MULTIVARIATE PORTMANTEAU TEST FOR AUTOREGRESSIVE MODELS WITH UNCORRELATED BUT NONINDEPENDENT ERRORS

CHRISTIAN FRANCQ ,* *GREMARS Université Lille 3*
HAMDI RAÏSSI,** *GREMARS Université Lille 3*

Abstract

In this paper we consider estimation and test of fit for multiple autoregressive time series models with nonindependent innovations. We derive the asymptotic distribution of the residual autocorrelations. It is shown that this asymptotic distribution can be quite different for models with iid innovations and models in which the innovations exhibit conditional heteroscedasticity or other forms of dependence. Consequently, the usual chi-square distribution does not provide adequate approximation to the distribution of the Box-Pierce goodness-of-fit portmanteau test in the presence of nonindependent innovations. We then 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: Vector weak AR model, Goodness-of-fit test, Residual autocorrelation, Diagnostic Checking, Box-Pierce and Ljung-Box portmanteau tests.

1. Introduction

In multivariate time series analysis, the Vectorial AutoRegressive (VAR) models are much employed (see Lütkepohl (1993)). The VAR models postulate that the d-dimensional vector X_t of the observations at time t can be represented as a linear combination of p past values $X_{t-1}, \ldots, X_{t-p}$ plus an error ϵ_t. A theoretical argument in favor of the VAR models is that any stationary process can be approximated by a VAR(p) model with sufficiently large p and uncorrelated errors. The reason of the success of these models is however of practical nature, and is certainly due to the fact that it is relatively easy to deal with a linear function of a finite number of past values.

It is however clear that the VAR models are not universal and that the choice of the order p is crucial. Thus it is important to check the validity of a VAR(p) model, for a given order p. In multivariate, the choice of p is particularly important because the number of parameters, pd^2, quickly increases with p, which entails statistical difficulties. This paper is devoted to the so-called portmanteau tests considered by Chitturi (1974) and Hosking (1980) for checking the overall significance of the residual autocorrelations of a VAR(p) model (see also Ahn (1988), Hosking (1981a, 1981b), Li and McLeod (1981)).

* Postal address: GREMARS, UFR MSES, Université Lille 3, Domaine du Pont de bois, BP 149, 59653 Villeneuve d'Ascq Cedex, France
** Postal address: GREMARS, UFR MSES, Université Lille 3, Domaine du Pont de bois, BP 149, 59653 Villeneuve d'Ascq Cedex, France

For the statistical analysis of VAR models, the errors ϵ_t are generally supposed to be independent (as in Lütkepohl (1993), Definition 3.1). This independence assumption is restrictive because it precludes conditional heteroscedasticity and/or other forms of nonlinearity (see Francq, Roy and Zakoïan (2005) and Francq and Zakoïan (2005) for the statistical inference of univariate ARMA models with uncorrelated but nonindependent errors, and see Dufour and Pelletier (2005) for weak VARMA modelling). The main goal of the present paper is to study the behaviour of the above-mentioned goodness-of-fit portmanteau tests when the ϵ_t's are not independent. We will see that the results obtained by the standard portmanteau tests can be quite misleading in this framework. A modified version of these tests is proposed.

Brüggemann, Lütkepohl and Saikkonen (2004) is an important recent reference dealing with portmanteau tests and other tests for residual autocorrelation in VAR models with iid innovations when some variables are cointegrated. In this paper we do not consider cointegrated variables, but the independence assumption of the ϵ_t's is relaxed.

The rest of the paper is organized as follows. Section 2 provides examples of AR models with uncorrelated but nonindependent errors, presents several expressions for the least squares (LS) estimator of the AR coefficients, and gives conditions ensuring the consistency and asymptotic normality of the LS estimator. Section 3 studies the asymptotic behaviour of the AR residuals. We obtain the asymptotic distribution of vectors of residual autocorrelations, under the assumption of the fitted AR(p) is an adequate linear model with a well chosen order p. The results are applied in Section 4 to obtain the asymptotic distribution of the portmanteau tests and to modify the critical values of these tests when they are applied to VAR models with nonindependent errors. Section 5 is devoted to the practical implementation of the modified version of the tests. Section 6 proposes numerical illustrations. The technical proofs are relegated to the appendix.

The following notations will be used throughout. For a matrix A of generic term $A(i,j)$ we use the norm $\|A\| = \sum |A(i,j)|$. The spectral radius of a square matrix A is denoted by $\rho(A)$, its trace is denoted by Tr. We denote by $A \otimes B$ the Kronecker product of two matrices A and B, vec A denotes the vector obtained by stacking the columns of A, and $A^{\otimes 2}$ stands for $A \otimes A$ (see *e.g.* Harville (1997) for more details about these matrix operators). The symbol $\Rightarrow$ denotes the convergence in distribution.

2. LS estimator of weak VAR models

Consider the vectorial AR(p) model

$$X_t = \sum_{i=1}^{p} A_{0i} X_{t-i} + \epsilon_t \qquad \text{for all } t \in \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\} \tag{2.1}$$

where the ϵ_t's are d-dimensional error terms, the X_t's are d-dimensional vectors, and the A_{0i}'s are $d \times d$ matrices. It is customary to say that (X_t) is a strong AR(p) model if (ϵ_t) is a strong white noise, that is, if it satisfies

The authors are grateful to the Professor Lütkepohl whose questions he asked during the Madrid EEA-ESEM 2004 congress motivated the present paper.

A1: (ϵ_t) is a sequence of independent and identically distributed (iid) random vectors, $E\epsilon_t = 0$ and $\mathrm{Var}\,(\epsilon_t) = \Sigma_\epsilon$.

We say that (2.1) is a weak AR(p) model if (ϵ_t) is a weak white noise, that is, if it satisfies

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

Assumption **A1** is clearly stronger than **A1'**. The class of strong AR models is often considered too restrictive by practitioners. Indeed Assumption **A1** amounts to assume that the best predictor of X_t is a linear combination of its p past values. If p is chosen large enough, it is reasonable to consider that the best predictor of X_t is well approximated by a function of $X_{t-1}, \dots, X_{t-p}$, but it is questionable to assume a linear form for this function. It is well known that numerous non linear processes admit weak linear representations (see Example 2.1 below). Weak linear representations are also obtained from transformations of strong linear processes (see Example 2.2 below). In Examples 2.1-2.3 below, Assumption **A1'** holds but **A1** is not satisfied. Other examples of univariate weak linear models can be found in Francq, Roy and Zakoïan (2005), and references therein.

For the statistical analysis of multivariate AR time series models, it is therefore of interest to replace the standard strong white noise assumption **A1** by the more flexible weak white noise assumption **A1'**. In this paper we focus on the estimation and validation stages of the statistical analysis of these weak multivariate AR models.

2.1. Examples of weak VAR models

The examples given in this section are mainly chosen for their simplicity. The first one is a weak white noise inspired by examples given by Romano and Thombs (1996) in the univariate case. The second is simply the causal representation of a non causal AR(1) process. This example shows that **A1** must by replaced by **A1'** when one wants to make, without loss of generality, the usual assumption that the roots of the AR polynomial are outside the unit circle. The third belongs to the class of GARCH models.

Example 2.1. In the univariate case, Romano and Thombs (1996) built weak white noises (ϵ_t) by setting $\epsilon_t = \eta_t \eta_{t-1} \cdots \eta_{t-k}$ where (η_t) is a strong white noise and $k \geq 1$. This approach can be extended to the d-multivariate framework by setting $\epsilon_t = B(\eta_t) \cdots B(\eta_{t-k+1}) \eta_{t-k}$, where $\{\eta_t = (\eta_{1\,t}, \dots, \eta_{d\,t})'\}_t$ is a d-dimensional strong white noise, and $B(\eta_t) = \{B_{ij}(\eta_t)\}$ is a $d \times d$ random matrix whose elements are linear combinations of the components of η_t. It is obvious to check that ϵ_t is a white noise, but in general this noise is not a strong one. Indeed, assuming for simplicity that $k = 1$ and $B_{11}(\eta_t) = \eta_{1\,t}$ and $B_{1j}(\eta_t) = 0$ for all $j > 1$, we have

$$\mathrm{Cov}\left(\epsilon_{1\,t}^2, \epsilon_{1\,t-1}^2\right) \quad = \quad \left\{E\eta_{1\,t}^2\right\}^2 \mathrm{Var}\left\{\eta_{1\,t}^2\right\} \neq 0,$$

which shows that the ϵ_t's are not independent.

Example 2.2. (Non causal AR(1)) Let the AR(1) model

$$X_t = AX_{t-1} + \epsilon_t, \quad \epsilon_t \quad \text{iid and} \quad E(\epsilon_t) = 0, \quad E(\epsilon_t \epsilon_t') = \Sigma_\epsilon,$$

where A is an invertible matrix whose all the eigenvalues $\lambda_i, 1 \leq i \leq d$, are greater than one in modulus. This equation has a stationary and anticipative solution of the form $X_t = -\sum_{i=1}^{\infty} A^{-i}\epsilon_{t+i}$. The autocovariance function of (X_t) is then given by

$$\Gamma_X(h) = A^{-h}\sum_{i=1}^{\infty} A^{-i}\Sigma_\epsilon A^{-i}, \qquad h \geq 0.$$

Let $\epsilon_t^* = X_t - A^{-1}X_{t-1}$. We have $E(\epsilon_t^*) = 0$, $\mathrm{Var}(\epsilon_t^*) = \Gamma_X(0) + \Gamma_X(1)(A^{-1})' + A^{-1}\Gamma_X(1) + A^{-1}\Gamma_X(1)(A^{-1})'$, and $\mathrm{Cov}(\epsilon_t^*, \epsilon_{t-h}^*) = 0$ for $h \neq 0$. Thus X_t admits the causal AR(1) representation $X_t = A^{-1}X_{t-1} + \epsilon_t^*$. However, in general, ϵ_t^* is not a martingale difference. To see this, assume for simplicity that the matrix A is such that $|a_{11}| > 1$ and $a_{1j} = a_{j1} = 0, \quad \forall j \in \{2, \cdots, d\}$. We then have $EX_{1\,t-1}^3 = (1 - a_{11}^3)^{-1}E\epsilon_{1\,t}^3, \quad EX_{1\,t}X_{1\,t-1}^2 = a_{11}^{-2}(1-a_{11}^3)^{-1}E\epsilon_{1\,t}^3$ and $E(\epsilon_{1\,t}^*X_{1\,t-1}^2) = EX_{1\,t}X_{1\,t-1}^2 - a_{11}^{-1}EX_{1\,t-1}^3 \neq 0$ when $E\epsilon_{1\,t}^3 \neq 0$. In this case ϵ_t^* is not a martingale difference because $E(\epsilon_{1\,t}^*X_{1\,t-1}^2) = E\left\{X_{1\,t-1}^2 E\left(\epsilon_{1\,t}^*|\epsilon_{t-1}^*, \dots\right)\right\} \neq 0$. Thus the white noise ϵ_t^* is not strong.

Example 2.3. In the univariate case, the GARCH models constitute important examples of weak white noises. These models have numerous extensions to the multivariate framework. The simplest of these extensions is certainly the multivariate GARCH model with constant correlation proposed by Jeantheau (1998). In this model the process (ϵ_t) verifies the following relation $\epsilon_t = \Delta_t\eta_t$ where $\eta_t = (\eta_{1\,t}, \dots, \eta_{d\,t})'$ is an iid centered process with $\mathrm{Var}(\eta_{i\,t}) = 1$, and Δ_t is a diagonal matrix whose elements $\sigma_{ii\,t}$ verify

$$\begin{pmatrix} \sigma_{11\,t}^2 \\ \vdots \\ \sigma_{dd\,t}^2 \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} + \sum_{i=1}^{q} A_i \begin{pmatrix} \epsilon_{1\,t-i}^2 \\ \vdots \\ \epsilon_{d\,t-i}^2 \end{pmatrix} + \sum_{j=1}^{p} B_j \begin{pmatrix} \sigma_{11\,t-j}^2 \\ \vdots \\ \sigma_{dd\,t-j}^2 \end{pmatrix}.$$

The elements of the matrices A_i and B_j, as well as the c_i's, are supposed to be positive. In addition suppose that the stationarity conditions hold (see Jeantheau (1998) for more details). The stationary solution of this GARCH equation satisfies **A1**$'$, but does not satisfy **A1** in general. Consider for instance the ARCH(1) case with A_1 such that $a_{11} \neq 0$ and $a_{12} = \cdots = a_{1n} = 0$. Then it is easy to see that $\mathrm{Cov}(\epsilon_{1\,t}^2, \epsilon_{1\,t-1}^2) = \mathrm{Cov}\left\{(c_1 + a_{11}\epsilon_{1\,t-1}^2)\eta_{1\,t}^2, \epsilon_{1\,t-1}^2\right\} = a_{11}\mathrm{Var}\epsilon_{1\,t}^2 \neq 0$, which shows that the iid assumption **A1** is not satisfied.

2.2. Derivation of the LS estimator

It is well known that (X_t) can be written as a MA(∞) of the form

$$X_t = \sum_{i=0}^{\infty}\psi_{0i}\epsilon_{t-i}, \quad \psi_{00} = I_d, \quad \sum_{i=0}^{\infty}\|\psi_{0i}\| < \infty, \tag{2.2}$$

under the assumption

A2: $\det A_0(z) \neq 0$ for all $|z| \leq 1$, where $A_0(z) = I_d - \sum_{i=1}^{p} A_{0i}z^i$.

Denote by $\theta_0 = \mathrm{vec}\ (A_{01} \cdots A_{0p})$ the vector of the unknown AR parameters. For any $\theta \in \mathbb{R}^{d^2p}$, let $A_1 = A_1(\theta), \dots, A_p = A_p(\theta)$ be $d \times d$ matrices such that $\theta = \mathrm{vec}\ (A_1, \dots, A_p)$. With this notation (2.1) can be rewritten as

$$X_t = \left\{\left(X'_{t-1}, \dots, X'_{t-p}\right) \otimes I_d\right\}\theta_0 + \epsilon_t \tag{2.3}$$

using the elementary relation vec $(ABC) = (C' \otimes A)$vec B for matrices of appropriate dimensions.

One of the most popular estimation procedure is that of the least squares (LS) estimator. For linear processes of the form (2.1), the LS estimator of θ coincides with the gaussian quasi-maximum likelihood (QML) estimator. Given the observations $X_1, \dots, X_n$, the LS/QML estimators of θ and Σ_ϵ are defined by

$$(\hat{\theta}_n, \hat{\Sigma}_\epsilon) = \arg\min_{\theta, \Sigma_\epsilon} \left\{ n \log(\det \Sigma_\epsilon) + \sum_{t=1}^{n} \epsilon_t'(\theta) \Sigma_\epsilon^{-1} \epsilon_t(\theta) \right\}$$

where

$$\epsilon_t(\theta) = X_t - \sum_{i=1}^{p} A_i X_{t-i}.$$

To give an explicit expression for these estimators, the d-dimensional AR(p) model (2.1) can be rewritten as the dp-dimensional AR(1) model

$$\tilde{X}_t = \tilde{A}_0 \tilde{X}_{t-1} + \tilde{\epsilon}_t, \tag{2.4}$$

where

$$\tilde{A}_0 = \begin{pmatrix} A_{01} & \cdots & A_{0p-1} & A_{0p} \\ I_d & 0 & & \\ & \ddots & & \\ & 0 & I_d & 0 \end{pmatrix}, \quad \tilde{X}_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix}, \quad \tilde{\epsilon}_t = \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Note that **A2** is equivalent to $\rho\left(\tilde{A}_0\right) < 1$. Let $\hat{\Sigma}_{\tilde{X}_t, \tilde{X}_{t-h}} = \frac{1}{n} \sum_{t=1}^{n} \tilde{X}_t \tilde{X}_{t-h}'$ and write $\hat{\Sigma}_{\tilde{X}_{t-1}}$ instead of $\hat{\Sigma}_{\tilde{X}_{t-1}, \tilde{X}_{t-1}}$. Note that $\hat{\Sigma}_{\tilde{X}_{t-1}}$ is a consistent estimator of $\Sigma_{\tilde{X}_t} = E\tilde{X}_t \tilde{X}_t'$, which is given by

$$\text{vec}\left(\Sigma_{\tilde{X}_t}\right) = \left(I_{(dp)^2} - \tilde{A}_0 \otimes \tilde{A}_0\right)^{-1} \text{vec}\left(\Sigma_{\tilde{\epsilon}_t}\right), \qquad \Sigma_{\tilde{\epsilon}_t} = \begin{pmatrix} \Sigma_\epsilon & 0_{d(p-1)}' \\ 0_{d(p-1)} & 0_{d(p-1) \times d(p-1)} \end{pmatrix}.$$

It is easy to see that the LS estimators of the AR parameters of models (2.1) and (2.4) are given by

$$\widehat{\tilde{A}} = \begin{pmatrix} \widehat{A}_1 & \cdots & \widehat{A}_{p-1} & \widehat{A}_p \\ I_d & 0 & & \\ & \ddots & & \\ & 0 & I_d & 0 \end{pmatrix} = \hat{\Sigma}_{\tilde{X}_t, \tilde{X}_{t-1}} \hat{\Sigma}_{\tilde{X}_{t-1}}^{-1}, \tag{2.5}$$

provided $\hat{\Sigma}_{\tilde{X}_{t-1}}$ is non singular, and that the LS estimators of the variances of the noises $(\tilde{\epsilon}_t)$ and (ϵ_t) are given by

$$\widehat{\Sigma}_{\tilde{\epsilon}} = \begin{pmatrix} \widehat{\Sigma}_\epsilon & 0_{d(p-1)}' \\ 0_{d(p-1)} & 0_{d(p-1) \times d(p-1)} \end{pmatrix} = \hat{\Sigma}_{\tilde{X}_t} - \hat{\Sigma}_{\tilde{X}_t, \tilde{X}_{t-1}} \hat{\Sigma}_{\tilde{X}_{t-1}}^{-1} \hat{\Sigma}_{\tilde{X}_{t-1}, \tilde{X}_t}, \tag{2.6}$$

with the convention that $X_t = 0$ when $t \leq 0$ or $t > n$.

