# Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms

Yacouba Boubacar Mainassara

7. December 2009

# Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms

Yacouba Boubacar Mainassara *

**Abstract**

We consider portmanteau tests for testing the adequacy of vector autoregressive moving-average (VARMA) models under the assumption that the errors are uncorrelated but not necessarily independent. We relax the standard independence assumption to extend the range of application of the VARMA models, and allow to cover linear representations of general nonlinear processes. We first study the joint distribution of the quasi-maximum likelihood estimator (QMLE) or the least squared estimator (LSE) and the noise empirical autocovariances. We then derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. We deduce the asymptotic distribution of the Ljung-Box (or Box-Pierce) portmanteau statistics for VARMA models with nonindependent innovations. It is shown that the asymptotic distribution of the portmanteau tests is that of a weighted sum of independent chi-squared random variables, which can be quite different from the usual chi-squared approximation used under iid assumptions on the noise. Hence we propose a method to adjust the critical values of the portmanteau tests. Monte carlo experiments illustrate the finite sample performance of the modified portmanteau test.

**Keywords:** Goodness-of-fit test, QMLE/LSE, Box-Pierce and Ljung-Box portmanteau tests, residual autocorrelation, Structural representation, weak VARMA models.

*Université Lille III, EQUIPPE, BP 60 149, 59653 Villeneuve d'Ascq cedex, France. E-mail: yacouba.boubacarmainassara@univ-lille3.fr

# 1 Introduction

The vector autoregressive moving-average (VARMA) models are used in time series analysis and econometrics to represent multivariate time series (see Reinsel, 1997, Lütkepohl, 2005). These VARMA models are a natural extension of the univariate ARMA models, which constitute the most widely used class of univariate time series models (see *e.g.* Brockwell and Davis, 1991). The sub-class of vector autoregressive (VAR) models has been studied in the econometric literature (see also Lütkepohl, 1993).

The validity of the different steps of the traditional methodology of Box and Jenkins, identification, estimation and validation, depends on the noises properties. After identification and estimation of the vector autoregressive moving-average processes, the next important step in the VARMA modeling consists in checking if the estimated model fits satisfactory the data. This adequacy checking step allows to validate or invalidate the choice of the orders $p$ and $q$. In VARMA$(p, q)$ models, the choice of $p$ and $q$ is particularly important because the number of parameters, $(p+q+2)d^2$, quickly increases with $p$ and $q$, which entails statistical difficulties.

In particular, the selection of too large orders $p$ and $q$ has the effect of introducing terms that are not necessarily relevant in the model, which generates statistical difficulties leads to a loss of precision in parameter estimation. Conversely, the selection of too small orders $p$ and $q$ causes loss some of information that can be detected by a correlation of residuals.

Thus it is important to check the validity of a VARMA$(p, q)$ model, for a given order $p$ and $q$. This paper is devoted to the problem of the validation step of VARMA representations of multivariate processes. This validation stage is not only based on portmanteau tests, but also on the examination of the autocorrelation function of the residuals. Based on the residual empirical autocorrelations, Box and Pierce (1970) (**BP** hereafter) derived a goodness-of-fit test, the portmanteau test, for univariate strong ARMA models. Ljung and Box (1978) (**LB** hereafter) proposed a modified portmanteau test which is nowadays one of the most popular diagnostic checking tool in ARMA modeling of time series. The multivariate version of the **BP** portmanteau statistic was introduced by Chitturi (1974). We use this so-called portmanteau tests considered by Chitturi (1974) and Hosking (1980) for checking the overall significance of the residual autocorrelations of a VARMA$(p, q)$ model (see also Hosking, 1981a,b; Li and McLeod, 1981; Ahn, 1988). Hosking (1981a) gave several equivalent forms of this statistic. Arbués (2008) proposed an extended portmanteau test for VARMA models with mixing nonlinear constraints.

The papers on the multivariate version of the portmanteau statistic are generally under the assumption that the errors $\epsilon_t$ are independent. This independence assumption is restrictive because it precludes conditional heteroscedasticity and/ or other forms of nonlinearity (see Francq and Zakoïan, 2005, for a review on weak univariate ARMA models). Relaxing this independence assumption allows to cover linear representations of general nonlinear processes and to extend the range of application of the VARMA models. VARMA models with nonindependent innovations (*i.e.* weak VARMA models) have been less studied than VARMA models with iid errors (*i.e.* strong VARMA models).

The asymptotic theory of weak ARMA model validation is mainly limited to the univariate framework (see Francq and Zakoïan, 2005).

In the multivariate analysis, notable exceptions are Dufour and Pelletier (2005) who study the choice of the order $p$ and $q$ of VARMA models under weak assumptions on the innovation process, Francq and Raïssi (2007) who study portmanteau tests for weak VAR models, Chabot-Hallé and Duchesne (2008) who study the asymptotic distribution of LSE and portmanteau test for periodic VAR in which the error term is a martingale difference sequence, and Boubacar Mainassara and Francq (2009) who study the consistency and the asymptotic normality of the QMLE for weak VARMA model. The main goal of the present article is to complete the available results concerning the statistical analysis of weak VARMA models by considering the adequacy problem under a general error terms, which have not been studied in the above-mentioned papers. We proceed to study the behaviour of the goodness-of fit portmanteau tests when the $\epsilon_t$ are not independent. We will see that the standard portmanteau tests can be quite misleading in the framework of non independent errors. A modified version of these tests is thus proposed.

The paper is organized as follows. Section 2 presents the structural weak VARMA models that we consider here. Structural forms are employed in econometrics in order to introduce instantaneous relationships between economic variables. Section 3 presents the results on the QMLE/LSE asymptotic distribution obtained by Boubacar Mainassara and Francq (2009) when $(\epsilon_t)$ satisfies mild mixing assumptions. Section 4 is devoted to the joint distribution of the QMLE/LSE and the noise empirical autocovariances. In Section 5 we derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. In Section 6 it is shown how the standard Ljung-Box (or Box-Pierce) portmanteau tests must be adapted in the case of VARMA models with nonindependent innovations. 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 $\text{vec}A$ the vector obtained by stacking the columns of $A$. The reader is refereed to Magnus and Neudecker (1988) for the properties of these operators. Denoting by $\|Z\|$ the Euclidean norm of $Z$. Let $\text{I}_r$ 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

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

where $\epsilon_t$ is a white noise, namely a stationary sequence of centered and uncorrelated random variables with a non singular variance $\Sigma_0$. It is customary to say that $(X_t)$ is a strong VARMA$(p,q)$ model if $(\epsilon_t)$ is a strong white noise, that is, if it satisfies

**A1:** $(\epsilon_t)$ is a sequence of independent and identically distributed (iid) random vectors, $E\epsilon_t = 0$ and $\text{Var}(\epsilon_t) = \Sigma_0$.

We say that (1) is a weak VARMA$(p,q)$ model if $(\epsilon_t)$ is a weak white noise, that is, if it satisfies

**A1':** $E\epsilon_t = 0$, $\text{Var}(\epsilon_t) = \Sigma_0$, and $\text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$ for all $t \in \mathbb{Z}$ and all $h \neq 0$.

Assumption **A1** is clearly stronger than **A1'**. The class of strong VARMA models is often considered too restrictive by practitioners. The standard VARMA$(p,q)$ form, which is sometimes called the reduced form, is obtained for $A_{00} = B_{00} = \text{I}_d$. Let $[A_{00} \dots A_{0p} B_{00} \dots B_{0q}]$ be the $d \times (p+q+2)d$ matrix of VAR and MA coefficients. The matrix $\Sigma_0$ is non singular and is considered as a nuisance parameter. The parameter of interest is denoted $\theta_0$, where $\theta_0$ belongs to the parameter space $\Theta \subset \mathbb{R}^{k_0}$, and $k_0$ is the number of unknown parameters, which is typically much smaller that $(p_0 + q_0 + 3)d^2$. The matrices $A_{00}, \dots A_{0p_0}, B_{00}, \dots B_{0q_0}$ involved in (1) and $\Sigma_0$ are specified by $\theta_0$. More precisely, we write $A_{0i} = A_i(\theta_0)$ and $B_{0j} = B_j(\theta_0)$ for $i = 0, \dots, p_0$ and $j = 0, \dots, q_0$, and $\Sigma_0 = \Sigma(\theta_0)$. We need the following assumptions used by Boubacar Mainassara and Francq (2009) to ensure the consistence and the asymptotic normality of the quasi-maximum likelihood estimator (QMLE).

