On Multiple Portmanteau Tests

Katayama, Naoya
Department of Economic Engineering, Faculty of Economics, Kyushu University

KYUSHU UNIVERSITY

ON MULTIPLE PORTMANTEAU TESTS [*]

BY NAOYA KATAYAMA

Faculty of Economics, Kyushu University, 6-19-1, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.

kata@kyudai.jp

Abstract. The portmanteau statistic based on the first m residual autocorrelations is used for diagnostic checks on the adequacy of fitting a model with varying m. In this paper, we propose an approximation of the joint probability of multiple portmanteau tests with different degrees of freedom (DF). This distribution is easy to compute when all DF are even integers; its empirical behavior is clarified in terms of asymptotic theory.

Keywords. Autoregressive-moving average model; Portmanteau test; Residual autocorrelations; Goodness-of-fit; Multiple tests.

1 INTRODUCTION

This paper provides a practical application of portmanteau tests. In time series analysis, the portmanteau test statistic is known as a goodness-of-fit test statistic and is defined by the sum of squares of the first m residual autocorrelations, $\hat{r}_i$, $i = 1, 2, \ldots, m$. First presented by Box & Pierce (1970), this statistic is supposed to behave as the sum of squares of the first m sample autocorrelations of the white noise process: r_i, $i = 1, 2, \ldots, m$. Li (2004) reviews applications of portmanteau statistics in various time series models.

From a practical standpoint, portmanteau tests are conducted by various values of m because we are unable to determine the optimal number of m for the following reasons: (i) The number of m needs to be moderately large in order to conduct chi-squared tests. (ii) For diagnosis of the residual process, we need to check many $\hat{r}_i$s, meaning m has to be large. (iii) Conversely, if the number of m is unnecessarily large, the size of this test is unstable and its power will decrease. For example, in the S-PLUS statistical software, the S-PLUS function `arima.diag` computes portmanteau statistics for an autoregressive integrated moving average (ARIMA) model and plots the p-values of the portmanteau tests for any possible value of m. To overcome (i), Ljung (1986) and Katayama (2008) propose modifications

---

[*]Discussion paper

of the test. Therefore, in practical analysis we can conduct multiple portmanteau tests for small to medium values of m regarding (ii) and (iii) (e.g., $m < 20$). However, there has been no attempt to establish the joint probability of the multiple tests.

For this reason we provide an approximated joint distribution of the portmanteau test statistics in Section 2. In addition, we review the recursive formula in Hosoya (1986) to compute the joint probability including the significance levels and p-values. This formula is useful when the degrees of freedom (DF) are even integers. The empirical performance in finite samples suggest that the approximation of the joint significance level of multiple portmanteau tests performs well when we conduct tests for small to medium even integer values of DF.

2 MAIN RESULTS

2.1 Univariate chi-squared approximation of the portmanteau statistic

First, we review the basic framework found in previous work. Suppose that a univariate time series $\{x_t\}$ is generated by an autoregressive-moving average, ARMA(p, q), model:

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

where $p + q > 0$, $L^i x_t = x_{t-i}$, $\{\varepsilon_t\}$ is iid $(0, \sigma_0^2)$, $\alpha(L; \boldsymbol{\alpha}_0) = 1 - \alpha_1^0 L - \cdots - \alpha_p^0 L^p$, $\beta(L; \boldsymbol{\beta}_0) = 1 + \beta_1^0 L + \cdots + \beta_q^0 L^q$, $\boldsymbol{\alpha}_0 = (\alpha_1^0, \ldots, \alpha_p^0)'$, and $\boldsymbol{\beta}_0 = (\beta_1^0, \ldots, \beta_q^0)'$. It is assumed that this 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), the nonlinear least squares estimator of $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0')'$, $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}', \widehat{\boldsymbol{\beta}}')'$, is obtained by minimizing the sum of squared residuals. Box & Jenkins (1976, Chapter 8) point out that it is important to check the assumption of the independence of $\{\varepsilon_t\}$ by using the residuals, $\widehat{e}_t = e_t(\widehat{\boldsymbol{\theta}})$, for $t = 1, \ldots, n$, which should behave in a manner that is consistent with the model. It follows that it is reasonable to check the adequacy of model fit by examining the residual autocorrelations as follows:

$$\widehat{r}_j = \widehat{r}_j(\{\widehat{e}_t\}) = \sum_{t=j+1}^n \widehat{e}_t \widehat{e}_{t-j} \left/ \sum_{t=1}^n \widehat{e}_t^2 \right., \qquad j = 1, 2, \ldots, n - 1. \tag{2}$$

The asymptotic joint distribution of $\widehat{\boldsymbol{r}}_m = (\widehat{r}_1, \ldots, \widehat{r}_m)'$ has been analyzed by Box & Pierce (1970) and McLeod (1978). Following the proof in McLeod (1978, equation (34)), by using a Taylor-series expansion of $\widehat{\boldsymbol{r}}_m$ around $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$, as $n \to \infty$, we have:

$$\widehat{\boldsymbol{r}}_m = \boldsymbol{r}_m + \boldsymbol{X}_m\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + O_p(1/n), \tag{3}$$

2

where $\boldsymbol{r}_m = (r_1, \ldots, r_m)'$, $r_j = \hat{r}_j(\{\varepsilon_t\})$, $\boldsymbol{X}_m = \boldsymbol{X}_m(\boldsymbol{\theta}_0)$ is a $m \times (p+q)$ matrix. Here each (i, j) element of the partitioned matrix of $\boldsymbol{X}_m$ is given in (16) of McLeod (1978), as follows:

$$\boldsymbol{X}_m = [\boldsymbol{d}_1(\boldsymbol{\theta}_0), \ldots, \boldsymbol{d}_m(\boldsymbol{\theta}_0)]' = \left[\begin{array}{c:c} -\alpha_{i-j}^{0*} & -\beta_{i-j}^{0*} \end{array} \right] \tag{4}$$

where $\boldsymbol{d}_i(\boldsymbol{\theta}_0)$ is $(p+q)$-vector consists of the ith row of $\boldsymbol{X}_m$ and α_i^{0*}s and β_i^{0*}s are defined by $1/\alpha(L; \boldsymbol{\alpha}_0) = \alpha_0^{0*} + \alpha_1^{0*}L + \alpha_2^{0*}L^2 + \cdots$, $1/\beta(L; \boldsymbol{\beta}_0) = \beta_0^{0*} + \beta_1^{0*}L + \beta_2^{0*}L^2 + \cdots$ and $\alpha_i^{0*} = \beta_i^{0*} = 0$ for $i < 0$. Note that

$$\boldsymbol{I}(\boldsymbol{\theta}_0) = \sum_{i=1}^{\infty} \boldsymbol{d}_i(\boldsymbol{\theta}_0)\boldsymbol{d}_i(\boldsymbol{\theta}_0)'$$

is the Fisher information matrix of the model (1), and $\sqrt{n}((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\ \boldsymbol{r}_m')'$ has an approximately normal distribution with a mean vector of $\boldsymbol{0}_m$ and the following covariance matrix:

$$\left[\begin{array}{cc} \boldsymbol{I}(\boldsymbol{\theta}_0)^{-1} & -\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{X}_m' \\ -\boldsymbol{X}_m\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1} & \mathbb{I}_m \end{array} \right], \tag{5}$$

from (28) of McLeod (1978), where $\boldsymbol{0}_m$ is the m-dimensional zero vector and $\mathbb{I}_m$ is an $m \times m$ identity matrix. It follows that $\hat{\boldsymbol{r}}_m$ is approximately normal with a mean vector of $\boldsymbol{0}_m$, and a covariance matrix of $(\mathbb{I}_m - \boldsymbol{C}_m)/n$ with rank $m - p - q$. Since $\lim_{m\to\infty} \boldsymbol{X}_m'\boldsymbol{X}_m = \boldsymbol{I}(\boldsymbol{\theta}_0)$, $\boldsymbol{C}_m = \boldsymbol{X}_m\boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{X}_m'$ is an idempotent matrix for moderately large m. For this, the portmanteau statistic:

$$Q_m = n\sum_{j=1}^{m} \hat{r}_j^2, \tag{6}$$

is distributed as χ_{m-p-q}^2, where χ_m^2 denotes a chi-squared variable with m DF. Ljung & Box (1978) showed that a better approximation could be achieved using the following modified portmanteau statistic:

$$Q_m^* = n(n+2)\sum_{j=1}^{m} \hat{r}_j^2/(n-j). \tag{7}$$

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 references therein.)

Ljung (1986) and Katayama (2008) propose two solutions for conducting chi-squared test with small values of m. As an approximate distribution of Q_m^* in place of χ_{m-p-q}^2, Ljung (1986) suggested an alternative chi-squared distribution, $a\chi_b^2$, where a and b generally depend on $\boldsymbol{\theta}_0$ and positive real values. Alternatively, Katayama (2008) proposed a modified Q_m^* with a correction term. Katayama (2008)

3

rewrote (3) as:

$$(\mathbb{I}_m - \boldsymbol{D}_m)\boldsymbol{T}_m\widehat{\boldsymbol{r}}_m = (\mathbb{I}_m - \boldsymbol{D}_m)\boldsymbol{T}_m\boldsymbol{r}_m + O_p\big(1/\sqrt{n}\big), \tag{8}$$

as $n \to \infty$, where $\boldsymbol{D}_m = \boldsymbol{D}_m(\boldsymbol{\theta}_0) = \boldsymbol{X}_m(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-1}\boldsymbol{X}'_m$ and

$$\boldsymbol{T}_m = \mathrm{diag}\left(\sqrt{\frac{n(n+2)}{n-1}}, \sqrt{\frac{n(n+2)}{n-2}}, \ldots, \sqrt{\frac{n(n+2)}{n-m}}\right) \tag{9}$$

$$= \sqrt{n}\mathbb{I}_m + O\big(m/\sqrt{n}\big).$$

Since $Q^*_m = \widehat{\boldsymbol{r}}'_m\boldsymbol{T}^2_m\widehat{\boldsymbol{r}}_m$ and $\boldsymbol{r}'_m\boldsymbol{T}_m(\mathbb{I}_m - \boldsymbol{D}_m)\boldsymbol{T}_m\boldsymbol{r}_m \xrightarrow{d} \chi^2_{m-p-q}$, we have:

$$Q^*_m - B^*_m \xrightarrow{d} \chi^2_{m-p-q}, \quad \text{as } n \to \infty, \tag{10}$$

where $B^*_m = \widehat{\boldsymbol{r}}'_m\boldsymbol{T}_m\boldsymbol{D}_m\boldsymbol{T}_m\widehat{\boldsymbol{r}}_m$. Therefore, Q^*_m has a positive extra random variable B^*_m for χ^2_{m-p-q} approximation. Katayama (2008, Theorem 1) proved that B^*_m is negligible as m, n increases. In addition, Katayama (2008, Theorem 2) proved that:

$$(\mathbb{I}_m - \widehat{\boldsymbol{D}}_m)\boldsymbol{T}_m\widehat{\boldsymbol{r}}_m = (\mathbb{I}_m - \boldsymbol{D}_m)\boldsymbol{T}_m\boldsymbol{r}_m + O_p\big(1/\sqrt{n}\big), \tag{11}$$

as $n \to \infty$, where $\widehat{\boldsymbol{D}}_m$ is an consistent estimator of $\boldsymbol{D}_m$ given by $\widehat{\boldsymbol{D}}_m = \boldsymbol{D}_m(\widehat{\boldsymbol{\theta}})$. Hence, Katayama (2008) proposed a modified portmanteau statistic with a correction term:

$$Q^{**}_m = Q^*_m - \widehat{B}^*_m, \tag{12}$$

where $\widehat{B}^*_m = \widehat{\boldsymbol{r}}'_m\boldsymbol{T}_m\widehat{\boldsymbol{D}}_m\boldsymbol{T}_m\widehat{\boldsymbol{r}}_m$. Unlike Q^*_m, Q^{**}_m is not required for large m for the χ^2_{m-p-q} approximation.

2.2 Multivariate chi-squared approximation of the portmanteau statistics

This section provides the asymptotic joint distribution of the portmanteau statistics. Let $p + q < m = m(1) < m(2) < \cdots < m(s) = M < n$. Then, from (8) and (9), we obtain:

$$\begin{bmatrix} (\mathbb{I}_{m(i)} - \boldsymbol{D}_{m(i)})\boldsymbol{T}_{m(i)}\widehat{\boldsymbol{r}}_{m(i)} \\ \boldsymbol{0}_{M-m(i)} \end{bmatrix} = \sqrt{n}\boldsymbol{E}_{m(i)}\boldsymbol{r}_M + O_p\big(1/\sqrt{n}\big), \tag{13}$$

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

$$\boldsymbol{E}_{m(i)} = \begin{bmatrix} \mathbb{I}_{m(i)} - \boldsymbol{D}_{m(i)} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix},$$

and $\boldsymbol{O}$ denotes an appropriate dimensional matrix where all elements are zeros. It follows that:

$$\left(Q^*_{m(1)} - B^*_{m(1)}, \ldots, Q^*_{m(s)} - B^*_{m(s)} \right)' \tag{14}$$
$$= n \left(\boldsymbol{r}'_M \boldsymbol{E}_{m(1)} \boldsymbol{r}_M, \ldots, \boldsymbol{r}'_M \boldsymbol{E}_{m(s)} \boldsymbol{r}_M \right)' + O_p \left(1/\sqrt{n} \right)$$
$$\xrightarrow{d} \left(\boldsymbol{z}'_M \boldsymbol{E}_{m(1)} \boldsymbol{z}_M, \ldots, \boldsymbol{z}'_M \boldsymbol{E}_{m(s)} \boldsymbol{z}_M \right)', \quad \text{as } n \to \infty,$$

where $\boldsymbol{z}_M \sim N_M(\boldsymbol{0}_M, \mathbb{I}_M)$. The convergence of (14) follows from $\sqrt{n}\boldsymbol{r}_M \xrightarrow{d} \boldsymbol{z}_M$ and Cramér–Wold device. Similarly, using (9) and (11), yields:

$$\left(Q^{**}_{m(1)}, \ldots, Q^{**}_{m(s)} \right)' \xrightarrow{d} \left(\boldsymbol{z}'_M \boldsymbol{E}_{m(1)} \boldsymbol{z}_M, \ldots, \boldsymbol{z}'_M \boldsymbol{E}_{m(s)} \boldsymbol{z}_M \right)', \quad \text{as } n \to \infty. \tag{15}$$

