Sankhyā : The Indian Journal of Statistics
2018, Volume 80-A, Part 1, pp. 121-137

A New Look at Portmanteau Tests

Fumiya Akashi, Hiroaki Odashima and Masanobu Taniguchi
Waseda University, Tokyo, Japan
Anna Clara Monti
University of Sannio, Benevento, Italy

Abstract

Portmanteau tests are some of the most commonly used statistical methods for model diagnostics. They can be applied in model checking either in the time series or in the regression context. The present paper proposes a portmanteau-type test, based on a sort of likelihood ratio statistic, useful to test general parametric hypotheses inherent to statistical models, which includes the classical portmanteau tests as special cases. Sufficient conditions for the statistic to be asymptotically chi-square distributed are elucidated in terms of the Fisher information matrix, and the results have very clear implications for the relationships between the parameter of interest and nuisance parameter. In addition, the power of the test is investigated when local alternative hypotheses are considered. Some interesting applications of the proposed test to various problems are illustrated, such as serial correlation tests where the proposed test is shown to be asymptotically equivalent to classical tests. Since portmanteau tests are widely used in many fields, it appears essential to elucidate the fundamental mechanism in a unified view.

AMS (2000) *subject classification*. 62F03, 62F05.
Keywords and phrases. Portmanteau test, Asymptotic local power, Serial correlation, Time series analysis, Variable selection

1 Introduction

Diagnostics is a central issue in statistical modeling, and one of the important tasks of diagnostics concerns verifying the absence of serial correlation of the error term. To check the adequacy of a fitted autoregressive moving average (ARMA) model, Box and Pierce (1970) proposed the test statistic

$$T_{BP} = n \sum_{k=1}^{M} \hat{r}_k^2, \tag{1.1}$$

The first and third author were supported by the JSPS Grant-in-Aid for Young Scientists (B): 16K16022 (Akashi, F., Waseda University) and JSPS Grant-in-Aid: 15H02061 (Taniguchi, M., Waseda University), respectively.

where $\hat{r}_k$ is the residual empirical autocorrelation at lag k, n is the sample size and M is a fixed integer. T_{BP} is approximately χ^2 distributed with $M-p-q$ degrees of freedom, where p and q are the order of the autoregressive and moving average polynomials, if n and M are moderately large (see Box and Pierce 1970). Ljung and Box (1978) improved T_{BP} to enhance the χ^2 approximation to its asymptotic distribution. The modified statistic, called the Ljung-Box test statistic, is defined as

$$T_{LB} = n(n+2)\sum_{k=1}^{M} \hat{r}_k^2/n-k. \tag{1.2}$$

Subsequently, many other modified statistics have been suggested and their powers have been evaluated (e.g. McLeod and Li 1983; Monti 1994; Pẽna and Rodríguez 2002). A nonparametric approach to testing for serial correlation was also suggested by Chan and Tran (1992). Li (2003) illustrates diagnostic methods and their developments in the last few decades comprehensively. Taniguchi and Amano (2009) elucidated the mechanism of portmanteau tests on the basis of a likelihood ratio statistic derived from the Whittle Likelihood. These authors showed that their statistic is asymptotically equivalent to the classical portmanteau test statistic, and that it is not asymptotically χ^2 distributed in specific models if M is fixed. Also in the context of linear regression models, various methods useful to detect serial correlation have been investigated, since when the errors are serially correlated, the ordinary least square estimators fail to be the best linear unbiased estimators. Durbin (1970) constructed naive tests of goodness of fit against AR(1) error correlation, which proved to be robust under various alternative hypotheses. Godfrey (1976) applied Durbin's procedure in testing for serial correlation in dynamic simultaneous equation models. Breusch (1978) compared Durbin's procedure with the Lagrange multiplier (LM) test in the context of dynamic linear models and illustrated the relationship between a portmanteau test and the LM test.

There are several advantages in using a portmanteau test. For example, the statistic is computationally much simpler to obtain than other test statistics, such as the LM and the Wald statistic. Consequentely several situations arise where a portmanteau statistic turns out to be more useful than alternative statistics. This evidence makes worth elucidating the properties of the portmanteau test, its theoretical background and its fundamental mechanism in a unified view.

The present paper proposes a portmanteau test which is widely applicable for the diagnostics of general statistical models. The statistic is obtained as a sort of likelihood ratio statistic, and its asymptotic properties are investigated. In particular, sufficient conditions are derived such that the statistic is asymptotically χ^2 distributed. The conditions are given in terms of the Fisher information matrix, and have very clear implications for the relationships between the parameter of interest and the nuisance parameter.

The paper is organized as follows. Section 2 introduces the portmanteau test statistic and gives the sufficient conditions which ensure an asymptotic χ^2 distribution. The same section investigates the limit distribution of the statistic under local contiguous alternatives, and evaluates the local power of the test. Section 3 shows some applications of the portmanteau test, which highlight that the framework developed for the test is widely applicable to various models. Finally, rigorous proofs of the theorems are relegated to Section 4.

As concerns notations and symbols used in this paper, the set of all integers is denoted as $\mathbb{Z}$. For any sequence of random vectors $\{A(t) : t \in \mathbb{Z}\}$, $A(t) \xrightarrow{p} A$ and $A(t) \xrightarrow{d} A$, respectively, denote the convergence to a random (or constant) vector A in probability and law. The transpose and complex transpose of a matrix M is denoted by $M^\top$. Moreover, the square root of a nonnegative definite matrix M is denoted by $M^{1/2}$. 0_i, $O_{j\times k}$ and I_l denote the i-dimensional zero vector, the $j \times k$ zero matrix and the $l \times l$ identity matrix, respectively. Finally $\chi^2_k(\mu)$ denotes the noncentral χ^2 random variable (r.v.) with k degrees of freedom and noncentrality parameter μ.

