Chi-Squared Portmanteau Statistics for Vector Autoregressive Models with Uncorrelated Errors

著者	Katayama Naoya
journal or publication title	Kansai University review of economics
volume	13
page range	1-23
year	2011-03
URL	http://hdl.handle.net/10112/00017190

Kansai University Review of Economics
No. 13 (March 2011), pp. 1–23

Chi-Squared Portmanteau Statistics for Vector Autoregressive Models with Uncorrelated Errors†

NAOYA KATAYAMA*

The portmanteau statistic based on the first *m* residual autocorrelations is used for testing the goodness-of-fit for vector autoregressive models with varying *m*. However, it is known that existing portmanteau statistics are approximately non chi-squared distributions in the presence of non-independent innovations. In this paper we propose a new portmanteau statistic that is asymptotically chi-squared even in the presence of non-independent innovations. We also study the joint probability of the multiple portmanteau statistics with different degrees of freedom.Monte Carlo experiments illustrate the finite sample performance for the proposed portmanteau test.

Keyword: Vector weak AR model; Portmanteau test; Residual autocorrelation; Goodness-of-fit test; Multiple tests.

1. Introduction

Portmanteau test statistics are commonly used in time series analysis, which is used as a goodness-of-fit test statistic and is defined by the sum of squares of the first m residual autocorrelations. This approach was first presented by Box & Pierce (1970) for the univariate autoregressive (AR) models. Chitturi (1974) extended the test statistics for vector AR models. Li (2004) reviewed applications of the portmanteau statistics in various time series models.

In these portmanteau statistics, the errors are generally supposed to be independent and identically distributed (i.i.d.), and are referred to as strong white noise (see, for example, Francq et al., (2005) (referred to as FRZ) and Francq and Raïssi (2007) (referred to as

* Faculty of Economics, Kansai University, 3-3-35 Yamate-cho, Suita, Osaka 564-8680, JAPAN. kata@kyudai.jp
† Revision Date: November 16, 2010. This work was supported by the Japan Society for the Promotion of Science.

FR)). The full versions of these manuscripts are available from the first author's website, and have frequent examples and expositions. These authors also claim the following: (i) the assumption of strong white noise is restrictive because it precludes conditional heteroskedasticity and/or other forms of nonlinearity where the errors are uncorrelated but non-independent and known as weak white noise. (ii) If we assume the errors are strong white noise, the existing portmanteau statistics are asymptotically chi-squared random variable. However, if the errors are weak white noise, they are not always asymptotically chi-squared random variable but are the weighted sums of the chi-squared random variables. (iii) Therefore if the errors are weak white noise, the chi-squared test by the portmanteau statistics results in an over-rejection problem. Therefore the present portmanteau test may unnecessarily provide a higher order (vectorized) AR model.

To solve (iii), they proposed estimation methods of the critical regions using the weighted sum of the chi-squared random variables, which uses Imhof's (1961) formula or other approximate methods (see, for example, Johnson et al., (1994, Section 8 in Chapter 18)). From the simulation experiments, they found that the methods worked favorably for the sample size $n \geq 1000$. Their methods are based on the approximation of the probability of the weighted sum of the chi-squared random variables and, consequently, the null distribution depends on the data and involves computational complexity. For example, for $n = 100$ and $m = 24$, the empirical relative rejection frequencies are far from the asymptotic significance level. Therefore, their proposed method is still constrained in terms of sample size.

Suppose that the data generating process (DGP) is a stationary and nonlinear process, where the aim of modeling is to forecast the data. If we have rich sample size, we are able to select more complicated parametric models by using an appropriate information criterion, which may improve the results of forecasting. If not, we would look for a simplified linear model such as the AR model with weak white noise. However, there is no appropriate method because of the small sample size. In this sense we have to develop a new method for a goodness-of-fit test when subject to weak white noise.

If we look at FRZ and FR from another standpoint, their methods are extensions of Ljung (1986) from the strong white noise case to the weak white noise case. Even if the noise is strong, it is known that the chi-squared approximation of the portmanteau statistic performs poorly when the degrees of freedom (DF) are small (equivalently, when the number of m is small). To solve this problem, Ljung (1986) proposed modifications of the portmanteau test, which compute critical regions from the weighted sum of the chi-squared random variables as an approximated random variable of the portmanteau statistic.

Alternatively, Katayama (2008) proposed modifications to the test. He proposed a modified portmanteau statistic with a correction term to conduct a chi-squared test with a small number of DF. Based on this modification, Katayama (2009) pointed out the advantage of multiple portmanteau tests. However, the portmanteau statistic by Katayama (2008, 2009) focused on the univariate autoregressive moving average (ARMA) model with strong white noise.

For this reason, in this paper, we extend the portmanteau statistic by Katayama (2008) from the strong white noise case to the weak white noise case in Section 2. We propose a new portmanteau statistic that is asymptotically chi-squared even in the presence of non-independent innovations. We also study the joint probability of the multiple portmanteau statistics with different DF. These distributions are easy to compute when all the DF are even integers.

For ease of understanding, we have adopted FR's notations and define O as an appropriate dimensional zero matrix. The mathematical proof is given in the Appendix.

2. MAIN THEORETICAL RESULTS

2.1 Chi-squared approximation of the portmanteau statistic

Suppose that the d-dimensional time series (X_t) is generated by a vector AR(p) model:

$$X_t = \sum_{i=1}^{p} A_i X_{t-i} + \epsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{1}$$

where (ϵ_t) is strictly stationary and ergodic; $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{Var}(\epsilon_t) = \Omega$ and Ω is positive definite; and

$$\det A(z) \neq 0 \quad \text{for all } |z| \leq 1, \text{ where } A(z) = I_d - \sum_{i=1}^{p} A_i z^i.$$

As in FR, if (ϵ_t) is i.i.d. (strong white noise), we call (X_t) a strong AR(p) model and if (ϵ_t) is uncorrelated (weak white noise), we call (X_t) a weak AR(p) model. In addition, to obtain the asymptotic normality of the least squares estimator of $\theta_0 = \mathrm{vec}[A_1 \ldots A_p]$, we assume the following conditions:

$$\sum_{i=0}^{\infty} \{\alpha_X(i)\}^{\nu/(2+\nu)} < \infty \quad \text{and} \quad \|X_t\|_{4+2\nu} < \infty$$

for some $\nu > 0$, where $\|\ \cdot\ \|$ denotes the Euclidian norm, $\|X\|_r = \mathrm{E}(\|X\|^r)^{1/r}$, and

$$\alpha_X(i) = \sup_{A \in \sigma(X_u, u \leq t), B \in \sigma(X_u, u \geq t+i)} |\Pr(A \cap B) - \Pr(A)\Pr(B)|.$$

Given the observations $X_{1-p}, \ldots, X_n$, the least squares estimator of θ_0, $\widehat{\theta}_n$, is obtained and its asymptotic distribution is given in Proposition 2 in FR.

Let $(\widehat{\epsilon}_t)$ be a residual sequence using $\widehat{\theta}_n$. If we fit the model (1) appropriately, $(\widehat{\epsilon}_t)$ should behave like (ϵ_t) as n increases. Therefore, as n increases, the residual autocovariances,

$$\widehat{\gamma}_m = \left[\{\mathrm{vec}\,\widehat{\Gamma}_\epsilon(1)\}', \ldots, \{\mathrm{vec}\,\widehat{\Gamma}_\epsilon(m)\}'\right]' \text{ where } \widehat{\Gamma}_\epsilon(j) = \frac{1}{n}\sum_{t=j+1}^{n} \widehat{\epsilon}_t \widehat{\epsilon}_{t-j}',$$

would behave like

$$c_m = [\{\mathrm{vec}\,C_1\}', \ldots, \{\mathrm{vec}\,C_m\}']' \text{ where } C_j = \frac{1}{n}\sum_{t=j+1}^{n} \epsilon_t \epsilon_{t-j}'.$$

However, testing for many $\widehat{\gamma}_m$s are cumbersome because we cannot compute the joint probability of the $\widehat{\gamma}_m$s. For this, the portmanteau statistic is commonly used which is the weighted sum of squares of $(\widehat{\gamma}_m)$ or the residual autocorrelations. FR showed that

$$\widehat{\gamma}_m = c_m + \Phi_m(\widehat{\theta}_n - \theta_0) + o_p(n^{-1/2}), \tag{2}$$

where Φ_m is $d^2m \times d^2p$ matrix is defined by:

$$\Phi_m = -\,\mathrm{E}\left[\epsilon_{t-1}', \ldots, \epsilon_{t-m}'\right]' \otimes \left[X_{t-1}', \ldots, X_{t-p}'\right] \otimes I_d,$$

and the asymptotic distribution of $\sqrt{n}\widehat{\gamma}_m$ is obtained from the joint asymptotic distribution of $\sqrt{n}(c_m', (\widehat{\theta}_n - \theta_0)')$. From FR (Theorem 1),

$$\sqrt{n}\begin{bmatrix} c_m \\ \widehat{\theta}_n - \theta_0 \end{bmatrix} \xrightarrow{d} N(O, \Xi), \quad \text{where} \quad \Xi = \begin{bmatrix} \Sigma_m & \Sigma_{c_m,\widehat{\theta}_n} \\ \Sigma_{c_m,\widehat{\theta}_n}' & \Sigma_{\widehat{\theta}_n} \end{bmatrix}; \tag{3}$$

$$\sqrt{n}\widehat{\gamma}_m \xrightarrow{d} N(O, \Sigma_{\widehat{\gamma}_m}), \quad \text{where} \quad \Sigma_{\widehat{\gamma}_m} = \Sigma_m + \Phi_m \Sigma_{\widehat{\theta}_n} \Phi_m' + \Sigma_{c_m,\widehat{\theta}_n}\Phi_m' + \Phi_m \Sigma_{c_m,\widehat{\theta}_n}'. \tag{4}$$