2.3   Computation of the asymptotic joint distribution

This section examines the asymptotic joint probability given by (14) and (15). Unfortunately, computation of the distribution is difficult because all elements are correlated. To overcome this difficulty, we provide an independence property of the distribution when m is moderately large. The mathematical proofs are given in the Appendix.

Lemma 1 *For $i < j \leq k < l$, as $m(i) \to \infty$, $\boldsymbol{E}_{m(j)}(\boldsymbol{E}_{m(l)} - \boldsymbol{E}_{m(k)}) = O(\epsilon^{m(k)+1-\max(p,q)})$ and $(\boldsymbol{E}_{m(j)} - \boldsymbol{E}_{m(i)})(\boldsymbol{E}_{m(l)} - \boldsymbol{E}_{m(k)}) = O(\epsilon^{m(i)+1-\max(p,q)})$, where $\epsilon \in (0,1)$ is such that ϵ is larger than the absolute value of any roots of $\alpha(z^{-1}; \boldsymbol{\alpha}_0)\beta(z^{-1}; \boldsymbol{\beta}_0) = 0$.*

From the Craig–Sakamoto Theorem (see, e.g., Provost, 1996), Lemma 1 indicates the asymptotic independence of $\boldsymbol{z}'_M \boldsymbol{E}_{m(j)} \boldsymbol{z}_M$ and $\boldsymbol{z}'_M (\boldsymbol{E}_{m(l)} - \boldsymbol{E}_{m(k)}) \boldsymbol{z}_M$, or $\boldsymbol{z}'_M (\boldsymbol{E}_{m(j)} - \boldsymbol{E}_{m(i)}) \boldsymbol{z}_M$ and $\boldsymbol{z}'_M (\boldsymbol{E}_{m(l)} - \boldsymbol{E}_{m(k)}) \boldsymbol{z}_M$, as m increase.

Theorem 1 *For $j = 1, \ldots, s$, as $m \to \infty$:*

$$\boldsymbol{z}'_M \boldsymbol{E}_{m(j)} \boldsymbol{z}_M = \psi(j) + O_p(\epsilon^{m+1-\max(p,q)}), \tag{16}$$

where $\psi(j) = \sum_{i=p+q+1}^{m(j)} Z_i^2$ and $\{Z_i\}$ is iid $N(0,1)$.

Theorem 1 provides an approximation to evaluate the significance level $\alpha \in (0,1)$ of multiple portmanteau tests by (14) or (15) when m, n are moderately large. Suppose that each portmanteau test has the same

significance level $\beta \in (0, 1)$, we can approximate α and β by:

$$\Pr\left(\psi(j) \le d_j, \ j = 1, 2, \ldots, s\right) = 1 - \alpha, \tag{17}$$

where d_js are given by:

$$\Pr\left(\psi(j) \le d_j\right) = 1 - \beta, \quad j = 1, 2, \ldots, s. \tag{18}$$

We can easily compute (17) and (18) from Hosoya (1986, Section 2) if $m(1) - p - q, m(i) - m(i - 1)$, $i = 2, \ldots, s$ are all even integers. However, since Hosoya (1986, Section 2) contains a few confusing typographical errors, we modify it briefly here. All integrations are easily verified with integration by parts.

Let d_is be constants such that $0 \le d_1 \le \ldots \le d_s$ and let $q_i, i = 1, \ldots, s$ be nonnegative integers. Set for a nonnegative integer j:

$$A(k, j) = \int_0^{d_1} dz_1 \int_{z_1}^{d_2} dz_2 \cdots \int_{z_{k-1}}^{d_k} z_1^{q_1}(z_2 - z_1)^{q_2} \cdots \tag{19}$$

$$(z_{k-1} - z_{k-2})^{q_{k-1}}(z_k - z_{k-1})^j \exp(-z_k/2)dz_k, \quad k = 2, \ldots, s,$$

$$B(k, j) = \int_0^{d_1} dz_1 \int_{z_1}^{d_2} dz_2 \cdots \int_{z_{k-2}}^{d_{k-1}} z_1^{q_1}(z_2 - z_1)^{q_2} \cdots \tag{20}$$

$$(z_{k-1} - z_{k-2})^{q_{k-1}}(d_k - z_{k-1})^j dz_{k-1}, \quad k = 2, \ldots, s.$$

Then those integrals are computed recursively according to the following steps. Set $p_k = q_k + q_{k+1} + \cdots + q_s + s - k + 1$ and then:

$$B(2, j) = \sum_{i=0}^j \frac{q_1!}{(q_1 + i + 1)!} \frac{j!}{(j - i)!} d_1^{q_1 + i + 1}(d_2 - d_1)^{j-i}, \quad j = 0, \ldots, p_2, \tag{21}$$

$$B(k, j) = \sum_{i=0}^j \frac{q_{k-1}!}{(q_{k-1} + i + 1)!} \frac{j!}{(j - i)!}(d_k - d_{k-1})^{j-i} B(k - 1, q_{k-1} + i + 1),$$

$j = 0, \ldots, p_k, \ k = 3, \ldots, s$, and then:

$$A(1, 0) = \int_0^{d_1} \exp(-z_1/2)dz_1 = -2\exp(-d_1/2) + 2, \tag{22}$$

$$A(1, j) = \int_0^{d_1} z_1^j \exp(-z_1/2)dz_1 = -2d_1^j \exp(-d_1/2) + 2jA(1, j - 1), \quad j = 1, \ldots, p_1,$$

$$A(k, 0) = -2\exp(-d_k/2)B(k, 0) + 2A(k - 1, q_{k-1}), \quad k = 2, \ldots, s,$$

$$A(k, j) = -2\exp(-d_k/2)B(k, j) + 2jA(k, j - 1), \quad j = 1, \ldots, p_k, \ k = 2, \ldots, s.$$

Let $m(1) - p - q = 2(q_1 + 1), m(i) - m(i-1) = 2(q_i + 1)$, $i = 2, \ldots, k$. We then obtain:

$$\Pr\left(\psi(j) \leq d_j, j = 1, 2, \ldots, k\right) = A(k, q_k) \bigg/ \left(\prod_{j=1}^{k} 2^{q_j + 1} q_j!\right) \tag{23}$$

and $\quad \Pr\left(\psi(j) \leq d_j, j = 1, 2, \ldots, k - 1 \,|\, \psi(k) = d_k\right) = B(k, q_k)/B^*(k, q_k), \tag{24}$

$k = 2, \ldots, s$, where $B^*(k, j)$ is defined by (20) with $d_i = d_k$, $i = 1, \ldots, k$.