**A2:** The functions $\theta \mapsto A_i(\theta)$ $i = 0, \dots, p$, $\theta \mapsto B_j(\theta)$ $j = 0, \dots, q$ 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 $\Sigma$ 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 $B_\theta(z) = B_0 - \sum_{i=1}^{q} B_i z^i$.

**A3:** For all $\theta \in \Theta$, we have $\det A_\theta(z) \det B_\theta(z) \neq 0$ for all $|z| \leq 1$.

**A4:** We have $\theta_0 \in \Theta$, where $\Theta$ is compact.

**A5:** The process $(\epsilon_t)$ is stationary and ergodic.

Note that **A5** is entailed by **A1**, but not by **A1′**. Note that $(\epsilon_t)$ can be replaced by $(X_t)$ in **A5**, 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.

**A6:** 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

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

**A7:** We have $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of $\Theta$.

**A8:** We have $E\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$,

where $\alpha_\epsilon(k)$, $k = 0, 1, \dots$, denotes the strong mixing coefficients of the process $(\epsilon_t)$. The reader is referred to Boubacar Mainassara and Francq (2009) for a discussion of these assumptions.

# 3 Least Squares Estimation under non-iid innovations

For all $\theta \in \Theta$, let $A_0 = A_0(\theta), \dots, A_p = A_p(\theta)$, $B_0 = B_0(\theta), \dots, B_q = B_q(\theta)$ and $\Sigma = \Sigma(\theta)$. Note that from **A3** 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

$$X_t - \sum_{i=1}^{p} A_{00}^{-1}A_{0i}X_{t-i} = e_t - \sum_{i=1}^{q} A_{00}^{-1}B_{0i}B_{00}^{-1}A_{00}e_{t-i}. \tag{2}$$

Note that $e_t(\theta_0) = e_t$. For simplicity, we will omit the notation $\theta$ in all quantities taken at the true value, $\theta_0$. For all $\theta \in \Theta$, the assumption on the MA polynomial (from **A3**) implies that there exists a sequence of constants matrices $(C_i(\theta))$ such that $\sum_{i=1}^{\infty} \|C_i(\theta)\| < \infty$ and

$$e_t(\theta) = X_t - \sum_{i=1}^{\infty} C_i(\theta)X_{t-i}. \tag{3}$$

Given a realization $X_1, X_2, \dots, X_n$ satisfying the VARMA representation (1), the variable $e_t(\theta)$ can be approximated, for $0 < t \le n$, by $\tilde{e}_t(\theta)$ defined recursively by

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

where the unknown initial values are set to zero: $\tilde{e}_0(\theta) = \cdots = \tilde{e}_{1-q}(\theta) = X_0 = \cdots = X_{1-p} = 0$. The gaussian quasi-likelihood is given by

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

A quasi-maximum likelihood (QML) of $\theta$ and $\Sigma_e$ are a measurable solution $(\hat{\theta}_n, \hat{\Sigma}_e)$ of

$$(\hat{\theta}_n, \hat{\Sigma}_e) = \arg\min_{\theta, \Sigma_e} \left\{\log(\det \Sigma_e) + \frac{1}{n}\sum_{t=1}^{n} \tilde{e}_t(\theta)\Sigma_e^{-1}\tilde{e}_t'(\theta)\right\}.$$

We now use the matrix $M_{\theta_0}$ of the coefficients of the reduced form to that made by Boubacar Mainassara and Francq (2009), 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 an assumption which specifies how this matrix depends on the parameter $\theta_0$. Let $\dot{M}_{\theta_0}$ be the matrix $\partial \mathrm{vec}(M_\theta)/\partial\theta'$ evaluated at $\theta_0$.

**A9:** The matrix $\dot{M}_{\theta_0}$ is of full rank $k_0$.

Under the following additional assumption, Boubacar Mainassara and Francq (2009) showed respectively in Theorem 1 and Theorem 2 the consistency and the asymptotic normality of the QML estimator of weak multivariate ARMA model. One of the most popular estimation procedure is that of the least squares estimator (LSE) minimizing

$$\log\det\hat{\Sigma}_e = \log\det\left\{\frac{1}{n}\sum_{t=1}^{n}\tilde{e}_t(\hat{\theta})\tilde{e}_t'(\hat{\theta})\right\},$$

or equivalently

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

For the processes of the form (2), under **A1'**, **A2**-**A9**, it can be shown (see *e.g.* Boubacar Mainassara and Francq 2009), that the LS estimator of $\theta$ coincides with the gaussian quasi-maximum likelihood estimator (QMLE). More precisely, $\hat{\theta}_n$ satisfies, almost surely,

$$\mathcal{O}_n(\hat{\theta}_n) = \min_{\theta\in\Theta}\mathcal{O}_n(\theta),$$

where

$$\mathcal{O}_n(\theta) = \log\det\left\{\frac{1}{n}\sum_{t=1}^{n}\tilde{e}_t(\theta)\tilde{e}_t'(\theta)\right\} \quad \text{or} \quad \mathcal{O}_n(\theta) = \det\left\{\frac{1}{n}\sum_{t=1}^{n}\tilde{e}_t(\theta)\tilde{e}_t'(\theta)\right\}.$$

To obtain the consistency and asymptotic normality of the QMLE/LSE, it will be convenient to consider the functions

$$O_n(\theta) = \log\det\Sigma_n \quad \text{or} \quad O_n(\theta) = \det\Sigma_n,$$

where $\Sigma_n = \Sigma_n(\theta) = n^{-1}\sum_{t=1}^n e_t(\theta)e_t'(\theta)$ and $(e_t(\theta))$ is given by (3). Under **A1'**, **A2**–**A9** or **A1**–**A4**, **A6**, **A8** and **A9**, let $\hat{\theta}_n$ be the LS estimate of $\theta_0$ by maximizing

$$O_n(\theta) = \log\det\left\{\frac{1}{n}\sum_{t=1}^n e_t(\theta)e_t'(\theta)\right\}.$$

In the univariate case, Francq and Zakoïan (1998) showed the asymptotic normality of the LS estimator under mixing assumptions. This remains valid of the multivariate LS estimator. Then under the assumptions **A1', A2**–**A9**, $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma_{\hat{\theta}_n} := J^{-1}IJ^{-1}$, where $J = J(\theta_0)$ and $I = I(\theta_0)$, with

$$J(\theta) = \lim_{n\to\infty}\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\mathcal{O}_n(\theta) \quad a.s.$$

and

$$I(\theta) = \lim_{n\to\infty}\mathrm{Var}\,\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\mathcal{O}_n(\theta).$$

In the standard strong VARMA case, *i.e.* when **A5** is replaced by the assumption **A1** that $(\epsilon_t)$ is iid, we have $I = J$, so that $\Sigma_{\hat{\theta}_n} = J^{-1}$.

# 4 Joint distribution of $\hat{\theta}_n$ and the noise empirical autocovariances

Let $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ be the LS 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, $\hat{\Sigma}_{e0} = \hat{\Gamma}_e(0) = n^{-1}\sum_{t=1}^n \hat{e}_t\hat{e}_t'$. We denote by

$$\gamma(h) = \frac{1}{n}\sum_{t=h+1}^n e_te_{t-h}' \quad\text{and}\quad \hat{\Gamma}_e(h) = \frac{1}{n}\sum_{t=h+1}^n \hat{e}_t\hat{e}_{t-h}'$$

the white noise "empirical" autocovariances and residual autocovariances. It should be noted that $\gamma(h)$ is not a statistic (unless if $p = q = 0$) because it depends on the unobserved innovations $e_t = e_t(\theta_0)$. For a fixed integer $m \geq 1$, let

$$\gamma_m = \left(\{\text{vec}\gamma(1)\}', \ldots, \{\text{vec}\gamma(m)\}'\right)'$$

and

$$\hat{\Gamma}_m = \left(\left\{\text{vec}\hat{\Gamma}_e(1)\right\}', \ldots, \left\{\text{vec}\hat{\Gamma}_e(m)\right\}'\right)',$$

and let

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

for $(\ell, \ell') \neq (0, 0)$. For the univariate ARMA model, Francq, Roy and Zakoïan (2005) have showed in Lemma A.1 that $|\Gamma(\ell, \ell')| \leq K \max(\ell, \ell')$ for some constant $K$, which is sufficient to ensure the existence of these matrices. We can generalize this result for the multivariate ARMA model. Then we obtain $\|\Gamma(\ell, \ell')\| \leq K \max(\ell, \ell')$ for some constant $K$. The proof is similar to the univariate case.

