Statistica Sinica Preprint No: SS-2016-0286.R1	
Title	AN IMPROVED MEASURE FOR LACK OF FIT IN TIME SERIES MODELS
Manuscript ID	SS-2016-0286.R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0286
Complete List of Authors	Thomas Fisher and Michael Robbins
Corresponding Author	Thomas Fisher
E-mail	fishert4@miamioh.edu
Notice: Accepted version subject to English editing.	

AN IMPROVED MEASURE FOR LACK OF FIT IN TIME SERIES MODELS

Thomas J. Fisher and Michael W. Robbins

Miami University and RAND Corporation

Abstract: The correlation structure of time series is of fundamental importance in diagnostic procedures. The squared autocorrelation function of the residuals of a fitted model is generally used as a measure of the goodness-of-fit; multivariate analogues are available for vector time series. As an alternative, we propose a logarithmic transformation of the determinant of a constructed Toeplitz matrix containing the typical measure of correlation. Theoretical results demonstrate the proposed measure is asymptotically more powerful than the typical measure of correlation (when used with or without the Ljung–Box correction) in the detection of a variety of residual dependence structures. The proposed method is shown to have utility when applied in conjunction to a host of methods used to diagnose the fit of strong and weak autoregressive moving average models and generalized autoregressive conditional heteroskedastic models. A simulation study demonstrates the effectiveness of the proposed method and illustrates its improvement over the existent procedures.

Key words and phrases: Autocorrelation; GARCH; Goodness-of-fit; Portmanteau; Vector ARMA

1. Introduction

With the recent explosion in the size and availability of data, accompanied by an interest in predictive modeling and analytics, the importance of the field of time series analysis continues to augment. Whether using time series regression or via the Box–Jenkins approach, it is well known that proper modeling of any serial correlation in a time series is essential for forecasting and likewise that proper modeling of the variability is essential for the accuracy of prediction intervals. Assessing the adequacy of a fitted model is an important diagnostic step in time series analysis.

In practice a time series is nearly always accompanied by a multitude of associated series that may provide supplemental information. Consider the possible inter-related economic indicators, for example. For analysis of multivariate time series, practitioners often assume a series has a stationary (vector) autoregressive moving average (VARMA/ARMA) representation. A d-dimensional time series $\{\mathbf{X}_t\}$ with mean vector $\boldsymbol{\mu}$ is said to have a VARMA representation if, for all $t \in \mathbb{Z}$,

$$\mathbf{X}_t - \boldsymbol{\mu} = \sum_{i=1}^{p} \boldsymbol{\Phi}_i \left(\mathbf{X}_{t-i} - \boldsymbol{\mu}\right) + \sum_{j=1}^{q} \boldsymbol{\Theta}_j \boldsymbol{\epsilon}_{t-j} + \boldsymbol{\epsilon}_t, \tag{1.1}$$

where $\{\boldsymbol{\epsilon}_t\}$ is a sequence of mean-zero error vectors, known as the innovations, with finite covariance $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. The terms $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Theta}_j$ are $d \times d$ matrices of vector autoregressive and moving average coefficients, respectively, for $i = 1, \ldots, p$ and $j = 1, \ldots, q$ where p is the autoregressive order and q is the order of the moving average. We note that for $d = 1$, we have the well-known ARMA model and in most multivariate applications practitioners use VAR models for ease-of-use and the lack of uniqueness in a VARMA covariance structure, see Wei (2006). When the innovations are an independent and identically distributed (iid) sequence, the model in (1.1) is called a *strong* VARMA; whereas, if the innovations are dependent but uncorrelated, it is referred to as a *weak* VARMA.

Assume that $\sqrt{n}$-consistent estimates $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Phi}}_1, \ldots, \hat{\boldsymbol{\Phi}}_p, \hat{\boldsymbol{\Theta}}_1, \ldots,$ and $\hat{\boldsymbol{\Theta}}_q$ have been calculated using the observed series $\{\mathbf{X}_1, \ldots \mathbf{X}_n\}$; it follows from Dunsmuir and Hannan (1976) that such estimates exist under the stated conditions. The adequacy of the fit is checked based on the serial correlation structure of the fitted residuals, $\hat{\boldsymbol{\epsilon}}_1, \ldots, \hat{\boldsymbol{\epsilon}}_n$, which are calculated to satisfy

$$\hat{\boldsymbol{\epsilon}}_t = (\mathbf{X}_t - \hat{\boldsymbol{\mu}}) - \sum_{i=1}^{p} \hat{\boldsymbol{\Phi}}_i \left(\mathbf{X}_{t-i} - \hat{\boldsymbol{\mu}}\right) - \sum_{j=1}^{q} \hat{\boldsymbol{\Theta}}_j \hat{\boldsymbol{\epsilon}}_{t-j},$$

for $t = 1, \ldots, n$. Equivalently, we look to statistically test H_0 : no serial correlation remains in the residuals, versus, H_1 : serial correlation remains. This can be accomplished by visually exploring the correlogram or performing a formal hypothesis test. Our focus here is the latter.

