ON THE BIAS OF THE PORTMANTEAU STATISTIC *

BY NAOYA KATAYAMA †

Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186, Japan.
pg01207@srv.cc.hit-u.ac.jp

Abstract. The portmanteau statistic is based on the first m residual autocorrelations, and is used for diagnostic checks on the adequacy of a fitting model. In this paper, we propose a modified portmanteau statistic with a correction factor that allows for the use of small values of m and eliminates the positively biased random variable for the chi-squared approximation. For this modification we take a different approach to that suggested by Ljung (1986). The power of the portmanteau statistic and the Lagrange multiplier test statistics are also examined, and their empirical behaviour is clarified in terms of asymptotic theory.

Keywords. Autoregressive-moving average model; Lagrange multiplier test; Portmanteau test; Residual autocorrelations; Time series model checking.

1 INTRODUCTION

Suppose that a univariate time series $\{x_t\}$ is generated by an autoregressive-moving average, ARMA(p,q), model:

$$\alpha(L;\boldsymbol{\alpha}_0)x_t = \beta(L;\boldsymbol{\beta}_0)\varepsilon_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{1.1}$$

where $p+q>0$, $L^i x_t = x_{t-i}$, $\{\varepsilon_t\}$ is iid $(0,\sigma_0^2)$, $\alpha(L;\boldsymbol{\alpha}_0) = 1-\sum_{i=1}^{p}\alpha_i^0 L^i$, $\beta(L;\boldsymbol{\beta}_0) = 1+\sum_{i=1}^{q}\beta_i^0 L^i$, $\boldsymbol{\alpha}_0 = (\alpha_1^0,\ldots,\alpha_p^0)'$, and $\boldsymbol{\beta}_0 = (\beta_1^0,\ldots,\beta_q^0)'$. It is assumed that the above model is stationary, invertible, and not redundant, so that the polynomials $\alpha(z;\boldsymbol{\alpha}_0)=0$ and $\beta(z;\boldsymbol{\beta}_0)=0$ have no roots in common, and that all the roots are outside the unit circle. Given a process, $\{x_t\}_{t=1}^n$, defined in (1.1), the non-linear least squares estimator of $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0')'$, $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\alpha}}_n{}', \widehat{\boldsymbol{\beta}}_n{}')'$, is obtained by minimizing the sum of squared residuals. The residuals, $\widehat{e}_t = e_t(\widehat{\boldsymbol{\theta}}_n)$, for $t=1,\ldots,n$, from the fitted models are given by $e_t(\widehat{\boldsymbol{\theta}}_n) = \sum_{i=0}^{t-1}\pi_i(\widehat{\boldsymbol{\theta}}_n)x_{t-i}$, and $\pi_i(\boldsymbol{\theta})$s are defined by $\alpha(L;\boldsymbol{\alpha})/\beta(L;\boldsymbol{\beta}) = \sum_{i=0}^{\infty}\pi_i(\boldsymbol{\theta})L^i$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ is any vector corresponding to $\boldsymbol{\theta}_0$ in the parameter space. Box & Jenkins (1976, Chapter 8) pointed out that it is important to check the assumption of the independence of $\{\varepsilon_t\}$ by using the residuals, which should behave in a manner that is consistent with the model. Li (2004) provides a good survey of this issue. It is reasonable to check the adequacy of the model fit by examining the residual autocorrelations, as follows:

$$\widehat{r}_k = \sum_{t=k+1}^{n}\widehat{e}_t\widehat{e}_{t-k}\Bigg/\sum_{t=1}^{n}\widehat{e}_t^2, \qquad k = 1, 2, \ldots, n-1. \tag{1.2}$$

The asymptotic joint distribution of $\widehat{\boldsymbol{r}}_n = (\widehat{r}_1,\ldots,\widehat{r}_m)'$ has been analysed by Box & Pierce (1970) and McLeod (1978). They have shown the following: for moderately large n and m, such that $p+q<m<n$, $\widehat{\boldsymbol{r}}_n$ is approximately normal with a mean vector of $\boldsymbol{0}_m$, and an idempotent covariance matrix of $(\mathbb{I}_m - \boldsymbol{C})/n$, with rank $m-p-q$, where $\boldsymbol{0}_m$ is the m-dimensional zero vector and $\mathbb{I}_m$ is an $m\times m$ identity matrix, which is a specific expression for $\boldsymbol{C}$ that is given in equation (15) of McLeod (1978); the portmanteau statistic:

$$Q_m = n\sum_{k=1}^{m}\widehat{r}_k^2, \tag{1.3}$$

*Revision Date: January 13, 2007.

†At Faculty of Economics, Kyushu University, 6-19-1, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan from April 1st, 2007.

is distributed as χ^2_{m-p-q}, where χ^2_m denotes a chi-squared variable with m degrees of freedom. They suggested using Q_m for a portmanteau test of model adequacy. Ljung & Box (1978) showed that a better approximation can be achieved by using the following modified portmanteau statistic:

$$Q^*_m = n(n+2)\sum_{k=1}^{m} \widehat{r}^2_k/(n-k). \tag{1.4}$$

These statistics are restricted to large m (e.g., $m \geq 20$) such that $\boldsymbol{C}$ is an idempotent matrix. However, many empirical studies have shown that as m increases, the empirical significance level also increases and the empirical power decreases. (See Ljung, 1986, and the references therein.) Therefore, the choice of the optimum value of m remains an issue.

In studying this problem, Ljung (1986) suggested using $a\chi^2_b$ in place of χ^2_{m-p-q} for small m, where a and b are calculated from the eigenvalues of $\mathbb{I}_m - \boldsymbol{C}$. Ljung's (1986) empirical test of the AR(1) model also revealed the following: the empirical significance level of the χ^2_{m-1} approximation exceeds the nominal significance level for values of the AR(1) parameters that are close to unity and for diminishing values of m; the empirical power of the χ^2_{m-1} approximation performs well when $m \geq 3$, and is virtually identical to that of the $a\chi^2_b$ approximation; the empirical significance level and power of the $a\chi^2_b$ approximation are adequate; the empirical power of the $a\chi^2_b$ and χ^2_{m-1} approximation matches that of the Lagrange multiplier (LM) test.

In this study, we construct a positively biased random variable of Q_m and Q^*_m for the chi-squared approximation and develop a theoretical link to Ljung's (1986) empirical study. In addition, we propose a modification of Q^*_m that allows for the use of small values of m; in doing so, we adopt a different approach to the one used by Ljung (1986) to develop his $a\chi^2_b$ approximation. Our modified statistics, Q^{**}_m, are defined by Q^*_m, from which we subtract a positive correction factor; this eliminates the positive bias in the Q^*_m for a chi-squared approximation. Under the null hypothesis, Q^{**}_m is distributed approximately χ^2_{m-p-q} for $m > p+q$, whereas Q^*_m is distributed approximately $a\chi^2_b$ for all $m \geq 1$ considered by Ljung (1986). However, a and b must be estimated, and in general, b is a positive real number. Therefore, the rejection region can be difficult to calculate and depends on the estimates. Empirical studies show that both the $Q^{**}_m \sim \chi^2_{m-p-q}$ and the $Q^*_m \sim a\chi^2_b$ tests perform equally well in terms of significance levels and power.

In Section 2, we discuss the asymptotic distribution of Q_m and Q^*_m, and show that these statistics yield a positively biased random variable for the chi-squared approximation; bias-corrected test statistics are then proposed. In Section 3, we compare the power of the portmanteau statistics and the LM test statistic, and show that using appropriate values of m makes the portmanteau statistics as powerful as the LM test under local alternatives. We also discuss empirical significance levels and power. Mathematical proofs are given in the appendix.

Throughout this paper, let $\partial f(\boldsymbol{x})/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{y}} = \partial f(\boldsymbol{y})/\partial \boldsymbol{x}$, $f^{(1)}(\boldsymbol{y}) = \partial f(\boldsymbol{y})/\partial \boldsymbol{x}$. In addition, let random-variable sequences be $\{X_n\}$ and $\{Y_n\}$. Then, $X_n \overset{\text{a}}{=} Y_n$ denotes $X_n - Y_n = o_p(1)$, as $n \to \infty$. This notation is also used when $\{X_n\}$ or $\{Y_n\}$ is a sequence. Let the continuous-type random variables X_n and Y_n have distribution functions of F_n and G_n, respectively. Then, $X_n \overset{\text{a}}{\sim} Y_n$ indicates that both have the same asymptotic distribution, as follows: $\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} G_n(x)$.