We are now able to state the following theorem, which is an extension of a result given in Francq, Roy and Zakoïan (2005).

**Theorem 4.1** *Assume $p > 0$ or $q > 0$. Under Assumptions* **A1'**-**A2**–**A9** *or* **A1**–**A4**, **A6**, **A8** *and* **A9**, *as $n \to \infty$, $\sqrt{n}(\gamma_m, \hat{\theta}_n - \theta_0)' \stackrel{d}{\Rightarrow} \mathcal{N}(0, \Xi)$ where*

$$\Xi = \begin{pmatrix} \Sigma_{\gamma_m} & \Sigma_{\gamma_m, \hat{\theta}_n} \\ \Sigma'_{\gamma_m, \hat{\theta}_n} & \Sigma_{\hat{\theta}_n} \end{pmatrix},$$

*with $\Sigma_{\gamma_m} = \{\Gamma(\ell, \ell')\}_{1 \leq \ell, \ell' \leq m}$, $\Sigma'_{\gamma_m, \hat{\theta}_n} = \lim_{n\to\infty} Cov(\sqrt{n}J^{-1}Y_n, \sqrt{n}\gamma_m)$ and $\Sigma_{\hat{\theta}_n} = \lim_{n\to\infty} Var(\sqrt{n}J^{-1}Y_n) = J^{-1}IJ^{-1}$ and $Y_n$ is given by (16) in the proof of this Theorem. The matrices $I$ and $J$ are defined in Section 2.*

# 5 Asymptotic distribution of residual empirical autocovariances and autocorrelations

Let the diagonal matrices

$$S_e = \text{Diag}\left(\sigma_e(1), \ldots, \sigma_e(d)\right) \quad \text{and} \quad \hat{S}_e = \text{Diag}\left(\hat{\sigma}_e(1), \ldots, \hat{\sigma}_e(d)\right),$$

where $\sigma_e^2(i)$ is the variance of the $i$-th coordinate of $e_t$ and $\hat{\sigma}_e^2(i)$ is its sample estimate, *i.e.* $\sigma_e(i) = \sqrt{Ee_{it}^2}$ and $\hat{\sigma}_e(i) = \sqrt{n^{-1}\sum_{t=1}^{n} \hat{e}_{it}^2}$. The theoretical and sample autocorrelations at lag $\ell$ are respectively defined by $R_e(\ell) = S_e^{-1}\Gamma_e(\ell)S_e^{-1}$ and $\hat{R}_e(\ell) = \hat{S}_e^{-1}\hat{\Gamma}_e(\ell)\hat{S}_e^{-1}$, with $\Gamma_e(\ell) := Ee_te'_{t-\ell} = 0$ for all $\ell \neq 0$. Consider the vector of the first $m$ sample autocorrelations

$$\hat{\rho}_m = \left(\left\{\mathrm{vec}\hat{R}_e(1)\right\}', \ldots, \left\{\mathrm{vec}\hat{R}_e(m)\right\}'\right)'.$$

**Theorem 5.1** *Under Assumptions in Theorem 4.1,*

$$\sqrt{n}\hat{\Gamma}_m \Rightarrow \mathcal{N}\left(0, \Sigma_{\hat{\Gamma}_m}\right) \quad \textit{and} \quad \sqrt{n}\hat{\rho}_m \Rightarrow \mathcal{N}\left(0, \Sigma_{\hat{\rho}_m}\right) \quad \textit{where,}$$

$$\Sigma_{\hat{\Gamma}_m} = \Sigma_{\gamma_m} + \Phi_m\Sigma_{\hat{\theta}_n}\Phi'_m + \Phi_m\Sigma_{\hat{\theta}_n,\gamma_m} + \Sigma'_{\hat{\theta}_n,\gamma_m}\Phi'_m \qquad (4)$$

$$\Sigma_{\hat{\rho}_m} = \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\}\Sigma_{\hat{\Gamma}_m}\left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\} \qquad (5)$$

*and $\Phi_m$ is given by (20) in the proof of this Theorem.*

# 6 Limiting distribution of the portmanteau statistics

Box and Pierce (1970) (**BP** hereafter) derived a goodness-of-fit test, the portmanteau test, for univariate strong ARMA models. Ljung and Box (1978) (**LB** hereafter) proposed a modified portmanteau test which is nowadays one of the most popular diagnostic checking tool in ARMA modeling of time series. The multivariate version of the **BP** portmanteau statistic was introduced by Chitturi (1974). Hosking (1981a) gave several equivalent forms of this statistic. Basic forms are

$$
\begin{aligned}
Q_m &= n\sum_{h=1}^{m} \mathrm{Tr}\left(\hat{\Gamma}_e'(h)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\hat{\Gamma}_e^{-1}(0)\right) \\
&= n\sum_{h=1}^{m} \mathrm{vec}\left(\hat{\Gamma}_e(h)\right)'\left(\hat{\Gamma}_e^{-1}(0)\otimes I_d\right)\mathrm{vec}\left(\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\right) \\
&= n\sum_{h=1}^{m} \mathrm{vec}\left(\hat{\Gamma}_e(h)\right)'\left(\hat{\Gamma}_e^{-1}(0)\otimes I_d\right)\left(I_d\otimes\hat{\Gamma}_e^{-1}(0)\right)\mathrm{vec}\left(\hat{\Gamma}_e(h)\right) \\
&= n\sum_{h=1}^{m} \mathrm{vec}\left(\hat{\Gamma}_e(h)\right)'\left(\hat{\Gamma}_e^{-1}(0)\otimes\hat{\Gamma}_e^{-1}(0)\right)\mathrm{vec}\left(\hat{\Gamma}_e(h)\right) \\
&= n\hat{\Gamma}_m'\left(I_m\otimes\left\{\hat{\Gamma}_e^{-1}(0)\otimes\hat{\Gamma}_e^{-1}(0)\right\}\right)\hat{\Gamma}_m \\
&= n\hat{\rho}_m'\left(I_m\otimes\left\{\hat{\Gamma}_e(0)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(0)\right\}\otimes\left\{\hat{\Gamma}_e(0)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(0)\right\}\right)\hat{\rho}_m \\
&= n\hat{\rho}_m'\left(I_m\otimes\left\{\hat{R}_e^{-1}(0)\otimes\hat{R}_e^{-1}(0)\right\}\right)\hat{\rho}_m.
\end{aligned}
$$

Where the equalities is obtained from the elementary relations $\mathrm{vec}(AB) = (I\otimes A)\,\mathrm{vec}\,B$, $(A\otimes B)(C\otimes D) = AC\otimes BD$ and $\mathrm{Tr}(ABC) = \mathrm{vec}(A')'(C'\otimes I)\,\mathrm{vec}\,B$. Similarly to the univariate **LB** portmanteau statistic, Hosking (1980) defined the modified portmanteau statistic

$$
\tilde{Q}_m = n^2\sum_{h=1}^{m}(n-h)^{-1}\mathrm{Tr}\left(\hat{\Gamma}_e'(h)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\hat{\Gamma}_e^{-1}(0)\right).
$$

These portmanteau statistics are generally used to test the null hypothesis

$$
H_0: \quad (X_t) \text{ satisfies a VARMA}(p,q) \text{ represntation}
$$

against the alternative

$$
H_1: \quad (X_t) \text{ does not admit a VARMA represntation or admits a VARMA}(p',q') \text{ represntation with } p'>p \text{ or } q'>q.
$$

These portmanteau tests are very useful tools for checking the overall significance of the residual autocorrelations. Under the assumption that the data generating process (**DGP**) follows a strong VARMA$(p,q)$ model, the asymptotic distribution of the statistics $Q_m$ and $\tilde{Q}_m$ is generally approximated by the $\chi^2_{d^2m-k_0}$ distribution ($d^2m > k_0$) (the degrees of freedom are obtained by subtracting the number of freely estimated VARMA coefficients from $d^2m$). When the innovations are gaussian, Hosking (1980) found that the finite-sample distribution of $\tilde{Q}_m$ is more nearly $\chi^2_{d^2(m-(p+q))}$ than that of $Q_m$. From Theorem 5.1 we deduce the following result, which gives the exact asymptotic distribution of the standard portmanteau statistics $Q_m$. We will see that the distribution may be very different from the $\chi^2_{d^2m-k_0}$ in the case of VARMA$(p,q)$ models.

