

# Diagnostic checking using subspace methods

Alfredo García-Hiernaux\*

January 26, 2009

## Abstract

The problem of diagnostic checking is tackled from the perspective of the subspace methods. Two statistics are presented and its asymptotic distributions are derived under the null. The procedures generalize the Box-Pierce statistic for single series and the Hoskings' statistic in the multivariate case. The performance of the proposals is illustrated via Monte Carlo simulations and an example with real data.

**Keywords:** Diagnostic checking, portmanteau tests, subspace methods

**JEL Classification:**

## 1 Introduction

Since the seminal work by Box and Pierce (1970) many studies have focused in the portmanteau tests and its ability to determine the adequacy of a model. Almost forty years later the subject is still open and many alternatives to this procedure, or the enhanced version by Ljung and Box (1978), are being proposed. Lately

---

\*Corresponding author. Universidad Complutense de Madrid, email: agarciah@ccee.ucm.es, tel: (+34) 91 394 25 11. The author gratefully acknowledges financial support from Ministerio de Educación y Ciencia, ref. ECO2008-02588/ECON.

the literature about diagnostic checking can be divided into two broad categories: i) those papers which relax assumptions of the original test, and ii) those which present some improvements in its finite properties.

Among others, the first group includes works by Lobato (2001) or Francq et al. (2005), who relax the assumption of independence of the errors, Jung and Tremayne (2003), suggesting a test for specific models of counts, Duchesne and Roy (2004), that generalize the test to multivariate autoregressive models with exogenous variables or, more recently, Escanciano and Lobato (2009) who design an automatic data-driven test, particularly suitable for financial data. On the other hand, the second class of methods is currently dominated by the use of bootstrap techniques which usually outperform the results in empirical size and power of the original proposal, see, *e.g.*, Horowitz et al. (2006).

The tests suggested in this paper belong to the previous second group as they do not relax any assumption of the original test, although they are not based on bootstrap methods either. However, they present some properties that are difficult to find in a single test, as: (i) they generalize the Box-Pierce statistic for single series and the Hoskings' statistic (Hosking, 1980) for multivariate processes, (ii) its asymptotic null distribution is known and hence, in comparison with bootstrap methods, there is no need of computationally expensive simulations to estimate the corresponding critical values, and (iii) they are, by construction, more robust in the presence of outliers that occur in the beginning or the end of the sample. The proposals are obtained by tackling the question from a subspace methods perspective.

The plan of the paper is as follows. Section 2 contains details of the subspace methods and the assumptions employed. Two general tests are derived in Section 3. Section 4 compares the performance of the proposals with the Ljung-Box test using Monte Carlo experiments and an application to real data. Some concluding remarks are offered in Section 5.

To express the results precisely, we introduce the following notation which will be used throughout the paper:  $\xrightarrow{d}$  means *converges in distribution to*,  $\xrightarrow{a.s.}$  means *converges almost surely to* and *plim* means *convergence in probability*. These three concepts are defined, e.g., in White (2001). Furthermore,  $\mathbf{I}_n$  will be a  $n$ -dimensional identity matrix and  $\mathbf{A}_m$  a square  $m$ -by- $m$  matrix, unless defined otherwise.

## 2 Model set and subspace estimation

Consider a linear fixed-coefficients system that can be described by the State Space (SS) model,

$$\mathbf{x}_{t+1} = \mathbf{\Phi}\mathbf{x}_t + \mathbf{E}\boldsymbol{\psi}_t \quad (1a)$$

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \boldsymbol{\psi}_t \quad (1b)$$

where  $\mathbf{x}_t$  is a state  $n$ -vector, being  $n$  the true order of the system. In addition,  $\mathbf{z}_t$  is an observable output  $m$ -vector, which is assumed to be zero-mean without loss of generality,  $\boldsymbol{\psi}_t$  is a noise  $m$ -vector (known as innovations), while  $\mathbf{\Phi}$ ,  $\mathbf{E}$  and  $\mathbf{H}$  are parametric matrices. Model (1a-1b) is called an “innovations model” and is used as it is simple and general, in the sense that any fixed-coefficients SS model can be written in this specific form (see e.g., Casals et al., 1999, Theorem 1). Moreover, some assumptions about the system and the noise must be established.

**Assumptions A.1.** Let  $\boldsymbol{\psi}_t$  be a sequence of independent and identically distributed random variables with  $E(\boldsymbol{\psi}_t) = 0$  and  $E(\boldsymbol{\psi}_t'\boldsymbol{\psi}_t) = \mathbf{Q}$ , being  $\mathbf{Q}$  a positive definite matrix. **A.2.** Let the system be stable and strictly minimum-phase, i.e., all the eigenvalues of  $\mathbf{\Phi}$  and  $(\mathbf{\Phi} - \mathbf{E}\mathbf{H})$  lie inside the unit circle.

Now we will show that the subspace methods can derive from the innovations

model. By substituting (1b) into (1a) in  $\boldsymbol{\psi}_t$  and solving by recursion we have:

$$\boldsymbol{x}_t = (\boldsymbol{\Phi} - \boldsymbol{E}\boldsymbol{H})^t \boldsymbol{x}_0 + \sum_{j=1}^t (\boldsymbol{\Phi} - \boldsymbol{E}\boldsymbol{H})^{t-j} \boldsymbol{E} \boldsymbol{z}_{j-1} \quad (2)$$

so that the states in time  $t$  depend on the initial state and past values of the output. We will use this equation afterward.

On the other hand, by recursive substitution in (1a) and replacing the result into the observation equation (1b), we get:

$$\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{\Phi}^t \boldsymbol{x}_0 + \boldsymbol{H} \sum_{j=0}^{t-1} \boldsymbol{\Phi}^j \boldsymbol{E} \boldsymbol{\psi}_{t-j-1} + \boldsymbol{\psi}_t \quad (3)$$

which means that the endogenous variable,  $\boldsymbol{z}_t$ , depends on the initial state vector,  $\boldsymbol{x}_0$ , and past and present innovations,  $\boldsymbol{\psi}_t$ . Equation (3) can be written in matrix form as,

$$\boldsymbol{Z}_p = \boldsymbol{O}\boldsymbol{X}_0 + \boldsymbol{V}\boldsymbol{\Psi}_p \quad (4)$$

where the subscript  $p$  is an integer that denotes the dimension of the row space of  $\boldsymbol{Z}_p$ , see Bauer (2005) for more explanations about  $p$ . In the following, we will define the matrices in equation (4):

1) Block-Hankel Matrices (BHM), which dimensions are determined by the integers  $p$  and  $f$ , such that:

$$\boldsymbol{Z}_p = \begin{pmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 & \dots & \boldsymbol{z}_{T-p-f+1} \\ \boldsymbol{z}_2 & \boldsymbol{z}_3 & \dots & \boldsymbol{z}_{T-p-f+2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{z}_p & \boldsymbol{z}_{p+1} & \dots & \boldsymbol{z}_{T-f} \end{pmatrix}; \quad \boldsymbol{Z}_f = \begin{pmatrix} \boldsymbol{z}_{p+1} & \boldsymbol{z}_{p+2} & \dots & \boldsymbol{z}_{T-f+1} \\ \boldsymbol{z}_{p+2} & \boldsymbol{z}_{p+3} & \dots & \boldsymbol{z}_{T-f+2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{z}_{p+f} & \boldsymbol{z}_{p+f+1} & \dots & \boldsymbol{z}_T \end{pmatrix} \quad (5)$$

In (4),  $\boldsymbol{\Psi}_p$  is as  $\boldsymbol{Z}_p$  but with  $\boldsymbol{\psi}_t$  instead of  $\boldsymbol{z}_t$ . For simplicity, in the following we will assume that the dimension of the past and future information sets is the same, *i.e.*,  $p = f = i$ . Notice that both,  $\boldsymbol{Z}_p$  and  $\boldsymbol{\Psi}_p$  are  $im \times (T - 2im + 1)$

matrices. In order to make the notation simpler we adopt, throughout the paper, the nomenclature  $T_* = T - 2im + 1$  to denote the number of columns of any BHM.

