Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 221 a 235

Generalized Portmanteau Tests Based on Subspace Methods

Tests de Portmanteau generalizados basados en métodos de subespacios

Alfredo García-Hiernaux[a]

Quantitative Economics Department, Universidad Complutense de Madrid, Spain

Abstract

The problem of diagnostic checking is tackled from the perspective of the subspace methods. Two statistics are presented and their asymptotic distributions are derived under the null hypothesis. The procedures are devised to deal with univariate and multivariate processes, are flexible and able to separately check regular and seasonal correlations. The performance in finite samples of the proposals is illustrated via Monte Carlo simulations and two examples with real data.

***Key words*:** Diagnostic checking, Portmanteau test, Residual autocorrelation, Residuals.

Resumen

Este artículo trata el problema de la diagnosis residual desde la perspectiva de los métodos de subespacios. Se presentan dos estadísticos y sus distribuciones asintóticas bajo la hipótesis nula. Ambos estadísticos pueden usarse con procesos univariantes o multivariantes, son flexibles y permiten contrastar separadamente las correlaciones regulares y estacionales. El comportamiento en muestras finitas de las dos propuestas se ilustra mediante simulaciones de Monte Carlo y dos ejemplos con datos reales.

***Palabras clave*:** autocorrelación residual, diagnosis de residuos, test de Portmanteau, residuos.

1. Introduction

Since the seminal work by Box & Pierce (1970), or the enhanced version by Ljung & Box (1978), many studies have focused in the ability of the statistical

[a]Professor. E-mail: agarciah@ucm.es

tests to determine the adequacy of a model. The procedures suggested in this paper cope with this problem from a novel perspective.

We use a subspace methods-based approach to derive two tests and their asymptotic distributions under the null of zero correlations up to order k. As subspace methods, the procedures are devised to deal with univariate and multivariate processes that leads to a generalization of Ljung & Box (1978) and Hosking (1980) -which is the Ljung-Box multivariate version- statistics, hereafter Q_{LB} and P_H, respectively.

The flexibility of the tests allows use to obtain gains in terms of statistical power and robustness against non-robust competitors as Q_{LB} and P_H. We propose that these gains can improve by tuning a specific matrix that may be modified by the user. Although this is not investigated in this paper, the question is briefly addressed in the conclusion. However, no comparison against robust statistics is performed as ours do not belong to this type of test. Our proposals are also able to separately test seasonal correlations. When applied to seasonal data, our tests present a gain in terms of degrees of freedom with respect to alternatives devised to cope with seasonality, as McLeod (1978) or Ursu & Duchesne (2009), and in terms of statistical power when compared to Q_{LB}. A Monte Carlo study shows that the finite sample properties of one of our tests outperform those of Q_{LB} in terms of nominal size, when the number of lags chosen grows, and in statistical power.

Finally, results in Aoki (1990), Casals, Sotoca & Jerez (1999) and Casals, García-Hiernaux & Jerez (2012) imply that Multiple-Source Error (MSE) state space, Single-Source Error (SSE) state space and VARMAX models are equally general and freely interchangeable. This means that our derivation of the distribution for the residuals of a VARMA model permits to test the adequacy of its equivalent MSE or SSE state space model. Consequently, our procedures can be sequentially used to determine the system order in a state space model (since the null hypothesis can always be written as residuals with system order equal to zero) which is a critical decision in the subspace methods literature and applied data modeling.

The plan of the paper is as follows. Section 2 presents previous results in subspace methods that will be used later. Some distributional results and the two tests proposed are derived in Sections 3 and 4, respectively. Lastly, Section 5 compares the performance of our proposals with Ljung-Box and Hoskings' tests using Monte Carlo experiments and two applications to real data.

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

2. Previous Results in Subspace Methods

Consider a linear fixed-coefficients system that can be described by the following state space model:

$$\boldsymbol{x}_{t+1} = \boldsymbol{\Phi}\boldsymbol{x}_t + \boldsymbol{E}\boldsymbol{\psi}_t \tag{1a}$$

$$\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{\psi}_t \tag{1b}$$

where $\boldsymbol{x}_t$ is a state n-vector, n being the true order of the system. In addition, $\boldsymbol{z}_t$ is an observable output m-vector, which is assumed to be zero-mean, $\boldsymbol{\psi}_t$ is an unobservable input m-vector, and $\boldsymbol{\Phi}$, $\boldsymbol{E}$ and $\boldsymbol{H}$ are parameter matrices with dimensions $(n \times n)$, $(m \times m)$ and $(n \times m)$, respectively. We suppose that the following assumptions hold in (1a-1b).

Assumptions. A.1: $\boldsymbol{\psi}_t$ is a sequence of zero-mean uncorrelated variables with $\mathrm{E}(\boldsymbol{\psi}_t\boldsymbol{\psi}_t') = \boldsymbol{\Gamma}$, $\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a positive definite matrix. A.2: The system is stable and strictly minimum-phase, *i.e.*, all the eigenvalues of $\boldsymbol{\Phi}$ and $(\boldsymbol{\Phi} - \boldsymbol{E}\boldsymbol{H})$ lie inside the unit circle.

We use the SSE, or also called innovations, form (1a-1b) since it is general and simpler than other representations. Its generality is discussed by Casals et al. (2012), who show that SSE, MSE and VARMAX models are equally general and freely interchangeable.

Additionally, throughout the paper we will also use $\bar{\boldsymbol{z}}_t$, a standardized version of $\boldsymbol{z}_t$, defined as $\bar{\boldsymbol{z}}_t = \hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{z}_t$, where $\hat{\boldsymbol{\Sigma}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t'$ and T is the sample size.

García-Hiernaux, Jerez & Casals (2010) show that model (1a-1b) can be transformed into a single equation in matrix form as $\boldsymbol{Z}_f = \boldsymbol{O}\boldsymbol{X}_f + \boldsymbol{V}\boldsymbol{\Psi}_f$, where: a) $\boldsymbol{Z}_f$ is a block Hankel matrix whose columns can be generally defined as $[\boldsymbol{z}_t', \dots, \boldsymbol{z}_{t+f-1}']'$ and each column is specified by a different value of t such that: $t = p+1, \dots, T-f+1$;[1] b) p and f are two integers chosen by the user, where $p > n$; and, c) $\boldsymbol{X}_f$ and $\boldsymbol{\Psi}_f$ are as $\boldsymbol{Z}_f$ but with $\boldsymbol{x}_t$ or $\boldsymbol{\psi}_t$, respectively, instead of $\boldsymbol{z}_t$. For simplicity, we assume $p = f$, denoting this integer by i. In this case, $\boldsymbol{Z}_f$ and $\boldsymbol{\Psi}_f$ are $im \times (T - 2i + 1)$ matrices. To simplify the notation, we denote the number of columns of both matrices by $T_* = T - 2i + 1$. Last, as it is detailed in García-Hiernaux et al. (2010), Section 2, matrices $\boldsymbol{O}$ and $\boldsymbol{V}$ are known functions of the original parameter matrices, $\boldsymbol{\Phi}$, $\boldsymbol{E}$ and $\boldsymbol{H}$:

$$\boldsymbol{O} := \begin{pmatrix}\boldsymbol{H}' & (\boldsymbol{H}\boldsymbol{\Phi})' & (\boldsymbol{H}\boldsymbol{\Phi}^2)' & \dots & (\boldsymbol{H}\boldsymbol{\Phi}^{i-1})'\end{pmatrix}'_{im \times n} \tag{2}$$

$$\boldsymbol{V} := \begin{pmatrix}\boldsymbol{I}_m & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{H}\boldsymbol{E} & \boldsymbol{I}_m & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{E} & \boldsymbol{H}\boldsymbol{E} & \boldsymbol{I}_m & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{H}\boldsymbol{\Phi}^{i-2}\boldsymbol{E} & \boldsymbol{H}\boldsymbol{\Phi}^{i-3}\boldsymbol{E} & \boldsymbol{H}\boldsymbol{\Phi}^{i-4}\boldsymbol{E} & \dots & \boldsymbol{I}_m\end{pmatrix}_{im} \tag{3}$$

[1]From now on all the block Hankel matrices will be defined in a similar way.

Given A.2 and for large values of i and T, $\boldsymbol{X}_f$ is to a close approximation representable as a linear combination of the past of the output, $\boldsymbol{M}\boldsymbol{Z}_p$, where $\boldsymbol{Z}_p := [\boldsymbol{z}'_{t-p}, \ldots, \boldsymbol{z}'_{t-1}]'$ with $t = p+1, \ldots, T-f+1$. Then, the relation between the past and the future of the output can be expressed by:

$$\boldsymbol{Z}_f \simeq \boldsymbol{\beta}\boldsymbol{Z}_p + \boldsymbol{V}\boldsymbol{\Psi}_f \tag{4}$$

where $\boldsymbol{\beta} = \boldsymbol{O}\boldsymbol{M}$. For a given system order n, subspace methods first solve a reduced-rank (as $\boldsymbol{\beta}$ is an im square matrix with rank $n < im$) weighted least squares problem by estimating $\boldsymbol{\beta}$ as:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{Z}_f\boldsymbol{Z}'_p(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-1} \tag{5}$$

and splitting it to estimate $\boldsymbol{O}$ and $\boldsymbol{M}$, and then $\boldsymbol{V}$. Finally, the parameter matrices in (1a-1b) can be obtained from the estimates $\hat{\boldsymbol{O}}$, $\hat{\boldsymbol{M}}$ and $\hat{\boldsymbol{V}}$, see, e.g., Katayama (2005).

3. Some Distributional Results

We begin by establishing the null hypothesis that $\boldsymbol{z}_t$ has no correlations different from zero up to lag order k, i.e., H_0: $\rho_j = 0$, $j = 1, 2, \ldots, k$, where ρ_j is the correlation coefficient of order j. It is common in the literature that the user just chooses k to conduct the hypothesis testing. Accordingly, we define i as a function of k, such that i is the integer rounded toward infinity of $(k+1)/2$. However, the tests could be directly adapted to any other value of i, or even different values of p and f.

The first proposal uses a generalized least squares approach. Using the previously defined standardized version of the output and input, we have $\bar{\boldsymbol{Z}}_f = \bar{\boldsymbol{\beta}}\bar{\boldsymbol{Z}}_p + \bar{\boldsymbol{V}}\bar{\boldsymbol{\Psi}}_f$, where $\bar{\boldsymbol{Z}}_p, \bar{\boldsymbol{\Psi}}_p$ are as $\boldsymbol{Z}_p, \boldsymbol{\Psi}_p$ but with $\bar{z}_t, \bar{\boldsymbol{\psi}}_t$ instead of the original $\boldsymbol{z}_t, \boldsymbol{\psi}_t$. Matrix $\bar{\boldsymbol{\beta}}$ can be estimated as (5), but with the standardized matrices $\bar{\boldsymbol{Z}}_p$ and $\bar{\boldsymbol{Z}}_f$ instead of $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_f$. Notice that an immediate consequence of the null hypothesis is that $\bar{\boldsymbol{\beta}} = \boldsymbol{0}_{im}$. By applying the vec operator, which stacks the columns of a matrix into a long vector, on $\hat{\bar{\boldsymbol{\beta}}}$ we state the following proposition:

Proposition 1. *Given A.1-A.2, under H_0, $\sqrt{T_*}vec(\hat{\bar{\boldsymbol{\beta}}}|\bar{\boldsymbol{Z}}_p) \xrightarrow{d} N(\boldsymbol{0}, \bar{\boldsymbol{\Pi}})$, where $\bar{\boldsymbol{\Pi}}$ is derived in the Appendix.*

The second test comes from a canonical correlation approach. This one is based on the information held in $\boldsymbol{O}$, which affects $\boldsymbol{Z}_f$ through $\boldsymbol{\beta}$, see (4). The canonical correlation analysis (CCA) between $\boldsymbol{Z}_f$ and $\boldsymbol{Z}_p$ is usually performed by analyzing the product $(\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\boldsymbol{Z}_f\boldsymbol{Z}'_p(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-\frac{1}{2}}$, see Katayama (2005) for a detailed description on CCA. From equation (5), one could get the product above from $(\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}$, estimating $\boldsymbol{O}$ as $\boldsymbol{Z}_f\boldsymbol{Z}'_p(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-\frac{1}{2}}$ and then $\boldsymbol{M}$ as $(\boldsymbol{Z}_p\boldsymbol{Z}'_p)^{-\frac{1}{2}}$, so that the equality $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{O}}\hat{\boldsymbol{M}}$ holds. This second alternative leads to Proposition 2:

Proposition 2. *Given A.1-A.2, under H_0, $\sqrt{T_*}vec\big((\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}|\boldsymbol{Z}_p\big) \xrightarrow{d} N\big(\boldsymbol{0}, \bar{\boldsymbol{\Pi}}\big)$.*

4. The Test Statistics

The covariance matrix $\bar{\boldsymbol{\Pi}}$ is not, generally, the identity matrix. In fact, it is only so when $i = 1$. For $i > 1$ some elements in $\hat{\boldsymbol{\beta}}$ and $(\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}$ are perfectly correlated by construction, see equation (8) in the Appendix. However, as the structure of $\bar{\boldsymbol{\Pi}}$ is known, the following proposition applies.

Proposition 3. *For any random matrix $\boldsymbol{A}$ such that $\sqrt{T_*}vec\boldsymbol{A} \xrightarrow{d} N\big(\boldsymbol{0}, \bar{\boldsymbol{\Pi}}\big)$, there is an idempotent matrix $\boldsymbol{P}_{(im)^2}$ of rank m^2k, such that $\mathcal{S}_A = T_*vec(\boldsymbol{A})'\boldsymbol{P}vec(\boldsymbol{A}) \xrightarrow{d} \chi^2_{m^2k}$.*

Corollary 1. Consequently, by combining Propositions 1, 2 and 3, we get that both, $\mathcal{S}_{\boldsymbol{\beta}} = T_*vec(\hat{\bar{\boldsymbol{\beta}}})'\boldsymbol{P}vec(\hat{\bar{\boldsymbol{\beta}}})$ and $\mathcal{S}_{\boldsymbol{O}} = T_*vec\big((\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}\big)'\boldsymbol{P}vec\big((\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}\big)$ converge to a chi-square distribution with m^2k degrees of freedom.