2.3. Asymptotic behavior of the LS estimator

To establish the strong consistency of the estimators defined in (2.5) and (2.6), we need the following assumptions.

A3: Matrix Σ_ϵ is positive definite.

A4: The sequence (ϵ_t) is strictly stationary and ergodic.

Note that **A4** is entailed by **A1**, but not by **A1′**. A straightforward consequence of the ergodic theorem is stated in the following proposition.

Proposition 2.1. *Under assumptions* **A1**-**A2**-**A3** *or* **A1**'-**A2**-**A3**-**A4**, *the matrix* $\hat{\Sigma}_{\tilde{X}_t}$ *is almost surely non singular, and almost surely*

$$\widehat{\tilde{A}} \to \tilde{A}_0, \quad \widehat{\Sigma}_{\tilde{\epsilon}} \to \Sigma_{\tilde{\epsilon}} \quad \text{and} \quad \hat{\theta}_n \to \theta_0$$

as $n \to \infty$.

To obtain the asymptotic normality of the LS estimator of θ, additional assumptions are needed when (ϵ_t) is not iid. In the univariate case, Francq and Zakoïan (1998) showed the asymptotic normality under mixing assumptions. We will extend this result to VAR models. The mixing coefficients of a stationary process $X = (X_t)$ are denoted by

$$\alpha_X(h) = \sup_{A\in\sigma(X_u, u\le t), B\in\sigma(X_u, u\ge t+h)} |P(A\cap B) - P(A)P(B)|\,.$$

The reader is referred to Davidson (1994) for details about mixing assumptions. Let $\|X\|_r = (E\|X\|^r)^{1/r}$, where $\|X\|$ denotes the Euclidean norm.

A5: The process $X = (X_t)$ is such that $\sum_{h=0}^{\infty}\{\alpha_X(h)\}^{\nu/(2+\nu)} < \infty$ and $\|X_t\|_{4+2\nu} < \infty$ for some $\nu > 0$.

The asymptotic distribution of the LS estimator is given in the following proposition.

Proposition 2.2. *Under assumptions* **A1**–**A3** *or under* **A1**' *and* **A2**–**A5**,

$$\sqrt{n} vec\left(\widehat{\tilde{A}} - \tilde{A}_0\right) \Rightarrow \mathcal{N}(0, \Omega), \tag{2.7}$$

where

$$\Omega = \sum_{h=-\infty}^{\infty} E\left\{\Sigma_{\tilde{X}_t}^{-1}\tilde{X}_{t-1}\tilde{X}'_{t-h-1}\Sigma_{\tilde{X}_t}^{-1} \otimes \tilde{\epsilon}_t\tilde{\epsilon}'_{t-h}\right\}. \tag{2.8}$$

Moreover

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \Rightarrow \mathcal{N}\left(0, \Sigma_{\hat{\theta}_n}\right), \tag{2.9}$$

where

$$\Sigma_{\hat{\theta}_n} = \sum_{h=-\infty}^{\infty} E\left\{\Sigma_{\tilde{X}_t}^{-1}\tilde{X}_{t-1}\tilde{X}'_{t-h-1}\Sigma_{\tilde{X}_t}^{-1} \otimes \epsilon_t\epsilon'_{t-h}\right\}. \tag{2.10}$$

Note that under **A1**, we have

$$\Omega = \Sigma_{\tilde{X}_t}^{-1} \otimes \Sigma_{\tilde{\epsilon}_t}.$$

and

$$\Sigma_{\hat{\theta}_n} = \Sigma_{\tilde{X}_t}^{-1} \otimes \Sigma_{\epsilon_t}. \tag{2.11}$$

This standard result can be found in Johansen (1995, Theorem 2.3 p. 19).

Example 2.4. The following example illustrates the difference between the asymptotic variance (2.10) in the weak case and the asymptotic variance (2.11) of the strong AR case. Consider a bivariate AR(1) model $X_t = AX_{t-1} + \epsilon_t$, with true AR(1) parameter $A_0 = 0$ and $\epsilon_t = B(\eta_t)\dots B(\eta_{t-k+1})\eta_{t-k}$ of the form given in Example 2.1. We have $X_t = \tilde{X}_t = \epsilon_t$. For simplicity assume that $B(\eta_t) = \mathrm{Diag}\,(\eta_{1\,t}, \eta_{2\,t})$ and that η is gaussian with $\mathrm{Var}(\eta_{1\,t}) = \mathrm{Var}(\eta_{2\,t}) = 1$ and $\mathrm{Cov}(\eta_{1\,t}, \eta_{2\,t}) = \rho \in (-1, 1)$. Thus, using $E\eta_{1\,t}^3\eta_{2\,t} = 3\rho$ and $E\eta_{1\,t}^2\eta_{2\,t}^2 = 1 + 2\rho^2$, we find that $\mathrm{Var}(\epsilon_{1\,t}) = \mathrm{Var}(\epsilon_{2\,t}) = 1$, $\mathrm{Cov}(\epsilon_{1\,t}, \epsilon_{2\,t}) = \rho^{k+1}$, and, using (2.10),

$$\begin{aligned}
\Sigma_{\hat{\theta}_n} &= E\left\{\Sigma_{\epsilon_t}^{-1}\epsilon_{t-1}\epsilon_{t-1}'\Sigma_{\epsilon_t}^{-1} \otimes \epsilon_t\epsilon_t'\right\} \\
&= \left(\Sigma_{\epsilon_t}^{-1} \otimes I_2\right) E\left\{\left(\epsilon_{t-1} \otimes \epsilon_t\right)\left(\epsilon_{t-1} \otimes \epsilon_t\right)'\right\}\left(\Sigma_{\epsilon_t}^{-1} \otimes I_2\right),
\end{aligned}$$

with $\Sigma_\epsilon = \begin{pmatrix} 1 & \rho^{k+1} \\ \rho^{k+1} & 1 \end{pmatrix}$ and $E\left\{\left(\epsilon_{t-1} \otimes \epsilon_t\right)\left(\epsilon_{t-1} \otimes \epsilon_t\right)'\right\}$ equal to

$$\begin{pmatrix}
3^k & \rho(3\rho)^k & \rho(3\rho)^k & \rho^2(1+2\rho^2)^k \\
\rho(3\rho)^k & (1+2\rho^2)^k & \rho^2(1+2\rho^2)^k & \rho(3\rho)^k \\
\rho(3\rho)^k & \rho^2(1+2\rho^2)^k & (1+2\rho^2)^k & \rho(3\rho)^k \\
\rho^2(1+2\rho^2)^k & \rho(3\rho)^k & \rho(3\rho)^k & 3^k
\end{pmatrix}.$$

When $k = 1$ the matrix $\Sigma_{\hat{\theta}_n}$ is equal to

$$\Sigma_W := \frac{1}{(\rho^2+1)(1-\rho^4)} \times$$
$$\begin{pmatrix}
-2\rho^4+3\rho^2+3 & \rho^2(\rho^2+3) & -3\rho^4-\rho^2 & \rho^2(-2\rho^4-3\rho^2+1) \\
\rho^2(\rho^2+3) & 3\rho^2+1 & \rho^2(-2\rho^4-3\rho^2+1) & -\rho^2(3\rho^2+1) \\
-\rho^2(3\rho^2+1) & \rho^2(-2\rho^4-3\rho^2+1) & 3\rho^2+1 & \rho^2(\rho^2+3) \\
\rho^2(-2\rho^4-3\rho^2+1) & -3\rho^4-\rho^2 & \rho^2(\rho^2+3) & -2\rho^4+3\rho^2+3
\end{pmatrix}.$$

By comparison, for the strong AR(1) with an iid noise with variance

$$\Sigma_\epsilon = \begin{pmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{pmatrix},$$

the asymptotic variance $\Sigma_{\hat{\theta}_n}$ is given by

$$\Sigma_S := \frac{1}{1-\rho^4}\begin{pmatrix}
1 & \rho^2 & -\rho^2 & -\rho^4 \\
\rho^2 & 1 & -\rho^4 & -\rho^2 \\
-\rho^2 & -\rho^4 & 1 & \rho^2 \\
-\rho^4 & -\rho^2 & \rho^2 & 1
\end{pmatrix}.$$

Figure 2.1 displays the ratio $\Sigma_W(1,1)/\Sigma_S(1,1)$ as function of ρ. It is clear from this example that the asymptotic variance of the LS estimator may be quite different when the noise is weak and when it is strong.

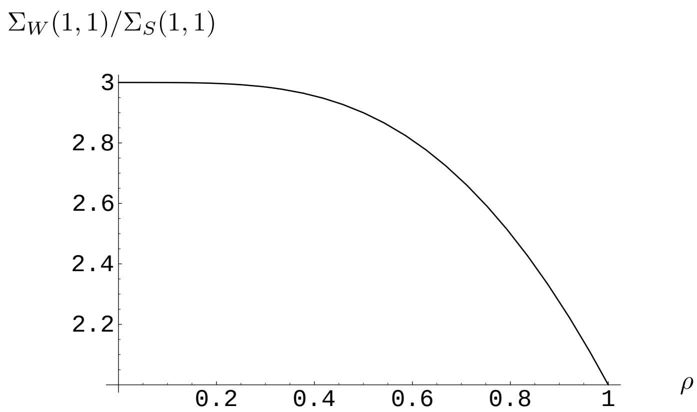

FIGURE 2.1: Asymptotic variance of the first component of LS estimator of the weak AR(1) model given in Example 2.4 divided by the analog asymptotic variance in the strong AR(1) case.

and

3. Asymptotic distribution of residual empirical autocorrelations

Let $\hat{\epsilon}_t = \epsilon_t(\hat{\theta}_n)$ be the LS residuals. We consider the white noise "empirical" autocovariances and residual autocovariances

$$C_h = \frac{1}{n}\sum_{t=h+1}^{n} \epsilon_t\epsilon'_{t-h}, \quad , \quad \hat{\Gamma}_\epsilon(h) = \frac{1}{n}\sum_{t=h+1}^{n} \hat{\epsilon}_t\hat{\epsilon}'_{t-h}, \quad 0 \le h < n. \tag{3.1}$$

It should be noted that C_h is not a computable statistic because it depends on the unobserved innovations $\epsilon_t = \epsilon_t(\theta_0)$. We will consider vectors of these autocovariances. For a fixed integer m, let

$$\hat{\gamma}_m = \left(\left\{\mathrm{vec}\,\hat{\Gamma}_\epsilon(1)\right\}', \ldots, \left\{\mathrm{vec}\,\hat{\Gamma}_\epsilon(m)\right\}'\right)' \quad \text{and} \quad c_m = \left(\{\mathrm{vec}\,C_1\}', \ldots, \{\mathrm{vec}\,C_m\}'\right)'.$$

Let the diagonal matrices

$$S_\epsilon = \mathrm{Diag}\left\{\sigma_\epsilon(1), \ldots, \sigma_\epsilon(d)\right\} \quad \text{and} \quad \hat{S}_\epsilon = \mathrm{Diag}\left\{\hat{\sigma}_\epsilon(1), \ldots, \hat{\sigma}_\epsilon(d)\right\},$$

where $\sigma_\epsilon^2(i)$ is the variance of the i-th coordinate of ϵ_t and $\hat{\sigma}_\epsilon^2(i)$ is its sample estimate, *i.e.* $\sigma_\epsilon(i) = \sqrt{E\epsilon_{i\,t}^2}$ and $\hat{\sigma}_\epsilon(i) = \sqrt{n^{-1}\sum_{t=1}^{n}\hat{\epsilon}_{i\,t}^2}$. The theoretical and sample autocorrelations at lag h are respectively defined by $R_\epsilon(h) = S_\epsilon^{-1}\Gamma_\epsilon(h)S_\epsilon^{-1}$ and $\hat{R}_\epsilon(h) = \hat{S}_\epsilon^{-1}\hat{\Gamma}_\epsilon(h)\hat{S}_\epsilon^{-1}$, with $\Gamma_\epsilon(h) := E\epsilon_t\epsilon'_{t-h} = 0$ for all $h \neq 0$. Consider the vector of the first m sample autocorrelations

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

Define

$$\begin{aligned}\Phi_m &= -E\begin{pmatrix}\epsilon_{t-1}\\ \vdots\\ \epsilon_{t-m}\end{pmatrix}\otimes \tilde{X}'_{t-1}\otimes I_d = -\sum_{i=0}^{\infty} E\begin{pmatrix}\epsilon_{t-1}\\ \vdots\\ \epsilon_{t-m}\end{pmatrix}\tilde{\epsilon}'_{t-i-1}\left(\tilde{A}^i\right)'\otimes I_d\\ &= -\sum_{i=0}^{m-1}\left\{\mathbf{1}_{m\times p}(i+1,1)\otimes \Sigma_{\epsilon_t}\right\}\left(\tilde{A}^i\right)'\otimes I_d,\end{aligned}$$

where $\mathbf{1}_{m\times p}(i,j)$ denotes the $m\times p$ matrix with 1 as ij-th element and 0 elsewhere. Define also the matrix

$$\Xi = \begin{pmatrix}\Sigma_{c_m} & \Sigma_{c_m,\hat{\theta}_n}\\ \Sigma'_{c_m,\hat{\theta}_n} & \Sigma_{\hat{\theta}_n}\end{pmatrix} = \sum_{h=-\infty}^{\infty} E\Upsilon_t\Upsilon'_{t-h}, \quad \text{where} \quad \Upsilon_t = \begin{pmatrix}u_t\\ v_t\end{pmatrix}, \tag{3.2}$$

with $u_t = (\epsilon'_{t-1},\dots,\epsilon'_{t-m})'\otimes\epsilon_t$ and $v_t = \Sigma^{-1}_{\tilde{X}_t}\tilde{X}_{t-1}\otimes\epsilon_t$. The following theorem gives the limiting distribution of the residual empirical autocovariances and autocorrelations.