(FR defined Σ_m as Σ_{c_m}, however we write Σ_m simply to avoid the double index). The asymptotic variance matrix of $\sqrt{n}c_m$, $\Sigma_m = [\Sigma_m(i,j)]_{i,j=1,\ldots,m}$, is given by:

$$\Sigma_m = \sum_{h=-\infty}^{\infty} \mathrm{E}(w_t w_{t-h}') \quad \text{and} \quad \Sigma_m(i,j) = \sum_{h=-\infty}^{\infty} E(\epsilon_{t-i} \otimes \epsilon_t)(\epsilon_{t-h-j} \otimes \epsilon_{t-h})',$$

where $w_t = (\epsilon_{t-1}', \ldots, \epsilon_{t-m}')' \otimes \epsilon_t$. Thus we have the FR (Theorem 2) versions of the portmanteau statistics by Chitturi (1974) and Hosking (1980):

$$Q_s(m) = n\sum_{j=1}^{m} \mathrm{tr}\left\{\widehat{\Gamma}_\epsilon(j)'\widehat{\Gamma}_\epsilon^{-1}(0)\widehat{\Gamma}_\epsilon(j)'\widehat{\Gamma}_\epsilon^{-1}(0)\right\} \tag{5}$$

$$\text{and } Q_s^*(m) = n^2\sum_{j=1}^{m} \frac{1}{n-j}\,\mathrm{tr}\left\{\widehat{\Gamma}_\epsilon(j)'\widehat{\Gamma}_\epsilon^{-1}(0)\widehat{\Gamma}_\epsilon(j)'\widehat{\Gamma}_\epsilon^{-1}(0)\right\} \tag{6}$$

which converge to a weighted sum of the chi-square variables under weak white noise, where the weights are the eigenvalues of $\Sigma_{\widehat{\gamma}_m}$. The statistic $Q_s^*(m)$ is a modified statistic of $Q_s(m)$ for a small sample distribution. On the other hand when (ϵ_t) is strong, these statistics are approximate to a $\chi^2_{d^2(m-p)}$ distribution. Therefore, we are unable to conduct a chi-square test by $Q_s(m)$ and $Q_s^*(m)$ without the condition of strong white noise.

Our aim is to produce an asymptotically chi-squared portmanteau statistic under not only strong white noise but also weak white noise. The idea is similar to (10) in Katayama (2008). Put $p < m < n$ and

$$D_m = \Phi_m(\Phi_m'\Sigma_m^{-1}\Phi_m)^{-1}\Phi_m'\Sigma_m^{-1}. \tag{7}$$

Multiplying both sides of (2) by $\sqrt{n}(I_{d^2m} - D_m)$, and using $(I_{d^2m} - D_m)\Phi_m = 0$ yields:

$$\begin{aligned}(I_{d^2m} - D_m)\sqrt{n}\widehat{\gamma}_m &= (I_{d^2m} - D_m)\sqrt{n}c_m + o_p(1)\\ &\xrightarrow{d} N(0,\ (I_{d^2m} - D_m)\Sigma_m),\end{aligned} \tag{8}$$

as $n \to \infty$, where the convergence follows from FR (p.468). By the Choleski decomposition, there exists a positive definite matrix, $\Sigma_m^{1/2}$, such that $\Sigma_m = \Sigma_m^{1/2}(\Sigma_m^{1/2})'$. Since $(I_{d^2m} - D_m)\Sigma_m = \Sigma_m - \Phi_m(\Phi_m'\Sigma_m^{-1}\Phi_m)^{-1}\Phi_m'$, it follows from (8) that:

$$\Sigma_m^{-1/2}(I_{d^2m} - D_m)\sqrt{n}\widehat{\gamma}_m \xrightarrow{d} N(0,\ I_{d^2m} - D_m^*), \tag{9}$$

as $n \to \infty$, where $D_m^* = F_m(F_m'F_m)^{-1}F_m'$ and $F_m = \Sigma_m^{-1/2}\Phi_m$. It follows that:

$$\begin{aligned}n\widehat{\gamma}_m'(I_{d^2m} - D_m')\Sigma_m^{-1}(I_{d^2m} - D_m)\widehat{\gamma}_m &= n\widehat{\gamma}_m'\Sigma_m^{-1}\widehat{\gamma}_m - B_w(m)\\ &\xrightarrow{d} \chi^2_{d^2(m-p)},\end{aligned} \tag{10}$$

as $n \to \infty$, where $B_w(m) = n\widehat{\gamma}_m'D_m'\Sigma_m D_m\widehat{\gamma}_m = n\widehat{\gamma}_m'(\Sigma_m^{-1/2})'D_m^*\Sigma_m^{-1/2}\widehat{\gamma}_m$ and the convergence follows from the fact that the $\operatorname{rank}\Phi_m = d^2p$ (see Appendix). FR pointed out there exist consistent estimators of Φ_m and Σ_m. Consequently, we propose a new portmanteau statistic:

$$\begin{aligned}Q_w(m) &= n\widehat{\gamma}_m'(I_{d^2m} - \widehat{D}_m')\widehat{\Sigma}_m^{-1}(I_{d^2m} - \widehat{D}_m)\widehat{\gamma}_m\\ &= n\widehat{\gamma}_m'\widehat{\Sigma}_m^{-1}\widehat{\gamma}_m - \widehat{B}_w(m),\end{aligned} \tag{11}$$

where $\widehat{B}_w(m) = n\widehat{\gamma}_m'\widehat{D}_m\widehat{\Sigma}_m\widehat{D}_m\widehat{\gamma}_m$ and $\widehat{D}_m$ and $\widehat{\Sigma}_m$ are consistent estimators of D_m and Σ_m.

Theorem 1 *Under the model* (1) *with weak white noise, for* $p < m < n$, $Q_w(m) \xrightarrow{d} \chi^2_{d^2(m-p)}$, *as* $n \to \infty$.

Note that the chi-squared approximation of the statistic $Q_w(m)$ is not required for large m, but only for $m > p$.

In addition to the chi-squared approximation, there are a few advantages to using $Q_w(m)$. First, our method reduces computational complexity (See steps 4–7 in FR (Section 4)). FR's method requires fitting the $(m+p)d^2$-vector AR(p) models to the estimate Ξ. While our method requires fitting, at most, the md^2-vector AR(p) models to the estimate Σ_m. Furthermore, reading FRZ and FR (including the longer versions) and other related articles, it seems that almost all examples of weak white noise show that Σ_m may be approximated by a block-diagonal (or tri-diagonal) matrix. In this case, the estimation of Σ_m does not require the estimation of the covariance structure of a high dimension multivariate process. It enables us to conduct a portmanteau test with large degree of freedom even if the sample size is small (e.g., $n = 100$ and $m = 20$) which is examined in Section 4. Finally, if we assume that Σ_m is block-diagonal, we are able to conduct multiple portmanteau tests which are examined in Sections 2.3 and 4.2.

2.2 Univariate case

FRZ considered the univariate case $d = 1$. In their paper, $\Gamma_{m,m} = \Sigma_m$, $\Gamma(i,j) = \Sigma_m(i,j)$ and $\Lambda_m = \Phi'_m\Omega^{-1}$. Therefore, for $d = 1$,

$$D_m = \Lambda'_m(\Lambda_m\Gamma^{-1}_{m,m}\Lambda'_m)^{-1}\Lambda_m\Gamma^{-1}_{m,m}.$$

From FRZ (Theorem 2),

$$\begin{aligned}\Sigma_{\widehat{\gamma}_m} =& \Gamma_{m,m} + \Lambda'_m(\Lambda_\infty\Lambda'_\infty)^{-1}\Lambda_\infty\Gamma_{\infty,\infty}\Lambda'_\infty(\Lambda_\infty\Lambda'_\infty)^{-1}\Lambda_m \\ &- \Gamma_{m,\infty}\Lambda'_\infty(\Lambda_\infty\Lambda'_\infty)^{-1}\Lambda_m - \Lambda'_m(\Lambda_\infty\Lambda'_\infty)^{-1}\Lambda_\infty\Gamma_{\infty,m},\end{aligned}$$

where $\Lambda_\infty\Lambda'_\infty = \lim_{m\to\infty}\Lambda_m\Lambda'_m$, $\Lambda_\infty\Gamma_{\infty,\infty}\Lambda'_\infty = \lim_{m,n\to\infty}\Lambda_m\Gamma_{m,n}\Lambda'_n$, $\Gamma_{m,\infty}\Lambda'_\infty = \lim_{n\to\infty}\Gamma_{m,n}\Lambda'_n$, and $\Lambda_\infty\Gamma_{\infty,m} = \lim_{n\to\infty}\Lambda_n\Gamma_{n,m}$.