2) The state sequence which is defined as  $\mathbf{X}_t = (\mathbf{x}_t \ \mathbf{x}_{t+1} \ \mathbf{x}_{t+2} \ \dots \ \mathbf{x}_{t+T-2i})$ . Specially, we will use the past and future state sequences, denoted, respectively, by  $\mathbf{X}_p = \mathbf{X}_0$  and  $\mathbf{X}_f = \mathbf{X}_i$ .

3) The Extended Observability matrix, which is:

$$\mathbf{O} = \left( \mathbf{H}' \ (\mathbf{H}\Phi)' \ (\mathbf{H}\Phi^2)' \ \dots \ (\mathbf{H}\Phi^{i-1})' \right)'_{im \times n} \quad (6)$$

4) The lower block triangular Toeplitz matrix, defined as:

$$\mathbf{V} = \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\mathbf{E} & \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Phi\mathbf{E} & \mathbf{H}\mathbf{E} & \mathbf{I}_m & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{i-2}\mathbf{E} & \mathbf{H}\Phi^{i-3}\mathbf{E} & \mathbf{H}\Phi^{i-4}\mathbf{E} & \dots & \mathbf{I}_m \end{pmatrix}_{im} \quad (7)$$

Given assumption A.2. and for large values of  $t$ , the first addend in equation (2) is negligible and  $\mathbf{X}_f$  is to a close approximation representable as a linear combination of the past of the output,  $\mathbf{M}\mathbf{Z}_p$ . Shifting time subscripts in (4) and substituting  $\mathbf{X}_f$  by  $\mathbf{M}\mathbf{Z}_p$  lead to,

$$\mathbf{Z}_f = \mathbf{O}\mathbf{M}\mathbf{Z}_p + \mathbf{V}\Psi_f \quad (8)$$

where  $\mathbf{Z}_f$ ,  $\mathbf{Z}_p$  and  $\Psi_f$  are as in (5), and  $\mathbf{O}$  and  $\mathbf{V}$ , respectively, as in (6) and (7).

Subspace methods estimate  $\mathbf{O}$ ,  $\mathbf{M}$  and  $\mathbf{V}$  in (8) by solving a reduced-rank weighted least square problem, as the product  $\mathbf{O}\mathbf{M}$ , which is an  $im$  square matrix, is of rank  $n < im$ . Then, an estimation of the parameter matrices in (1a-1b) can be obtained from  $\mathbf{O}$ ,  $\mathbf{M}$  and  $\mathbf{V}$ , see, *e.g.*, Katayama (2005).

### 3 Statistics and its distribution

Firstly, the following hypothesis of the data,  $\mathbf{z}_t$ , should be defined.

**Hypothesis** *Let the null hypothesis, hereafter  $H_0$ , be that there are no correlations up to lag order  $k$  and let the alternative hypothesis,  $H_1$ , be that there exist correlations up to lag order  $k$ .*

Moreover, hereafter we will consider  $i$  as a function of  $k$ , such that  $i$  will be the integer rounded toward infinity of  $(k + 1)/2$ . Nevertheless, this is for the sake of simplicity and the tests could be directly adapted to any suitable value of  $i$ , or even different values of  $p$  and  $f$ . From equation (8) and by denoting  $\boldsymbol{\beta} = \mathbf{O}\mathbf{M}$ , we can write:

$$\mathbf{Z}_f = \boldsymbol{\beta}\mathbf{Z}_p + \mathbf{V}\boldsymbol{\Psi}_f, \quad (9)$$

where  $\boldsymbol{\beta}$  is rank deficient. Ignoring this restriction, which is not of interest in our aim,  $\boldsymbol{\beta}$  can be estimated using least squares as:

$$\hat{\boldsymbol{\beta}} = \mathbf{Z}_f\mathbf{Z}_p'(\mathbf{Z}_p\mathbf{Z}_p')^{-1}. \quad (10)$$

#### 3.1 By exploiting an estimate of the $\boldsymbol{\beta}$ matrix

From the least square estimate of  $\boldsymbol{\beta}$  and by means of the *vec* operator, which stacks the columns of a matrix into a long vector, we state the following proposition:

**Proposition 1** *Under  $H_0$  and given assumptions A.1. and A.2., the covariance matrix of  $\text{vec}(\hat{\boldsymbol{\beta}}|\mathbf{Z}_p)$  can be formulated as  $\mathbf{H}^{-1}\mathbf{A}(\boldsymbol{\Omega} \otimes \mathbf{Q})\mathbf{A}'\mathbf{H}^{-1}$ , where  $\mathbf{A} = \mathbf{Z}_p \otimes \mathbf{I}_{im}$ ,  $\mathbf{H} = \mathbf{A}'\mathbf{A}$  and the structure of  $\boldsymbol{\Omega}$  is represented in Appendix A, equation (20).*

The proof is given in Appendix A.

In the univariate case, *i.e.* when  $m = 1$  and the noise covariance,  $Q$ , is a scalar, Kolmogorov's strong law of large numbers (see, White, 2001, Theorem 3.1) and

$H_0$  ensure that  $T_*\mathbf{H}^{-1} \xrightarrow{a.s.} Q^{-1}\mathbf{I}_{i^2}$  and  $T_*^{-1}Q\mathbf{A}'\mathbf{\Omega}\mathbf{A} \xrightarrow{a.s.} Q^2\bar{\mathbf{\Pi}}$ , where  $\bar{\mathbf{\Pi}}$  presents the following structure:

$$\bar{\mathbf{\Pi}} = \begin{pmatrix} \mathbf{I}_{im^2} & \mathbf{\Pi}_{i-1} & \mathbf{\Pi}_{i-2} & \dots & \mathbf{\Pi}_1 \\ \mathbf{\Pi}'_{i-1} & \mathbf{I}_{im^2} & \mathbf{\Pi}_{i-1} & \dots & \mathbf{\Pi}_2 \\ \mathbf{\Pi}'_{i-2} & \mathbf{\Pi}'_{i-1} & \mathbf{I}_{im^2} & \dots & \mathbf{\Pi}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Pi}'_1 & \mathbf{\Pi}'_2 & \mathbf{\Pi}'_3 & \dots & \mathbf{I}_{im^2} \end{pmatrix}_{(im)^2} \quad (11)$$

$$\text{with } \mathbf{\Pi}_{i-j} = \begin{pmatrix} \boldsymbol{\pi}_{i-j} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\pi}_{i-j} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\pi}_{i-j} \end{pmatrix}_{im^2} \quad \text{and } \boldsymbol{\pi}_{i-j} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{m(i-j)} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{im},$$

where  $j = 1, 2, \dots, i-1$ . In those cases, following Proposition 1,  $\sqrt{T_*}vec(\hat{\boldsymbol{\beta}}|\mathbf{Z}_p) \xrightarrow{d} N(\mathbf{0}, \bar{\mathbf{\Pi}})$ .