Matrix $\boldsymbol{P}$ is the product of two weighting matrices that average the perfectly correlated elements of $vec(\boldsymbol{A})$ in a vector of m^2k uncorrelated elements. This point deserves further discussion, as it makes the procedure flexible by tuning matrix $\boldsymbol{P}$ according on the specific case. For instance, some $\boldsymbol{P}$ could be chosen with the aim of reducing the effects of outliers or increasing the statistical power of the tests.

We have seen that, when $i > 1$ some elements in $\hat{\bar{\boldsymbol{\beta}}}$ and $(\boldsymbol{Z}_f\boldsymbol{Z}'_f)^{-\frac{1}{2}}\hat{\boldsymbol{O}}$ are perfectly correlated. Matrix $\boldsymbol{P}$, as it is proposed in the proof of Proposition 3 averages the perfectly correlated elements to obtain a vector of uncorrelated components. The procedure computes each k-order correlation for different non-disjoint subsamples and averages them to obtain a single one. In this way, the effect of an outlier will be mitigated, provided that it only affects a small proportion of the weighted correlations. This will be more likely the more subsamples we use, *i.e.*, the higher i is. Obviously, our statistics do not use robust estimation methods, as M-estimators or MM-estimators, and therefore they are not robust statistics and will perform worse than those methods in the presence of outliers. However, we expect that they present a better performance than non-robust statistics as Q_{LB} in such cases; specifically, innovational outliers, additive outliers or level changes (see, for definitions, Tsay 1988). An example illustrates this feature in the next section.

An interesting point that deserves more attention is that one could easily tune the matrix $\boldsymbol{P}$ according to the data. If we are suspicious about the presence of outliers then, instead of calculating the mean of several k-order correlation (which is the proposal here), the median or the minimum could be used. In these cases, the distribution of the statistics should be derived but the statistics are likely to be more robust.

On the other hand, often in practice, only the low-order correlations are of interest to analysts. Consequently, the possibility of modifying $\boldsymbol{P}$ by increasing the weights of low lags (either *ad-hoc* or using a more sophisticated mechanism) should increase the power of the tests.

In any case, a standard use of the Portmanteau tests is to check the residuals obtained from fitting Vector Autoregressive Moving Average, VARMA, models.

Here we adopt the usual definition of a stationary m-variate ARMA(p,q) process (see, *e.g.*, Liu 2006, p. 14.2). Nevertheless, when $\boldsymbol{z}_t$ are the residuals from a VARMA model, the asymptotic distribution of $\mathcal{S}_\beta$ and $\mathcal{S}_O$ is not as it has been shown. The reason is that A.1 does not hold, as residuals, contrary to innovations, present some linear constraints inherit from the VARMA estimation (see, e.g., Mauricio 2007). In these circumstances, the following proposition establishes the asymptotic distribution of both statistics.

Proposition 4. *When $\boldsymbol{z}_t$ in (1b) are the residuals from a fitted m-vector ARMA(p,q) model, then, under H_0, $\mathcal{S}_\beta$ and $\mathcal{S}_O$ converge in distribution to a $\chi^2_{m^2(k-p-q)}$.*

At this point, notice that testing H_0 in any m-variate process requires (if the Ljung-Box test is used) a Q-matrix that leads to m^2 different statistics. As Hosking (1980) test, ours offer a more natural scalar statistic instead. Further, it is straightforward to see that for $p = 1$ and $f = k + 1$ both, $\mathcal{S}_\beta$ and $\mathcal{S}_O$, are equivalent to: (i) Ljung-Box statistic when $m = 1$ and (ii) Hoskings' statistic when $m \geq 1$ (see, Hosking 1980, p. 605). In short, our procedures generalize Ljung-Box and Hosking's procedures, allowing for different values of p and f.

Furthermore, these results are extended to multiplicative seasonal VARMA(p,q) $\times(P,Q)_s$ models, where s is the seasonal period and (P,Q) are the seasonal autoregressive and moving average orders, respectively (see, Liu 2006, p. 14.36). Regarding this, McLeod (1978), for the univariate case ($m = 1$), and Ursu & Duchesne (2009), for multivariate processes, proved that an adjusted version of the Q-statistic follows a $\chi^2_{m^2(k-p-q-P-Q)}$. With our proposals, if one only identifies and estimates the seasonal parameters (P,Q), $\mathcal{S}_\beta$ or $\mathcal{S}_O$ and Proposition 4 could easily be used to check whether there is seasonal correlation in the residuals, testing H_0: $\rho_j = 0$, $j = s, 2s, \dots, ks$. The statistics should be computed by replacing $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_f$ by their seasonal counterparts $\boldsymbol{Z}^s_p := [\boldsymbol{z}'_{t-si}, \boldsymbol{z}'_{t-s(i-1)}, \dots, \boldsymbol{z}'_{t-s}]'$ and $\boldsymbol{Z}^s_f := [\boldsymbol{z}'_t, \boldsymbol{z}'_{t+s}, \dots, \boldsymbol{z}'_{t+s(i-1)}]'$, where $t = si+1, s(i+1)+1, \dots, T-s(i-1)$. In those cases $\mathcal{S}_\beta$ and $\mathcal{S}_O$ follow a $\chi^2_{m^2(k-P-Q)}$. Hence, the adequacy of a VARMA$(p,q)\times(P,Q)_s$ model can be checked by sequentially identifying, estimating and applying the tests using the seasonal matrices, $\boldsymbol{Z}^s_p$ and $\boldsymbol{Z}^s_f$, and then the regular ones, $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_f$. The sequential procedure implies a gain in terms of degrees of freedom with respect to Ursu & Duchesne (2009) when testing for seasonal correlation, as we only consider the seasonal part and not the complete model. This may be a great advantage in very short samples.

5. Numerical Examples

In this section we investigate the finite sample properties of the proposed tests. Its performance is compared with that of Ljung-Box (Q_{LB}) and Hosking (P_H) statistics, as they are the most common and cited diagnostic tests in the literature for the univariate and the multivariate case, respectively. As said previously, no comparison against robust methods is made as ours do not fulfill those characteristics. However, in order to analyze its behavior in different situations, we split the study into some Monte Carlo simulations of univariate processes without outliers contamination and two applications to real data in which, at least the first one, contains documented additive outliers.

5.1. Monte Carlo Simulations

Firstly, we will study how the autocorrelation structure affects the empirical size and power of the tests.