We now consider the asymptotic variance of $\sqrt{n}\Phi'_m{\Sigma_m}^{-1}\widehat{\gamma}_m$, which is a part of $B_w(m)$. For $d = 1$, the asymptotic variance of $\mathrm{Var}(\sqrt{n}\Phi'_m{\Sigma_m}^{-1}\widehat{\gamma}_m)$ is $\mathcal{I}_m\mathcal{J}^{-1}\mathcal{I}_m\mathcal{J}^{-1}\mathcal{I}_m - \mathcal{I}_m$, as $m \to \infty$, where $\mathcal{I}_m = \Lambda_m\Gamma^{-1}_{m,m}\Lambda'_m$ and $\mathcal{J} = \Lambda_\infty\Lambda'_\infty$. As noted by FRZ and Katayama (2008), for the case of a univariate strong AR model, $\lim_{m\to\infty}\mathcal{I}_m = \mathcal{J}$ and this variance vanishes as m increases. However, for the case of the weak AR model, this equation does not hold in general.

2.3. Multiple portmanteau tests

In practical analysis, we conduct multiple portmanteau tests in which we vary m because we are unable to find an optimal number for m. Katayama (2009) proposed multiple portmanteau tests for the univariate ARMA models with strong white noise errors.

Let $p < m = m(1) < m(2) < \cdots < m(s) = M < n$. Then, from (8) and (9), we obtain:

$$\begin{bmatrix} \Sigma_{m(i)}^{-1/2}(I_{d^2m(i)} - D_{m(i)})\sqrt{n}\widehat{\gamma}_m \\ O \end{bmatrix} = \sqrt{n}E_{m(i)}c_M + o_p(1), \tag{12}$$

for $i = 1, 2, \ldots, s$, where $E_{m(i)}$ is a $d^2M \times d^2M$ matrix defined by:

$$E_{m(i)} = \begin{bmatrix} \Sigma_{m(i)}^{-1/2}(I_{d^2m(i)} - D_{m(i)}) & O \\ O & O \end{bmatrix},$$

which yields the following theorem:

Theorem 2 *Under the model* (1) *with weak white noise, it holds that, as* $\to \infty$*:*

$$[Q_w(m(1)), \ldots, Q_w(m(s))] \xrightarrow{d} \left[(Z_M^*)'E_{m(1)}^*Z_M^*, \ldots, (Z_M^*)'E_{m(s)}^*Z_M^*\right] \tag{13}$$

where Z_M^* *is a* Md^2*-dimensional random variable such that* $Z_M^* \sim N(0, \Sigma_M)$ *and*

$$E_{m(i)}^* = (E_{m(i)}^*)'E_{m(i)}^* = \begin{bmatrix} (\Sigma_{m(i)}^{-1/2})'(I_{d^2m(i)} - D_{m(i)}^*)\Sigma_{m(i)}^{-1/2} & O \\ O & O \end{bmatrix}.$$

The convergence of (13) follows from $\sqrt{n}c_M \xrightarrow{d} Z_M^*$ and the Cramér-Wold device.

In general, computation of the probability of the asymptotic joint distribution of (13) is difficult because all the elements are correlated. However, if Σ_M is a block-diagonal matrix, we obtain the following corollary:

Corollary 1 *Under assumptions in Theorem 2, suppose that* Σ_M *is a block-diagonal such that*

$$\Sigma_M = \text{diag}[\Sigma_M(1,1), \ldots, \Sigma_M(M,M)], \tag{14}$$

where $\Sigma(i,i)$, $i = 1, \ldots, M$, *is a* $d^2 \times d^2$ *symmetric matrix. Then, for* $j = 1, \ldots, s$, *as* $m \to \infty$,

$$(Z_M^*)'E_{m(j)}^*Z_M^* = \sum_{i=d^2p+1}^{d^2m(j)} Z_i^2 + O_p(\lambda^{m+1-p}), \tag{15}$$

where $\lambda \in (0,1)$ *and* (Z_i) *is i.i.d.* $N(0,1)$.

In general, (15) does not hold without (14) because Lemma 2 in the Appendix does not hold. The assumption of Corollary 1 is stronger than that of Theorem 2. However, there are some nonlinear models satisfying this assumption, including the m-dependent process by Romano and Thombs (1996, Example 2.1) and the generalized autoregressive conditionally heteroskedastic (GARCH) process, (see the longer versions of FRZ and FR).

From this corollary, if we have strong evidence that Σ_M is block-diagonal, since we are able to conduct multiple portmanteau tests easily as in Katayama (2009) when all the DF are even integers. The finite sample performance of the multiple tests are examined in Section 0.4.

3. Note on the strong white noise case

The following remarks focus on the case where (ϵ_t) is strong white noise.

3.1 Relationship between $Q_w(m)$ and $Q_s(m)$

The idea of the chi-squared approximation of $Q_w(m)$ is similar to that of Hosking (1980, Section 4). Hosking (1980) considered the portmanteau statistics for the multivariate ARMA models subject to errors from strong white noise. When the (ϵ_t) is strong, $\Sigma_m = I_m \otimes \Omega \otimes \Omega$ and $n\widehat{\gamma}'_m \Sigma_m^{-1} \widehat{\gamma}_m$ in (10) becomes:

$$\begin{aligned} n\widehat{\gamma}'_m \Sigma_m^{-1} \widehat{\gamma}_m &= n\widehat{\gamma}'_m (I_m \otimes \Omega^{-1} \otimes \Omega^{-1}) \widehat{\gamma}_m \\ &= n \sum_{j=1}^{m} \{\mathrm{vec}\, \widehat{\Gamma}_\epsilon(j)\}' (\Omega^{-1} \otimes \Omega^{-1}) \{\mathrm{vec}\, \widehat{\Gamma}_\epsilon(j)\} \\ &= n \sum_{j=1}^{m} \mathrm{tr}\left(\widehat{\Gamma}_\epsilon(j)' \Omega^{-1} \widehat{\Gamma}_\epsilon(j)' \Omega^{-1} \right), \end{aligned}$$

which is approximately equal to the portmanteau statistics $Q_s(m)$ and $Q_s^*(m)$. We note that Hosking (1980, Appendix) makes additional assumptions, which are extensions of Box and Pierce's (1970) assumptions for the multivariate case. Accordingly Hosking (1980, Section 4) ignores the term corresponding to $B_w(m)$ in (10) and restricts $Q_s(m)$ and $Q_s^*(m)$ to large m for the chi-squared approximation. However, Ljung (1986) and Katayama (2008) pointed out that their assumptions are inappropriate both theoretically and empirically because the convergence depends on not only the asymptotic behavior of m but also the true value of the AR parameter, θ_0.

3.2 A modified portmanteau statistic under strong white noise

If we have firm evidence that (ϵ_t) is strong white noise, we have the alternative portmanteau statistics with a correction term. From $D_m'\Sigma_m D_m = \Sigma_m^{-1}\Phi_m(\Phi_m'\Sigma_m^{-1}\Phi_m)^{-1}\Phi_m'\Sigma_m^{-1}$ and $\Sigma_m = I_m \otimes \Omega \otimes \Omega$, we have

$$Q_s^{**}(m) = Q_s^*(m) - \widehat{B}_s^*(m), \tag{16}$$

where

$$\begin{aligned}
\widehat{B}_s^*(m) &= (\widehat{\gamma}_m^*)'(\widehat{\Sigma}_m^s)^{-1}\widehat{\Phi}_m\left\{\widehat{\Phi}_m'(\widehat{\Sigma}_m^s)^{-1}\widehat{\Phi}_m\right\}^{-1}\widehat{\Phi}_m'(\widehat{\Sigma}_m^s)^{-1}\widehat{\gamma}_m^*,\\
\widehat{\gamma}_m^* &= \left[\frac{n}{\sqrt{n-1}}\{\mathrm{vec}\,\widehat{\Gamma}_\epsilon(1)\}',\ldots,\frac{n}{\sqrt{n-m}}\{\mathrm{vec}\,\widehat{\Gamma}_\epsilon(m)\}'\right]'\\
(\widehat{\Sigma}_m^s)^{-1} &= I_m \otimes \widehat{\Gamma}_\epsilon^{-1}(0)\otimes \widehat{\Gamma}_\epsilon^{-1}(0),
\end{aligned}$$

Unlike $Q_s(m)$ and $Q_s^*(m)$, the chi-squared approximation of the statistic $Q_s^{**}(m)$ is not required for large m, but only for $m > p$. This statistic is an extension of the modified portmanteau statistic by Katayama (2008) to the multivariate case.

As in (14) in Katayama (2008), $\widehat{B}_s^*(m) \to 0$ as m increases (see Appendix).

3.3 Independence property of the portmanteau statistics

We determine an independence property under strong white noise. Let $\widehat{\gamma}(i) = \mathrm{vec}\,\widehat{\Gamma}_\epsilon(i)$, $i = 1,\ldots,M$ and $p < m \le M < n$. From Hosking (1980) and FR, the asymptotic variance of $\widehat{\gamma}_M$ under strong white noise is

$$\Sigma_{\widehat{\gamma}_M} = I_M \otimes \Omega \otimes \Omega - \Phi_M \Sigma_{\widehat{\theta}_n}\Phi_M'. \tag{17}$$

By Lemma 1, as $m \to \infty$, (17) becomes:

$$\Sigma_{\widehat{\gamma}_M} = \mathrm{diag}\,[\Omega\otimes\Omega,\ldots,\Omega\otimes\Omega] - \begin{bmatrix}\Phi_m\Sigma_{\widehat{\theta}_n}\Phi_m' & O\\ O & O\end{bmatrix} + O(\lambda^{m+1-p}), \tag{18}$$

which approximately indicates the independence of $\widehat{\gamma}_m, \widehat{\gamma}(m+1),\ldots,\widehat{\gamma}(M)$. In other words, $Q_s^{**}(m)$ and $Q_s^{**}(M) - Q_s^{**}(m)$ are approximately independent and $Q_s^{**}(M) - Q_s^{**}(m) \sim \chi^2_{d^2(M-m)}$ if n, m is large. As an alternative to Corollary 1, it establishes multiple portmanteau tests under strong white noise. Similar results can be applied to $Q_s(m)$ and $Q_s^*(m)$ because $\widehat{B}_s^*(m) \to 0$ as m increases.