In the case of a *weak* VARMA process, the residuals are often assumed to follow a vector generalized autoregressive conditional heteroskedastic (GARCH) process

$$\boldsymbol{\epsilon}_t = \mathbf{H}_t^{1/2}\boldsymbol{\eta}_t$$

where the $d \times d$ matrix $\mathbf{H}_t$ is the conditional covariance matrix of $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ is an iid vector process such that $E\left[\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top\right] = \mathbf{I}_d$, where $\mathbf{I}_j$ denotes the $j \times j$ identity matrix and $\mathbf{A}^\top$ represents the transpose of matrix $\mathbf{A}$. Many parametric formulations for the matrix process $\mathbf{H}_t$ exists and a review can be found in Silvennoinen and Teräsvirta (2009). When $d = 1$ (where $\mathbf{H}_t = h_t$), this is the GARCH process of Engle (1982) and Bollerslev (1986):

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}. \tag{1.2}$$

We note that (1.2) is essentially an ARMA process on ϵ_t^2 terms. When modeling and assessing the fit of a GARCH process, one typically concentrates on the square of the residual series from the (V)ARMA model.

In this article, we explore the problem of goodness-of-fit testing for a fitted time series. Our primary goal is to introduce a new measure of correlation that is used to enhance the power of a plethora of extant test statistics that have been proposed for the purpose of assessing the goodness-of-fit for time series models in a wide variety of settings. In Section 2, we introduce the pertinent methods for measuring serial correlation in a time series and provide our new, more powerful, measure for serial correlation. Section 3 reviews several members of the class of so-called portmanteau tests, and introduces analogues of these tests that are based on the proposed measure of correlation. Section 4 provides simulations showing the proposed method can provide substantial power increases while retaining type I error rates and discussion follows in Section 5.

2. Measures of Correlation

2.1 Traditional Measure

The autocovariance function is arguably the foundational tool of time series analysis. The value of this function realized at lag k in a d-dimensional stationary time series $\{\mathbf{X}_t\}$ with mean vector $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\Gamma}_k(\mathbf{X}_t) = E[(\mathbf{X}_t - \boldsymbol{\mu})\,(\mathbf{X}_{t-k} - \boldsymbol{\mu})^\top].$$

Note that the operand (indicated as $\mathbf{X}_t$ above) in this quantity (and in those defined below) is used to indicate the process over which the quantity is being calculated. In practice, the preferred tool for monitoring intra-series dependence is the autocorrelation function, which is defined here using $\mathbf{R}_k(\mathbf{X}_t) = \mathbf{L}(\mathbf{X}_t)^\top \boldsymbol{\Gamma}_k(\mathbf{X}_t)\mathbf{L}(\mathbf{X}_t)$, where $\mathbf{L}(\mathbf{X}_t)$ is the lower Cholesky decomposition of $\mathbf{\Gamma}_0^{-1}(\mathbf{X}_t)$ (note that the usefulness of other manners of defining multivariate autocorrelation are discussed at the end of Section 2.2). When $d = 1$, these two components are estimated in the typical way,

$$\hat{\gamma}_k(x_t) = \frac{1}{n}\sum_{t=k+1}^{n}\left(x_t - \bar{x}\right)\left(x_{t-k} - \bar{x}\right) \qquad \text{and} \qquad \hat{\rho}_k(x_t) = \frac{\hat{\gamma}_k(x_t)}{\hat{\gamma}_0(x_t)},$$

for sample mean $\bar{x}$; in the multivariate setting, the estimators are

$$\hat{\mathbf{\Gamma}}_k(\mathbf{X}_t) = \frac{1}{n}\sum_{t=k+1}^{n}\left(\mathbf{X}_t - \bar{\mathbf{X}}\right)\left(\mathbf{X}_{t-k} - \bar{\mathbf{X}}\right)^{\top}$$

and

$$\hat{\mathbf{R}}_k(\mathbf{X}_t) = \hat{\mathbf{L}}(\mathbf{X}_t)^{\top}\hat{\mathbf{\Gamma}}_k(\mathbf{X}_t)\hat{\mathbf{L}}(\mathbf{X}_t), \tag{2.1}$$

for $\hat{\mathbf{L}}(\mathbf{X}_t)$, the lower Cholesky decomposition of $\hat{\mathbf{\Gamma}}_0^{-1}(\mathbf{X}_t)$, and sample mean vector $\bar{\mathbf{X}}$. For the purposes of diagnostic procedures, we look at the correlation structure of the fitted residual series $\{\hat{\boldsymbol{\epsilon}}_t\}$. To simplify notation, in the remainder of this article we will use $\hat{\rho}_k = \hat{\rho}_k(\hat{\epsilon}_t)$, $\hat{\mathbf{\Gamma}}_k = \hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\epsilon}}_t)$, and $\hat{\mathbf{R}}_k = \hat{\mathbf{R}}_k(\hat{\boldsymbol{\epsilon}}_t)$ unless otherwise noted.