2 THE POSITIVE BIAS OF THE PORTMANTEAU STATISTIC

2.1 Derivation of the positive bias

In this section, we discuss the bias of Q_m and Q^*_m under the null hypothesis. First, we review the results of McLeod (1978). Following the proof of McLeod (1978, equation (34)), by using a Taylor-series expansion of $\widehat{\boldsymbol{r}}_n$ around $\widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$, as $n \to \infty$, we have:

$$\widehat{\boldsymbol{r}}_n = \boldsymbol{r}_n + \boldsymbol{X}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) + O_p(1/n), \tag{2.1}$$

where $\boldsymbol{r}_n = (r_1, \ldots, r_m)'$, $r_k = \sum_{t=k+1}^{n} \varepsilon_t\varepsilon_{t-k}/\sum_{t=1}^{n}\varepsilon^2_t$, $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{\theta}_0) = (\boldsymbol{d}_1(\boldsymbol{\theta}_0), \ldots, \boldsymbol{d}_m(\boldsymbol{\theta}_0))'$, $\boldsymbol{d}_i(\boldsymbol{\theta}_0) = \sum_{k=0}^{i} \pi^{(1)}_k(\boldsymbol{\theta}_0)\psi_{i-k}(\boldsymbol{\theta}_0)$, and $\{\psi_i(\boldsymbol{\theta}_0)\}$ is given by $\beta(L;\boldsymbol{\beta}_0)/\alpha(L;\boldsymbol{\alpha}_0) = \sum_{i=0}^{\infty}\psi_i(\boldsymbol{\theta}_0)L^i$. That is, $\{\boldsymbol{d}_i(\boldsymbol{\theta}_0)\}$ are given by: $\varepsilon_t(\boldsymbol{\theta}) = \sum_{i=0}^{\infty}\pi_i(\boldsymbol{\theta})x_{t-i}$ and $\varepsilon^{(1)}_t(\boldsymbol{\theta}_0) = \sum_{i=0}^{\infty}\pi^{(1)}_i(\boldsymbol{\theta}_0)x_{t-i} = \sum_{i=1}^{\infty}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\varepsilon_{t-i}$. Each (i,j) element of the partitioned matrix of $\boldsymbol{X}$ is given in (16) of McLeod (1978), as follows:

$$\boldsymbol{X} = \left(-\alpha^{0*}_{i-j} \vdots -\beta^{0*}_{i-j} \right), \tag{2.2}$$

where α_i^{0*}s and β_i^{0*}s are defined by $1/\alpha(L;\boldsymbol{\alpha}_0) = \sum_{i=0}^{\infty}\alpha_i^{0*}L^i$, $1/\beta(L;\boldsymbol{\beta}_0) = \sum_{i=0}^{\infty}\beta_i^{0*}L^i$ and $\alpha_i^{0*} = \beta_i^{0*} = 0$ for $i < 0$. Note that $\mathrm{rank}(\boldsymbol{X}) = p+q$, $\boldsymbol{I}(\boldsymbol{\theta}_0) = \sum_{i=1}^{\infty}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)'$ is the Fisher information matrix of the model (1.1), and $\sqrt{n}((\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)'\ \boldsymbol{r}_n{}')'$ has an approximately normal distribution with a mean vector of $\boldsymbol{0}_m$ and the following covariance matrix:

$$\begin{pmatrix} \boldsymbol{I}(\boldsymbol{\theta}_0)^{-1} & -\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{X}' \\ -\boldsymbol{X}\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1} & \mathbb{I}_m \end{pmatrix} \tag{2.3}$$

from (28) of McLeod (1978). By using these results, McLeod (1978, Theorem 1) showed that $\widehat{\boldsymbol{r}}_n$ is approximately normal, with a mean vector of $\boldsymbol{0}_m$ and a covariance matrix of $(\mathbb{I}_m - \boldsymbol{C})/n$, where:

$$\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{\theta}_0) = \boldsymbol{X}\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{X}'. \tag{2.4}$$

Let:

$$\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{\theta}_0) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'. \tag{2.5}$$

If m is sufficiently large that $\boldsymbol{X}'\boldsymbol{X} = \sum_{i=1}^{m}\boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)' \overset{\mathrm{a}}{=} \boldsymbol{I}(\boldsymbol{\theta}_0)$, then $\mathbb{I}_m - \boldsymbol{C}$ is approximately equivalent to $\mathbb{I}_m - \boldsymbol{D}$ and is idempotent with a rank of $m-p-q$. Thus, Q_m and Q_m^* are distributed approximately χ^2_{m-p-q}.

Now, we follow Box & Pierce (1970, Section 2.4). Multiplying both sides of (2.1) by $\sqrt{n}(\mathbb{I}_m - \boldsymbol{D})$, and using $(\mathbb{I}_m - \boldsymbol{D})\boldsymbol{X} = \boldsymbol{0}_m$ yields:

$$\sqrt{n}(\mathbb{I}_m - \boldsymbol{D})\widehat{\boldsymbol{r}}_n = \sqrt{n}(\mathbb{I}_m - \boldsymbol{D})\boldsymbol{r}_n + O_p\big(1/\sqrt{n}\big). \tag{2.6}$$

Similarly, using:

$$\begin{aligned}\boldsymbol{T}_n &= \mathrm{diag}\left(\sqrt{\frac{n(n+2)}{n-1}}, \sqrt{\frac{n(n+2)}{n-2}}, \ldots, \sqrt{\frac{n(n+2)}{n-m}}\right) \\ &= \sqrt{n}\mathbb{I}_m + O\big(m/\sqrt{n}\big),\end{aligned} \tag{2.7}$$

yields:

$$(\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n = (\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\boldsymbol{r}_n + O_p\big(1/\sqrt{n}\big). \tag{2.8}$$

Because $Q_m = n\widehat{\boldsymbol{r}}_n{}'\widehat{\boldsymbol{r}}_n$ and $Q_m^* = \widehat{\boldsymbol{r}}_n{}'\boldsymbol{T}_n^2\widehat{\boldsymbol{r}}_n$ and the first terms on the right-hand sides of (2.6) and (2.8) are approximately $N(\boldsymbol{0}_m, (\mathbb{I}_m - \boldsymbol{D}))$, (2.6) and (2.8) respectively indicate that Q_m and Q_m^* have positively biased random variables:

$$B_{m,n} = n\widehat{\boldsymbol{r}}_n{}'\boldsymbol{D}\widehat{\boldsymbol{r}}_n \quad \text{and} \quad B_{m,n}^* = \widehat{\boldsymbol{r}}_n{}'\boldsymbol{T}_n\boldsymbol{D}\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n, \tag{2.9}$$

for a χ^2_{m-p-q} approximation.

2.2 The convergence of the positively biased random variables

We now prove that $B_{m,n}$ and $B_{m,n}^*$ are negligible as $m, n \to \infty$. To prove this, Box & Pierce (1970, p. 1513) assumed that, for the AR(p) model, $m = m_n$ increases as n increases, for all α_j^{0*}, where $j \geq m_n - p$ are of the order $1/\sqrt{n}$; the ratio m_n/n is of the order $1/\sqrt{n}$. However, in an empirical study, Ljung (1986, Section 3) found these assumptions to be inappropriate. This is because the empirical significance level of Q_m^* was found to be stable even for small m, and this stability was found to depend on the value of the AR(1) parameter. This finding is theoretically plausible because $B_{m,n} = n\widehat{\boldsymbol{r}}_n{}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{r}}_n$ and $B_{m,n}^* = \widehat{\boldsymbol{r}}_n{}'\boldsymbol{T}_n\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n$, and, from McLeod (1978, Theorem 1), the asymptotic variances of $\sqrt{n}\boldsymbol{X}'\widehat{\boldsymbol{r}}_n$ and $\boldsymbol{X}'\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n$ are:

$$\begin{aligned}\boldsymbol{X}'(\mathbb{I}_m - \boldsymbol{C})\boldsymbol{X} &= \boldsymbol{X}'\boldsymbol{X}\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}\big(\boldsymbol{I}(\boldsymbol{\theta}_0) - \boldsymbol{X}'\boldsymbol{X}\big) \\ &= O\left(\sum_{j=m+1}^{\infty}\boldsymbol{d}_j(\boldsymbol{\theta}_0)\boldsymbol{d}_j(\boldsymbol{\theta}_0)'\right) \\ &= O\big(\epsilon^{2m}\big),\end{aligned} \tag{2.10}$$

as $m \to \infty$, where $\epsilon \in (0,1)$ is such that ϵ is larger than the absolute value of any roots of $\alpha(z^{-1};\boldsymbol{\alpha}_0)\beta(z^{-1};\boldsymbol{\beta}_0) = 0$. Note that, for the AR(1) model, ϵ is the absolute value of the AR(1) parameter.

Hence, we assume the following.

