Working Paper 11 - 14
Statistics and Econometrics Series 09
April 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

A BASIC GOODNESS-OF-FIT PROCESS FOR VARMA(p,q) MODELS

Santiago Velilla[a] and Huong Nguyen[a]

Abstract

This Working Paper presents some preliminary results for a new goodness-of-fit method for VARMA(p,q) models. Relations between least squares residuals and true errors are re-examined, and a new family of statistics is proposed. A new goodness-of-fit process is also suggested, that can be seen as an extension of a previously proposed technique in Ubierna and Velilla (2007). The limit behavior of this last random object is obtained as a consequence of a collection of asymptotic results that generalize those obtained previously in Hosking (1981) and Ubierna and Velilla (2007).

Keywords: Brownian bridge, error sample correlation matrix, goodness-of-fit process, VARMA(p,q) models, weak convergence.

[a]Universidad Carlos III de Madrid, Department of Statistics, Facultad de Ciencias Sociales y Jurdicas, Campus de Getafe, Madrid, Spain. E-mail addresses: santiago.velilla@uc3m.es (Santiago Velilla) and huong.nguyen@uc3m.es. (Huong Nguyen)

1 NOTATIONS AND BASIC ELEMENTS

This section considers parameter estimates of multivariate time series. The commonly used model, autoregressive moving average *VARMA(p,q)* process, is defined in section 1.1. Different from the univariate case, the parameter estimation depends heavily on a estimator of the covariance matrix. Section 1.4 contains some results related to the character of the information matrix as a limit.

1.1 Model definition

Consider a causal and invertible $m-$variate autoregressive moving average *VARMA(p,q)* process

$$\mathbf{\Phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \mathbf{\Theta}(B)\boldsymbol{\varepsilon}_t \ , \tag{1}$$

where B is backward shift operator $B\mathbf{X}_t = \mathbf{X}_{t-1}$, $\boldsymbol{\mu}$ is the $m \times 1$ mean vector and $\{\boldsymbol{\varepsilon}_t : t \in \mathbf{Z}\}$ is a zero mean white noise sequence $WN(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a $m \times m$ positive definite matrix. Additionally, $\mathbf{\Phi}(z) = \mathbf{I}_m - \mathbf{\Phi}_1 z - \cdots - \mathbf{\Phi}_p z^p$ and $\mathbf{\Theta}(z) = \mathbf{I}_m + \mathbf{\Theta}_1 z + \cdots + \mathbf{\Theta}_q z^q$ are matrix polynomials, where $\mathbf{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p, \mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_q$ are $m \times m$ real matrices such that the roots of the determinantal equations $|\mathbf{\Phi}(z)| = 0$ and $|\mathbf{\Theta}(z)| = 0$ all lie outside the unit circle. We also assume that both $\mathbf{\Phi}_p$ and $\mathbf{\Theta}_q$ are non-null matrices, and that the identifiability condition of Hannan (1969), $r(\mathbf{\Phi}_p, \mathbf{\Theta}_q) = m$, holds. Furthermore, it will be convenient to put $P = \max(p, q)$, and to define the $m \times mp$ matrix $\mathbf{\Phi} = (\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p)$, the $m \times mq$ matrix $\mathbf{\Theta} = (\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_q)$, and the $m^2(p+q) \times 1$ vector of parameters $\mathbf{\Lambda} = \mathrm{vec}(\mathbf{\Phi}, \mathbf{\Theta})$.

1.2 Parameter estimation and residuals

Given n observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from model (1), the mean vector $\boldsymbol{\mu}$ can be estimated by the sample mean $\overline{\mathbf{X}}_n = n^{-1} \sum_{t=1}^{n} \mathbf{X}_t$. To estimate the remaining parameters $(\mathbf{\Phi}, \mathbf{\Theta}, \mathbf{\Sigma})$, we consider the collection of $m \times 1$ vectors $\{\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \mathbf{\Theta}, \boldsymbol{\mu}) : 1 \le t \le n\}$, that are defined recursively, using the equations

$$\mathbf{\Phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \mathbf{\Theta}(B)\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \mathbf{\Theta}, \boldsymbol{\mu}) \ , \tag{2}$$

and the initial conditions $\mathbf{X}_t - \boldsymbol{\mu} \equiv \mathbf{0} \equiv \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\mu})$, $t \leq 0$. In practice, the maximum likelihood (ML) estimators of $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$ are considered. Following Lütkepohl (2005, sec.12.2), these are obtained by maximizing the Gaussian likelihood function. Numerically speaking, this problem is equivalent to minimizing the objective function

$$l_n(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma}) = \frac{n}{2} \log(|\boldsymbol{\Sigma}|) + \frac{1}{2} \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t'(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \overline{\mathbf{X}}_n) \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \overline{\mathbf{X}}_n) \; . \tag{3}$$

For a fixed value of the $m \times m$ matrix $\boldsymbol{\Sigma}$, minimizing the second summand of (3) would be a generalized least squares (GLS) criterion for estimating the parameters $(\boldsymbol{\Phi}, \boldsymbol{\Theta})$. In the literature, this is sometimes written (Hosking, 1980, p. 604) in the form

$$\mathrm{vec}(\boldsymbol{\Xi})'(\mathbf{I}_{n-P} \otimes \boldsymbol{\Sigma})^{-1} \mathrm{vec}(\boldsymbol{\Xi}) \; , \tag{4}$$

where $\boldsymbol{\Xi} = (\boldsymbol{\varepsilon}_{P+1}(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \overline{\mathbf{X}}_n), \cdots, \boldsymbol{\varepsilon}_n(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \overline{\mathbf{X}}_n))$ is a $m \times (n-P)$ matrix. We denote the optimizers of (4) as $(\overline{\boldsymbol{\Phi}}(\boldsymbol{\Sigma}), \overline{\boldsymbol{\Theta}}(\boldsymbol{\Sigma}))$. In the univariate case, i.e. $m = 1$, $\boldsymbol{\Sigma}$ is a scalar parameter $\sigma^2 > 0$. Thus, it is easy to see that the quantities $(\overline{\boldsymbol{\Phi}}(\sigma^2), \overline{\boldsymbol{\Theta}}(\sigma^2))$ that minimize the second summand of (3) do not depend on σ^2. In fact, $\overline{\boldsymbol{\Phi}}(\sigma^2)$ and $\overline{\boldsymbol{\Theta}}(\sigma^2)$ are the ML estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$, respectively. However, in the multivariate case, i.e. $m > 1$, $(\overline{\boldsymbol{\Phi}}(\boldsymbol{\Sigma}), \overline{\boldsymbol{\Theta}}(\boldsymbol{\Sigma}))$ depend in general on $\boldsymbol{\Sigma}$, and therefore they must be used in combination with the proper choice of $\widehat{\boldsymbol{\Sigma}}_n = \arg\min_{\boldsymbol{\Sigma}} l_n(\overline{\boldsymbol{\Phi}}(\boldsymbol{\Sigma}), \overline{\boldsymbol{\Theta}}(\boldsymbol{\Sigma}), \boldsymbol{\Sigma})$. Optimization of the log-likelihood function (3) must then performed with respect to all the parameters of $\boldsymbol{\Phi}, \boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ simultaneously. In general, finding the ML estimates $(\widehat{\boldsymbol{\Phi}}_n, \widehat{\boldsymbol{\Theta}}_n, \widehat{\boldsymbol{\Sigma}}_n)$ of $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$, where $\widehat{\boldsymbol{\Phi}}_n = \overline{\boldsymbol{\Phi}}_n(\widehat{\boldsymbol{\Sigma}}_n)$ and $\widehat{\boldsymbol{\Theta}}_n = \overline{\boldsymbol{\Theta}}_n(\widehat{\boldsymbol{\Sigma}}_n)$ is a complex nonlinear optimization problem, affected by the potentially large number of parameters involved, and that must be solved using an adequate efficient algorithm. According to Lütkepohl (2005, p. 408), the vector of ML estimators $\widehat{\boldsymbol{\Lambda}} = \mathrm{vec}(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}})$ is consistent and asymptotically normal for $\boldsymbol{\Lambda} = \mathrm{vec}(\boldsymbol{\Phi}, \boldsymbol{\Theta})$. In particular,

$$\sqrt{n}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}) \xrightarrow{D} N_{m^2(p+q)}[\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\Lambda})] \; , \tag{5}$$

where $\mathbf{I}(\boldsymbol{\Lambda}) = E[\partial^2 l_n(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma}) / \partial \boldsymbol{\Lambda} \partial \boldsymbol{\Lambda}']$ is the $m^2(p+q) \times m^2(p+q)$ information matrix.

Once that the set of estimates $\widehat{\boldsymbol{\Lambda}} = \mathrm{vec}(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}})$ has been determined, the residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$, $t = 1, \ldots, n$, are defined recursively, using equation (2), in the form

$$\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\Phi}}, \overline{\mathbf{X}}_n) = (\mathbf{X}_t - \overline{\mathbf{X}}_n) - \sum_{i=1}^{p} \widehat{\boldsymbol{\Phi}}_i (\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n) - \sum_{j=1}^{q} \widehat{\boldsymbol{\Theta}}_j \widehat{\boldsymbol{\varepsilon}}_{t-j} \; , \quad t = 1, \ldots, n \; , \tag{6}$$

with the usual conditions $\mathbf{X}_t - \overline{\mathbf{X}}_n \equiv \mathbf{0} \equiv \widehat{\boldsymbol{\varepsilon}}_t$, for $t \leq 0$. In practice, only residual vectors for $t > P = \max(p, q)$ are considered. The $\widehat{\boldsymbol{\varepsilon}}_t$ are the natural estimates of the true errors $\boldsymbol{\varepsilon}_t$ of the model, that are unobservable. The relation between the $\widehat{\boldsymbol{\varepsilon}}_t$ and the $\boldsymbol{\varepsilon}_t$ will be studied later in Section 2.1

1.3 Computation of the ML estimates

In the particular case of a *VAR(p)* model, the functions $\boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\mu})$ coincide, for $t > p$, with the true error vectors: $\boldsymbol{\varepsilon}_t$:

$$\boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \boldsymbol{\mu}) = (\mathbf{X}_t - \boldsymbol{\mu}) - \sum_{i=1}^{p} \boldsymbol{\Phi}_i (\mathbf{X}_{t-1} - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t \ . \tag{7}$$

Consequently, the Gaussian log-likelihood objective function (3) is now

$$l_n(\boldsymbol{\Phi}, \boldsymbol{\Sigma}) = \frac{n}{2} \log(|\boldsymbol{\Sigma}|) + \frac{1}{2} \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t'(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n) \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n) \ . \tag{8}$$

Taking partial derivatives in (8) with respect to $\boldsymbol{\Sigma}$, using expressions (15) and (17) in Lütkepohl (2005, p. 670), the equations $\partial l_n(\boldsymbol{\Phi}, \boldsymbol{\Sigma}) / \partial \boldsymbol{\Sigma} = \mathbf{0}$ are equivalent to

$$\frac{n}{2} \boldsymbol{\Sigma}^{-1} - \frac{1}{2} \boldsymbol{\Sigma}^{-1} [\sum_{t>P}^{n} \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n) \boldsymbol{\varepsilon}_t'(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n)] \boldsymbol{\Sigma}^{-1} = \mathbf{0} \ , \tag{9}$$

where, since $q = 0$, $P = \max(p, q) = p$. These lead to the solution

$$\boldsymbol{\Sigma}_n(\boldsymbol{\Phi}) = \frac{1}{n} \sum_{t>P}^{n} \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n) \boldsymbol{\varepsilon}_t'(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n) \ . \tag{10}$$

Therefore, in the *VAR(p)* case the computational ML estimation problem reduces to estimating the autoregressive parameters $\boldsymbol{\Phi}$.