**Theorem 6.1** *Under Assumptions in Theorem 5.1, the statistics $Q_m$ and $\tilde{Q}_m$ converge in distribution, as $n \to \infty$, to*

$$Z_m(\xi_m) = \sum_{i=1}^{d^2m} \xi_{i,d^2m} Z_i^2$$

*where $\xi_m = (\xi_{1,d^2m}, \dots, \xi_{d^2m,d^2m})'$ is the vector of the eigenvalues of the matrix*

$$\Omega_m = \left(I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}\right) \Sigma_{\hat{\Gamma}_m} \left(I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}\right),$$

*and $Z_1, \dots, Z_m$ are independent $\mathcal{N}(0,1)$ variables.*

It is seen in Theorem 6.1, that the asymptotic distribution of the **BP** and **LB** portmanteau tests depends of the nuisance parameters involving $\Sigma_e$, the matrix $\Phi_m$ and the elements of the matrix $\Xi$. We need an consistent estimator of the above unknown matrices. The matrix $\Sigma_e$ can be consistently estimate by its sample estimate $\hat{\Sigma}_e = \hat{\Gamma}_e(0)$. The matrix $\Phi_m$ can be easily estimated by its empirical counterpart

$$\hat{\Phi}_m = \frac{1}{n} \sum_{t=1}^{n} \left\{ \left(\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m}\right)' \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}_{\theta_0 = \hat{\theta}_n}.$$

In the econometric literature the nonparametric kernel estimator, also called heteroskedastic autocorrelation consistent (HAC) estimator (see Newey and West, 1987, or Andrews, 1991), is widely used to estimate covariance matrices of the form $\Xi$. An alternative method consists in using a parametric AR estimate of the spectral density of $\Upsilon_t = \left(\Upsilon'_{1,t}, \Upsilon'_{2,t}\right)'$, where $\Upsilon_{1,t} = \left(e'_{t-1}, \dots, e'_{t-m}\right)' \otimes e_t$ and $\Upsilon_{2,t} = -2J^{-1} \left(\partial e'_t(\theta_0)/\partial\theta\right) \Sigma_{e0}^{-1} e_t(\theta_0)$. Interpreting $(2\pi)^{-1}\Xi$ as the spectral density of the stationary process $(\Upsilon_t)$

evaluated at frequency 0 (see Brockwell and Davis, 1991, p. 459). This approach, which has been studied by Berk (1974) (see also den Hann and Levin, 1997). So we have

$$\Xi = \mathbf{\Phi}^{-1}(1)\Sigma_u \mathbf{\Phi}^{-1}(1)$$

when $(\Upsilon_t)$ satisfies an AR($\infty$) representation of the form

$$\mathbf{\Phi}(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty} \Phi_i \Upsilon_{t-i} = u_t, \tag{6}$$

where $u_t$ is a weak white noise with variance matrix $\Sigma_u$. Since $\Upsilon_t$ is not observable, let $\hat{\Upsilon}_t$ be the vector obtained by replacing $\theta_0$ by $\hat{\theta}_n$ in $\Upsilon_t$. Let $\hat{\mathbf{\Phi}}_r(z) = \mathrm{I}_{k_0+d^2m} + \sum_{i=1}^{r} \hat{\Phi}_{r,i} z^i$, where $\hat{\Phi}_{r,1}, \ldots, \hat{\Phi}_{r,r}$ denote the coefficients of the LS regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}, \ldots, \hat{\Upsilon}_{t-r}$. Let $\hat{u}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance of $\hat{u}_{r,1}, \ldots, \hat{u}_{r,n}$.

We are now able to state the following theorem, which is an extension of a result given in Francq, Roy and Zakoïan (2005).

**Theorem 6.2** *In addition to the assumptions of Theorem 4.1, assume that the process* $(\Upsilon_t)$ *admits an AR*($\infty$) *representation (6) in which the roots of* $\det \mathbf{\Phi}(z) = 0$ *are outside the unit disk,* $\|\Phi_i\| = o(i^{-2})$*, and* $\Sigma_u = Var(u_t)$ *is non-singular. Moreover we assume that* $E\|\epsilon_t\|^{8+4\nu} < \infty$ *and* $\sum_{k=0}^{\infty} \{\alpha_{X,\epsilon}(k)\}^{\nu/(2+\nu)} < \infty$ *for some* $\nu > 0$*, where* $\{\alpha_{X,\epsilon}(k)\}_{k\geq 0}$ *denotes the sequence of the strong mixing coefficients of the process* $(X_t', \epsilon_t')'$*. Then the spectral estimator of* $\Xi$

$$\hat{\Xi}^{\mathrm{SP}} := \hat{\mathbf{\Phi}}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\mathbf{\Phi}}_r'^{-1}(1) \to \Xi$$

*in probability when* $r = r(n) \to \infty$ *and* $r^3/n \to 0$ *as* $n \to \infty$*.*

Let $\hat{\Omega}_m$ be the matrix obtained by replacing $\Xi$ by $\hat{\Xi}$ and $\Sigma_e$ by $\hat{\Sigma}_e$ in $\Omega_m$. Denote by $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \ldots, \hat{\xi}_{d^2m,d^2m})'$ the vector of the eigenvalues of $\hat{\Omega}_m$. At the asymptotic level $\alpha$, the **LB** test (resp. the **BP** test) consists in rejecting the adequacy of the weak VARMA$(p,q)$ model when

$$\tilde{Q}_m > S_m(1-\alpha) \quad (\text{resp.} \quad Q_m > S_m(1-\alpha))$$

where $S_m(1-\alpha)$ is such that $P\left\{Z_m(\hat{\xi}_m) > S_m(1-\alpha)\right\} = \alpha$.

# 7 Implementation of the goodness-of-fit portmanteau tests

Let $X_1, \dots, X_n$, be observations of a $d$-multivariate process. For testing the adequacy of weak VARMA$(p,q)$ model, we use the following steps to implement the modified version of the portmanteau test.

1. Compute the estimates $\hat{A}_1, \dots, \hat{A}_p, \hat{B}_1, \dots, \hat{B}_q$ by QMLE/LSE.

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

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

for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$.

3. Compute the residual autocovariances $\hat{\Gamma}_e(0) = \hat{\Sigma}_{e0}$ and $\hat{\Gamma}_e(h)$ for $h = 1, \dots, m$ and $\hat{\Gamma}_m = \left( \left\{ \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \hat{\Gamma}_e(m) \right\}' \right)'$.

4. Compute the matrix $\hat{J} = 2n^{-1} \sum_{t=1}^{n} \left( \partial \hat{e}_t' / \partial \theta \right) \hat{\Sigma}_{e0}^{-1} \left( \partial \hat{e}_t / \partial \theta' \right)$.

5. Compute $\hat{\Upsilon}_t = \left( \hat{\Upsilon}_{1,t}', \hat{\Upsilon}_{2,t}' \right)'$, where $\hat{\Upsilon}_{1,t} = \left( \hat{e}_{t-1}', \dots, \hat{e}_{t-m}' \right)' \otimes \hat{e}_t$ and $\hat{\Upsilon}_{2,t} = -2\hat{J}^{-1} \left( \partial \hat{e}_t' / \partial \theta \right) \hat{\Sigma}_{e0}^{-1} \hat{e}_t$.

6. Fit the VAR$(r)$ model

$$\hat{\mathbf{\Phi}}_r(L) \hat{\Upsilon}_t := \left( I_{d^2 m + k_0} + \sum_{i=1}^{r} \hat{\Phi}_{r,i}(L) \right) \hat{\Upsilon}_t = \hat{u}_{r,t}.$$

   The VAR order $r$ can be fixed or selected by AIC/BIC information criteria.