For the univariate case, (18) provides multiple tests using individual residual autocorrelations. We reject the fitted model if $Q_m^{**} > c_m$ or $\{a_{ni}^{-1}\widehat{\gamma}(i)/\widehat{\gamma}(0)\}^2 > c$, $\exists i = m+1, \ldots, M$, where a_{ni} are the asymptotic standard errors of $\widehat{\gamma}(i)/\widehat{\gamma}(0)$ and c and c_m are given by

$$\Pr(\chi^2_{m-p-q} \le c_m) = (1-\beta_m) \quad \text{and} \quad \Pr(\chi^2_1 \le c) = 1-\beta_1.$$

where $\beta_m, \beta_1 \in (0,1)$. Since Q_m^{**} and $\widehat{\gamma}(i)$, $i > m$ are approximately independent, the overall significance level of this test is approximately given by

$$1 - \Pr\left(\chi^2_{m-p-q} \le c_m\right)\left\{\Pr\left(\chi^2_1 \le c\right)\right\}^{M-m} = 1-(1-\beta_m)(1-\beta_1)^{M-m}, \qquad (19)$$

when m and n are large. The approximate significance level of multiple tests by higher-order lag individual residual autocorrelations is $1-(1-\beta)^{M-m}$. This probability is approximately equal to: For $M-m = 10, 20, 30$, $1-0.95^{M-m} = 0.40, 0.64, 0.79$ and $1-0.99^{M-m} = 0.10, 0.18, 0.26$. These results give us an interpretation of the use of model diagnostics as follows: In standard model diagnostics using statistical software, the residual autocorrelation functions, $\widehat{\gamma}(i)$s, are plotted with 95% confidence intervals. However, a few higher-order lag $\widehat{\gamma}(i)$s are sometimes out of the 95% confidence limits even if the residual process behaves as the white noise process. However, this phenomenon is natural as this probability is more than 40% when $M-m \ge 10$. Except for the lower order residual autocorrelations, although the plot is a useful tool to check the regularity of autocorrelations, we should not pay attention to the 95% confidence intervals but to the 99% confidence intervals from the standpoint of multiple testing.

4. Some simulation studies

This section examines the finite sample properties of the proposed portmanteau statistics by simulation studies.

4.1 Estimation of Σ_m

FRZ and FR pointed out there are two main approaches for estimating the matrix Ξ in (3). One is the model dependent method which assumes a model for the noise (e.g., i.i.d., GARCH, Markov-switching), and a plug-in approach which leads to a consistent estimator. However, if we have strong evidence about the nonlinear structure and a large enough sample size, we should be able build this information into the model. The second approach is a fully nonparametric one in which no specific time series model is assumed for the noise. For the second approach, they provided two proposals. One is to use a heteroskedasticity and autocorrelation consistent (HAC) estimator (see Andrews, 1991, and references therein). The other approach is to obtain a consistent estimator by fitting the noise distribution to an auxiliary AR model of large order (see, for example, Section 4 in FR).

As argued in Section 2.1, we only estimate Σ_m in Ξ, which is the asymptotic variance of $\sqrt{n}c_m$ and may be approximated by block tri-diagonal matrix. We will now suppose that the matrix is (possibly approximately) a block tri-diagonal: there exists $\tau = 1, \ldots, m-2$ such that $\Sigma_m(i,j) = O$ for $|i-j| > \tau$. Here we discuss how to estimate this matrix.

One reasonable way to compute a consistent estimator of Σ_m follows from FR (Section 4). Assume $\widehat{\Sigma}_m(i,j)$, $i,j = 1, \ldots, m$ is a $d^2 \times d^2$ matrix, $\widehat{\Sigma}_m = [\widehat{\Sigma}_m(i,j)]_{i,j=1,\ldots,m}$, and

$$\widehat{w}_{i,t} = (\widehat{\epsilon}'_{t-i}, \ldots, \widehat{\epsilon}'_{t-i-\tau+1})' \otimes \widehat{\epsilon}_t, \quad \text{for} \quad i = 1, \ldots, m-\tau+1, \ t = 1, \ldots, n.$$

step1. Put $\widehat{\Sigma}_m(i,j) = O$ for all i,j and $k = m-\tau+1$.

step2. Using the Durbin-Levinson algorithm fit for the AR(r_k) model:

$$\widehat{w}_{k,t} - \sum_{j=1}^{r_k} \widehat{\mathcal{A}}_{k,j}\widehat{w}_{k,t-j} = \widehat{u}_{k,t}.$$

where r_k is fixed or determined by BIC information criteria.

step3. Update $\widehat{\Sigma}_m = [\widehat{\Sigma}_m(i,j)]$ by substituting a matrix:

$$[\widehat{\Sigma}_m(i,j)]_{i,j=k,\ldots,k+h-1} = \left(I_{d^2h} - \sum_{j=1}^{r_k} \widehat{\mathcal{A}}_{k,j}\right)^{-1} \widehat{\Sigma}_{k,u} \left(I_{d^2h} - \sum_{j=1}^{r_k} \widehat{\mathcal{A}}'_{k,j}\right)^{-1},$$

where $\widehat{\Sigma}_{k,u} = \frac{1}{n}\sum_{t=1}^{n} \widehat{u}_{k,t}\widehat{u}'_{k,t}$.

step4. Repeat steps 2 and 3 for $k = m-\tau, m-\tau-1, \ldots, 1$.

In the rest of this section we examine the empirical size and power of the various portmanteau statistics presented above. Here we also examine finite sample behavior for different values of τ.

4.2 Empirical significance level

We examine the following DGPs with $n = 100, 1000$ with 1000 replications:

DGP 1: $X_t = A(0.95, 0)X_{t-1} + \epsilon_t^L(0)$; DGP 2: $X_t = A(0.6, 0.3)X_{t-1} + \epsilon_t^L(0.5)$;

DGP 3: $X_t = A(0.6, 0)X_{t-1} + \epsilon_t^{NL}(0)$; DGP 4: $X_t = A(0.95, 0)X_{t-1} + \epsilon_t^{NL}(0)$;

DGP 5: $X_t = A(0.6, 0.3)X_{t-1} + \epsilon_t^{NL}(0.5)$; DGP 6: $X_t = A(0.95, 0)X_{t-1} + \epsilon_t^{NL}(0.5)$;

where

$$A(a,b) = \begin{bmatrix} a & b \\ b & a \end{bmatrix}, \quad \epsilon_t^L(\rho) = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \text{ i.i.d. } N\left(O, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \quad \epsilon_t^{NL}(\rho) = \begin{bmatrix} \epsilon_{1t}\epsilon_{1t-1}\epsilon_{1t-2} \\ \epsilon_{2t}\epsilon_{2t-1}\epsilon_{2t-2}. \end{bmatrix}.$$

FR showed that the Σ_ms are block-diagonals in each DGP and conducted simulation studies for DGPs 1 and 4. The studies of DGPs 1 and 4 are also conducted using FR (Section 5.1). To compute $Q_w(m)$, we used the method given in FR (Section 4) and Section 0.4.1 for $\widehat{\Sigma}_m$ with $\tau = 1$. In each model, we computed the statistics for $m = 2, 3, \ldots, 24$, $n = 100$ and for $m = 2, 3, \ldots, 48$, $n = 1000$.

The empirical significance level of $Q_s^*(m), Q_s^{**}(m)$ and $Q_w(m)$ corresponding to 5% for the AR(1) model are given in Figures 1 and 2. As discussed in FR, $Q_s^*(m)$ tends to exceed the significance level in each case. For the strong white noise case (DGPs 1 and 2), $Q_s^{**}(m)$ performs favorably, however it shows over-rejection behavior in the weak white noise case. Whereas $Q_w(m)$ performs well as a whole even for $n = 100$, $m > 10$, and the nearly non-stationary case.