On the other hand, using the formula (2) in Lütkepohl (2005, p. 667), it is easy to check that the conditions $\partial l_n(\boldsymbol{\Phi}, \boldsymbol{\Sigma}) / \partial \mathrm{vec}(\boldsymbol{\Phi}_i)' = \mathbf{0}$, $i = 1, \ldots, p$, lead to

$$\sum_{t>p}^{n} \boldsymbol{\varepsilon}_t'(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n) \boldsymbol{\Sigma}^{-1} \mathbf{D}_{t,i} = \mathbf{0} \ , \quad i = 1, \ldots, p \ , \tag{11}$$

where, taking into account the matrix formula $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A}) \mathrm{vec}(\mathbf{X})$,

$$\mathbf{D}_{t,i} = \frac{\partial \boldsymbol{\varepsilon}_t(\boldsymbol{\Phi}, \overline{\mathbf{X}}_n)}{\partial \mathrm{vec}(\boldsymbol{\Phi}_i)'} = -[(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)' \otimes \mathbf{I}_m] \ , \quad i = 1, \ldots, p \ .$$

The expressions (11) coincide, although with a different notation, with the blocks of equation (12.3.2) in Lütkepohl (2005, p. 468). Applying (11), the ML estimates $\widehat{\mathbf{\Phi}}_i$, $i = 1, \ldots, p$, satisfy the equations

$$\begin{aligned}
\mathbf{0} &= \sum_{t>p}^{n}[(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n) \otimes \mathbf{I}_m]\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n) \quad = \\
&= \sum_{t>p}^{n}[(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n) \otimes \mathbf{I}_m]\mathrm{vec}[\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n)] \quad = \\
&= \sum_{t>p}^{n}\mathrm{vec}[\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)'] \quad = \\
&= (\mathbf{I}_m \otimes \mathbf{\Sigma}^{-1})\mathrm{vec}[\sum_{t>p}^{n}\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)')] \ . \qquad (12)
\end{aligned}$$

Multiplying finally on the left by the matrix $(\mathbf{I}_m \otimes \mathbf{\Sigma}^{-1})^{-1} = (\mathbf{I}_m \otimes \mathbf{\Sigma})$, the equations (12) are equivalent to the exact orthogonality conditions

$$\sum_{t>p}^{n}\boldsymbol{\varepsilon}_t(\mathbf{\Phi}, \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)' = \mathbf{0} \ , \quad i = 1, \ldots, p \ , \qquad (13)$$

that are independent of $\mathbf{\Sigma}$. In particular, expanding expressions (13) the estimates $\widehat{\mathbf{\Phi}}_i$ satisfy the normal or Yule-Walker type equations

$$\sum_{t>p}^{n}(\mathbf{X}_t - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)' = \sum_{j=1}^{p}\mathbf{\Phi}_j[\sum_{t>p}^{n}(\mathbf{X}_{t-j} - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n)'] \ , \quad i = 1, \ldots, p \ . \qquad (14)$$

Example. In the case of a *VAR(1)* process, (14) reduce to

$$\sum_{t>p}^{n}(\mathbf{X}_t - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)' = \mathbf{\Phi}_1\sum_{t>p}^{n}(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)' \ . \qquad (15)$$

Thus, the ML or Yule-Walker type estimate of $\mathbf{\Phi}_1$ is

$$\widehat{\mathbf{\Phi}}_1 = [\sum_{t>p}^{n}(\mathbf{X}_t - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)'][\sum_{t>p}^{n}(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)(\mathbf{X}_{t-1} - \overline{\mathbf{X}}_n)']^{-1} = \widehat{\mathbf{R}}_1 \ , \qquad (16)$$

where $\widehat{\mathbf{R}}_1$ is the first correlation matrix of the sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, as defined by Chitturi (1974). ■

In conclusion, to determine the ML estimates in the *VAR(p)* case amounts to solving the linear system given in (14). However, for the general *VARMA(p,q)* case the orthogonality conditions (11) depend on $\mathbf{\Sigma}$. Hence, an appropriate algorithm is needed. See Lütkepohl (2005, sec.12.3).

1.4 The information matrix as a limit

Consider the $m \times m$ coefficients of the series expansions $\mathbf{\Phi}^{-1}(z)\mathbf{\Theta}(z) = \sum_{j=0}^{\infty} \mathbf{\Omega}_j z^j$ and $\mathbf{\Theta}^{-1}(z) = \sum_{j=0}^{\infty} \mathbf{L}_j z^j$ where $\mathbf{\Omega}_0 = \mathbf{L}_0 = \mathbf{I}_m$. Define also the collection of matrices $\mathbf{G}_k = \sum_{j=0}^{k} (\mathbf{\Sigma}\mathbf{\Omega}_j' \otimes \mathbf{L}_{k-j})$ and $\mathbf{F}_k = \mathbf{\Sigma} \otimes \mathbf{L}_k$, $k \geq 0$. Additionally, construct the sequence of $km^2 \times m^2(p+q)$ matrices $\mathbf{Z}_k = (\mathbf{X}_k, \mathbf{Y}_k)$, $k \geq 1$, where

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{G}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_1 & \mathbf{G}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_2 & \mathbf{G}_1 & \mathbf{G}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{G}_{k-1} & \mathbf{G}_{k-2} & \mathbf{G}_{k-3} & \cdots & \mathbf{G}_{k-p} \end{pmatrix} ,$$

and

$$\mathbf{Y}_k = \begin{pmatrix} \mathbf{F}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_1 & \mathbf{F}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_2 & \mathbf{F}_1 & \mathbf{F}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{F}_{k-1} & \mathbf{F}_{k-2} & \mathbf{F}_{k-3} & \cdots & \mathbf{F}_{k-q} \end{pmatrix} .$$

Following Dunsmuir and Hannan (1976, sec. 4), an explicit expression of the $(i,j)^{th}$ entry of the information matrix $\mathbf{I}(\mathbf{\Lambda}) = (I_{i,j}(\mathbf{\Lambda}))$, $i,j = 1,\dots,m^2(p+q)$, can be obtained in the form

$$I_{i,j}(\mathbf{\Lambda}) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{tr}[\mathbf{f}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{f}(\omega,\mathbf{\Lambda})/\partial\lambda_i)\mathbf{f}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{f}(\omega,\mathbf{\Lambda})/\partial\lambda_j)]d\omega \ , \qquad (17)$$

where $\mathbf{\Lambda} = (\lambda_1,\dots,\lambda_{m^2(p+q)})'$, and

$$\mathbf{f}(\omega,\mathbf{\Lambda}) = \frac{1}{2\pi}\mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{\Theta}(e^{i\omega})\mathbf{\Sigma}\mathbf{\Theta}'(e^{-i\omega})\mathbf{\Phi}'^{-1}(e^{-i\omega}) \ , \ -\pi \leq \omega \leq \pi \ , \qquad (18)$$

is the spectral density matrix of the process (1). Notice that the form (17) generalizes the situation of the univariate context. See e.g Brockwell and Davis (1987, p.376). On the other hand, again by Dunsmuir and Hannan (1976, p 358), the expression (17) can be rewritten as follows

$$I_{i,j}(\mathbf{\Lambda}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{k}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{k}(\omega,\mathbf{\Lambda})/\partial\lambda_i)\mathbf{\Sigma}[\mathbf{k}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{k}(\omega,\mathbf{\Lambda})/\partial\lambda_j)]^*\mathbf{\Sigma}^{-1}\}d\omega \ , \tag{19}$$

where $\mathbf{k}(\omega,\mathbf{\Lambda}) = \mathbf{\Phi}^{-1}(e^{i\omega})\mathbf{\Theta}(e^{i\omega})$ and * denotes the conjugate transpose matrix. Unlike the univariate case, the noise covariance matrix $\mathbf{\Sigma}$ does not cancel in (19). Therefore, for $m > 1$, the information matrix is not scale free.

Introduce the notation $\mathbf{\Phi}_r = (\phi_{jk,r} : j,k = 1,\ldots,m), r = 1,\ldots,p$ and $\mathbf{\Theta}_s = (\phi_{jk,r} : j,k = 1,\ldots,m), s = 1,\ldots,q$, for the $m\times m$ autoregressive and moving average matrices of model (1). It is convenient now to partition the information matrix as follows

$$\mathbf{I}(\mathbf{\Lambda}) = \begin{pmatrix} \mathbf{I}_{11}(\mathbf{\Lambda}) & \mathbf{I}_{12}(\mathbf{\Lambda}) \\ \mathbf{I}_{21}(\mathbf{\Lambda}) & \mathbf{I}_{22}(\mathbf{\Lambda}) \end{pmatrix} .$$

As a consequence, $\mathbf{I}_{11}(\mathbf{\Lambda})$ is a $m^2p \times m^2p$ matrix, whose (r,R) block has coordinates that are of the form

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{k}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{k}(\omega,\mathbf{\Lambda})/\partial\boldsymbol{\phi}_{jk,r})\mathbf{\Sigma}[\mathbf{k}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{k}(\omega,\mathbf{\Lambda})/\partial\boldsymbol{\phi}_{JK,R}]^*\mathbf{\Sigma}^{-1}\}d\omega \ ,$$

The remaining blocks of $\mathbf{I}(\mathbf{\Lambda})$ are defined similarly. For example, $\mathbf{I}_{12}(\mathbf{\Lambda})$ is a $m^2p\times m^2q$ matrix, whose (r,s) block has entries

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{k}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{k}(\omega,\mathbf{\Lambda})/\partial\boldsymbol{\phi}_{jk,r})\mathbf{\Sigma}[\mathbf{k}^{-1}(\omega,\mathbf{\Lambda})(\partial\mathbf{k}(\omega,\mathbf{\Lambda})/\partial\boldsymbol{\theta}_{JK,s}]^*\mathbf{\Sigma}^{-1}\}d\omega \ .$$

After some algebra it can be seen that, for $r,R = 1,\ldots,p$ and $s,S = 1,\ldots,q$, the (r,R) block of $\mathbf{I}_{11}(\mathbf{\Lambda})$ is $\sum_{a=0}^{\infty}\mathbf{G}'_{a-r}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{G}_{a-R}$, where $\mathbf{G}_a = \sum_{u=0}^{a}(\mathbf{\Sigma}\mathbf{\Omega}'_u\otimes\mathbf{L}_{a-u})$ for $a\geq 0$ and $\mathbf{G}_a = 0$ for $a<0$; the (r,s) block of $\mathbf{I}_{12}(\mathbf{\Lambda})$ is $\sum_{a=0}^{\infty}\mathbf{G}'_{a-r}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{F}_{a-s}$, where $\mathbf{F}_a = \mathbf{\Sigma}\otimes\mathbf{L}_a$; the (s,r) block of $\mathbf{I}_{21}(\mathbf{\Lambda})$ is the transpose of the (r,s) block of $\mathbf{I}_{12}(\mathbf{\Lambda})$; finally the (s,S) block of $\mathbf{I}_{22}(\mathbf{\Lambda})$ is $\sum_{a=0}^{\infty}\mathbf{F}'_{a-s}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{F}_{a-S}$.

2 RELATIONS BETWEEN THE RESIDUALS AND THE TRUE ERROR VECTORS

In this section, relations between the residuals $\widehat{\boldsymbol{\varepsilon}}_t$ of (6) and the true error vectors $\boldsymbol{\varepsilon}_t$ of model (1) are given. These extend in a natural way to multivariate time series some well-known results in the univariate case. New problems then are also taken into account. The associated correlation matrix of the residuals $\widehat{\boldsymbol{\varepsilon}}_t$ is also an important tool in identification and diagnostic checking. Thus, this section derives also some connections to its counterpart for the true errors $\boldsymbol{\varepsilon}_t$.

2.1 A linear relation between the residual and the true error vectors

Recall that, as defined in (6), $\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}}, \overline{\mathbf{X}}_n)$, where $(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}})$ are the ML estimates of $(\boldsymbol{\Phi}, \boldsymbol{\Theta})$. Proceeding as in Lemma 1 (Hosking, 1980, p. 603), a Taylor's expansion of $\widehat{\boldsymbol{\varepsilon}}_t$ about $(\boldsymbol{\Phi}, \boldsymbol{\Theta})$ leads to the following linear relation between the residuals and the true error vectors:

$$\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t - \sum_{i=1}^{p}\sum_{r=0}^{\infty} \mathbf{L}_r(\widehat{\boldsymbol{\Phi}}_i - \boldsymbol{\Phi}_i)(\mathbf{X}_{t-i-r} - \boldsymbol{\mu}) - \sum_{j=1}^{q}\sum_{r=0}^{\infty} \mathbf{L}_r(\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j)\boldsymbol{\varepsilon}_{t-j-r} + \widehat{\mathbf{h}}_t \ , \quad (20)$$

where $\widehat{\mathbf{h}}_t$ is a $m \times 1$ remainder term, suitably bounded in probability.

2.2 A linear relation between the residual and error covariance matrices

We write the $m \times m$ sample error covariance matrix at lag k with the notation

$$\mathbf{C}_k = \frac{1}{n}\sum_{t=1}^{n-k} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+k} \ , \quad 0 \leq k \leq n-1 \ . \quad (21)$$

Similarly, the $m \times m$ kth residual covariance matrix is given by

$$\widehat{\mathbf{C}}_k = \frac{1}{n}\sum_{t>P}^{n-k} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}'_{t+k} \ , \quad 0 \leq k \leq n-(P+1) \ , \quad (22)$$

where the $\widehat{\boldsymbol{\varepsilon}}_t$ are as in the left-hand side of expression (20). As before, following Lemma 2 (Hosking, 1980, p. 603), the relation between the residual and error covariance matrices is

$$\widehat{\mathbf{C}}'_k = \mathbf{C}'_k - \sum_{i=1}^{p}\sum_{r=0}^{k-i} \mathbf{L}_{k-i-r}(\widehat{\boldsymbol{\Phi}}_i - \boldsymbol{\Phi}_i)\boldsymbol{\Omega}_r\boldsymbol{\Sigma} - \sum_{j=1}^{q} \mathbf{L}_{k-j}(\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j)\boldsymbol{\Sigma} + \mathbf{U}_k \ , \quad (23)$$

where $\mathbf{U}_k$ is a $m \times m$ remainder term bounded in probability at the rate n^{-1}. After taking vecs in both sides of (23), it follows that, for each $k \geq 1$,

$$\begin{pmatrix} \text{vec}(\widehat{\mathbf{C}}_1') \\ \text{vec}(\widehat{\mathbf{C}}_2') \\ \vdots \\ \text{vec}(\widehat{\mathbf{C}}_k') \end{pmatrix} = \begin{pmatrix} \text{vec}(\mathbf{C}_1') \\ \text{vec}(\mathbf{C}_2') \\ \vdots \\ \text{vec}(\mathbf{C}_k') \end{pmatrix} - \mathbf{Z}_k \text{vec}[(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}) - (\mathbf{\Phi}, \mathbf{\Theta})] + \mathbf{V}_k \ , \tag{24}$$

where the $km^2 \times m^2(p+q)$ matrix $\mathbf{Z}_k = (\mathbf{X}_k, \mathbf{Y}_k)$ is as defined in section 1.4, and $\mathbf{V}_k = (\text{vec}(\mathbf{U}_1)', \ldots, \text{vec}(\mathbf{U}_k)')'$ is a $km^2 \times 1$ random vector.