7. Define the estimator

$$\hat{\Xi}^{\mathrm{SP}} := \hat{\mathbf{\Phi}}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\mathbf{\Phi}}_r'^{-1}(1) = \begin{pmatrix} \hat{\Sigma}_{\gamma_m} & \hat{\Sigma}_{\gamma_m, \hat{\theta}_n} \\ \hat{\Sigma}_{\gamma_m, \hat{\theta}_n}' & \hat{\Sigma}_{\hat{\theta}_n} \end{pmatrix}, \quad \hat{\Sigma}_{\hat{u}_r} = \frac{1}{n} \sum_{t=1}^{n} \hat{u}_{r,t} \hat{u}_{r,t}'.$$

8. Define the estimator

$$\hat{\Phi}_m = \frac{1}{n}\sum_{t=1}^{n}\left\{\left(\hat{e}'_{t-1},\dots,\hat{e}'_{t-m}\right)'\otimes\frac{\partial e_t(\theta_0)}{\partial\theta'}\right\}_{\theta_0=\hat{\theta}_n}.$$

9. Define the estimators

$$\hat{\Sigma}_{\hat{\Gamma}_m} = \hat{\Sigma}_{\gamma_m} + \hat{\Phi}_m\hat{\Sigma}_{\hat{\theta}_n}\hat{\Phi}'_m + \hat{\Phi}_m\hat{\Sigma}_{\hat{\theta}_n,\gamma_m} + \hat{\Sigma}'_{\hat{\theta}_n,\gamma_m}\hat{\Phi}'_m$$

$$\hat{\Sigma}_{\hat{\rho}_m} = \left\{I_m\otimes(\hat{S}_e\otimes\hat{S}_e)^{-1}\right\}\hat{\Sigma}_{\hat{\Gamma}_m}\left\{I_m\otimes(\hat{S}_e\otimes\hat{S}_e)^{-1}\right\}$$

10. Compute the eigenvalues $\hat{\xi}_m = (\hat{\xi}_{1,d^2m},\dots,\hat{\xi}_{d^2m,d^2m})'$ of the matrix

$$\hat{\Omega}_m = \left(I_m\otimes\hat{\Sigma}_{e0}^{-1/2}\otimes\hat{\Sigma}_{e0}^{-1/2}\right)\hat{\Sigma}_{\hat{\Gamma}_m}\left(I_m\otimes\hat{\Sigma}_{e0}^{-1/2}\otimes\hat{\Sigma}_{e0}^{-1/2}\right).$$

11. Compute the portmanteau statistics

$$Q_m = n\hat{\rho}'_m\left(I_m\otimes\left\{\hat{R}_e^{-1}(0)\otimes\hat{R}_e^{-1}(0)\right\}\right)\hat{\rho}_m \quad \text{and}$$

$$\tilde{Q}_m = n^2\sum_{h=1}^{m}\frac{1}{(n-h)}\mathrm{Tr}\left(\hat{\Gamma}'_e(h)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\hat{\Gamma}_e^{-1}(0)\right).$$

12. Evaluate the $p$-values

$$P\left\{Z_m(\hat{\xi}_m) > Q_m\right\} \quad \text{and} \quad P\left\{Z_m(\hat{\xi}_m) > \tilde{Q}_m\right\}, \quad Z_m(\hat{\xi}_m) = \sum_{i=1}^{d^2m}\hat{\xi}_{i,d^2m}Z_i^2,$$

using the Imhof algorithm (1961). The **BP** test (resp. the **LB** test) rejects the adequacy of the weak VARMA$(p,q)$ model when the first (resp. the second) $p$-value is less than the asymptotic level $\alpha$.

# 8 Numerical illustrations

In this section, by means of Monte Carlo experiments, we investigate the finite sample properties of the test introduced in this paper. For illustrative purpose, we only present the results of the modified and standard versions of the **LB** test. The results concerning the **BP** test are not presented here, because they are very close to those of the **LB** test. The numerical illustrations of this section are made with the free statistical softwares R (see http://cran.r-project.org/) and FORTRAN (to compute the $p$-values using the Imohf algorithm, 1961).

We used the spectral estimator $\hat{I} = \hat{I}^{\mathrm{SP}}$ defined in Theorem 6.2, and the AR order $r = r(n)$ is automatically selected by BIC criterion in the weak VARMA models (in this case, Theorem 6.2 requires that $r \to \infty$), using the function `VARselect()` of the **vars** R package. In the strong VARMA case we can be shown that, the AR spectral estimator is consistent with any fixed value of $r$ (or $r = o(n^{1/3})$ as in Theorem 6.2). In Table 1 we took $r = 1$.

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%. When the relative rejection frequencies are outside the significant limits, they are displayed in bold type in Tables 1,...,4.

## 8.1 Empirical size

To generate the strong and the weak VARMA models, we consider the bivariate model of the form

$$\begin{aligned} \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & a_1(2,2) \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \\ &\quad - \begin{pmatrix} 0 & 0 \\ b_1(2,1) & b_1(2,2) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix}, \end{aligned} \tag{7}$$

where $(a_1(2,2), b_1(2,1), b_1(2,2)) = (0.950, -0.313, 0.250)$. This model (7) is a VARMA(1,1) model in echelon form.

### 8.1.1 Strong VARMA model case

We first consider the strong VARMA case. To generate this model, we assume that in (7) the innovation process $(\epsilon_t)$ is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \mathrm{IID}\,\mathcal{N}(0, I_2). \tag{8}$$

We simulated $N = 1,000$ independent trajectories of size $n = 100$, $n = 500$ and $n = 2,000$ of Model (7) with the strong Gaussian noise (8). For each of these $N$ replications we estimated the coefficients $(a_1(2,2), b_1(2,1), b_1(2,2))$

and we applied portmanteau tests to the residuals for different values of $m$. For the standard **LB** test the model is therefore rejected when the statistic $\tilde{Q}_m$ is greater than $\chi^2_{(4m-3)}(0.95)$, where $m$ is the number of autocorrelations used in the **LB** statistic. This corresponds to a nominal asymptotic level $\alpha = 5\%$ in the standard case. We know that the asymptotic level of the standard **LB** test is indeed $\alpha = 5\%$ when $(a_1(2,2), b_1(2,1), b_1(2,2)) = (0,0,0)$. Note however that, even when the noise is strong, the asymptotic level is not exactly $\alpha = 5\%$ when $(a_1(2,2), b_1(2,1), b_1(2,2)) \neq (0,0,0))$.

For the modified **LB** test the model is therefore rejected when the statistic $\tilde{Q}_m$ is greater than $S_m(0.95)$ *i.e.* when the $p$-value $(P\{Z_m(\hat{\xi}_m) > \tilde{Q}_m\})$ is less than the asymptotic level $\alpha = 0.05$. Let $A$ and $B$ the $(2 \times 2)$-matrices with non zero elements $a_1(2,2)$, $b_1(2,1)$ and $b_1(2,2)$. When the roots of $\det(I_2 - Az)\det(I_2 - Bz) = 0$ are near the unit disk, so the asymptotic distribution of $\tilde{Q}_m$ is likely to be far from its $\chi^2_{(4m-3)}$ approximation. Table 1 displays the relative rejection frequencies of the null hypothesis $H_0$ that the **DGP** follows an VARMA$(1,1)$ model in (7), over the $N = 1,000$ independent replications. As expected the observed relative rejection frequency of the standard **LB** test is very far from the nominal $\alpha = 5\%$ when the number $m$ of autocorrelations used in the **LB** statistic is small. This is in accordance with the results in the literature on the standard VARMA models. In particular, Hosking (1980) showed that the statistic $\tilde{Q}_m$ has approximately the chi-squared distribution $\chi^2_{d^2(m-(p+q))}$ without any identifiability contraint. The theory that the $\chi^2_{(4m-3)}$ approximation is better for larger $m$ is confirmed. Thus the error of first kind is well controlled by all the tests in the strong case, except for the standard **LB** test when $m \leq p+q$. We draw the somewhat surprising conclusion that, even in the strong VARMA case, the modified version may be preferable to the standard one, when the number $m$ of autocorrelations used is small.

### 8.1.2 Weak VARMA model case

We now repeat the same experiments on different weak VARMA$(1,1)$ models. We first assume that in (7) the innovation process $(\epsilon_t)$ is an ARCH(1) (*i.e.* $p = 0,\ q = 1$) model

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} h_{11,t} & 0 \\ 0 & h_{22,t} \end{pmatrix} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \tag{9}$$

Table 1: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA$(1, 1)$ model (7)-(8).

| | $m = 1$ | | | $m = 2$ | | | $m = 3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Length $n$ | 100 | 500 | 2,000 | 100 | 500 | 2,000 | 100 | 500 | 2,000 |
| modified **LB** | 5.5 | 5.6 | 4.0 | 3.7 | 4.4 | 4.8 | **2.6** | 4.1 | **3.3** |
| standard **LB** | **16.2** | **16.3** | **15.5** | **8.2** | **8.0** | **7.7** | **6.7** | **6.8** | 5.8 |
| | $m = 4$ | | | $m = 6$ | | | $m = 10$ | | |
| Length $n$ | 100 | 500 | 2,000 | 100 | 500 | 2,000 | 100 | 500 | 2,000 |
| modified **LB** | **2.2** | 3.9 | 4.4 | **2.3** | 3.7 | **3.4** | **6.8** | 4.2 | **3.5** |
| standard **LB** | 5.6 | 6.0 | 6.2 | 5.1 | 5.9 | 4.5 | 5.0 | 5.7 | 4.9 |

where

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

with $c_1 = 0.3$, $c_2 = 0.2$, $a_{11} = 0.45$, $a_{21} = 0.4$ and $a_{22} = 0.25$. As expected, Table 2 shows that the standard **LB** test poorly performs to assess the adequacy of this weak VARMA$(1, 1)$ model. In view of the observed relative rejection frequency, the standard **LB** test rejects very often the true VARMA$(1, 1)$ and all the relative rejection frequencies are definitely outside the significant limits. By contrast, the error of first kind is well controlled by the modified version of the **LB** test. We draw the conclusion that, at least for this particular weak VARMA model, the modified version is clearly preferable to the standard one.

