : \\ &\qquad \mathbf{B}^*_{h-1}\left(I_{d^2} \otimes e_{t-h}(\theta)\right) : \cdots : \mathbf{B}^*_{h-q}\left(I_{d^2} \otimes e_{t-h}(\theta)\right)\Big]\end{aligned}$$

$$\begin{aligned} &= \sum_{h=0}^{\infty} \left[-\mathbf{A}^*_{h-1} : \cdots : -\mathbf{A}^*_{h-p} : \mathbf{B}^*_{h-1} : \cdots : \mathbf{B}^*_{h-q} \right] \begin{bmatrix} I_{d^2p} \otimes e_{t-h}(\theta) \\ I_{d^2q} \otimes e_{t-h}(\theta) \end{bmatrix} \\ &= \sum_{h=1}^{\infty} \lambda_h(\theta) \left(I_{d^2(p+q)} \otimes e_{t-h}(\theta) \right). \end{aligned}$$

Hence, at $\theta = \theta_0$ we have

$$\frac{\partial e_t}{\partial \theta'} = \sum_{i \geq 1} \lambda_i \left(I_{d^2(p+q)} \otimes e_{t-i} \right).$$

It thus remains to prove the following two Lemmas. □

Lemma A.1. *We have*

$$\frac{\partial e_t(\theta)}{\partial \mathbf{a}'} = -\sum_{h=0}^{\infty} \mathbf{A}^*_{\theta,h} \left(I_{d^2p} \otimes e_{t-h}(\theta) \right),$$

where $\mathbf{A}^*_{\theta,h} = [\mathbf{A}^*_{h-1} : \mathbf{A}^*_{h-2} : \cdots : \mathbf{A}^*_{h-p}]$ *is a* $d \times d^3p$ *matrix.*

Proof of Lemma A.1. Differentiating the two terms of the following equality

$$\mathbf{A}_\theta(L) X_t = \mathbf{B}_\theta(L) e_t(\theta),$$

with respect to the AR coefficients, we obtain

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{a}_{ij,\ell}} &= -\mathbf{B}_\theta^{-1}(L) E_{ij} X_{t-\ell} = -\mathbf{B}_\theta^{-1}(L) E_{ij} \mathbf{A}_\theta^{-1}(L) \mathbf{B}_\theta(L) e_{t-\ell}(\theta) \\ &= -M_{ij}(L) e_{t-\ell}(\theta), \qquad \ell = 1, \dots, p, \end{aligned}$$

where $E_{ij} = \partial \mathbf{A}_\ell / \partial \mathbf{a}_{ij,\ell}$ is the $d \times d$ matrix with 1 at position (i,j) and 0 elsewhere and

$$e_t(\theta) = \mathbf{B}_\theta^{-1}(L) \mathbf{A}_\theta(L) X_t.$$

Then we have

$$\frac{\partial e_t(\theta)}{\partial \mathbf{a}_{ij,\ell}} = -\sum_{h=0}^{\infty} \mathbf{A}^*_{ij,h} e_{t-\ell-h}(\theta). \tag{A.1}$$

Hence, for any $\mathbf{a}_\ell$ writing the multivariate noise derivatives

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{a}'_\ell} &= -\underbrace{\left[M_{11}(L) e_{t-\ell}(\theta) : M_{21}(L) e_{t-\ell}(\theta) : \cdots : M_{dd}(L) e_{t-\ell}(\theta) \right]}_{d \times d^2} \\ &= -\left[M_{11}(L) : M_{21}(L) : \cdots : M_{dd}(L) \right] \left(I_{d^2} \otimes e_{t-\ell}(\theta) \right) \\ &= -M(L) \left(I_{d^2} \otimes e_{t-\ell}(\theta) \right) = -M(L) \mathbf{e}_{t-\ell}(\theta) = V_{t-\ell}(\theta), \end{aligned}$$

where $M(L) = [M_{11}(L) : M_{21}(L) : \cdots : M_{dd}(L)]$, $\mathbf{e}_{t-\ell}(\theta) = I_{d^2} \otimes e_{t-\ell}(\theta)$ and $V_{t-\ell}(\theta)$ are respectively the $d \times d^3$, $d^3 \times d^2$ and $d \times d^2$ matrices. Hence, we have

$$\frac{\partial e_t(\theta)}{\partial \mathbf{a}'_\ell} = -\sum_{h=0}^{\infty} \mathbf{A}^*_h \mathbf{e}_{t-\ell-h}(\theta) = -\sum_{k-\ell=0}^{\infty} \mathbf{A}^*_{k-\ell} \mathbf{e}_{t-k}(\theta) = -\sum_{k=\ell}^{\infty} \mathbf{A}^*_{k-\ell} \mathbf{e}_{t-k}(\theta)$$

$$= \quad -\sum_{k=0}^{\infty}\mathbf{A}^*_{k-\ell}\mathbf{e}_{t-k}(\theta) = V_{t-\ell}(\theta),$$

with $\mathbf{A}^*_{k-\ell} = 0$ when $k < \ell$. With these notations, we obtain

$$\begin{aligned}
\frac{\partial e_t(\theta)}{\partial \mathbf{a}'} &= -\sum_{k=0}^{\infty}\underbrace{\left[\mathbf{A}^*_{k-1}\mathbf{e}_{t-k}(\theta) : \mathbf{A}^*_{k-2}\mathbf{e}_{t-k}(\theta) : \cdots : \mathbf{A}^*_{k-p}\mathbf{e}_{t-k}(\theta)\right]}_{d\times d^2p} \\
&= -\sum_{k=0}^{\infty}\left[\mathbf{A}^*_{k-1} : \mathbf{A}^*_{k-2} : \cdots : \mathbf{A}^*_{k-p}\right](I_p \otimes \mathbf{e}_{t-k}(\theta)) \\
&= -\sum_{k=0}^{\infty}\mathbf{A}^*_{\theta,k}\left(I_{d^2p} \otimes e_{t-k}(\theta)\right),
\end{aligned}$$

where $\mathbf{A}^*_{\theta,k} = [\mathbf{A}^*_{k-1} : \mathbf{A}^*_{k-2} : \cdots : \mathbf{A}^*_{k-p}]$ is the $d \times d^3q$ matrix. The conclusion follows. □

Lemma A.2. *We have*

$$\frac{\partial e_t(\theta)}{\partial \mathbf{b}'} = \sum_{h=0}^{\infty}\mathbf{B}^*_{\theta,h}\left(I_{d^2q} \otimes e_{t-h}(\theta)\right),$$

where $\mathbf{B}^*_{\theta,h} = [\mathbf{B}^*_{h-1} : \mathbf{B}^*_{h-2} : \cdots : \mathbf{B}^*_{h-q}]$ *is a* $d \times d^3q$ *matrix.*

Proof of Lemma A.2. The proof is similar to that given in Lemma A.1. However, we will give the derivatives which are different to that in the previous Lemma. Differentiating the two terms of the following equality

$$\mathbf{A}_\theta(L)X_t = \mathbf{B}_\theta(L)e_t(\theta),$$

with respect to the MA coefficients, we obtain

$$\begin{aligned}
\frac{\partial e_t(\theta)}{\partial \mathbf{b}_{ij,\ell'}} &= \mathbf{B}_\theta^{-1}(L)E_{ij}\mathbf{B}_\theta^{-1}(L)\mathbf{A}_\theta(L)X_{t-\ell'}(\theta) = \mathbf{B}_\theta^{-1}(L)E_{ij}e_{t-\ell'}(\theta) \\
&= N_{ij}(L)e_{t-\ell'}(\theta), \qquad \ell' = 1,\dots,q,
\end{aligned}$$

where $E_{ij} = \partial\mathbf{B}_{\ell'}/\partial\mathbf{b}_{ij,\ell'}$ is the $d \times d$ matrix with 1 at position (i,j) and 0 elsewhere. We then have

$$\frac{\partial e_t(\theta)}{\partial \mathbf{b}_{ij,\ell'}} = \sum_{h=0}^{\infty}\mathbf{B}^*_{ij,h}e_{t-\ell'-h}(\theta). \tag{A.2}$$