However, when  $m > 1$  and  $\mathbf{Q}$  is a square  $m$ -matrix, keeping the same distribution for  $vec(\hat{\boldsymbol{\beta}}|\mathbf{Z}_p)$  requires some standardization of the data. In these cases, under  $H_0$ , we have  $E(\mathbf{z}_t\mathbf{z}'_t) = E(\boldsymbol{\psi}_t\boldsymbol{\psi}'_t) = \mathbf{I}_T \otimes \mathbf{Q}$ . Thus, one can define  $\bar{\mathbf{z}}_t = \mathbf{z}_t\hat{\mathbf{Q}}^{-\frac{1}{2}}$ , where  $\hat{\mathbf{Q}} = T^{-1}\mathbf{z}'_t\mathbf{z}_t$ , such that under the null  $E(\bar{\mathbf{z}}_t\bar{\mathbf{z}}'_t) = E(\bar{\boldsymbol{\psi}}_t\bar{\boldsymbol{\psi}}'_t) = \mathbf{I}_{Tm}$ . Then, the results put forward in the univariate case can be generalized, since:

$$(i) \quad T_*\bar{\mathbf{H}}^{-1} \xrightarrow{a.s.} \mathbf{I}_{(im)^2}, \text{ where } \bar{\mathbf{H}} = (\bar{\mathbf{Z}}_p \otimes \mathbf{I}_{im})(\bar{\mathbf{Z}}'_p \otimes \mathbf{I}_{im}), \text{ and} \quad (12)$$

$$(ii) \quad T_*^{-1}\bar{\mathbf{A}}'(\mathbf{\Omega} \otimes \mathbf{I}_m)\bar{\mathbf{A}} \xrightarrow{a.s.} \bar{\mathbf{\Pi}}, \text{ where } \bar{\mathbf{A}} = (\bar{\mathbf{Z}}'_p \otimes \mathbf{I}_{im}), \quad (13)$$

being  $\bar{\mathbf{Z}}_p$  the BHM computed as in (5) but with (the standardized)  $\bar{\mathbf{z}}_t$  instead of the original  $\mathbf{z}_t$ .

Obviously,  $\bar{\mathbf{\Pi}}$  is not, in general, the identity matrix. In fact, it is only so for the specific case when  $i = 1$ . For  $i > 1$ , some elements of  $\hat{\boldsymbol{\beta}}$ , the parameter matrix estimated in (10) but using the standardized BHMs  $\bar{\mathbf{Z}}_p$  and  $\bar{\mathbf{Z}}_f$ , are (perfectly)

correlated. In both cases, the following proposition and corollary can be stated.

**Proposition 2** *For any random matrix  $\mathbf{A}$  such that  $\sqrt{T_*} \text{vec} \mathbf{A} \xrightarrow{d} N(\mathbf{0}, \bar{\mathbf{\Pi}})$ , there is an idempotent matrix  $\mathbf{P}_{(im)^2}$  of rank  $m^2k$ , such that:*

$$s(\mathbf{A}) = T_* \text{vec}(\mathbf{A})' \mathbf{P} \text{vec}(\mathbf{A}) \xrightarrow{d} \chi_{m^2k}^2$$

The proof is given in Appendix A.

**Corollary 1** *By applying Proposition 2, under  $H_0$  and being  $\hat{\beta} = \bar{\mathbf{Z}}_f \bar{\mathbf{Z}}_p' (\bar{\mathbf{Z}}_p \bar{\mathbf{Z}}_p')^{-1}$ , then  $s(\hat{\beta}) \xrightarrow{d} \chi_{m^2k}^2$  holds.*

### 3.2 By exploiting an estimate of the $\mathbf{O}$ matrix

A second intuitive idea is to use the information held in an estimate of matrix  $\mathbf{O}$ , defined in equation (6) and used in (8). Typically, subspace methods split  $\hat{\beta}$ , in order to estimate  $\mathbf{O}$  and  $\mathbf{M}$ . As we showed in Section 2, for a given system order  $n$ , subspace methods look for a  $n$ -rank approximation of  $\hat{\beta}$ . This is done by means of the singular value decomposition of  $\mathbf{W}\hat{\beta}$ , being  $\mathbf{W}$  some weighting matrix. Once again, as we are not worried about  $n$  and the rank restriction, one of the (infinite) possible decompositions of  $\mathbf{W}\hat{\beta}$  could be  $\mathbf{W}\hat{\mathbf{O}} = \mathbf{W}\mathbf{Z}_f \mathbf{Z}_p' (\mathbf{Z}_p \mathbf{Z}_p')^{-\frac{1}{2}}$  and  $\hat{\mathbf{M}} = (\mathbf{Z}_p \mathbf{Z}_p')^{-\frac{1}{2}}$ , which leads to the following proposition and corollary.

**Proposition 3** *Under  $H_0$  and given assumptions A.1. and A.2., if  $\mathbf{W} = (\mathbf{Z}_f \mathbf{Z}_f')^{-\frac{1}{2}}$ , then  $\sqrt{T_*} \text{vec}(\mathbf{W}\hat{\mathbf{O}} | \mathbf{Z}_p) \xrightarrow{d} N(\mathbf{0}, \bar{\mathbf{\Pi}})$ .*

The proof is given in Appendix A.

**Corollary 2** *By applying Proposition 2, under  $H_0$ , then  $s(\mathbf{W}\hat{\mathbf{O}}) \xrightarrow{d} \chi_{m^2k}^2$  holds.*

### 3.3 Univariate and multivariate ARMA residuals

One of the most common use of the portmanteau tests is to check the residuals obtained from fitting (vector) autoregressive moving-average, (V)ARMA, models. Here we will adopt the usual definition of a stationary  $m$ -variate ARMA( $p, q$ )

process (*e.g.*, see Peña et al., 2001, p. 368). When  $\mathbf{z}_t$  are the residuals from a VARMA model, the asymptotic distribution of  $s(\hat{\boldsymbol{\beta}})$  and  $s(\mathbf{W}\hat{\boldsymbol{\Omega}})$  is not as in Corollaries 1 and 2, respectively. The reason is that assumption A.1. does not hold, as residuals, contrary to innovations, present some linear constraints inherit from the VARMA estimation, see Box and Pierce (1970) for the univariate case or Hosking (1980) for the multivariate case. In these circumstances, the following proposition establishes the asymptotic distribution of both statistics.

**Proposition 4** *If  $\mathbf{z}_t$  in (1b) are the residuals from a fitted  $m$ -vector ARMA( $p, q$ ) model, then, under  $H_0$ ,  $s(\hat{\boldsymbol{\beta}})$  and  $s(\mathbf{W}\hat{\boldsymbol{\Omega}})$  converge in distribution to a  $\chi_{m^2(k-p-q)}^2$ .*

The proof is given in Appendix A.