When $k = 0$, the relation (23) reduces to $\widehat{\mathbf{C}}_0' = \mathbf{C}_0' + O_P(1/n)$, where $\mathbf{C}_0' \xrightarrow{P} \mathbf{\Sigma}$. Therefore, $\widehat{\mathbf{C}}_0'$ is also consistent for the matrix $\mathbf{\Sigma}$. More precisely, $\sqrt{n}(\widehat{\mathbf{C}}_0' - \mathbf{\Sigma}) = \sqrt{n}(\mathbf{C}_0' - \mathbf{\Sigma}) + O_P(1/\sqrt{n}) = O_P(1)$. On the other hand, as it will be seen later in Section 3.2, it can be shown that, under appropriate regularity conditions,

$$\sqrt{n} \begin{pmatrix} \text{vec}(\mathbf{C}_1') \\ \text{vec}(\mathbf{C}_2') \\ \vdots \\ \text{vec}(\mathbf{C}_k') \end{pmatrix} \xrightarrow{D} \mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix} , \quad k \geq 1 \ , \tag{25}$$

where $\mathbf{V}_j^*$, $j = 1, \ldots, k$ are i.i.d. $\mathbf{N}_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$ and $\mathcal{W} = \mathbf{I}_k \otimes \mathbf{\Sigma} \otimes \mathbf{\Sigma}$.

Define now the $km^2 \times km^2$ block diagonal matrix

$$\mathbf{W} = \text{diag}(\mathbf{C}_0 \otimes \mathbf{C}_0, \overset{(k)}{\cdots}, \mathbf{C}_0 \otimes \mathbf{C}_0) = \mathbf{I}_k \otimes \mathbf{C}_0 \otimes \mathbf{C}_0 \ , \tag{26}$$

and its residual counterpart $\widehat{\mathbf{W}} = \mathbf{I}_k \otimes \widehat{\mathbf{C}}_0 \otimes \widehat{\mathbf{C}}_0$. Recall the notation $\mathbf{\Lambda} = \text{vec}(\mathbf{\Phi}, \mathbf{\Theta})$ and $\widehat{\mathbf{\Lambda}} = \text{vec}(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}})$. Notice first that

$$\widehat{\mathbf{W}}^{-1/2} \begin{pmatrix} \text{vec}(\widehat{\mathbf{C}}_1') \\ \text{vec}(\widehat{\mathbf{C}}_2') \\ \vdots \\ \text{vec}(\widehat{\mathbf{C}}_k') \end{pmatrix} = \mathcal{W}^{-1/2} \begin{pmatrix} \text{vec}(\widehat{\mathbf{C}}_1') \\ \text{vec}(\widehat{\mathbf{C}}_2') \\ \vdots \\ \text{vec}(\widehat{\mathbf{C}}_k') \end{pmatrix} + (\widehat{\mathbf{W}}^{-1/2} - \mathcal{W}^{-1/2}) \begin{pmatrix} \text{vec}(\widehat{\mathbf{C}}_1') \\ \text{vec}(\widehat{\mathbf{C}}_2') \\ \vdots \\ \text{vec}(\widehat{\mathbf{C}}_k') \end{pmatrix} . \tag{27}$$

As a consequence of (23), (24) and (25), the second summand of (27) is $O_P(1/n)$. On the other hand, according to Hosking (1980, Appendix, p.607), the following orthogonality condition holds

$$\mathbf{Z}_k'\mathcal{W}^{-1}\begin{pmatrix}\operatorname{vec}(\widehat{\mathbf{C}}_1')\\ \operatorname{vec}(\widehat{\mathbf{C}}_2')\\ \vdots\\ \operatorname{vec}(\widehat{\mathbf{C}}_k')\end{pmatrix} = O_P(\frac{1}{n}) \ . \tag{28}$$

Therefore, using (24) and (25) again, and taking into account that $\mathbf{W}^{-1/2} - \mathcal{W}^{-1/2} = O_P(1/\sqrt{n})$, it can be written

$$\widehat{\mathbf{W}}^{-1/2}\begin{pmatrix}\operatorname{vec}(\widehat{\mathbf{C}}_1')\\ \operatorname{vec}(\widehat{\mathbf{C}}_2')\\ \vdots\\ \operatorname{vec}(\widehat{\mathbf{C}}_k')\end{pmatrix} = (\mathbf{I}_{km^2} - \mathbf{P})\mathbf{W}^{-1/2}\begin{pmatrix}\operatorname{vec}(\mathbf{C}_1')\\ \operatorname{vec}(\mathbf{C}_2')\\ \vdots\\ \operatorname{vec}(\mathbf{C}_k')\end{pmatrix} + O_P(\frac{1}{n}) \ , \tag{29}$$

where $\mathbf{P} = \mathcal{W}^{-1/2}\mathbf{Z}_k(\mathbf{Z}_k'\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\mathbf{Z}_k'\mathcal{W}^{-1/2}$ is the $km^2 \times km^2$ orthogonal projection matrix onto the subspace spanned by the columns of $\mathcal{W}^{-1/2}\mathbf{Z}_k$. Notice finally that, from (25) and (29), it follows that

$$\sqrt{n}\widehat{\mathbf{W}}^{-1/2}\begin{pmatrix}\operatorname{vec}(\widehat{\mathbf{C}}_1')\\ \operatorname{vec}(\widehat{\mathbf{C}}_2')\\ \vdots\\ \operatorname{vec}(\widehat{\mathbf{C}}_k')\end{pmatrix} \overset{D}{\cong} (\mathbf{I}_{km^2} - \mathbf{P})\mathbf{N}_{km^2}(\mathbf{0}, \mathbf{I}_{km^2}) \ . \tag{30}$$

2.3 Relation with the results of the univariate case

When $m = 1$, $\widehat{\mathbf{C}}_k = \widehat{c}_k = \sum_{t=1}^{n-k}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t+k}/n$ and $\mathbf{C}_k = c_k = \sum_{t=1}^{n-k}\varepsilon_t\varepsilon_{t+k}/n$ are the sample covariances of the univariate residuals $\widehat{\varepsilon}_t$ and errors ε_t, respectively. On the other hand, $\mathbf{\Sigma} = \sigma^2 > 0$ and $\mathcal{W} = \sigma^4\mathbf{I}_k$. Therefore, relation (29) becomes

$$\begin{pmatrix}\widehat{r}_1\\ \widehat{r}_2\\ \vdots\\ \widehat{r}_k\end{pmatrix} = (\mathbf{I}_k - \mathbf{P})\begin{pmatrix}r_1\\ r_2\\ \vdots\\ r_k\end{pmatrix} + O_P(\frac{1}{n}) \ , \tag{31}$$

where $\widehat{r}_j = \widehat{c}_j/\widehat{c}_0$ is the jth correlation of the residuals, and $r_j = c_j/c_0$ that of the errors, $j = 1, \ldots, k$. Moreover, the $k \times k$ projection matrix $\mathbf{P} = \mathbf{Z}_k(\mathbf{Z}_k'\mathbf{Z}_k)^{-1}\mathbf{Z}_k'$ depends on the components of $\mathbf{Z}_k = (\mathbf{X}_k, \mathbf{Y}_k)$, where $\mathbf{X}_k$ and $\mathbf{Y}_k$ are as given in Section 1.4, $k \geq 1$. More precisely, in the notation introduced there, it is relatively easy to check that $\mathbf{G}_k = \sigma^2 h_k$ and $\mathbf{F}_k = \sigma^2 l_k$, $k \geq 0$, where the series expansions $\phi^{-1}(z) = \sum_{j=0}^{\infty} h_j z^j$ and $\theta^{-1}(z) = \sum_{j=0}^{\infty} l_j z^j$ are the reciprocals of the autoregressive and moving average polynomials, $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, respectively. Therefore, it follows that

$$\mathbf{X}_k = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ h_1 & 1 & 0 & \cdots & 0 \\ h_2 & h_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{k-1} & h_{k-2} & h_{k-3} & \cdots & h_{k-p} \end{pmatrix},$$

and

$$\mathbf{Y}_k = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_1 & 1 & 0 & \cdots & 0 \\ l_2 & l_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{k-1} & l_{k-2} & l_{k-3} & \cdots & l_{k-q} \end{pmatrix}.$$

Consider now the coefficients $\{a_r : r \geq 0\}$ of the series expansion $a(z) = [\phi(z)\theta(z)]^{-1} = [\theta(z)\phi(z)]^{-1} = \sum_{r=0}^{\infty} a_r z^r$. Consider also the following identities

$$l_k = \phi(B) a_k = a_k - \sum_{s=1}^{p} \phi_s a_{k-s} \ , \tag{32}$$

and

$$h_k = \theta(B) a_k = a_k + \sum_{r=1}^{q} \theta_r a_{k-r} \ . \tag{33}$$

From (32) and (33), it can be written $\mathbf{Z}_k = \mathbf{A}_k \mathbf{B}$, where

$$\mathbf{A}_k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_1 & 1 & \cdots & 0 \\ a_2 & a_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{k-1} & a_{k-2} & \cdots & a_{k-(p+q)} \end{pmatrix},$$

is a $k \times (p+q)$ matrix, and $\mathbf{B} = \mathbf{B}(\boldsymbol{\phi}, \boldsymbol{\theta}) = (\mathbf{B}_1(\boldsymbol{\theta}), \mathbf{B}_2(\boldsymbol{\phi}))$ is a $(p+q) \times (p+q)$ matrix, partitioned in blocks $\mathbf{B}_1(\boldsymbol{\theta})$ and $\mathbf{B}_2(\boldsymbol{\phi})$ of orders $(p+q) \times q$ and $(p+q) \times p$ respectively, where

$$\mathbf{B}_1(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \theta_1 & 1 & \cdots & \vdots \\ \theta_2 & \theta_1 & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ \theta_q & \theta_{q-1} & \cdots & \vdots \\ 0 & \theta_q & & \vdots \\ \vdots & 0 & & \vdots \\ & & & 1 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \theta_{q-1} \\ 0 & 0 & \cdots & \theta_q \end{pmatrix}, \quad \mathbf{B}_2(\boldsymbol{\phi}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\phi_1 & 1 & \cdots & \vdots \\ -\phi_2 & -\phi_1 & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ -\phi_p & -\phi_{p-1} & \cdots & \vdots \\ 0 & -\phi_p & & \vdots \\ \vdots & 0 & & \vdots \\ & & & 1 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & -\phi_{p-1} \\ 0 & 0 & \cdots & -\phi_p \end{pmatrix}.$$

Assuming that the $p+q$ roots, $s_1, \ldots, s_p$ of $\phi(z)$ and $t_1, \ldots, t_q$ of $\theta(z)$, are all different, it can be established that $\mathbf{B}$ is invertible. Therefore, $\mathbf{P} = \mathbf{A}_k(\mathbf{A}_k'\mathbf{A}_k)^{-1}\mathbf{A}_k'$. Accordingly, the univariate version of (29) is given by

$$\begin{pmatrix} \widehat{r}_1 \\ \widehat{r}_2 \\ \vdots \\ \widehat{r}_k \end{pmatrix} = [\mathbf{I}_k - \mathbf{A}_k(\mathbf{A}_k'\mathbf{A}_k)^{-1}\mathbf{A}_k'] \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{pmatrix} + O_P(\frac{1}{n}) \ . \tag{34}$$

Expression (34) was derived by Box and Pierce (1970, Section 5). It can be used to obtain the asymptotic distribution of the portmanteau statistic $Q = n\sum_{j=1}^{k} \widehat{r}_j^2$ for a univariate *ARMA* process. However, according to Hosking (1980, Section 8), this argument is not available in the multivariate context, since matrix multiplication is not commutative. Therefore, the arguments given in (32) and (33) do not apply. See Hosking (1980, Section 8) for further discussion.

Relations (29) and (34) are similar in spirit. However, the former is a very highly dimensional one, because its left-hand side is a $km^2 \times 1$ vector. Therefore, unless a very large sample is available, (29) is not entirely convenient in applications. A question is how to simplify (29) so that its both sides are, as in expression (34) by Box and Pierce (1970), of the order $k \times 1$. A possible solution is discussed in the following section.