Similarly to Lemma A.2, we have

$$\begin{aligned}
\frac{\partial e_t(\theta)}{\partial \mathbf{b}'_{\ell'}} &= \underbrace{\left[N_{11}(L)e_{t-\ell'}(\theta) : N_{21}(L)e_{t-\ell'}(\theta) : \cdots : N_{dd}(L)e_{t-\ell'}(\theta)\right]}_{d\times d^2} \\
&= N(L)\left(I_{d^2} \otimes e_{t-\ell'}(\theta)\right) = N(L)\mathbf{e}_{t-\ell'}(\theta) = U_{t-\ell'}(\theta),
\end{aligned}$$

where $N(L) = [N_{11}(L) : N_{21}(L) : \cdots : N_{dd}(L)]$ and $U_{t-\ell'}$ are respectively the $d \times d^3$ and $d \times d^2$ matrices. Then, we have

$$\frac{\partial e_t(\theta)}{\partial \mathbf{b}'_{\ell'}} = \sum_{h=0}^{\infty} \mathbf{B}_h^* \mathbf{e}_{t-\ell'-h}(\theta) = \sum_{k=0}^{\infty} \mathbf{B}_{k-\ell'}^* \mathbf{e}_{t-k}(\theta) = U_{t-\ell'}(\theta),$$

where $\mathbf{B}_{k-\ell'}^* = 0$ when $k < \ell'$. With these notations, we obtain

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{b}'} &= \sum_{k=0}^{\infty} \underbrace{\left[\mathbf{B}_{k-1}^* \mathbf{e}_{t-k}(\theta) : \mathbf{B}_{k-2}^* \mathbf{e}_{t-k}(\theta) : \cdots : \mathbf{B}_{k-q}^* \mathbf{e}_{t-k}(\theta)\right]}_{d \times d^2 q} \\ &= \sum_{k=0}^{\infty} \mathbf{B}_{\theta,k}^* \left(I_{d^2 q} \otimes e_{t-k}(\theta)\right), \end{aligned}$$

where $\mathbf{B}_{\theta,k}^* = [\mathbf{B}_{k-1}^* : \mathbf{B}_{k-2}^* : \cdots : \mathbf{B}_{k-q}^*]$ is the $d \times d^3 q$ matrix. The conclusion follows. □

Proof of Proposition 4.1. Let

$$\begin{aligned} \tilde{\ell}_n(\theta, \Sigma_e) &= -\frac{2}{n} \log \tilde{\mathrm{L}}_n(\theta, \Sigma_e) \\ &= \frac{1}{n} \sum_{t=1}^{n} \left\{ d \log(2\pi) + \log\det \Sigma_e + \tilde{e}'_t(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta) \right\}. \end{aligned}$$

In [9], it is shown that $\ell_n(\theta, \Sigma_e) = \tilde{\ell}_n(\theta, \Sigma_e) + o(1)$ a.s., where

$$\ell_n(\theta, \Sigma_e) = \frac{1}{n} \sum_{t=1}^{n} \left\{ d \log(2\pi) + \log\det \Sigma_e + e'_t(\theta) \Sigma_e^{-1} e_t(\theta) \right\}.$$

It is also shown uniformly in $\theta \in \Theta$ that

$$\frac{\partial \ell_n(\theta, \Sigma_e)}{\partial \theta} = \frac{\partial \tilde{\ell}_n(\theta, \Sigma_e)}{\partial \theta} + o(1) \quad a.s.$$

The same equality holds for the second-order derivatives of $\tilde{\ell}_n$. We thus have

$$J = \lim_{n\to\infty} \mathbf{J}_n, \quad \text{where} \quad \mathbf{J}_n = \frac{\partial^2 \ell_n(\theta_0, \Sigma_{e0})}{\partial \theta \partial \theta'}.$$

Using well-known results on matrix derivatives, we have

$$\frac{\partial \ell_n(\theta_0, \Sigma_{e0})}{\partial \theta} = \frac{2}{n} \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \theta} e'_t(\theta_0) \right\} \Sigma_{e0}^{-1} e_t(\theta_0), \tag{A.3}$$

where

$$\frac{\partial}{\partial \theta'} e_t(\theta_0) = \left(\frac{\partial}{\partial \theta_1} e_t(\theta_0), \ldots, \frac{\partial}{\partial \theta_{k_0}} e_t(\theta_0) \right).$$

In view of (A.3), we have

$$\begin{aligned}
\mathbf{J}_n &= \frac{2}{n}\sum_{t=1}^{n}\left(\frac{\partial e_t'(\theta_0)}{\partial\theta}\Sigma_{e0}^{-1}\frac{\partial e_t(\theta_0)}{\partial\theta'}+\frac{\partial^2 e_t'(\theta_0)}{\partial\theta\partial\theta'}\Sigma_{e0}^{-1}e_t(\theta_0)\right)\\
&\to 2\mathbb{E}\left\{\frac{\partial^2 e_t'(\theta_0)}{\partial\theta\partial\theta'}\right\}\Sigma_{e0}^{-1}e_t+2\mathbb{E}\left\{\frac{\partial}{\partial\theta}e_t'(\theta_0)\right\}\Sigma_{e0}^{-1}\left\{\frac{\partial}{\partial\theta'}e_t(\theta_0)\right\},\quad a.s.
\end{aligned}$$

by the ergodic theorem. Using the orthogonality between e_t and any linear combination of the past values of e_t, we have $2\mathbb{E}\{\partial^2 e_t'(\theta_0)/\partial\theta\partial\theta'\}\Sigma_{e0}^{-1}e_t=0$. We then have

$$J=2\mathbb{E}\left\{\frac{\partial}{\partial\theta}e_t'(\theta_0)\Sigma_{e0}^{-1}\frac{\partial}{\partial\theta'}e_t(\theta_0)\right\}.$$

Using $\mathrm{vec}(ABC)=(C'\otimes A)\,\mathrm{vec}(B)$ with $C=I_{d^2(p+q)}$ and in view of Proposition 3.1, we obtain

$$\begin{aligned}
\mathrm{vec}\,J &= 2\mathbb{E}\left\{\frac{\partial e_t'(\theta_0)}{\partial\theta}\otimes\frac{\partial e_t'(\theta_0)}{\partial\theta}\right\}\mathrm{vec}\,\Sigma_{e0}^{-1}\\
&= 2\sum_{i\geq1}\mathbb{E}\left\{\left(I_{d^2(p+q)}\otimes e_{t-i}'\right)\lambda_i'\right\}\otimes\left\{\left(I_{d^2(p+q)}\otimes e_{t-i}'\right)\lambda_i'\right\}\mathrm{vec}\,\Sigma_{e0}^{-1}.
\end{aligned}$$