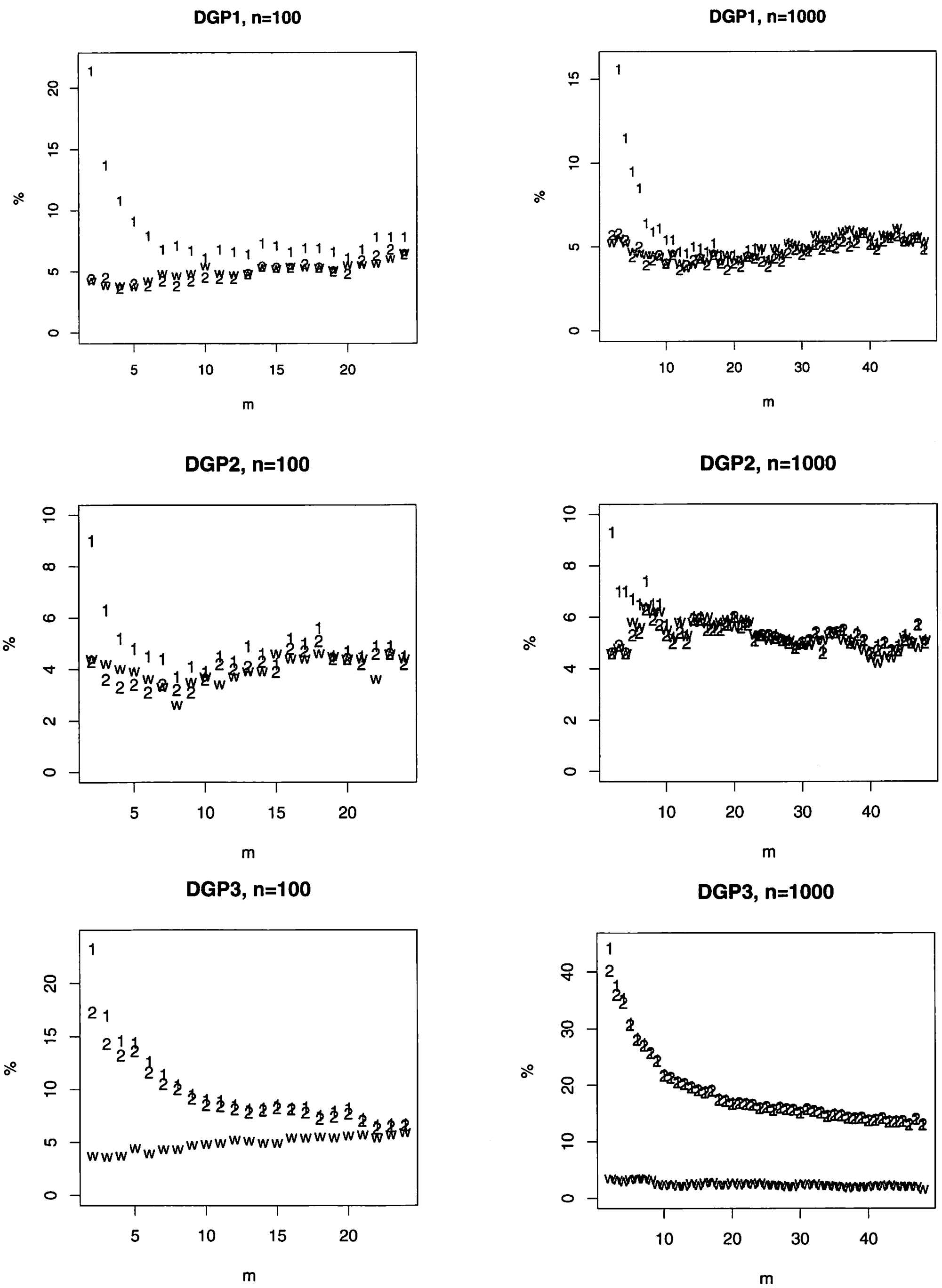

Figure 1: Percentages of the empirical significance level of the statistics $Q_s^*(m) \sim \chi^2_{4(m-1)}$ (denoted by 1), $Q_s^{**}(m) \sim \chi^2_{4(m-1)}$ (denoted by 2) and $Q_w(m) \sim \chi^2_{4(m-1)}$ (denoted by w).

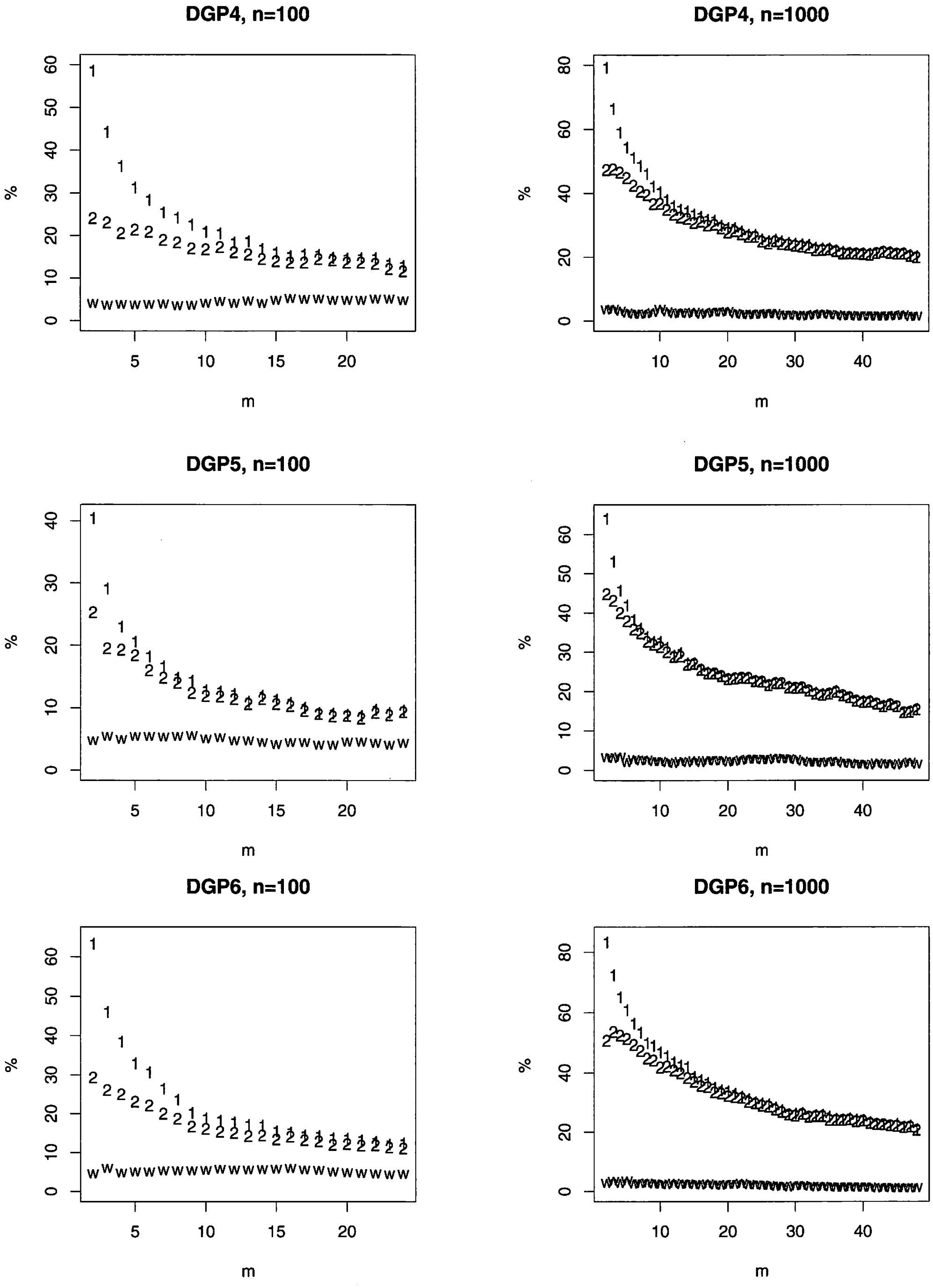

Figure 2: Percentages of the empirical significance level of the statistics $Q_s^*(m) \sim \chi^2_{4(m-1)}$ (denoted by 1), $Q_s^{**}(m) \sim \chi^2_{4(m-1)}$ (denoted by 2) and $Q_w(m) \sim \chi^2_{4(m-1)}$ (denoted by w).

Since these DGPs' Σ_ms are block-diagonal matrices, we are able to conduct the multiple portmanteau tests discussed in Section 2.3. Table 1 shows the empirical significance levels of the multiple portmanteau tests, where the marginal significance levels are 5%, $m = 2, 3, \ldots, 24$ for $n = 100$, and $m = 2, 3, \ldots, 48$ for $n = 1000$. From Katayama (2009), the approximate significance levels of these multiple portmanteau tests are 20.8% and 25.0%, respectively. Table 1: reveals that the empirical joint significance levels perform similarly to the asymptotic significance levels for the strong AR(1) models, while it reveals under-rejection behavior towards the asymptotic significance levels for the weak AR(1) models.

Table 1: Empirical significance levels of the multiple portmanteau tests given marginal significance level 5%, where asymptotic joint significance levels are 20.8% for $n = 100$ and 25.0% for $n = 1000$, respectively

DGP	1	2	3	4	5	6
$n = 100$	18.4	17.1	13.5	12.1	12.0	12.7
$n = 1000$	24.6	26.1	16.3	13.7	14.0	13.3

Figures 1 and 2 show the empirical size of $Q_s^*(m)$ and $Q_s^{**}(m)$ tend to be consistent as m increases, which implies that the correction term $\widehat{B}_s(m)$ in $Q_s^{**}(m)$ converges with zero as m increases. Based on these, we also examined $Q_w^1(m) = n\widehat{\gamma}_m'\widehat{\Sigma}_m^{-1}\widehat{\gamma}_m$ which is the first term of $Q_w(m)$ in (11). We guessed $\widehat{B}_w(m)$ in $Q_w(m)$ might converge to zero as m increases. However, we could not prove this theoretically from Section 0.2.2 even if we used Lemma 1. Therefore we examined the finite sample behavior of $\widehat{B}_w(m)$ as m increases. We also examined a modified version of the portmanteau statistics for the small sample approximation of $Q_w(m)$ corresponding to $Q_s^*(m)$:

$$Q_w^*(m) = \widehat{\gamma}_m'(T_n \otimes I_{d^2})(I_{d^2m} - \widehat{D}_m')\widehat{\Sigma}_m^{-1}(I_{d^2m} - \widehat{D}_m)(T_n \otimes I_{d^2})\widehat{\gamma}_m, \tag{20}$$

where $T_n = n\,\mathrm{diag}[(n-1)^{-1/2}, \ldots, (n-m)^{-1/2}]$. Figures 3 and 4 show the finite sample behavior of these statistics. It reveals that though $Q_w^1(m)$ approaches $Q_w(m)$ as m increases, this is not at an exponential rate. It implies that the correction term $\widehat{B}_w(m)$ does not converge to zero rapidly as m increases. In addition, $Q_w^*(m)$ shows over-rejection behavior as m increases, which implies the use of $Q_w(m)$ is more favorable than $Q_w^*(m)$ for conducting chi-squared tests with small sample sizes.