2 Portmanteau Test for General Statistical Models

Let $x^{(n)} = (X_1^\top, \ldots, X_n^\top)^\top$ be a collection of d-dimensional random vectors and let $p_n(x^{(n)}; \theta)$ be the probability density function of $x^{(n)}$ with $\theta = (\theta_1^\top, \theta_2^\top)^\top \in \Theta \subset \mathbb{R}^{q+p}$, where $\theta_1 = (\theta_{1,1}, \ldots, \theta_{1,q})^\top \in \mathbb{R}^q$ $(q \geq 0)$ and $\theta_2 = (\theta_{2,1}, \ldots, \theta_{2,p})^\top \in \mathbb{R}^p$ $(p \geq 1)$. We also assume that Θ is a compact subset of $\mathbb{R}^{q+p}$. The paper focuses on the following testing problem

$$H : \theta_2 = \theta_2^0 \text{ against } A : \theta_2 \neq \theta_2^0. \tag{2.1}$$

This setting is very general and encompasses various classical testing problems in time series analysis, multivariate analysis, and so on. In particular this setting can address the following three important applications:

(i) testing serial correlation in stationary time series models,

(ii) testing serial correlation in linear regression models,

(iii) variable selection problems in linear regression models.

Hence widely recurrent diagnostic analyses are embedded in our unifying setting. We will illustrate the details of some of the applications in Section 3.

In what follows, the portmanteau type test statistic is introduced. Hereafter, we adopt the notation $l(\theta) = l(\theta_1, \theta_2) = n^{-1} \log p_n\{x^{(n)}; \theta\}$ for the log-likelihood function, where $\theta^0 = (\theta_1^\top, \theta_2^{0\top})^\top$.

The Fisher information matrix is given by

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix},$$

where

$$F_{ij} = \lim_{n\to\infty} nE_{\theta^0}\left[\partial l(\theta)/\partial\theta_i \partial l(\theta)/\partial\theta_j^\top\Big|_{\theta=\theta^0}\right] \quad (i, j \in \{1, 2\})$$

and E_θ denotes the expectation with respect to $p\{x^{(n)}; \theta\}$.

We also define the constrained maximum likelihood estimator of θ_1 under H, $\hat{\theta}_1 = \arg\max_{\theta_1\in\Theta_1} l(\theta_1, \theta_2^0)$, and the constrained estimator of θ_2 when $\theta_1 = \hat{\theta}_1$, $\hat{\theta}_2(\hat{\theta}_1) = \arg\max_{\theta_2\in\Theta_2} l(\hat{\theta}_1, \theta_2)$. The general portmanteau test, which is a sort of likelihood ratio test between H and A, is based on the following statistic

$$T_P = 2n\left[l\left\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\right\} - l(\hat{\theta}_1, \theta_2^0)\right],$$

which compares the values of the log-likelihood function when θ_2 varies between θ_2^0 and $\hat{\theta}_2(\hat{\theta}_1)$.

Remark 1. *Unlike the classical likelihood ratio theory, we initially compute the constrained estimator $\hat{\theta}_1$ under H, then we maximize the likelihood function with respect to θ_2 under the constraint $\theta_1 = \hat{\theta}_1$ and obtain $l\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\}$. Hence in both stages, the value of one or the other parameter is fixed, and we never consider the global maximum "$\max_{(\theta_1,\theta_2)\in\Theta} l(\theta_1, \theta_2)$". This procedure has undeniable benefits from the computational viewpoint, since the optimization is always carried out over a space of smaller dimension than Θ, and it is crucial for the asymptotic expansion of the test statistic T_P given in Theorems 1 and 2. Moreover, due to this approach, the proposed test statistic asymptotically coincides with T_{BP}. Further analyses on the effects of this procedure are developed in Section 3.2.*

To derive the asymptotic distribution of T_P, the conditions summarized in the next assumption are required.

Assumption 1.

(i) F_{11} and F_{22} are nonsingular.

(ii) *$l(\theta)$ is continuously three times differentiable with respect to θ, and the partial derivative $\partial/\partial\theta$ and the expectation E_θ are interchangeable.*

(iii) *The cumulants of $U = (U(1), \ldots, U(p+q))^\top = n^{1/2}\partial l(\theta)/\partial\theta|_{\theta=\theta^0}$ of any order exist, and satisfy*

$$cum\{U(i_1), \ldots, U(i_J)\} = O\left(n^{-J/2+1}\right)$$

for each $J = 2, 3, \ldots$ and any $i_1, \ldots, i_J \in \{1, \ldots, p\}$. Here the cumulant $cum\{Y(1), \ldots, Y(J)\}$ of $(Y(1), \ldots, Y(J))^\top$ is defined as

$$cum\{Y(1), \ldots, Y(J)\} = \sum_{p=1}^{r}(-1)^{p-1}(p-1)!E\left[\prod_{j\in\nu_1} Y(j)\right]\cdots E\left[\prod_{j\in\nu_p} Y(j)\right],$$

where the summation extends over all partitions $\{\nu_1, \ldots, \nu_p\}$, $p = 1, \ldots, J$, of $\{1, \ldots, J\}$.

Remark 2. *These conditions are pretty mild and hold for most of the models of interest in statistical applications. The second and third conditions, in particular, are needed to guarantee the asymptotic normality of the score $\partial l(\theta)/\partial\theta$. From Brillinger (1981, Section 2), $\partial l(\theta)/\partial\theta$ is asymptotically normally distributed if and only if the joint cumulants of $\partial l(\theta)/\partial\theta$ of order $J \geq 3$ vanish, since the third and higher-order cumulants of the normal distribution are zero (see also Theorem 2.3.1 of Brillinger (1981)).*

Since condition (iii) of Assumption 1 is fairly technical, we illustrate an example of a model where (iii) holds.