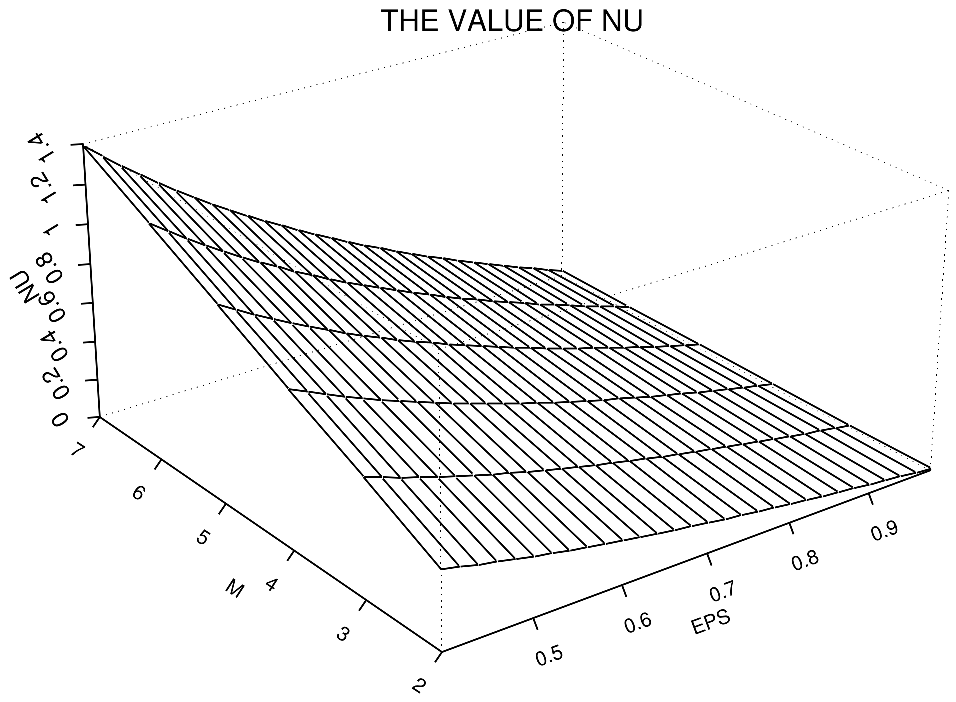

Figure 1: The value of ν. EPS denotes ϵ and $\nu = -M \log \epsilon / \log n$ is given by (2.12) with $n = 100$.

Assumption 1 *For $m = m_n$ given by $p + q < m < n$, there exist values of $\nu > 0$ and $M \geq 2$ such that:*

$$m_n = O(\log n), \tag{2.11}$$

$$M = m_n + 1 - \max(p, q) = -\nu \log n / \log \epsilon. \tag{2.12}$$

Thus, from the stationary and invertibility assumptions, $|\alpha_j^{0*}| = O(\epsilon^j)$ and $|\beta_j^{0*}| = O(\epsilon^j)$, it follows that:

$$|\boldsymbol{d}_j(\boldsymbol{\theta}_0)| = O\big(|\alpha_{j-p}^{0*}|\big) + O\big(|\beta_{j-q}^{0*}|\big) = O\Big(\epsilon^{j-\max(p,q)}\Big), \tag{2.13}$$

for $j > m$ and by (2.12):

$$\sum_{j=m+1}^{\infty} |\boldsymbol{d}_j(\boldsymbol{\theta}_0)| = O(n^{-\nu}) \quad \text{and} \quad \sum_{j=m+1}^{\infty} \boldsymbol{d}_j(\boldsymbol{\theta}_0)\boldsymbol{d}_j(\boldsymbol{\theta}_0)' = O(n^{-2\nu}), \tag{2.14}$$

as $m, n \to \infty$. Using (2.14), we obtain the following theorem.

Theorem 2.1 *Under the Assumption 1, it holds that, as $m, n \to \infty$:*

$$B_{m,n} = O_p\Big(n^{-2\min(\nu,0.5)}\Big) \quad \textit{and} \quad B_{m,n}^* = O_p\Big(n^{-2\min(\nu,0.5)}\Big), \tag{2.15}$$

where $B_{m,n}$ and $B_{m,n}^*$ are given by (2.9).

Thus, the chi-squared approximation of Q_m and Q_m^* performs poorly when ν is small.

In Figure 1, we represent the value of ν by using the equality $\nu = -M \log \epsilon / \log n$, with EPS used to denote ϵ, and $n = 100$. The figure indicates that $B_{m,n}$ and $B_{m,n}^*$ converges slowly, as $M \to 2$ and $\epsilon \to 1$, which is consistent with the empirical study of Ljung (1986, Section 3).

However, the results implied by (2.15) and illustrated in Figure 1 are theoretically interesting but not practically useful. This is because the values of ν and ϵ are generally unknown. Although we have shown that a sufficient condition for a chi-squared approximation is that $m = O(\log n)$ (in which case, the rate of increase is less than the order of $\sqrt{n}$), our analysis has not revealed the optimal choice of m for the portmanteau test.

For an intuitive understanding of $B_{m,n}$, we provide a proof of the convergence of $B_{m,n}$ without making Assumption 1. Because $B_{m,n} = n\widehat{\boldsymbol{r}}_n{}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{r}}_n$, from (A.5) and (A.6), we have:

$$\sqrt{n}\boldsymbol{X}'\widehat{\boldsymbol{r}}_n = \sqrt{n}\sum_{k=1}^{m}\boldsymbol{d}_k(\boldsymbol{\theta}_0)\widehat{r}_k = \sqrt{n}\sum_{k=1}^{m}\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{r}_k + O_p\left(\frac{1}{\sqrt{n}}\right), \qquad \text{as } n\to\infty, \tag{2.16}$$

and

$$\sqrt{n}\sum_{k=1}^{m}\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{r}_k \to \sum_{k=1}^{n-1}\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{r}_k \propto \sum_{k=2}^{n}\frac{\partial e_k(\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}e_k(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{0}_{p+q}, \qquad \text{as } m\to n-1. \tag{2.17}$$

This expression implies that $B_{m,n}$ becomes negligible as $m,n\to\infty$. The last equality of (2.17) follows from the fact that $\widehat{\boldsymbol{\theta}}_n$ satisfies the first-order conditions of the objective function. The analysis of $B^*_{m,n}$ is similar.

2.3 The bias-corrected portmanteau statistics

Ljung & Box (1978) showed that the Q^*_m statistic produces much better approximations than does the Q_m. Therefore, we focus on the Q^*_m statistic and propose a bias-corrected Q^*_m statistic. Let $\widehat{\boldsymbol{X}} = \boldsymbol{X}(\widehat{\boldsymbol{\theta}}_n)$. Because $\widehat{\boldsymbol{D}} = \boldsymbol{D}(\widehat{\boldsymbol{\theta}}_n) = \widehat{\boldsymbol{X}}(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'$ is a consistent estimator of $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{\theta}_0)$ in $B^*_{m,n}$, it is reasonable to consider a bias-corrected portmanteau statistic that is obtained from (2.8) and (2.9):

$$Q^{**}_m = Q^*_m - \widehat{B}^*_{m,n}, \tag{2.18}$$

where Q^*_m is given by (1.4) and $\widehat{B}^*_{m,n} = \widehat{\boldsymbol{r}}_n{}'\boldsymbol{T}_n\widehat{\boldsymbol{D}}\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n$. The following theorem shows the statistic Q^{**}_m eliminates the disadvantageous effect on Q^*_m, $B^*_{m,n}$, for the χ^2_{m-p-q} approximation.

Theorem 2.2 *For a fixed $m > p+q$, it holds that, as $n\to\infty$:*

$$(\mathbb{I}_m - \widehat{\boldsymbol{D}})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n = (\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\boldsymbol{r}_n + O_p\big(1/\sqrt{n}\big). \tag{2.19}$$

Unlike Q_m and Q^*_m, the χ^2_{m-p-q} approximation of the statistic Q^{**}_m is not required for large m, but only for $m > p+q$.