As Hosoya (1986) pointed out, when the number of DF is odd, the numerical computation of (17) and (18) is barely tractable. Hosoya (1986) proposed a transformation of test statistics to apply for odd DF. However, we do not discuss this as the limitation of an even integer DF is not a serious problem in practical usage of multiple portmanteau tests.

We now summarize three cases to evaluate significance levels and p-values of multiple portmanteau tests when all values of DF are even integers:

(I) When we conduct each portmanteau test with the same marginal significance level β by (18), the joint significance level α by (17) is computed from (23).

(II) When we set the joint significance level α by (17), the critical values d_js and marginal significance level β in (18) are calculated from the Newton-Raphson iteration method in Hosoya (1986, Section 2). Since $d_j = d_j(\beta)$s are defined by (18), the Newton-Raphson iteration method can be conducted as:

$$\beta^{(i+1)} = \beta^{(i)} - \frac{\Pr\{\psi(j) \leq d_j(\beta^{(i)}), \ j = 1, \ldots, s\} - (1 - \alpha)}{\partial \Pr\{\psi(j) \leq d_j(\beta^{(i)}), \ j = 1, \ldots, s\}/\partial \beta^{(i)}}, \tag{25}$$

where $\beta^{(i)}$ denotes the ith-step value of the marginal significance level β. Using (23) and (24), the derivative term in (25) is computed by:

$$\partial \Pr\{\psi(j) \leq d_j(\beta), \ j = 1, \ldots, s\}/\partial \beta \tag{26}$$
$$= -\Pr\{\phi(1, j) \leq d_j - d_1, \ j = 2, \ldots, s\}$$
$$- I(s \geq 3) \sum_{k=2}^{s-1} \Pr\{\psi(j) \leq d_j, \ j = 1, \ldots, k - 1 \,|\, \psi(k) = d_k\}$$
$$\times \Pr\{\phi(k, l) \leq d_l - d_k, \ l = k + 1, \ldots, s\}$$
$$- \Pr\{\psi(j) \leq d_j, \ j = 1, \ldots, s - 1 \,|\, \psi(s) = d_s\},$$

where $\phi(i, j) = \psi(j) - \psi(i) = Z_{m(i)+1}^2 + \cdots + Z_{m(j)}^2$ and $I(s \geq 3)$ is the indicator function such that $I = 1$ if $s \geq 3$ and $I = 0$ if $s = 2$. One reasonable starting value is $\beta^{(0)} = \alpha$. Equations (25) and (26) are computed from (23) and (24).

(III) When we compute the joint p-value from the marginal p-values, we also apply Hosoya (1989, Section 4.2.1). Let $\overline{Q}_{m(j)}$ be a computed portmanteau statistic with $m(j) - p - q$ DF. The marginal p-value is then given by $p^*(j) = \Pr(\psi(j) > \overline{Q}_{m(j)})$. Putting $p^*_{min} = \min\{p^*(1), \ldots, p^*(s)\}$, the critical values d^*_js are defined by:

$$p^*_{min} = \Pr(\psi(j) > d^*_j), \quad j = 1, \ldots, s.$$

Then the joint p-value, $p^*(s)$, is given by:

$$p^*(s) = 1 - \Pr(\psi(j) \leq d^*_j, \ j = 1, \ldots, s), \tag{27}$$

where $p^*(s)$ is computed from (23).

Table 1 gives Four examples of approximate significance levels and d_js on the multiple portmanteau tests by (I) and (II). To conduct the Newton-Raphson iteration method given in (25), we chose α as the starting value. The iterations continued until $|\beta^{(i+1)} - \beta^{(i)}| < 0.1^5$. A few iterations were enough for all cases. It would seem some monotonicity properties exist between the set of DF and α or β.

Table 1: Approximate significance levels of the multiple portmanteau tests given $100\alpha = 1, 5, 10$ (%) or $100\beta = 1, 5, 10$ (%) defined by (17) and (18) and the corresponding critical points, d_js, where DF$(\cdot)$ denotes a vector of DF.

DF(2, 6, 10)					DF(12, 18)			
$100\alpha(\%)$	$100\beta(\%)$	d_1	d_2	d_3	$100\alpha(\%)$	$100\beta(\%)$	d_1	d_2
1.0000	0.4289	10.9034	18.9269	25.6183	1.0000	0.6425	27.5543	36.3169
5.0000	2.3083	7.5373	14.6588	20.7264	5.0000	3.4030	22.3253	30.3666
10.000	4.8442	6.0548	12.6781	18.4091	10.0000	7.0490	19.8239	27.4760
2.2573	1.0000	9.2103	16.8119	23.2092	1.5363	1.0000	26.2169	34.8053
10.2972	5.0000	5.9915	12.5916	18.3070	7.2188	5.0000	21.0260	28.8693
19.3697	10.0000	4.6052	10.6446	15.9872	13.9029	10.0000	18.5493	25.9894

DF(4, 8, 12)					DF(18, 24)			
$100\alpha(\%)$	$100\beta(\%)$	d_1	d_2	d_3	$100\alpha(\%)$	$100\beta(\%)$	d_1	d_2
1.0000	0.4595	15.0520	22.1791	28.5489	1.0000	0.6726	36.1620	44.4691
5.0000	2.4915	11.1513	17.5443	23.3476	5.0000	3.5604	30.1938	37.8874
10.0000	5.2448	9.3719	15.3637	20.8613	10.0000	7.3650	27.2937	34.6571
2.1064	1.0000	13.2767	20.0902	26.2169	1.4694	1.0000	34.8053	42.9798
9.5687	5.0000	9.4877	15.5073	21.0260	6.9182	5.0000	28.8693	36.4150
18.0009	10.0000	7.7794	13.3616	18.5493	13.3534	10.0000	25.9894	33.1962

The empirical significance levels corresponding to $\alpha = 0.05$ for the AR(1) model, $(1 - \alpha^0 L)x_t = \varepsilon_t$, are given in Figure 1, where $\{\varepsilon_t\}$ is iid $N(0, 1)$. The solid (dashed) line denotes the empirical significance level of Q^*_m (Q^{**}_m). We conducted 4 cases of the set of DF given in Table 1 with $n = 100$. Figure 1 reveals that: (i) Q^*_m and Q^{**}_m perform similarly unless m is very small and $|\alpha^0|$ close to 1. Note that approximation of the marginal and joint significance levels depend on ϵ and m, and $\epsilon = |\alpha^0|$ in the