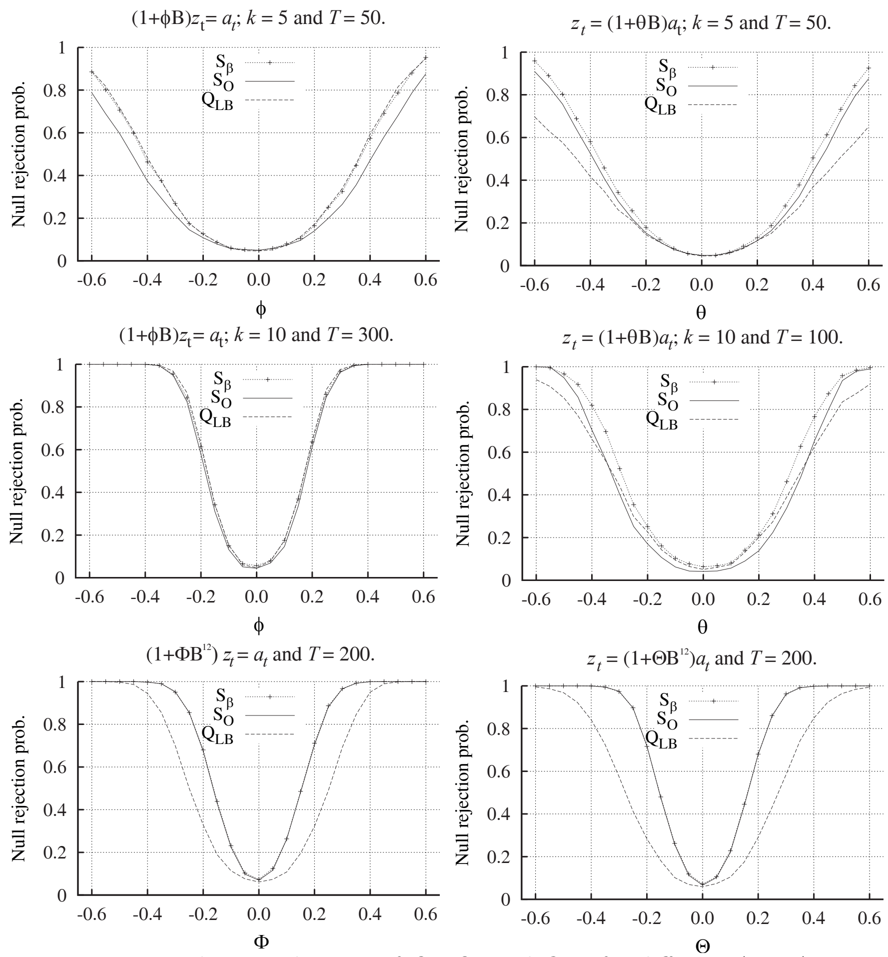

FIGURE 1: Empirical size and power of $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} for different ARMA processes (computed with a χ^2_k at 5% and 5000 replications). The graphs at the bottom depict the size and power for two seasonal processes. In these cases, Q_{LB} is computed with $k = 24$ to be able to capture the seasonal structure, while $\mathcal{S}_\beta$ and $\mathcal{S}_O$ are computed with the seasonal matrices $\boldsymbol{Z}_p^s$ and $\boldsymbol{Z}_f^s$ and $k^s = 2$.

Figure 1 presents the empirical size and power of $\mathcal{S}_O$, $\mathcal{S}_\beta$ and Q_{LB} for alternative AR(1) and MA(1) processes, with different k (lags) and T (sample size).

Hosking's test is omitted as it coincides with Q_{LB} in univariate processes.[2] The most noticeable features of this exercise are:

1. In processes without seasonality and short samples ($T = 50$):

   a) Q_{LB} and $\mathcal{S}_\beta$ perform very similarly with autoregressive structures, both being slightly more powerful than $\mathcal{S}_O$.

   b) The empirical power of Q_{LB} is clearly outperformed by our two proposals when MA structures. This result partially coincides with Monti (1994) who proposes a test using the residual partial autocorrelations whose behavior is better than that of Q_{LB} if the order of the MA is understated. However, in that case it was shown that Q_{LB} was more powerful if the order of the AR part was understated. In contrast, we did not find any evidence of this when applying $\mathcal{S}_\beta$.

2. The asymptotically equivalence of the three tests is observed when T grows. For $T = 300$ and a AR(1) process the performance of the three tests is almost identical. When $T = 200$ and a MA(1) process our tests still outperform Q_{LB}, although less evidently than when $T = 50$.

3. In seasonal processes, $\mathcal{S}_O$ and $\mathcal{S}_\beta$ clearly outperform Q_{LB} in terms of statistical power. Not surprisingly, this enhancement is even bigger with seasonal MA(1) processes. The explanation comes from the fact that $\mathcal{S}_O$ and $\mathcal{S}_\beta$ are computed with the seasonal matrices $\boldsymbol{Z}_p^s$ and $\boldsymbol{Z}_f^s$ defined in Section 4 and the test is then computed with $k^s = 2$. However, Q_{LB} is computed with $k = 24$ to be able to capture the seasonal correlation.

Secondly, we analyze the empirical distribution of the statistics under H_0 for white noise samples and increasing values of k. Notice that in those cases the null distribution follows a χ^2_k. In this context, Figure 2 shows that $\mathcal{S}_\beta$ better fits the theoretical distribution than Q_{LB} and $\mathcal{S}_O$, when $k = 15$ and $T = 50$. Interestingly enough, the simulations evidence that Q_{LB} and $\mathcal{S}_O$ empirical distributions get further away from the theoretical one when k increases for a given T. Nevertheless, the distribution of $\mathcal{S}_\beta$ correctly fits its theoretical counterpart regardless of the value of k.[3]

5.2. Two Examples with Real Data

The first example with real data considers the Residence Telephone Extensions Inward Movement known as RESEX series (y_t). The left plot of Figure 3 shows the original monthly series that goes from January 1966 to May 1973, where observations $t = 83, 84$ are larger than the rest. These two outliers have a known cause, namely a bargain month, in which residence extensions could be requested free of

[2]Simulations with higher lags in pure autoregressive, pure moving average or ARMA models show similar or mixed results that do not suggest additional conclusions and, consequently, are not presented here. However, they are available from the author upon request.

[3]Additional simulations not shown here are available from the author upon request.