Using also $AC\otimes BD=(A\otimes B)(C\otimes D)$, we have

$$\begin{aligned}
\mathrm{vec}\,J &= 2\sum_{i\geq1}\mathbb{E}\left\{\left(I_{d^2(p+q)}\otimes e_t'\right)\otimes\left(I_{d^2(p+q)}\otimes e_t'\right)\right\}\left\{\lambda_i'\otimes\lambda_i'\right\}\mathrm{vec}\,\Sigma_{e0}^{-1}\\
&= 2\sum_{i\geq1}\mathbb{E}\left\{\left(I_{d^2(p+q)}\otimes e_t'\right)^{\otimes2}\right\}\lambda_i^{'\otimes2}\,\mathrm{vec}\,\Sigma_{e0}^{-1}=2\sum_{i\geq1}\mathcal{M}\lambda_i^{'\otimes2}\,\mathrm{vec}\,\Sigma_{e0}^{-1}.
\end{aligned}$$

The conclusion is complete. □

Proof of Proposition 4.2. In view of (4.2), let

$$\Upsilon_t=\frac{\partial}{\partial\theta}\mathcal{L}_t(\theta_0,\Sigma_{e0})=\left(\frac{\partial}{\partial\theta_1}\mathcal{L}_t(\theta,\Sigma_{e0}),\dots,\frac{\partial}{\partial\theta_{k_0}}\mathcal{L}_t(\theta,\Sigma_{e0})\right)'_{\theta=\theta_0}$$

where

$$\mathcal{L}_t(\theta,\Sigma_{e0})=\log\det\Sigma_{e0}+e_t'(\theta)\Sigma_{e0}^{-1}e_t(\theta).$$

We have

$$\Upsilon_t=2\frac{\partial e_t'(\theta_0)}{\partial\theta}\Sigma_{e0}^{-1}e_t(\theta_0)=2\left\{e_t'(\theta_0)\otimes\frac{\partial e_t'(\theta_0)}{\partial\theta}\right\}\mathrm{vec}\,\Sigma_{e0}^{-1}.$$

In view of (4.1), we have

$$I=\sum_{h=-\infty}^{+\infty}\mathrm{Cov}\left(2\left\{\frac{\partial}{\partial\theta}e_t'(\theta_0)\right\}\Sigma_{e0}^{-1}e_t,2\left\{\frac{\partial}{\partial\theta}e_{t-h}'(\theta_0)\right\}\Sigma_{e0}^{-1}e_{t-h}\right)$$

$$= \sum_{h=-\infty}^{+\infty} \mathrm{Cov}\left(2\left\{e'_t \otimes \frac{\partial e'_t}{\partial\theta}\right\} \mathrm{vec}\,\Sigma_{e0}^{-1}, 2\left\{e'_{t-h} \otimes \frac{\partial e'_{t-h}}{\partial\theta}\right\} \mathrm{vec}\,\Sigma_{e0}^{-1}\right)$$
$$= 4\sum_{h=-\infty}^{+\infty} \mathbb{E}\left(\left\{e'_t \otimes \frac{\partial e'_t}{\partial\theta}\right\} \mathrm{vec}\,\Sigma_{e0}^{-1}\left\{\mathrm{vec}\,\Sigma_{e0}^{-1}\right\}'\left\{e'_{t-h} \otimes \frac{\partial e'_{t-h}}{\partial\theta}\right\}'\right).$$

Using the elementary relation $\mathrm{vec}(ABC) = (C' \otimes A)\,\mathrm{vec}\,B$, we have

$$\mathrm{vec}\,I = 4\sum_{h=-\infty}^{+\infty} \mathbb{E}\left(\left\{e'_{t-h} \otimes \frac{\partial e'_{t-h}}{\partial\theta}\right\} \otimes \left\{e'_t \otimes \frac{\partial e'_t}{\partial\theta}\right\}\right) \mathrm{vec}\left(\mathrm{vec}\,\Sigma_{e0}^{-1}\left\{\mathrm{vec}\,\Sigma_{e0}^{-1}\right\}'\right).$$

By Proposition 3.1, we obtain

$$\begin{aligned}\mathrm{vec}\,I &= 4\sum_{h=-\infty}^{+\infty}\sum_{i,j=1}^{+\infty} \mathbb{E}\left(\left\{e'_{t-h} \otimes \left(I_{d^2(p+q)} \otimes e'_{t-j-h}\right)\lambda'_j\right\}\right.\\ &\quad \left.\otimes\left\{e'_t \otimes \left(I_{d^2(p+q)} \otimes e'_{t-i}\right)\lambda'_i\right\}\right) \mathrm{vec}\left(\mathrm{vec}\,\Sigma_{e0}^{-1}\left\{\mathrm{vec}\,\Sigma_{e0}^{-1}\right\}'\right).\end{aligned}$$

Using $AC \otimes BD = (A \otimes B)(C \otimes D)$, we have

$$\begin{aligned}\mathrm{vec}\,I &= 4\sum_{h=-\infty}^{+\infty}\sum_{i,j=1}^{+\infty} \mathbb{E}\left(\left\{e'_{t-h} \otimes \left(I_{d^2(p+q)} \otimes e'_{t-j-h}\right)\right\}\left\{I_d \otimes \lambda'_j\right\}\right.\\ &\quad \left.\otimes\left\{e'_t \otimes \left(I_{d^2(p+q)} \otimes e'_{t-i}\right)\right\}\left\{I_d \otimes \lambda'_i\right\}\right) \mathrm{vec}\left(\mathrm{vec}\,\Sigma_{e0}^{-1}\left\{\mathrm{vec}\,\Sigma_{e0}^{-1}\right\}'\right).\end{aligned}$$

Using also $AC \otimes BD = (A \otimes B)(C \otimes D)$, we have

$$\begin{aligned}\mathrm{vec}\,I &= 4\sum_{h=-\infty}^{+\infty}\sum_{i,j=1}^{+\infty} \mathbb{E}\left(\left\{e'_{t-h} \otimes \left(I_{d^2(p+q)} \otimes e'_{t-j-h}\right)\right\} \otimes \left\{e'_t \otimes \left(I_{d^2(p+q)} \otimes e'_{t-i}\right)\right\}\right)\\ &\quad \left(\left\{I_d \otimes \lambda'_j\right\} \otimes \left\{I_d \otimes \lambda'_i\right\}\right) \mathrm{vec}\left(\mathrm{vec}\,\Sigma_{e0}^{-1}\left\{\mathrm{vec}\,\Sigma_{e0}^{-1}\right\}'\right)\\ &= 4\sum_{i,j=1}^{+\infty} \mathbf{\Gamma}(i,j)\left(\left\{I_d \otimes \lambda'_j\right\} \otimes \left\{I_d \otimes \lambda'_i\right\}\right) \mathrm{vec}\left(\mathrm{vec}\,\Sigma_{e0}^{-1}\left\{\mathrm{vec}\,\Sigma_{e0}^{-1}\right\}'\right),\end{aligned}$$

where

$$\mathbf{\Gamma}(i,j) = \sum_{h=-\infty}^{+\infty} \mathbb{E}\left(\left\{e'_{t-h} \otimes \left(I_{d^2(p+q)} \otimes e'_{t-j-h}\right)\right\} \otimes \left\{e'_t \otimes \left(I_{d^2(p+q)} \otimes e'_{t-i}\right)\right\}\right).$$