Notice that checking the null hypothesis in any  $m$ -variate process requires, if the Ljung-Box test is used, the typical Q-matrix which leads to  $m^2$  different statistics. Our tests offer a more natural scalar statistic instead. Further, it is straightforward to see that for  $m = 1$ , both  $s(\hat{\boldsymbol{\beta}})$  and  $s(\mathbf{W}\hat{\boldsymbol{\Omega}})$  are equivalent to the Box-Pierce statistic if  $p = 1$  and  $f = k + 1$  in the BHM defined in (5). Analogously, when  $m > 1$ , the Hoskings' statistic is a particular case of the tests proposed. As in the previous case, they are equivalent if  $p = 1$  and  $f = k + 1$  (see Hosking, 1980, p. 605). In short, our proposals generalize Box-Pierce's and Hosking's procedures, since they allow for different values of  $p$ . Furthermore, this fact entails benefits in terms of robustness. As we have previously seen, if  $p > 1$  then  $\bar{\boldsymbol{\Pi}}$  is not the identity matrix, which means that some elements in  $s(\hat{\boldsymbol{\beta}})$  and  $s(\mathbf{W}\hat{\boldsymbol{\Omega}})$  are correlated. Consequently, those elements will be weighted (as it is shown in the proof of proposition 2) to obtain a vector of uncorrelated components. In this way, the effect of an outlier will be mitigated, provided that it only affects some, but not all, of the weighted correlations. This usually occurs with outliers located in the beginning or the end of the sample.

## 4 Numerical examples

In this section we investigate the finite sample properties of the proposed tests. All the results are compared with those obtained with the  $Q_{LB}$  statistic by Ljung-Box as it is the most common and cited diagnostic test in the literature. We will split the exercise into the analysis of some Monte Carlo simulations of univariate and bivariate processes and an application to real data.

### 4.1 Monte Carlo Simulations

In this first part we will study three aspects of the tests: 1) its empirical power, 2) how the chosen order of lags influences over its performance and 3) if its empirical size coincides with the designated significance level (nominal size).

[FIGURE 1 SHOULD BE AROUND HERE]

Figure 1 presents the empirical power of  $s(\mathbf{W}\hat{\mathbf{O}})$ ,  $s(\hat{\boldsymbol{\beta}})$  and  $Q_{LB}$ . The sample size is 50 observations and the data generating processes and lags employed are: i) an AR(1) with  $k = 1$  and  $k = 5$ , ii) a MA(1) with  $k = 5$ , and iii) an ARMA(1,1) with the autoregressive parameter fixed to  $-.8$  and  $k = 3$ . In the ARMA model, an AR(1) is misspecified and estimated so that the statistics computed from the residuals are distributed as a  $\chi_{k-1}^2$ . The most noticeable features are:

1. In the four cases, the empirical power achieved by  $s(\mathbf{W}\hat{\mathbf{O}})$  is higher or equal to the one shown by  $Q_{LB}$  and higher than the one exhibited by  $s(\hat{\boldsymbol{\beta}})$ . On the other hand, the three tests perform particularly badly with the ARMA(1,1) model when the moving average parameter is negative, specifically when the AR and MA parameters are close to be cancelled out.
2. When  $k = 5$  the empirical size of  $s(\mathbf{W}\hat{\mathbf{O}})$  and  $Q_{LB}$  is (statistically) higher than the nominal one. For the AR(1) model the behavior of  $s(\mathbf{W}\hat{\mathbf{O}})$  and  $Q_{LB}$  is almost identical, however for the MA(1) process  $Q_{LB}$  is clearly outperformed by our two proposals.

3.  $s(\hat{\beta})$  is the only test which holds an empirical size close to the nominal size in the four different situations.

On the other hand, Table 1 shows the empirical size and power of the three tests, calculated for nine VARMA bivariate models. In these cases, the  $\mathbf{Q}_{LB}$  is a matrix of statistics whereas  $s(\hat{\beta})$  and  $s(\mathbf{W}\hat{\mathbf{O}})$  remain being a scalar. The main conclusions are:

1. Both tests,  $s(\hat{\beta})$  and  $s(\mathbf{W}\hat{\mathbf{O}})$ , usually present a higher empirical power than the highest value obtained in the  $\mathbf{Q}_{LB}$  matrix (see models 2-5, 7 and 8).
2. The empirical size, which is reasonably good for all the tests when  $k = 1$ , becomes higher than the nominal size for  $\mathbf{Q}_{LB}$  when  $k$  increases (compare models 1 and 6).
3. Even if the empirical size is close to the nominal size for all the tests and  $k = 1$  when there are no correlations at all (see model 1), it is clearly higher in the elements of the  $\mathbf{Q}_{LB}$  matrix corresponding to zero-parameters when the data has some kind of structure (see models 2,3,6,7 and 8).

[TABLE 1 SHOULD BE AROUND HERE]

## 4.2 An example with real data

The second example deals with the logarithms of indices of monthly flour prices in the cities of Buffalo, Minneapolis and Kansas City, over the period from August 1972 to November 1980, which give us 100 observations at each site. The series have been previously studied in Tiao and Tsay (1989), Grubb (1992) and Lütkepohl and Poskitt (1996) among others. The aim of modelling these data is to illustrate the performance of the proposed test statistics, as specification tools, and compare it with the most commonly used  $Q_{LB}$ .

As the series all appear non-stationary, we use the log-difference transformation  $\mathbf{z}_t = \nabla \log(\mathbf{y}_t)$ , where  $\mathbf{y}_t$  are the original series and  $\nabla = 1 - B$  is the difference operator. Table 2 shows the results obtained by applying the statistics with different

lags to this process. The first conclusion is that even if all the tests suggest that there are significant correlations, at least up to order one, the  $Q$ -statistic presents very low power when a (not-so) large lag is chosen. This is because the fact that the significant correlations at lag one are diluted by insignificant correlations at other lags seems to be more important in  $Q_{LB}$  than in  $s(\hat{\beta})$  or  $s(\mathbf{W}\hat{\mathbf{O}})$ . In addition,  $Q_{LB}$  only reveals 5 out of 9 correlations statistically significant at 5% in lag one.

[TABLE 2 SHOULD BE AROUND HERE]

Following the results obtained with the  $Q_{LB}$  at a significance level of 5% in Table 2 with  $k = 1$ , a restricted VAR(1) model  $(\mathbf{I} - \Phi_1 B)\mathbf{z}_t = \mathbf{a}_t$  is tentatively specified. The model was fitted to the data by maximum likelihood, resulting:

$$\hat{\Phi}_1 = \begin{pmatrix} 0 & -.188^* & -.035 \\ 0 & -.289^* & 0 \\ -.401^* & .117 & 0 \end{pmatrix}, \quad \hat{\Sigma}_a = \begin{pmatrix} 2.263 & 2.296 & 2.202 \\ & 2.496 & 2.364 \\ & & 2.770 \end{pmatrix} \times 10^{-3} \quad (14)$$

where, in the parameter matrices, an entry constrained to be zero is denoted by ‘0’ and ‘\*’ means that the parameter is significant at a 5%. The results of using  $Q_{LB}$ ,  $s(\mathbf{W}\hat{\mathbf{O}})$  and  $s(\hat{\beta})$  as diagnostic tests over the residuals of this model are in Table 3.