In two other sets of experiments, we assume that in (7) the innovation process $(\epsilon_t)$ is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}\eta_{2,t-1}\eta_{1,t-2} \\ \eta_{2,t}\eta_{1,t-1}\eta_{2,t-2} \end{pmatrix}, \quad \text{with } \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID}\,\mathcal{N}(0, I_2), \tag{10}$$

and then by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}(|\eta_{1,t-1}| + 1)^{-1} \\ \eta_{2,t}(|\eta_{2,t-1}| + 1)^{-1} \end{pmatrix}, \quad \text{with } \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID}\,\mathcal{N}(0, I_2), \tag{11}$$

These noises are direct extensions of the weak noises defined by Romano and Thombs (1996) in the univariate case. As expected, Table 3 shows that

Table 2: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA$(1,1)$ model (7)-(9).

| | $m=1$ | | | $m=2$ | | | $m=3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Length $n$ | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 |
| modified **LB** | **6.9** | **8.5** | **7.4** | 5.9 | 6.4 | 6.3 | 4.2 | 6.1 | 5.3 |
| standard **LB** | **38.5** | **39.7** | **43.1** | **32.0** | **38.2** | **42.9** | **27.6** | **35.6** | **42.1** |
| | $m=4$ | | | $m=6$ | | | $m=10$ | | |
| Length $n$ | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 |
| modified **LB** | 3.9 | 4.8 | 5.5 | **3.3** | 3.8 | 6.0 | **2.7** | **3.5** | 3.8 |
| standard **LB** | **24.9** | **32.3** | **39.2** | **21.2** | **27.3** | **32.1** | **17.0** | **21.2** | **25.4** |

the standard **LB** test poorly performs to assess the adequacy of this weak VARMA$(1,1)$ model. In view of the observed relative rejection frequency, the standard **LB** test rejects very often the true VARMA$(1,1)$, as in Table 2. By contrast, the error of first kind is well controlled by the modified version of the **LB** test. We draw again the conclusion that, for this particular weak VARMA model, the modified version is clearly preferable to the standard one.

By contrast, Table 4 shows that the error of first kind is well controlled by all the tests in this particular weak VARMA model, except for the standard **LB** test when $m=1$. We draw the conclusion that, the modified version may be preferable to the standard one.

## 8.2 Empirical power

In this part, we simulated $N=1,000$ independent trajectories of size $n=500$, $n=1,000$ and $n=2,000$ of a weak VARMA$(2,2)$ defined by

$$\begin{aligned}\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & a_1(2,2) \end{pmatrix}\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & a_2(2,2) \end{pmatrix}\begin{pmatrix} X_{1,t-2} \\ X_{2,t-2} \end{pmatrix} \\ &\quad + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ b_1(2,1) & b_1(2,2) \end{pmatrix}\begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\ &\quad - \begin{pmatrix} 0 & 0 \\ b_2(2,1) & b_2(2,2) \end{pmatrix}\begin{pmatrix} \epsilon_{1,t-2} \\ \epsilon_{2,t-2} \end{pmatrix}, \end{aligned} \tag{12}$$

Table 3: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA$(1,1)$ model (7)-(10).

| | $m=1$ | | | $m=2$ | | | $m=3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Length $n$ | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 |
| modified **LB** | 4.7 | 3.9 | 5.3 | **3.4** | **2.8** | 4.7 | **3.1** | **2.5** | 4.7 |
| standard **LB** | **58.7** | **58.3** | **62.9** | **59.2** | **57.7** | **64.2** | **48.0** | **53.2** | **57.7** |
| | $m=4$ | | | $m=6$ | | | $m=10$ | | |
| Length $n$ | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 |
| modified **LB** | **2.2** | **2.2** | 5.3 | **1.9** | **2.0** | 4.6 | 3.6 | **3.1** | 5.3 |
| standard **LB** | **41.4** | **46.4** | **51.8** | **33.9** | **40.3** | **44.9** | **25.8** | **32.4** | **37.3** |

where the innovation process $(\epsilon_t)$ is an ARCH(1) model given by (9) and where

$$\begin{aligned} & \{a_1(2,2), a_2(2,2), b_1(2,1), b_1(2,2), b_2(2,1), b_2(2,2)\} \\ = & \ (0.225, 0.100, -0.313, 0.250, -0.140, -0.160)\,. \end{aligned}$$

For each of these $N = 1,000$ replications we fitted an VARMA$(1,1)$ model and perform standard and modified **LB** test based on $m = 1, \dots, 4$, 6 and 10 residual autocorrelations. The adequacy of the VARMA$(1,1)$ model is rejected when the $p$-value is less than 5%. Table 5 displays the relative rejection frequencies of over the $N = 1,000$ independent replications. In this example, the standard and modified versions of the **LB** test have similar powers, except for $n = 500$ and $n = 1000$. One could think that the modified version is slightly less powerful that the standard version. Actually, the comparison made in Table 5 is not fair because the actual level of the standard version is generally very greater than the 5% nominal level for this particular weak VARMA model (see Table 3).

# 9 Appendix

For the proof of Theorem 4.1, we need the following lemma on the standard matrices derivatives.

Table 4: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA$(1,1)$ model (7)-(11).

| | $m=1$ | | | $m=2$ | | | $m=3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Length $n$ | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 |
| modified **LB** | 5.0 | 5.1 | 4.3 | 5.2 | 5.0 | 5.0 | 4.3 | 5.6 | 5.4 |
| standard **LB** | **7.7** | **8.1** | 6.4 | **6.6** | 5.6 | 6.2 | 5.3 | 6.1 | 5.6 |
| | $m=4$ | | | $m=6$ | | | $m=10$ | | |
| Length $n$ | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 | 500 | 2,000 | 10,000 |
| modified **LB** | 4.2 | 5.6 | 5.2 | 3.9 | 4.4 | 4.8 | 3.8 | 4.1 | 4.9 |
| standard **LB** | 4.8 | 6.3 | 5.5 | 4.6 | 4.7 | 4.9 | 4.8 | 4.3 | 4.9 |

**Lemma 1** *If $f(A)$ is a scalar function of a matrix $A$ whose elements $a_{ij}$ are function of a variable $x$, then*

$$\frac{\partial f(A)}{\partial x} = \sum_{i,j} \frac{\partial f(A)}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial x} = Tr\left\{ \frac{\partial f(A)}{\partial A'} \frac{\partial A}{\partial x} \right\}. \tag{13}$$

*When $A$ is invertible, we also have*

$$\frac{\partial \log |\det(A)|}{\partial A'} \quad = \quad A^{-1} \tag{14}$$

**Proof of Theorem 4.1**. Recall that

$$\mathcal{O}_n(\theta) = \log\det\left\{ \frac{1}{n}\sum_{t=1}^{n} \tilde{e}_t(\theta)\tilde{e}'_t(\theta) \right\} \quad \text{and} \quad O_n(\theta) = \log\det\left\{ \frac{1}{n}\sum_{t=1}^{n} e_t(\theta)e'_t(\theta) \right\}.$$