The conclusion is complete. □

Proof of Remark 4.1. For $d = 1$, we have

$$\mathcal{M} := \mathbb{E}\left\{\left(I_{(p+q)} \times e_t\right)^{\otimes 2}\right\} = \sigma^2 I_{(p+q)^2},$$

where σ^2 is the variance of the univariate process. We also have

$$\begin{aligned}\mathbf{\Gamma}(i,j) &= \sum_{h=-\infty}^{+\infty} \mathbb{E}\left(\left\{e_{t-h}e_{t-j-h}I_{(p+q)}\right\} \otimes \left\{e_t e_{t-i} I_{(p+q)}\right\}\right) \\ &= \sum_{h=-\infty}^{+\infty} \mathbb{E}\left(e_t e_{t-i} e_{t-h} e_{t-j-h}\right) I_{(p+q)^2} = \gamma(i,j) I_{(p+q)^2}. \qquad \text{(A.4)}\end{aligned}$$

In view of Proposition 4.1, we have

$$\operatorname{vec} J = 2\sum_{i\geq 1} \mathcal{M}\left\{\lambda_i' \otimes \lambda_i'\right\}\sigma^{-2}.$$

Replacing $\mathcal{M}$ by $\sigma^2 I_{(p+q)^2}$ in $\operatorname{vec} J$, we have

$$\operatorname{vec} J = 2\sum_{i\geq 1} \left\{\lambda_i \otimes \lambda_i\right\}'.$$

Using (A.4) and in view of Proposition 4.2, we have

$$\operatorname{vec} I = \frac{4}{\sigma^4}\sum_{i,j=1}^{+\infty} \mathbf{\Gamma}(i,j)\left\{\lambda_j' \otimes \lambda_i'\right\} = \frac{4}{\sigma^4}\sum_{i,j=1}^{+\infty} \gamma(i,j)\left\{\lambda_j \otimes \lambda_i\right\}'.$$

The conclusion is complete. □

Proof of Theorem 5.1. For any multiplicative norm, we have

$$\begin{aligned}\left\|\operatorname{vec} J - \operatorname{vec} \hat{J}_n\right\| \leq\ & 2\sum_{i\geq 1}\left\{\left\|\mathcal{M} - \hat{\mathcal{M}}_n\right\| \left\|\lambda_i'^{\otimes 2}\right\| \left\|\operatorname{vec}\left(\Sigma_{e0}^{-1}\right)\right\|\right. \\ & + \left\|\hat{\mathcal{M}}_n\right\| \left\|\lambda_i'^{\otimes 2} - \hat{\lambda}_i'^{\otimes 2}\right\| \left\|\operatorname{vec}\Sigma_{e0}^{-1}\right\| \\ & \left. + \left\|\hat{\mathcal{M}}_n\right\| \left\|\hat{\lambda}_i'^{\otimes 2}\right\| \left\|\operatorname{vec}\left(\hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1}\right)\right\|\right\}.\end{aligned}$$

The proof will thus follow from Lemmas A.3, A.4 and A.6 below. □

Lemma A.3. *Under Assumptions* **A1**–**A8**, *we have*

$$\left\|\operatorname{vec}\left\{\hat{\lambda}_i - \lambda_i(\theta_0)\right\}\right\| \leq K\rho^i \times o_{a.s.}(1) \text{ a.s. as } n \to \infty,$$

where ρ is a constant belonging to $[0,1[$, and $K > 0$.

Proof of Lemma A.3. Boubacar Maïnassara and Francq [9] have showed the strong consistency of $\hat{\theta}_n$ ($\hat{\theta}_n \to \theta_0$ *a.s.* as $n \to \infty$), which entails

$$\left\|\hat{\theta}_n - \theta_0\right\| = o_{a.s}(1). \qquad \text{(A.5)}$$

We have $\mathbf{A}^*_{ij,h} = O(\rho^h)$ and $\mathbf{B}^*_{ij,h} = O(\rho^h)$ uniformly in $\theta \in \Theta$ for some $\rho \in [0,1[$. In view of (3.1), we thus have $\sup_{\theta\in\Theta} \|\lambda_h(\theta)\| \le K\rho^h$. Similarly for any $m \in \{1,\dots,k_0\}$, we have

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\lambda_h(\theta)}{\partial\theta_m}\right\| \le K\rho^h. \tag{A.6}$$

Using a Taylor expansion of $\operatorname{vec}\hat{\lambda}_i$ about θ_0, we obtain

$$\operatorname{vec}\hat{\lambda}_i \;=\; \operatorname{vec}\lambda_i + \frac{\partial \operatorname{vec}\lambda_i(\theta_n^*)}{\partial\theta'}(\hat{\theta}_n - \theta_0),$$

where θ_n^* is between $\hat{\theta}_n$ and θ_0. For any multiplicative norm, we have

$$\left\|\operatorname{vec}(\hat{\lambda}_i - \lambda_i)\right\| \;\le\; \left\|\frac{\partial \operatorname{vec}\lambda_i(\theta_n^*)}{\partial\theta'}\right\| \left\|\hat{\theta}_n - \theta_0\right\|.$$

In view of (A.5) and (A.6), the proof is complete. □

Lemma A.4. *Under Assumptions* **A1**–**A8***, we have*

$$\hat{\mathcal{M}}_n \to \mathcal{M} \text{ a.s. as } n \to \infty.$$

Proof of Lemma A.4. We have

$$e_t(\theta) = X_t - \sum_{i=1}^{p} A_0^{-1}A_iX_{t-i} + \sum_{i=1}^{q} A_0^{-1}B_iB_0^{-1}A_0e_{t-i}(\theta) \qquad \forall t \in \mathbb{Z}. \tag{A.7}$$

For any $\theta \in \Theta$, let

$$\mathcal{M}_n(\theta) := \frac{1}{n}\sum_{t=1}^{n}\left\{\left(I_{d^2(p+q)} \otimes e_t'(\theta)\right)^{\otimes 2}\right\} \text{ and } \mathcal{M}(\theta) := \mathbb{E}\left\{\left(I_{d^2(p+q)} \otimes e_t'(\theta)\right)^{\otimes 2}\right\}.$$

Now the ergodic theorem shows that almost surely

$$\mathcal{M}_n(\theta) \to \mathcal{M}(\theta).$$

In view of (A.7), using **A2** and the compactness of Θ, we have

$$e_t(\theta) = X_t + \sum_{i=1}^{\infty} C_i(\theta)X_{t-i}, \qquad \sup_{\theta\in\Theta}\|C_i(\theta)\| \le K\rho^i.$$

We thus have

$$\mathbb{E}\sup_{\theta\in\Theta}\|e_t(\theta)\|^2 < \infty, \tag{A.8}$$

by Assumption **A7**. Now, we will consider the norm defined by: $\|Z\|_2 = \sqrt{\mathbb{E}\|Z\|^2}$, where Z is a d_1 random vector. In view of Proposition 3.1, (A.8) and $\sup_{\theta\in\Theta}\|\lambda_h(\theta)\| \le K\rho^h$, we have

$$\left\|\sup_{\theta\in\Theta}\frac{\partial e_t(\theta)}{\partial\theta'}\right\|_2 \le \sum_{i\ge 1}\sup_{\theta\in\Theta}\|\lambda_i(\theta)\| \times \left\|\sup_{\theta\in\Theta}\left\{I_{d^2(p+q)} \otimes e_t(\theta)\right\}\right\|_2 < \infty. \tag{A.9}$$