Note that (2.7), (A.6) and (A.7) can be used to replace the bias-corrected term for Q^{**}_m, given by $\widehat{B}^*_{m,n}$, by the following approximately equivalent statistic:

$$n\boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n)'\left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1}\boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n), \tag{2.20}$$

where

$$\boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n) = \sum_{k=m+1}^{n-1}\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{r}_k = -\sum_{k=1}^{m}\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{r}_k = -\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{r}}_n. \tag{2.21}$$

The empirical significance levels corresponding to 5% and 10% for the AR(1) model, $(1-\alpha^0L)x_t = \varepsilon_t$, are given in Table 1, where $\{\varepsilon_t\}$ is iid $N(0,1)$. For each cell in the three columns of Table 1, headed $m = 2,3,5,25$, the first number is the empirical significance level of $Q^*_m \sim \chi^2_{m-1}$, the number in parentheses is the empirical significance level of $Q^*_m \sim a\chi^2_b$, which is based on Ljung (1986), and the number in the square brackets is the empirical significance level of $Q^{**}_m \sim \chi^2_{m-1}$. The values of a and b were calculated from a non-linear least squares estimator of α^0, $\widehat{\alpha}$, and $\widehat{B}^*_{m,n}$ is given by:

$$\widehat{B}^*_{m,n} = n(n+2)\frac{1-\widehat{\alpha}^2}{1-\widehat{\alpha}^{2m}}\left(\sum_{k=1}^{m}\frac{\widehat{\alpha}^{k-1}}{\sqrt{n-k}}\widehat{r}_k\right)^2.$$

The results for $Q^*_m \sim a\chi^2_b$ and $Q^{**}_m \sim \chi^2_{m-1}$ provide approximately valid tests uniformly for α^0 and m, whereas the results for $Q^*_m \sim \chi^2_{m-1}$ were poor when m was small and α^0 was close to unity. In addition, three approximations converged to similar empirical significance levels as m increased and ϵ decreased. These results are consistent with those of Ljung (1986, Section 3) and the theoretical results discussed in this section. However, as many researchers have pointed out, for large m, the empirical significance levels of portmanteau test statistics tend to exceed those predicted by asymptotic theory.

Table 1: Empirical significance level of the statistic $Q_m^* \sim \chi^2_{m-1}$, $Q_m^* \sim a\chi^2_b$ (in parentheses), and $Q_m^{**} \sim \chi^2_{m-1}$ (in square brackets) for the AR(1) model, $(1-\alpha^0 L)x_t = \varepsilon_t$; $n = 100$, $m = 2, 3, 5, 25$, and $\alpha^0 = 0.4, 0.7, 0.8, 0.9$, with 10,000 replications. The significance levels were 5% and 10%.

%	α^0	$m=2$			$m=3$			$m=5$			$m=25$		
5%	0.4	4.92	(4.84)	[4.82]	4.65	(4.63)	[4.63]	4.47	(4.47)	[4.47]	6.18	(6.18)	[6.18]
	0.7	5.58	(4.48)	[4.65]	4.96	(4.65)	[4.62]	4.88	(4.82)	[4.82]	6.26	(6.26)	[6.26]
	0.8	6.89	(4.39)	[4.60]	5.46	(4.58)	[4.71]	4.93	(4.73)	[4.68]	6.50	(6.50)	[6.50]
	0.9	8.37	(4.11)	[4.37]	6.00	(4.07)	[4.26]	4.87	(4.13)	[4.31]	5.99	(5.98)	[5.99]
10%	0.4	10.10	(9.73)	[9.88]	9.97	(9.92)	[9.94]	9.48	(9.48)	[9.48]	10.49	(10.49)	[10.49]
	0.7	11.47	(8.55)	[9.38]	10.23	(9.47)	[9.66]	9.64	(9.47)	[9.47]	11.12	(11.12)	[11.12]
	0.8	14.28	(9.35)	[9.62]	11.12	(9.27)	[9.37]	9.89	(9.34)	[9.36]	10.90	(10.90)	[10.90]
	0.9	17.22	(8.69)	[8.98]	12.15	(8.21)	[8.62]	9.41	(7.92)	[8.23]	9.91	(9.91)	[9.91]

3 THE POWER OF THE PORTMANTEAU STATISTICS

3.1 The relationship between the power of the portmanteau and LM tests under local alternatives

We compare the power of the portmanteau statistics, Q_m^* and Q_m^{**}, to the power of the LM test statistic of Godfrey (1979) under local alternatives. As an example of an alternative model, we considered an ARMA$(p, q+r)$ model:

$$\alpha(L;\boldsymbol{\alpha}_0)x_t = \gamma(L;\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)\varepsilon_t, \tag{3.1}$$

where $\gamma(L;\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) = 1 + \sum_{i=1}^{q+r}\beta_i^0 L^i$, $\boldsymbol{\gamma}_0 = (\beta^0_{q+1},\dots,\beta^0_{q+r})' = \boldsymbol{c}/\sqrt{n}$, $\boldsymbol{c}$ is a constant vector, $\alpha(z;\boldsymbol{\alpha}_0) = 0$ and $\gamma(z;\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) = 0$ have no roots in common, and the absolute value of any root of $\alpha(z^{-1};\boldsymbol{\alpha}_0)\gamma(z^{-1};\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) = 0$ is less than $\epsilon \in (0,1)$. To examine the power of these statistics, we introduce the following notations. Let $\boldsymbol{\delta}_0 = (\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0', \boldsymbol{\gamma}_0')' = (\boldsymbol{\theta}_0', \boldsymbol{\gamma}_0')'$, $\boldsymbol{\delta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}')'$ be any vector corresponding to $\boldsymbol{\delta}_0$ in the parameter space, and let $\boldsymbol{I}(\boldsymbol{\delta}_0)$ be a Fisher information matrix of model (3.1), as follows:

$$\boldsymbol{I}(\boldsymbol{\delta}_0) = \begin{pmatrix} \boldsymbol{I}_{\theta\theta}(\boldsymbol{\delta}_0) & \boldsymbol{I}_{\theta\gamma}(\boldsymbol{\delta}_0) \\ \boldsymbol{I}_{\theta\gamma}(\boldsymbol{\delta}_0)' & \boldsymbol{I}_{\gamma\gamma}(\boldsymbol{\delta}_0) \end{pmatrix} = \sum_{i=1}^{\infty} \begin{pmatrix} \boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)\boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)' & \boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)\boldsymbol{d}_{\gamma,i}(\boldsymbol{\delta}_0)' \\ \boldsymbol{d}_{\gamma,i}(\boldsymbol{\delta}_0)\boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)' & \boldsymbol{d}_{\gamma,i}(\boldsymbol{\delta}_0)\boldsymbol{d}_{\gamma,i}(\boldsymbol{\delta}_0)' \end{pmatrix}, \tag{3.2}$$

where $\{\boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)\}$ and $\{\boldsymbol{d}_{\gamma,i}(\boldsymbol{\delta}_0)\}$ are given by: $\varepsilon_t(\boldsymbol{\delta}) = \{\alpha(L;\boldsymbol{\alpha})/\gamma(L;\boldsymbol{\beta},\boldsymbol{\gamma})\}x_t$,

$$\left[\frac{\partial\varepsilon_t(\boldsymbol{\delta}_0)}{\partial\boldsymbol{\alpha}'} \quad \frac{\partial\varepsilon_t(\boldsymbol{\delta}_0)}{\partial\boldsymbol{\beta}'}\right]' = -\left[\frac{(L, L^2,\dots,L^p)}{\alpha(L;\boldsymbol{\alpha}_0)} \quad \frac{(L, L^2,\dots,L^q)}{\gamma(L;\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)}\right]'\varepsilon_t = \sum_{i=1}^{\infty}\boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)\varepsilon_{t-i},$$

$$\frac{\partial\varepsilon_t(\boldsymbol{\delta}_0)}{\partial\boldsymbol{\gamma}} = -\frac{(L^{q+1}, L^{q+2},\dots,L^{q+r})'}{\gamma(L;\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)}\varepsilon_t = \sum_{i=1}^{\infty}\boldsymbol{d}_{\gamma,i}(\boldsymbol{\delta}_0)\varepsilon_{t-i},$$

respectively. Let $\boldsymbol{X}_\theta = \boldsymbol{X}_\theta(\boldsymbol{\delta}_0) = (\boldsymbol{d}_{\theta,1}(\boldsymbol{\delta}_0),\dots,\boldsymbol{d}_{\theta,m}(\boldsymbol{\delta}_0))'$, $\boldsymbol{X}_\gamma = \boldsymbol{X}_\gamma(\boldsymbol{\delta}_0) = (\boldsymbol{d}_{\gamma,1}(\boldsymbol{\delta}_0),\dots,\boldsymbol{d}_{\gamma,m}(\boldsymbol{\delta}_0))'$, $\boldsymbol{D}_\theta = \boldsymbol{X}_\theta(\boldsymbol{X}_\theta{}'\boldsymbol{X}_\theta)^{-1}\boldsymbol{X}_\theta{}'$ and let the restricted estimate be $\widehat{\boldsymbol{\delta}}_n = (\widehat{\boldsymbol{\theta}}_n{}', \boldsymbol{0}_r')'$.

Using $\boldsymbol{D} = \boldsymbol{D}_\theta + O(1/\sqrt{n})$, we obtain the following theorem.