Theorem 3.1. *Under assumptions* **A1–A3** *or* **A1'–A5**,

$$\sqrt{n}\hat{\gamma}_m \Rightarrow \mathcal{N}(0,\Sigma_{\hat{\gamma}_m}) \quad \textit{where} \quad \Sigma_{\hat{\gamma}_m} = \Sigma_{c_m} + \Phi_m\Sigma_{\hat{\theta}_n}\Phi'_m + \Sigma_{c_m,\hat{\theta}_n}\Phi'_m + \Phi_m\Sigma'_{c_m,\hat{\theta}_n} \tag{3.3}$$

and

$$\sqrt{n}\hat{\rho}_m \Rightarrow \mathcal{N}(0,\Sigma_{\hat{\rho}_m}) \quad \textit{where} \quad \Sigma_{\hat{\rho}_m} = \left\{I_m\otimes(S_\epsilon\otimes S_\epsilon)^{-1}\right\}\Sigma_{\hat{\gamma}_m}\left\{I_m\otimes(S_\epsilon\otimes S_\epsilon)^{-1}\right\}. \tag{3.4}$$

Remark 3.1. In the strong AR case, we have $\Sigma_{c_m} = Eu_tu'_t = I_m\otimes\Sigma_\epsilon^{\otimes 2}$, $\Phi_m\Sigma_{\hat{\theta}_n}\Phi'_m = \Phi_m\left\{\Sigma^{-1}_{\tilde{X}_t}\otimes\Sigma_{\epsilon_t}\right\}\Phi'_m$ and

$$\begin{aligned}\Sigma_{c_m,\hat{\theta}_n} &= Eu_tv'_t = \sum_{i\geq 0} E\begin{pmatrix}\epsilon_{t-1}\\ \vdots\\ \epsilon_{t-m}\end{pmatrix}\tilde{\epsilon}'_{t-i-1}(\tilde{A}^i)'\Sigma^{-1}_{\tilde{X}_t}\otimes\epsilon_t\epsilon'_t\\ &= \sum_{i=0}^{m-1}\left\{\mathbf{1}_{m\times p}(i+1,1)\otimes\Sigma_{\epsilon_t}\right\}(\tilde{A}^i)'\Sigma^{-1}_{\tilde{X}_t}\otimes\Sigma_{\epsilon_t}\\ &= \sum_{i=0}^{m-1}\left[\left\{\mathbf{1}_{m\times p}(i+1,1)\otimes\Sigma_{\epsilon_t}\right\}(\tilde{A}^i)'\otimes I_d\right]\left(\Sigma^{-1}_{\tilde{X}_t}\otimes\Sigma_{\epsilon_t}\right)\\ &= -\Phi_m\Sigma_{\hat{\theta}_n}.\end{aligned}$$

Thus $\Sigma_{\hat{\gamma}_m} = \Sigma_{c_m} - \Phi_m\Sigma_{\hat{\theta}_n}\Phi'_m$. Using straightforward computations, we have

$$\Sigma_{\hat{\gamma}_m} = I_m\otimes\Sigma_\epsilon^{\otimes 2} - E\left[\begin{pmatrix}\epsilon_{t-1}\\ \vdots\\ \epsilon_{t-m}\end{pmatrix}\tilde{X}'_{t-1}\right]\Sigma^{-1}_{\tilde{X}_t}E\left[\begin{pmatrix}\epsilon_{t-1}\\ \vdots\\ \epsilon_{t-m}\end{pmatrix}\tilde{X}'_{t-1}\right]'\otimes\Sigma_\epsilon,$$

which is the result obtained by Brüggemann, Lütkepohl and Saikkonen (2004).

Remark 3.2. Francq, Roy and Zakoïan (2005) considered the univariate case $d=1$. In their paper

$$\Sigma_{c_m} = \sum_{h=-\infty}^{+\infty} E(u_t u'_{t-h}) = \sum_{h=-\infty}^{+\infty} \left\{E(\epsilon_t\epsilon_{t-i}\epsilon_{t-h}\epsilon_{t-i'-h})\right\}_{1\le i,i'\le m}$$

is denoted by $\Gamma_{m,m'} = \{\Gamma(i,i')\}_{1\le i,i'\le m}$. They also introduce the vectors $\lambda_i = -(\psi_{0i-1},\dots,\psi_{0i-p})'$, with the convention $\psi_{0i}=0$ for $i<0$, and the $p\times m$ matrices $\Lambda_m = (\lambda_1,\dots,\lambda_m)$. Noting that $\tilde{X}_t = -\sum_{i=1}^{\infty}\epsilon_{t+1-i}\lambda_i$, we have

$$\Sigma_{\hat{\theta}_n} = \sum_{h=-\infty}^{+\infty} E(v_t v'_{t-h}) \quad = \quad \Sigma_{\tilde{X}_t}^{-1}\sum_{i,i'=1}^{\infty}\lambda_i\Gamma(i,i')\lambda'_{i'}\Sigma_{\tilde{X}_t}^{-1} \;:=\; \Sigma_{\tilde{X}_t}^{-1}\Lambda_\infty\Gamma_{\infty,\infty}\Lambda'_\infty\Sigma_{\tilde{X}_t}^{-1}$$

where the matrix $\Sigma_{\tilde{X}_t}$ is such that

$$\Sigma_{\tilde{X}_t} = \sigma_\epsilon^2\sum_{i=1}^{\infty}\lambda_i\lambda'_i \;:=\; \sigma_\epsilon^2\Lambda_\infty\Lambda'_\infty$$

and σ_ϵ^2 is the variance of the univariate process (ϵ_t). We also have

$$\Phi_m = \sum_{i=1}^{m} E\begin{pmatrix}\epsilon_{t-1}\\ \vdots\\ \epsilon_{t-m}\end{pmatrix}\epsilon_{t-i}\lambda'_i = \sigma_\epsilon^2\Lambda'_m.$$

Similarly we obtain

$$\Sigma_{c_m,\hat{\theta}_n} = -\sum_{i'=1}^{\infty}\begin{pmatrix}\Gamma(1,i')\\ \vdots\\ \Gamma(m,i')\end{pmatrix}\lambda'_{i'}\Sigma_{\tilde{X}_t}^{-1} := -\sigma_\epsilon^{-2}\Gamma_{m,\infty}\Lambda'_\infty\left(\Lambda_\infty\Lambda'_\infty\right)^{-1}$$

in the univariate case.

Using these notations, Theorem 3.1 gives in the case $d=1$

$$\begin{aligned}\Sigma_{\hat{\rho}_m} \quad &= \quad \sigma_\epsilon^{-4}\Sigma_{\hat{\gamma}_m} = \sigma_\epsilon^{-4}\Big\{\Gamma_{m,m} + \Lambda'_m\left(\Lambda_\infty\Lambda'_\infty\right)^{-1}\Lambda_\infty\Gamma_{\infty,\infty}\Lambda'_\infty\left(\Lambda_\infty\Lambda'_\infty\right)^{-1}\Lambda_m\\ &\quad -\Gamma_{m,\infty}\Lambda'_\infty\left(\Lambda_\infty\Lambda'_\infty\right)^{-1}\Lambda_m - \Lambda'_m\left(\Lambda_\infty\Lambda'_\infty\right)^{-1}\Lambda_\infty\Gamma_{\infty,m}\Big\},\end{aligned}$$

which is the result given in Francq, Roy and Zakoïan (2005, Theorem 3.2).

Example 3.1. For the weak AR(1) models considered in Example 2.4, we have

$$\Phi_1 = -\Sigma_\epsilon\otimes I_2 = -\begin{pmatrix}1 & 0 & \rho^{k+1} & 0\\ 0 & 1 & 0 & \rho^{k+1}\\ \rho^{k+1} & 0 & 1 & 0\\ 0 & \rho^{k+1} & 0 & 1\end{pmatrix}, \quad \Phi_m = \begin{pmatrix}\Phi_1\\ 0_{4m\times 4}\end{pmatrix}.$$

Matrix Σ_{c_m} is the variance of the noise $u_t = \left(\epsilon'_{t-1},\dots,\epsilon'_{t-m}\right)'\otimes\epsilon'_t$. Thus Σ_{c_m} is a block-diagonal matrix of the form $\Sigma_{c_m} = \mathrm{Diag}\left\{\Sigma_{c_m}(1,1),\dots,\Sigma_{c_m}(m,m)\right\}$, with diagonal elements $\Sigma_{c_m}(i,i)$ equal to

$$\begin{pmatrix} 3^{k-i+1} & \rho^i(3\rho)^{k-i+1} & \rho^i(3\rho)^{k-i+1} & \rho^{2i}(1+2\rho^2)^{k-i+1} \\ \rho^i(3\rho)^{k-i+1} & (1+2\rho^2)^{k-i+1} & \rho^{2i}(1+2\rho^2)^{k-i+1} & \rho^i(3\rho)^{k-i+1} \\ \rho^i(3\rho)^{k-i+1} & \rho^{2i}(1+2\rho^2)^{k-i+1} & (1+2\rho^2)^{k-i+1} & \rho^i(3\rho)^{k-i+1} \\ \rho^{2i}(1+2\rho^2)^{k-i+1} & \rho^i(3\rho)^{k-i+1} & \rho^i(3\rho)^{k-i+1} & 3^{k-i+1} \end{pmatrix}$$

for $i=1,\dots,k$ and $\Sigma_{c_m}(i,i)=\Sigma_\epsilon^{\otimes 2}$ for $i\geq k+1$. Since $v_t=\left(\Sigma_\epsilon^{-1}\epsilon_{t-1}\right)\otimes\epsilon_t=\left\{e'_m(1)\otimes\Sigma_\epsilon^{-1}\otimes I_2\right\}u_t$, we obtain

$$\Sigma'_{c_m,\hat\theta_n}=\left\{e'_m(1)\otimes\Sigma_\epsilon^{-1}\otimes I_2\right\}\Sigma_{c_m} \tag{3.5}$$

and

$$\Sigma'_{\hat\theta_n}=\left\{e'_m(1)\otimes\Sigma_\epsilon^{-1}\otimes I_2\right\}\Sigma_{c_m}\left\{e_m(1)\otimes\Sigma_\epsilon^{-1}\otimes I_2\right\}. \tag{3.6}$$

Finally, using (3.5), (3.6), $\Phi_m=-\left\{e_m(1)\otimes\Sigma_\epsilon\otimes I_2\right\}$ and $S_\epsilon=I_2$, we obtain

$$\begin{aligned}\Sigma_{\hat\rho_m} &= \Sigma_{\hat\gamma_m}=\Sigma_{c_m}+\Phi_m\Sigma_{\hat\theta_n}\Phi'_m+\Sigma_{c_m,\hat\theta_n}\Phi'_m+\Phi_m\Sigma'_{c_m,\hat\theta_n} \\ &= \Sigma_{c_m}+\left\{\mathbf{1}_{m\times m}(1,1)\otimes I_4\right\}\Sigma_{c_m}\left\{\mathbf{1}_{m\times m}(1,1)\otimes I_4\right\} \\ &\quad -\Sigma_{c_m}\left\{\mathbf{1}_{m\times m}(1,1)\otimes I_4\right\}-\left\{\mathbf{1}_{m\times m}(1,1)\otimes I_4\right\}\Sigma_{c_m} \\ &= \mathrm{Diag}\left\{0_{4\times 4},\Sigma_{c_m}(2,2),\dots,\Sigma_{c_m}(m,m)\right\}. \end{aligned} \tag{3.7}$$

By comparison, for the strong AR(1) with the same variance $\Sigma_\epsilon=\begin{pmatrix}1 & \rho^{k+1}\\ \rho^{k+1} & 1\end{pmatrix}$ we obtain $\Sigma_{c_m}=I_m\otimes\Sigma_\epsilon^{\otimes 2}$, and because the derivations made in (3.7) remain valid in the strong case, $\Sigma_{\hat\rho_m}=\mathrm{Diag}\left\{0,\Sigma_\epsilon^{\otimes 2},\dots,\Sigma_\epsilon^{\otimes 2}\right\}$. It is interesting to note that when $k=1$ the asymptotic variance $\Sigma_{\hat\rho_m}$ is the same in the strong and weak cases, thought the asymptotic variance $\Sigma_{\hat\theta_n}$ is different.

4. Portmanteau test

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 modelling 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

$$\begin{aligned} Q_m &= n\sum_{h=1}^{m}\mathrm{Tr}\left(\hat\Gamma'_\epsilon(h)\hat\Gamma_\epsilon^{-1}(0)\hat\Gamma_\epsilon(h)\hat\Gamma_\epsilon^{-1}(0)\right) \\ &= n\sum_{h=1}^{m}\mathrm{Tr}\left(\mathrm{vec}\,\hat\Gamma_\epsilon(h)\right)'\left(\hat\Gamma_\epsilon^{-1}(0)\otimes\hat\Gamma_\epsilon^{-1}(0)\right)\left(\mathrm{vec}\,\hat\Gamma_\epsilon(h)\right) \\ &= n\hat\gamma'_m\left(I_m\otimes\hat\Gamma_\epsilon^{-1}(0)\otimes\hat\Gamma_\epsilon^{-1}(0)\right)\hat\gamma_m \\ &= n\hat\rho'_m\left(I_m\otimes\hat R_\epsilon^{-1}(0)\otimes\hat R_\epsilon^{-1}(0)\right)\hat\rho_m. \end{aligned}$$

Similarly to the univariate LB portmanteau statistic, Hosking (1980) defined the modified portmanteau statistic

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

Under the assumption that the data generating process (DGP) follows a strong AR(p) model, the asymptotic distribution of the statistics Q_m and $\tilde{Q}_m$ is well approximated by the $\chi^2_{d^2(m-p)}$ distribution ($m > p$). When the innovations are gaussian, Hosking (1980) found that the finite-sample distribution of $\tilde{Q}_m$ is more nearly $\chi^2_{d^2(m-p)}$ than that of Q_m.

From Theorem 3.1 we can deduce the following result, which gives the limiting distribution of the standard portmanteau statistics under general assumptions on the innovation process of the fitted AR(p) model.

Theorem 4.1. *Under assumptions* **A1**–**A3** *or* **A1'**–**A5***, the statistics* Q_m *and* $\tilde{Q}_m$ *converge in distribution, as* $n \to \infty$*, to*

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

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

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

and the Z_i*'s are independent* $\mathcal{N}(0,1)$ *variables.*

The following examples show that, for the asymptotic distribution of Q_m and $\tilde{Q}_m$, the $\chi^2_{d^2(m-p)}$ approximation is no longer valid in the framework of weak AR(p) models.

Example 4.1. In the case of a strong AR(1) model with $A_0 = 0$, it is easy to see that the eigenvalues of Ω_m are 0 with algebraic multiplicity d^2 and 1 with multiplicity $d^2(m-1)$. Thus the asymptotic distribution of Q_m is exactly a $\chi^2_{d^2(m-1)}$ for this strong AR(1) model. When $A_0 \neq 0$, the $\chi^2_{d^2(m-1)}$ law is only an approximation of the asymptotic distribution.

For the weak AR(1) model considered in Example 2.4-3.1, with $\Sigma_\epsilon = I_2$, which corresponds to $\rho = 0$, we have $\Omega_m = \mathrm{Diag}\left\{0_{4\times 4}, \Sigma_{c_m}(2,2), \ldots, \Sigma_{c_m}(m,m)\right\}$, with $\Sigma_{c_m}(i,i) = \mathrm{Diag}\left(3^{k-i+1}, 1, 1, 3^{k-i+1}\right)$ for $i = 2, \ldots, k$ and $\Sigma_{c_m}(i,i) = I_4$ for $i \geq k+1$. Thus in the case $k \geq m > 1$ the asymptotic distribution of Q_m is that of

$$\Gamma\left(\frac{1}{2}, m-1\right) + \sum_{i=2}^{m} \Gamma\left(\frac{1/2}{3^{k-i+1}}, 1\right), \tag{4.1}$$

where the $\Gamma(b,p)$'s denote independent gamma distributions with parameters b and p, *i.e.* with density probability $x \mapsto b^p \left(\int_0^\infty \exp(-y) y^{p-1} dy\right)^{-1} \exp(-bx) x^{p-1} \mathbb{I}_{[0,\infty)}(x)$, $b,\ p > 0$. In the case $1 < k \leq m$, the asymptotic distribution in that of

$$\Gamma\left(\frac{1}{2}, 2m-k-1\right) + \sum_{i=2}^{k} \Gamma\left(\frac{1/2}{3^{k-i+1}}, 1\right). \tag{4.2}$$

As illustrated in Figure 4.1, it is clear that the distributions of (4.1) and (4.2) obtained in the weak case may be quite different from the $\chi^2_{d^2(m-1)}$ obtained in the strong case.