In view of Theorem 1 in Boubacar Mainassara and Francq (2009) and **A6**, we have almost surely $\hat{\theta}_n \to \theta_0 \in \overset{\circ}{\Theta}$. Thus $\partial\mathcal{O}_n(\hat{\theta}_n)/\partial\theta = 0$ for sufficiently large $n$, and a standard Taylor expansion of the derivative of $\mathcal{O}_n$ about $\theta_0$, taken at $\hat{\theta}_n$, yields

$$\begin{aligned} 0 = \sqrt{n}\frac{\partial \mathcal{O}_n(\hat{\theta}_n)}{\partial\theta} &= \sqrt{n}\frac{\partial \mathcal{O}_n(\theta_0)}{\partial\theta} + \frac{\partial^2 \mathcal{O}_n(\theta^*)}{\partial\theta\partial\theta'}\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \\ &= \sqrt{n}\frac{\partial O_n(\theta_0)}{\partial\theta} + \frac{\partial^2 O_n(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) + o_P(1), \end{aligned} \tag{15}$$

Table 5: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(2, 2) model (7)-(10).

| | $m=1$ | | | $m=2$ | | | $m=3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Length $n$ | 500 | 1,000 | 5,000 | 500 | 1,000 | 5,000 | 500 | 1,000 | 5,000 |
| modified **LB** | 12.5 | 23.9 | 96.9 | 48.5 | 78.2 | 99.9 | 45.0 | 78.0 | 99.9 |
| standard **LB** | 94.9 | 99.8 | 100.0 | 98.1 | 99.9 | 100.0 | 97.4 | 99.9 | 100.0 |
| | $m=4$ | | | $m=6$ | | | $m=10$ | | |
| Length $n$ | 500 | 1,000 | 5,000 | 500 | 1,000 | 5,000 | 500 | 1,000 | 5,000 |
| modified **LB** | 43.5 | 77.2 | 99.9 | 39.2 | 77.6 | 99.9 | 31.0 | 74.0 | 100.0 |
| standard **LB** | 96.2 | 99.8 | 100.0 | 92.0 | 99.7 | 100.0 | 85.4 | 99.1 | 100.0 |

using arguments given in FZ (proof of Theorem 2), where $\theta^*$ is between $\theta_0$ and $\theta_n$. Thus, by standard arguments, we have from (15):

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) &= -J^{-1}\sqrt{n}\frac{\partial O_n(\theta_0)}{\partial\theta}+o_P(1)\\ &= J^{-1}\sqrt{n}Y_n+o_P(1)\end{aligned}$$

where

$$Y_n=-\frac{\partial O_n(\theta_0)}{\partial\theta}=-\frac{\partial}{\partial\theta}\log\det\left\{\frac{1}{n}\sum_{t=1}^{n}e_t(\theta_0)e_t'(\theta_0)\right\}. \tag{16}$$

Showing that the initial values are asymptotically negligible, and using (13) and (14), we have

$$\frac{\partial O_n(\theta)}{\partial\theta_i}=\mathrm{Tr}\left\{\frac{\partial\log|\Sigma_n|}{\partial\Sigma_n}\frac{\partial\Sigma_n}{\partial\theta_i}\right\}=\mathrm{Tr}\left\{\Sigma_n^{-1}\frac{\partial\Sigma_n}{\partial\theta_i}\right\}, \tag{17}$$

with

$$\frac{\partial\Sigma_n}{\partial\theta_i}=\frac{2}{n}\sum_{t=1}^{n}e_t(\theta)\frac{\partial e_t'(\theta)}{\partial\theta_i}.$$

Then, for $1\leq i\leq k_0=(p+q+2)d^2$, the $i$-th coordinate of the vector $Y_n$ is of the form

$$Y_n^{(i)}=-\mathrm{Tr}\left\{\frac{2}{n}\sum_{t=1}^{n}\Sigma_{e0}^{-1}e_t(\theta_0)\frac{\partial e_t'(\theta_0)}{\partial\theta_i}\right\},\qquad \Sigma_{e0}=\Sigma_n(\theta_0).$$

It is easily shown that for $\ell, \ell' \geq 1$,

$$\begin{aligned} \mathrm{Cov}(\sqrt{n}\,\mathrm{vec}\,\gamma(\ell), \sqrt{n}\,\mathrm{vec}\,\gamma(\ell')) &= \frac{1}{n}\sum_{t=\ell+1}^{n}\sum_{t'=\ell'+1}^{n} E\left(\{e_{t-\ell}\otimes e_t\}\{e_{t'-\ell'}\otimes e_{t'}\}'\right) \\ &\to \Gamma(\ell,\ell') \quad \text{as} \quad n\to\infty, \end{aligned}$$

Then, we have

$$\Sigma_{\gamma_m} = \{\Gamma(\ell,\ell')\}_{1\leq \ell,\ell'\leq m}$$

$$\begin{aligned} \lim_{n\to\infty}\mathrm{Cov}(\sqrt{n}J^{-1}Y_n, \sqrt{n}\,\mathrm{vec}\,\gamma(\ell)) &= -\sum_{t=\ell+1}^{n} J^{-1}E\left(\frac{\partial O_n}{\partial\theta}\{e_{t-\ell}\otimes e_t\}'\right) \\ &\to -\sum_{h=-\infty}^{+\infty} 2J^{-1}E\left(\mathcal{E}_t\{e_{t-h-\ell}\otimes e_{t-h}\}'\right), \end{aligned}$$

where $\mathcal{E}_t = \left(\left(\mathrm{Tr}\left\{\Sigma_{e0}^{-1}e_t(\theta_0)\frac{\partial e_t'(\theta_0)}{\partial\theta_1}\right\}\right)', \dots, \left(\mathrm{Tr}\left\{\Sigma_{e0}^{-1}e_t(\theta_0)\frac{\partial e_t'(\theta_0)}{\partial\theta_{k_0}}\right\}\right)'\right)'$. Then, we have

$$\begin{aligned} \lim_{n\to\infty}\mathrm{Cov}(\sqrt{n}J^{-1}Y_n, \sqrt{n}\gamma_m) &\to -\sum_{h=-\infty}^{+\infty} 2J^{-1}E\left\{\mathcal{E}_t\left(\begin{pmatrix} e_{t-h-1} \\ \vdots \\ e_{t-h-m}\end{pmatrix}\otimes e_{t-h}\right)'\right\} \\ &= \Sigma'_{\gamma_m,\hat{\theta}_n} \end{aligned}$$

Applying the central limit theorem (CLT) for mixing processes (see Herrndorf, 1984) we directly obtain

$$\begin{aligned} \lim_{n\to\infty}\mathrm{Var}(\sqrt{n}J^{-1}Y_n) &= J^{-1}IJ^{-1} \\ &= \Sigma_{\hat{\theta}_n} \end{aligned}$$

which shows the asymptotic covariance matrix of Theorem 4.1. It is clear that the existence of these matrices is ensured by the Davydov (1968) inequality. The proof is complete. □

**Proof of Theorem 5.1.** Recall that

$$e_t(\theta) = X_t - \sum_{i=1}^{\infty} C_i(\theta) X_{t-i} = \mathbf{B}_\theta^{-1}(L)\mathbf{A}_\theta(L)X_t$$

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 $\mathbf{A}_\ell = (a_{ij,\ell})$ and $\mathbf{B}_{\ell'} = (b_{ij,\ell'})$.

We define the matrices $\mathbf{A}^*_{ij,h}$ and $\mathbf{B}^*_{ij,h}$ by

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

for $h \ge 0$, where $E_{ij} = \partial \mathbf{A}_\ell / \partial a_{ij,\ell} = \partial \mathbf{B}_{\ell'} / \partial b_{ij,\ell'}$ is the $d \times d$ matrix with 1 at position $(i,j)$ and 0 elsewhere. Take $\mathbf{A}^*_{ij,h} = \mathbf{B}^*_{ij,h} = 0$ when $h < 0$. For any $a_{ij,\ell}$ and $b_{ij,\ell'}$ writing respectively the multivariate noise derivatives

$$\frac{\partial e_t}{\partial a_{ij,\ell}} = -\mathbf{B}_\theta^{-1}(L)E_{ij}X_{t-\ell} = -\sum_{h=0}^{\infty} \mathbf{A}^*_{ij,h} X_{t-h-\ell} \tag{18}$$

and

$$\frac{\partial e_t}{\partial b_{ij,\ell'}} = \mathbf{B}_\theta^{-1}(L)E_{ij}\mathbf{B}_\theta^{-1}(L)\mathbf{A}_\theta(L)X_{t-\ell'} = \sum_{h=0}^{\infty} \mathbf{B}^*_{ij,h} X_{t-h-\ell'}. \tag{19}$$