[TABLE 3 SHOULD BE AROUND HERE]

The  $Q_{LB}$  suggests that there are no correlations for  $k = 2, 5, 10, 15$  at 10% level of significance, implying that model presented in (14) could be appropriate. However,  $s(\mathbf{W}\hat{\mathbf{O}})$  and  $s(\hat{\beta})$  reject the null hypothesis for  $k = 2$  and  $k = 2, 5, 10, 15$ , respectively, at 5% level of significance. So, may be the  $Q$ -statistic has lead to an inappropriate model. For instance, if we had followed our proposals in Table 2, we could have specified an unrestricted VAR(1). The estimation of this model

returns:

$$\hat{\Phi}_1 = \begin{pmatrix} 1.226^* & -1.355^* & .005 \\ .830^* & -1.027^* & .035 \\ .463 & -.813^* & .142 \end{pmatrix}, \quad \hat{\Sigma}_a = \begin{pmatrix} 2.033 & 2.140 & 2.039 \\ & 2.390 & 2.253 \\ & & 2.647 \end{pmatrix} \times 10^{-3} \quad (15)$$

To check if the estimated residual correlations of model (15) are approximately zero, the three tests are again used. The results are shown in Table 4. This time, none of the tests reject the null hypothesis that there are no correlations up to order  $k = 2, 5, 10, 15$ . However, notice that our proposals show more evidence in favour of the null than  $Q_{LB}$ , at least in the first lags. Finally, note that model (15) was also proposed by Lütkepohl and Poskitt (1996) and is better than many other alternatives, in particular model (14), as it was shown in Grubb (1992).

[TABLE 4 SHOULD BE AROUND HERE]

## 5 Concluding remarks

This work tackles the problem of diagnostic checking from an original point of view. Two statistics based on the subspace methods are presented and its asymptotic distributions are derived under the null. The procedures generalize the Box-Pierce statistic for single series and the Hoskings' statistic in the multivariate case. Moreover, they are more robust than the alternatives in the presence of outliers. Monte Carlo simulations and an example with real data show that our proposals perform better than the common Ljung-Box Q-statistic in many situations.

## References

- Bauer, D. (2005). Asymptotic properties of subspace estimators. *Automatica*, 41:359–376.
- Box, G. E. P. and Pierce, D. A. (1970). Distribution of residuals autocorrelations

- in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.
- Casals, J., Sotoca, S., and Jerez, M. (1999). A fast and stable method to compute the likelihood of time invariant state space models. *Economics Letters*, 65(3):329–337.
- Duchesne, P. and Roy, R. (2004). On consistent testing for serial correlation of unknown form in vector time series models. *Journal of Multivariate Analysis*, 89:148–180.
- Escanciano, J. C. and Lobato, I. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, forthcoming.
- Francq, C., Roy, R., and Zakoian, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association*, 100(13):532–544.
- Grubb, H. (1992). A multivariate time series analysis of some flour price data. *Applied Statistics*, 41:95–107.
- Horowitz, J. L., Lobato, I., Nankervis, J. C., and Savin, N. (2006). Bootstrapping the box-pierce q test: A robust test of uncorrelatedness. *Journal of Econometrics*, 133(2):841–862.
- Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371):602–608.
- Jung, R. C. and Tremayne, A. R. (2003). Testing for serial dependence in time series models of counts. *Journal of Time Series Analysis*, 24(1):65–84.
- Katayama, T. (2005). *Subspace Methods for System Identification*. Springer Verlag, London.
- Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:297–303.

- Lobato, I. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96:1066–1076.
- Lütkepohl, H. and Poskitt, D. S. (1996). Specification of echelon form VARMA models. *Journal of Business and Economic Statistics*, 14(1):69–79.
- Peña, D., Tiao, G. C., and Tsay, R. C. (2001). *A Course in Time Series Analyses*. Wiley Inter-science.
- Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society, B Series*, 51(2):157–213.
- White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press.

## Appendix A

**Proof of Proposition 1.** By applying the *vec* operator, equation (9) can be formulated as:

$$\text{vec}\mathbf{Z}_f = (\mathbf{Z}'_p \otimes \mathbf{I}_{im})\text{vec}\boldsymbol{\beta} + \text{vec}\boldsymbol{\Psi}_f, \quad (16)$$

where we have used that  $\mathbf{V} = \mathbf{I}_{im}$  under  $H_0$ . From this,  $\hat{\boldsymbol{\beta}}$  will be computed as,

$$\text{vec}\hat{\boldsymbol{\beta}} = [(\mathbf{Z}'_p \otimes \mathbf{I}_{im})(\mathbf{Z}'_p \otimes \mathbf{I}_{im})]^{-1}(\mathbf{Z}'_p \otimes \mathbf{I}_{im})'\text{vec}\mathbf{Z}_f, \quad (17)$$

or substituting equation (16) in  $\text{vec}\mathbf{Z}_f$  and rearranging,

$$\text{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{H}^{-1}\mathbf{A}'\text{vec}\boldsymbol{\Psi}_f, \quad (18)$$

where  $\mathbf{H} = \mathbf{A}'\mathbf{A}$  and  $\mathbf{A} = \mathbf{Z}'_p \otimes \mathbf{I}_{im}$ . Therefore, the covariance matrix of  $\text{vec}\hat{\boldsymbol{\beta}}$  conditional to  $\mathbf{Z}_p$  will be formulated as:

$$\text{cov}[\text{vec}\hat{\boldsymbol{\beta}}|\mathbf{Z}_p] = \text{E}[\mathbf{H}^{-1}\mathbf{A}'\text{vec}(\boldsymbol{\Psi}_f)\text{vec}(\boldsymbol{\Psi}_f)'\mathbf{A}\mathbf{H}^{-1}|\mathbf{Z}_p] \quad (19a)$$

$$= \mathbf{H}^{-1}\mathbf{A}'(\boldsymbol{\Omega} \otimes \mathbf{Q})\mathbf{A}\mathbf{H}^{-1}, \quad (19b)$$

where  $\Omega$ , which is not an identity matrix, presents the form:

$$\Omega = \begin{pmatrix} \mathbf{I}_{im} & \boldsymbol{\omega}'_{i-1} & \boldsymbol{\omega}'_{i-2} & \cdots & \boldsymbol{\omega}'_{i-(T^*-1)} \\ \boldsymbol{\omega}_{i-1} & \mathbf{I}_{im} & \boldsymbol{\omega}'_{i-1} & \cdots & \boldsymbol{\omega}'_{i-(T^*-2)} \\ \boldsymbol{\omega}_{i-2} & \boldsymbol{\omega}_{i-1} & \mathbf{I}_{im} & \cdots & \boldsymbol{\omega}'_{i-(T^*-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\omega}_{i-(T^*-1)} & \boldsymbol{\omega}_{i-(T^*-2)} & \boldsymbol{\omega}_{i-(T^*-3)} & \cdots & \mathbf{I}_{im} \end{pmatrix}_{imT^*}, \quad (20)$$

where

$$\boldsymbol{\omega}_{i-j} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{m(i-j)} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{im} \quad \text{and } j = 1, 2, \dots, T^* - 1. \quad (21)$$

Moreover, when  $j \geq i$ ,  $\boldsymbol{\omega}_{i-j}$  is an  $im$ -by- $im$  zero-matrix. This particular composition of  $\Omega$  is inherited from the structure of the BHM  $\Psi_f$ , defined in (5).