Theorem 3.1 *Under the model* (3.1), *for a fixed $m > p+q$, it follows that, as $n \to \infty$:*

$$(\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n = (\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\boldsymbol{r}_n - (\mathbb{I}_m - \boldsymbol{D}_\theta)\boldsymbol{X}_\gamma\boldsymbol{c} + O_p\big(1/\sqrt{n}\big), \tag{3.3}$$

and

$$(\mathbb{I}_m - \widehat{\boldsymbol{D}})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n = (\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\boldsymbol{r}_n - (\mathbb{I}_m - \boldsymbol{D}_\theta)\boldsymbol{X}_\gamma\boldsymbol{c} + O_p\big(1/\sqrt{n}\big). \tag{3.4}$$

As $n \to \infty$, it follows that:

$$Q_m^{**} \overset{a}{\sim} \chi^2_{m-p-q}\big(\boldsymbol{c}'\boldsymbol{X}_\gamma{}'(\mathbb{I}_m - \boldsymbol{D}_\theta)\boldsymbol{X}_\gamma\boldsymbol{c}\big), \tag{3.5}$$

where $\chi^2_m(\lambda)$ denotes a non-central chi-squared variable with m degrees of freedom and a non-centrality parameter, λ. This implies that the power of Q_m^{**} increases as $\boldsymbol{c}'\boldsymbol{X}_\gamma{}'(\mathbb{I}_m - \boldsymbol{D}_\theta)\boldsymbol{X}_\gamma\boldsymbol{c}$ increases.

Theorem 3.2 *Theorem 2.1 still holds if the model* (1.1) *is replaced by the model* (3.1).

It follows from (3.3) that, when m, n is large, Q_m^* is also distributed approximately $\chi^2_{m-p-q}(\boldsymbol{c}'\boldsymbol{X}_\gamma{}'(\mathbb{I}_m - \boldsymbol{D}_\theta)\boldsymbol{X}_\gamma\boldsymbol{c})$ under the local alternatives.

The LM test statistic, $LM(r)$, used for testing the hypothesis that $r = 0$, is presented by Godfrey (1979). It follows from, for example, Kohn (1979, Theorem 5) and Poskitt & Tremayne (1982, Section 2) that:

$$LM(r) \overset{a}{\sim} \chi^2_r(\boldsymbol{c}'\boldsymbol{\Sigma}(\boldsymbol{\delta}_0)\boldsymbol{c}), \tag{3.6}$$

as $n \to \infty$, where $\boldsymbol{\Sigma}(\boldsymbol{\delta}_0) = \boldsymbol{I}_{\gamma\gamma}(\boldsymbol{\delta}_0) - \boldsymbol{I}_{\theta\gamma}(\boldsymbol{\delta}_0)'\boldsymbol{I}_{\theta\theta}(\boldsymbol{\delta}_0)^{-1}\boldsymbol{I}_{\theta\gamma}(\boldsymbol{\delta}_0)$.

Note that, as $m, n \to \infty$, $\boldsymbol{X}_\theta{}'\boldsymbol{X}_\gamma \overset{a}{=} \boldsymbol{I}_{\theta\gamma}(\boldsymbol{\delta}_0)$, $\boldsymbol{X}_\theta{}'\boldsymbol{X}_\theta \overset{a}{=} \boldsymbol{I}_{\theta\theta}(\boldsymbol{\delta}_0)$, $\boldsymbol{X}_\gamma{}'\boldsymbol{X}_\gamma \overset{a}{=} \boldsymbol{I}_{\gamma\gamma}(\boldsymbol{\delta}_0)$, and

$$\boldsymbol{c}'\boldsymbol{X}_\gamma{}'(\mathbb{I}_m - \boldsymbol{D}_\theta)\boldsymbol{X}_\gamma\boldsymbol{c} \overset{a}{=} \boldsymbol{c}'\boldsymbol{\Sigma}(\boldsymbol{\delta}_0)\boldsymbol{c}. \tag{3.7}$$

It follows from (3.5) and (3.6) that when both m and r are fixed, but these are moderately large, so that (3.7) holds and $m - p - q = r$, then the power of the portmanteau tests compares favourably with that of $LM(r)$, despite a lack of concrete parametric alternative models. In other words, when the alternative models assumed by $LM(r)$ comprise true models, the portmanteau tests based on a χ^2_{m-p-q} approximation and the LM tests based on $LM(r) \sim \chi^2_r$ lose power in equal measure as m, r increases.

3.2 The empirical power of the portmanteau and LM tests

We conducted empirical experiments to examine the finite sample performance of the power of the portmanteau and LM tests for testing the AR(1) model. The data were generated from the following ARMA$(1,2)$ models:

$$(1 - \alpha^0 L)x_t = (1 + \beta_1^0 L + \beta_2^0 L^2)\varepsilon_t, \qquad t = 1, 2, \ldots, n, \tag{3.8}$$

where $\{\varepsilon_t\}$ is iid $N(0,1)$, $\alpha^0 = 0.7, 0.9$, $(\beta_1^0, \beta_2^0) = (0,0), (\pm 0.2, 0), (\pm 0.4, 0), (-0.2, -0.4), (0.2, 0.4)$, $n = 50, 100$. The test statistics are Q_m^* and Q_m^{**}, which are reported in Table 1, and Godfrey's $LM(r)$ statistic, which is used to test the AR(1) model against an alternative AR$(1+r)$ model. Using this test is equivalent to testing the AR(1) model against an ARMA$(1, r)$ model (see, e.g., Godfrey, 1979, and Poskitt & Tremayne, 1980). Figures 2, 3 and 4 illustrate typical percentage rejection rates for these test statistics, with $r = m - 1 = 1, 2, \ldots, 10$, and are based on 10,000 replications.

Figures 2 and 3 show the empirical significance levels and powers of $Q_m^* \sim \chi^2_{m-1}$ (the classical portmanteau test, labelled P), $Q_m^* \sim a\chi^2_b$ (Ljung's portmanteau test, labelled L), $Q_m^{**} \sim \chi^2_{m-1}$ (the bias-corrected portmanteau test, labelled B), and $LM(r) \sim \chi^2_r$ (Godfrey's LM test, labelled G). Note that the empirical significance level of $Q_m^* \sim a\chi^2_b$ is omitted from Figure 2 because it is almost identical to that of $Q_m^{**} \sim \chi^2_{m-1}$. Figure 2 reveals that the empirical significance level of $Q_m^* \sim \chi^2_{m-1}$ is high when m is small, as was discussed in Section 2. In addition, the empirical significance level of $LM(r)$ decreases as r increases, which is consistent with the empirical study of Kwan (1993). Hence, Figure 3 reveals that, as $r = m-1$ increases, the LM test based on $LM(r)$ has less power than expected, whereas the three portmanteau tests have similar powers, which decline slowly.

For this reason, we also empirically analysed the size-adjusted power of the portmanteau and LM tests. Specifically, critical points were estimated from empirical distributions consisting of 10,000 values of each test statistic computed under the null hypothesis, which specifies an AR(1) model, by using the estimated AR(1) parameter values. Power was then estimated, for each ARMA$(1,2)$ model, from 10,000 values of each test statistic. Figure 4 illustrates the size-adjusted power of the portmanteau and LM tests based on the ARMA$(1,2)$ models given in Figure 3. Figure 4 reveals that the portmanteau tests are typically more powerful when m is small and are approximately as powerful as the $LM(r)$ test when $r = m - 1$ is large and the ARMA$(1, r)$ is the correct model. These results are consistent with those of Ljung (1986, Section 4) and with our theoretical results discussed in this section.

Overall, the portmanteau tests based on $Q_m^{**} \sim \chi^2_{m-p-q}$ and $Q_m^* \sim a\chi^2_b$ for small m (e.g., for $m = 5, 10$) are better than the portmanteau tests based on $Q_m^* \sim \chi^2_{m-p-q}$ for large m (e.g., for m greater than 20), in terms of their significance levels and power, both theoretically and empirically.