2.4 A simplified version of the basic multivariate relation

Chitturi (1974) defines the kth sample correlation matrix of the errors $\boldsymbol{\varepsilon}_t$ as

$$\mathbf{R}_k = \mathbf{C}_k' \mathbf{C}_0^{-1} \ , \tag{35}$$

where the $m \times m$ matrices $\mathbf{C}_k$, $k \geq 0$, are as given in (21). The residual analogue of (35) is

$$\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}_k' \widehat{\mathbf{C}}_0^{-1} \ , \tag{36}$$

where the $\widehat{\mathbf{C}}_k$, $k \geq 0$, are as those considered in (22). Expressions (35) and (36) above generalize the usual definitions of the error and residual correlograms, r_k and $\widehat{r}_k$ respectively, when $m = 1$.

Since the trace of a square matrix is a single number, it is naturally to consider, for dimension-reduction purposes, the properties of the $k \times 1$ random vector

$$\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{R}}_1) \\ \mathrm{tr}(\widehat{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{R}}_k) \end{pmatrix} \ , \quad k \geq 1 \ , \tag{37}$$

as derived from the complex expression (29). Using standard properties of the trace operator, as well as the formulae $\mathrm{tr}(\mathbf{AB}) = [\mathrm{vec}(\mathbf{A}')]'\mathrm{vec}(\mathbf{B})$ and $\mathrm{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A})\mathrm{vec}(\mathbf{X})$, it can be written

$$\begin{aligned} \frac{1}{\sqrt{m}} \mathrm{tr}(\widehat{\mathbf{R}}_k) \quad &= \quad \frac{1}{\sqrt{m}} \mathrm{tr}(\widehat{\mathbf{C}}_0^{-1/2} \widehat{\mathbf{C}}_k' \widehat{\mathbf{C}}_0^{-1/2}) \quad = \quad \frac{1}{\sqrt{m}} \mathrm{tr}(\mathbf{I}_m \widehat{\mathbf{C}}_0^{-1/2} \widehat{\mathbf{C}}_k' \widehat{\mathbf{C}}_0^{-1/2}) \quad = \\ &= \quad \mathbf{a}' \mathrm{vec}(\widehat{\mathbf{C}}_0^{-1/2} \widehat{\mathbf{C}}_k' \widehat{\mathbf{C}}_0^{-1/2}) \quad = \quad \mathbf{a}' (\widehat{\mathbf{C}}_0^{-1/2} \otimes \widehat{\mathbf{C}}_0^{-1/2}) \mathrm{vec}(\widehat{\mathbf{C}}_k') \ , \end{aligned} \tag{38}$$

where $\mathbf{a} = \mathrm{vec}(\mathbf{I}_m)/\sqrt{m}$ is a unit $m^2 \times 1$ vector. Combining (37) and (38) leads to the linear relationship

$$\frac{1}{\sqrt{m}}\begin{pmatrix}\mathrm{tr}(\widehat{\mathbf{R}}_1)\\ \mathrm{tr}(\widehat{\mathbf{R}}_2)\\ \vdots\\ \mathrm{tr}(\widehat{\mathbf{R}}_k)\end{pmatrix} = (\mathbf{I}_k\otimes\mathbf{a}')\widehat{\mathbf{W}}^{-1/2}\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{C}}'_1)\\ \mathrm{vec}(\widehat{\mathbf{C}}'_2)\\ \vdots\\ \mathrm{vec}(\widehat{\mathbf{C}}'_k)\end{pmatrix}, \quad k\geq 1\ . \tag{39}$$

Consequently, the expressions (29) and (39) give

$$\frac{1}{\sqrt{m}}\begin{pmatrix}\mathrm{tr}(\widehat{\mathbf{R}}_1)\\ \mathrm{tr}(\widehat{\mathbf{R}}_2)\\ \vdots\\ \mathrm{tr}(\widehat{\mathbf{R}}_k)\end{pmatrix} = (\mathbf{I}_k\otimes\mathbf{a}')(\mathbf{I}_{km^2}-\mathbf{P})\mathbf{W}^{-1/2}\begin{pmatrix}\mathrm{vec}(\mathbf{C}'_1)\\ \mathrm{vec}(\mathbf{C}'_2)\\ \vdots\\ \mathrm{vec}(\mathbf{C}'_k)\end{pmatrix} + O_P(\frac{1}{n})\ . \tag{40}$$

Expression (40) can be reexpressed using the $km^2\times km^2$ orthogonal projection matrix onto the subspace spanned by the columns of the $km^2\times k$ matrix $\mathbf{I}_k\otimes\mathbf{a}$, $\mathbf{H}=(\mathbf{I}_k\otimes\mathbf{a})(\mathbf{I}_k\otimes\mathbf{a}')$. Writing $(\mathbf{I}_{km^2}-\mathbf{P})=(\mathbf{I}_{km^2}-\mathbf{P})\mathbf{H}+(\mathbf{I}_{km^2}-\mathbf{P})(\mathbf{I}_{km^2}-\mathbf{H})$ and taking into account that

$$\frac{1}{\sqrt{m}}\begin{pmatrix}\mathrm{tr}(\mathbf{R}_1)\\ \mathrm{tr}(\mathbf{R}_2)\\ \vdots\\ \mathrm{tr}(\mathbf{R}_k)\end{pmatrix} = (\mathbf{I}_k\otimes\mathbf{a}')\mathbf{W}^{-1/2}\begin{pmatrix}\mathrm{vec}(\mathbf{C}'_1)\\ \mathrm{vec}(\mathbf{C}'_2)\\ \vdots\\ \mathrm{vec}(\mathbf{C}'_k)\end{pmatrix}, \quad k\geq 1\ , \tag{41}$$

expression (40) above transforms into

$$\begin{aligned}\frac{1}{\sqrt{m}}\begin{pmatrix}\mathrm{tr}(\widehat{\mathbf{R}}_1)\\ \mathrm{tr}(\widehat{\mathbf{R}}_2)\\ \vdots\\ \mathrm{tr}(\widehat{\mathbf{R}}_k)\end{pmatrix} &= (\mathbf{I}_k-\mathbf{Q})\frac{1}{\sqrt{m}}\begin{pmatrix}\mathrm{tr}(\mathbf{R}_1)\\ \mathrm{tr}(\mathbf{R}_2)\\ \vdots\\ \mathrm{tr}(\mathbf{R}_k)\end{pmatrix}+\\ &+(\mathbf{I}_k\otimes\mathbf{a}')(\mathbf{I}_{km^2}-\mathbf{P})(\mathbf{I}_{km^2}-\mathbf{H})\mathbf{W}^{-1/2}\begin{pmatrix}\mathrm{vec}(\mathbf{C}'_1)\\ \mathrm{vec}(\mathbf{C}'_2)\\ \vdots\\ \mathrm{vec}(\mathbf{C}'_k)\end{pmatrix}+O_P(\frac{1}{n})\ ,\end{aligned} \tag{42}$$

where $\mathbf{Q}=(\mathbf{I}_k\otimes\mathbf{a}')\mathbf{P}(\mathbf{I}_k\otimes\mathbf{a})$ is a $k\times k$ symmetric matrix. Notice that when $m=1$, $\mathbf{a}=1$. Thus, $\mathbf{H}=\mathbf{I}_k$. From the results of Section 2.4, it also follows that $\mathbf{Q}=$

$\mathbf{A}_k(\mathbf{A}_k'\mathbf{A}_k)^{-1}\mathbf{A}_k'$. Consequently, the second term of (42) vanishes, and expression above is just the previously derived (34) by Box and Pierce (1970, section 5).

2.5 The structure of the simplified multivariate relation

In (34), for each $j = 1, \ldots, k$, one has

$$\widehat{r}_j = r_j - (a_{j-1}, \ldots, a_{j-(p+q)})(\mathbf{A}_k'\mathbf{A}_k)^{-1}\mathbf{u} + O_P(\frac{1}{n}) \ , \tag{43}$$

where $\mathbf{u} = (u_1, \ldots, u_{p+q})'$ is a $(p+q) \times 1$ vector of ith coordinate $u_i = \sum_{j=i}^k a_{j-i} r_j$, $i = 1, \ldots, p+q$. In the multivariate case, it is interesting to analyze the structure of the simplified relation (40), in order to get a result similar to (43).

Consider first the sequences of $m^2 \times m^2$ matrices $\mathbf{G}_r = \sum_{s=0}^r (\mathbf{\Sigma}\mathbf{\Omega}_s' \otimes \mathbf{L}_{r-s})$ and $\mathbf{F}_r = \mathbf{\Sigma} \otimes \mathbf{L}_r$, $r \geq 0$, where $\mathbf{\Omega}_s$ and $\mathbf{L}_s$ are as defined in Section 1.4. With this notation, the jth $m^2 \times m^2(p+q)$ row block of $\mathcal{W}^{-1/2}\mathbf{Z}_k$ can be written in the form

$$\mathbf{\Xi}_j' = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{G}_{j-1}, \ldots, \mathbf{G}_{j-p}\, ; \mathbf{F}_{j-1}, \ldots, \mathbf{F}_{j-q}) \ , \tag{44}$$

where $\mathbf{Z}_k = (\mathbf{X}_k, \mathbf{Y}_k)$ is also as defined in Section 1.4. On the other hand, define $\boldsymbol{\xi}_j' = \mathbf{a}'\mathbf{\Xi}_j' = (\mathbf{g}_{j-1}', \ldots, \mathbf{g}_{j-p}'\, ; \mathbf{f}_{j-1}', \ldots, \mathbf{f}_{j-q}')$, where

$$\begin{aligned} \mathbf{g}_j' \quad &= \quad \mathbf{a}'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{G}_j \quad = \quad \mathbf{a}' \sum_{k=0}^{j} (\mathbf{\Sigma}^{1/2}\mathbf{\Omega}_k' \otimes \mathbf{\Sigma}^{-1/2}\mathbf{L}_{j-k}) \quad = \\ &= \quad \frac{1}{\sqrt{m}} \sum_{k=0}^{j} [\mathrm{vec}(\mathbf{L}_{j-k}'\mathbf{\Omega}_k')]' \ , \end{aligned} \tag{45}$$

and

$$\mathbf{f}_j' = \mathbf{a}'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{F}_j = \mathbf{a}'(\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2}\mathbf{L}_j) = \frac{1}{\sqrt{m}}[\mathrm{vec}(\mathbf{L}_j')]' \ , \tag{46}$$

are row vectors of $1 \times m^2$. Finally, consider the $m^2(p+q) \times 1$ vector $\mathbf{U} = (\mathbf{u}_1', \ldots, \mathbf{u}_p'; \mathbf{u}_{p+1}', \ldots, \mathbf{u}_{p+q}')'$, where, for $i = 1, \ldots, p$,

$$\mathbf{u}_i = \sum_{j=i}^{k} \mathbf{G}_{j-i}'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2})\mathrm{vec}(\mathbf{C}_j') \ , \tag{47}$$

where $\mathbf{G}_r'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) = \sum_{s=0}^r (\mathbf{\Omega}_s\mathbf{\Sigma} \otimes \mathbf{L}_{r-s}')(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) = \sum_{s=0}^r (\mathbf{\Omega}_s\mathbf{\Sigma}^{1/2} \otimes \mathbf{L}_{r-s}'\mathbf{\Sigma}^{-1/2})$, and, for $i = 1, \ldots, q$,

$$\mathbf{u}_{p+i} = \sum_{j=i}^{k} \mathbf{F}_{j-i}'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2})\mathrm{vec}(\mathbf{C}_j') \ , \tag{48}$$

where $\mathbf{F}_r'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) = (\mathbf{\Sigma} \otimes \mathbf{L}_r')(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) = \mathbf{\Sigma}^{1/2} \otimes \mathbf{L}_r'\mathbf{\Sigma}^{-1/2}$. Therefore, going back to (40), and taking into account (41), it follows that, for each $j = 1, \ldots, k$,

$$\frac{1}{\sqrt{m}}\,\mathrm{tr}(\widehat{\mathbf{R}}_j) = \frac{1}{\sqrt{m}}\,\mathrm{tr}(\mathbf{R}_j) - \boldsymbol{\xi}_j'(\mathbf{Z}_k'\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}\mathbf{U} + O_P(\frac{1}{n}) \ . \tag{49}$$

Expression (49) is a multivariate generalization of (43), $\widehat{r}_j = r_j - (a_{j-1}, \ldots, a_{j-(p+q)})$ $(\mathbf{A}_k'\mathbf{A}_k)^{-1}\mathbf{u} + O_P(1/n)$, where $\mathbf{u} = (u_1, \ldots, u_{p+q})'$ and $u_i = \sum_{j=i}^k a_{j-i}r_j$, $i = 1, \ldots, p+q$. Its structure is similar to this last expression. The role of $\widehat{r}_j$ and r_j is adopted by $\mathrm{tr}(\widehat{\mathbf{R}}_j)/\sqrt{m}$ and $\mathrm{tr}(\mathbf{R}_j)/\sqrt{m}$, respectively; the constants a_{j-i} are replaced by either the row vector $\mathbf{g}_{j-i}'$ or $\mathbf{f}_{j-i}'$; the matrix $(\mathbf{A}_k'\mathbf{A}_k)^{-1}$ by $(\mathbf{Z}_k'\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}$; moreover, the blocks of $\mathbf{U}$ are linear combinations of $(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2})\mathrm{vec}(\mathbf{C}_j')$ with coefficients given by the transposes of the components of $\mathbf{\Xi}_j'$ in (44).