Example 1. *Consider the regression model*

$$y = Z\beta + \epsilon, \tag{2.2}$$

where Z is an $n \times m$ matrix of observations on the regressors such that rank$(Z) = m$, y is an $n \times 1$ vector of observations on the regressand, β is an $m \times 1$ vector-valued unknown coefficient, and ϵ is an $n \times 1$ random vector distributed as $N(0_n, \Sigma_n)$, where Σ_n is an $n \times n$ positive definite matrix. We assume that Σ_n depends on an unknown scale parameter $\xi > 0$ and an unknown correlation coefficient $b \in (-1, 1)$ as follows

$$\Sigma_n(\xi, b) = \begin{cases} \left(\xi(-b)^{|i-j|} : i, j = 1, \ldots, n\right) & (b \neq 0) \\ \xi I_n & (b = 0) \end{cases}.$$

In other words, we suppose that $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^\top$ *is generated by an AR(1) process* $\epsilon_t = -b\epsilon_{t-1} + u_t$ *with* $\epsilon_0 = 0$ *and* $\{u_t\}$ *is a sequence of independent and identically distributed (i.i.d.)* $N(0, \xi)$ *random variables.*

Now, suppose that we are interested in whether the noise ϵ *is uncorrelated. In this case, the nuisance parameter is* $\theta_1 = (\beta^\top, \xi)^\top$ *and the parameter of interest is* $\theta_2 = b$*. That is, the testing problem* (2.1) *becomes*

$$H : b = 0 \text{ against } A : b \neq 0. \tag{2.3}$$

Simple algebra shows that the quantities $U(j)$ *in Assumption 1 are given by*

$$U(j) = \begin{cases} -n^{-1/2}\xi^{-1} \sum_{i=1}^{n} z^{ij} u_i(\beta) & (j = 1, \ldots, m) \\ -n^{1/2}(2\xi)^{-1} + n^{-1/2}(2\xi^2)^{-1} \sum_{i=1}^{n} u_i(\beta)^2 & (j = m + 1) \\ -n^{-1/2}\xi^{-1} \sum_{i=1}^{n-1} u_i(\beta)u_{i+1}(\beta) & (j = m + 2) \end{cases}, \tag{2.4}$$

where z^{ij} *is the* (i, j)*th component of* Z *and* $u_i(\beta)$ *is the* i*th component of* $y - Z\beta$*. Based on* (2.4)*, it is easily seen that for* $U = (U(1), \ldots, U(m+2))^\top$*,* $E[U] = 0_{m+2}$ *and* $Var[U] = O(1)$*. Further, from Brillinger (1981), the cumulants have the following three useful properties:*

(i) *for any set of random variables* $Y(1), \ldots, Y(J)$ *and given constants* $a_1, \ldots, a_J$*,*

$$cum\{a_1Y(1), \ldots, a_JY(J)\} = a_1 \cdots a_J cum\{Y(1), \ldots, Y(J)\};$$

(ii) *if two groups of random variables* $\{Y(1), \ldots, Y(r)\}$ *and* $\{Y(r+1), \ldots, Y(J)\}$ *are independent, then*

$$cum\{Y(1), \ldots, Y(J)\} = 0;$$

(iii) *if* $(Y(1), \ldots, Y(J))$ *is a normal vector for* $J \geq 3$*, then*

$$cum\{Y(1), \ldots, Y(J)\} = 0.$$

Therefore, the joint cumulant is

$$\begin{aligned} &cum\{U(j_1), \ldots, U(j_J)\} \\ &= 1/n^{J/2} \sum_{i_1=1}^{n} \cdots \sum_{i_J=1}^{n} A(i_1, j_1) \cdots A(i_J, j_J) cum\{V(i_1), \ldots, V(i_J)\}, \end{aligned}$$

for $J \geq 3$ *and any* $j_1, \ldots, j_J \in \{1, \ldots, m + 2\}$*, where* $V(i)$ *is one of* $u_i(\beta)$*,* $u_i(\beta)^2$ *and* $u_i(\beta)u_{i+1}(\beta)$ *for* $i = 1, \ldots, n$ *and* $u_{n+1}(\beta) = 0$*. The coefficient* $A(i, j)$ *is given by*

$$A(i, j) = \begin{cases} -\xi^{-1} z^{ij} & (j = 1, \ldots, m) \\ (2\xi^2)^{-1} & (j = m + 1) \\ -\xi^{-1} & (j = m + 2) \end{cases}$$

and is obviously $O(1)$ uniformly in $i = i_1, \ldots, i_J \in \{1, \ldots, n\}$ and $j = j_1, \ldots, j_J \in \{1, \ldots, m+2\}$. Moreover, by recalling the fact that $u_1(\beta), \ldots, u_n(\beta)$ are i.i.d. normal r.v.s under H and using Theorem 2.3.2 by Brillinger (1981), it can be shown that

$$\sum_{i_1=1}^{n} \cdots \sum_{i_J=1}^{n} A(i_1, j_1) \cdots A(i_J, j_J) cum\{V(i_1), \ldots, V(i_J)\} = O(n).$$

Hence, Assumption 1 is satisfied. We will revisit this model later on.

The asymptotic expansion of T_P is given by the following theorem.

Theorem 1. *Suppose Assumption 1 holds. Then, under H,*

$$T_P = N^\top F_{22\cdot 1}^{1/2} F_{22}^{-1} F_{22\cdot 1}^{1/2} N + o_p(1), \tag{2.5}$$

where N is a p-dimensional standard normal random vector and $F_{22\cdot 1} = F_{22} - F_{21} F_{11}^{-1} F_{12}$.