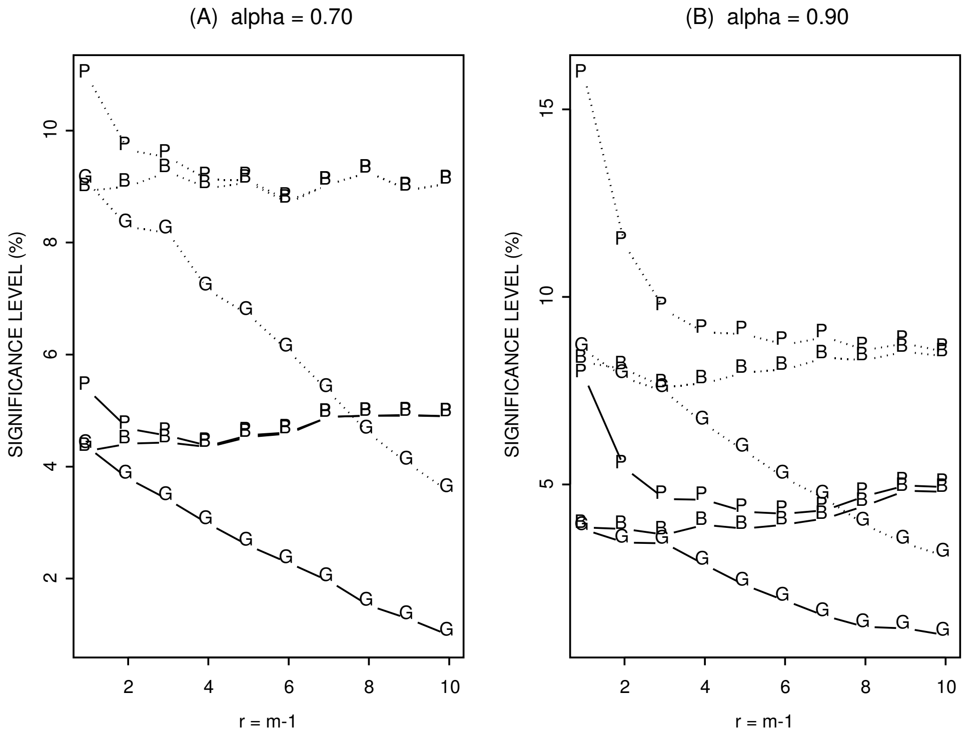

Figure 2: Empirical significance level of the test statistics (with P denoting $Q_m^* \sim \chi^2_{m-1}$, B denoting $Q_m^{**} \sim \chi^2_{m-1}$, and G denoting $LM(r) \sim \chi^2_r$) for the AR(1) model, $(1-\alpha^0 L)x_t = \varepsilon_t$: $n = 50$, and alpha denotes $\alpha^0 = 0.7, 0.9$. The significance levels are 5% (solid line) and 10% (dashed line).

ACKNOWLEDGEMENTS

The author wishes to thank the editors, the referees, and Assistant Professor Eiji Kurozumi for their valuable comments and suggestions. This work was supported by the Japan Society for the Promotion of Science.

APPENDIX

Proof of $\boldsymbol{X} = (\boldsymbol{d}_1(\boldsymbol{\theta}_0), \ldots, \boldsymbol{d}_m(\boldsymbol{\theta}_0))'$

Let $\widehat{r}_k$ be $\widehat{r}_k = r_k(\widehat{\boldsymbol{\theta}}_n)$. Similarly to (32) and (33) of McLeod (1978), we obtain:

$$\begin{aligned}\frac{\partial r_k(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} &= \frac{1}{n\sigma_0^2}\sum_{t=k+1}^{n} \varepsilon_t^{(1)}(\boldsymbol{\theta}_0)\varepsilon_{t-k}(\boldsymbol{\theta}_0) + O_p\left(\frac{1}{\sqrt{n}}\right) \qquad \text{(A.1)}\\ &= \frac{1}{\sigma_0^2}\,\mathrm{E}\left[\sum_{i=1}^{\infty} \boldsymbol{d}_i(\boldsymbol{\theta}_0)\varepsilon_{t-i}\varepsilon_{t-k}\right] + O_p\left(\frac{1}{\sqrt{n}}\right)\\ &= \boldsymbol{d}_k(\boldsymbol{\theta}_0) + O_p\big(1/\sqrt{n}\big),\end{aligned}$$

for $k = 1, 2, \ldots, m$. □

We present the following lemma, which is required to prove that $\mathrm{rank}(\boldsymbol{X}) = p + q$.

Lemma A.1 *The matrix* $\boldsymbol{X}$ *is* $\mathrm{rank}(\boldsymbol{X}) = p + q$ *if and only if a stationary and invertible ARMA(p,q) model is not redundant.*

Proof. An outline of the proof is given by McLeod (1999, Theorem 1), who establishes that a necessary and sufficient condition for the Fisher information matrix of a stationary and invertible ARMA model to be non-singular is that the model is not redundant.

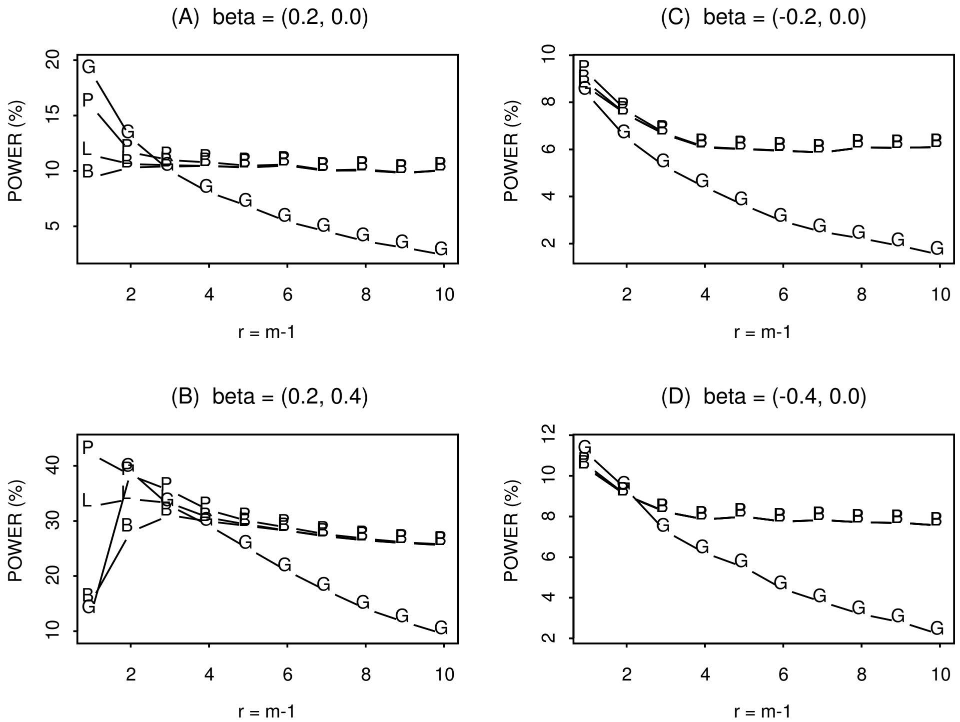

Figure 3: Empirical power based on asymptotic critical values of the test statistics (with P denoting $Q_m^* \sim \chi^2_{m-1}$, L denoting $Q_m^* \sim a\chi^2_b$, B denoting $Q_m^{**} \sim \chi^2_{m-1}$, and G denoting $LM(r) \sim \chi^2_r$) after an AR(1) model was fitted to data generated from ARMA(1, 2) models given by (3.8): $n = 50$, $\alpha^0 = 0.7$, and beta denotes (β_1^0, β_2^0). The significance level is 5%.

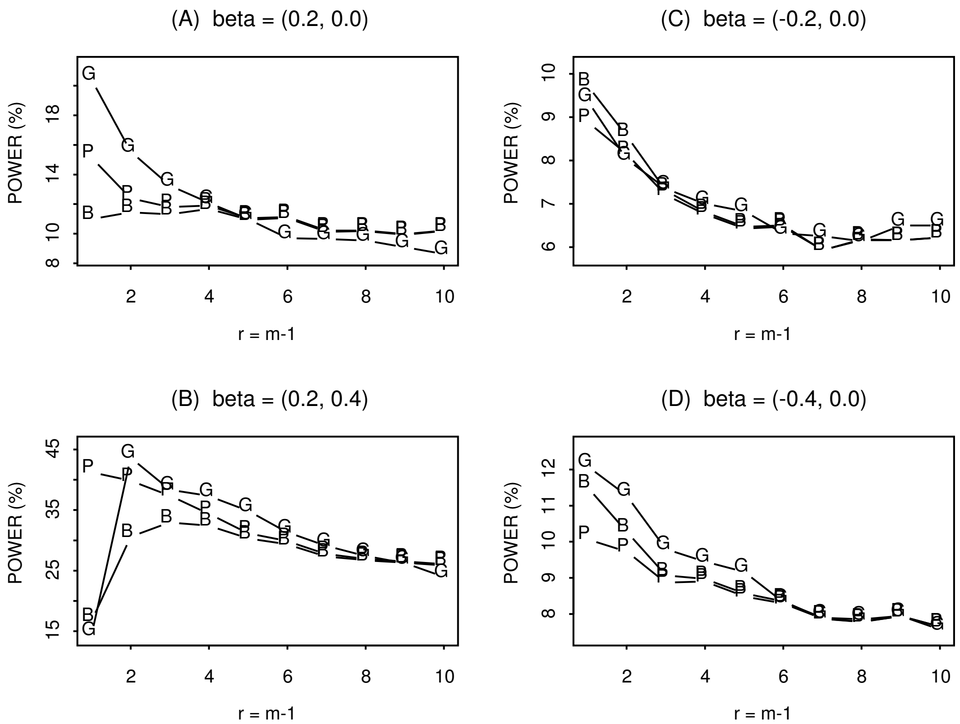

Figure 4: Size-adjusted power of the test statistics (with P denoting Q_m^*, B denoting Q_m^{**}, and G denoting $LM(r)$) after an AR(1) model was fitted to data generated from the ARMA(1, 2) models presented in Figure 3. beta denotes (β_1^0, β_2^0). The significance level is 5%.