8

AR(1) model. Therefore, departure of Q_m^* an Q_m^{**} was effected by a correction term $\widehat{B}_m^*$ in this case. (ii) The empirical marginal (joint) significance levels increase as the number of DF increase. This is also apparent from previous simulation studies. (iii) However, the empirical joint significance level of the multiple portmanteau tests performs well if the set of DF contains a small number of DF. It would then appear that the empirical marginal significance levels smoothes the empirical joint significance level. Together, these suggest that use of the joint significance level of the multiple portmanteau tests for small to medium values of DF is highly effective because an approximation of the empirical marginal significance levels occasionally performs poorly depending on the values of m and ϵ (the case where m is large or small and ϵ is close to 1).
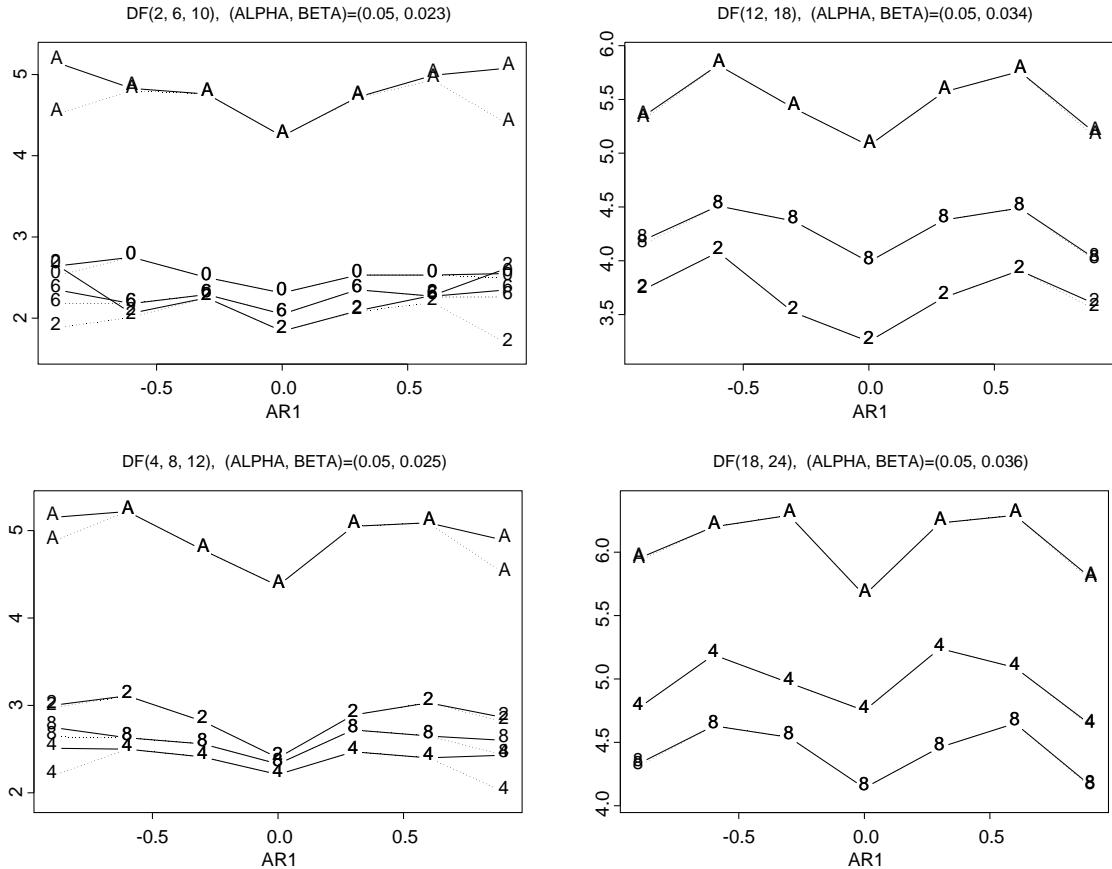
Figure 1: Percentages of the empirical significance level of the statistics $Q_m^* \sim \chi_{m-1}^2$ (solid line) and $Q_m^{**} \sim \chi_{m-1}^2$ (dashed line) for the AR(1) model, $(1 - \alpha^0 L)x_t = \varepsilon_t$; $n = 100$, and $\alpha^0 = 0, \pm0.3, \pm0.6, \pm0.9$, with 10,000 replications. DF($\cdot$) denotes a vector of DF. The significance levels are given in Table 1. Lines with A denote the empirical joint significance level and lines with a number denote the empirical marginal significance level, where the numbers denote single digit DF. For example, in the figure DF(2,6,10), 2 denotes 2 DF and 0 denotes 10 DF.

From these analysis, and to avoid the arbitrary choice of DF, researchers may conduct multiple portmanteau tests with any possible even DF, $m(i) - p - q = 2i, i = 1, 2, \ldots, s$, similarly to the S-PLUS function `arima.diag`. Figure 2 depicts these joint approximate significance levels α by varying $2s = 2, 4, \ldots, 100$, when $\beta = 0.01, 0.05, 0.10$ are given. The computations are given by (23). Figure 2

reveals that α increases like a logarithm function and, for $2s \leq 40$, α is at most $0.05, 0.20, 0.34$ when $\beta = 0.01, 0.05, 0.10$, respectively. Note that this figure also shows the upper bound of α if we conduct multiple portmanteau tests with even DF, $m(i) - p - q$, $i = 1, \ldots, k$, $k \leq s$ and $m(k) - p - q \leq 2s$. The joint empirical significance levels for the case of $s = 12$ is examined with Figure 3, where we use an AR(1) model as in Figure 1. The joint empirical significance performs well uniformly in α^0 and β. This is consistent with the conjectures obtained by Figure 1.
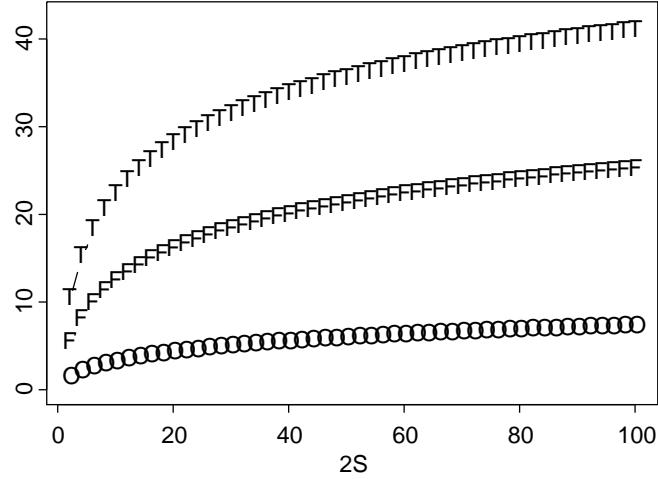
Figure 2: Joint significance level $100\alpha\%$ varying $2s = 2, 4, \ldots, 100$, where O denotes $100\beta = 1\%$, F denotes $100\beta = 5\%$ and T denotes $100\beta = 10\%$.
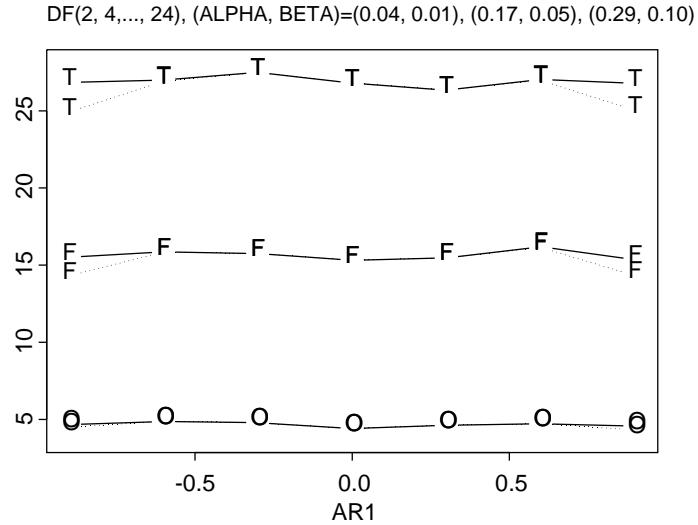
Figure 3: Percentages of the joint empirical significance level of the statistics $Q_m^* \sim \chi_{m-1}^2$ (solid line) and $Q_m^{**} \sim \chi_{m-1}^2$ (dashed line) for the AR(1) model, $(1 - \alpha^0 L)x_t = \varepsilon_t$; $n = 100$, and $\alpha^0 = 0, \pm 0.3, \pm 0.6, \pm 0.9$, with 10,000 replications, where DF($\cdot$) denotes a vector of DF, O denotes $100\beta = 1\%$ ($100\alpha = 4.08\%$), F denotes $100\beta = 5\%$ ($100\alpha = 16.6\%$) and T denotes $100\beta = 10\%$ ($100\alpha = 29.1\%$).