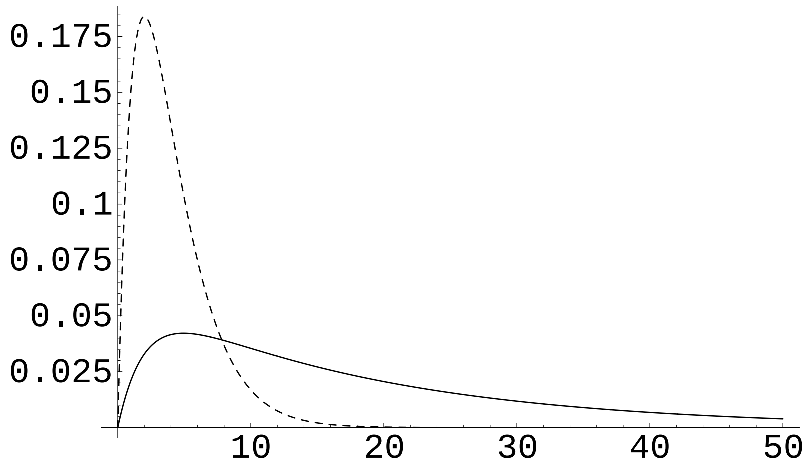

FIGURE 4.1: Comparison of the asymptotic distribution of the portmanteau statistic Q_m in the weak AR case (full line) and in standard strong AR case (dotted line). The AR models are the bivariate AR(1) models defined in Example 4.1, with $k = 3$ and $m = 2$ (thus the dotted line corresponds to the χ^2_4).

Example 4.2. Consider a bivariate AR(1) model $X_t = AX_{t-1} + \epsilon_t$, with $A = 0$. Assume that the innovation process (ϵ_t) is an ARCH(1) model with constant correlation described in Example 2.3:

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

where

$$\begin{pmatrix} \sigma^2_{11,t} \\ \sigma^2_{22,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \epsilon^2_{1,t-1} \\ \epsilon^2_{2,t-1} \end{pmatrix}.$$

Suppose for simplicity that $(\eta_{1,t}, \eta_{2,t})$ is gaussian with variance I_2. We obtain when $m = 2$, $c_1 = 0.3$, and $c_2 = 0.2$

	non zero eigenvalues of $\Sigma_{\hat{\rho}_m}$	Distribution of $Z_m\,(\xi_m)$
$b_{11} = b_{21} = b_{22} = 0$	(1,1,1,1)	χ^2_4
$b_{11} = 0.45$, $b_{21} = 0.4$, $b_{22} = 0.25$	(2.03, 2.44, 1.16, 1.52)	$2.03\chi^2_1 + 2.44\chi^2_1 + 1.16\chi^2_1 + 1.52\chi^2_1$

This table shows that the asymptotic distribution of the goodness-of-fit portmanteau tests may be quite different for AR models with GARCH innovations and strong AR models.

It is seen in Theorem 4.1 that the asymptotic distribution of the BP and LB portmanteau tests depends of nuisance parameters involving Σ_ϵ and the elements of Ξ. The matrix Σ_ϵ can be consistently estimated by its sample estimate $\hat{\Sigma}_\epsilon = \hat{\Gamma}_\epsilon(0)$. To obtain a consistent estimator of Ξ we will use an autoregressive spectral estimator, as in Francq, Roy and Zakoïan (2004). In view of (3.2), the matrix Ξ can be interpreted as 2π times the spectral density of Υ_t evaluated at the frequency zero (see *e.g.* Brockwell and Davis, 1991, p.459). So we have

$$\Xi = \mathcal{A}^{-1}(1)\Sigma_u\mathcal{A}'^{-1}(1),$$

when Υ_t admits the AR(∞) representation

$$\mathcal{A}(B)\Upsilon_t = \Upsilon_t - \sum_{i=1}^{\infty}\mathcal{A}_i\Upsilon_{t-i} = u_t \tag{4.3}$$

where (u_t) is a white noise of variance Σ_u. Since Υ_t is not observable, we have to consider the vectors $\hat{\Upsilon}_t$ obtained by replacing ϵ_t by $\hat{\epsilon}_t$ in Υ_t, with the convention $\hat{\epsilon}_t = 0$ when $t \leq 0$ or $t > n$. Let $\hat{\mathcal{A}}_r(z) = I_{d^2(m+p)} - \sum_{i=1}^{r}\hat{\mathcal{A}}_{r,i}z^i$, where $\hat{\mathcal{A}}_{r,1}, \cdots, \hat{\mathcal{A}}_{r,r}$ denote the coefficients of the LS regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}, \cdots, \hat{\Upsilon}_{t-r}$. Let $\hat{u}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance of $\hat{u}_{r,1}, \dots, \hat{u}_{r,n}$. We need the following assumption.

A7: The innovation process (ϵ_t) of the VAR(p) model (2.1) is such that the process (Υ_t) defined in (3.2) admits an AR(∞) representation (4.3) in which the roots of $\det(\mathcal{A}(z)) = 0$ are outside the unit disk, $\|A_i\| = o(i^{-2})$, and $\Sigma_u = \mathrm{Var}(u_t)$ is non-singular. Moreover we assume that $\|\epsilon_t\|_{8+4\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_\epsilon(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$.

We are now able to state the following theorem, which is an extension of a result given in Francq, Roy and Zakoïan (2004).

Theorem 4.2. *Under Assumptions* **A1-A7***,*

$$\hat{\Xi} := \hat{\mathcal{A}}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\mathcal{A}}_r'^{-1}(1) \to \Xi$$

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 Ξ by $\hat{\Xi}$ and Σ_ϵ by $\hat{\Sigma}_\epsilon$ in Ω_m. Denote by $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \cdots, \hat{\xi}_{d^2m,d^2m})$ the eigenvalues of $\hat{\Omega}_m$. At the asymptotic level α, the LB test (resp. the BP test) consists in rejecting the adequacy of the weak AR(p) model when

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

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

5. Implementation of the goodness-of-fit portmanteau tests

Given d-multivariate observations $X_1, \dots, X_n$, one can use the following steps to implement the modified version of the portmanteau test for testing the AR(p) model adequacy.

1) Compute the empirical autocovariances $\hat{\Gamma}_X(h)$, for $h = 0, \dots, p$;
2) Compute the estimates $\hat{A}_i, \dots, \hat{A}_p$ using (2.5) or using the Durbin-Levinson algorithm;

3) Compute the residuals $\hat{\epsilon}_t = X_t - \sum_{i=1}^{p} \hat{A}_i X_{t-i} \mathbb{I}_{\{t-i>0\}}$, for $t = 1, \dots, n$, and the residual autocovariances $\hat{\Gamma}_\epsilon(0) = \hat{\Sigma}_\epsilon$ and $\hat{\Gamma}_\epsilon(h)$, for $h = 1, \dots, m$;
4) Compute

$$\widehat{\Upsilon}_t = \left(\begin{array}{c} \left(\begin{array}{c} \hat{\epsilon}_{t-1} \\ \vdots \\ \hat{\epsilon}_{t-m} \end{array} \right) \otimes \hat{\epsilon}_t \\ \hat{\Sigma}^{-1}_{\tilde{X}_{t-1}} \tilde{X}_{t-1} \otimes \hat{\epsilon}_t \end{array} \right) \mathbb{I}_{\{t-m>0\}} \quad \text{for } t = 1, \dots, n;$$

5) Compute the first $p_0 + 1$ autocovariances of $\widehat{\Upsilon}_1, \dots, \widehat{\Upsilon}_n$;
6) Using the Durbin-Levinson algorithm, fit the AR(p_0) model

$$\left(I - \sum_{i=1}^{p_0} \hat{\mathcal{A}}_i B^i \right) \widehat{\Upsilon}_t = \hat{U}_t;$$

7) Define the estimator

$$\hat{\Xi} = \left(I - \sum_{i=1}^{p_0} \hat{\mathcal{A}}_i \right)^{-1} \hat{\Sigma}_U \left(I - \sum_{i=1}^{p_0} \hat{\mathcal{A}}_i \right)^{\prime -1}, \quad \hat{\Sigma}_U = \frac{1}{n} \sum_{t=1}^{n} \hat{U}_t \hat{U}_t';$$

8) Define the estimator

$$\hat{\Sigma}_{\hat{\gamma}_m} = \hat{\Sigma}_{c_m} + \hat{\Phi}_m \hat{\Sigma}_{\hat{\theta}_n} \hat{\Phi}'_m + \hat{\Sigma}_{c_m, \hat{\theta}_n} \hat{\Phi}'_m + \hat{\Phi}_m \hat{\Sigma}'_{c_m, \hat{\theta}_n},$$

where

$$\hat{\Xi} = \left(\begin{array}{cc} \hat{\Sigma}_{c_m} & \hat{\Sigma}_{c_m, \hat{\theta}_n} \\ \hat{\Sigma}'_{c_m, \hat{\theta}_n} & \hat{\Sigma}_{\hat{\theta}_n} \end{array} \right), \quad \hat{\Phi}_m = -\sum_{i=0}^{m-1} \left\{ \mathbf{1}_{m \times p}(i+1, 1) \otimes \hat{\Sigma}_\epsilon \right\} \left(\widehat{\tilde{A}}^i \right)' \otimes I_d;$$

9) Compute the eigenvalues $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \cdots, \hat{\xi}_{d^2m,d^2m})$ of

$$\hat{\Omega}_m = \left(I_m \otimes \hat{\Sigma}_\epsilon^{-1/2} \otimes \hat{\Sigma}_\epsilon^{-1/2} \right) \hat{\Sigma}_{\hat{\gamma}_m} \left(I_m \otimes \hat{\Sigma}_\epsilon^{-1/2} \otimes \hat{\Sigma}_\epsilon^{-1/2} \right);$$

10) Compute the portmanteau statistics Q_m and $\tilde{Q}_m$, and evaluate the p-values

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

using the Imhof algorithm (1961). The BP test (resp. the LB test) rejects the null hypothesis of a weak AR(p) model when the first (resp. the second) p-value is less than the asymptotic level α.

An alternative to the Imhof algorithm consists in approximating the distribution of $Z_m(\hat{\xi}_m)$ by a gamma distribution $\Gamma(b, p)$ with parameter

$$b = \frac{\sum_{i=1}^{d^2m} \hat{\xi}_{i,d^2m}}{2 \sum_{i=1}^{d^2m} \hat{\xi}_{i,d^2m}^2} \quad \text{and} \quad p = \frac{\left(\sum_{i=1}^{d^2m} \hat{\xi}_{i,d^2m} \right)^2}{2 \sum_{i=1}^{d^2m} \hat{\xi}_{i,d^2m}^2}.$$

6. Numerical illustrations

In this section, by means of Monte Carlo experiments, we investigate the finite sample properties of the modified and standard versions of the portmanteau tests. We only present the results 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.

6.1. Empirical size

First consider the strong bivariate AR(1) model defined by

$$X_t = AX_{t-1} + \epsilon_t, \quad \epsilon_t \text{ iid } \mathcal{N}(0, I_2), \quad A = 0.95 I_2. \tag{6.1}$$

We simulated $N = 1000$ independent trajectories of length $n = 100$ and $n = 1000$ of this strong AR(1) model. For each of these N replications we estimated the AR(1) matrix coefficient and we applied portmanteau tests to the residuals. The nominal asymptotic level of the tests is $\alpha = 5\%$. For the standard LB test the model is therefore rejected when $\tilde{Q}_m$ is greater than $\chi^2_{4(m-1)}(0.95)$. We know that the asymptotic level of the standard LB test is indeed $\alpha = 5\%$ when $A = 0$, but this is only an approximation in the case $A \neq 0$. For the model (6.1), the roots of $|I - Az| = 0$ are near the unit disk, so the asymptotic distribution of $\tilde{Q}_m$ is likely to be far from its $\chi^2_{4(m-1)}$ approximation. Table 1 displays the relative rejection frequencies of the null hypothesis H_0 that the DGP follows an AR(1) model, over the N replications. When they are outside the 5% significant limits 3.65% and 6.35%, the relative rejection frequencies are displayed in bold type. As expected the observed relative rejection frequency of the standard LB test is very far from the nominal $\alpha = 5\%$. The results are worse for $n = 1000$ than for $n = 100$. This is not surprising because, as we have seen, the asymptotic level is not $\alpha = 5\%$ when $A \neq 0$. In accordance with the theoretical, the $\chi^2_{4(m-1)}$ approximation is better for larger m. In contrast, the modified version of the LB test well controls the error of first kind.

From this example we draw the conclusion that, for strong AR models with coefficients far from 0, the modified version may be preferable to the standard one.

TABLE 1: Empirical size (in %) of the standard and modified versions of the LB test in the case of the strong AR(1) model (6.1).

	$m = 2$		$m = 3$		$m = 6$	
	$n = 100$	$n = 1000$	$n = 100$	$n = 1000$	$n = 100$	$n = 1000$
modified LB	4.6	4.1	4.7	5.4	4.4	6.1
standard LB	**13.8**	**22.6**	**8.4**	**14.2**	4.3	**10.0**

We now repeat the same experiments on a weak AR(1) model of the form given in Example 2.4-3.1-4.1, defined by

$$X_t = AX_{t-1} + \epsilon_t, \quad \epsilon_t = \begin{pmatrix} \eta_{1\,t}\eta_{1\,t-1}\eta_{1\,t-2} \\ \eta_{2\,t}\eta_{2\,t-1}\eta_{2\,t-2} \end{pmatrix}, \quad \begin{pmatrix} \eta_{1\,t} \\ \eta_{2\,t} \end{pmatrix} \text{ iid } \mathcal{N}(0, I_2), \tag{6.2}$$

with $A = 0.5 I_2$. As expected, the standard LB test poorly performs to assess the adequacy of this weak AR(1) model. The true AR(1) model is over-rejected by the standard version of the LB test. Based on the standard LB test, the practitioner is likely to select a too complicated AR model. This could entail erroneous interpretations and a loss of efficiency in terms of linear predictions.

TABLE 2: Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak AR(1) model (6.2).

	$m=2$		$m=3$		$m=6$	
	$n=5000$	$n=10000$	$n=5000$	$n=10000$	$n=5000$	$n=10000$
modified LB	4,9	4,6	4,6	4,2	3.8	4,0
standard LB	**48,8**	**46,0**	**44,1**	**40,9**	**38,8**	**34,0**

6.2. Empirical power

In this part, we consider $N = 1000$ replications of size $n = 1000$ and $n = 2000$ of a weak AR(2) model defined by

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \epsilon_t, \quad \epsilon_t = \begin{pmatrix} \eta_{1\,t}\eta_{1\,t-1}\eta_{1\,t-2} \\ \eta_{2\,t}\eta_{2\,t-1}\eta_{2\,t-2} \end{pmatrix}, \tag{6.3}$$

where $\begin{pmatrix} \eta_{1\,t} \\ \eta_{2\,t} \end{pmatrix}$ iid $\mathcal{N}(0, I_2)$, $A_1 = \begin{pmatrix} 0,2 & 0,1 \\ 0,1 & 0,2 \end{pmatrix}$ and $A_2 = \begin{pmatrix} 0,1 & 0 \\ 0 & 0,1 \end{pmatrix}$.
For each of these N replications we adjusted an AR(1) model and perform standard and modified portmanteau tests based on $m = 2, 3$ or 6 residual autocorrelations. The adequacy of the AR(1) model is rejected when the p-value is less than 5%. Table 3 displays the relative frequency of rejection over the N replications. In this example the power of the portmanteau tests is not very high because A_2 is close to zero. At first sight one could think that the modified version is slightly less powerful that the standard version. Actually, the comparison made in Table 3 is not fair because the actual level of the standard version is generally greater than the 5% nominal level.

TABLE 3: Empirical power (in %) of the standard and modified versions of the 5% nominal level LB test in the case of the weak AR(2) model (6.3).