Let $\boldsymbol{X} = (\boldsymbol{A}\ \boldsymbol{B})$, $\boldsymbol{A} = (-\alpha^{0*}_{i-j})$ and $\boldsymbol{B} = (-\beta^{0*}_{i-j})$. Then:

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} \boldsymbol{A}'\boldsymbol{A} & \boldsymbol{A}'\boldsymbol{B} \\ \boldsymbol{B}'\boldsymbol{A} & \boldsymbol{B}'\boldsymbol{B} \end{pmatrix} = \mathrm{E}\left[\boldsymbol{w}\boldsymbol{w}'\right],$$

where $\boldsymbol{w} = (u_m, \dots, u_{m-p+1}, v_m, \dots, v_{m-q+1})'$, $\alpha(L;\boldsymbol{\alpha}^0)u_t = -z_t$, $\beta(L;\boldsymbol{\beta}^0)v_t = -z_t$, $\{z_t\}_{t=1}^{\infty} \sim$ iid $(0,1)$, $z_t = 0$ for $t \le 0$. Thus, $u_t = -\sum_{i=0}^{t-1} \alpha_i^{0*} z_{t-i}$ and $v_t = -\sum_{i=0}^{t-1} \beta_i^{0*} z_{t-i}$ are approximate versions of the AR(p) and AR(q) models, respectively, truncated at $t = 0$.

If $p = 0$ or $q = 0$, the model is not redundant, and $\mathrm{rank}(\boldsymbol{X}) = p + q$ because $\mathrm{rank}(\boldsymbol{A}) = p$ and $\mathrm{rank}(\boldsymbol{B}) = q$.

When $p, q > 0$, $\boldsymbol{X}'\boldsymbol{X}$ is non-negative definite and it is singular if and only if $\boldsymbol{l}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{l} = \mathrm{E}[\boldsymbol{l}'\boldsymbol{w}]^2 = 0$ for some $p+q$-vector $\boldsymbol{l} \neq \boldsymbol{0}_{p+q}$. Let $\boldsymbol{l} = (a_0, \dots, a_{p-1}, b_0, \dots, b_{q-1})'$ and note that, because $\boldsymbol{A}'\boldsymbol{A}$ and $\boldsymbol{B}'\boldsymbol{B}$ are positive definite, we must have $a_i \neq 0$ for some i as well as $b_j \neq 0$ for some j. It follows that:

$$\boldsymbol{l}'\boldsymbol{w} = a(L)u_m + b(L)v_m = 0, \tag{A.2}$$

where $a(L) = \sum_{i=0}^{p-1} a_i L^i$ and $b(L) = \sum_{i=0}^{q-1} b_i L^i$. Equation (A.2) can now be written as: $-\{a(L)/\alpha(L;\boldsymbol{\alpha}_0)\}z_m - \{b(L)/\beta(L;\boldsymbol{\beta}_0)\}z_m = 0$ and

$$\frac{\beta(L;\boldsymbol{\beta}_0)}{\alpha(L;\boldsymbol{\alpha}_0)} z_m = -\frac{b(L)}{a(L)} z_m. \tag{A.3}$$

Because the maximum degrees of the non-zero polynomials $a(L)$ and $b(L)$ are $p-1$ and $q-1$, respectively, it follows that the polynomials $\alpha(z;\boldsymbol{\alpha}_0)$ and $\beta(z;\boldsymbol{\beta}_0)$ contain a common factor and thus the model is redundant. The proof of the converse proposition is obtained by the same reasoning as that used by McLeod (1999) to prove Theorem 1. Hence, the proof is omitted. □

Proof of (2.7)***.***

Because:

$$\sqrt{\frac{n(n+2)}{n-i}} - \sqrt{n} = \sqrt{n}\left\{\frac{1+2/n}{1-i/n} - 1\right\}, \quad i = 1, 2, \dots, m,$$

(2.7) follows from the Taylor-series expansion of $f(x,y) = \sqrt{(1+y)/(1-x)}$ around $(x,y) = (0,0)$. □

We now prove that the following lemma is needed to prove (2.15).

Lemma A.2 *Let $X_{i,t}$ and $Y_{i,t}$ be random variables such that $\mathrm{E}[X_{i,t}]^2 < \infty$ and $\mathrm{E}[Y_{i,t}]^2 < C_1\epsilon^{2t}$ for $i = 1, 2$, $t = 1, 2 \dots, n$, $C_1 > 0$, and $\epsilon \in (0,1)$. Let $\{a_k\}$ be positive sequences, such that $a_k < C_2\epsilon^k$ for $k > 1$ and $C_2 > 0$, and let $m = m_n$ be a positive integer that satisfies Assumption 1. Then, as $n \to \infty$, $\sum_{t=m+2}^{n}\sum_{k=m+1}^{t-1} a_k X_{1,t-k}X_{2,t} = O_p(n^{1-\nu})$, $\sum_{t=m+2}^{n}\sum_{k=m+1}^{t-1} a_k X_{1,t-k}Y_{2,t} = O_p(n^{-2\nu})$, $\sum_{t=m+2}^{n}\sum_{k=m+1}^{t-1} a_k Y_{1,t-k}X_{2,t} = O_p(\log n n^{-\nu})$, and $\sum_{t=m+2}^{n}\sum_{k=m+1}^{t-1} a_k Y_{1,t-k}Y_{2,t} = O_p(\log n n^{-2\nu})$.*

Proof. Using $\sum_{k=m+1}^{\infty} a_k = O(\epsilon^{m+1}) = O(n^{-\nu})$ and the Cauchy–Schwarz inequality yields the above result. □

Proof of Theorem 2.1.

To show (2.15), it is sufficient to show that:

$$\sqrt{n}\boldsymbol{X}'\widehat{\boldsymbol{r}}_n = O_p\Big(n^{-\min(\nu,0.5)}\Big) \quad \text{and} \quad \boldsymbol{X}'\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n = O_p\Big(n^{-\min(\nu,0.5)}\Big). \tag{A.4}$$

Let $S_n(\boldsymbol{\theta}) = \{2n\sigma_0^2\}^{-1}\sum_{t=1}^{n} e_t(\boldsymbol{\theta})^2$, $\boldsymbol{a}_n(\widehat{\boldsymbol{\theta}}_n) = \sum_{k=1}^{n-1}\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{r}_k$, $\boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n) = \sum_{k=m+1}^{n-1}\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{r}_k$, $\boldsymbol{c}_{m,n}(\widehat{\boldsymbol{\theta}}_n) = \sum_{k=1}^{m}\partial\boldsymbol{d}_k(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0'\widehat{r}_k\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, and $\widehat{\sigma}_n^2 = n^{-1}\sum_{t=1}^{n}\widehat{e}_t^2$. Then:

$$\boldsymbol{X}'\widehat{\boldsymbol{r}}_n = \sum_{k=1}^{m}\boldsymbol{d}_k(\boldsymbol{\theta}_0)\widehat{r}_k = \boldsymbol{a}_n(\widehat{\boldsymbol{\theta}}_n) - \boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n) - \boldsymbol{c}_{m,n}(\widehat{\boldsymbol{\theta}}_n)/\sqrt{n} + O_p\Big(n^{-3/2}\Big),$$

as $n \to \infty$, where we have used the fact that $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_p(1/\sqrt{n})$, $\widehat{r}_k = O_p(1/\sqrt{n})$, and

$$\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{d}_k(\boldsymbol{\theta}_0) + \frac{\partial\boldsymbol{d}_k(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0'}\Big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\Big) + O_p\big(n^{-1}\big), \qquad k = 1, 2, \dots, m. \tag{A.5}$$