As compared to the univariate case, the multivariate relation (49) depends on the scale parameter $\mathbf{\Sigma}$. The following example illustrates the situation.

Example. In the case of a *VAR(p)* process, $q = 0$. Moreover, $\mathbf{G}_j = (\mathbf{\Sigma}\mathbf{H}_j' \otimes \mathbf{I}_m)$, where the $\{\mathbf{H}_j : j \geq 0\}$ are the coefficients of the series expansion $\mathbf{\Phi}^{-1}(z) = \sum_{j=0}^{\infty} \mathbf{H}_j\mathbf{z}^j$, and $\mathbf{H}_0 = \mathbf{I}_m$. Thus, $\boldsymbol{\xi}_j' = (\mathbf{g}_{j-1}', \ldots, \mathbf{g}_{j-p}')$, where

$$\mathbf{g}_j' = \mathbf{a}'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{G}_j \quad = \quad \mathbf{a}'(\mathbf{\Sigma}^{1/2}\mathbf{H}_j' \otimes \mathbf{\Sigma}^{-1/2}) = \frac{1}{\sqrt{m}}[\mathrm{vec}(\mathbf{H}_j')]' \ . \tag{50}$$

Consequently, the terms $\boldsymbol{\xi}_j'$, $\mathrm{tr}(\widehat{\mathbf{R}}_j)/\sqrt{m}$ and $\mathrm{tr}(\mathbf{R}_j)/\sqrt{m}$ in (49) do not depend on $\mathbf{\Sigma}$. However, $\mathbf{\Sigma}$ does not cancel in both the structures of $(\mathbf{Z}_k'\mathcal{W}^{-1}\mathbf{Z}_k)^{-1}$ and $\mathbf{U}$. This is because the $m^2p \times 1$ vector $\mathbf{U} = (\mathbf{u}_1', \ldots, \mathbf{u}_p')'$ is formed by blocks of the form

$$\mathbf{u}_i = \sum_{j=i}^{k} \mathbf{G}_{j-i}'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2})\mathrm{vec}(\mathbf{C}_j') \ ,$$

$i = 1, \ldots, p$, where $\mathbf{G}_r'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) = (\mathbf{H}_r\mathbf{\Sigma} \otimes \mathbf{I}_m)(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) = \mathbf{H}_r\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{-1/2}$.
■

3 AN INITIAL GOODNESS-OF-FIT PROCESS

3.1 Introduction

In univariate *ARMA(p, q)* models, instead of the standard goodness-of-fit method proposed by Box and Pierce (1970), an alternative possibility is to consider the residual version of the process $\{W_n(u) : 0 \leq u \leq 1\}$, where

$$W_n(u) = \frac{\sqrt{2}}{\pi} \sqrt{n} \sum_{k=1}^{n-1} r_k \frac{\sin(k\pi u)}{k} \ , \tag{51}$$

being the $\{r_k : 1 \leq k \leq n-1\}$, as defined in Section 2.3, the sample autocorrelations of the errors $\{\varepsilon_t : 1 \leq t \leq n\}$. Under adequate regularity conditions, the process of (51) converges weakly as $n \to \infty$ in $C[0,1]$, the space of continuous functions in $[0,1]$, to the Brownian bridge $\{B(u) : 0 \leq u \leq 1\}$. See Durlauf (1991, section 2) and Anderson (1993, section 2) for details.

A natural extension of the process of (51) for *VARMA(p,q)* models is given by $\{W_n^m(u) : 0 \leq u \leq 1\}$, where

$$W_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1} \frac{\operatorname{tr}(\mathbf{R}_k)}{\sqrt{m}} \frac{\sin(k\pi u)}{k} \tag{52}$$

depends on the correlation matrix of the errors $\{\boldsymbol{\varepsilon}_t : 1 \leq t \leq n\}$ given in (35), $\mathbf{R}_k = \mathbf{C}_k' \mathbf{C}_0^{-1}$, introduced by Chitturi (1974). Similarly as (51), $W_n^m(u)$ is not feasible, because it depends on the unobservable errors $\{\boldsymbol{\varepsilon}_t : 1 \leq t \leq n\}$. However, it will serve as a building block for a new procedure of goodness-of-fit in multivariate time series. The basic task now is to proof that, as $n \to \infty$, $W_n^m(u)$ converges weakly in $C[0,1]$ to the Brownian bridge $\{B(u) : 0 \leq u \leq 1\}$. The corresponding derivations require a collection of auxiliary asymptotic tools, that are collected in the next section.

3.2 Auxiliary asymptotic results

The first result refers to convergence (25) in Section 2.2.

Theorem 3.1 Suppose that the error vectors $\{\boldsymbol{\varepsilon}_t\}$ are i.i.d. with $\mathrm{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$, $\mathrm{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma} > 0$, and finite fourth order moments $\mathrm{E}[\|\boldsymbol{\varepsilon}_t\|^4] < +\infty$. Then, as $n \longrightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \operatorname{vec}(\mathbf{C}_1') \\ \operatorname{vec}(\mathbf{C}_2') \\ \vdots \\ \operatorname{vec}(\mathbf{C}_k') \end{pmatrix} \xrightarrow{D} \mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix} , \quad k \geq 1 \ , \tag{30}$$

where $\mathbf{V}_j^*$, $j = 1, \ldots, k$ are i.i.d. $\mathbf{N}_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$, and $\mathcal{W} = \mathbf{I}_k \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$.

Proof. The proof of this result is given in the appendix. ■

This result is related to Chitturi (1976, p. 206 Theorem 1). The assumption $\mathrm{E}[\|\boldsymbol{\varepsilon}_t\|^4] < +\infty$ is really stronger than actually needed. It is included in the statement of the theorem because it is the one that is needed to establish the asymptotic normality of $\sqrt{n}(\mathbf{C}_0' - \boldsymbol{\Sigma})$. Define now the statistics $\overline{\mathbf{R}}_k = \mathbf{C}_k'\boldsymbol{\Sigma}^{-1}$, $k \geq 1$. These are a modification of the $\mathbf{R}_k = \mathbf{C}_k'\mathbf{C}_0^{-1}$ of Chitturi (1974) given in (35), obtained by replacing $\mathbf{C}_0$ by the constant $m \times m$ covariance matrix $\boldsymbol{\Sigma}$.

Theorem 3.2 Under the same assumptions of Theorem 3.1, as $n \to \infty$,

$$\sqrt{n}[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}}_1) \\ \mathrm{tr}(\overline{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\overline{\mathbf{R}}_k) \end{pmatrix}] \xrightarrow{D} N_k(\mathbf{0}, \mathbf{I}_k) \ . \tag{53}$$

Proof. Convergence (53) follows from (25) using the fact

$$\sqrt{n}[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}}_1) \\ \mathrm{tr}(\overline{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\overline{\mathbf{R}}_k) \end{pmatrix}] = (\mathbf{I}_k \otimes \mathbf{a}')\mathcal{W}^{-1/2}[\sqrt{n}\begin{pmatrix} \mathrm{vec}(\mathbf{C}_1') \\ \mathrm{vec}(\mathbf{C}_2') \\ \vdots \\ \mathrm{vec}(\mathbf{C}_k') \end{pmatrix}] \xrightarrow{D} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{pmatrix} \ , \qquad k \geq 1 \ .$$

where the $v_j = \mathbf{a}'\mathbf{V}_j^*$ are i.i.d. $N(0,1)$ random variables, $j = 1, \ldots, k$. ■

Corollary 3.1 Under the same assumptions of Theorem 3.1, as $n \to \infty$,

$$\sqrt{n}[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\mathbf{R}_1) \\ \mathrm{tr}(\mathbf{R}_2) \\ \vdots \\ \mathrm{tr}(\mathbf{R}_k) \end{pmatrix}] \xrightarrow{D} N_k(\mathbf{0}, \mathbf{I}_k) \ . \tag{54}$$

Proof. Using the $km^2 \times km^2$ block diagonal matrix $\mathbf{W} = \mathbf{I}_k \otimes \mathbf{C}_0 \otimes \mathbf{C}_0$, defined in expression (26) of Section 2.2, it can be written

$$\sqrt{n}[\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\mathbf{R}_1) \\ \mathrm{tr}(\mathbf{R}_2) \\ \vdots \\ \mathrm{tr}(\mathbf{R}_k) \end{pmatrix}] = (\mathbf{I}_k \otimes \mathbf{a}')\mathbf{W}^{-1/2}[\sqrt{n} \begin{pmatrix} \mathrm{vec}(\mathbf{C}_1') \\ \mathrm{vec}(\mathbf{C}_2') \\ \vdots \\ \mathrm{vec}(\mathbf{C}_k') \end{pmatrix}] \,.$$

Since $\mathbf{C}_0 \xrightarrow{P} \mathbf{\Sigma}$, it follows that $\mathbf{W} \xrightarrow{P} \mathcal{W}$. On the other hand, using standard properties of the Kronecker product (Fang and Zhang, 1990, p.13), the matrix $\mathcal{W}$ is positive definite. Thus, using a continuity argument similarly as in Brockwell and Davis (1991, Proposition 6.1.4), it can be checked that $\mathbf{W}^{-1/2} \xrightarrow{P} \mathcal{W}^{-1/2}$. Therefore, from Theorem 3.1 and Slutsky's theorem,

$$\sqrt{n}[\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\mathbf{R}_1) \\ \mathrm{tr}(\mathbf{R}_2) \\ \vdots \\ \mathrm{tr}(\mathbf{R}_k) \end{pmatrix}] \xrightarrow{D} (\mathbf{I}_k \otimes \mathbf{a}')\mathcal{W}^{-1/2}\mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{pmatrix} \,,$$

where the v_j, $j = 1, \ldots, k$, are as in Theorem 3.2. ∎

3.3 A result on the convergence of a stochastic process in $C[0, 1]$

The following lemma will be used for establishing the convergence properties of the process of (52). The proof is given in Ubierna and Velilla (2007, section 2).

Lemma 3.1 Let $\{A_n(u) : 0 \leq u \leq 1\}$, $n = 1, 2, \ldots$, and $\{A(u) : 0 \leq u \leq 1\}$ be processes in $C[0, 1]$. Consider a fixed integer $d_0 \geq 1$, and suppose that for each $d \geq d_0$ it can be written $A_n(u) = A_n^d(u) + R_n^d(u)$ and $A(u) = A^d(u) + R^d(u)$, $0 \leq u \leq 1$. Assume also the following three conditions for the processes $\{A_n^d(u) : 0 \leq u \leq 1\}$, $\{A^d(u) : 0 \leq u \leq 1\}$, $\{R_n^d(u) : 0 \leq u \leq 1\}$, and $\{R^d(u) : 0 \leq u \leq 1\}$:

(C.1) For each $d \geq d_0$, the finite-dimensional distributions of the sequence $\{A_n^d(u) : 0 \leq u \leq 1\}$ converge weakly, as $n \to \infty$, to those of $\{A^d(u) : 0 \leq u \leq 1\}$;

(C.2) For each $d \geq d_0$, the probability distributions of the sequence $\{A_n^d(u) : 0 \leq u \leq 1\}$ are tight;

(C.3) For each $\varepsilon > 0$

$$\limsup_d (\limsup_n P[\sup_{0\le u\le 1} |R_n^d(u)| > \varepsilon]) = 0 \ ,$$

and $\limsup_d P[\sup_{0\le u\le 1} |R^d(u)| > \varepsilon] = 0$.

Then, $\{A_n(u) : 0 \le u \le 1\}$ converges weakly in $C[0,1]$, as $n \to \infty$, to the process $\{A(u) : 0 \le u \le 1\}$.

3.4 Convergence of the process $\{W_n^m(u) : 0 \le u \le 1\}$

The convergence properties of (52) will be studied analyzing first those of the auxiliary process $\{\overline{W}_n^m(u) : 0 \le u \le 1\}$, where

$$\overline{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1} \frac{\mathrm{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}} \frac{\sin(k\pi u)}{k} \ , \tag{55}$$

depends on the matrices $\overline{\mathbf{R}}_k = \mathbf{C}_k'\boldsymbol{\Sigma}^{-1}$, rather than $\mathbf{R}_k = \mathbf{C}_k'\mathbf{C}_0^{-1}$, $k \ge 1$.