Since the necessary and sufficient conditions for the quadratic form Eq. 2.5 to have an asymptotic χ^2 distribution are discussed in Tziritas (1987) in detail, we confine ourselves to state the following Corollary, which is a straightforward consequence of Theorem 1.

Corollary 1. *Suppose Assumption 1 holds.*

(i). *If $q < p$, $F_{22} = I_p$ and $F_{21} F_{11}^{-1} F_{12}$ is idempotent with rank r, then, under H, $T_P \xrightarrow{d} \chi^2_{p-r}$ as $n \to \infty$.*

(ii). *If $F_{22} \neq I_p$ and $F_{12} = O_{q \times p}$, then, under H, $T_P \xrightarrow{d} \chi^2_p$ as $n \to \infty$.*

Theorem 1 and Corollary 1 make clear that the asymptotic distribution of T_P depends on the structure of the information matrix and provide two alternative conditions which ensure an asymptotic χ^2 distribution. In this regard Theorem 1 elucidates the mechanism of portmanteau tests in a general framework, and grasps the stream of portmanteau-works as special cases. In the case of time series, in particular, Taniguchi and Amano (2009) introduced the Whittle likelihood-based test statistic, and showed that the test based on this statistic is asymptotically equivalent to the Box-Pierce test based on T_{BP} and to the Ljung-Box test based on T_{LB}. Our approach is applicable not only to Taniguchi and Amano (2009)'s situation (when a Whittle likelihood is available) but also to more general contexts. In what follows we shall refer to the test based on T_P as the T_P-test.

Next, the goodness of the T_P-test is investigated in terms of local power. Consider the sequence of local alternatives

$$A_n : \theta_2 = \theta_2^{(n)} = \theta_2^0 + n^{-1/2} h,$$

where $h = (h_1, \dots, h_p)^\top \in \mathbb{R}^p$, and denote the log-likelihood ratio between H and A_n by $\Lambda_n(\theta^0, \theta^{(n)})$; that is,

$$\Lambda_n \left\{\theta^0, \theta^{(n)}\right\} = \log p_n \left\{x^{(n)}, \theta^{(n)}\right\} / p_n \left\{x^{(n)}, \theta^0\right\},$$

where $\theta^{(n)} = (\theta_1^\top, \theta_2^{(n)\top})^\top$. It is known that $\Lambda_n(\theta^0, \theta^{(n)})$ is asymptotically normal for a sufficiently rich class of regular statistical models (e.g. Taniguchi and Kakizawa 2000). Hence, the following assumption is considered to hold.

Assumption 2. *Under H, the log-likelihood ratio $\Lambda_n\{\theta^0, \theta^{(n)}\}$ admits the stochastic expansion*

$$\Lambda_n \left\{\theta^0, \theta^{(n)}\right\} = h^\top \Delta_n - 1/2 h^\top F_{22} h + o_p(1)$$

as $n \to \infty$, where $\{\Delta_n\}$ is a sequence of random vectors such that $\Delta_n \xrightarrow{d} \Delta$, and Δ is a p-dimensional normal random vector with zero mean vector and covariance matrix F_{22}.

A family $\{p_n(x^{(n)}, \theta)\}$ satisfying Assumption 2 is said to be locally asymptotically normal. By using LeCam's so-called third lemma, we have the following theorems on the asymptotic distribution of the test statistics under A_n.

Theorem 2. *Suppose Assumptions 1 and 2 hold. Then, under A_n,*

$$T_P = \left(N + F_{22\cdot1}^{1/2} h\right)^\top F_{22\cdot1}^{1/2} F_{22}^{-1} F_{22\cdot1}^{1/2} \left(N + F_{22\cdot1}^{1/2} h\right) + o_p(1),$$

where N is a p-dimensional standard normal random vector and $F_{22\cdot1} = F_{22} - F_{21} F_{11}^{-1} F_{12}$.

Corollary 2. *Suppose Assumptions 1 and 2 hold.*

1. *If $q < p$, $F_{22} = I_p$ and $F_{21} F_{11}^{-1} F_{12}$ is idempotent with rank r, then, under A_n, $T_P \xrightarrow{d} \chi^2_{p-r}(\mu)$ as $n \to \infty$, where $\mu = h^\top F_{22\cdot1} h$.*
2. *If $F_{22} \neq I_p$ and $F_{12} = O_{q\times p}$, then, under A_n, $T_P \xrightarrow{d} \chi^2_p(\mu)$ as $n \to \infty$, where $\mu = h^\top F_{22} h$.*

Theorem 2 and Corollary 2 allow assessing the power of the T_P-test under local alternatives.

3 Applications and Numerical Examples

The current section shows some interesting applications of the general results of Section 2, which refer to cases (i) and (ii), respectively, of Corollaries 1 and 2.