	$m=2$		$m=3$		$m=6$	
	$n=1000$	$n=2000$	$n=1000$	$n=2000$	$n=1000$	$n=2000$
modified LB	54,9	84,8	45,8	82,6	38.6	74,5
standard LB	84,7	97,5	77,9	96,8	64.2	92.1

Appendix A. Complementary proofs

We need the following lemma for the proof of (2.5) and (2.6):

Lemma A.1. *Let A be symmetric positive definite and B symmetric positive semi-definite of the same order m. Then*

$$Tr\,(A^{-1}B) - \log\det(A^{-1}B) \geq Tr\,(A^{-1}A) - \log\det(A^{-1}A) = m.$$

Proof of Lemma A.1. There exists a non singular matrix P and a positive semi definite diagonal matrix $\Lambda = \mathrm{diag}\{\lambda_1, \dots, \lambda_m\}$ such that

$$A = PP', \qquad B = P\Lambda P'$$

(see Magnus and Neudecker, Theorem 1.23, 1988). Then $A^{-1}B = (P')^{-1}\Lambda P'$. Using Tr $(CD) =$ Tr (DC) for matrices C and D of size $d_1 \times d_2$ and $d_2 \times d_1$, we obtain

$$\mathrm{Tr}(A^{-1}B) = \mathrm{Tr}(\Lambda) = \sum_{i=1}^{m} \lambda_i, \quad \det(A^{-1}B) = \det(\Lambda) = \prod_{i=1}^{m} \lambda_i.$$

Note that the λ_i are non-negative. Hence, using $x - \log x \geq 1$,

$$\mathrm{Tr}(A^{-1}B) - \log\det(A^{-1}B) \geq \sum_{i=1}^{m} (\lambda_i - \log\lambda_i) \geq m. \quad \Box$$

Proof of (2.5) and (2.6). Using elementary properties of the trace operator,

$$\begin{aligned}
(\hat{\theta}_n, \hat{\Sigma}_\epsilon) &= \arg\min_{\theta, \Sigma_\epsilon} \left[n \log\left(\det \Sigma_\epsilon\right) + \mathrm{Tr}\ \left\{ \Sigma_\epsilon^{-1} \sum_{t=1}^{n} \epsilon_t(\theta)\epsilon_t'(\theta) \right\} \right] \\
&= \arg\min_{\theta, \Sigma_\epsilon} \left[n \log\left(\det \Sigma_\epsilon\right) + \mathrm{Tr}\ \left\{ \Sigma_\epsilon^{-1} D \sum_{t=1}^{n} \tilde{\epsilon}_t(\theta)\tilde{\epsilon}_t'(\theta) D' \right\} \right]
\end{aligned}$$

where $D = (I_d, 0_{d\times d}, \dots, 0_{d\times d})$ is such that $\epsilon_t(\theta) = D\tilde{\epsilon}_t(\theta)$.

The parameter θ enters only in the last term, which can be written as

$$\mathrm{Tr}\ \left\{ D'\Sigma_\epsilon^{-1} D \sum_{t=1}^{n} (\tilde{X}_t - \tilde{A}\tilde{X}_{t-1})(\tilde{X}_t - \tilde{A}\tilde{X}_{t-1})' \right\} \quad = \quad c_1 + c_2 + c_3 + c_4$$

where $\tilde{A}^* = \hat{\Sigma}_{\tilde{X}_t, \tilde{X}_{t-1}} \hat{\Sigma}^{-1}_{\tilde{X}_{t-1}, \tilde{X}_{t-1}}$,

$$\begin{aligned}
c_1 &= \mathrm{Tr}\ \left\{ D\Sigma_\epsilon^{-1} D' \sum_{t=1}^{n} (\tilde{X}_t - \tilde{A}^*\tilde{X}_{t-1})(\tilde{X}_t - \tilde{A}^*\tilde{X}_{t-1})' \right\} \\
c_2 &= \mathrm{Tr}\ \left\{ D\Sigma_\epsilon^{-1} D' \sum_{t=1}^{n} (\tilde{X}_t - \tilde{A}^*\tilde{X}_{t-1})\tilde{X}_{t-1}'(\tilde{A}^* - \tilde{A})' \right\} \\
c_3 &= \mathrm{Tr}\ \left\{ D\Sigma_\epsilon^{-1} D' \sum_{t=1}^{n} (\tilde{A}^* - \tilde{A})\tilde{X}_{t-1}(\tilde{X}_t - \tilde{A}^*\tilde{X}_{t-1})' \right\} \\
c_4 &= \mathrm{Tr} \left\{ D\Sigma_\epsilon^{-1} D' \sum_{t=1}^{n} (\tilde{A}^* - \tilde{A})\tilde{X}_{t-1}\tilde{X}_{t-1}'(\tilde{A}^* - \tilde{A})' \right\}.
\end{aligned}$$

We can remark that

$$n^{-1} \sum_{t=1}^{n} (\tilde{X}_t - \tilde{A}^*\tilde{X}_{t-1})\tilde{X}_{t-1}' = \hat{\Sigma}_{\tilde{X}_t, \tilde{X}_{t-1}} - \hat{\Sigma}_{\tilde{X}_t, \tilde{X}_{t-1}} \hat{\Sigma}^{-1}_{\tilde{X}_{t-1}, \tilde{X}_{t-1}} \hat{\Sigma}_{\tilde{X}_{t-1}, \tilde{X}_{t-1}} = 0,$$

so $c_2 = c_3 = 0$. Moreover, c_4 is nonnegative for all values of $\tilde{A}$, and it is equal to zero if and only if $\tilde{A} = \tilde{A}^*$. Thus $\hat{\tilde{A}} = \tilde{A}^*$ is the LS estimator of $\tilde{A}$, which shows (2.5).

Now in view of lemma (A.1) we find

$$\begin{aligned} & n\log\left(\det\Sigma_\epsilon\right) + \mathrm{Tr}\ \left\{\Sigma_\epsilon^{-1}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)\epsilon_t'(\hat{\theta}_n)\right\} = n\log\left(\det\Sigma_\epsilon\right) + n\mathrm{Tr}\ \left(\Sigma_\epsilon^{-1}\Sigma_\epsilon^*\right) \\ = \ & n\log\left(\det\Sigma_\epsilon^*\right) + n\left\{\mathrm{Tr}\ \left(\Sigma_\epsilon^{-1}\Sigma_\epsilon^*\right) - \log\det\left(\Sigma_\epsilon^{-1}\Sigma_\epsilon^*\right)\right\} \\ \geq \ & n\log(\det(\Sigma_\epsilon{}^*)) + dn, \end{aligned}$$

where $\Sigma_\epsilon^* = n^{-1}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)\epsilon_t'(\hat{\theta}_n)$. Since

$$n\log\left(\det\Sigma_\epsilon^*\right) + \mathrm{Tr}\ \left\{(\Sigma_\epsilon^*)^{-1}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)\epsilon_t'(\hat{\theta}_n)\right\} = n\log(\det(\Sigma_\epsilon{}^*)) + dn,$$

the LS estimator of Σ_ϵ is $\hat{\Sigma}_\epsilon = \Sigma_\epsilon^*$, and (2.6) follows.

Proof of Proposition 2.1. First note that if $\hat{\Sigma}_{\tilde{X}_t}$ is singular, then there exists $\lambda \in \mathbb{R}^{dp}$, $\lambda \neq 0$, such that

$$0 = \lambda'\hat{\Sigma}_{\tilde{X}_t}\lambda = \frac{1}{n}\sum_{t=1}^n \lambda'\tilde{X}_t\tilde{X}_t'\lambda = \frac{1}{n}\sum_{t=1}^n\left(\lambda'\tilde{X}_t\right)^2,$$

which entails $\lambda'\tilde{X}_t = 0$ for $t = 1, 2, \dots, n$. Writing $\lambda' = (\lambda_1', \dots, \lambda_p')$ and using (2.2), we have

$$\lambda'\tilde{X}_t = \sum_{i=1}^p \lambda_i' X_{t+1-i} = \lambda_1'\epsilon_t + R_t$$

where R_t is not correlated with ϵ_t. If $\lambda_1 \neq 0$ then $\mathrm{Var}(\lambda'\tilde{X}_t) = \mathrm{Var}(\lambda_1'\epsilon_1) + \mathrm{Var}(R_t) \geq \lambda_1'\Sigma_\epsilon\lambda_1 > 0$, in view of **A3**. Similarly, if $\lambda_1 = \cdots = \lambda_{r-1} = 0$ and $\lambda_r \neq 0$ then $\mathrm{Var}(\lambda'\tilde{X}_t) = \mathrm{Var}(\lambda_r'\epsilon_{t+1-r}) + \mathrm{Var}(R_{t+1-r}) \geq \lambda_r'\Sigma_\epsilon\lambda_r > 0$. Therefore $\lambda'\tilde{X}_t$ is not almost surely equal to 0, and $\hat{\Sigma}_{\tilde{X}_t}$ is almost surely invertible.

In view of (2.2) and Assumptions **A2** and **A4**, (X_t) and $(\tilde{X}_t)$ are stationary and ergodic (see Billingsley, 1995, , Theorem 36.4.). The ergodic theorem implies that, almost surely,

$$\hat{\Sigma}_{\tilde{X}_t,\tilde{X}_{t-1}} \to \Sigma_{\tilde{X}_t,\tilde{X}_{t-1}} := E\tilde{X}_t\tilde{X}_{t-1}' \quad \text{and} \quad \hat{\Sigma}_{\tilde{X}_t} \to \Sigma_{\tilde{X}_t} := E\tilde{X}_t\tilde{X}_t'$$

as $n \to \infty$. Therefore

$$\widehat{\tilde{A}} \to E\tilde{X}_t\tilde{X}_{t-1}'\left(E\tilde{X}_t\tilde{X}_t'\right)^{-1} = E\left(\tilde{A}\tilde{X}_{t-1} + \tilde{\epsilon}_t\right)\tilde{X}_{t-1}'\left(E\tilde{X}_t\tilde{X}_t'\right)^{-1} = \tilde{A},$$

using $E\tilde{\epsilon}_t\tilde{X}_{t-1}' = 0$. The rest of the proof follows by the same arguments. □

Proof of Proposition 2.2. Using the Slutsky lemma,

$$\sqrt{n}\left(\widehat{\tilde{A}} - \tilde{A}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \tilde{\epsilon}_t\tilde{X}_{t-1}'\widehat{\Sigma}_{\tilde{X}_t}^{-1} = \frac{1}{\sqrt{n}}\sum_{t=1}^n \tilde{\epsilon}_t\tilde{X}_{t-1}'\Sigma_{\tilde{X}_t}^{-1} + o_P(1).$$

Let $Y = (Y_t)$ defined by $Y_t = \mathrm{vec}\left(\tilde{\epsilon}_t \tilde{X}'_{t-1}\right)$. Since $\tilde{\epsilon}_t$ is a measurable function of $\tilde{X}_t$ and $\tilde{X}_{t-1}$, it easy to see that $\alpha_Y(|h|+1) \leq \alpha_{\tilde{X}}(|h|) \leq \alpha_X(|h|-p)$, setting $\alpha_X(\ell) = 1/4$ for $\ell \leq 0$. Thus **A5** implies $\sum_{h=0}^{\infty}\{\alpha_Y(h)\}^{\nu/(2+\nu)} < \infty$ and, using the Hölder inequality, $\| Y_t \|_{2+\nu} < \infty$ for some $\nu > 0$.

Moreover by the Lebesgue theorem and the stationarity of (Y_t) we have

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathrm{cov}(Y_t, Y_s) &= \lim_{n\to\infty} \frac{1}{n} \sum_{|h|<n} (n- \mid h \mid) \mathrm{cov}(Y_t, Y_{t-h}) \\ &= \sum_{h=-\infty}^{\infty} \mathrm{cov}\left(Y_t, Y_{t-h}\right). \qquad \text{(A.1)}\end{aligned}$$

The existence of the last sum is justified by the Davydov inequality (1968): there exists a constant K such that

$$\left\| E\left\{\mathrm{vec}\left(\tilde{\epsilon}_t \tilde{X}'_{t-1}\right)\right\}\left\{\mathrm{vec}\left(\tilde{\epsilon}_{t-h} \tilde{X}'_{t-h-1}\right)\right\}' \right\| \leq K \|\mathrm{vec}\left(\tilde{\epsilon}_t \tilde{X}'_{t-1}\right)\|_{2+\nu}^2 \alpha_Y(|h|)^{\nu/(2+\nu)}.$$

The asymptotic normality of $\sqrt{n}\left(\widehat{\tilde{A}} - \tilde{A}\right)$ follows from the central limit theorem (CLT) for mixing processes given by Herrndorf (1984). To obtain the form of the asymptotic variance Ω, note that

$$\begin{aligned}\sqrt{n}\mathrm{vec}\left(\widehat{\tilde{A}} - \tilde{A}\right) &= \left(\Sigma_{\tilde{X}_t}^{-1} \otimes I_{dp}\right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + o_P(1) \\ \mathrm{cov}\left(Y_t, Y_{t-h}\right) &= E\left(\tilde{X}_{t-1}\tilde{X}'_{t-h-1} \otimes \tilde{\epsilon}_t \tilde{\epsilon}'_{t-h}\right),\end{aligned}$$

where the last equality is obtained from the elementary relations $\mathrm{vec}\,(ab') = b \otimes a$ for any vectors a and b, and $(A\otimes B)(C \otimes D) = (AC)\otimes(BD)$ for matrices of appropriate sizes. In view of (A.1), it follows that

$$\Omega = \left(\Sigma_{\tilde{X}_t}^{-1} \otimes I_{dp}\right) \sum_{h=-\infty}^{\infty} E\left\{\tilde{X}_{t-1}\tilde{X}'_{t-h-1} \otimes \tilde{\epsilon}_t \tilde{\epsilon}'_{t-h}\right\} \left(\Sigma_{\tilde{X}_t}^{-1} \otimes I_{dp}\right),$$

which entails (2.8).

The asymptotic distribution (2.9) comes from (2.7) and the fact that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}\left\{I_{dp} \otimes (I_d, 0_{d\times d(p-1)})\right\} \mathrm{vec}\left(\widehat{\tilde{A}} - \tilde{A}_0\right).$$

Note that (2.3) yields

$$\underline{Y} = (\underline{X} \otimes I_d)\,\theta_0 + \underline{\epsilon},$$

where $\underline{Y} = \mathrm{vec}\,(X_1, \ldots, X_n)$, $\underline{X}' = \left(\tilde{X}_0, \ldots, \tilde{X}_{n-1}\right)$ and $\underline{\epsilon} = \mathrm{vec}\,(\epsilon_1, \ldots, \epsilon_n)$. Then it can be seen that

$$\hat{\theta}_n = \left\{\left(\underline{X}'\underline{X}\right)^{-1} \underline{X}' \otimes I_d\right\} \underline{Y} = n^{-1} \sum_{t=1}^{n} \left(\hat{\Sigma}_{\tilde{X}_{t-1}}^{-1} \tilde{X}_{t-1} \otimes I_d\right) X_t$$

and

$$\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)=\sqrt{n}\left\{\left(\underline{X}'\underline{X}\right)^{-1}\underline{X}'\otimes I_d\right\}\underline{\epsilon}=n^{-1/2}\sum_{t=1}^{n}\hat{\Sigma}_{\tilde{X}_{t-1}}^{-1}\tilde{X}_{t-1}\otimes\epsilon_t, \tag{A.2}$$

which gives (2.10). □

Proof of Theorem 3.1. First we will show the asymptotic normality of the joint distribution of c_m and $\hat{\theta}_n-\theta_0$. Using (3.1) and (A.2), we have

$$\sqrt{n}\begin{pmatrix} c_m \\ \hat{\theta}_n-\theta_0 \end{pmatrix}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Upsilon_t+o_P(1), \qquad \Upsilon_t=\begin{pmatrix} \begin{pmatrix} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-m} \end{pmatrix}\otimes\epsilon_t \\ \Sigma_{\tilde{X}_{t-1}}^{-1}\tilde{X}_{t-1}\otimes\epsilon_t \end{pmatrix}.$$

It is clear that Υ_t is a measurable function of $X_t,\dots,X_{t-p-m}$. Thus the CLT applies to the stationary mixing process (Υ_t), and we obtain

$$\sqrt{n}\begin{pmatrix} c_m \\ \hat{\theta}_n-\theta_0 \end{pmatrix}\Rightarrow\mathcal{N}(0,\Xi).$$

To bring out the existence of Ξ, note that from (3.2) we have

$$\Sigma_{c_m}=\sum_{h=-\infty}^{\infty}E\left[\begin{pmatrix} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-m} \end{pmatrix}\otimes\epsilon_t\right]\left[\left(\epsilon'_{t-1-h},\dots,\epsilon'_{t-m-h}\right)\otimes\epsilon'_{t-h}\right]=\sum_{h=-\infty}^{\infty}E\left(u_tu'_{t-h}\right)$$

and

$$\Sigma_{c_m,\hat{\theta}_n}=\sum_{h=-\infty}^{\infty}E\left[\begin{pmatrix} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-m} \end{pmatrix}\otimes\epsilon_t\right]\left[\Sigma_{\tilde{X}_t}^{-1}\tilde{X}'_{t-h-1}\otimes\epsilon'_{t-h}\right]=\sum_{h=-\infty}^{\infty}E\left(u_tv'_{t-h}\right)$$

It is clear that the existence of these matrices is ensured by the Davydov inequality (1968). Then the result follows.