Let $e_t = (e_{1t}, \dots, e_{dt})'$. The non zero components of the vector $\operatorname{vec} \mathcal{M}_n(\theta)$ are of the form $n^{-1} \sum_{t=1}^n e_{it}(\theta) e_{jt}(\theta)$, for $(i,j) \in \{1, \dots, d\}^2$. We deduce that the elements of the matrix $\partial \operatorname{vec} \mathcal{M}_n(\theta)/\partial\theta'$ are linear combinations of

$$\frac{2}{n} \sum_{t=1}^n e_{it}(\theta) \frac{\partial e_{jt}(\theta)}{\partial \theta}.$$

By the Cauchy-Schwartz inequality we have

$$\begin{aligned}\left\| \frac{2}{n} \sum_{t=1}^n \sup_{\theta\in\Theta} \left\{ e_{it}(\theta) \frac{\partial e_{jt}(\theta)}{\partial \theta} \right\} \right\| &\leq \sqrt{\frac{2}{n} \sum_{t=1}^n \sup_{\theta\in\Theta} \{e_{it}^2(\theta)\} \times \frac{2}{n} \sum_{t=1}^n \left\| \sup_{\theta\in\Theta} \frac{\partial e_{jt}(\theta)}{\partial\theta} \right\|^2} \\ &\leq 2 \sqrt{\frac{1}{n} \sum_{t=1}^n \sup_{\theta\in\Theta} \|e_t(\theta)\|^2 \times \frac{1}{n} \sum_{t=1}^n \left\| \sup_{\theta\in\Theta} \frac{\partial e_t(\theta)}{\partial\theta'} \right\|^2}.\end{aligned}$$

The ergodic theorem shows that almost surely

$$\lim_{n\to\infty} \frac{2}{n} \sum_{t=1}^n \left\| \sup_{\theta\in\Theta} \left\{ e_{it}(\theta) \frac{\partial e_{jt}(\theta)}{\partial\theta} \right\} \right\| \leq 2 \sqrt{\mathbb{E} \sup_{\theta\in\Theta} \|e_t(\theta)\|^2 \times \mathbb{E} \left\| \sup_{\theta\in\Theta} \frac{\partial e_t(\theta)}{\partial\theta'} \right\|^2}.$$

Now using (A.8) and (A.9), the right-hand side of the inequality is bounded. We then deduce that

$$\sup_{\theta\in\Theta} \left\| \frac{\partial \operatorname{vec} \mathcal{M}_n(\theta)}{\partial\theta'} \right\| = O_{a.s}(1). \tag{A.10}$$

A Taylor expansion of $\operatorname{vec} \hat{\mathcal{M}}_n$ about θ_0 gives

$$\operatorname{vec} \hat{\mathcal{M}}_n \quad = \quad \operatorname{vec} \mathcal{M}_n + \frac{\partial \operatorname{vec} \mathcal{M}_n(\theta_n^*)}{\partial\theta'} (\hat{\theta}_n - \theta_0),$$

where θ_n^* is between $\hat{\theta}_n$ and θ_0. Using the strong consistency of $\hat{\theta}_n$ and (A.10), it is easily seen that

$$\operatorname{vec} \hat{\mathcal{M}}_n = \operatorname{vec} \mathcal{M}_n(\hat{\theta}_n) \to \operatorname{vec} \mathcal{M}(\theta_0) = \operatorname{vec} \mathcal{M}, \text{ a.s.}$$

The conclusion is complete. □

Lemma A.5. *Under Assumptions* **A1**–**A8**, *we have*

$$\hat{\Sigma}_{e0} \to \Sigma_{e0} \ \textit{a.s., as } n \to \infty.$$

Proof of Lemma A.5. We have $\hat{\Sigma}_{e0} = \Sigma_n(\hat{\theta}_n)$ with $\Sigma_n(\theta) = n^{-1} \sum_{t=1}^n e_t(\theta) e_t'(\theta)$. By the ergodic theorem

$$\Sigma_n(\theta) \to \Sigma_e(\theta) := \mathbb{E} e_t(\theta) e_t'(\theta), \text{ a.s.}$$

Using the elementary relation $\mathrm{vec}(aa') = a \otimes a$, where a is a vector, we have $\mathrm{vec}\,\hat{\Sigma}_{e0} = n^{-1}\sum_{t=1}^{n} e_t(\hat{\theta}_n) \otimes e_t(\hat{\theta}_n)$ and $\mathrm{vec}\,\Sigma_n(\theta_0) = n^{-1}\sum_{t=1}^{n} e_t(\theta_0) \otimes e_t(\theta_0)$. Using a Taylor expansion of $\mathrm{vec}\,\hat{\Sigma}_{e0}$ around θ_0 and (2.2), we obtain

$$\mathrm{vec}\,\hat{\Sigma}_{e0} \quad = \quad \mathrm{vec}\,\Sigma_n(\theta_0) + \frac{1}{n}\sum_{t=1}^{n}\left\{e_t \otimes \frac{\partial e_t}{\partial \theta'} + \frac{\partial e_t}{\partial \theta'} \otimes e_t\right\}(\hat{\theta}_n - \theta_0) + O_P\left(\frac{1}{n}\right).$$

Using the strong consistency of $\hat{\theta}_n$,

$$\mathbb{E}\sup_{\theta\in\Theta}\|e_t(\theta)\|^2 < \infty \text{ and } \left\|\sup_{\theta\in\Theta}\frac{\partial e_t(\theta)}{\partial \theta'}\right\|_2 < \infty,$$

it is easily seen that

$$\mathrm{vec}\,\hat{\Sigma}_{e0} \to \mathrm{vec}\,\Sigma_e(\theta_0) = \mathrm{vec}\,\Sigma_{e0}, \text{ a.s.}$$

The conclusion is complete. □

Lemma A.6. *Under Assumptions* **A1**–**A8**, *we have*

$$\hat{\Sigma}_{e0}^{-1} \to \Sigma_{e0}^{-1} \;\; a.s. \;\; as n \to \infty.$$

Proof of Lemma A.6. For any multiplicative norm, we have

$$\left\|\hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1}\right\| \quad = \quad \left\|-\hat{\Sigma}_{e0}^{-1}\left(\hat{\Sigma}_{e0} - \Sigma_{e0}\right)\Sigma_{e0}^{-1}\right\| \le \left\|\hat{\Sigma}_{e0}^{-1}\right\| \left\|\hat{\Sigma}_{e0} - \Sigma_{e0}\right\| \left\|\Sigma_{e0}^{-1}\right\|.$$

In view of Lemma A.5 and $\|\Sigma_{e0}^{-1}\| < \infty$ (because the matrix Σ_{e0} is nonsingular), we have

$$\left\|\hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1}\right\| \to 0 \quad a.s.$$

The conclusion is complete. □

Proof of Theorem 5.2. Let the matrices

$$\hat{\Lambda}_{ij} = \left(\left\{I_d \otimes \hat{\lambda}_i'\right\} \otimes \left\{I_d \otimes \hat{\lambda}_j'\right\}\right), \quad \Lambda_{ij} = \left(\{I_d \otimes \lambda_i'\} \otimes \{I_d \otimes \lambda_j'\}\right),$$

$$\hat{\Delta}_{e0} = \left(\mathrm{vec}\,\hat{\Sigma}_{e0}^{-1}\left\{\mathrm{vec}\,\hat{\Sigma}_{e0}^{-1}\right\}'\right) \quad \text{and} \quad \Delta_{e0} = \left(\mathrm{vec}\,\Sigma_{e0}^{-1}\left\{\mathrm{vec}\,\Sigma_{e0}^{-1}\right\}'\right).$$