APPENDIX

Proof of Lemma 1: For simplicity, we shall only prove $\boldsymbol{E}_{m(1)}(\boldsymbol{E}_{m(2)} - \boldsymbol{E}_{m(1)}) = O(\epsilon^{m+1-\max(p,q)})$. However, other cases can be similarly treated. First, we prove that:

$$\boldsymbol{D}_{m(2)} = \begin{bmatrix} \boldsymbol{D}_m & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} + O\left(\epsilon^{m+1-\max(p,q)}\right). \tag{A.1}$$

Let:

$$\boldsymbol{X}_{m(2)} = \begin{bmatrix} \boldsymbol{X}_m \\ \boldsymbol{X}_{1,2} \end{bmatrix} \quad \text{where} \quad \boldsymbol{X}_{1,2} = \left[\boldsymbol{d}_{m+1}(\boldsymbol{\theta}_0), \dots, \boldsymbol{d}_{m(2)}(\boldsymbol{\theta}_0)\right]'.$$

Then we obtain:

$$\boldsymbol{D}_{m(2)} = \begin{bmatrix} \boldsymbol{X}_m(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_m & \boldsymbol{X}_m(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_{1,2} \\ \boldsymbol{X}_{1,2}(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_m & \boldsymbol{X}_{1,2}(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_{1,2} \end{bmatrix}.$$

Since $\boldsymbol{X}'_m\boldsymbol{X}_m \to \boldsymbol{I}(\boldsymbol{\theta}_0)$,

$$\boldsymbol{d}_m(\boldsymbol{\theta}_0) = O\left(\epsilon^{m-\max(p,q)}\right) \quad \text{and} \quad \boldsymbol{X}_{1,2} = O(\boldsymbol{d}_{m+1}(\boldsymbol{\theta}_0)) = O\left(\epsilon^{m+1-\max(p,q)}\right) \tag{A.2}$$

from Katayama (2008, equation (16)), we obtain:

$$\boldsymbol{X}_m(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_m - \boldsymbol{D}_m$$

$$= \boldsymbol{X}_m(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\left\{\mathbb{I}_{p+q} - (\boldsymbol{X}'_m\boldsymbol{X}_m + \boldsymbol{X}'_{1,2}\boldsymbol{X}_{1,2})(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-1}\right\}\boldsymbol{X}'_m$$

$$= -\boldsymbol{X}_m(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_{1,2}\boldsymbol{X}_{1,2}(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-1}\boldsymbol{X}'_m$$

$$= O\left(\boldsymbol{X}'_{1,2}\boldsymbol{X}_{1,2}\right)$$

$$= O\left(\sum_{i=m+1}^{\infty} \epsilon^{i-\max(p,q)}\right)$$

$$= O\left(\epsilon^{m+1-\max(p,q)}\right).$$

Similarly, $\boldsymbol{X}_m(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_{1,2} = O(\epsilon^{m+1-\max(p,q)})$ and

$$\boldsymbol{X}_{1,2}(\boldsymbol{X}'_{m(2)}\boldsymbol{X}_{m(2)})^{-1}\boldsymbol{X}'_{1,2} = O(\boldsymbol{d}_{m+1}(\boldsymbol{\theta}_0)'\boldsymbol{d}_{m+1}(\boldsymbol{\theta}_0))$$

$$= O\left(\epsilon^{2\{m+1-\max(p,q)\}}\right),$$

which proves (A.1). It follows that:

$$
\boldsymbol{E}_{m(2)} - \boldsymbol{E}_m = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \mathbb{I}_{m(2)-m} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} + O\left(\epsilon^{m+1-\max(p,q)}\right), \tag{A.3}
$$

which proves $\boldsymbol{E}_{m(1)}\left(\boldsymbol{E}_{m(2)} - \boldsymbol{E}_{m(1)}\right) = O(\epsilon^{m+1-\max(p,q)})$. $\qquad\square$

Proof of Theorem 1: From Lemma 1 and (A.3), we obtain:

$$
\begin{bmatrix} \boldsymbol{z}'_M \boldsymbol{E}_{m(1)} \boldsymbol{z}_M \\ \boldsymbol{z}'_M \boldsymbol{E}_{m(2)} \boldsymbol{z}_M \\ \vdots \\ \boldsymbol{z}'_M \boldsymbol{E}_{m(s)} \boldsymbol{z}_M \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{z}'_M \boldsymbol{E}_{m(1)} \boldsymbol{z}_M \\ \boldsymbol{z}'_M (\boldsymbol{E}_{m(2)} - \boldsymbol{E}_{m(1)}) \boldsymbol{z}_M \\ \vdots \\ \boldsymbol{z}'_M (\boldsymbol{E}_{m(s)} - \boldsymbol{E}_{m(s-1)}) \boldsymbol{z}_M \end{bmatrix}
$$

$$
= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sum_{i=p+q+1}^{m(1)} Z_i^2 \\ \sum_{i=m(1)+1}^{m(2)} Z_i^2 \\ \vdots \\ \sum_{i=m(s-1)+1}^{m(s)} Z_i^2 \end{bmatrix} + O(\epsilon^{m+1-\max(p,q)}),
$$

where the lower triangular matrices are $s \times s$ matrices. $\qquad\square$

ACKNOWLEDGMENTS

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.