3.1. Stationary time series models. The first example applies the portmanteau test in the context of general linear processes. We consider an interesting structure for the innovation process as in Taniguchi and Amano (2009), and illustrate a case when condition (i) of Corollaries 1 and 2 holds. Suppose that $\{X_t : t \in \mathbb{Z}\}$ is a stationary linear process with a spectral density function $f_\theta(\omega)$ of the form

$$f_\theta(\omega) = g_{\theta_1}(\omega)\sigma_u^2/2\pi \left\{ \sum_{j=-p}^{p} \psi_j \exp(-\mathbf{i}j\omega) \right\}, \tag{3.1}$$

where $\mathbf{i}$ denotes the imaginary unit, $\sigma_u > 0$, $\psi_0 = 1$ and $\psi_{-j} = \psi_j$ for $j = 1, \ldots, p$. Moreover, suppose $g_{\theta_1}(\omega)$ is given by

$$g_{\theta_1}(\omega) = 1/2\pi \exp\left\{ \sum_{j=1}^{q} \theta_{1,j} \cos(j\omega) \right\},$$

that is $g_{\theta_1}(\omega)$ is an exponential spectral density (Bloomfield 1973). We assume that the parameter $\psi = (\psi_1, \ldots, \psi_p)^\top$ is constrained to belong to

$$\Theta_2 = \left\{ \psi \in I\!R^p : \min_{\omega \in [-\pi,\pi)} \left(1 + 2\sum_{j=1}^{p} \psi_j \cos(j\omega) \right) \geq e \right\} \quad (e > 0 \text{ is a constant})$$

such that $f_\theta(\omega)$ does not take zero or negative value. Now, let us consider the following hypotheses

$$H : \psi = 0_p \text{ against } A : \psi \neq 0_p. \tag{3.2}$$

The purpose is to test whether the innovation process has null serial correlations, since under H the spectral density function coincides with that of a linear process generated from uncorrelated innovations. Under the alternative A, instead, $f_\theta(\omega)$ is the spectral density function of a process generated by a p-dependent innovation sequence with autocovariance ψ_j at lag j, $-p \leq j \leq p$. The focus here is on the correlation of the innovation sequence (and not on the correlation of the whole process).

There is an additional reason of interest on the test for the hypotheses (3.2) related to model (3.1), since as shown in Remark 3, the proposed test statistic T_P is asymptotically equivalent to the classical Box-Pierce test statistic (1.1). This result is consistent with the nature of the test which deals with the serial correlation of the innovation process.

Notice that, by putting $\theta_2 = \psi$, the hypotheses in Eq. 3.2 coincide with those in Eq. 2.1 of Section 2. The Fisher information matrix can be expressed in terms of the spectral density function of the process; namely,

$$F = 1/4\pi \int_{-\pi}^{\pi} \partial \log f_\theta(\omega)/\partial\theta \partial \log f_\theta(\omega)/\partial\theta^\top \Big|_{\theta=(\theta_1^\top, 0_p^\top)^\top} d\omega.$$

By simple algebra, we obtain $F_{11} = 1/4 I_q$, $F_{22} = I_p$ and

$$F_{12} = \begin{pmatrix} 2^{-1} I_q & O_{q\times(p-q)} \end{pmatrix}.$$

Therefore, $F_{21}F_{11}^{-1}F_{12}$ is given by

$$F_{21}F_{11}^{-1}F_{12} = \begin{pmatrix} I_q & O_{q\times(p-q)} \\ O_{(p-q)\times q} & O_{(p-q)\times(p-q)} \end{pmatrix}.$$

This implies that $F_{21}F_{11}^{-1}F_{12}$ is idempotent with $\text{rank}(F_{21}F_{11}^{-1}F_{12}) = q$. Thus, the conditions for (i) of Corollary 1 are satisfied in this case, and hence T_P converges to a χ^2_{p-q} r.v. without need of any modification. Furthermore, under the local alternatives

$$A_n : \psi = n^{-1/2} h,$$

T_P converges to a $\chi^2_{p-q}(\mu)$ r.v. with $\mu = h_{q+1}^2 + \cdots h_p^2$ (see (i) of Corollary 2).

Remark 3. *In the context of model* (3.1)*, an interesting point to be explored is the relationship between the* T_P *statistic and the Box-Pierce statistic* (1.1) *when the hypotheses in* Eq. 3.2 *are tested. To avoid unnecessarily complicated notations and discussion, in what follows, we restrict ourselves to the case when the process is Gaussian. Under the Gaussian assumption, it is shown that the log-likelihood is approximated by*

$$D(\theta) := -1/4\pi \int_{-\pi}^{\pi} \{\log f_\theta(\omega) + I_n(\omega)/f_\theta(\omega)\} \, d\omega$$

(e.g., Taniguchi and Kakizawa 2000, Section 7.2). The proposed test statistic T_p *admits, under* H*, the stochastic expansion*

$$T_P = nV_n(\theta_2^0)^\top F_{22}^{-1} V_n(\theta_2^0) + o_p(1), \tag{3.3}$$

where

$$V_n(\theta_2) = \partial D(\hat{\theta}_1, \theta_2)/\partial\theta_2,$$

$f_\theta(\omega)$ *is defined in* Eq. 3.1 *and* $I_n(\omega)$ *is the periodogram defined as*

$$I_n(\omega) = 1/2\pi n \left| \sum_{t=1}^{n} X_t \exp(\mathrm{i}t\omega) \right|^2$$

(For details of the expansion (3.3)*, see Proof of Theorem 1). Then, as illustrated by Taniguchi and Amano (*2009*, pp. 186-188), we have*

$$T_P = n \sum_{k=1}^{p} \hat{r_k^2} + o_p(1),$$

where $\hat{r}_k$ *is the residual empirical autocorrelation at lag* k *based on* $\hat{\theta}_1$*. That is* $T_P = T_{BP} + o_p(1)$ *and hence* $T_P = T_{LB} + o_p(1)$*.*

3.2. Serial correlation in linear regression models. The second example deals with the linear regression model (2.2) of Example 1. The focus is testing whether the noise ϵ is uncorrelated through the hypotheses (2.3) which, as mentioned before, coincide with Eq. 2.1 of Section 2.