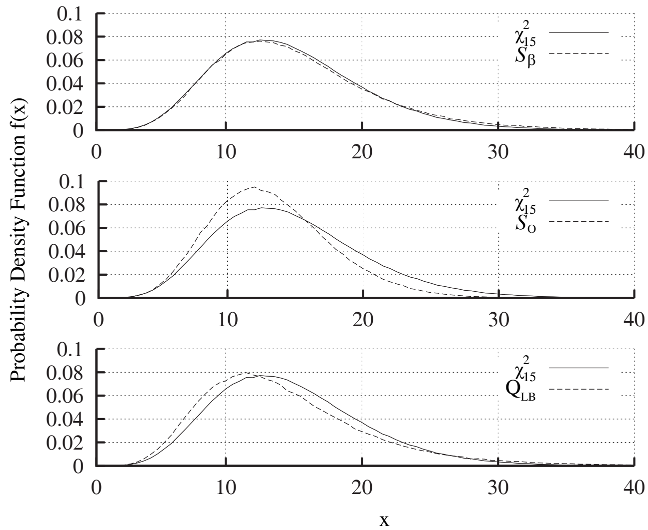

FIGURE 2: Empirical distribution for $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} compared to a theoretical χ^2_{15}; 250,000 replications for $T = 50$ and $k = 15$.

charge. Robust methods identify an AR(1) in the regularly and seasonally differenced transformation ($\nabla\nabla_{12}y_t$), see, e.g., Rousseeuw & Leroy (1987) or Li (2004). On the contrary, standard methods usually do not capture the autocorrelation structure due to the effect of the outliers.

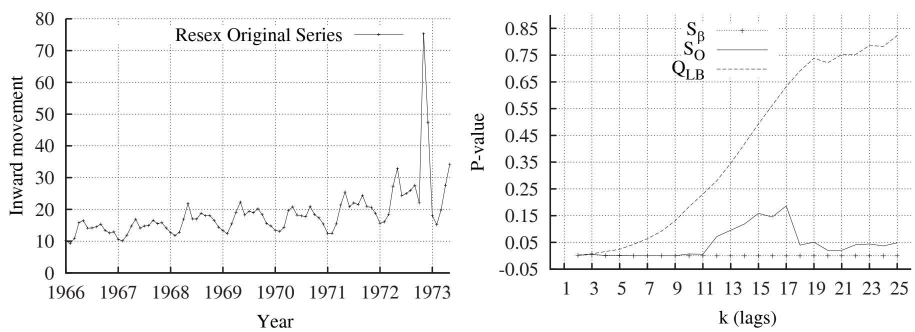

FIGURE 3: Top plot: Original RESEX series (y_t). Bottom plot: P-values of $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} for lags (k) from 1 to 25 obtained by applying the statistics to the transformed series $\nabla\nabla_{12}\log y_t$.

When we apply $\mathcal{S}_\beta$, $\mathcal{S}_O$ and Q_{LB} to the transformed series $\nabla\nabla_{12}\log y_t$, we find that Q_{LB} does not reject the null from $k = 7$ at 5% of significance and from $k = 8$ at 10%. However, $\mathcal{S}_O$ rejects the null at a 5% for all k except when $k = 12 - 17$, where the p-values always remain below 16%. Finally, $\mathcal{S}_\beta$ behaves much better than Q_{LB} and $\mathcal{S}_O$ with this data, rejecting the null at 1% of significance for all k studied. This example is relevant as most empirical works only show the Q_{LB} values for high lags (usually $10, 15$ or 20) without paying attention to the loss of power when k increases, that can grow dramatically in the presence of outliers. $\mathcal{S}_\beta$ behavior explanation lies in the fact that i has been defined as a positive function of k (see Section 2), so when k grows, i increases. As i is the number of subsamples to compute the autocorrelations of the same order, when i increases, the weight of the contaminated subsamples diminishes.

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 aim of modeling these data is to illustrate the performance of the proposed statistics, as specification tools, and compare it with Q_{LB} and P_H.

Since all series appear non-stationary, we use the log-difference transformation $\boldsymbol{z}_t = \nabla \log(\boldsymbol{y}_t)$, where $\boldsymbol{y}_t$ are the original series. Table 1 shows the results of applying the statistics to $\boldsymbol{z}_t$ with different lags. The first conclusion is that even if all the tests suggest that there are significant correlations, at least up to order one, Q_{LB} presents very low power when a (not-so) large lag is chosen. It seems that the significant correlations at lag one are diluted by insignificant correlations at other lags, and this effect is much more important in Q_{LB} than in $\mathcal{S}_\beta$, $\mathcal{S}_O$ or P_H. In this context, notice that $\mathcal{S}_\beta$ is the only statistic that keeps its p-value under 5% for $k = 5, 10$. Additionally, Q_{LB} only reveals 5 out of 9 correlations statistically significant at 5%, when $k = 1$.

Table 1: P-value of the statistics. H_0: There are no correlations up to lag k in $\boldsymbol{z}_t$.

k (lag)	$\mathcal{S}_O$	$\mathcal{S}_\beta$	P_H	Q_{LB}
1	.000*	.000*	.000*	$\begin{pmatrix} .172 & .026^* & .047^* \\ .103 & .027^* & .056 \\ .045^* & .018^* & .066 \end{pmatrix}$
5	.241	.035*	.072	$\begin{pmatrix} .822 & .416 & .506 \\ .716 & .421 & .493 \\ .470 & .309 & .549 \end{pmatrix}$
10	.155	.003*	.082	$\begin{pmatrix} .954 & .744 & .632 \\ .918 & .734 & .545 \\ .779 & .682 & .573 \end{pmatrix}$

* rejects at 5%.

Following the results obtained with Q_{LB} at 5% in Table 1 when $k = 1$, a restricted VAR(1) model $(\boldsymbol{I} - \boldsymbol{\Phi}_1 B)\boldsymbol{z}_t = \boldsymbol{a}_t$ is tentatively specified. Parameter estimates result in:

$$\hat{\boldsymbol{\Phi}}_1 = \begin{pmatrix} 0 & -.188^* & -.035 \\ 0 & -.289^* & 0 \\ -.401^* & .117 & 0 \end{pmatrix}, \quad \hat{\boldsymbol{\Gamma}}_{\boldsymbol{a}} = \begin{pmatrix} 2.263 & 2.296 & 2.202 \\ & 2.496 & 2.364 \\ & & 2.770 \end{pmatrix} \times 10^{-3}, \quad (6)$$

where '0' denotes an entry constrained to be zero and '*' means the parameter is significant at 5%. Table 2 presents the p-value of the diagnostic tests on the residuals of model (6).

Table 2: P-value of the statistics. H_0: There are no correlations up to lag k in model (6) residuals.

Statistic	k (lags)			
	2	5	10	15
$\mathcal{S}_O$	.003*	.200	.110	.202
$\mathcal{S}_\beta$	.000*	.003*	.006*	.007*
P_H	.000*	.037*	.052	.256
$Q^\dagger_{LB}$	.429	.869	.792	.884

$Q^\dagger_{LB}$ is to the lowest p-value among all the elements of the Q_{LB} matrix.
* rejects at 5%.