HOSOYA, Y. (1986) A Simultaneous test in the presence of nested alternative hypotheses. *Journal of Applied Probability* **23A**, 187–200.

——— (1989) Hierarchical statistical models and a generalized likelihood ratio test. *Journal of the Royal Statistical Society* **B51**, 435–447.

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

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.

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

OMITTED ARGUMENTS

In this section, we present proofs and arguments not included in the main text of the paper and not to be published.

Typos in Hosoya (1986, pp193–194):

1. On equation (2.6), 1st term on RHS. $\sum_{k=2}^{p-2}$ should be $I(p \geq 3) \sum_{k=2}^{p-1}$.

2. On equations (2.7) and (2.8), these should be defined by $k = 2, \ldots, p$ because the cases of $k = 1$ are not well defined.

3. On equation (2.9), 1st term on RHS of 1st equation. $-d_1^k \exp(-d_1/2)$ should be $-2d_1^k \exp(-d_1/2)$.

4. On equation (2.9), 3rd equation is not well defined from (2.8).

5. On equation (2.11), 2nd term on RHS. $B(k-2, m_{k-1} + j + 1)$ should be $B(k-1, m_{k-1} + j + 1)$.

Figures showing rates of rejection for the empirical experiments: We present figures that illustrate empirical results for the testing for AR(1) model, $(1 - \alpha^0)x_t = \varepsilon_t$; $n = 100$, and $\alpha^0 = 0, \pm 0.1, \ldots, \pm 0.9$, with 10,000 replications. Notations are the same as used for Figures 1 and 3.
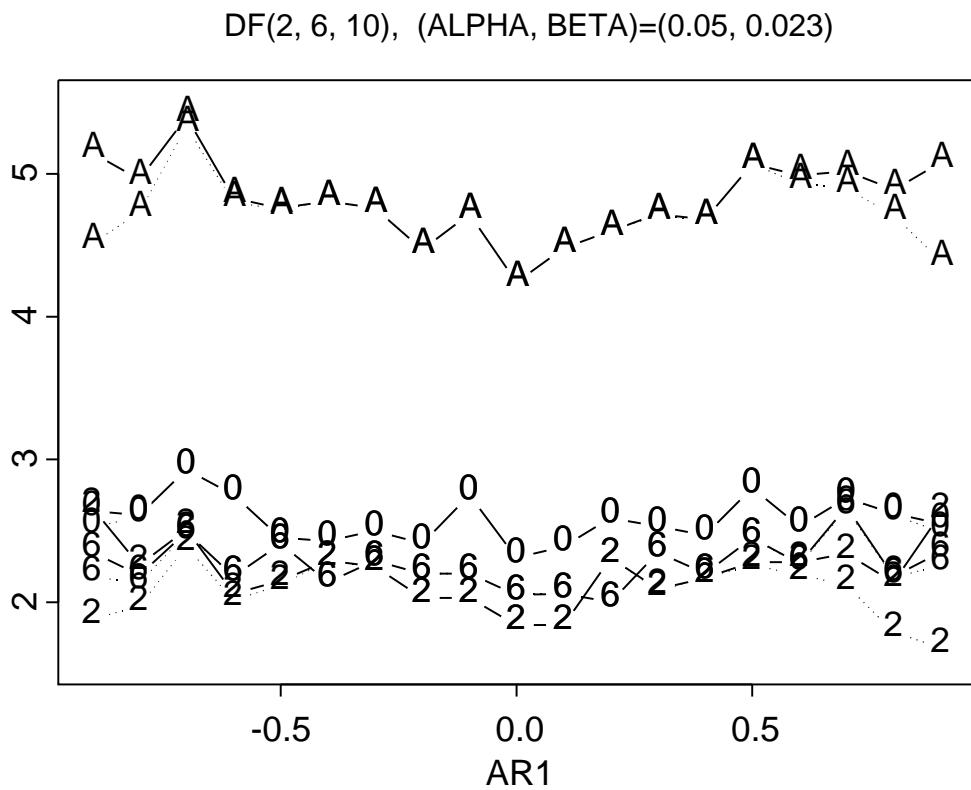

Figure 4: Empirical significance level (%).

DF(4, 8, 12), (ALPHA, BETA)=(0.05, 0.025)
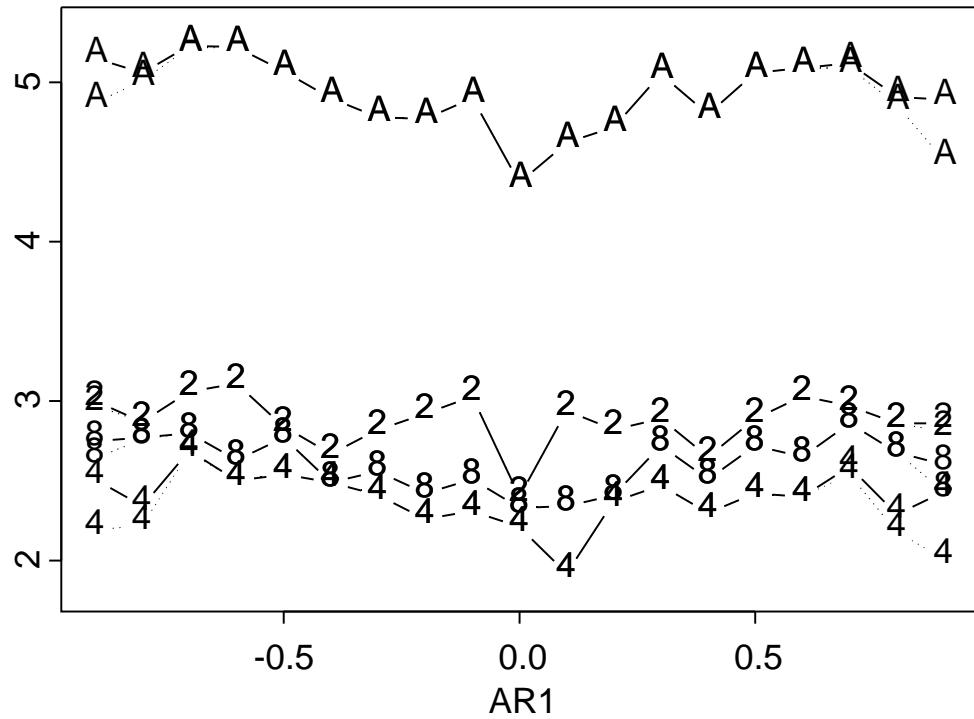
Figure 5: Empirical significance level (%).