On the other hand, considering $\hat{\Gamma}(h)$ and $\gamma(h)$ as values of the same function at the points $\hat{\theta}_n$ and $\theta_0$, a Taylor expansion about $\theta_0$ gives

$$\begin{aligned}
\operatorname{vec} \hat{\Gamma}_e(h) &= \operatorname{vec} \gamma(h) + \frac{1}{n} \sum_{t=h+1}^{n} \left\{ e_{t-h}(\theta) \otimes \frac{\partial e_t(\theta)}{\partial \theta'} \right. \\
&\qquad \left. + \frac{\partial e_{t-h}(\theta)}{\partial \theta'} \otimes e_t(\theta) \right\}_{\theta = \theta_n^*} (\hat{\theta}_n - \theta_0) + O_P(1/n) \\
&= \operatorname{vec} \gamma(h) + E\left( e_{t-h}(\theta_0) \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right) (\hat{\theta}_n - \theta_0) + O_P(1/n),
\end{aligned}$$

where $\theta_n^*$ is between $\hat{\theta}_n$ and $\theta_0$. The last equality follows from the consistency of $\hat{\theta}_n$ and the fact that $(\partial e_{t-h}/\partial \theta')(\theta_0)$ is not correlated with $e_t$ when $h \ge 0$. Then for $h = 1, \dots, m$,

$$\hat{\Gamma}_m := \left( \left\{ \operatorname{vec}\hat{\Gamma}_e(1) \right\}', \dots, \left\{ \operatorname{vec}\hat{\Gamma}_e(m) \right\}' \right)' = \gamma_m + \Phi_m(\hat{\theta}_n - \theta_0) + O_P(1/n),$$

where

$$\Phi_m = E\left\{\begin{pmatrix} e_{t-1} \\ \vdots \\ e_{t-m} \end{pmatrix} \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'}\right\}. \tag{20}$$

In $\Phi_m$, one can express $(\partial e_t/\partial\theta')(\theta_0)$ in terms of the multivariate derivatives (18) and (19). From Theorem 4.1, we have obtained the asymptotic joint distribution of $\gamma_m$ and $\hat{\theta}_n - \theta_0$, which shows that the asymptotic distribution of $\sqrt{n}\hat{\Gamma}_m$, is normal, with mean zero and covariance matrix

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{Var}(\sqrt{n}\hat{\Gamma}_m) &= \lim_{n\to\infty} \mathrm{Var}(\sqrt{n}\gamma_m) + \Phi_m \lim_{n\to\infty} \mathrm{Var}(\sqrt{n}(\hat{\theta}_n - \theta_0))\Phi_m' \\
&\quad + \Phi_m \lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}(\hat{\theta}_n - \theta_0), \sqrt{n}\gamma_m) \\
&\quad + \lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\gamma_m, \sqrt{n}(\hat{\theta}_n - \theta_0))\Phi_m' \\
&= \Sigma_{\gamma_m} + \Phi_m \Sigma_{\hat{\theta}_n} \Phi_m' + \Phi_m \Sigma_{\hat{\theta}_n,\gamma_m} + \Sigma'_{\hat{\theta}_n,\gamma_m} \Phi_m'.
\end{aligned}$$

From a Taylor expansion about $\theta_0$ of $\mathrm{vec}\,\hat{\Gamma}_e(0)$ we have, $\mathrm{vec}\,\hat{\Gamma}_e(0) = \mathrm{vec}\,\gamma(0) + O_P(n^{-1/2})$. Moreover, $\sqrt{n}(\mathrm{vec}\,\gamma(0) - E\,\mathrm{vec}\,\gamma(0)) = O_P(1)$ by the CLT for mixing processes. Thus $\sqrt{n}(\hat{S}_e \otimes \hat{S}_e - S_e \otimes S_e) = O_P(1)$ and, using (4) and the ergodic theorem, we obtain

$$\begin{aligned}
& n\left\{\mathrm{vec}(\hat{S}_e^{-1}\hat{\Gamma}_e(h)\hat{S}_e^{-1}) - \mathrm{vec}(S_e^{-1}\hat{\Gamma}_e(h)S_e^{-1})\right\} \\
=\ & n\left\{(\hat{S}_e^{-1} \otimes \hat{S}_e^{-1})\,\mathrm{vec}\,\hat{\Gamma}_e(h) - (S_e^{-1} \otimes S_e^{-1})\,\mathrm{vec}\,\hat{\Gamma}_e(h)\right\} \\
=\ & n\left\{(\hat{S}_e \otimes \hat{S}_e)^{-1}\,\mathrm{vec}\,\hat{\Gamma}_e(h) - (S_e \otimes S_e)^{-1}\,\mathrm{vec}\,\hat{\Gamma}_e(h)\right\} \\
=\ & (\hat{S}_e \otimes \hat{S}_e)^{-1}\sqrt{n}(S_e \otimes S_e - \hat{S}_e \otimes \hat{S}_e)(S_e \otimes S_e)^{-1}\sqrt{n}\,\mathrm{vec}\,\hat{\Gamma}_e(h) \\
=\ & O_P(1).
\end{aligned}$$

In the previous equalities we also use $\mathrm{vec}(ABC) = (C' \otimes A)\,\mathrm{vec}(B)$ *and* $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ when A and B are invertible. It follows that

$$\begin{aligned}
\hat{\rho}_m &= \left(\left\{\mathrm{vec}\hat{R}_e(1)\right\}', \dots, \left\{\mathrm{vec}\hat{R}_e(m)\right\}'\right)' \\
&= \left(\left\{(\hat{S}_e \otimes \hat{S}_e)^{-1}\mathrm{vec}\hat{\Gamma}_e(1)\right\}', \dots, \left\{(\hat{S}_e \otimes \hat{S}_e)^{-1}\mathrm{vec}\hat{\Gamma}_e(m)\right\}'\right)' \\
&= \left\{I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1}\right\}\hat{\Gamma}_m = \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\}\hat{\Gamma}_m + O_P(n^{-1}).
\end{aligned}$$

We now obtain (5) from (4). Hence, we have

$$\mathrm{Var}(\sqrt{n}\hat{\rho}_m) = \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\} \Sigma_{\hat{\Gamma}_m} \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\}.$$

The proof is complete. □

**Proof of Theorem 6.2.** The proof is similar to that given by Francq, Roy and Zakoïan (2005) for Theorem 5.1. □

# References

**Ahn, S. K.** (1988) Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika* 75, 590–93.

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

**Arbués, I.** (2008) An extended portmanteau test for VARMA models with mixing nonlinear constraints, *Journal of Time Series Analysis* 29, 741–761.

**Boubacar Mainassara, Y. and Francq, C.** (2009) Estimating structural VARMA models with uncorrelated but non-independent error terms. *Working Papers,* http://perso.univ-lille3.fr/ cfrancq/pub.html.

**Box, G. E. P. and Pierce, D. A.** (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–26.

**Brockwell, P. J. and Davis, R. A.** (1991) *Time series: theory and methods.* Springer Verlag, New York.

**Chabot-Hallé, D. and Duchesne, P.** (2008) Diagnostic checking of multivariate nonlinear time series models with martingale difference errors, *Statistics and Probability Letters* 78, 997–1005.

**Chitturi, R. V.** (1974) Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association* 69, 928–934.

**Davydov, Y. A.** (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Applications* 13, 691–696.

**den Hann, W.J. and Levin, A.** (1997) A Practitioner's Guide to Robust Covariance Matrix Estimation. *In Handbook of Statistics* 15, Rao, C.R. and G.S. Maddala (eds), 291–341.

**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.*

**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.

**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.

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

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

**Herrndorf, N.** (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.

**Hosking, J. R. M.** (1980) The multivariate portmanteau statistic, *Journal of the American Statistical Association* 75, 602–608.

**Hosking, J. R. M.** (1981a) Equivalent forms of the multivariate portmanteau statistic, *Journal of the Royal Statistical Society* B 43, 261–262.

**Hosking, J. R. M.** (1981b) Lagrange-tests of multivariate time series models, *Journal of the Royal Statistical Society* B 43, 219–230.

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

**Li, W. K. and McLeod, A. I.** (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models, *Journal of the Royal Statistical Society* B 43, 231–239.

**Lütkepohl, H.** (1993) *Introduction to multiple time series analysis.* Springer Verlag, Berlin.

**Lütkepohl, H.** (2005) *New introduction to multiple time series analysis.* Springer Verlag, Berlin.

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

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

**Reinsel, G. C.** (1997) *Elements of multivariate time series Analysis.* Second edition. Springer Verlag, New York.

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