It is easily shown that $\det[\Sigma_n(\xi, b)] = \xi^n(1 - b^2)^{n-1}$. Hence the log-likelihood function $l(\theta_1, \theta_2)$ is given by

$$l(\theta_1, \theta_2) = -1/2 \log(2\pi) - n - 1/2n \log(1 - b^2) - 1/2 \log \xi - Q_n(\beta, \xi, b),$$

where

$$Q_n(\beta, \xi, b) = 1/n(1 - b^2)\xi \left\{ 1/2 \sum_{i=1}^{n} u_i(\beta)^2 + b \sum_{i=1}^{n-1} u_i(\beta)u_{i+1}(\beta) + b^2/2 \sum_{i=2}^{n-1} u_i(\beta)^2 \right\}$$

and $u_i(\beta)$ is the ith component of $y - Z\beta$. The maximum likelihood estimator of $\theta_1 = (\beta^\top, \xi)^\top$ under H is

$$\hat{\theta}_1 = \begin{pmatrix} (Z^\top Z)^{-1} Z^\top y \\ n^{-1}\{y - Z(Z^\top Z)^{-1} Zy\}^\top \{y - Z(Z^\top Z)^{-1} Zy\} \end{pmatrix}.$$

Given Z and y, the quantity $l\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\}$ can be computed as $\sup_{\theta_2 \in \Theta_2} l(\hat{\theta}_1, \theta_2)$. Since the exact form of $l\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\}$ is complicated, the details are omitted here. However, we have

$$nE_{\theta^0}\left\{\partial l(\theta_1, \theta_2)/\partial\theta \partial l(\theta_1, \theta_2)/\partial\theta^\top\right\} = \begin{pmatrix} (n\xi)^{-1} Z^\top Z & O_{m\times 1} & O_{m\times 1} \\ O_{1\times m} & (2\xi^2)^{-1} & 0 \\ O_{1\times m} & 0 & (1 - n^{-1}) \end{pmatrix}.$$

Then, $F_{12} = O_{q\times p}$, where $q = \dim(\theta_1) = m + 1$ and $p = \dim(\theta_2) = 1$. Therefore, from (ii) of Corollary 1, it follows that T_P converges to a χ_1^2 r.v. Remarkably the condition $q < p$ is not required in this case.

Next we focus on the local power of the T_P-test. Consider the sequence of local alternatives

$$A_n : b = n^{-1/2}h, \tag{3.4}$$

so that $T_P \xrightarrow{d} \chi_1^2(h^2)$ as $n \to \infty$ by (ii) of Corollary 2. In this case, $F_{22\cdot 1} = F_{22}$ does not depend on F_{11}, and the theoretical local power of the T_P-test, when the significance level is 5%, is given by

$$LP_h^{95} = \int_{p_{95}}^{\infty} f_{NC}(x; 1, h^2)dx,$$

where p_{95} is the 95th-percentile of the $\chi_1^2(0)$ distribution and $f_{NC}(x; 1, \mu)$ is the probability density function of a $\chi_1^2(\mu)$ r.v. This result is quite natural since the parameters θ_1 and θ_2 are orthogonal to each other (off-diagonal blocks of the Fisher information matrix are zero). Table 1 shows the theoretical power of the T_P-test when the local alternatives (3.4) are considered for various h. It can be appreciated that the power increases remarkably when the value of h under the alternative hypothesis moves away from 0.

The rest of this Section compares the finite sample performance of the test statistics T_P, T_{BP} and T_{LB} by computing the empirical type-I error rates under $H : b = 0$ and the powers under the alternatives $A : b = 0.1, 0.3, 0.5$.

We consider model (2.2) with the design matrix $Z = (i^j : 1 \leq i \leq n$ and $j = 0, 1)$ (i.e. a linear trend). The nominal significance level of the test is 0.05 and the true values of the nuisance parameters are $\beta = (0, 1)^\top$ and $\xi = 1$. The maximum number of lags for T_{BP} and T_{LB} is $M = 30$. All the simulation results (reported in Table 2) are based on 100·000 samples of various sizes ($n = 50, 200, 400, 600$).

From Table 2, we observe that the empirical type-I error rate of the T_P-test is fairly close to the nominal level even for small sample sizes. By comparison, the T_{BP}-test is definitely less accurate, while the simulated type-I error rate of the T_{LB}-test is remarkably larger than the nominal level.

Table 1: Theoretical power of the T_P-test with local alternatives (3.4)

$h = 1$	$h = 2$	$h = 3$
0.1701	0.5160	0.8508

Table 2: Empirical type-I error rates and powers of T_P, T_{BP} and T_{LB}-tests

$H: b = 0.0$			
n	T_P	T_{BP}	T_{LB}
50	0.05544	0.01343	0.12666
200	0.05088	0.04914	0.09146
400	0.05171	0.05726	0.08012
600	0.05073	0.06037	0.07608
$A: b = 0.1$			
50	0.18667	0.01702	0.14419
200	0.76158	0.08800	0.14576
400	0.97372	0.14727	0.18432
600	0.99775	0.20951	0.23816
$A: b = 0.3$			
50	0.97270	0.08250	0.32064
200	1.00000	0.63464	0.71004
400	1.00000	0.95705	0.96491
600	1.00000	0.99788	0.99830
$A: b = 0.5$			
50	1.00000	0.34752	0.66214
200	1.00000	0.99545	0.99714
400	1.00000	1.00000	1.00000
600	1.00000	1.00000	1.00000

Furthermore, the power of the T_P-test rapidly increases, and it systematically outperforms – by a large amount – the alternative tests.