For any multiplicative norm, we have

$$\begin{aligned}\left\|\mathrm{vec}\,I - \mathrm{vec}\,\hat{I}_n\right\| \quad \le \quad 4\sum_{i,j\ge 1}\Big\{&\left\|\mathbf{\Gamma}(i,j) - \hat{\mathbf{\Gamma}}_n(i,j)\right\| \|\Lambda_{ij}\| \|\mathrm{vec}\,\Delta_{e0}\| \\ &+ \left\|\hat{\mathbf{\Gamma}}_n(i,j)\right\| \left\|\Lambda_{ij} - \hat{\Lambda}_{ij}\right\| \|\mathrm{vec}\,\Delta_{e0}\| \\ &+ \left\|\hat{\mathbf{\Gamma}}_n(i,j)\right\| \left\|\hat{\Lambda}_{ij}\right\| \left\|\mathrm{vec}\left(\Delta_{e0} - \hat{\Delta}_{e0}\right)\right\|\Big\}.\end{aligned}$$

Lemma A.3 and $\|\lambda_i\| = O(\rho^i)$ entail

$$\begin{aligned}\left\|\operatorname{vec}\hat{\Lambda}_{ij} - \operatorname{vec}\Lambda_{ij}\right\| &\leq \left\|\operatorname{vec}\left\{I_d \otimes (\hat{\lambda}_i - \lambda_i) \otimes (I_d \otimes \hat{\lambda}_j)\right\}\right\| \\ &\quad + \left\|\operatorname{vec}\left\{(I_d \otimes \lambda_i) \otimes \left(I_d \otimes (\hat{\lambda}_j - \lambda_j)\right)\right\}\right\| \\ &\leq K\rho^{i+j} \times o_{a.s}(1).\end{aligned}$$

We also have $\|\Lambda_{ij}\| = O(\rho^{i+j})$. In view of Lemma A.6 and $\|\Sigma_{e0}^{-1}\| < \infty$, we have

$$\begin{aligned}\left\|\hat{\Delta}_{e0} - \Delta_{e0}\right\| &\leq \left\|\operatorname{vec}\left(\hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1}\right)\right\| \left\|\left\{\operatorname{vec}\hat{\Sigma}_{e0}^{-1}\right\}'\right\| \\ &\quad + \left\|\operatorname{vec}\Sigma_{e0}^{-1}\right\| \left\|\left\{\operatorname{vec}\left(\hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1}\right)\right\}'\right\| \to 0 \quad a.s.\end{aligned}$$

We consider the white noise “empirical” autocovariances

$$\Gamma_e(h) = \frac{1}{n}\sum_{t=h+1}^{n} e_t e'_{t-h}, \quad \text{for} \quad 0 \leq h < n.$$

For $k, k', m, m' = 1, \dots, \infty$, let

$$\Gamma(k, k') = \sum_{h=-\infty}^{\infty} \mathbb{E}\left(\{e_{t-k} \otimes e_t\}\{e_{t-h-k'} \otimes e_{t-h}\}'\right).$$

In the univariate case Francq, Roy and Zakoïan [21] have showed (see the proofs of their Lemmas A.1 and A.3) that $\sup_{\ell,\ell'>0} |\Gamma(\ell, \ell')| < \infty$. This can be directly extended in the multivariate case. The non zero elements of $\mathbf{\Gamma}(i, j)$ are of the form

$$\sum_{h=-\infty}^{\infty} \mathbb{E}\left(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j}\right), \quad \text{for} \quad (i_1, i_2, i_3, i_4) \in \{1, \dots, d\}^4.$$

We thus have

$$\sup_{i,j\geq 1}\left|\sum_{h=-\infty}^{\infty} \mathbb{E}\left(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j}\right)\right| \leq \sup_{i,j\geq 1} \|\Gamma(i,j)\| < \infty. \tag{A.11}$$

We then deduce that $\sup_{i,j\geq 1} \|\mathbf{\Gamma}(i,j)\| = O(1)$. The proof will thus follow from Lemma A.7 below, in which we show the consistency of $\hat{\mathbf{\Gamma}}_n(i,j)$ uniformly in i and j. □

Lemma A.7. *Under Assumptions* **A1**–**A8***, we have*

$$\sup_{i,j}\left\|\hat{\mathbf{\Gamma}}_n(i,j) - \mathbf{\Gamma}(i,j)\right\| \to 0 \textit{ in probability as } n \to \infty.$$

Proof of Lemma A.7. For any $\theta \in \Theta$, let

$$\begin{aligned}\mathcal{M}_{n\ ij,h}(\theta) \quad := \quad & \frac{1}{n}\sum_{t=1}^{n-|h|}\left(\left\{e'_{t-h}(\theta)\otimes\left(I_{d^2(p+q)}\otimes e'_{t-j-h}(\theta)\right)\right\}\right.\\ & \left.\otimes\left\{e'_t(\theta)\otimes\left(I_{d^2(p+q)}\otimes e'_{t-i}(\theta)\right)\right\}\right).\end{aligned}$$

By the ergodic theorem, we have

$$\begin{aligned}\mathcal{M}_{n\ ij,h}(\theta) \quad \to \quad & \mathcal{M}_{ij,h}(\theta) := \mathbb{E}\left(\left\{e'_{t-h}(\theta)\otimes\left(I_{d^2(p+q)}\otimes e'_{t-j-h}(\theta)\right)\right\}\right.\\ & \left.\otimes\left\{e'_t(\theta)\otimes\left(I_{d^2(p+q)}\otimes e'_{t-i}(\theta)\right)\right\}\right) \ a.s.\end{aligned}$$

A Taylor expansion of $\operatorname{vec}\hat{\mathcal{M}}_{n\ ij,h}$ around θ_0 and (2.2) give

$$\operatorname{vec}\hat{\mathcal{M}}_{n\ ij,h} \quad = \quad \operatorname{vec}\mathcal{M}_{n\ ij,h} + \left\{\frac{\partial \operatorname{vec}\mathcal{M}_{n\ ij,h}}{\partial\theta'}\right\}_{\theta_0}(\hat{\theta}_n-\theta_0) + O_P\left(\frac{1}{n}\right).$$

In view of (A.11), we then deduce that

$$\lim_{n\to\infty}\sup_{i,j\geq 1}\sup_{|h|<n}\left\|\frac{\partial\operatorname{vec}\mathcal{M}_{n\ ij,h}}{\partial\theta'}\right\| < \infty, \quad a.s. \tag{A.12}$$

By the ergodic theorem, (2.2) and (A.12), for any multiplicative norm, we have

$$\begin{aligned}\sup_{i,j\geq 1}\sup_{|h|<n}\left\|\operatorname{vec}\left(\hat{\mathcal{M}}_{n\ ij,h}-\mathcal{M}_{ij,h}\right)\right\| \leq & \lim_{n\to\infty}\sup_{i,j\geq 1}\sup_{|h|<n}\left\|\frac{\partial \operatorname{vec}\mathcal{M}_{n\ ij,h}}{\partial\theta'}\right\|\left\|\hat{\theta}_n-\theta_0\right\|\\ & + O_P\left(\frac{1}{n}\right) = O_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned} \tag{A.13}$$