On the other hand, considering C_h and $\hat{\Gamma}_\epsilon(h)$ as values of the same function at the points θ_0 and $\hat{\theta}_n$, a Taylor expansion about θ_0 gives

$$\operatorname{vec}\hat{\Gamma}_\epsilon(h)=\operatorname{vec}C_h+\frac{1}{n}\sum_{t=h+1}^{n}\left\{\epsilon_{t-h}(\theta)\otimes\frac{\partial\epsilon_t(\theta)}{\partial\theta'}+\frac{\partial\epsilon_{t-h}(\theta)}{\partial\theta'}\otimes\epsilon_t(\theta)\right\}_{\theta=\theta^*}\left(\hat{\theta}_n-\theta_0\right), \tag{A.3}$$

where θ^* is between $\hat{\theta}_n$ and θ_0. In view of the consistency of $\hat{\theta}_n$ and the fact that $\partial\epsilon_{t-h}/\partial\theta'$ is not correlated with ϵ_t when $h\geq 0$, it is easy to see that, under mild assumptions,

$$\operatorname{vec}\hat{\Gamma}_\epsilon(h)=\operatorname{vec}C_h+E\left\{\epsilon_{t-h}\otimes\frac{\partial\epsilon_t}{\partial\theta'}(\theta_0)\right\}\left(\hat{\theta}_n-\theta_0\right)+o_P(n^{-1/2}). \tag{A.4}$$

Similarly to (A.4), it follows

$$\hat{\gamma}_m = c_m + \Phi_m \left(\hat{\theta}_n - \theta_0\right) + o_P(n^{-1/2}), \qquad \Phi_m = E\left\{ \begin{pmatrix} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-m} \end{pmatrix} \otimes \frac{\partial \epsilon_t}{\partial \theta'}(\theta_0) \right\}.$$

The expressions of Φ_m are obtained noting that in view of (2.3) we have

$$\frac{\partial \epsilon_t}{\partial \theta'} = \frac{\partial}{\partial \theta'}\left[X_t - \left\{\left(X'_{t-1}, \dots, X'_{t-p}\right) \otimes I_d\right\}\theta\right] = -\left(X'_{t-1}, \dots, X'_{t-p}\right) \otimes I_d,$$

using the elementary relations $a \otimes b' = ab'$ for any vectors a and b, and $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ for matrices of appropriate sizes. Then the asymptotic joint distribution of c_m and $\widehat{\theta}_n - \theta_0$ shows that the asymptotic distribution of $\sqrt{n}\widehat{\gamma}_m$, defined by (A.4), is normal, with mean zero, and covariance matrix

$$\mathrm{Var}_{as}(\sqrt{n}\widehat{\gamma}_m) = \Sigma_{c_m} + \Phi_m \Sigma_{\hat{\theta}_n} \Phi'_m + \Sigma_{c_m, \hat{\theta}_n} \Phi'_m + \Phi_m \Sigma'_{c_m, \hat{\theta}_n}.$$

From (A.3) we have $\sqrt{n}\mathrm{vec}\,\hat{\Gamma}_\epsilon(0) = \sqrt{n}\mathrm{vec}\,C_0 + o_P(1)$. Moreover $\sqrt{n}\,(\mathrm{vec}\,C_0 - E\mathrm{vec}\,C_0) = O_P(1)$ by the CLT for mixing processes. Thus $\sqrt{n}\left(\hat{S}_\epsilon \otimes \hat{S}_\epsilon \; - S_\epsilon \otimes S_\epsilon\right) = O_P(1)$ and, using (3.3) and the ergodic theorem, we obtain

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

In the previous equalities we also use $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ when A and B are invertible. It follows that $\hat{\rho}_m = I_m \otimes \left(\hat{S}_\epsilon \otimes \hat{S}_\epsilon\right)^{-1} \hat{\gamma}_m = I_m \otimes (S_\epsilon \otimes S_\epsilon)^{-1}\hat{\gamma}_m + O_P(n^{-1})$. We now obtain (3.4) from (3.3). $\square$

Proof of Theorem 4.2. The proof is similar to that given by Francq, Roy and Zakoïan (2004) for Theorem 5.1. However, we will give the complete proof of the corresponding Lemma A3 and A4 which are somewhat different from the original proof. This is due to the fact that elements of X_t are involved in the expression of Υ_t. Let $\underline{\Upsilon}_{r,t} = (\Upsilon'_{t-1}, ..., \Upsilon'_{t-r})'$, we define with obvious notations

$$\Sigma_{\Upsilon,\Upsilon} = E\Upsilon_t\Upsilon'_t, \quad \Sigma_{\Upsilon,\underline{\Upsilon}_r} = E\Upsilon_t\underline{\Upsilon}'_{r,t}, \quad \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} = E\underline{\Upsilon}_{r,t}\underline{\Upsilon}'_{r,t}.$$

and

$$\hat{\Sigma}_{\Upsilon,\Upsilon} = \frac{1}{n-r}\sum_{t=l}^{n}\Upsilon_t\Upsilon'_t, \quad \hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} \; = \; \frac{1}{n-r}\sum_{t=1}^{n}\Upsilon_t\underline{\Upsilon}'_{r,t}, \quad \hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} = \frac{1}{n-r}\sum_{t=1}^{n}\underline{\Upsilon}_{r,t}\underline{\Upsilon}'_{r,t}.$$

Now let us state the following Lemma. The proof of this Lemma is the same of Lemma A2 in Francq, Roy and Zakoïan (2004).

Lemma A.2. *Under Assumptions* **A1**-**A7**,

$$\sup_{r\geq 1}\max\left\{\left\|\Sigma_{\Upsilon,\underline{\Upsilon}_r}\right\|, \left\|\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|, \left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|\right\} \leq \infty.$$

Note that to obtain this result we have to consider the multiplicative matrix norm defined by: $\|A\| = \sup_{\|x\|\leq 1}\|Ax\| = \rho^{1/2}(A'A)$, where A is a $d_1 \times d_2$ matrix, $\|x\|$ is the Euclidean norm of the vector $x \in \mathbb{R}^{d_2}$, and $\rho(\cdot)$ denotes the spectral radius. This norm satisfies

$$\|A\|^2 \leq \sum_{i,j} a_{i,j}^2$$

with obvious notations.

Lemma A.3. *Let* $v_{i,t} = x_{j_1,t-1-i}\epsilon_{j_2,t}$, $v'_{i,t} = x_{j'_1,t-1-i}\epsilon_{j'_2,t}$, $u_{i,t} = \epsilon_{j_1,t-i}\epsilon_{j_2,t}$ *and* $u'_{i,t} = \epsilon_{j'_1,t-i}\epsilon_{j'_2,t}$*; where* $j_1, j_2, j'_1, j'_2, \in \{1,\cdots,d\}$*. Suppose that* $\sum_{k=0}^{\infty}\{\alpha_X(k)\}^{\frac{\nu}{2+\nu}} < \infty$*. Then for* $i_1, i_2 \geq 1$

$$\sup_{s\in\mathbb{Z}} \sum_{h=-\infty}^{\infty} \left|Cov\left(v_{i_1,t}v'_{i_2,t-s}, v_{i_1,t-h}v'_{i_2,t-h-s}\right)\right| < \infty. \tag{A.5}$$

$$\sup_{s\in\mathbb{Z}} \sum_{h=-\infty}^{\infty} \left|Cov\left(v_{i_1,t}u'_{i_2,t-s}, v_{i_1,t-h}u'_{i_2,t-h-s}\right)\right| < \infty. \tag{A.6}$$

$$\sup_{s\in\mathbb{Z}} \sum_{h=-\infty}^{\infty} \left|Cov\left(u_{i_1,t}u'_{i_2,t-s}, u_{i_1,t-h}u'_{i_2,t-h-s}\right)\right| < \infty. \tag{A.7}$$

Proof. We will only give the proof of (A.5), the proofs of (A.6) and (A.7) are similar. Note also that without loss of generality, we can take the supremum over the integers $s > 0$, and consider the sum for positive h. Let $i_0 = (i_1+1)\wedge(i_2+1)$. Because the process (ϵ_t) is a measurable function of $(X_t, \cdots, X_{t-p})$ we have, using the Davydov (1968) inequality

$$\begin{aligned}\sum_{h=s+i_0+p}^{\infty} \left|\mathrm{Cov}\left(v_{i_1,t}v'_{i_2,t-s}, v_{i_1,t-h}v'_{i_2,t-h-s}\right)\right| \quad &\leq \quad K_0 \sum_{h=s+i_0+p}^{\infty} \|X_t\|_{8+4\nu}^8 \\ &\qquad \{\alpha_X(h-s-i_0-p)\}^{\frac{\nu}{2+\nu}}.\end{aligned}$$

By Assumption **A7**, the sum in the right hand side of the last inequality is bounded by a constant independent of s. To deal with the terms for $h < s+i_0+p$, we write

$$\begin{aligned}\mathrm{Cov}\left(v_{i_1,t}v'_{i_2,t-s}, v_{i_1,t-h}v'_{i_2,t-h-s}\right) \quad &= \quad \mathrm{Cov}\left(v_{i_1,t}v_{i_1,t-h}, v'_{i_2,t-s}v'_{i_2,t-h-s}\right) \\ &\quad +E\left\{v_{i_1,t}v_{i_1,t-h}\right\}E\left\{v'_{i_2,t-s}v'_{i_2,t-h-s}\right\} \\ &\quad -E\left\{v_{i_1,t}v'_{i_2,t-s}\right\}E\left\{v_{i_1,t-h}v'_{i_2,t-h-s}\right\}\end{aligned}$$

With the convention $\alpha_X(k) = \frac{1}{4}$ for $k \leq 0$, we have

$$\begin{aligned}\sum_{h=0}^{s+i_0+p-1} \left|\mathrm{Cov}\left(v_{i_1,t}v_{i_1,t-h}, v'_{i_2,t-s}v'_{i_2,t-h-s}\right)\right| \quad &\leq \quad K_0 \sum_{k=0}^{s+i_0+p-1} \|X_t\|_{8+4\nu}^8 \\ &\qquad \{\alpha_X(k+1-i_0-i_1-p)\}^{\frac{\nu}{2+\nu}},\end{aligned}$$

$$\sum_{h=0}^{s+i_0+p-1} |E\{v_{i_1,t}v_{i_1,t-h}\}| \quad \leq \quad K_0 \sum_{h=0}^{s+i_0+p-1} \|X_t\|_{4+2\nu}^4 \{\alpha_X(h-i_1-p)\}^{\frac{\nu}{2+\nu}},$$

$$\sum_{h=0}^{s+i_0+p-1} \left|E\left\{v_{i_1,t}v'_{i_2,t-s}\right\}\right| \quad \leq \quad (s+i_0+p)K_0\|X_t\|_{4+2\nu}^4 \{\alpha_X(s-i_1-p)\}^{\frac{\nu}{2+\nu}},$$

Noting that it can be shown that $\sup_{h\geq 1} h\{\alpha_X(h)\}^{\frac{\nu}{2+\nu}} < \infty$, the right hand sides of the three last inequalities are clearly bounded by constants independent of s. Then, the expression (A.5) is bounded. $\Box$

Let $\hat{\underline{\Upsilon}}_{r,t} = (\hat{\Upsilon}'_{t-1}, ..., \hat{\Upsilon}'_{t-r})'$ and define

$$\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}} := \frac{1}{n-r}\sum_{t=l}^{n} \hat{\Upsilon}_t\hat{\Upsilon}'_t, \quad \hat{\Sigma}_{\hat{\Upsilon},\hat{\underline{\Upsilon}}_r} := \frac{1}{n-r}\sum_{t=1}^{n} \hat{\Upsilon}_t\hat{\underline{\Upsilon}}'_{r,t}, \quad \hat{\Sigma}_{\hat{\underline{\Upsilon}}_r,\hat{\underline{\Upsilon}}_r} := \frac{1}{n-r}\sum_{t=1}^{n} \hat{\underline{\Upsilon}}_{r,t}\hat{\underline{\Upsilon}}'_{r,t}.$$

We are now able to state the following Lemma.

Lemma A.4. *Under Assumptions* **A1'-A7** *or* **A1-A3** *and* **A7**, $\sqrt{r}\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r,\hat{\underline{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}} - \Sigma_{\Upsilon,\Upsilon}\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\underline{\Upsilon}}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\|$ *tend to zero in probability as* $n\to\infty$ *when* $r = o(n^{1/3})$.

Proof of Lemma A.4. First we need to define the following vectors

$$\Upsilon_t^* = \left(\begin{array}{c} \left(\begin{array}{c} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-m} \end{array}\right) \otimes \epsilon_t \\ \tilde{X}_{t-1}\otimes\epsilon_t \end{array}\right) \quad \text{and} \quad \underline{\Upsilon}^*_{r,t} = \left(\begin{array}{c} \Upsilon^*_{t-1} \\ \vdots \\ \Upsilon^*_{t-r} \end{array}\right)$$

and let

$$\Sigma_r = I_r \otimes \left(\begin{array}{cc} I_{d^2m} & 0_{d^2m\times d^2p} \\ 0_{d^2p\times d^2m} & \Sigma_{\tilde{X}}^{-1}\otimes I_d \end{array}\right).$$

In addition we denote by $\hat{\Sigma}_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}$ and $\Sigma_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}$ the matrix obtained by replacing $\Upsilon_{r,t}$ by $\Upsilon^*_{r,t}$ in the expressions of respectively $\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$, and $\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$. Note that using the notations introduced below, we have

$$\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} = \Sigma_r\hat{\Sigma}_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}\Sigma_r - \Sigma_r\Sigma_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}\Sigma_r = \Sigma_r\left(\hat{\Sigma}_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r} - \Sigma_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}\right)\Sigma_r. \quad (A.8)$$

Now we will determine the elements of the matrices $\hat{\Sigma}_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}$ and $\Sigma_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}$. Let $x_{j,t}$ the $j-th$ element the vector X_t. For $1\leq i_1,i_2\leq m$ and $1\leq r_1,r_2\leq r$, the elements of the $\{d^2(m+p)(r_1-1)+d^2(i_1-1)+d(j_1-1)+j_2\}-th$ row and $\{d^2(m+p)(r_2-1)+d^2(i_2-1)+d(j'_1-1)+j'_2\}-th$ column of $\hat{\Sigma}_{\underline{\Upsilon}^*_r,\underline{\Upsilon}^*_r}$, is of the form $\frac{1}{n-r}\sum_{t=r_0+1}^{n}\epsilon_t^{(4)}$,where $\epsilon_t^{(4)} = \epsilon_{j_1,t-i_1-r_1}\epsilon_{j_2,t-r_1}\epsilon_{j'_1,t-i_2-r_2}\epsilon_{j'_2,t-r_2}$ and $r_0 = (r_1+i_1)\wedge(r_2+i_2)$. In the same way the element of the $\{d^2(m+p)(r_1-1)+d^2m+d^2(i_1-1)+d(j_1-1)+j_2\}-th$ row and $\{d^2(m+p)(r_2-1)+d^2(i_2-1)+d(j'_1-1)+j'_2\}-th$ column, the $\{d^2(m+p)(r_1-$