DF(12, 18), (ALPHA, BETA)=(0.05, 0.034)
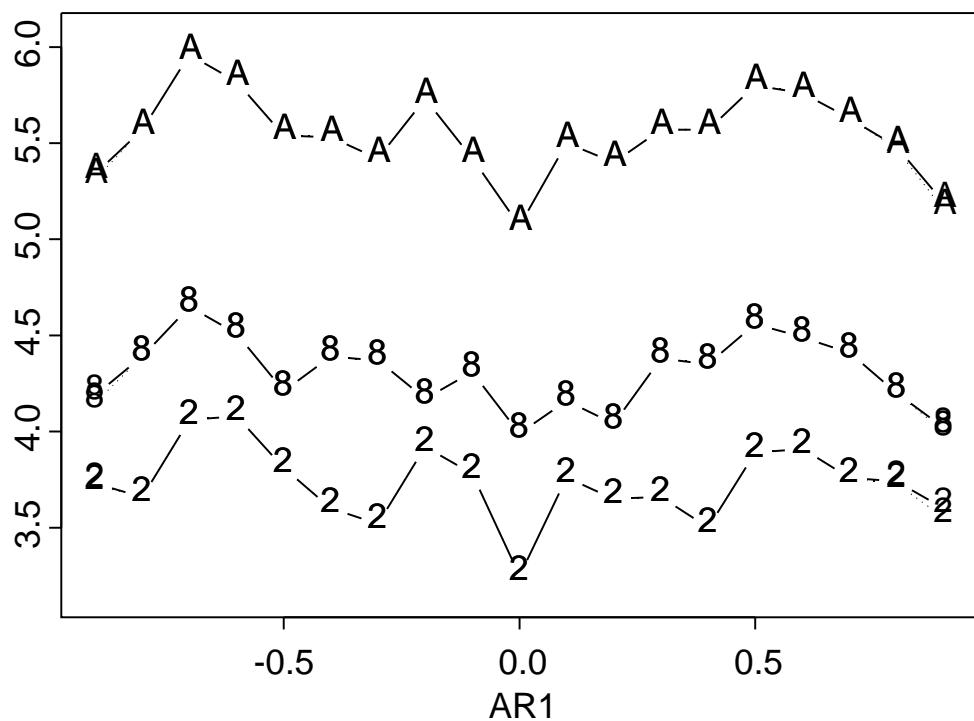
Figure 6: Empirical significance level (%).

DF(18, 24), (ALPHA, BETA)=(0.05, 0.036)
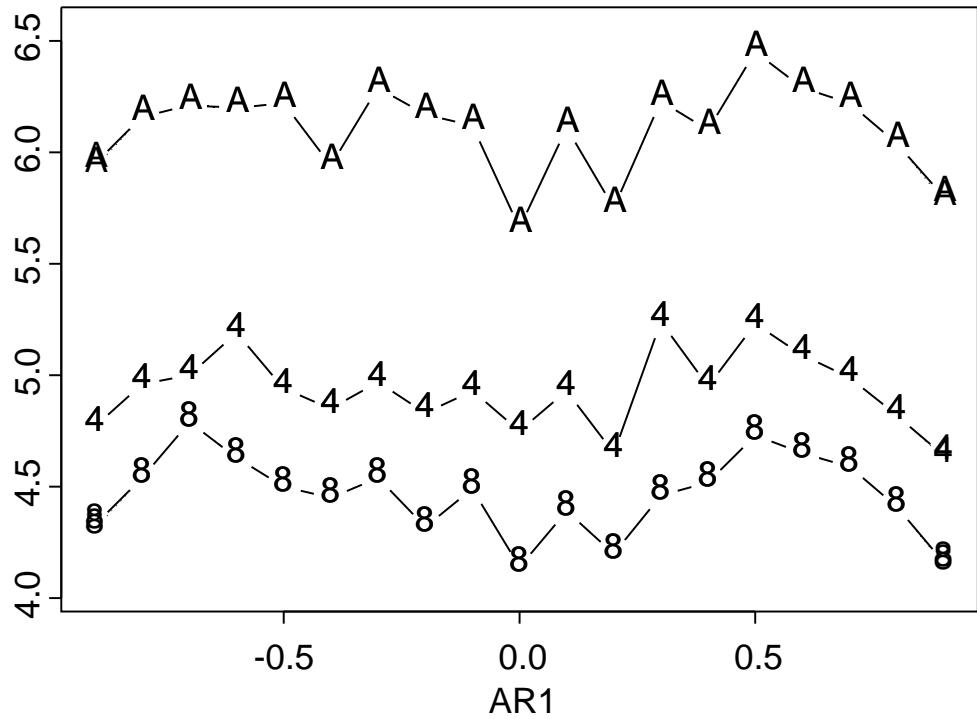
Figure 7: Empirical significance level (%).

DF(2, 4,..., 24), (ALPHA, BETA)=(0.04, 0.01), (0.17, 0.05), (0.29, 0.10)
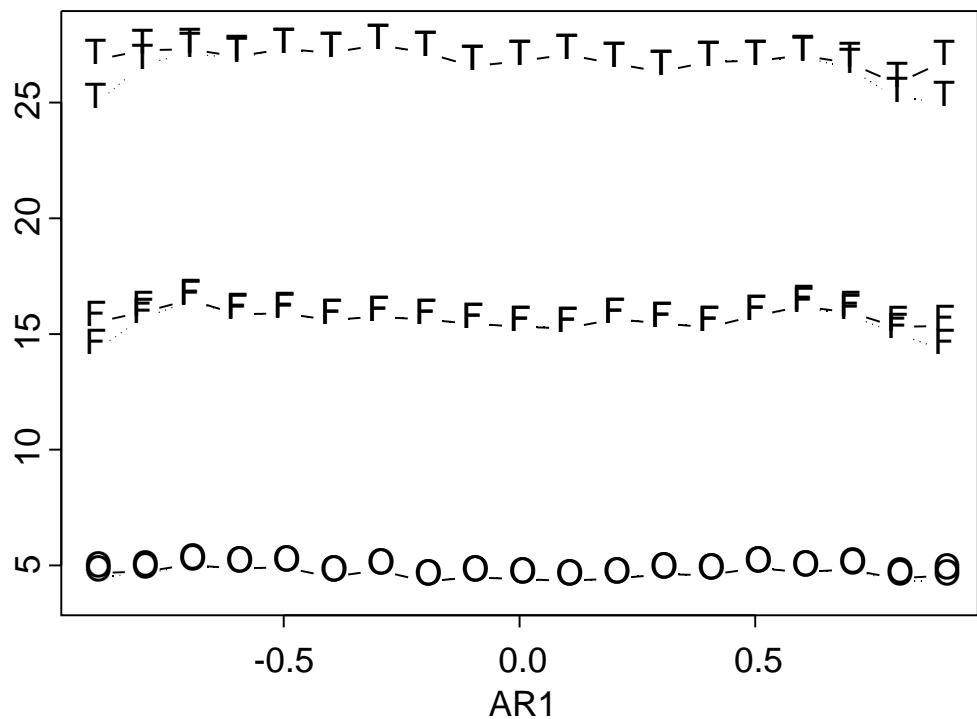
Figure 8: Empirical significance level (%).