We have

$$\begin{aligned}\hat{\mathbf{\Gamma}}_n(i,j)-\mathbf{\Gamma}(i,j) \quad = \quad & \sum_{h=-T_n}^{+T_n} f(hb_n)\left(\hat{\mathcal{M}}_{n\ ij,h}-\mathcal{M}_{ij,h}\right)\\ & + \sum_{h=-T_n}^{+T_n}\left\{f(hb_n)-1\right\}\mathcal{M}_{ij,h} - \sum_{|h|>T_n}\mathcal{M}_{ij,h}.\end{aligned}$$

By the triangular inequality, for any multiplicative norm, we have

$$\sup_{i,j\geq 1}\left\|\hat{\mathbf{\Gamma}}_n(i,j)-\mathbf{\Gamma}(i,j)\right\| \quad \leq \quad g_1+g_2+g_3,$$

where

$$g_1 = \sup_{i,j\geq 1}\sup_{|h|<n}\left\|\hat{\mathcal{M}}_{n\ ij,h}-\mathcal{M}_{ij,h}\right\|\sum_{|h|\leq T_n}\left|f(hb_n)\right|,$$

$$g_2 = \sum_{|h|\leq T_n}\left|f(hb_n)-1\right|\left\|\mathcal{M}_{ij,h}\right\| \quad \text{and} \quad g_3 = \sum_{|h|>T_n}\left\|\mathcal{M}_{ij,h}\right\|.$$

The non zero elements of $\mathcal{M}_{ij,h}$ are of the form $\mathbb{E}(e_{i_1t}e_{i_2t-i}e_{i_3t-h}e_{i_4t-h-j})$, with $(i_1,i_2,i_3,i_4)\in\{1,\dots,d\}^4$. Now, using the covariance inequality obtained by Davydov [13], it is easy to show that

$$\begin{aligned}\left|\mathbb{E}\left(e_{i_1t}e_{i_2t-i}e_{i_3t-h}e_{i_4t-h-j}\right)\right| &= \left|\mathrm{Cov}\left(e_{i_1t}e_{i_2t-i},e_{i_3t-h}e_{i_4t-h-j}\right)\right|\\ &\leq K\alpha_\epsilon^{\nu/(2+\nu)}\left(h\right).\end{aligned}$$

We then deduce that

$$\|\mathcal{M}_{ij,h}\|\leq K\alpha_\epsilon^{\nu/(2+\nu)}\left(h\right). \tag{A.14}$$

In view of **A7**, we thus have $g_3\to 0$ as $n\to\infty$. Let m be a fixed integer and we write $g_2\leq s_1+s_2$, where

$$s_1=\sum_{|h|\leq m}|f(hb_n)-1|\,\|\mathcal{M}_{ij,h}\| \quad\text{and}\quad s_2=\sum_{m<|h|\leq T_n}|f(hb_n)-1|\,\|\mathcal{M}_{ij,h}\|.$$

For $|h|\leq m$, we have $hb_n\to 0$ as $n\to\infty$ and $f(hb_n)\to 1$, it follows that $s_1\to 0$. If we choose m sufficiently large, s_2 becomes small, using (A.14) and the fact that $f(\cdot)$ is bounded. It follows that $g_2\to 0$. In view of (5.2) and (A.13), we have

$$g_1\leq\frac{1}{b_n}\sup_{i,j\geq 1}\sup_{|h|<n}\left\|\hat{\mathcal{M}}_{n\,ij,h}-\mathcal{M}_{ij,h}\right\|O(1)=O_P\left(\frac{1}{b_n\sqrt{n}}\right)=o_p(1),$$

since $nb_n^2\to\infty$, in view of (5.1). The conclusion is complete. □

Proof of Proposition 6.1. For $i,j\in\{1,\dots,d\}$, let $\tilde{\mathbf{A}}^*_{ij,h}=\mathbf{A}^*_{ij,h}$ for $0\leq h\leq M$ and $\tilde{\mathbf{A}}^*_{ij,h}=0_{d\times d}$ for $h>M$. Similarly, we defined $\tilde{\mathbf{B}}^*_{ij,h}$ and $\tilde{\lambda}_h$. We have

$$\begin{aligned}\mathrm{vec}\,J-\mathrm{vec}\,J^M &= 2\sum_{h=1}^{\infty}\Big\{\mathcal{M}\left(\lambda_h'\otimes\lambda_h'-\tilde{\lambda}_h'\otimes\tilde{\lambda}_h'\right)\mathrm{vec}\,\Sigma_{e0}^{-1}\\ &\quad+\mathcal{M}\left(\tilde{\lambda}_h'\otimes\tilde{\lambda}_h'\right)\mathrm{vec}\,\Sigma_{e0}^{-1}\Big\}.\end{aligned}$$

Recall that the polynomials

$$\mathbf{A}_\theta(z)=I_d-\sum_{i=1}^{p}\mathbf{A}_iz^i \text{ and } \mathbf{B}_\theta(z)=I_d-\sum_{i=1}^{q}\mathbf{B}_iz^i,$$

with $\mathbf{A}_i=A_0^{-1}A_i$ and $\mathbf{B}_i=A_0^{-1}B_iB_0^{-1}A_0$.

By **A2**, the zeroes of $\mathbf{A}_\theta(z)$ and $\mathbf{B}_\theta(z)$ are of modulus strictly greater than one:

$$\det A_\theta(z)\det B_\theta(z)=0\Rightarrow|z|>1.$$

Thus,

$$\mathbf{A}_\theta^{-1}(z)=\sum_{h=0}^{\infty}\mathcal{A}_hz^h \qquad\text{and}\quad \mathbf{B}_\theta^{-1}(z)=\sum_{h=0}^{\infty}\mathcal{B}_hz^h.$$

Consequently each the d^2 elements (rational function of $|z|$) of $\mathbf{A}_\theta^{-1}(z)$ and $\mathbf{B}_\theta^{-1}(z)$ are in the form

$$\frac{1}{\prod_{i=1}^{k}(1-\rho_i z)} = \sum_{h=0}^{\infty} d_h z^h \qquad \text{with} \quad |d_h| \leq (h+1)^{k-1}\overline{\rho}^h,$$

for all $|z| \leq 1$ and if

$$\max_{i=1,\dots,k} |\rho_i| \leq \overline{\rho} < 1,$$

where $\overline{\rho}$ is the inverse of the largest modulus of the zeroes of the polynomials $\det A_\theta(z)$ and $\det B_\theta(z)$.

We thus have

$$\max\{\|\mathcal{A}_h\|, \|\mathcal{B}_h\|\} \leq d^2(p+q)(h+1)^{k_0}\overline{\rho}^h \leq K_0\overline{\rho}^{h/2} \text{ and } \|\lambda_h\| \leq K_1\overline{\rho}^{h/2} \tag{A.15}$$

with

$$k_0 = d(p+q), \quad K_0 = d^2(p+q)\left(\frac{-2k_0}{\log\overline{\rho}}\right)^{k_0} \overline{\rho}^{\,-0.5-k_0/\log\overline{\rho}} \text{ and } K_1 = d^4(p+q)^2K_0.$$