□

**Proof of Proposition 2.** Once the structure of the covariance matrix  $\bar{\Pi}$  is known, it is straightforward to see that not all the random elements in  $\mathbf{A}$  are independent, except when  $i = 1$ . Therefore, the aim is to find where the independent components are located. Given the structure of  $\bar{\Pi}$  and using the submatrix Matlab notation:

- (i) The first  $im$  elements of  $\text{vec}\mathbf{A}$ , which are  $\mathbf{A}_{1:im,1:m}$ , are uncorrelated as the square submatrix  $\bar{\Pi}_{1:im} = \mathbf{I}_{im^2}$ .
- (ii) As the first  $m$  rows of  $\bar{\Pi}'_{i-1}$  are zeros, then the elements of the submatrix  $\mathbf{A}_{1:m,m+1:m+2}$  are also uncorrelated with those of  $\mathbf{A}_{1:im,1:m}$ . In the same way, this occurs for every element in the submatrix  $\mathbf{A}_{1:m,m+1:im}$  due to the structure of zeros in matrices  $\bar{\Pi}'_{i-k}$  for  $k = 1, 2, \dots, i - 1$ . Then the elements in  $\mathbf{A}_{1:m,m+1:im}$  are uncorrelated with those of  $\mathbf{A}_{1:im,1:m}$  and, therefore,  $\bar{\Pi}$  is of rank  $m^2(2i - 1)$ .

In order to extract  $m^2k$  independent elements from  $\mathbf{A}$ , we use the singular value decomposition of  $\bar{\mathbf{\Pi}}$ , creating a matrix  $\mathbf{B}_{(im)^2 \times m^2k}$  such that  $\bar{\mathbf{\Pi}} = \mathbf{B}\mathbf{B}'$  as:

$$\bar{\mathbf{\Pi}} \stackrel{\text{svd}}{=} \underbrace{\mathbf{U}\mathbf{S}^{\frac{1}{2}}}_{\mathbf{B}} \underbrace{\mathbf{S}^{\frac{1}{2}}\mathbf{V}'}_{\mathbf{B}'} \quad (22)$$

$$= \mathbf{B} \cdot \mathbf{B}'. \quad (23)$$

Consequently, from (23) we have that  $\mathbf{B}^\dagger \bar{\mathbf{\Pi}} \mathbf{B}^\dagger = \mathbf{I}_{m^2k}$ , where the symbol ' $\dagger$ ' denotes the Moore-Penrose inverse. Therefore,  $\mathbf{B}^\dagger \text{vec}(\mathbf{A}) \xrightarrow{d} N(\mathbf{0}, T_*^{-1} \mathbf{I}_{m^2k})$  which leads directly to:

$$s(\mathbf{A}) = T_* \text{vec}(\mathbf{A})' \mathbf{P} \text{vec}(\mathbf{A}) \xrightarrow{d} \chi_{m^2k}^2$$

being  $\mathbf{P} = \mathbf{B}^\dagger \mathbf{B}^\dagger$  a weighting idempotent matrix with rank  $m^2k$  that averages the perfectly correlated elements of  $\text{vec}(\mathbf{A})$  in a vector of  $m^2k$  uncorrelated elements.

□

**Proof of Proposition 3.** Let us estimate  $\mathbf{W}\hat{\mathbf{O}}$  as:

$$\mathbf{W}\hat{\mathbf{O}} = \mathbf{W}\mathbf{Z}_f \mathbf{Z}'_p (\mathbf{Z}_p \mathbf{Z}'_p)^{-\frac{1}{2}} \quad (24)$$

As  $\mathbf{V} = \mathbf{I}_{im}$  under  $H_0$ , then (24) becomes:

$$\mathbf{W}\hat{\mathbf{O}} = \mathbf{W}(\mathbf{O}\mathbf{M}\mathbf{Z}_p + \mathbf{\Psi}_f) \mathbf{Z}'_p (\mathbf{Z}_p \mathbf{Z}'_p)^{-\frac{1}{2}} \quad (25)$$

and substituting  $\mathbf{M} = (\mathbf{Z}_p \mathbf{Z}'_p)^{-\frac{1}{2}}$  we obtain:

$$\mathbf{W}(\hat{\mathbf{O}} - \mathbf{O}) = \mathbf{W}\mathbf{\Psi}_f \mathbf{Z}'_p (\mathbf{Z}_p \mathbf{Z}'_p)^{-\frac{1}{2}}. \quad (26)$$

Applying the  $\text{vec}$  operator in the equation above, we have:

$$\text{vec}[\mathbf{W}(\hat{\mathbf{O}} - \mathbf{O})] = [((\mathbf{Z}_p \mathbf{Z}'_p)^{-\frac{1}{2}} \mathbf{Z}'_p) \otimes \mathbf{W}] \text{vec}\mathbf{\Psi}_f. \quad (27)$$

Thus, the covariance matrix of  $vec[\mathbf{W}(\hat{\mathbf{O}}-\mathbf{O})]$  conditional to  $\mathbf{Z}_p$  will be formulated as:

$$\mathbb{E}\left[\left[\left(\mathbf{Z}_p\mathbf{Z}'_p\right)^{-\frac{1}{2}}\mathbf{Z}_p\right]\otimes\mathbf{W}\right]vec\Psi_fvec\Psi'_f\left[\left(\mathbf{Z}'_p\left(\mathbf{Z}_p\mathbf{Z}'_p\right)^{-\frac{1}{2}}\right)\otimes\mathbf{W}\right]|\mathbf{Z}_p\right]. \quad (28)$$

By replacing  $\mathbf{W} = \left(\mathbf{Z}_p\mathbf{Z}'_p\right)^{-\frac{1}{2}}$  and using that, under  $H_0$ ,  $\mathbf{Z}_f|\mathbf{Z}_p = \mathbf{Z}_f$ , then equation (28) can be written as:

$$\left[\left(\mathbf{Z}_p\mathbf{Z}'_p\right)^{-\frac{1}{2}}\mathbf{Z}_p\right]\otimes\left(\mathbf{Z}_f\mathbf{Z}'_f\right)^{-\frac{1}{2}}\left(\boldsymbol{\Omega}\otimes\mathbf{Q}\right)\left[\left(\mathbf{Z}'_p\left(\mathbf{Z}_p\mathbf{Z}'_p\right)^{-\frac{1}{2}}\right)\otimes\left(\mathbf{Z}_f\mathbf{Z}'_f\right)^{-\frac{1}{2}}\right]. \quad (29)$$