4 Proof of Theorems

4.1. Proof of Theorem 1. Throughout this section, we adopt the notation

$$\partial_i l(\theta_1, \theta_2) = \partial l(\eta_1, \eta_2)/\partial \eta_i |_{(\eta_1,\eta_2)=(\theta_1,\theta_2)}$$

and

$$\partial_{ij} l(\theta_1, \theta_2) = \partial^2 l(\eta_1, \eta_2)/\partial \eta_i \partial \eta_j^\top \Big|_{(\eta_1,\eta_2)=(\theta_1,\theta_2)}$$

for $i, j = 1, 2$. Note that $\hat{\theta}_2(\hat{\theta}_1)$ is the maximizer of $l(\hat{\theta}_1, \theta_2)$, hence we have

$$\partial_2 l\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\} = \partial l(\hat{\theta}_1, \theta_2)/\partial \theta_2 \Big|_{\theta_2=\hat{\theta}_2(\hat{\theta}_1)} = 0_p. \tag{4.1}$$

By Eq. 4.1 and expanding $l(\theta_1, \theta_2^0)$ around $\hat{\theta}_2(\hat{\theta}_1)$, it yields

$$\begin{aligned} T_P &= 2n\left[l\left\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\right\} - l\left(\hat{\theta}_1, \theta_2^0\right)\right] \\ &= 2n\left[\partial_2 l\left\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\right\}^\top \left\{\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\right\}\right. \\ &\quad \left. -1/2\left\{\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\right\}^\top \partial_{22} l(\hat{\theta}_1, \tilde{\theta}_2^*)\left\{\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\right\}\right] \\ &= n\left\{\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\right\}^\top \left\{-\partial_{22} l(\hat{\theta}_1, \tilde{\theta}_2^*)\right\}\left\{\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\right\}, \end{aligned} \tag{4.2}$$

where $\tilde{\theta}_2^*$ is an intermediate point between $\hat{\theta}_2(\hat{\theta}_1)$ and θ_2^0. To derive the asymptotic distribution of $n^{1/2}\{\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\}$, we expand $\partial_2 l\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\}$ around θ_2^0 and obtain

$$\begin{aligned} \partial_2 l\left\{\hat{\theta}_1, \hat{\theta}_2(\hat{\theta}_1)\right\} &= \partial_2 l(\hat{\theta}_1, \theta_2^0) + \partial_{22} l(\hat{\theta}_1, \theta_2^0)\left\{\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\right\} \\ &\quad + O_p(\|\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\|^2). \end{aligned} \tag{4.3}$$

Eqs. 4.3 and 4.1 lead to

$$\begin{aligned} \hat{\theta}_2(\hat{\theta}_1) - \theta_2^0 &= -\partial_{22} l(\hat{\theta}_1, \theta_2^0)^{-1} \partial_2 l(\hat{\theta}_1, \theta_2^0) + O_p(\|\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\|^2) \\ &= F_{22}^{-1} \partial_2 l(\hat{\theta}_1, \theta_2^0) + O_p(\|\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0\|^2) \end{aligned} \tag{4.4}$$

and similarly,

$$\hat{\theta}_1 - \theta_1 = F_{11}^{-1} \partial_1 l(\theta_1, \theta_2^0) + O_p(\|\hat{\theta}_1 - \theta_1\|^2). \tag{4.5}$$

On the other hand, an expansion of $\partial_2 l(\hat{\theta}_1, \theta_2^0)$ around θ_1 yields

$$\begin{aligned} \partial_2 l(\hat{\theta}_1, \theta_2^0) &= \partial_2 l(\theta_1, \theta_2^0) + \partial_{21} l(\theta_1, \theta_2^0)(\hat{\theta}_1 - \theta_1) + O_p(\|\hat{\theta}_1 - \theta_1\|^2) \\ &= \partial_2 l(\theta_1, \theta_2^0) - F_{21}(\hat{\theta}_1 - \theta_1) + O_p(\|\hat{\theta}_1 - \theta_1\|^2). \end{aligned} \tag{4.6}$$

Substituting Eq. 4.5 into Eq. 4.6 gives

$$\partial_2 l(\hat{\theta}_1, \theta_2^0) = \partial_2 l(\theta_1, \theta_2^0) - F_{21} F_{11}^{-1} \partial_1 l(\theta_1, \theta_2^0) + O_p(\|\hat{\theta}_1 - \theta_1\|^2),$$

and then Eq. 4.4 becomes

$$\hat{\theta}_2(\hat{\theta}_1) - \theta_2^0 = F_{22}^{-1}\left\{\partial_2 l(\theta_1, \theta_2^0) - F_{21} F_{11}^{-1} \partial_1 l(\theta_1, \theta_2^0)\right\} + \text{(lower order terms)}. \tag{4.7}$$

Furthermore, from (iii) of Assumption 1, the third and higher order cumulants of the score $n^{1/2}\partial l(\theta_1, \theta_2^0)/\partial\theta$ vanish as $n \to \infty$. Then, the following asymptotic normality holds:

$$n^{1/2} \begin{pmatrix} \partial_1 l(\theta_1, \theta_2^0) \\ \partial_2 l(\theta_1, \theta_2^0) \end{pmatrix} \xrightarrow{d} N(0, F).$$

Hence, by Eq. 4.2, Eq. 4.7 and the asymptotic normality of the score under H, the following asymptotic expansion is obtained for T_P:

$$T_P = N^\top F_{22\cdot 1}^{1/2} F_{22}^{-1} F_{22\cdot 1}^{1/2} N + o_p(1), \tag{4.8}$$

where N is a p-dimensional standard normal random vector and $F_{22\cdot 1} = F_{22} - F_{21}F_{11}^{-1}F_{12}$.