We then obtain

$$\begin{aligned}\left\|\operatorname{vec} J - \operatorname{vec} J^M\right\| &\leq 2\sum_{h=M+1}^{\infty} \|\mathcal{M}\| \left\|\lambda_h' \otimes \lambda_h'\right\| \left\|\operatorname{vec} \Sigma_{e0}^{-1}\right\| \\ &\leq 2\overline{\pi}\overline{\Gamma}\sum_{h=M+1}^{\infty} \left\|\lambda_h'\right\|^2 \leq 2\overline{\pi}\overline{\Gamma}\left(\frac{K_1}{1-\overline{\rho}^{1/2}}\right)^2 \overline{\rho}^{M+1}\end{aligned}$$

and the result follows. □

Proof of Proposition 6.2. The proof is similar to that given in Proposition 6.1. However, we will give the expression which are different to that in the previous Proposition. Let

$$\overline{\mathbf{\Gamma}} = \max_{i,j\geq 0} \|\mathbf{\Gamma}(i,j)\|, \qquad \overline{\pi}_1 = \left\|\operatorname{vec}\left(\operatorname{vec}\Sigma_{e0}^{-1}\left\{\operatorname{vec}\Sigma_{e0}^{-1}\right\}'\right)\right\|.$$

Using (A.15), we then obtain

$$\begin{aligned}\left\|\operatorname{vec} I - \operatorname{vec} I^M\right\| &\leq 4\overline{\pi}_1\sum_{i,j=M+1}^{\infty} \|\mathbf{\Gamma}(i,j)\| \left\|\left\{I_d \otimes \lambda_j'\right\} \otimes \left\{I_d \otimes \lambda_i'\right\}\right\| \\ &\leq 4\overline{\pi}_1 d^2\overline{\mathbf{\Gamma}}\sum_{i,j=M+1}^{\infty} \left\|\lambda_i'\right\| \left\|\lambda_j'\right\| \leq 4\overline{\pi}_1 d^2\overline{\mathbf{\Gamma}}\left(\frac{K_1}{1-\overline{\rho}^{1/2}}\right)^2 \overline{\rho}^{M+1}\end{aligned}$$

and the result follows. □

Acknowledgements

I would like to thank the editor and the reviewer for helpful remarks and Julien Yves Rolland for his assistance on the simulation studies.

The work was partially supported by a BQR (Bonus Qualité Recherche) of the Université de Franche-Comté.

References

[1] Amendola, A. and Francq, C. (2009). Concepts and tools for nonlinear time series modelling. *Handbook of Computational Econometrics*, D. Belsley and E. Kontoghiorghes (eds), Wiley.

[2] Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858. MR1106513

[3] Andrews, B., Davis, R. A. and Breidt, F. J. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* **97**, 1638–1659. MR2256234

[4] Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics* **2**, 489–502. MR0421010

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327. MR0853051

[6] Boubacar Maïnassara, Y. (2011). Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms. *Journal of Statistical Planning and Inference* **141**, 2961–2975. MR2787759

[7] Boubacar Maïnassara, Y. (2012). Selection of weak VARMA models by modified Akaike's information criteria. *Journal of Time Series Analysis* **33**, 121–130. MR2877612

[8] Boubacar Maïnassara, Y., Carbon, M. and Francq, C. (2012). Computing and estimating information matrices of weak ARMA models. *Computational Statistics and Data Analysis* **56**, 345–361. MR2842343

[9] Boubacar Maïnassara, Y. and Francq, C. (2011). Estimating structural VARMA models with uncorrelated but non-independent error terms. *Journal of Multivariate Analysis* **102**, 496–505. MR2755011

[10] Bouhaddioui, C. and Roy, R. (2006). A generalized Portmanteau test for independence of two infinite-order vector autoregressive series. *Journal of Time Series Analysis* **27**, 505–544. MR2245711

[11] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. 2nd ed., Holden-Day, San Francisco. MR0436499

[12] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Verlag, New York. MR1093459

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

[14] den Hann, W. J. and Levin, A. (1997). A practitioner's guide to robust covariance matrix estimation. *Handbook of Statistics*, **15**, C. R. Rao and G. S. Maddala (eds), 291–341. MR1492705

[15] Duchesne, P. and Roy, R. (2004). On consistent testing for serial correlation of unknown form in vector time series models. *Journal of Multivariate Analysis* **89**, 148–180. MR2041214

[16] Dufour, J.-M. and Pelletier, D. (2005). Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada*.

[17] Durbin, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika* **46**, 306–316. MR0114283

[18] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987–1007. MR0666121

[19] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Verlag, New York. MR1964455

[20] Francq, C. and Raïssi, H. (2007). Multivariate Portmanteau test for autoregressive models with uncorrelated but nonindependent errors. *Journal of Time Series Analysis* **28**, 454–470. MR2355317

[21] Francq, C., Roy, R. and Zakoïan, J.-M. (2003). Goodness-of-fit tests for ARMA models with uncorrelated errors, *Discussion Paper CRM-2925, Centre de recherches mathématiques, Université de Montréal*. `http://www.crm.umontreal.ca/pub/Rapports/2900-2999/2925.pdf`.

[22] Francq, C., Roy, R. and Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* **100**, 532–544. MR2160557

[23] Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* **68**, 145–165. MR1627612

[24] Francq, C. and Zakoïan, J.-M. (2000). Covariance matrix estimation of mixing weak ARMA models. *Journal of Statistical Planning and Inference* **83**, 369–394. MR1748016

[25] Francq, C. and Zakoïan, J.-M. (2005). Recent results for linear time series models with non independent innovations. *Statistical Modeling and Analysis for Complex Data Problems*, Chap. **12**, P. Duchesne and B. Rémillard (eds). Springer Verlag, New York, 241–265. MR2189539

[26] Francq, C. and Zakoïan, J.-M. (2007). HAC estimation and strong linearity testing in weak ARMA models. *Journal of Multivariate Analysis* **98**, 114–144. MR2292920

[27] Godolphin, E. J and Bane, S. R. (2006). On the evaluation of the information matrix for multiplicative seasonal time series. *Journal of Time Series Analysis* **27**, 167–190. MR2235843

[28] Hannan, E. J. (1976). The identification and parametrization of ARMAX and state space forms. *Econometrica* **44**, 713–723. MR0479509

[29] Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed of Autoregressive Moving Average order. *Biometrika* **69**, 81–94. MR0655673

[30] Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864. MR1399220

[31] Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419–426.

[32] Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**, 70–86. MR1613694

[33] Katayama, N. (2012). Chi-squared portmanteau tests for structural VARMA models with uncorrelated errors. *Journal of Time Series Analysis* **33**, 863–872. MR2991904

[34] Klein, A. and Mélard, G. (1990). Fisher's information matrix for seasonal moving average models. *Journal of Time Series Analysis* **11**, 231–237. MR1064439

[35] Klein, A. and Mélard, G. (2004). An algorithm for computing the asymptotic Fisher information matrix for seasonal SISO models. *Journal of Time Series Analysis* **25**, 627–648. MR2089188

[36] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin. MR2172368

[37] Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Application in Statistics and Econometrics*. New-York, Wiley. MR0940471

[38] McLeod, A. I. (1978). On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society* **B 40**, 296–302. MR0522212

[39] McLeod, A. I. (1984). Duality and other properties of multiplicative seasonal autoregressive-moving average models. *Biometrika* **71**, 207–211. MR0738345

[40] Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708. MR0890864

[41] Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*, 2nd ed., Springer Verlag, New York. MR1451875

[42] Romano, J. L. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* **91**, 590–600. MR1395728

[43] Siddiqui, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autogressive process. *Annals of Mathematical Statistics* **29**, 585–588. MR0095568

[44] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*, Clarendon Press Oxford. MR1079320

[45] van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

[46] Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för Matematik* **2**, 423–434. MR0060797

[47] Whittle, P. (1963). On the fitting of multivariate autoregressions and the approximate canonical factorization of a special density matrix. *Biometrika* **40**, 129–134. MR0161430