It follows from $\sum_{j=1}^{t-1} \pi_j^{(1)}(\widehat{\boldsymbol{\theta}}_n)x_{t-j} = \sum_{j=1}^{t-1} \boldsymbol{d}_j(\widehat{\boldsymbol{\theta}}_n)e_{t-j}(\widehat{\boldsymbol{\theta}}_n)$ that:

$$\boldsymbol{a}_n(\widehat{\boldsymbol{\theta}}_n) = \sigma_0^2 S_n^{(1)}(\widehat{\boldsymbol{\theta}}_n)/\widehat{\sigma}_n^2 = \mathbf{0}_{p+q}, \qquad \boldsymbol{c}_{m,n}(\widehat{\boldsymbol{\theta}}_n) = O_p\big(1/\sqrt{n}\big) \tag{A.6}$$

and

$$\sqrt{n}\boldsymbol{X}'\widehat{\boldsymbol{r}}_n = -\sqrt{n}\boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n) + O_p(1/\sqrt{n}). \tag{A.7}$$

Therefore, to show that (A.4) follows from (2.7), we must show that $\boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n) = O_p(n^{-\nu-0.5})$. From the Taylor-series expansion of $\widehat{e}_t = e_t(\widehat{\boldsymbol{\theta}}_n)$ around $\widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$, we obtain:

$$e_t(\widehat{\boldsymbol{\theta}}_n) = \varepsilon_t + \varepsilon_t^{(1)}(\boldsymbol{\theta}_0)'\Big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\Big) - w_t(\widehat{\boldsymbol{\theta}}_n) + O_p\big(n^{-1}\big), \qquad t = 1, 2, \ldots, n, \tag{A.8}$$

where $w_t(\widehat{\boldsymbol{\theta}}_n) = \sum_{j=t}^{\infty} \pi_j(\widehat{\boldsymbol{\theta}}_n)x_{t-j}$. Using (A.5), (A.8), (2.13), (2.14), and Lemma A.2, we obtain:

$$\boldsymbol{b}_{m,n}(\widehat{\boldsymbol{\theta}}_n) = \frac{1}{n\widehat{\sigma}_n^2} \sum_{t=m+2}^{n} \left\{ \sum_{k=m+1}^{t-1} \boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)\widehat{e}_{t-k} \right\} \widehat{e}_t = O_p\big(n^{-\nu-0.5}\big). \tag{A.9}$$

Hence, (A.4) follows from (A.9). □

Proof of Theorem 2.2.

From (2.8), we obtain:

$$\begin{aligned} (\mathbb{I}_m - \widehat{\boldsymbol{D}})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n &= (\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n + (\boldsymbol{D} - \widehat{\boldsymbol{D}})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n \\ &= (\mathbb{I}_m - \boldsymbol{D})\boldsymbol{T}_n\boldsymbol{r}_n + (\boldsymbol{D} - \widehat{\boldsymbol{D}})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n + O_p\big(1/\sqrt{n}\big). \end{aligned} \tag{A.10}$$

Therefore, it is sufficient to show that $(\boldsymbol{D} - \widehat{\boldsymbol{D}})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n = O_p(1/\sqrt{n})$. From (A.5) and (A.6), we have:

$$\begin{aligned} (\boldsymbol{D} - \widehat{\boldsymbol{D}})\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big(\boldsymbol{X} - \widehat{\boldsymbol{X}}\Big)'\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n + O_p\big(1/\sqrt{n}\big) \\ &= -\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}_{m,n}(\widehat{\boldsymbol{\theta}}_n) + O_p\big(1/\sqrt{n}\big) \\ &= O_p\big(1/\sqrt{n}\big), \end{aligned} \tag{A.11}$$

which establishes (2.19). □

Proof of Theorem 3.1.

First, we consider the limiting distribution of $\widehat{\boldsymbol{\theta}}_n$. Because $\widehat{\boldsymbol{\theta}}_n$ minimizes $S_n(\boldsymbol{\theta}) = \{2n\sigma_0^2\}^{-1}\sum_{t=1}^{n} e_t(\boldsymbol{\theta})^2$, as $n \to \infty$, we have:

$$\begin{aligned} &n^{-1/2}S_n^{(1)}(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}_{p+q} \overset{\text{a}}{=} n^{-1/2}S_n^{(1)}(\boldsymbol{\theta}_0) + \boldsymbol{I}_{\theta\theta}(\boldsymbol{\delta}_0)\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\Big) - \boldsymbol{I}_{\theta\gamma}(\boldsymbol{\delta}_0)\boldsymbol{c}, \\ &\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\Big) \overset{\text{a}}{\sim} N\big(\boldsymbol{I}_{\theta\theta}(\boldsymbol{\delta}_0)^{-1}\boldsymbol{I}_{\theta\gamma}(\boldsymbol{\delta}_0)\boldsymbol{c},\ \boldsymbol{I}_{\theta\theta}(\boldsymbol{\delta}_0)^{-1}\big), \end{aligned} \tag{A.12}$$

and, hence, $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_p(1/\sqrt{n})$. Because $\boldsymbol{d}_i(\boldsymbol{\theta}_0) = \boldsymbol{d}_{\theta,i}((\boldsymbol{\theta}_0', \mathbf{0}_r')')$, using a Taylor-series expansion yields, as $n \to \infty$:

$$\begin{aligned} &\boldsymbol{d}_i(\boldsymbol{\theta}_0) = \boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0) + \frac{\partial \boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\gamma}_0'}(\mathbf{0}_r - \boldsymbol{\gamma}_0) + O\big(n^{-1}\big), \qquad i = 1, 2, \ldots, m, \\ &\boldsymbol{X} = \boldsymbol{X}_\theta + O\big(1/\sqrt{n}\big) \quad \text{and} \quad \boldsymbol{D} = \boldsymbol{D}_\theta + O\big(1/\sqrt{n}\big). \end{aligned}$$

Using $\boldsymbol{D} = \boldsymbol{D}_\theta + O(1/\sqrt{n})$, (2.7), and the argument used in the context of (A.1), it follows that, as $n \to \infty$:

$$\boldsymbol{T}_n\widehat{\boldsymbol{r}}_n = \boldsymbol{T}_n\boldsymbol{r}_n + \boldsymbol{X}_\theta\sqrt{n}\Big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\Big) - \boldsymbol{X}_\gamma\boldsymbol{c} + O_p\big(1/\sqrt{n}\big) \tag{A.13}$$

and (3.3) holds.

From equations (A.12) and (A.13), it follows that $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_p(1/\sqrt{n})$ and $\widehat{\boldsymbol{r}}_n = O_p(1/\sqrt{n})$. Therefore, the proof of (3.4) follows trivially from (A.10) and (A.11). □

Proof of Theorem 3.2.

We use the notation already defined in the proof of Theorem 2.1. Using a Taylor-series expansion of $\boldsymbol{d}_k(\widehat{\boldsymbol{\theta}}_n)$ and $\widehat{e}_t = e_t(\widehat{\boldsymbol{\theta}}_n)$, we obtain (A.5) and the following expression:

$$e_t(\widehat{\boldsymbol{\theta}}_n) = \varepsilon_t + \sum_{i=1}^{\infty} \boldsymbol{d}_{\theta,i}(\boldsymbol{\delta}_0)'\varepsilon_{t-i}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) - \sum_{i=1}^{\infty} \boldsymbol{d}_{\gamma,i}(\boldsymbol{\delta}_0)'\varepsilon_{t-i}\frac{\boldsymbol{c}}{\sqrt{n}} - w_t(\widehat{\boldsymbol{\theta}}_n) + O_p\left(n^{-1}\right), \qquad \text{(A.14)}$$

for $t = 1, 2, \ldots, n$. Therefore, the argument already used in the proof of (A.4), (A.12), and (A.13) yields (A.7). Because $\widehat{\sigma}_n^2 \stackrel{\text{a}}{=} \sigma_0^2$, the remainder of the proof is obvious from (A.14) and the proof of (A.9). Hence, this proof is omitted. □

REFERENCES

BOX, G. E. P. & JENKINS, G. M. (1976) *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

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.

GODFREY, L. G. (1979) Testing the adequacy of a time series model. *Biometrika* **66**, 67–72.

KOHN, R. (1979) Asymptotic estimation and hypothesis testing results for vector linear time series models. *Econometrica* **47**, 1005–30.

KWAN, A.C.C. (1993) A note on the finite-sample distribution of Lagrange multiplier tests for univariate time series models. *Communications in Statistics. Simulation and Computation* **22(4)**, 1135–1160.

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.

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

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

MCLEOD, A. I. (1999) Necessary and sufficient condition for non-singular Fisher information matrix in ARMA and Fractional ARIMA models. *The American Statistician* **53**, 71–2.

POSKITT, D. S. & TREMAYNE, A. R. (1980) Testing the specification of a fitted autoregressive-moving average model. *Biometrika* **67**, 359–63.

POSKITT, D. S. & TREMAYNE, A. R. (1982) Diagnostic tests for multiple time series models. *Annals of Statistics* **10**, 114–20.