Q_{LB} suggests that the correlations are zero for $k = 2, 5, 10, 15$ at 10% level of significance, implying that model (6) is appropriate. However, $\mathcal{S}_O$, P_H and $\mathcal{S}_\beta$ reject H_0 for $k = 2$, $k = 2, 5, 10$ and $k = 2, 5, 10, 15$, respectively, at 5% level. Hence, $\mathcal{S}_O$, P_H and particularly $\mathcal{S}_\beta$ strongly evidence that Q_{LB} leads to an inappropriate specification. Instead, if we specify an unrestricted VAR(1), the estimation returns:

$$\hat{\mathbf{\Phi}}_1 = \begin{pmatrix} 1.226^* & -1.355^* & .005 \\ .830^* & -1.027^* & .035 \\ .463 & -.813^* & .142 \end{pmatrix}, \quad \hat{\mathbf{\Gamma}}_{\boldsymbol{a}} = \begin{pmatrix} 2.033 & 2.140 & 2.039 \\ & 2.390 & 2.253 \\ & & 2.647 \end{pmatrix} \times 10^{-3} \tag{7}$$

To check if the residual correlations of model (7) are zero, the four procedures are again employed. Table 3 shows these results. None of the tests rejects H_0 for any value of k. Surprisingly, Q_{LB} presents the smallest evidence in favor of the null out of the four alternative for $k = 2, 5$. Model (7) was proposed by Lütkepohl & Poskitt (1996) and, as it was shown in Grubb (1992), is better than many other alternatives, in particular model (6).

Table 3: P-value of the statistics. H_0: There are no correlations up to lag k in model (7) residuals.

Statistic	k (lags)			
	2	5	10	15
$\mathcal{S}_O$	.953	.952	.480	.454
$\mathcal{S}_\beta$	.937	.952	.445	.506
P_H	.945	.951	.601	.838
$Q^\dagger_{LB}$	.455	.756	.736	.858

$Q^\dagger_{LB}$ is to the lowest p-value among all the elements of the Q_{LB} matrix.
* rejects at 5%.

From this exercise with multiple series we conclude that: (i) multivariate Portmanteau statistics, $\mathcal{S}_{\boldsymbol{\beta}}, \mathcal{S}_{\boldsymbol{O}}$ and P_H, perform better than the multiple Q_{LB}, and (ii) $\mathcal{S}_\beta$ seems to be more powerful than $\mathcal{S}_O$ and P_H when k grows.

6. Concluding Remarks

This work tackles the problem of diagnostic checking from an original viewpoint. Two statistics based on subspace methods are presented and their asymptotic distributions are derived under the null. They generalize the Box-Pierce statistic for single series, the Hoskings' statistic in the multivariate case and are able to separately test seasonal and regular correlations. Monte Carlo simulations and two examples with real data show that our proposals perform better than the common Ljung-Box Q-statistic in many different situations. The procedures can sequentially be used to determine the system order, as the null hypothesis can always be written as $n = 0$, which is a critical decision in the subspace methods literature and the applied data modeling.

Moreover, the subspace structure and the possibility of tuning a weight matrix make the tests more flexible and robust against outliers than non-robust alternatives. In this paper we just propose a particular form for this matrix $\boldsymbol{P}$ (see proof of Proposition 3), but others are possible and could be fitted to the characteristics of the data. A deeper analysis of this point with the suggestion of different matrices $\boldsymbol{P}$ could be the core of a next research.

Finally, the procedures used in the numerical examples and described in the paper are implemented in a MATLAB toolbox for time series modeling called `E4` that can be downloaded from the webpage `www.ucm.es/info/icae/e4`. The source code for all the functions in the toolbox is freely provided under the terms of the GNU General Public License. This site also includes a complete user manual and other materials.

Acknowledgment

Manuel Domínguez, Miguel Jerez and two anonymous referees made useful comments and suggestions to previous versions of this work. The author gratefully acknowledges financial support from Ministerio de Educación y Ciencia, ref. ECO2011-23972 and the Ramón Areces Foundation.

[Recibido: noviembre de 2012 — Aceptado: mayo de 2013]

References

Aoki, M. (1990), *State Space Modelling of Time Series*, Springer Verlag, New York.

Box, G. E. P. & 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., García-Hiernaux, A. & Jerez, M. (2012), 'From general State-Space to VARMAX models', *Mathematics and Computers in Simulation* **80**(5), 924–936.

Casals, J., Sotoca, S. & Jerez, M. (1999), 'A fast and stable method to compute the likelihood of time invariant state space models', *Economics Letters* **65**(3), 329–337.

García-Hiernaux, A., Jerez, M. & Casals, J. (2010), 'Unit roots and cointegration modeling through a family of flexible information criteria', *Journal of Statistical Computation and Simulation* **80**(2), 173–189.

Grubb, H. (1992), 'A multivariate time series analysis of some flour price data', *Applied Statistics* **41**, 95–107.

Hosking, J. R. M. (1980), 'The multivariate Pormanteau statistic', *Journal of the American Statistical Association* **75**(371), 602–608.

Katayama, T. (2005), *Subspace Methods for System Identification*, Springer Verlag, London.

Li, W. K. (2004), *Diagnostic Checks in Time Series*, Chapman and Hall/CRC, Florida.

Liu, L. M. (2006), *Time Series Analysis and Forecasting*, 2 edn, Scientific Computing Associates Corporation, Illinois.

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

Lütkepohl, H. & Poskitt, D. S. (1996), 'Specification of echelon form VARMA models', *Journal of Business and Economic Statistics* **14**(1), 69–79.

Mauricio, J. A. (2007), 'Computing and using residuals in time series models', *Computational Statistics and Data Analysis* **52**(3), 1746–1763.

McLeod, A. I. (1978), 'On the distribution of residual autocorrelations in Box-Jenkins model', *Journal of the Royal Statistics Society B* **40**, 296–302.

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

Rousseeuw, P. J. & Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, John Wiley, New York.

Tsay, R. S. (1988), 'Outliers, Level shifts, and variance changes in time series', *Journal of Forecasting* **7**, 1–20.

Ursu, E. & Duchesne, P. (2009), 'On multiplicative seasonal modelling for vector time series', *Statistics and Probability Letters* **79**(19), 2045–2052.

White, H. (2001), *Asymptotic Theory for Econometricians*, Academic Press.