$1)+d^2m+d^2(i_1-1)+d(j_1-1)+j_2\}-th$ row and $\{d^2(m+p)(r_2-1)+d^2m+d^2(i_2-1)+d(j_1'-1)+j_2'\}-th$ column of $\hat{\Sigma}_{\underline{\Upsilon}_r^*,\underline{\Upsilon}_r^*}$ are of the form respectively $\frac{1}{n-r}\sum_{t=r_0+2}^{n} w_t^{(4)}$ and $\frac{1}{n-r}\sum_{t=r_0+2}^{n} v_t^{(4)}$, where $w_t^{(4)} = x_{j_1,t-1-i_1-r_1}\epsilon_{j_2,t-r_1}\epsilon_{j_1',t-i_2-r_2}\epsilon_{j_2',t-r_2}$ and $v_t^{(4)} = x_{j_1,t-1-i_1-r_1}\epsilon_{j_2,t-r_1}x_{j_1',t-1-i_2-r_2}\epsilon_{j_2',t-r_2}$. By stationarity of $(\epsilon_t^{(4)})$, we have

$$\begin{aligned}
& E\left\{\left(\frac{1}{n-r}\sum_{t=r_0+1}^{n}\epsilon_t^{(4)}\right)-E\epsilon_t^{(4)}\right\}^2 \\
= \; & E\left\{\frac{1}{n-r}\sum_{t=r_0+1}^{n}(\epsilon_t^{(4)}-E\epsilon_t^{(4)})\right\}^2+\left(\frac{r-r_0}{n-r}\right)^2\left(E\epsilon_t^{(4)}\right)^2 \\
= \; & \frac{1}{(n-r)^2}\sum_{h=-(n-r_0-1)}^{n-r_0-1}(n-r_0)\operatorname{Cov}\left(\epsilon_t^{(4)},\epsilon_{t-h}^{(4)}\right)+\left(\frac{r-r_0}{n-r}\right)^2\left(E\epsilon_t^{(4)}\right)^2 \\
\leq \; & \frac{n}{(n-r)^2}\sum_{h=-\infty}^{\infty}\left|\operatorname{Cov}(\epsilon_t^{(4)},\epsilon_{t-h}^{(4)})\right|+\left(\frac{r-r_0}{n-r}\right)^2\left(E\epsilon_t^{(4)}\right)^2, \\
\leq \; & K_0\left(\frac{n}{(n-r)^2}+\frac{r^2}{(n-r)^2}\right),
\end{aligned}$$

for some constant K_0 independent of r_1,r_2,i_1,i_2 and r,n. Similarly, by stationarity of $(v_t^{(4)})$ and $(w_t^{(4)})$ we obtain

$$E\left\{\left(\frac{1}{n-r}\sum_{t=r_0+1}^{n}v_t^{(4)}\right)-Ev_t^{(4)}\right\}^2\leq K_1\left(\frac{n}{(n-r)^2}+\frac{r^2}{(n-r)^2}\right),$$

$$E\left\{\left(\frac{1}{n-r}\sum_{t=r_0+1}^{n}w_t^{(4)}\right)-Ew_t^{(4)}\right\}^2\leq K_2\left(\frac{n}{(n-r)^2}+\frac{r^2}{(n-r)^2}\right),$$

where K_1 and K_2 are constants independent of r_1,r_2,i_1,i_2. The last three inequalities hold because by Lemma A.4 $\sum_{h=-\infty}^{\infty}\left|\operatorname{Cov}(\epsilon_t^{(4)},\epsilon_{t-h}^{(4)})\right|$, $\sum_{h=-\infty}^{\infty}\left|\operatorname{Cov}(v_t^{(4)},v_{t-h}^{(4)})\right|$ and $\sum_{h=-\infty}^{\infty}\left|\operatorname{Cov}(w_t^{(4)},w_{t-h}^{(4)})\right|$ are uniformly bounded in r_1 and r_2. Let σ_{ij} be the element of the $i-th$ row and $j-th$ column of the matrix $(\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}-\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r})$. From the moment assumption on the process (X_t) we know that the elements of Σ_r are bounded. Thus using (A.8) we can deduce that

$$E(\sigma_{ij})^2\leq Kd^4(m+p)^2\left(\frac{n}{(n-r)^2}+\frac{r^2}{(n-r)^2}\right), \tag{A.9}$$

where K is a constant independent of r_1,r_2,i_1,i_2. In view of (A.9) and Lemma A.2 we have

$$\begin{aligned}
& E\left\{r\|\hat{\Sigma}_{\Upsilon,\Upsilon}-\Sigma_{\Upsilon,\Upsilon}\|^2\right\}\leq E\left\{r\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}-\Sigma_{\Upsilon,\underline{\Upsilon}_r}\|^2\right\} \\
& \leq E\left\{r\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}-\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\|^2\right\}\leq Kd^6(m+p)^4r^3\left(\frac{n}{(n-r)^2}+\frac{r^2}{(n-r)^2}\right)=o(1)
\end{aligned}$$

as $n \to \infty$ when $r = r(n) = o(n^{1/3})$. Then, when $r = o(n^{1/3})$

$$\begin{aligned} \sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\| &= o_P(1), \\ \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\Upsilon} - \Sigma_{\Upsilon,\Upsilon}\| &= o_P(1), \quad \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\| = o_P(1). \end{aligned} \tag{A.10}$$

Recall that $\hat{\epsilon}_t = e_t(\hat{\theta})$ are the residuals of the model (2.1). Let $\tilde{\epsilon}_t = \epsilon_t(\hat{\theta})$. We define $\tilde{\epsilon}_t^{(4)}$, $\tilde{v}_t^{(4)}$, $\tilde{w}_t^{(4)}$ and $\tilde{\Upsilon}_t$ by replacing ϵ_t by $\epsilon_t(\hat{\theta})$, respectively, in $\epsilon_t^{(4)}$, $v_t^{(4)}$, $w_t^{(4)}$, Υ_t. Similarly we define $\hat{\epsilon}_t^{(4)}$, $\hat{v}_t^{(4)}$, $\hat{w}_t^{(4)}$ with obvious notations. In addition we denote by $\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}$, $\hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}$ and $\hat{\Sigma}_{\tilde{\Upsilon},\tilde{\Upsilon}}$ the matrices obtained by replacing Υ_t by $\tilde{\Upsilon}_t$ in $\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$, $\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}$ and $\hat{\Sigma}_{\Upsilon,\Upsilon}$. Now, we will show that replacing $\hat{\epsilon}_t$ by $\tilde{\epsilon}_t = \epsilon_t(\hat{\theta})$ does not modify the asymptotic behaviour of the estimators. It can be easily shown that, almost surely, there exist constants $K > 0$ and $\rho \in]0,1[$ such that, $\sup_{\theta\in\Theta^*}|\epsilon_t(\theta) - e_t(\theta)| \leq K\rho^t$. Then we have for $t \geq r_0$,

$$\begin{aligned} & \left|\hat{\epsilon}_t^{(4)} - \tilde{\epsilon}_t^{(4)}\right| \\ \leq\ & K\rho^{t-i_1-r_1}\left|\hat{\epsilon}_{j_2,t-r_1}\hat{\epsilon}_{j_1',t-i_2-r_2}\hat{\epsilon}_{t-r_2}\right| + K\rho^{t-r_1}\left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\hat{\epsilon}_{j_1',t-i_2-r_2}\hat{\epsilon}_{t-r_2}\right| \\ & +K\rho^{t-i_2-r_2}\left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}\hat{\epsilon}_{j_2',t-r_2}\right| + K\rho^{t-r_2}\left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}\tilde{\epsilon}_{j_1',t-i_2-r_2}\right| \\ \leq\ & K\rho^{t-r_0}\left\{\left|\tilde{\epsilon}_{j_2,t-r_1}\right| + K\rho^{t-r_0}\right\}\left\{\left|\tilde{\epsilon}_{j_1',t-i_2-r_2}\right| + K\rho^{t-r_0}\right\}\left\{\left|\tilde{\epsilon}_{j_2,t-r_2}\right| + K\rho^{t-r_0}\right\} \\ & +K\rho^{t-r_0}\left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\right|\left\{\left|\tilde{\epsilon}_{j_1',t-i_2-r_2}\right| + K\rho^{t-r_0}\right\}\left\{\left|\tilde{\epsilon}_{j_2,t-r_2}\right| + K\rho^{t-r_0}\right\} \\ & +K\rho^{t-r_0}\left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}\right|\left\{\left|\tilde{\epsilon}_{j_2',t-r_2}\right| + K\rho^{t-r_0}\right\} \\ & +K\rho^{t-r_0}\left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_2}\tilde{\epsilon}_{j_1',t-i_2-r_2}\right| \\ \leq\ & K\rho^{t-r_0}\left\{K^3\rho^{3(t-r_0)} + \sum_{l=1}^{3}K^{(3-l)}\rho^{(3-l)(t-r_0)}\sum_{t_1,\dots,t_l\in\mathcal{T}_i}\prod_{k=1}^{l}\|\ \tilde{\epsilon}_{t_k}\ \|\right\}, \end{aligned}$$

where $\mathcal{T}_l = \mathcal{T}_l(t, r_1, r_2, i_1, i_2)$ denotes a set of indices $t_1, \dots, t_l$ such that

$$t_k \in \{t - i_1 - r_1, t - r_1, t - i_2 - r_2, t - r_2\} \qquad \text{for } 1 \leq k \leq l.$$

In the same way, we have

$$\begin{aligned} & \left|\hat{v}_t^{(4)} - \tilde{v}_t^{(4)}\right| \\ \leq\ & \left|x_{j_1,t-1-i_1-r_1}x_{j_1',t-1-r_2}\right|\left|\hat{\epsilon}_{j_2,t-r_1}\hat{\epsilon}_{j_2',t-r_2} - \tilde{\epsilon}_{j_2,t-r_1}\tilde{\epsilon}_{j_2',t-r_2}\right| \\ \leq\ & \left|x_{j_1,t-1-i_1-r_1}x_{j_1',t-1-r_2}\right|\{K\rho^{t-r_1}\left|\hat{\epsilon}_{j_2',t-r_2}\right| + K\rho^{t-r_2}\left|\tilde{\epsilon}_{j_2,t-r_1}\right|\} \\ \leq\ & \left|x_{j_1,t-1-i_1-r_1}x_{j_1',t-1-r_2}\right|K\rho^{t-r_0}\{K\rho^{t-r_0} + \left|\tilde{\epsilon}_{j_2',t-r_2}\right| + \left|\tilde{\epsilon}_{j_2,t-r_1}\right|\} \\ \leq\ & \left|x_{j_1,t-1-i_1-r_1}x_{j_1',t-1-r_2}\right|\{K^2\rho^{2(t-r_0)} + K\rho^{t-r_0}\left|\tilde{\epsilon}_{j_2',t-r_2}\right| + K\rho^{t-r_0}\left|\tilde{\epsilon}_{j_2,t-r_1}\right|\} \end{aligned}$$

and,

$$\begin{aligned}
&\left|\hat{w}_t^{(4)}-\tilde{w}_t^{(4)}\right|\\
\leq\ & |x_{j_1,t-1-i_1-r_1}|\left|\hat{\epsilon}_{j_2,t-r_1}\hat{\epsilon}_{j_1',t-i_2-r_2}\hat{\epsilon}_{j_2',t-r_2}-\tilde{\epsilon}_{j_2,t-r_1}\tilde{\epsilon}_{j_1',t-i_2-r_2}\tilde{\epsilon}_{j_2',t-r_2}\right|\\
\leq\ & |x_{j_1,t-1-i_1-r_1}|\{K\rho^{t-r_1}\left|\hat{\epsilon}_{j_1',t-i_2-r_2}\hat{\epsilon}_{j_2',t-r_2}\right|+K\rho^{t-r_2-i_2}\left|\tilde{\epsilon}_{j_2,t-r_1}\hat{\epsilon}_{j_2',t-r_2}\right|+\\
& K\rho^{t-r_2}\left|\tilde{\epsilon}_{j_2,t-r_1}\tilde{\epsilon}_{j_2',t-i_2-r_2}\right|\\
\leq\ & |x_{j_1,t-1-i_1-r_1}|K\rho^{t-r_0}\{(K\rho^{t-r_0}+\left|\tilde{\epsilon}_{j_1',t-i_2-r_2}\right|)(K\rho^{t-r_0}+\left|\tilde{\epsilon}_{j_2',t-r_2}\right|)\\
& +\left|\tilde{\epsilon}_{j_2,t-r_1}\right|(K\rho^{t-r_0}+\left|\tilde{\epsilon}_{j_2',t-r_2}\right|)+K\rho^{t-r_0}\left|\tilde{\epsilon}_{j_2,t-r_1}\tilde{\epsilon}_{j_2',t-i_2-r_2}\right|\}\\
\leq\ & K\rho^{t-r_0}\left|x_{j_1,t-1-i_1-r_1}\right|\left\{K^2\rho^{2(t-r_0)}+\sum_{l=1}^{2}K^{(2-l)}\rho^{(2-l)(t-r_0)}\sum_{t_1,\dots,t_l\in\mathcal{T}_i}\prod_{k=1}^{l}\parallel\tilde{\epsilon}_{t_k}\parallel\right\},
\end{aligned}$$

By the Hölder and Lyapunov inequalities,

$$E\prod_{k=1}^{l}\parallel\tilde{\epsilon}_{t_k}\parallel^2\leq E\sup_{\theta\in\Theta^*}\parallel\epsilon_t(\theta)\parallel^6<\infty. \tag{A.11}$$

Then, for finite constants K_1^*,K_2^* and K_3^* independent of t,r_1,r_2,i_1 and i_2, we have the following inequalities

$$\begin{aligned}
\left\|\hat{\epsilon}_t^{(4)}-\tilde{\epsilon}_t^{(4)}\right\|_2 &\leq K_1^*\rho^{t-r_0}\\
\left\|\hat{v}_t^{(4)}-\tilde{v}_t^{(4)}\right\|_2 &\leq K_2^*\rho^{t-r_0}\\
\left\|\hat{v}_t^{(4)}-\tilde{v}_t^{(4)}\right\|_2 &\leq K_3^*\rho^{t-r_0}
\end{aligned}$$

and, when $r=o(n)$,

$$\left\|\frac{1}{n-r}\sum_{t=r_0+1}^{n-m}\left(\hat{\epsilon}_t^{(4)}-\tilde{\epsilon}_t^{(4)}\right)\right\|_2 \leq \frac{1}{n-r}K_1^*\sum_{k=1}^{\infty}\rho^k=O\left(n^{-1}\right). \tag{A.12}$$

$$\left\|\frac{1}{n-r}\sum_{t=r_0+1}^{n-m}\left(\hat{v}_t^{(4)}-\tilde{v}_t^{(4)}\right)\right\|_2 \leq \frac{1}{n-r}K_2^*\sum_{k=1}^{\infty}\rho^k=O\left(n^{-1}\right). \tag{A.13}$$

$$\left\|\frac{1}{n-r}\sum_{t=r_0+1}^{n-m}\left(\hat{w}_t^{(4)}-\tilde{w}_t^{(4)}\right)\right\|_2 \leq \frac{1}{n-r}K_3^*\sum_{k=1}^{\infty}\rho^k=O\left(n^{-1}\right). \tag{A.14}$$

Similarly to the vectors $\tilde{\Upsilon}_t$ and $\hat{\Upsilon}_t$, we denote by $\tilde{\Upsilon}_t^*$, $\underline{\tilde{\Upsilon}}_{r,t}^*$ and $\hat{\Upsilon}_t^*$, $\underline{\hat{\Upsilon}}_{r,t}^*$ the vectors obtained by replacing ϵ_t by $\tilde{\epsilon}_t$ and $\hat{\epsilon}_t$ in Υ_t^* and $\underline{\Upsilon}_{r,t}^*$. Thus, we can consider the matrices $\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*}$, $\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}$ which are obtained by replacing $\Upsilon_{r,t}^*$ by respectively $\tilde{\Upsilon}_{r,t}^*$ and $\hat{\Upsilon}_{r,t}^*$ in the expression of respectively $\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$ and $\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}$. In addition we need to define the following matrix

$$\hat{\Sigma}_r=I_r\otimes\begin{pmatrix} I_{d^2m} & 0_{d^2m\times d^2p}\\ 0_{d^2p\times d^2m} & \hat{\Sigma}_{\tilde{X}}^{-1}\otimes I_d\end{pmatrix}.$$

Using these notations we have,

$$\begin{aligned}
& \sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}\| = \sqrt{r}\|\hat{\Sigma}_r\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r - \Sigma_r\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*}\Sigma_r \| \\
\leq\ & \sqrt{r}\|\hat{\Sigma}_r\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r - \Sigma_r\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r + \Sigma_r\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r - \Sigma_r\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*}\Sigma_r \| \\
\leq\ & \sqrt{r} \parallel (\hat{\Sigma}_r - \Sigma_r)\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r \parallel + \sqrt{r} \parallel \Sigma_r(\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*}\Sigma_r) \parallel \\
\leq\ & \sqrt{r} \parallel \Sigma_r(\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r - \hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\Sigma_r + \hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\Sigma_r - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*}\Sigma_r) \parallel \\
& + \sqrt{r} \parallel (\hat{\Sigma}_r - \Sigma_r)\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r \parallel \\
\leq\ & \sqrt{r} \parallel (\hat{\Sigma}_r - \Sigma_r)\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r \parallel + \sqrt{r} \parallel \Sigma_r\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}(\hat{\Sigma}_r - \Sigma_r) \parallel \\
& + \sqrt{r} \parallel \hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*} \parallel \parallel \Sigma_r \parallel^2
\end{aligned}$$

The terms in the right hand side of the last inequality tend to zero when $r = o(n^{1/3})$. Indeed, from standard arguments, inequality (A.11) and the moment assumption on the process X_t we have

$$\lim_{n\to\infty} \sqrt{r} \parallel \Sigma_r\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}(\hat{\Sigma}_r - \Sigma_r) \parallel = \lim_{n\to\infty} \sqrt{r} \parallel (\hat{\Sigma}_r - \Sigma_r)\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*}\hat{\Sigma}_r \parallel = 0.$$