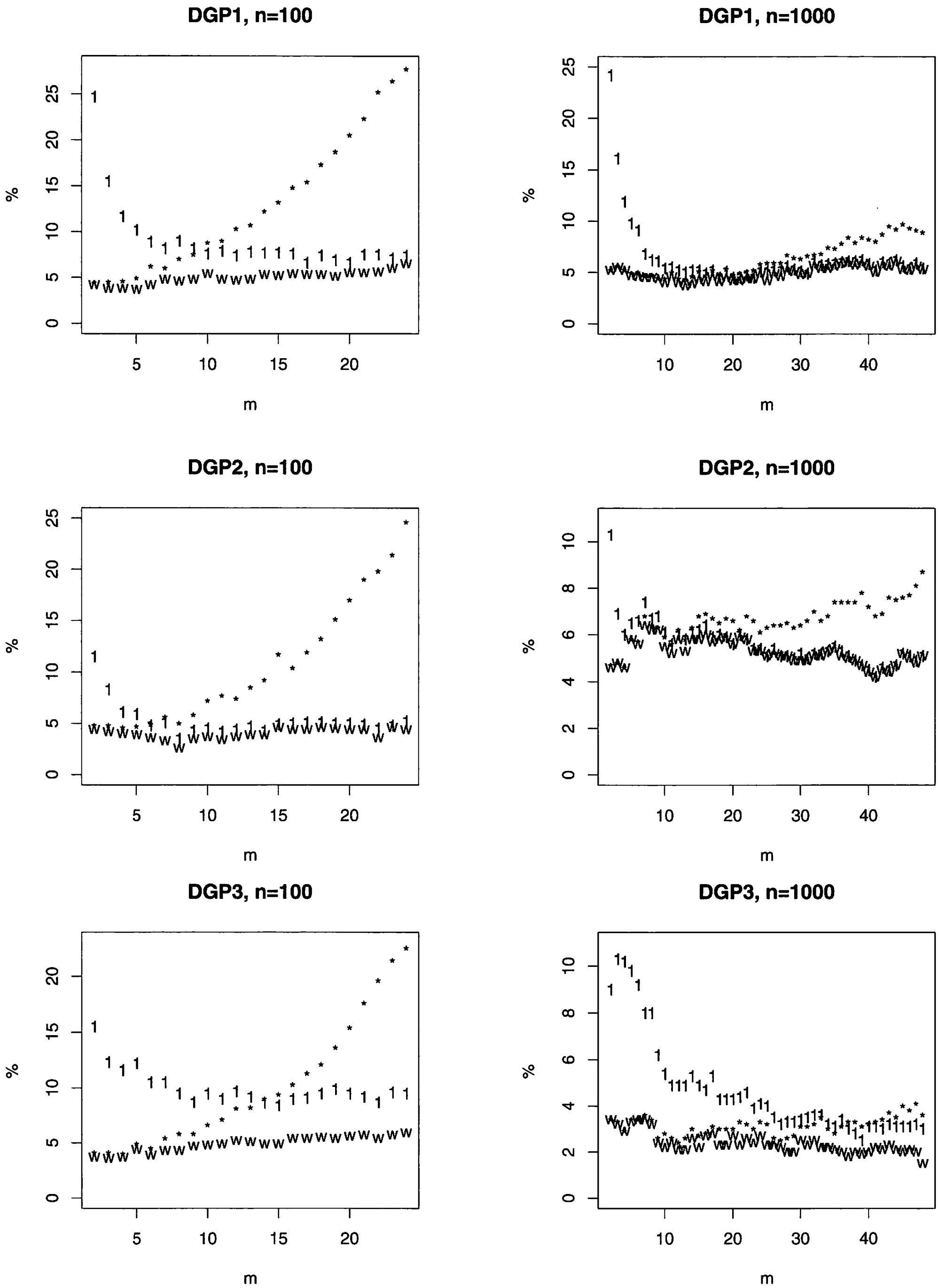

Figure 3: Percentages of the empirical significance level of $Q_w^1(m) \sim \chi^2_{4(m-1)}$ (denoted by 1) $Q_w(m) \sim \chi^2_{4(m-1)}$ (denoted by w) and $Q_w^*(m) \sim \chi^2_{4(m-1)}$ (denoted by *).

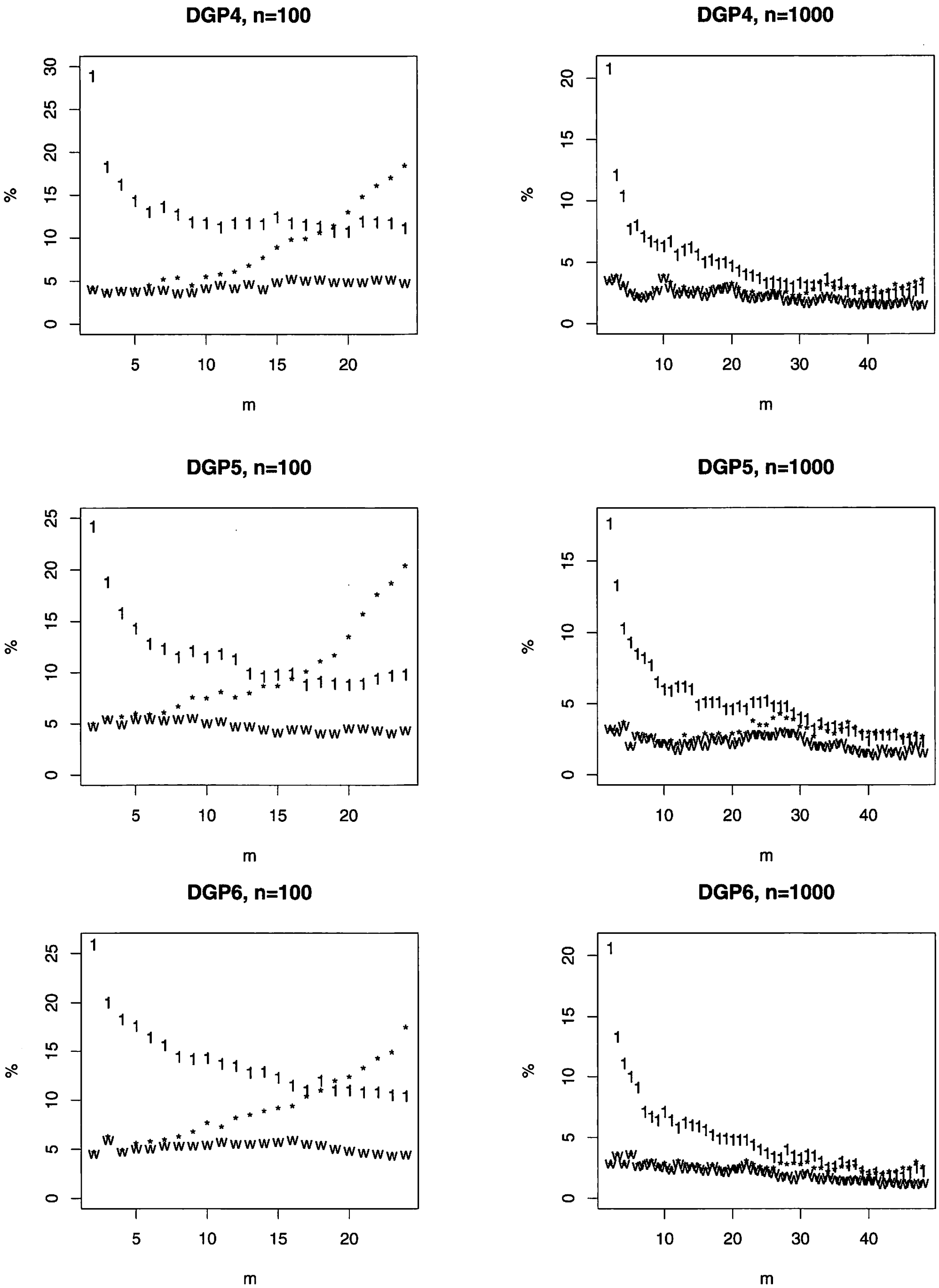

Figure 4: Percentages of the empirical significance level of $Q_w^1(m) \sim \chi^2_{4(m-1)}$ (denoted by 1) $Q_w(m) \sim \chi^2_{4(m-1)}$ (denoted by w) and $Q_w^*(m) \sim \chi^2_{4(m-1)}$ (denoted by *).