Again under  $H_0$ ,  $\sqrt{T_*}\left(\mathbf{Z}_f\mathbf{Z}'_f\right)^{-\frac{1}{2}}\xrightarrow{a.s.}\mathbf{I}_i\otimes\mathbf{Q}^{-\frac{1}{2}}$  and  $\sqrt{T_*}\left(\mathbf{Z}_p\mathbf{Z}'_p\right)^{-\frac{1}{2}}\xrightarrow{a.s.}\mathbf{I}_i\otimes\mathbf{Q}^{-\frac{1}{2}}$  hold. Using the properties of the Kronecker product, we can finally write the asymptotic covariance matrix of  $vec(\mathbf{W}\hat{\mathbf{O}})$  as the following expression:

$$cov[vec(\mathbf{W}\hat{\mathbf{O}})]\xrightarrow{a.s.}T_*^{-2}\left[\left[\left(\mathbf{I}_i\otimes\mathbf{Q}^{-\frac{1}{2}}\right)\mathbf{Z}_p\right]\otimes\mathbf{I}_i\right]\boldsymbol{\Omega}\left[\left(\mathbf{Z}'_p\left(\mathbf{I}_i\otimes\mathbf{Q}^{-\frac{1}{2}}\right)\right)\otimes\mathbf{I}_i\right]\otimes\mathbf{I}_m. \quad (30)$$

On the other hand, from equations (12-13) the covariance matrix of  $vec(\hat{\boldsymbol{\beta}}|\mathbf{Z}_p)$  is computed as:

$$cov[vec(\hat{\boldsymbol{\beta}}|\mathbf{Z}_p)] = \bar{\mathbf{H}}^{-1}\left(\bar{\mathbf{Z}}_p\otimes\mathbf{I}_{im}\right)\left(\boldsymbol{\Omega}\otimes\mathbf{I}_m\right)\left(\bar{\mathbf{Z}}'_p\otimes\mathbf{I}_{im}\right)\bar{\mathbf{H}}'^{-1} \quad (31)$$

$$= T_*^{-1}\underbrace{\left(T_*\bar{\mathbf{H}}^{-1}\right)}_{a.s.}\underbrace{\left[T_*^{-1}\left[\left(\bar{\mathbf{Z}}_p\otimes\mathbf{I}_i\right)\boldsymbol{\Omega}\left(\bar{\mathbf{Z}}'_p\otimes\mathbf{I}_i\right)\right]\otimes\mathbf{I}_m\right]}_{a.s.}\underbrace{\left(T_*\bar{\mathbf{H}}'^{-1}\right)}_{a.s.} \quad (32)$$

$$\xrightarrow{a.s.} T_*^{-1}\cdot\mathbf{I}_{(im)^2}\cdot\bar{\boldsymbol{\Pi}}\cdot\mathbf{I}_{(im)^2} \quad (33)$$

$$= T_*^{-1}\bar{\boldsymbol{\Pi}} \quad (34)$$

Finally, as  $\lim_{T\rightarrow\infty}|\bar{\mathbf{Z}}_p-\left(\mathbf{I}_i\otimes\mathbf{Q}^{-\frac{1}{2}}\right)\mathbf{Z}_p| = \mathbf{0}$ , then both covariances, (30) and (31), tend asymptotically to  $T_*^{-1}\bar{\boldsymbol{\Pi}}$ .

□

**Proof of Proposition 4.** Let the  $r$ th autocovariance matrix of the innovations be  $\mathbf{C}_r = T^{-1}\boldsymbol{\psi}'_t\boldsymbol{\psi}_{t-r}$  and the  $r$ th residual autocovariance matrix be  $\hat{\mathbf{C}}_r = T^{-1}\hat{\boldsymbol{\psi}}'_t\hat{\boldsymbol{\psi}}_{t-r}$ . Further, define  $\mathbf{C} = (\mathbf{C}_1\ \mathbf{C}_2\ \dots\ \mathbf{C}_k)$  and similary  $\hat{\mathbf{C}}$ . Then, Hosking (1980) proved

that:

$$vec(\hat{\mathbf{C}}) = \mathbf{D}vec(\mathbf{C}), \quad (35)$$

where  $\mathbf{D}$  is idempotent of rank  $m^2(k - p - q)$ .

Let now define  $\hat{\boldsymbol{\beta}}_* = \bar{\mathbf{Z}}_f \bar{\mathbf{Z}}_p' (\bar{\mathbf{Z}}_p \bar{\mathbf{Z}}_p')^{-1}$  as the  $\boldsymbol{\beta}$  matrix estimated using the (standardized) residuals of a VARMA( $p, q$ ) model. As these residuals have been previously standardized (see Section 3.1),  $\hat{\boldsymbol{\beta}}_* \xrightarrow{a.s.} \hat{\mathbf{C}}(\mathbf{I}_i \otimes \mathbf{I}_m)^{-1} = \hat{\mathbf{C}}$  where

$$\hat{\mathbf{C}} = \begin{pmatrix} \hat{\mathbf{C}}_{\bar{k}-i+1} & \hat{\mathbf{C}}_{\bar{k}-i} & \cdots & \hat{\mathbf{C}}_1 \\ \hat{\mathbf{C}}_{\bar{k}-i+2} & \hat{\mathbf{C}}_{\bar{k}-i+1} & \cdots & \hat{\mathbf{C}}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{C}}_{\bar{k}} & \hat{\mathbf{C}}_{\bar{k}-1} & \cdots & \hat{\mathbf{C}}_{\bar{k}-i+1} \end{pmatrix}_{im} \quad \text{being } \bar{k} \equiv \begin{cases} k & \text{if } k \text{ is odd} \\ k + 1 & \text{if } k \text{ is even.} \end{cases} \quad (36)$$

Then, we can write  $\mathbf{B}^\dagger vec(\hat{\boldsymbol{\beta}}_*) = \bar{\mathbf{D}}\mathbf{B}^\dagger vec(\hat{\boldsymbol{\beta}})$  as it was done in equation (35), since  $\mathbf{B}^\dagger vec(\hat{\boldsymbol{\beta}}_*)$  and  $\mathbf{B}^\dagger vec(\hat{\boldsymbol{\beta}})$  have, asymptotically, the same elements than  $vec(\hat{\mathbf{C}})$  and  $vec(\mathbf{C})$ , respectively, although sorted in different order. Likewise,  $\bar{\mathbf{D}}$  has the same rows than  $\mathbf{D}$ , but ordered differently, and hence  $rank(\bar{\mathbf{D}}) = rank(\mathbf{D}) = m^2(k - p - q)$ .