3.4.1 Convergence of $\{\overline{W}_n^m(u) : 0 \le u \le 1\}$

Proposition 3.1 If the error vectors $\{\boldsymbol{\varepsilon}_t\}$ are i.i.d. with $\mathrm{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$, $\mathrm{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma} > 0$ and finite eighth order moments $\mathrm{E}[\|\boldsymbol{\varepsilon}_t\|^8] < +\infty$, then, as $n \to \infty$,

$$\overline{W}_n^m(u) \to_\omega B(u) \ .$$

Proof. The proof is based on checking the conditions of Lemma 3.1 above. Recall first the standard representation of the Brownian bridge

$$B(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty} v_k \frac{\sin(k\pi u)}{k} \ ,$$

where $\{v_k : k \ge 1\}$ is a sequence of i.i.d. $N(0,1)$ random variables. See Ash and Gardner (1975, Chapter 1). Then, the natural choice for the processes considered in Lemma 3.1 are $A_n(u) = \overline{W}_n^m(u)$ and $A(u) = B(u)$. Moreover, $d_0 = 1$, and

$$A_n^d(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{d} \frac{\mathrm{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}} \frac{\sin(k\pi u)}{k} \quad , \quad R_n^d(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=d+1}^{n-1} \frac{\mathrm{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}} \frac{\sin(k\pi u)}{k} \ .$$

Also

$$A^d(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{d} v_k \frac{\sin(k\pi u)}{k} \quad , \quad R^d(u) = \frac{\sqrt{2}}{\pi}\sum_{k=d+1}^{\infty} v_k \frac{\sin(k\pi u)}{k} \ .$$

In this manner, $A_n(u) = A_n^d(u) + R_n^d(u)$, and $A(u) = A^d(u) + R^d(u)$, $d \geq 1$. It is also convenient to consider the family of $d \times 1$ vectors

$$\mathbf{a}_d(u) = \frac{\sqrt{2}}{\pi}(\sin(\pi u),\ \sin(2\pi u)/2,\ \ldots,\ \sin(d\pi u)/d)' \ , \quad 0 \leq u \leq 1 \ .$$

(C.1) Pick an integer $r \geq 1$, and select $u_1,\ u_2,\ \ldots,\ u_r$ in $[0,1]$. Clearly,

$$A_n^d(u) = \mathbf{a}'_d(u)\sqrt{n}\ [\mathrm{tr}(\overline{\mathbf{R}}_1), \mathrm{tr}(\overline{\mathbf{R}}_2), \cdots, \mathrm{tr}(\overline{\mathbf{R}}_d)]'/\sqrt{m} \ .$$

Therefore, by Theorem 3.2 and the Cramér-Wold device,

$$\begin{pmatrix} A_n^d(u_1) \\ A_n^d(u_2) \\ \vdots \\ A_n^d(u_r) \end{pmatrix} = \mathbf{M}\sqrt{n}[\frac{1}{\sqrt{m}}\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}}_1) \\ \mathrm{tr}(\overline{\mathbf{R}}_2) \\ \vdots \\ \mathrm{tr}(\overline{\mathbf{R}}_d) \end{pmatrix}] \xrightarrow{D} \mathbf{M}\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} = \begin{pmatrix} A^d(u_1) \\ A^d(u_2) \\ \vdots \\ A^d(u_r) \end{pmatrix} ,$$

where $\mathbf{M}$ is a $r \times d$ constant matrix whose jth row is $\mathbf{a}'_d(u_j)$, $j = 1,\ \ldots,\ r$; ■

(C.2) For checking tightness, we use the necessary and sufficient well-known conditions *(i)* and *(ii)* of Theorem 7.3 in Billingsley (1999, p 82). Since $A_n^d(0) = 0$, the sequence of random variables $\{A_n^d(0)\}$, $n = 1,\ 2,\ \ldots$, is clearly bounded in probability. Accordingly, condition *(i)* holds. On the other hand, by the mean value theorem, it is easy to check the inequality

$$|\sin(k\pi u)/k - \sin(k\pi v)/k| \leq \pi\,|u - v| \ , \quad 0 \leq u,\ v \leq 1 \ . \tag{56}$$

As a result of (56), it follows that $|A_n^d(u) - A_n^d(v)| \leq Z_n^d\,|u - v|$, where, as a consequence of Theorem 3.2, the sequence $Z_n^d = (\sqrt{2}/\pi)\sqrt{n}\sum_{k=1}^{d}\left|\mathrm{tr}(\overline{\mathbf{R}}_k)/\sqrt{m}\right|$ is $O_p(1)$. Thus, given $\varepsilon,\ \eta > 0$, there exists $M(\eta) > 0$ such that $P[Z_n^d > M(\eta)] \leq \eta$, $n \geq 1$. Hence, if $0 < \delta = \delta(\varepsilon, \eta) < \min[1, \varepsilon/M(\eta)]$, it follows that

$$P[\sup_{|u-v|<\delta}\left|A_n^d(u) - A_n^d(v)\right| \geq \varepsilon] \leq P[Z_n^d > M(\eta)] \leq \eta \ , \quad n \geq 1 \ .$$

As a conclusion, condition *(ii)* holds. ■

(C.3) First part. For dealing with $R^d(u)$, consider for $d_1 > d$, the following inequality:

$$R^2(d,d_1) = \sup_{0\le u\le 1}\left|\frac{\sqrt{2}}{\pi}\sum_{k=d+1}^{d_1} v_k \frac{\sin(k\pi u)}{k}\right|^2 \le \frac{2}{\pi^2}\sup_{0\le u\le 1}\left|\sum_{k=d+1}^{d_1} v_k \frac{\exp(ik\pi u)}{k}\right|^2 , \qquad (57)$$

where $\exp(ik\pi u) = \cos(k\pi u) + i\sin(k\pi u)$. Introducing the index $r = j - k$, then

$$A(u) = \left|\sum_{k=d+1}^{d_1} v_k \frac{\exp(ik\pi u)}{k}\right|^2 = (\sum_{j=d+1}^{d_1} v_j \frac{\exp(ij\pi u)}{j})(\sum_{k=d+1}^{d_1} v_k \frac{\exp(-ik\pi u)}{k}) ,$$

can written in the form $A(u) = \sum_{r=(d+1)-d_1}^{d_1-(d+1)} Y_r(u)$, where $Y_0(u) = \sum_{j=d+1}^{d_1}(v_j/j)^2$ and

$$Y_r(u) = \sum_{k=d+1}^{d_1-r} \exp[ir\pi u]\frac{v_k v_{k+r}}{k(k+r)} , \; r>0; \quad Y_r(u) = \sum_{j=d+1}^{d_1-|r|} \exp[ir\pi u]\frac{v_j v_{j+|r|}}{j(j+|r|)} , \; r<0 .$$

Notice then that $|A(u)| \le A = Y_0 + 2\sum_{r=1}^{d_1-(d+1)}|Y_r|$, where $Y_0 = Y_0(1)$ and

$$Y_r = \sum_{k=d+1}^{d_1-r}\frac{v_k v_{k+r}}{k(k+r)} \quad , \quad r>0 .$$

Then, using (57), $E[R^2(d,d_1)] \le (2/\pi^2)E[A] \le (4/\pi^2)(E[Y_0] + \sum_{r=1}^{d_1-(d+1)} E[|Y_r|])$. On the other hand, $E[Y_0] = \sum_{j=d+1}^{d_1} 1/j^2$. Also, for $r>0$,

$$E[|Y_r|] \le E^{1/2}[|Y_r|^2] \le \left[\sum_{k=d+1}^{d_1-r}\sum_{j=d+1}^{d_1-r}\frac{|E[v_k v_{k+r} v_j v_{j+r}]|}{k(k+r)j(j+r)}\right]^{1/2} = \left[\sum_{k=d+1}^{d_1-r}\frac{1}{k^2(k+r)^2}\right]^{1/2} .$$

Accordingly,

$$E[R^2(d,d_1)] \le \frac{4}{\pi^2}\sum_{k=d+1}^{d_1}\frac{1}{k^2} + \frac{4}{\pi^2}\sum_{r=1}^{d_1-(d+1)}[\sum_{k=d+1}^{d_1-r}\frac{1}{k^2(k+r)^2}]^{1/2} . \qquad (58)$$

Proceeding now as in Grenander and Rosenblatt (1957, p.189), take $d = 2^p$ and $d_1 = 2^{p+1}$. Then, $d_1 - d = 2^{p+1} - 2^p = 2^p$. Hence, it follows that

$$\sum_{k=d+1}^{d_1}\frac{1}{k^2} = \sum_{k=2^p+1}^{2^{p+1}}\frac{1}{k^2} = \frac{1}{(2^p+1)^2} + \frac{1}{(2^p+2)^2} + ... + \frac{1}{(2^{p+1})^2} \le$$

$$\le \frac{1}{2^{2p}} + \frac{1}{2^{2p}} + ... + \frac{1}{2^{2p}} \le \frac{2^p}{2^{2p}} = \frac{1}{2^p} \le \frac{1}{2^{p/2}} ,$$

and

$$\sum_{r=1}^{d_1-(d+1)} [\sum_{k=d+1}^{d_1-r} \frac{1}{k^2(k+r)^2}]^{1/2} \leq \sum_{r=1}^{2^p}[\sum_{k=2^p+1}^{2^{p+1}} \frac{1}{k^2(k+r)^2}]^{1/2} \leq \sum_{r=1}^{2^p}[\sum_{k=2^p+1}^{2^{p+1}} \frac{1}{k^4}]^{1/2} \leq$$

$$\leq \sum_{r=1}^{2^p}[\frac{2^p}{2^{4p}}]^{1/2} = \frac{2^p}{2^{3p/2}} = \frac{1}{2^{p/2}} .$$

Thus, going back to (58), $E[R(d,d_1)] \leq E^{1/2}[R^2(d,d_1)] \leq A_1 2^{-p/4}$, where $A_1 = (8/\pi^2)^{1/2}$ is a positive constant. Note that $R(d,d_1) = \sup_{0\leq u\leq 1} \left|R^{d_1}(u) - R^d(u)\right|$. Now let's define $Z_d = \sup_{0\leq u\leq 1} \left|R^d(u)\right| \leq \sum_{j=0}^{\infty} R(2^{p+j}, 2^{p+j+1})$. Then,

$$E(Z_d) \leq \sum_{j=0}^{\infty} E[R(2^{p+j}, 2^{p+j+1})] \leq A_1 \sum_{j=p}^{\infty} \frac{1}{2^{j/4}} \longrightarrow 0 \quad , \quad p \longrightarrow \infty .$$

This argument above shows that the sequence of random variables $\{Z_d\}$ converges to zero in probability as $d \to \infty$. This is enough to guarantee that $\limsup_d P[\sup_{0\leq u\leq 1} |R^d(u)| > \varepsilon] = 0$ for all $\varepsilon > 0$; ∎

Second part. The proof of this part is based on the following result:

Lemma 3.2 Consider a sequence of i.i.d random vectors $\{\mathbf{u}_t\}$ with $E[\mathbf{u}_t] = \mathbf{0}$ and $\text{Var}[\mathbf{u}_t] = \mathbf{I}_m$. Writing $\mathbf{u}_t = (u_{t,1}, \ldots, u_{t,m})'$, consider the following empirical covariances

$$g_{k,IJ} = \frac{1}{n}\sum_{t=1}^{n-k} u_{t,I}u_{t+k,J} ,$$

for $I, J = 1, \ldots, m$. If the vectors $\mathbf{u}_t$ have finite eighth order moments $\text{E}[\|\mathbf{u}_t\|^8] < +\infty$, then, there exist constants $A_1 > 0$ and $A_2 > 0$ such that, for $k, l, I, J \geq 1$,

$$|E[g_{k,IJ}g_{k+r,IJ}g_{l,IJ}g_{l+r,IJ}]| \leq \begin{cases} A_1/n^3, & k \neq l; \\ A_2/n^2, & k = l. \end{cases}$$

Proof. The proof of this result is obtained proceeding as in Grenander and Rosenblatt (1957, p. 186-189). ∎

To apply Lemma 3.2 to the task of bounding the tail sums of the process $\{R_n^d(u) : 0 \leq u \leq 1\}$, let's consider the particular case in which $\mathbf{u}_t = \mathbf{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t = (u_{t,1}, \ldots, u_{t,m})'$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,m})'$. Under the assumptions of Proposition 3.1 for the moments of the error vectors $\{\boldsymbol{\varepsilon}_t\}$, the assumptions of Lemma 3.2 are satisfied. To verify this, notice that each component of the error vector $\mathbf{u}_t$ is a linear combination of the original $\boldsymbol{\varepsilon}_t$. More precisely, if $\boldsymbol{\Sigma}^{-1/2} = (\sigma^{ij} : i, j = 1, \ldots, m)$, then

$$u_{t,i} = \sigma^{i1}\varepsilon_{t,1} + \ldots + \sigma^{im}\varepsilon_{t,m} \quad , \quad i = 1, \ldots, m \ .$$

As a result, $|u_{t,i}| \leq \sigma^i(\sum_{j=1}^m |\varepsilon_{t,j}|)$, where $\sigma^i = \max_{1\leq j\leq m} |\sigma^{ij}|$. Then, as a standard consequence of Hölder's inequality, it follows that $E(|u_{t,i}|^8) < +\infty$, $i = 1, \ldots, m$.