Appendix

Proof of Proposition 1. Equation (4) can be written as an equality by including a term that tends to zero at an exponential rate as a result of the minimum-phase assumption. For the lack of simplicity, we neglect this term during the proof and treat equation (4) as an equality. By applying the *vec* operator to the standardized version of equation (4), we have $vec\bar{\boldsymbol{Z}}_f = (\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})vec\bar{\boldsymbol{\beta}} + vec\bar{\boldsymbol{\Psi}}_f$, where we use that, under H_0, $\bar{\boldsymbol{V}} = \boldsymbol{I}_{im}$. From this, $vec\hat{\bar{\boldsymbol{\beta}}} = [(\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})'(\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})]^{-1}(\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})'vec\bar{\boldsymbol{Z}}_f$, and hence we get $vec(\hat{\bar{\boldsymbol{\beta}}} - \bar{\boldsymbol{\beta}}) = \bar{\boldsymbol{H}}^{-1}\bar{\boldsymbol{A}}'vec\bar{\boldsymbol{\Psi}}_f$, where $\bar{\boldsymbol{H}} = \bar{\boldsymbol{A}}'\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{A}} = \bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im}$. Therefore, the covariance of $vec\hat{\bar{\boldsymbol{\beta}}}$ conditional to $\bar{\boldsymbol{Z}}_p$ is $cov[vec\hat{\bar{\boldsymbol{\beta}}}|\bar{\boldsymbol{Z}}_{\boldsymbol{p}}] = \bar{\boldsymbol{H}}^{-1}\bar{\boldsymbol{A}}'(\boldsymbol{\Omega} \otimes \boldsymbol{I}_m)\bar{\boldsymbol{A}}\bar{\boldsymbol{H}}^{-1}$, where $(\boldsymbol{\Omega} \otimes \boldsymbol{I}_m)$ denotes de covariance of $vec\bar{\boldsymbol{\Psi}}$ and we use that, under H_0, $\mathrm{E}(\bar{\boldsymbol{z}}_t\bar{\boldsymbol{z}}_t') = \mathrm{E}(\bar{\boldsymbol{\psi}}_t\bar{\boldsymbol{\psi}}_t') = \boldsymbol{I}_m$. Asymptotically, the Ergodic Theorem (see, Theorem 3.34, White 2001) and H_0 ensure that $T_*^{-1}\bar{\boldsymbol{A}}'(\boldsymbol{\Omega} \otimes \boldsymbol{I}_m)\bar{\boldsymbol{A}} \overset{a.s.}{\to} \bar{\boldsymbol{\Pi}}$ and $T_*\bar{\boldsymbol{H}}^{-1} \overset{a.s.}{\to} \boldsymbol{I}_{(im)^2}$, where $\bar{\boldsymbol{\Pi}}$ has the following structure:

$$\bar{\boldsymbol{\Pi}} = \begin{pmatrix} \boldsymbol{I}_{im^2} & \boldsymbol{\Pi}_{i-1} & \boldsymbol{\Pi}_{i-2} & \dots & \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_{i-1}' & \boldsymbol{I}_{im^2} & \boldsymbol{\Pi}_{i-1} & \dots & \boldsymbol{\Pi}_2 \\ \boldsymbol{\Pi}_{i-2}' & \boldsymbol{\Pi}_{i-1}' & \boldsymbol{I}_{im^2} & \dots & \boldsymbol{\Pi}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Pi}_1' & \boldsymbol{\Pi}_2' & \boldsymbol{\Pi}_3' & \dots & \boldsymbol{I}_{im^2} \end{pmatrix}_{(im)^2} \tag{8}$$

where $\boldsymbol{\Pi}_{i-j}$ is a diagonal im^2 matrix with $\boldsymbol{\omega}_{i-j}$ in the main diagonal,

$$\boldsymbol{\omega}_{i-j} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{m(i-j)} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}_{im} \quad \text{and } j = 1, 2, \dots, T_* - 1 \tag{9}$$

Moreover, when $j \geq i$, $\boldsymbol{\omega}_{i-j}$ is an im zero-matrix. This particular composition of $\bar{\boldsymbol{\Pi}}$ is inherited from the structure of $\boldsymbol{\Psi}_f$. Consequently, $\sqrt{T_*}vec(\hat{\bar{\boldsymbol{\beta}}}|\bar{\boldsymbol{Z}}_{\boldsymbol{p}}) \overset{d}{\to} N\big(\boldsymbol{0}, \bar{\boldsymbol{\Pi}}\big)$. ■

Proof of Proposition 2. Let $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}} = (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\boldsymbol{Z}_f\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}$, which becomes $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}} = (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}(\boldsymbol{O}\boldsymbol{M}\boldsymbol{Z}_p + \boldsymbol{\Psi}_f)\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}$ under the null. Substituting $\boldsymbol{M} = (\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}$ and vectorizing, we get $vec[(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}(\hat{\boldsymbol{O}} - \boldsymbol{O})] = \big[\big((\boldsymbol{Z}_p\boldsymbol{Z}_p)^{-\frac{1}{2}}\boldsymbol{Z}_p'\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]vec\boldsymbol{\Psi}_f$.

The covariance matrix of $vec[(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}(\hat{\boldsymbol{O}} - \boldsymbol{O})]$ conditional to $\boldsymbol{Z}_p$ is written $\mathrm{E}\Big[\big[\big((\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\boldsymbol{Z}_p\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]vec\boldsymbol{\Psi}_f vec\boldsymbol{\Psi}_f'\big[\big(\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]|\boldsymbol{Z}_p\Big]$. By replacing $(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}} = (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}$ and using that, under H_0, $\boldsymbol{Z}_f|\boldsymbol{Z}_p = \boldsymbol{Z}_f$, the covariance becomes $\big[\big((\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\boldsymbol{Z}_p\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big](\boldsymbol{\Omega} \otimes \boldsymbol{Q})\big[\big(\boldsymbol{Z}_p'(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}}\big) \otimes (\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\big]$. Again under the null hypothesis, $\sqrt{T_*}(\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}} \overset{a.s.}{\to} \boldsymbol{I}_i \otimes \boldsymbol{\Gamma}^{-\frac{1}{2}}$ and $\sqrt{T_*}(\boldsymbol{Z}_p\boldsymbol{Z}_p')^{-\frac{1}{2}} \overset{a.s.}{\to} \boldsymbol{I}_i \otimes \boldsymbol{\Gamma}^{-\frac{1}{2}}$ hold. Using the properties of the Kronecker product, we can finally write $cov[vec((\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}})] \overset{a.s.}{\to} T_*^{-2}\Big[\big[\big((\boldsymbol{I}_i \otimes \boldsymbol{\Gamma}^{-\frac{1}{2}})\boldsymbol{Z}_p\big) \otimes \boldsymbol{I}_i\big]\boldsymbol{\Omega}\big[\big(\boldsymbol{Z}_p'(\boldsymbol{I}_i \otimes \boldsymbol{\Gamma}^{-\frac{1}{2}})\big) \otimes \boldsymbol{I}_i\big]\Big] \otimes \boldsymbol{I}_m$.

On the other hand, the covariance of $vec(\hat{\boldsymbol{\beta}}|\boldsymbol{Z}_p)$ is $\bar{\boldsymbol{H}}^{-1}(\bar{\boldsymbol{Z}}_p \otimes \boldsymbol{I}_{im})(\boldsymbol{\Omega} \otimes \boldsymbol{I}_m)(\bar{\boldsymbol{Z}}_p' \otimes \boldsymbol{I}_{im})\bar{\boldsymbol{H}}'^{-1} \overset{a.s.}{\to} T_*^{-1}\bar{\boldsymbol{\Pi}}$. Finally, as $\lim_{T\to\infty} |\bar{\boldsymbol{Z}}_{\boldsymbol{p}} - (\boldsymbol{I}_i \otimes \boldsymbol{\Gamma}^{-\frac{1}{2}})\boldsymbol{Z}_p| = \boldsymbol{0}$, then both, $vec(\hat{\bar{\boldsymbol{\beta}}}|\boldsymbol{Z}_p)$ and $cov[vec((\boldsymbol{Z}_f\boldsymbol{Z}_f')^{-\frac{1}{2}}\hat{\boldsymbol{O}})]$, tend asymptotically to $T_*^{-1}\bar{\boldsymbol{\Pi}}$. ■