Figure 5: shows the finite sample behavior of $Q_w(m)$ with $\tau = 1, 2, 3$ and $n = 100$. Since Σ_m is a block-diagonal matrix, $\tau = 1$ is a desirable method for estimating Σ_m. These figures show the over-rejection behavior for the cases of $\tau = 2$ or $\tau = 3$, which indicates the selection of τ is sensitive to the consistency of Σ_m.

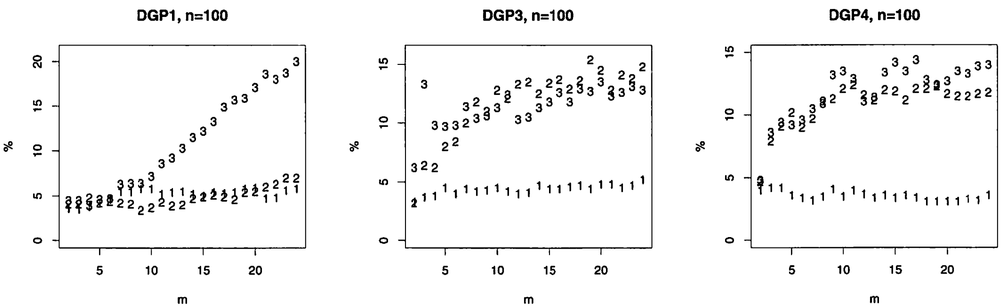

Figure 5: Percentages of the empirical significance level of $Q_w(m) \sim \chi^2_{4(m-1)}$, where the cases of $\tau = 1, 2, 3$ are denoted by 1, 2, 3.

4.3 Empirical power

We consider 1000 replications of size $n = 100, 1000$ of the bivariate AR(2) models defined by:

$$\text{DGP 7:} \quad X_t = A(0.2, 0.1)X_{t-1} + A(c, 0)X_{t-2} + \epsilon_t^L(0);$$
$$\text{DGP 8:} \quad X_t = A(0.2, 0.1)X_{t-1} + A(c, 0)X_{t-2} + \epsilon_t^{NL}(0);$$
$$\text{DGP 9:} \quad X_t = A(0.6, 0)X_{t-1} + A(c, 0)X_{t-2} + \epsilon_t^{NL}(0);$$

where $c = 0.0, 0.1, 0.2, 0.3$ and $A(a, b), \epsilon_t^L(0), \epsilon_t^{NL}(0)$ are defined similarly to DGPs 1–6 in Section 4.2. DGP 8 is also conducted by FR (Section 5.2) and DGP 9 with $c = 0$ corresponds to DGP3 in Section 4.2. The results are given in Figure 6 and Table 2. We fitted an AR(1) model and conducted the tests using $Q_w(m) \sim \chi^2_{4(m-1)}$ with a 5% significance level for $m = 2, 3, \ldots, 24$, $n = 100$ and for $m = 2, 3, \ldots, 48$, $n = 1000$ with $\tau = 1$. Figure 6 indicates the test using $Q_w(m)$ is more powerful as n, c increases and m decreases, which are natural results compared with the previous simulation studies of the portmanteau tests. The results using DGP 8, $n = 1000$ indicate the tests using $Q_w(m)$ possess power comparable to the portmanteau tests suggested by FR. Table 2 gives the results of the multiple portmanteau tests, where the tests are conducted similarly to Table 1. Table 2 indicates the multiple portmanteau tests possess sufficient power, as we expected.

As a whole, the tests by $Q_w(m)$ perform well even for the small number of n and are comparable to the tests suggested by FR. However, we note that these simulation studies are based on Σ_m as a block-diagonal and the problem of the choice of τ still remains.

Table 2: Empirical power of multiple portmanteau tests given a marginal significance level 5%, where the asymptotic joint significant levels are 20.8% for $n = 100$ and 25.0% for $n = 1000$

DGP 7 $c =$	0.0	0.1	0.2	0.3
$n = 100$	15.9	21.0	52.2	86.8
$n = 1000$	23.9	97.0	100.0	100.0
DGP 8 $c =$	0.0	0.1	0.2	0.3
$n = 100$	14.3	14.9	28.5	47.7
$n = 1000$	15.0	59.7	99.0	100.0
DGP 9 $c =$	0.0	0.1	0.2	0.3
$n = 100$	13.5	14.7	22.7	35.5
$n = 1000$	16.3	41.6	90.6	99.5

APPENDIX

Proof of $\operatorname{rank} \Phi_m = d^2p$: From (2.1.13) in Lütkepohl (2006), we have:

$$X_t = \sum_{i=0}^{\infty} A_i^* \epsilon_{t-i}, \tag{A.1}$$

where $A_i^* = J\widetilde{A}^i J'$, J is a $d \times dp$ matrix such that $J = [I_d \; O \ldots \; O]$, $\widetilde{A} = A_1$ for $p = 1$ and

$$\widetilde{A} = \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_d & O & \ldots & O & O \\ O & I_d & \ddots & \vdots & O \\ \vdots & \ddots & \ddots & O & \vdots \\ O & \ldots & O & I_d & O \end{bmatrix} \quad \text{for } p > 1.$$

It follows that

$$\Phi_m = -G_m \otimes I_d, \tag{A.2}$$

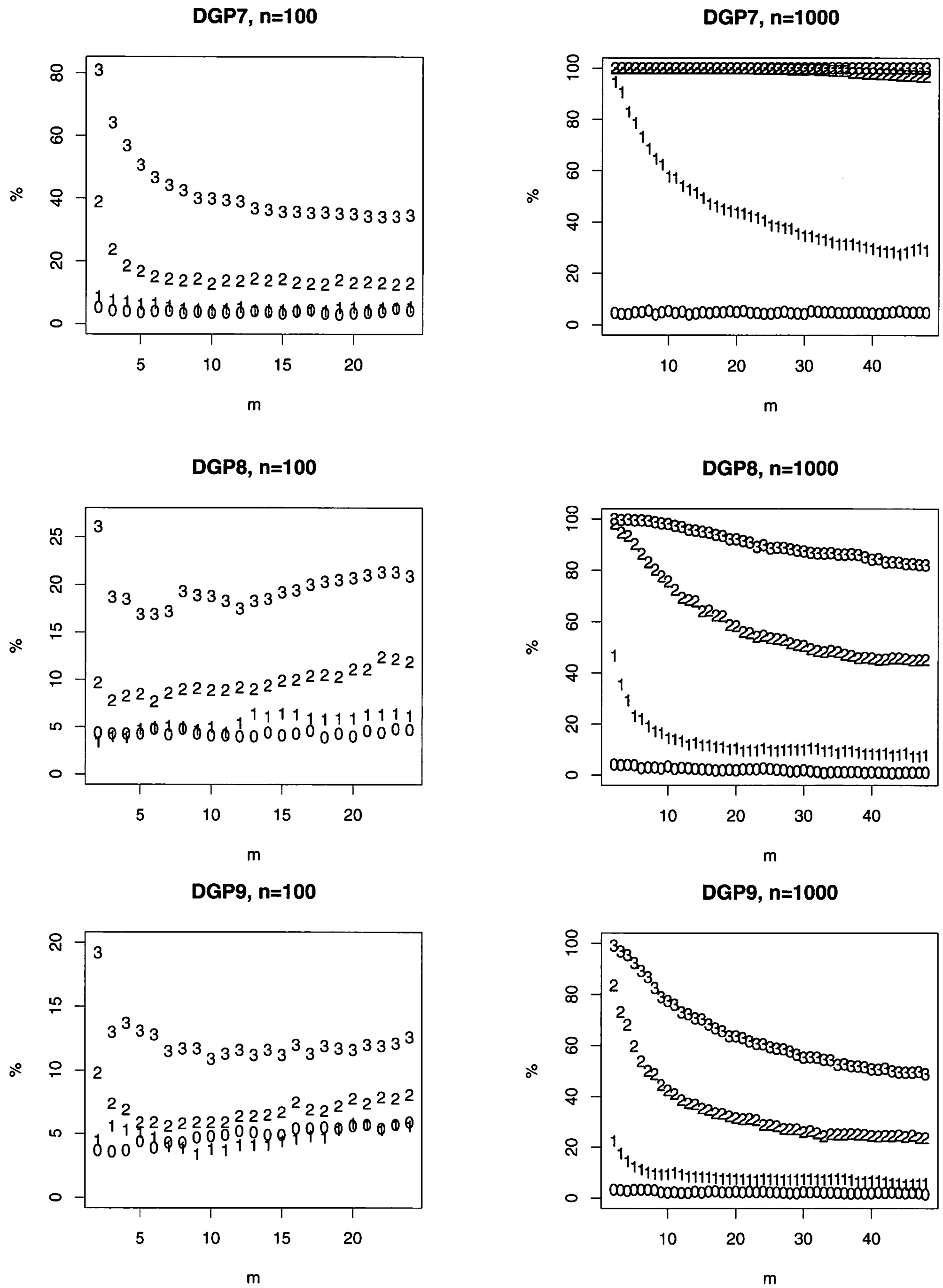

Figure 6: Percentages of the empirical power of $Q_w(m) \sim \chi^2_{4(m-1)}$, where the cases of $c = 0.0, 0.1, \ldots, 0.3$ are denoted by $0, 1, \ldots, 3$.