On the other hand, $\mathrm{tr}(\overline{\mathbf{R}}_k) = \mathrm{tr}(\mathbf{C}_k'\boldsymbol{\Sigma}^{-1}) = \mathrm{tr}(\boldsymbol{\Sigma}^{-1/2}\mathbf{C}_k\boldsymbol{\Sigma}^{-1/2})$. Thus,

$$\mathrm{tr}(\overline{\mathbf{R}}_k) = \frac{1}{n}\mathrm{tr}[\sum_{t=1}^{n-k}(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t)(\boldsymbol{\varepsilon}_{t+k}'\boldsymbol{\Sigma}^{-1/2})] = \frac{1}{n}\sum_{I=1}^{m}\sum_{t=1}^{n-k} u_{t,I}u_{t+k,I} = \sum_{I=1}^{m} g_{k,II} \ ,$$

where $g_{k,II} = (1/n)\sum_{t=1}^{n-k} u_{t,I}u_{t+k,I}$. Therefore, for $n-1 > d$,

$$R_n^2(d,n) = \sup_{0\leq u\leq 1}\left|\frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=d+1}^{n-1}\frac{\mathrm{tr}(\overline{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k}\right|^2 \leq \sum_{I=1}^{m}\frac{2n}{m\pi^2}\sup_{0\leq u\leq 1}\left|\sum_{k=d+1}^{n-1} g_{k,II}\frac{\sin(k\pi u)}{k}\right|^2 \ .$$

Using the same arguments to those considered in the first part, it can be then seen that $E[R_n^2(d,n)] \leq m^{-1}\sum_{I=1}^m (4n/\pi^2)\sum_{r=0}^{n-1-(d+1)} E[|X_r|]$, where

$$X_r = \sum_{k=d+1}^{n-1-r}\exp[ir\pi u]\frac{g_{k,II}g_{k+r,II}}{k(k+r)} \ , \quad r \geq 0; \quad X_r = \sum_{j=d+1}^{n-1-|r|}\exp[ir\pi u]\frac{g_{j,II}g_{j+|r|,II}}{j(j+|r|)} \ , \quad r \leq 0 \ .$$

But, for $r \geq 0$,

$$E[|X_r|^2] \leq \sum_{k=d+1}^{n-1-r}\sum_{j=d+1}^{n-1-r}\frac{|E[g_{k,II}g_{k+r,II}g_{j,II}g_{j+r,II}]|}{k(k+r)j(j+r)} \ ,$$

where, by Lemma 3.2,

$$|E[g_{k,II}g_{k+r,II}g_{j,II}g_{j+r,II}]| \leq \begin{cases} B_1/n^3, & k \neq j; \\ B_2/n^2, & k = j. \end{cases} \ ,$$

where B_1 and B_2 are positive constants. Therefore, applying Cauchy-Schwartz inequality, it follows that

$$\begin{aligned}
E[|X_r|^2] &\leq \frac{B_2}{n^2}\sum_{k=d+1}^{n-1-r}\frac{1}{k^2(k+r)^2}+\frac{B_1}{n^3}\sum_{k=d+1}^{n-1-r}\frac{1}{k(k+r)}\sum_{j=d+1}^{n-1-r}\frac{1}{j(j+r)} \leq \\
&\leq \frac{B_2}{n^2}\sum_{k=d+1}^{n-1-r}\frac{1}{k^2(k+r)^2}+\frac{B_1}{n^3}(n-1-r-d)\sum_{k=d+1}^{n-1-r}\frac{1}{k^2(k+r)^2} \leq \\
&\leq \frac{B_3}{n^2}\sum_{k=d+1}^{n-1-r}\frac{1}{k^2(k+r)^2} ,
\end{aligned}$$

where B_3 is positive constant. This leads to the inequality,

$$E[|X_r|] \leq E^{1/2}[|X_r|^2] \leq \frac{B_3^{1/2}}{n}[\sum_{k=d+1}^{n-1-r}\frac{1}{k^2(k+r)^2}]^{1/2} . \tag{59}$$

Using (59),

$$\begin{aligned}
E[R_n^2(d,n)] &\leq \frac{4n}{m\pi^2}\sum_{I=1}^{m}\sum_{r=0}^{n-1-(d+1)}\frac{B_3^{1/2}}{n}[\sum_{k=d+1}^{n-1-r}\frac{1}{k^2(k+r)^2}]^{1/2} \leq \\
&\leq \frac{4B_3^{1/2}}{\pi^2}\sum_{r=0}^{n-1-(d+1)}[\sum_{k=d+1}^{n-1-r}\frac{1}{k^2(k+r)^2}]^{1/2} .
\end{aligned}$$

Taking now $d = 2^p$ and $n-1 = 2^{p+1}$, and proceeding as in the first part, then $E[R_n(d,n)] \leq B2^{-p/4}$, where B is a positive constant. To finish the proof, define $Z_{d,n} = \sup_{0\leq u\leq 1}\left|R_n^d(u)\right| \leq \sum_{j=0}^{\infty} R_n(2^{p+j}, 2^{p+j+1})$. But $E(Z_{d,n}) \leq \sum_{j=0}^{\infty} E[R_n(2^{p+j}, 2^{p+j+1})]$ $\leq B\sum_{j=p}^{\infty} 2^{-j/4} \longrightarrow 0$ as both $p, n \longrightarrow \infty$. Using finally the elementary inequality $P[\sup_{0\leq u\leq 1}|R_n^d(u)| > \varepsilon] \leq E(Z_{d,n})/\varepsilon$, the proof of **(C.3)** is finished. ■

3.4.2 Application to the convergence of $\{W_n^m(u) : 0 \leq u \leq 1\}$

In this section, the goal is to establish the convergence of the process of (52) to the Brownian bridge. Since $\mathbf{R}_k = \mathbf{C}_k'\mathbf{C}_0^{-1}$, proceeding as in expression (38), it follows that

$$\frac{1}{\sqrt{m}}\,\mathrm{tr}(\mathbf{R}_k) = \frac{1}{\sqrt{m}}\,\mathrm{tr}(\mathbf{C}_0^{-1/2}\mathbf{C}_k\mathbf{C}_0^{-1/2}) = \mathbf{a}'(\mathbf{C}_0^{-1/2}\otimes\mathbf{C}_0^{-1/2})\mathrm{vec}(\mathbf{C}_k) ,$$

where $\mathbf{a} = \mathrm{vec}(\mathbf{I}_m)/\sqrt{m}$. Therefore, the processes $\{W_n^m(u) : 0 \leq u \leq 1\}$ and $\{\overline{W}_n^m(u) : 0 \leq u \leq 1\}$ can be written in the form

$$W_n^m(u) = \mathbf{a}_n'\mathbf{U}_n^m(u) \quad \text{and} \quad \overline{W}_n^m(u) = \mathbf{a}'\mathbf{U}_n^m(u) , \tag{60}$$

where $\mathbf{a}_n = (\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{1/2})(\mathbf{C}_0^{-1/2} \otimes \mathbf{C}_0^{-1/2})\mathbf{a}$ and

$$\mathbf{U}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\,(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2}) \sum_{k=1}^{n-1} \text{vec}(\mathbf{C}_k)\frac{\sin(k\pi u)}{k} \tag{61}$$

are random vectors of $m^2 \times 1$. Define now $D_n(u) = W_n^m(u) - \overline{W}_n^m(u)$, $0 \leq u \leq 1$.

Lemma 3.3 Under the same assumptions for the errors given in the statement of Proposition 3.1, $\sup\limits_{0\leq u\leq 1} |D_n(u)| = o_P(1)$.

Proof. Using (60), it can be written

$$D_n(u) = W_n^m(u) - \overline{W}_n^m(u) = (\mathbf{a}_n - \mathbf{a})'\mathbf{U}_n^m(u) \ ,$$

where $\mathbf{a}_n$ and $\mathbf{U}_n^m(u)$ are as given in (61). Taking into account that $\mathbf{C}_0 \xrightarrow{P} \mathbf{\Sigma} > 0$, it easily follows that $\mathbf{a}_n \xrightarrow{P} (\mathbf{\Sigma}^{1/2} \otimes \mathbf{\Sigma}^{1/2})(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\mathbf{a} = (\mathbf{I}_m \otimes \mathbf{I}_m)\mathbf{a} = \mathbf{I}_{m^2}\mathbf{a} = \mathbf{a}$. To finish the proof of the lemma, it is then enough to check that all the coordinates $[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u)$, $I, J = 1, \ldots, m$, of the $m^2 \times 1$ process $\{\mathbf{U}_n^m(u) : 0 \leq u \leq 1\}$ are bounded in probability. In fact, for each choice of $I, J = 1, \ldots, m$, it follows that

$$[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1}[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\text{vec}(\mathbf{C}_k)\frac{\sin(k\pi u)}{k} \ . \tag{62}$$

Using well-known properties of the trace and vec operators, it can be written $[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'$ $(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\text{vec}(\mathbf{C}_k) = \text{tr}(\mathbf{\Sigma}^{-1/2}\mathbf{C}_k\mathbf{\Sigma}^{-1/2}\mathbf{e}_J\mathbf{e}_I') = \mathbf{e}_I'\mathbf{\Sigma}^{-1/2}\mathbf{C}_k\mathbf{\Sigma}^{-1/2}\mathbf{e}_J$. But, considering definition (21) for the matrix $\mathbf{C}_k$ and putting again $\mathbf{u}_t = \mathbf{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t = (u_{t,1}, \ldots, u_{t,m})'$, it follows that $[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{\Sigma}^{-1/2})\text{vec}(\mathbf{C}_k) = \sum_{t=1}^{n-k} u_{t,I}u_{t+k,J}/n = g_{k,IJ}$, where $g_{k,IJ}$ is as considered in Lemma 3.2. Therefore, the left-hand side of (62) is just

$$[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1} g_{k,IJ}\frac{\sin(k\pi u)}{k} \ . \tag{63}$$

Since $\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'$ is a unit vector, the same type of techniques used in the proof of Proposition 3.1 lead to $[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u) \to_\omega B(u)$. Notice also the identity $\overline{W}_n^m(u) = \sum_{I=1}^m [\text{vec}(\mathbf{e}_I\mathbf{e}_I')]'\mathbf{U}_n^m(u)/\sqrt{m}$.

Finally, the probability bound $M_n = \max_{I,J}\sup_{0\leq u\leq 1} |[\text{vec}(\mathbf{e}_I\mathbf{e}_J')]'\mathbf{U}_n^m(u)| = O_P(1)$ is obtained. Consequently,

$$\sup_{0\leq u\leq 1} |D_n(u)| \leq m^2 \sup_{I,J} |\mathbf{a}_n(I, J) - \mathbf{a}(I, J)| M_n = o_P(1) \ . \quad \blacksquare$$

Theorem 3.3 Under the same assumptions for the errors given in the statement of Proposition 3.1, as $n \to \infty$

$$W_n^m(u) \to_\omega B(u) \ .$$

Proof. By Theorem 7.1 in Billingsley (1999, p.80), it suffices to prove that: *a)* the finite dimensional distributions of $\{W_n^m(u) : 0 \leq u \leq 1\}$ converge weakly to those of $\{B(u) : 0 \leq u \leq 1\}$; and *b)* the sequence of probability distributions of $\{W_n^m(u) : 0 \leq u \leq 1\}$ is tight. Write $W_n^m(u) = \overline{W}_n^m(u) + D_n(u)$, , $0 \leq u \leq 1$. For *a)*, simply write

$$\begin{pmatrix} W_n(u_1) \\ W_n(u_2) \\ \vdots \\ W_n(u_r) \end{pmatrix} = \begin{pmatrix} \overline{W}_n(u_1) \\ \overline{W}_n(u_2) \\ \vdots \\ \overline{W}_n(u_r) \end{pmatrix} + \begin{pmatrix} D_n(u_1) \\ D_n(u_2) \\ \vdots \\ D_n(u_r) \end{pmatrix} \ . \tag{64}$$

From Lemma 3.1 and Proposition 3.1, the first summand at the left-hand side of (64) converges in distribution to the $r \times 1$ random vector $(B(u_1), \ldots, B(u_r))'$. On the other hand, from Lemma 3.3 the second summand of (64) is $o_P(1)$. Therefore, by Slutsky's theorem, it also follows that $(W_n(u_1), \ldots, W_n(u_r))' \to_\omega (B(u_1), \ldots, B(u_r))'$.

On the other hand, tightness of $\{W_n^m(u) : 0 \leq u \leq 1\}$, part *(b)*, can be checked using the necessary and sufficient conditions *(i)* and *(ii)* of Theorem 7.3 in Billingsley (1999, p 82). Condition *(i)* holds trivially, since $W_n^m(0) = 0 = O_P(1)$. On the other hand, it is easy to verify the inequality

$$P[\sup_{|u-v|<\delta} |W_n^m(u) - W_n^m(v)| \geq \varepsilon] \leq$$

$$\leq P[\sup_{|u-v|<\delta} \left|\overline{W}_n^m(u) - \overline{W}_n^m(v)\right| \geq \varepsilon/2] + P[\sup_{0\leq u\leq 1} |D_n(u)| \geq \varepsilon/4] \ . \tag{65}$$

From Lemma 3.1 and Proposition 3.1 the probability distributions of $\{\overline{W}_n^m(u) : 0 \leq u \leq 1\}$ are tight. Moreover, from Lemma 3.3 $\sup_{0\leq u\leq 1} |D_n(u)| = o_P(1)$. Therefore, given $\varepsilon, \eta > 0$, there exist $0 < \delta < 1$ and n_1 such that the first summand of (65) is bounded above by $\eta/2$ for $n \geq n_1$. There exists also n_2 such that the second summand of (65) is below $\eta/2$ for $n \geq n_2$. In summary, condition *(ii)* of Theorem 7.3 is satisfied for δ and $n \geq \max(n_1, n_2)$. ■

4 WORK IN PROGRESS

In applications, the error vectors $\boldsymbol{\varepsilon}_t$ are not observable, and must be then replaced by the corresponding residuals $\widehat{\boldsymbol{\varepsilon}}_t$. Therefore, a natural process to be considered in applications is $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$, where

$$\widehat{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(k\pi u)}{k} \tag{66}$$

depends on $\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}_k\widehat{\mathbf{C}}_0^{-1}$, the correlation matrix of the residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$.