Finally, we have previously shown that  $\mathbf{B}^\dagger vec(\hat{\boldsymbol{\beta}}|\mathbf{Z}_p) \xrightarrow{d} N(\mathbf{0}, T_*^{-1}\mathbf{I}_{m^2k})$  and, consequently,  $\mathbf{B}^\dagger vec(\hat{\boldsymbol{\beta}}_*|\mathbf{Z}_p) \xrightarrow{d} N(\mathbf{0}, T_*^{-1}\bar{\mathbf{D}})$ , which leads to:

$$T_* vec(\hat{\boldsymbol{\beta}}_*)' \mathbf{P} vec(\hat{\boldsymbol{\beta}}_*) \xrightarrow{d} \chi_{m^2(k-p-q)}^2$$

□

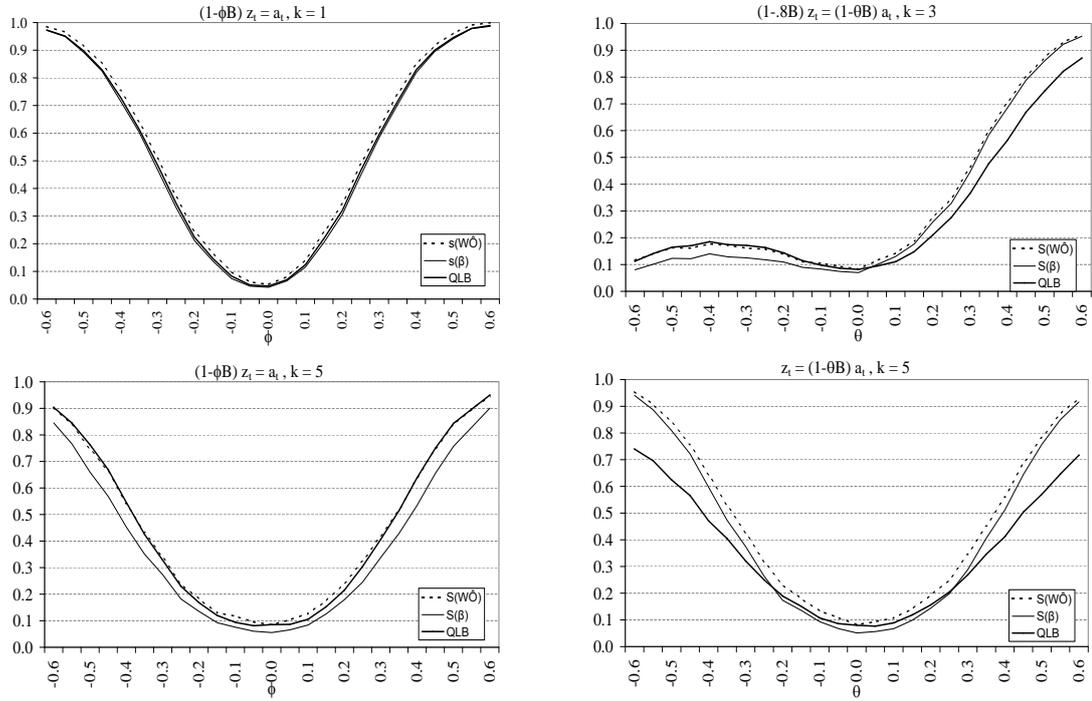


Figure 1: Empirical power of  $s(\mathbf{W}\hat{\mathbf{O}})$ ,  $s(\hat{\beta})$  and  $Q_{LB}$  for a sample size  $T = 50$ , different ARMA processes and lags ( $k$ ). The empirical powers are computed with a  $\chi_k^2$  distribution except for the ARMA(1,1) process where a  $\chi_{k-1}^2$  is employed as  $H_0$  is tested over the residuals of a misspecified AR(1) model. Results computed with 5000 replications.

Table 1: Empirical size and power of  $s(\mathbf{W}\hat{\mathbf{O}})$ ,  $s(\hat{\boldsymbol{\beta}})$  and  $Q_{LB}$  for a sample size  $T = 50$ , different bivariate VARMA processes and lags ( $k$ ). Results computed with 5000 replications.

Model	$k$ (lag)	AR	MA	$\Sigma$	$s(\mathbf{W}\hat{\mathbf{O}})$	$s(\hat{\boldsymbol{\beta}})$	$Q_{LB}$
1	1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$	.049	.046	$\begin{pmatrix} .052 & .055 \\ .057 & .053 \end{pmatrix}$
2	1	$\begin{pmatrix} 1 & -.2B \\ 0 & 1-.4B \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -.7 \\ -.7 & 2 \end{pmatrix}$	.957	.950	$\begin{pmatrix} .106 & .607 \\ .135 & .732 \end{pmatrix}$
3	1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1-.8B & 0 \\ .5B & 1-.5B \end{pmatrix}$	$\begin{pmatrix} 1 & .5 \\ .5 & 2 \end{pmatrix}$	.981	.980	$\begin{pmatrix} .967 & .195 \\ .320 & .607 \end{pmatrix}$
4	1	$\begin{pmatrix} 1-.4B & 0 \\ 0 & 1-.8B \end{pmatrix}$	$\begin{pmatrix} 1-.7B & .2B \\ -.3B & 1-.7B \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$	.196	.187	$\begin{pmatrix} .082 & .093 \\ .105 & .055 \end{pmatrix}$
5	2	$\begin{pmatrix} 1-.4B+.3B^2 & -.4B^2 \\ .1B & 1-.4B-.2B^2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 5 & 0 \\ 0 & .5 \end{pmatrix}$	.915	.962	$\begin{pmatrix} .718 & .368 \\ .526 & .847 \end{pmatrix}$
6	4	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$	.041	.059	$\begin{pmatrix} .066 & .074 \\ .072 & .066 \end{pmatrix}$
7	4	$\begin{pmatrix} 1 & -.2B^4 \\ 0 & 1-.4B^4 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -.7 \\ -.7 & 2 \end{pmatrix}$	.600	.669	$\begin{pmatrix} .090 & .382 \\ .119 & .584 \end{pmatrix}$
8	8	$\begin{pmatrix} 1-.3B & 0 \\ .4B & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -.5 \\ -.5 & 2 \end{pmatrix}$	.286	.481	$\begin{pmatrix} .334 & .130 \\ .356 & .152 \end{pmatrix}$

Table 2: P-value of the statistics.  $H_0$  : there are no correlations up to lag  $k$  in  $\mathbf{z}_t$ .

$k$ (lag)	$s(\mathbf{W}\hat{\mathbf{O}})$	$s(\hat{\boldsymbol{\beta}})$	$\mathbf{Q}_{LB}$
1	.000*	.000*	$\begin{pmatrix} .168 & .025^* & .045^* \\ .099 & .026^* & .053 \\ .043^* & .017^* & .063 \end{pmatrix}$
5	.281	.042*	$\begin{pmatrix} .813 & .404 & .491 \\ .704 & .408 & .475 \\ .453 & .295 & .533 \end{pmatrix}$
10	.184	.000*	$\begin{pmatrix} .938 & .707 & .552 \\ .896 & .698 & .460 \\ .737 & .644 & .475 \end{pmatrix}$

Table 3: P-value of the statistics.  $H_0$ : there are no correlations up to lag  $k$  in the residuals of model (14).

Statistic	$k$ (lags)			
	2	5	10	15
$Q_{LB}^\dagger$	.413	.855	.698	.779
$s(\mathbf{W}\hat{\mathbf{O}})$	.003*	.219	.150	.359
$s(\hat{\boldsymbol{\beta}})$	.000*	.028*	.002*	.000*

$Q_{LB}^\dagger$  presents the lowest p-value among all the elements of the  $\mathbf{Q}_{LB}$  matrix

Table 4: P-value of the statistics.  $H_0$ : there are no correlations up to lag  $k$  in the residuals of model (15).

Statistic	$k$ (lags)			
	2	5	10	15
$Q_{LB}^\dagger$	.447	.729	.642	.744
$s(\mathbf{W}\hat{\mathbf{O}})$	.944	.953	.611	.704
$s(\hat{\boldsymbol{\beta}})$	.942	.934	.331	.288

$Q_{LB}^\dagger$  presents the lowest p-value among all the elements of the  $\mathbf{Q}_{LB}$  matrix