Proof of Proposition 3. As matrix $\bar{\boldsymbol{\Pi}}$ is known, it is straightforward to see that not all the elements in $\boldsymbol{A}$ are independent, except when $i = 1$, that implies $\bar{\boldsymbol{\Pi}} = \boldsymbol{I}_{m^2}$. Given the structure of $\bar{\boldsymbol{\Pi}}$ and using the submatrix Matlab notation: (i) The first im elements of $vec\boldsymbol{A}$, which are $\boldsymbol{A}_{1:im,1:m}$, are uncorrelated as the square submatrix $\bar{\boldsymbol{\Pi}}_{1:im} = \boldsymbol{I}_{im^2}$, and (ii) as the first m rows of $\bar{\boldsymbol{\Pi}}_{i-1}'$ are zeros, then the elements of the submatrix $\boldsymbol{A}_{1:m,m+1:m+2}$ are also uncorrelated with those of $\boldsymbol{A}_{1:im,1:m}$. This occurs for every element in the submatrix $\boldsymbol{A}_{1:m,m+1:im}$ due to the structure of zeros in $\bar{\boldsymbol{\Pi}}_{i-k}'$, $k = 1, 2, \dots, i-1$. Then the elements in $\boldsymbol{A}_{1:m,m+1:im}$ are uncorrelated with those of $\boldsymbol{A}_{1:im,1:m}$ and, therefore, $\bar{\boldsymbol{\Pi}}$ is of rank $m^2(2i-1)$. In order to extract m^2k independent elements from $\boldsymbol{A}$, we use the singular value decomposition (SVD) of $\bar{\boldsymbol{\Pi}}$, yielding a matrix $\boldsymbol{B}_{(im)^2\times m^2k}$ such that $\bar{\boldsymbol{\Pi}} \overset{\text{svd}}{=} \boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{V}' = \boldsymbol{B}\boldsymbol{B}'$. Consequently, we have $\boldsymbol{B}^\dagger\bar{\boldsymbol{\Pi}}\boldsymbol{B}'^\dagger = \boldsymbol{I}_{m^2k}$, where '†' denotes the Moore-Penrose pseudo inverse, and $\boldsymbol{B}^\dagger vec(\boldsymbol{A}) \overset{d}{\to} N\big(\boldsymbol{0}, T_*^{-1}\boldsymbol{I}_{m^2k}\big)$ which leads to $\mathcal{S}_{\boldsymbol{A}} = T_*vec(\boldsymbol{A})'\boldsymbol{P}vec(\boldsymbol{A}) \overset{d}{\to} \chi^2_{m^2k}$, $\boldsymbol{P} = \boldsymbol{B}'^\dagger\boldsymbol{B}^\dagger$ being a symmetric idempotent matrix of rank m^2k. ■

Proof of Proposition 4. Let the rth autocovariance matrix of the innovations be $\boldsymbol{C}_r = T^{-1}\boldsymbol{\psi}_t\boldsymbol{\psi}_{t-r}'$ and the rth residual autocovariance matrix be $\hat{\boldsymbol{C}}_r = T^{-1}\hat{\boldsymbol{\psi}}_t\hat{\boldsymbol{\psi}}_{t-r}'$. Further, define $\boldsymbol{C} = (\boldsymbol{C}_1\,\boldsymbol{C}_2\,\dots\,\boldsymbol{C}_k)$ and similary $\hat{\boldsymbol{C}}$. (Hosking 1980) proved that $vec(\hat{\boldsymbol{C}}) = \boldsymbol{D}vec(\boldsymbol{C})$ where $\boldsymbol{D}$ is idempotent of rank $m^2(k-p-q)$. Let $\hat{\bar{\boldsymbol{\beta}}}_{\boldsymbol{*}}$ be as in (5) but using $\bar{\boldsymbol{z}}_t$ instead of $\boldsymbol{z}_t$ and assuming that $\bar{\boldsymbol{z}}_t$ are the standardized residuals from a VARMA(p,q) model. In such a case, $\hat{\bar{\boldsymbol{\beta}}}_{\boldsymbol{*}} \overset{a.s.}{\to} \hat{\mathbb{C}}(\boldsymbol{I}_i \otimes \boldsymbol{I}_m)^{-1} = \hat{\mathbb{C}}$ where:

$$\hat{\mathbb{C}} = \begin{pmatrix} \hat{\boldsymbol{C}}_{\bar{k}-i+1} & \hat{\boldsymbol{C}}_{\bar{k}-i} & \dots & \hat{\boldsymbol{C}}_1 \\ \hat{\boldsymbol{C}}_{\bar{k}-i+2} & \hat{\boldsymbol{C}}_{\bar{k}-i+1} & \dots & \hat{\boldsymbol{C}}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{C}}_{\bar{k}} & \hat{\boldsymbol{C}}_{\bar{k}-1} & \dots & \hat{\boldsymbol{C}}_{\bar{k}-i+1} \end{pmatrix}_{im} \quad \text{with } \bar{k} \equiv \begin{cases} k \text{ if } k \text{ is odd} \\ k+1 \text{ if } k \text{ is even.} \end{cases} \tag{10}$$

Then, we can write $\boldsymbol{B}^\dagger vec(\hat{\bar{\boldsymbol{\beta}}}_*) = \bar{\boldsymbol{D}}\boldsymbol{B}^\dagger vec(\hat{\bar{\boldsymbol{\beta}}})$ as it was done by (Hosking 1980), since $\boldsymbol{B}^\dagger vec(\hat{\bar{\boldsymbol{\beta}}}_{\boldsymbol{*}})$ and $\boldsymbol{B}^\dagger vec(\hat{\bar{\boldsymbol{\beta}}})$ have, asymptotically, the same elements as $vec(\hat{\boldsymbol{C}})$ and $vec(\boldsymbol{C})$, respectively, but sorted in different order. Likewise, $\bar{\boldsymbol{D}}$ has the same rows as $\boldsymbol{D}$, but ordered differently, that yields $rank(\bar{\boldsymbol{D}}) = rank(\boldsymbol{D}) = m^2(k-p-q)$. Finally, we previously showed that $\boldsymbol{B}^\dagger vec(\hat{\bar{\boldsymbol{\beta}}}|\boldsymbol{Z}_p) \overset{d}{\to} N\big(\boldsymbol{0}, T_*^{-1}\boldsymbol{I}_{m^2k}\big)$ and, consequently, $\boldsymbol{B}^\dagger vec(\hat{\bar{\boldsymbol{\beta}}}_*|\boldsymbol{Z}_p) \overset{d}{\to} N\big(\boldsymbol{0}, T_*^{-1}\bar{\boldsymbol{D}}\big)$, which leads to $T_*vec(\hat{\bar{\boldsymbol{\beta}}}_*)'\boldsymbol{P}vec(\hat{\bar{\boldsymbol{\beta}}}_*) \overset{d}{\to} \chi^2_{m^2(k-p-q)}$. ■