The expression (66) generalizes the process defined in Ubierna and Velilla (2007, Section 1.), $\{\widehat{W}_n(u) : 0 \leq u \leq 1\}$, where

$$\widehat{W}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \widehat{r}_k \frac{\sin(k\pi u)}{k} \ , \tag{67}$$

and $\{\widehat{r}_k : 1 \leq k \leq n - (P+1)\}$ is the univariate residual correlogram. Put $f(\omega) = (2\pi)^{-1} \mid \theta(e^{-i\omega}) \mid^2 / \mid \phi(e^{-i\omega}) \mid^2$, $-\pi \leq \omega \leq \pi$, for the standardized spectral density of a univariate *ARMA(p,q)* process of the form $\phi(B)(X_t - \mu) = \theta(B)\varepsilon_t$. Let also $\mathbf{I}(\boldsymbol{\Lambda})$ denote the $(p+q) \times (p+q)$ information matrix for the $(p+q) \times 1$ vector of parameters $\boldsymbol{\Lambda} = (\phi_1, \ldots, \phi_p; \theta_1, \ldots, \theta_q)'$ of this process. Ubierna and Velilla (2007, Theorem 2.1, p. 2905) established that, as $n \to \infty$, $\widehat{W}_n(u) \to_w G(u)$, where $\{G(u) : 0 \leq u \leq 1\}$ is a zero mean Gaussian process with covariance function

$$\gamma(u, v) = [\min(u, v) - uv] - \frac{1}{2\pi^2}\,\mathbf{g}(\pi u)'\mathbf{I}^{-1}(\boldsymbol{\Lambda})\mathbf{g}(\pi v) \ , \tag{68}$$

where $\mathbf{g}(\pi u) = \int_0^{\pi u}[\partial \log f(t)/\partial\boldsymbol{\Lambda}]dt$. The first summand in (68) is the covariance function of the Brownian bridge $\{B(u) : 0 \leq u \leq 1\}$. The second summand is a parametric p.d. quadratic form that does not depend on σ^2.

In the $m > 1$ case, considering the spectral density matrix of the model (1), $\mathbf{f}(\omega, \boldsymbol{\Lambda}) = \frac{1}{2\pi}\boldsymbol{\Phi}^{-1}(e^{i\omega})\boldsymbol{\Theta}(e^{i\omega})\boldsymbol{\Sigma}\boldsymbol{\Theta}'(e^{-i\omega})\boldsymbol{\Phi}'^{-1}(e^{-i\omega})$, $-\pi \leq \omega \leq \pi$, it can be motivated and conjectured that the process of (66) converges weekly to a zero mean Gaussian process $\{G^m(u) : 0 \leq u \leq 1\}$ with covariance function of the form

$$\gamma^m(u, v) = [\min(u, v) - uv] - \frac{1}{2\pi^2 m}\,\boldsymbol{\Gamma}(\pi u)'\mathbf{I}^{-1}(\boldsymbol{\Lambda})\boldsymbol{\Gamma}(\pi v) \ , \tag{69}$$

where $\mathbf{\Gamma}(\pi u)$ is a $m^2(p+q) \times 1$ vector of coordinates

$$\int_0^{\pi u} \mathrm{tr}[\mathbf{f}^{-1}(\omega, \mathbf{\Lambda}) \partial \mathbf{f}(\omega, \mathbf{\Lambda}) / \partial \lambda_i] d\omega \ , \qquad i = 1, \ldots, p+q \ . \tag{70}$$

The structure of (70) is similar to that of the univariate case, given in (68). However, by the results of Section 1.4, both $\mathbf{\Gamma}(\pi u)$ and the $m^2(p+q) \times m^2(p+q)$ information matrix $\mathbf{I}(\mathbf{\Lambda})$ are not scale free, and they both depend on the covariance matrix $\mathbf{\Sigma}$ of the errors. Further work is in progress.

APPENDIX: PROOF OF THEOREM 3.1

Recall the definition $\mathbf{C}_j = n^{-1} \sum_{t=1}^{n-j} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j}$ given in (21). We proof

$$\sqrt{n} \begin{pmatrix} \mathrm{vec}(\mathbf{C}_1) \\ \mathrm{vec}(\mathbf{C}_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}_k) \end{pmatrix} \xrightarrow{D} \mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix} \ , \qquad k \geq 1 \ ,$$

by showing that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+1}) \\ \mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+2}) \\ \vdots \\ \mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+k}) \end{pmatrix} \xrightarrow{D} \mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix} \ . \tag{A.1}$$

Consider the sequence of random variables $\{X_t : t \in \mathbb{Z}\}$ such that

$$X_t = \sum_{j=1}^{k} [\mathrm{vec}(\boldsymbol{\xi}_j)]' \mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j}) = \sum_{j=1}^{k} \mathrm{tr}(\boldsymbol{\xi}'_j \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j}) = \sum_{j=1}^{k} \boldsymbol{\varepsilon}'_t \boldsymbol{\xi}_j \boldsymbol{\varepsilon}_{t+j} \ ,$$

where $\boldsymbol{\xi}_j$ is a constant $m \times m$ matrix, $j = 1, \ldots, k$. Under the i.i.d. assumption on the $\{\boldsymbol{\varepsilon}_t\}$, the sequence $\{X_t : t \in \mathbb{Z}\}$ is strictly stationary. On the other hand, the sets $\{X_t : t \leq 0\}$ and $\{X_t : t \geq k+1\}$ are independent. Therefore, the sequence $\{X_t : t \in \mathbb{Z}\}$ is also k-dependent. Moreover,

$$E[X_t] = E[\boldsymbol{\varepsilon}'_t \boldsymbol{\xi}_j \boldsymbol{\varepsilon}_{t+j}] = E[\mathrm{tr}(\boldsymbol{\xi}'_j \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j})] = \mathrm{tr}[\boldsymbol{\xi}'_j E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j})] = \mathrm{tr}[\boldsymbol{\xi}'_j \mathrm{Cov}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j})] = 0 \ .$$

Additionally, the covariance function is given by

$$\gamma(h) = E[X_t X_{t+h}] =$$

$$= [\mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k)]' E\left[\begin{pmatrix} \mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+1}) \\ \mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+2}) \\ \vdots \\ \mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+k}) \end{pmatrix} \begin{pmatrix} \mathrm{vec}(\boldsymbol{\varepsilon}_{t+h} \boldsymbol{\varepsilon}'_{t+h+1}) \\ \mathrm{vec}(\boldsymbol{\varepsilon}_{t+h} \boldsymbol{\varepsilon}'_{t+h+2}) \\ \vdots \\ \mathrm{vec}(\boldsymbol{\varepsilon}_{t+h} \boldsymbol{\varepsilon}'_{t+h+k}) \end{pmatrix}'\right] \mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k) \ . \tag{A.2}$$

Recalling that $\mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j}) = \boldsymbol{\varepsilon}_{t+j} \otimes \boldsymbol{\varepsilon}_t$, it follows that

$$E\{\mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j})[\mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+h})]'\} = E[(\boldsymbol{\varepsilon}_{t+j} \otimes \boldsymbol{\varepsilon}_t)(\boldsymbol{\varepsilon}'_{t+h} \otimes \boldsymbol{\varepsilon}'_t)] = E[\boldsymbol{\varepsilon}_{t+j} \boldsymbol{\varepsilon}'_{t+h} \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t] \ .$$

Accordingly, by the law of iterated expectations, $E\{\mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+j})[\mathrm{vec}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+h})]'\} = E[\boldsymbol{\varepsilon}_{t+j} \boldsymbol{\varepsilon}'_{t+h} \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t] = E(E[\boldsymbol{\varepsilon}_{t+j} \boldsymbol{\varepsilon}'_{t+h} \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \mid \boldsymbol{\varepsilon}_t]) = E[\mathrm{Cov}(\boldsymbol{\varepsilon}_{t+j}, \boldsymbol{\varepsilon}_{t+h}) \otimes \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t] = \mathrm{Cov}(\boldsymbol{\varepsilon}_{t+j}, \boldsymbol{\varepsilon}_{t+h}) \otimes E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t]$. Note that this expectation is $\mathbf{0}$, when $j \neq h$ and is $\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$, when $j = h$. Thus, from expression (A.2), it follows that $\gamma(h) = 0$ for $h \geq 1$, and $\gamma(0) = [\mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k)]' \mathcal{W} \mathrm{vec}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k)$. Using now Theorem 6.4.2 in Brockwell and Davis (1987, p.206), the convergence below holds:

$$[\mathrm{vec}\,(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k)]' \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \mathrm{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+1}\right) \\ \mathrm{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+2}\right) \\ \vdots \\ \mathrm{vec}\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+k}\right) \end{pmatrix}\right] = \sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} X_t\right) \xrightarrow{D}$$

$$\xrightarrow{D} [\mathrm{vec}\,(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_k)]' \, \mathcal{W}^{1/2} \begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix} . \tag{A.3}$$

By the Cramér-Wold device, (A.3) finishes the proof of (A.1).

Now consider a $m^2 \times m^2$ commutation matrix $\mathbf{K}_{mm}$ of order m, and the $km^2 \times km^2$ matrix $\mathbf{K} = \mathrm{diag}(\mathbf{K}_{mm}, \overset{(k)}{\cdots}, \mathbf{K}_{mm})$. Using the identity $\mathrm{vec}(\mathbf{C}'_k) = \mathbf{K}_{mm} \mathrm{vec}(\mathbf{C}_k)$, it follows that

$$\sqrt{n}\begin{pmatrix} \mathrm{vec}(\mathbf{C}'_1) \\ \mathrm{vec}(\mathbf{C}'_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}'_k) \end{pmatrix} = \mathbf{K}\left[\sqrt{n}\begin{pmatrix} \mathrm{vec}(\mathbf{C}_1) \\ \mathrm{vec}(\mathbf{C}_2) \\ \vdots \\ \mathrm{vec}(\mathbf{C}_k) \end{pmatrix}\right] \xrightarrow{D} \mathbf{K}\mathcal{W}^{1/2}\begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix} \stackrel{D}{\equiv} \mathcal{W}^{1/2}\begin{pmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \\ \vdots \\ \mathbf{V}_k^* \end{pmatrix}. \tag{A.4}$$

The equivalence in distribution at the right-hand side of (A.4) follows from the identity $\mathbf{K}_{mm}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{K}_{mm} = \boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}$, that is a consequence of equation (24) in Lütkepohl (2005, p. 664). Then, since $\mathbf{K}_{mm} = \mathbf{K}'_{mm}$, both $\mathbf{K}_{mm}(\boldsymbol{\Sigma}^{1/2}\otimes\boldsymbol{\Sigma}^{1/2})\mathbf{V}_j$ and $(\boldsymbol{\Sigma}^{1/2}\otimes\boldsymbol{\Sigma}^{1/2})\mathbf{V}_j$ have the same distribution $\mathbf{N}_{m^2}[\mathbf{0}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}]$, $j = 1, \ldots, k$. ∎

5 References

Anderson, T. W., 1993. Goodness of fit tests for spectral distributions. The Annals of Statistics 21 (2), pp. 830–847.

Ash, R. B., Gardner, M. F., 1975. Topics in stochastic processes. Academic Press, New York.

Billingsley, P., 1999. Convergence of Probability Measures. Wiley, New York.

Box, G. E. P., Pierce, D. A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association 65 (332), pp. 1509–1526.

Brockwell, P. J., Davis, R. A., 1987. Time series: Theory and methods. SpringerVerlag, New York.

Brockwell, P. J., Davis, R. A., 1991. Time series: Theory and methods. 2nd Edition, SpringerVerlag, New York.

Chitturi, R. V., 1974. Distribution of residual autocorrelations in multiple autoregressive schemes. Journal of the American Statistical Association 69 (348), 928–934.

Chitturi, R. V., 1976. Distribution of multivariate white noise autocorrelations. Journal of the American Statistical Association 71 (353), pp. 223–226.

Dunsmuir, W., Hannan, E. J., 1976. Vector linear time series models. Advances in Applied Probability 8 (2), 339–364.

Durlauf, S. N., 1991. Spectral based testing of the martingale hypothesis. Journal of Econometrics 50 (3), 355 – 376.

Fang, K.-T., Zhang, Y.-T., 1990. Generalized Multivariate Analysis. Springer-Verlag.

Grenander, U., Rosenblatt, M., 1957. Statistical analysis of stationary time series. John Wiley and Sons, New York.

Hannan, E., 1969. The indentification of mixed vector-autoregressive moving average systems. Biometrica 56, 223–225.

Hosking, J. R. M., 1980. The multivariate portmanteau statistic. Journal of the American Statistical Association 75 (371), 602–608.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer.

Ubierna, A., Velilla, S., 2007. A goodness-of-fit process for arma(p,q) models based on a modified residual autocorrelation sequence. Journal of Statistical Planning and Inference 137 (9), 2903 – 2919.