4.2. Proof of Corollary 1. First, suppose that $F_{22} = I_p$. Then, a necessary and sufficient condition which ensures that $N^\top F_{22\cdot 1}^{1/2} F_{22}^{-1} F_{22\cdot 1}^{1/2} N$ has an asymptotic χ^2 distribution is that $F_{22\cdot 1}^{1/2} F_{22}^{-1} F_{22\cdot 1}^{1/2}$ is idempotent (see Rao 1973, p.186). This condition is equivalent to the idempotence of $F_{21}F_{11}^{-1}F_{12}$ when $F_{22} = I_p$, and the degrees of freedom of the limit distribution are given by

$$\begin{aligned} \operatorname{rank}(F_{22\cdot 1}) &= \operatorname{rank}(I_p - F_{21}F_{11}^{-1}F_{12}) \\ &= \operatorname{tr}(I_p - F_{21}F_{11}^{-1}F_{12}) \\ &= p - \operatorname{tr}(F_{21}F_{11}^{-1}F_{12}) \\ &= p - \operatorname{rank}(F_{21}F_{11}^{-1}F_{12}). \end{aligned}$$

Notice that the condition $q < p$ is needed to avoid degeneracy of the limit distribution.

Second, we prove assertion (ii) of Theorem 1. By the definition, $F_{22\cdot 1} = F_{22}$ when $F_{12} = O_{q\times p}$, and Eq. 4.8 becomes $T_P = N^\top N + o_p(1)$, which has an asymptotic χ_p^2 distribution.

4.3. Proof of Theorem 2 and Corollary 2. Under Assumption 2, expanding $l\{\theta^{(n)}\}$ around θ^0 and solving gives

$$\Lambda_n \left\{\theta^0, \theta^{(n)}\right\} = n^{1/2}\partial_2 l(\theta_1, \theta_2^0)^\top h - 1/2 h^\top F_{22} h + o_p(1).$$

By recalling (4.7), we can see that

$$\begin{pmatrix} n^{1/2}\left\{\hat\theta_2(\hat\theta_1) - \theta_2^0\right\} \\ \Lambda_n(\theta^0, \theta^{(n)}) \end{pmatrix} \xrightarrow{d} N\left[\begin{pmatrix} 0_p \\ -1/2h^\top F_{22}h \end{pmatrix}, \begin{pmatrix} F_{22}^{-1}F_{22\cdot 1}F_{22}^{-1} & \sigma_A \\ \sigma_A^\top & h^\top F_{22} h \end{pmatrix}\right],$$

where $\sigma_A = F_{22}^{-1} F_{22\cdot 1} h$. From LeCam's third lemma, we have

$$n^{1/2} \left\{ \hat{\theta}_2(\hat{\theta}_1) - \theta_2^0 \right\} \xrightarrow{d} N \left(\sigma_A, F_{22}^{-1} F_{22\cdot 1} F_{22}^{-1} \right) \tag{4.9}$$

under the contiguous alternatives A_n. Therefore, under A_n, T_P can be written as

$$T_P = \left(N + F_{22\cdot 1}^{1/2} h \right)^{\top} F_{22\cdot 1}^{1/2} F_{22}^{-1} F_{22\cdot 1}^{1/2} \left(N + F_{22\cdot 1}^{1/2} h \right) + o_p(1).$$

From the same argumentin the proof of Theorem 1 and Corollary 1, we have the desired results.

Acknowledgements. Authors thank the Associate Editor and an anonymous Referee for their constructive comments on an earlier version of the paper. The third author thanks the Research Institute for Science & Engineering, Waseda University, for their supports.

References

BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217–226.

BOX, G.E. and PIERCE, D.A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* **65**, 1509–1526.

BREUSCH, T.S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers* **17**, 334–355.

BRILLINGER, D.R. (1981). Time Series: Data Analysis and Theory, Vol. 36 SIAM.

CHAN, N.H. and TRAN, L.T. (1992). Nonparametric tests for serial dependence. *Journal of Time Series Analysis* **13**, 19–28.

DURBIN, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. Econometrica pp. 410–421.

GODFREY, L.G. (1976). Testing for serial correlation in dynamic simultaneous equation models. Econometrica pp. 1077–1084.

LI, W.K. (2003). Diagnostic checks in time series. Chapman and Hall.

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

MCLEOD, A.I. and LI, W.K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* **4**, 269–273.

MONTI, A.C. (1994). A proposal for a residual autocorrelation test in linear models. *Biometrika* **81**, 776–780.

PẼNA, D. and RODRÍGUEZ, J. (2002). A powerful portmanteau test of lack of fit for time series. *Journal of the American Statistical Association* **97**, 601–610.

RAO, C.R. (1973). Linear statistical inference and its applications. Wiley Series in Probability and Mathematical Statistics (EUA).

TANIGUCHI, M. and KAKIZAWA, Y. (2000). *Asymptotic theory of statistical inference for time series*. Springer, New York.

TANIGUCHI, M. and AMANO, T. (2009). Systematic approach for portmanteau tests in view of whittle likelihood ratio. *Journal of the Japan Statistical Society* **39**, 177–192.

TZIRITAS, G. (1987). On the distribution of positive-definite Gaussian quadratic forms. *IEEE Transactions on Information Theory* **33**, 6, 895–906.

FUMIYA AKASHI
HIROAKI ODASHIMA
MASANOBU TANIGUCHI
DEPARTMENT OF PURE AND APPLIED
MATHEMATICS, SCHOOL OF FUNDAMENTAL
SCIENCE AND ENGINEERING,
WASEDA UNIVERSITY,
3-4-1 OKUBO, SHINJUKU-KU,
TOKYO 169-8555, JAPAN
E-mail: f.akashi@aoni.waseda.jp
aki-waseda.re116@akane.waseda.jp
taniguchi@waseda.jp

ANNA CLARA MONTI
DEPARTMENT OF LAW, ECONOMICS,
MANAGEMENT AND QUANTITATIVE
METHODS,
UNIVERSITY OF SANNIO,
82100 BENEVENTO, ITALY
E-mail: acmonti@unisannio.it

Paper received: 25 January 2016.