where

$$G_m = \mathrm{E}\begin{bmatrix}\epsilon_{t-1}\\ \vdots \\ \epsilon_{t-m}\end{bmatrix} \otimes \left[X'_{t-1},\dots,X'_{t-p}\right] = \begin{bmatrix} \Omega(A_0^*)' & O & \dots & \dots & O \\ \Omega(A_1^*)' & \Omega(A_0^*)' & O & \dots & O \\ \vdots & \vdots & & & \Omega(A_0^*)' \\ \vdots & \vdots & & & \vdots \\ \Omega(A_{m-1}^*)' & \Omega(A_{m-2}^*)' & \dots & \dots & \Omega(A_{m-p}^*)' \end{bmatrix}$$

$$= \left[\Omega(A_{i-j}^*)'\right]_{i=1,\dots,m,\ j=1,\dots,p}$$

and $A_k^* = 0$ for $k < 0$. Therefore, $\operatorname{rank}\Phi_m = \operatorname{rank}G_m \cdot \operatorname{rank}I_d = d^2p$. □

We present the following Lemmas, which are required to prove Corollary 1:

Lemma 1 *Under the assumptions in Theorem 2, it holds that, as $m \to \infty$:*

$$A_m^* = O(\lambda^m), \tag{A.3}$$

$$\Phi_M = [\Phi'_m \quad O]' + O(\lambda^{m+1-p}), \tag{A.4}$$

$$D_M^* = \begin{bmatrix} D_m^* & O \\ O & O \end{bmatrix} + O(\lambda^{m+1-p}), \tag{A.5}$$

where $\lambda \in (0,1)$ is such that λ is larger than the absolute value of any eigenvalues of $\widetilde{A}$ and $\Sigma_M^{-1/2}$ in D_M^* is a block upper triangular matrix.

Note that except for the strong white noise case, $\sqrt{n}\widehat{\gamma}_m, \sqrt{n}\operatorname{vec}\widehat{\Gamma}_\epsilon(m+1),\dots,\sqrt{n}\operatorname{vec}\widehat{\Gamma}_\epsilon(M)$ are not independent in general as m, n increase, even if we apply (A.4) (see (4)).

Proof of Lemma 1: We have (A.3) from Lütkepohl (2006, Section 2.1.1 and A.9.1). In addition, (A.4) holds from (A.2). According to Lütkepohl (2006, Section A.9.3), there exists an upper triangular matrix, $\Sigma_m^{1/2}$ such that $\Sigma_m^{1/2}(\Sigma_m^{1/2})' = \Sigma_m$. Furthermore, from Lütkepohl (2006, Section A.10), we have

$$\Sigma_M^{1/2} = \begin{bmatrix} \Sigma_m^{1/2} & C_{12} \\ O & C_{22} \end{bmatrix} \quad \text{and} \quad \Sigma_M^{-1/2} = \begin{bmatrix} \Sigma_m^{-1/2} & -\Sigma_m^{-1/2}C_{12}C_{22}^{-1} \\ O & C_{22}^{-1} \end{bmatrix}.$$

Then, from (A.4), we have:

$$F_M = \left[\Sigma_m^{-1/2}\Phi'_m \quad O\right]' + O(\lambda^{m-p}), \tag{A.6}$$

as $m \to \infty$, which proves (A.5). □

Lemma 2 *Under the assumptions in Corollary 1, for $i < j \le k < l$, as $m(i) \to \infty$, $(\Sigma_M^{1/2})'E_{m(j)}^*\Sigma_M(E_{m(l)}^* - E_{m(k)}^*)\Sigma_M^{1/2} = O(\lambda^{m(k)+1-p})$ and $(\Sigma_M^{1/2})'(E_{m(j)}^* - E_{m(i)}^*)\Sigma_M(E_{m(l)}^* - E_{m(k)}^*)\Sigma_M^{1/2} = O(\lambda^{m(i)+1-p})$.*

From the Craig-Sakamoto Theorem (see, for example, Provost, 1996), Lemma 2 indicates the asymptotic independence of $(Z_M^*)' E_{m(j)} Z_M^*$ and $(Z_M^*)'(E_{m(l)} - E_{m(k)}) Z_M^*$, or $(Z_M^*)'(E_{m(j)} - E_{m(i)}) Z_M^*$ and $(Z_M^*)'(E_{m(l)} - E_{m(k)}) Z_M^*$, as m increases.

Note that Lemma 2 does not hold in general when Σ_M is not block-diagonal. Σ_M may be approximated by a block tri-diagonal matrix, Σ_M^*. However, $(\Sigma_M^*)^{-1}$ is not a block tri-diagonal matrix. See, e.g., Magnus and Neudecker (1999, Theorem 15.9).

Proof of Lemma 2: Similar to the proof of Katayama (2009, Lemma 1), we shall only prove that, as $m(1) \to \infty$:

$$(\Sigma_M^{1/2})' E_{m(1)}^* \Sigma_M (E_{m(2)}^* - E_{m(1)}^*) \Sigma_M^{1/2} = O(\lambda^{m(1)+1-p}). \tag{A.7}$$

However, other cases can be similarly treated. From Lemma 1, we have $E_{m(2)}^* = \text{diag}[E_{m(1)}^*,\ \Sigma(2,2)^{-1},\ O] + O(\lambda^{m(1)+1-p})$. It follows that

$$\begin{aligned}
&(\Sigma_M^{1/2})' E_{m(1)}^* \Sigma_M^{1/2} \cdot (\Sigma_M^{1/2})'(E_{m(2)}^* - E_{m(1)}^*) \Sigma_M^{1/2} \\
&= \begin{bmatrix} I_{m(1)} - D_{m(1)}^* & O & O \\ O & O & O \\ O & O & O \end{bmatrix} \begin{bmatrix} O & O & O \\ O & I_{m(2)-m(1)} & O \\ O & O & O \end{bmatrix} + O(\lambda^{m(1)+1-p}),
\end{aligned}$$

which proves (A.7). □

Proof of Corollary 1: From Lemma 2, the proof of Corollary 1 is treated in a similar manner to the proof of Katayama (2009, Theorem 1). □

Proof of $\widehat{B}_s^*(m) \to 0$ as $m \to \infty$: Since $(\widehat{\Sigma}_m^s)^{-1}$ and $\widehat{\Phi}_m$ are consistent estimators of Σ_m^{-1} and Φ_m, it is enough to examine $\Phi_m' \Sigma_m^{-1} \widehat{\gamma}_m^*$. From (17), the asymptotic variance of $\Phi_m' \Sigma_m^{-1} \widehat{\gamma}_m^*$ is

$$\Phi_m' \Sigma_m^{-1} (\Sigma_m - \Phi_m \Sigma_{\widehat{\theta}_n} \Phi_m') \Sigma_m^{-1} \Phi_m = \Phi_m' \Sigma_m^{-1} \Phi_m \Sigma_{\widehat{\theta}_n} \left(\Sigma_{\widehat{\theta}_n}^{-1} - \Phi_m' \Sigma_m^{-1} \Phi_m \right).$$

It follows that it is enough to show that

$$\Sigma_{\widehat{\theta}_n}^{-1} - \Phi_m' \Sigma_m^{-1} \Phi_m = O(\lambda^{2m}), \tag{A.8}$$

as $m \to \infty$. From (A.2), Hosking (1980) and FR, we have

$$\begin{aligned}
\Sigma_{\widehat{\theta}_n}^{-1} &= \left[\sum_{k=1}^{\infty} A_{k-i}^* \Omega (A_{k-j}^*)' \right]_{i,j=1,\ldots,p} \otimes \Omega^{-1} \\
\Phi_m' \Sigma_m^{-1} \Phi_m &= \left[\sum_{k=1}^{m} A_{k-i}^* \Omega (A_{k-j}^*)' \right]_{i,j=1,\ldots,p} \otimes \Omega^{-1}.
\end{aligned}$$

By these and (A.3), we obtain (A.8). □

References

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

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**, 1509–26.

CHITTURI, R.V. (1974): Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association* **69**, 928–34.

FRANCQ, C., and RAÏSSI, H. (2007): Multivariate portmanteau test for autoregressive models with uncorrelated but non-independent errors. *Journal of Time Series Analysis*, **28(3)**, 454–470.

FRANCQ, C., ROY, R. and ZAKOÏAN, J.-M. (2005): Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* **100**, 532–44.

HOSKING, J. R. M. (1980): The multivariate portmanteau statistic. *Journal of the American Statistical Association* **75**, 343–86.

JOHNSON, N.L., KOTZ, and BALAKRISHANAN, N. (1994): *Continuous Univariate Distributions*, vol.1 (2nd edn). New York: Wiley.

KATAYAMA, N. (2008): An improvement of the portmanteau statistic. *Journal of Time Series Analysis*, **29(2)**, 359–370.

———— (2009): On multiple portmanteau tests. *Journal of Time Series Analysis*, **30(5)**, 487–504.

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

LI, W.K. (2004): *Diagnostic Checks in Time Series*. New York: Chapman & Hall.

LJUNG, G. M. (1986): Diagnostic testing of univariate time series models. *Biometrika* **73**, 725–30.

LÜTKEPOHL, H. (2006): *New Introduction to Multiple Time Series Analysis*. Berlin: Springer Verlag.

MAGNUS, J.R. & NEUDECKER, H. (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edn, New York: Wiley.

PROVOST, S.B. (1996): On Craig's theorem and its generalizations. *Journal of Statistical Planning and Inference*, **53**, 311–321.

ROMANO, J.L. & THOMBS, L.A. (1996): Inference for Autocorrelations under weak assumptions. *Journal of the American Statistical Association*, **91**, 590–600.