In addition, from the inequalities (A.12), (A.13) and (A.14) we obtain

$$\begin{aligned}
& \lim_{n\to\infty} \sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*}\| \parallel \Sigma_r \parallel^2 \\
=\ & \lim_{n\to\infty} \sqrt{r} \parallel \hat{\Sigma}_r - \Sigma_r \parallel \parallel \hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*} \parallel \parallel \hat{\Sigma}_r \parallel \\
=\ & \lim_{n\to\infty} \sqrt{r} \parallel \Sigma_r \parallel \parallel \hat{\Sigma}_{\underline{\hat{\Upsilon}}_r^*,\underline{\hat{\Upsilon}}_r^*} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*} \parallel \parallel \hat{\Sigma}_r - \Sigma_r \parallel = 0.
\end{aligned}$$

Then we can conclude that when $r = o(n^{1/3})$

$$\begin{aligned}
\lim_{n\to\infty} \sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}\| &= \lim_{n\to\infty} \sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}} - \hat{\Sigma}_{\tilde{\Upsilon},\tilde{\Upsilon}}\| \\
&= \lim_{n\to\infty} \sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}\| = 0. \qquad \text{(A.15)}
\end{aligned}$$

Now we will show that we can replace $\tilde{\epsilon}_t$ by ϵ_t. First using the notations introduced, it is easy to see that

$$\parallel \hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r} \parallel = \parallel \Sigma_r(\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*} - \hat{\Sigma}_{\underline{\Upsilon}_r^*,\underline{\Upsilon}_r^*})\Sigma_r \parallel$$

Note that the general term of $(\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r^*,\underline{\tilde{\Upsilon}}_r^*} - \hat{\Sigma}_{\underline{\Upsilon}_r^*,\underline{\Upsilon}_r^*})$ is of the form $(\epsilon_t^{(4)} - \tilde{\epsilon}_t^{(4)})$, $(v_t^{(4)} - \tilde{v}_t^{(4)})$ or $(w_t^{(4)} - \tilde{w}_t^{(4)})$. A Taylor expansion about θ_0 yields $\left|\epsilon_t(\hat{\theta}) - \epsilon_t\right| \leq \epsilon_t^* \left\|\hat{\theta} - \theta_0\right\|$, $\epsilon_t^* = \left\|\frac{\partial}{\partial\theta'}\epsilon_t(\theta^*)\right\|$ where $\theta^* = \theta^*(t,n)$ is between $\hat{\theta}$ and θ_0.

Thus, for $t \geq r_0$,

$$\begin{aligned}
\left|\epsilon_t^{(4)} - \tilde{\epsilon}_t^{(4)}\right| \leq\ & \left\|\hat{\theta} - \theta_0\right\| \Big\{\epsilon_{j_1,t-i_1-r_1}^* \left|\epsilon_{j_2,t-r_1}\epsilon_{j_1',t-i_2-r_2}\epsilon_{j_2',t-r_2}\right| \\
& + \epsilon_{j_2,t-r_1}^* \left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\epsilon_{j_1',t-i_2-r_2}\epsilon_{j_2',t-r_2}\right| \\
& + \epsilon_{j_1',t-i_2-r_2}^* \left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}\epsilon_{j_2',t-r_2}\right| \\
& + \epsilon_{j_2',t-r_2}^* \left|\tilde{\epsilon}_{j_1,t-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}\tilde{\epsilon}_{j_1',t-i_2-r_2}\right|\Big\}.
\end{aligned}$$

In the same way we write

$$\begin{aligned}\left|v_t^{(4)}-\tilde{v}_t^{(4)}\right| &\leq \left\|\hat{\theta}-\theta_0\right\|\left\{\epsilon^*_{j_2,t-r_1}\left|x_{j_1,t-1-r_1-i_1}x_{j_1',t-1-i_2-r_2}\epsilon_{j_2',t-r_2}\right|\right.\\ &\quad\left.+\epsilon^*_{j_2',t-r_2}\left|x_{j_1,t-1-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}x_{j_1',t-1-r_2-i_2}\right|\right\}\end{aligned}$$

and

$$\begin{aligned}\left|w_t^{(4)}-\tilde{w}_t^{(4)}\right| &\leq \left\|\hat{\theta}-\theta_0\right\|\left\{\epsilon^*_{j_2,t-r_1}\left|x_{j_1,t-1-i_1-r_1}\epsilon_{j_1',t-i_2-r_2}\epsilon_{j_2',t-r_2}\right|\right.\\ &\quad+\epsilon^*_{j_1',t-i_2-r_2}\left|x_{j_1,t-1-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}\epsilon_{j_2',t-r_2}\right|\\ &\quad\left.+\epsilon^*_{j_2',t-r_2}\left|x_{j_1,t-1-i_1-r_1}\tilde{\epsilon}_{j_2,t-r_1}\tilde{\epsilon}_{j_1',t-i_2-r_2}\right|\right\}.\end{aligned}$$

Note that, in the previous inequalities, the L^2-norm of the terms into brackets is bounded, uniformly in t,n,r_1,r_2,i_1 and i_2 because

$$E\left|\epsilon_t^*\right|^8\leq E\sup_{\theta\in\Theta^*}\left\|\frac{\partial\epsilon_t}{\partial\theta'}(\theta)\right\|^8<\infty,\qquad E\parallel\epsilon_t\parallel^8<\infty,\qquad\text{and}\qquad E\parallel\tilde{\epsilon}_t\parallel^8<\infty.$$

Using the Jensen inequality, it follows

$$\left\{\frac{1}{n-r}\sum_{t=r_0+1}^{n}\left(\epsilon_t^{(4)}-\tilde{\epsilon}_t^{(4)}\right)\right\}^2\leq\left\|\hat{\theta}-\theta_0\right\|^2D^{(1)}_{n,r_1,i_1,r_2,i_2},$$

$$\left\{\frac{1}{n-r}\sum_{t=r_0+1}^{n}\left(v_t^{(4)}-\tilde{v}_t^{(4)}\right)\right\}^2\leq\left\|\hat{\theta}-\theta_0\right\|^2D^{(2)}_{n,r_1,i_1,r_2,i_2},$$

$$\left\{\frac{1}{n-r}\sum_{t=r_0+1}^{n}\left(v_t^{(4)}-\tilde{v}_t^{(4)}\right)\right\}^2\leq\left\|\hat{\theta}-\theta_0\right\|^2D^{(3)}_{n,r_1,i_1,r_2,i_2},$$

where $E\left|D^{(k)}_{n,r_1,i_1,r_2,i_2}\right|\leq K^*,\quad k\quad\in\{1,2,3\}$, for some constant K^* independent of n,r_1,r_2,i_1 and i_2. We deduce that $r\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}-\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}\|^2$, $r\|\hat{\Sigma}_{\Upsilon,\Upsilon}-\hat{\Sigma}_{\tilde{\Upsilon},\tilde{\Upsilon}}\|^2$, and $r\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}-\hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}\|$ are respectively bounded by $r^3\left\|\hat{\theta}-\theta_0\right\|^2O_P(1)$, $r\left\|\hat{\theta}-\theta_0\right\|^2O_P(1)$, and $r^2\left\|\hat{\theta}-\theta_0\right\|^2O_P(1)$. Since $\left\|\hat{\theta}-\theta_0\right\|=O_P\left(n^{-1/2}\right)$, we obtain for $r=o(n^{1/3})$

$$\begin{aligned}\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}-\hat{\Sigma}_{\underline{\tilde{\Upsilon}}_r,\underline{\tilde{\Upsilon}}_r}\| &= o_P(1),\\ \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\Upsilon}-\hat{\Sigma}_{\tilde{\Upsilon},\tilde{\Upsilon}}\| &= o_P(1),\\ \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}-\hat{\Sigma}_{\tilde{\Upsilon},\underline{\tilde{\Upsilon}}_r}\| &= o_P(1).\end{aligned}\tag{A.16}$$

The proof of the lemma follows from (A.10), (A.15) and (A.16). □

Recall that we have the AR(∞) model

$$\mathcal{A}(B)\Upsilon_t:=\Upsilon_t-\sum_{i=1}^{\infty}\mathcal{A}_i\Upsilon_{t-i}=u_t,$$

where $\sum_{i=1}^{\infty}\|\mathcal{A}_i\|<\infty$ and $\det(\mathcal{A}(z))\neq 0$ if $|z|\leq 1$, and that we have the regression

$$\hat{\mathcal{A}}_r(B)\hat{\Upsilon}_t := \hat{\Upsilon}_t - \sum_{i=1}^{r}\hat{\mathcal{A}}_{r,i}\hat{\Upsilon}_{t-i} = \hat{\Upsilon}_t - \underline{\hat{\mathcal{A}}}_r\underline{\hat{\Upsilon}}_{r,t} = \hat{u}_{r,t}$$

where $\underline{\hat{\mathcal{A}}}_r = \left(\hat{\mathcal{A}}_{r,1}\cdots\hat{\mathcal{A}}_{r,r}\right) = \hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}$. Now introduce the regression of Υ_t on $\Upsilon_{t-1},...,\Upsilon_{t-r}$ defined by

$$\mathcal{A}_r(B)\Upsilon_t := \Upsilon_t - \sum_{i=1}^{r}\mathcal{A}_{r,i}\Upsilon_{t-i} = \Upsilon_t - \underline{\mathcal{A}}_r\underline{\Upsilon}_{r,t} = u_{r,t}, \qquad u_{r,t}\perp\{\Upsilon_{t-1}\cdots\Upsilon_{t-r}\}$$

where $\underline{\mathcal{A}}_r = (\mathcal{A}_{r,1}\cdots\mathcal{A}_{r,r}) = \Sigma_{\Upsilon,\underline{\Upsilon}_r}\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}$. Writing $\underline{\mathcal{A}}_r^* = (\mathcal{A}_1...\mathcal{A}_r)$ and using standard computations we have

$$\left\|\hat{\mathcal{A}}_r(1)-\mathcal{A}(1)\right\| \leq \sqrt{m}\sqrt{r}\left\{\left\|\underline{\hat{\mathcal{A}}}_r-\underline{\mathcal{A}}_r\right\|+\|\underline{\mathcal{A}}_r^*-\underline{\mathcal{A}}_r\|\right\}+\left\|\sum_{i=r+1}^{\infty}\mathcal{A}_i\right\| \quad \text{(A.17)}$$

and

$$\begin{aligned}\left\|\hat{\Sigma}_{\hat{u}_r}-\Sigma_u\right\| \leq\ & \left\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\Upsilon}}-\Sigma_{\Upsilon,\Upsilon}\right\|+\left\|\left(\underline{\hat{\mathcal{A}}}_r-\underline{\mathcal{A}}_r^*\right)\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right)\right\| \\ & +\left\|\left(\underline{\hat{\mathcal{A}}}_r-\underline{\mathcal{A}}_r^*\right)\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right\|+\left\|\underline{\mathcal{A}}_r^*\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right)\right\| \\ & +\left\|\sum_{i=r+1}^{\infty}\mathcal{A}_iE\Upsilon_{t-i}\Upsilon'_t\right\|. \qquad \text{(A.18)}\end{aligned}$$

The arguments used to show that the terms in the right hand side of (A.17) and (A.18) tend to zero when $r=o(n^{1/3})$ can be extended directly from Francq, Roy and Zakoïan (2004). Indeed, using some computations and Lemma A.2 and A.4 we have for every $\epsilon>0$

$$\begin{aligned}& P\left(\sqrt{r}\left\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}-\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|>\varepsilon\right) \\ \leq\ & P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|>\frac{\varepsilon}{\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|^2+\varepsilon r^{-1/2}\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|}\right) \\ & +P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|\geq\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|^{-1}\right)=o(1).\end{aligned}$$

Then we obtain

$$\sqrt{r}\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}-\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\| = o_P(1). \qquad \text{(A.19)}$$

Using (A.19) and Lemma A.2 we have

$$\left\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}\right\|\leq\left\|\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}-\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|+\left\|\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\|=O_p(1)$$

we can deduce that

$$\begin{aligned}
\sqrt{r}\left\|\underline{\hat{\mathcal{A}}}_r-\underline{\mathcal{A}}_r\right\| &= \sqrt{r}\left\|\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}-\Sigma_{\Upsilon,\underline{\Upsilon}_r}\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right\| \\
&= \sqrt{r}\left\|\left(\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}-\Sigma_{\Upsilon,\underline{\Upsilon}_r}\right)\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}+\Sigma_{\Upsilon,\underline{\Upsilon}_r}\left(\hat{\Sigma}^{-1}_{\underline{\hat{\Upsilon}}_r,\underline{\hat{\Upsilon}}_r}-\Sigma^{-1}_{\underline{\Upsilon}_r,\underline{\Upsilon}_r}\right)\right\|=o_P(1).
\end{aligned}$$

Moreover, using Lemma A.5 in Francq, Roy and Zakoïan (2004), it can be shown that

$$\sqrt{r}\left\|\underline{\mathcal{A}}_r^*-\underline{\mathcal{A}}_r\right\|\to 0 \quad \text{and} \quad \left\|\sum_{i=r+1}^{\infty}\mathcal{A}_i\right\|\to 0,$$

as $r\to\infty$. In view of (A.17) we can conclude that $\hat{\mathcal{A}}_r(1)\to\mathcal{A}(1)$. Similarly it can be shown that $\hat{\Sigma}_{\hat{u}_r,\hat{u}_r}\to\Sigma_{u,u}$ in probability.

References

Ahn, S. K. (1988) Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika*, 75, 590–593.

Billingsley, P. (1961) *Probability and measure.* John Wiley, New York.

Box, G. E. P. and Pierce, D. A. (1970) Distribution of residual autocorrelations in autoregressive-integreted moving average time series models. *J. Amer. Statist. Soc.*, 65, 1509–1526.

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

Brüggeman, R., Lütkepohl, H. and Saikkonen, P. (2004) Residual autocorrelation testing for error correction models. Working document. European university institute.

Chitturi, R. V. (1974) Distribution of residual autocorrelations in multiple autoregressive schemes. *J. Amer. Statist. Soc.*, 65, 1509–1526.

Davidson (1994) *Stochastic limit theory.* Oxford University Press. New York.

Davydov, Y. A. (1968) Convergence of distributions generated by stationary stochastic process. *Theory Prob. Appl.*, 13, 691–696.

Dufour, J-M., Pelletier, D. (2005) Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application. Preprint.

Francq, C., Roy, R. and Zakoïan, J-M. (2004) Goodness-of-fit tests for ARMA models with uncorrelated errors. Working document, http://gremars.univ-lille3.fr/ francq/papiers/portmanteaulong.ps

Francq, C., Roy, R. and Zakoïan, J-M. (2005) Diagnostic checking in ARMA models with uncorrelated errors. *J. Amer. Statist. Soc.*, 100, 532–544.

Francq, C. and Zakoïan, J-M. (1998) Estimating Linear Representations of Nonlinear Processes. *J. Statist. Plan. Infer.*, 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. P. Duchesne and B. Rémillard Editors, Springer Verlag, New York.

HARVILLE, D.A. (1997) *Matrix Algebra From a Statistician's Perspective.* Springer-Verlag, New York.

HERRNDORF, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Prob.*, 12, 141–153.

HOSKING, J. R. M. (1980) The multivariate portmanteau statistic. *J. Amer. Statist. Soc.*, 75, 343–386.

HOSKING, J. R. M. (1981a) Equivalent forms of the multivariate portmanteau statistic. *J. R. Statist. Soc.*, 43, 261–262.

HOSKING, J. R. M. (1981b) Lagrange-tests of multivariate time series models, *J. R. Statist. Soc.*, 43, 219–230.

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

JOHANSEN, S. (1995) *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press, New York.

JEANTHEAU, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14, 70–86.

LI, W. K., AND MCLEOD, A. I. (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models, *J. R. Statist. Soc.*, 43, 231–239.

LJUNG, G. M. AND BOX, G. E. P. (1978) On measure of lack of fit in time series models. *Biometrika*, 65, 297–303.

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

MAGNUS, J. R. AND NEUDECKER, H. (1988) *Matrix differential calculus with applications in statistics and econometrics.* Chichester: John Wiley, New York.

ROMANO, J. L. AND THOMBS, L. A. (1996) Inference for Autocorrelations under Weak Assumptions. *J. Amer. Statist. Soc.*, 91, 590–600.