Since the term $\hat{\rho}_k^2$ effectively indicates the presence of residual serial correlation at lag k, Box and Pierce (1970) construct a goodness-of-fit statistic for univariate time series using a sum of the squared sample autocorrelation function. In that vein, for the purpose of diagnosing the fit of a VARMA model, it is useful to condense all the terms of the matrix $\hat{\mathbf{R}}_k$ into a single value that gauges the magnitude of serial correlations at lag k. Such a quantity can serve as a statistic for testing whether or not at least one of the elements of $\hat{\boldsymbol{\Gamma}}_k$ (or $\hat{\mathbf{R}}_k$) is nonzero. Hosking (1980) suggests

$$\tilde{h}_k = (\text{vec}\hat{\boldsymbol{\Gamma}}_k)^\top(\hat{\boldsymbol{\Gamma}}_0^{-1} \otimes \hat{\boldsymbol{\Gamma}}_0^{-1})\text{vec}\hat{\boldsymbol{\Gamma}}_k = (\text{vec}\hat{\mathbf{R}}_k)^\top \text{vec}\hat{\mathbf{R}}_k = \text{tr}(\hat{\mathbf{R}}_k^\top \hat{\mathbf{R}}_k), \quad (2.2)$$

where vec$\mathbf{A}$ represents the columns of matrix $\mathbf{A}$ stacked on top of one another, $\mathbf{A} \otimes \mathbf{B}$ represents the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$ and tr$(\mathbf{A})$ indicates the trace of matrix $\mathbf{A}$. Note that with univariate data, $\tilde{h}_k = \hat{\rho}_k^2$; therefore, $\tilde{h}_k$ is a multivariate generalization of the measure used by Box and Pierce (1970). For VARMA models, it follows that $n\tilde{h}_k$ in (2.2) is asymptotically distributed as a linear combination of d^2 iid χ^2_1 random variables, where χ^2_ν denotes a chi-squared distribution with ν degrees of freedom and the coefficients in the combination are set as the eigenvalues of the covariance matrix of vec$\hat{\mathbf{R}}_k$ (see Hosking, 1980; Li and McLeod, 1981).

In moderate sample sizes, the distribution of $n\tilde{h}_k$ is known to be poorly approximated by its limiting distribution—test statistics that invoke the measure can be highly conservative in practical settings. Ljung and Box (1978) suggest that the performance can be improved by multiplying the squared correlation by a correction factor that depends on k. The multivariate analogue of the Ljung–Box corrected measure is $\tilde{h}_k^* = n\tilde{h}_k/(n-k)$; see Hosking (1980). Herein, any statistic that employs $\tilde{h}_k^*$ is referred to an LB-type statistic, whereas one that utilizes $\tilde{h}_k$ is called BP-type.

For univariate time series, an alternative measure of correlation is given by the partial autocorrelation function, which measures the remaining correlation at lag k after accounting for correlation at lower lags. Monti (1994) justifies a LB-type correction on the partial autocorrelation–many of the goodness-of-fit statistics described in the following section can be constructed using this measure as well.

The state-of-the-art for goodness-of-fit testing in time series has advanced well beyond the findings above. Nonetheless, most goodness-of-fit test statistics are calculated using the classic BP-type or LB-type measure of correlation. Below, we introduce a new measure that provides more power in detecting serial correlation while retaining the same asymptotic distribution under the null hypothesis.

2.2 Proposed Measure

We propose a block Toeplitz matrix for the purpose of gauging the magnitude of autocorrelation at the k^{th} lag within the residuals of a fitted time series. We present our contribution in the multivariate setting (which encompasses the univariate). For lag-k autocorrelation matrix $\hat{\mathbf{R}}_k$, consider

$$\tilde{\mathbf{R}}_k = \begin{bmatrix} \mathbf{I}_d & \hat{\mathbf{R}}_k \\ \hat{\mathbf{R}}_k^\top & \mathbf{I}_d \end{bmatrix}. \tag{2.3}$$

Under the null hypothesis of no residual series correlation, the matrix $\tilde{\mathbf{R}}_k$ should be, for $k \neq 0$, statistically equivalent to $\mathbf{I}_{2d}$.

Borrowing from the framework of Robbins and Fisher (2015), establishment of relevant properties regarding $\tilde{\mathbf{R}}_k$ mandates the following lemma. Akin to Hosking (1980), assume that the observed series $\{\mathbf{X}_t\}$ obeys the model in (1.1) and that the sequence of innovations has finite variance.

Lemma 1. *The eigenvalues of* $\tilde{\mathbf{R}}_k$ *are symmetric about 1.*

Proof. Let $\tilde{\mathbf{R}}_k^0 = \tilde{\mathbf{R}}_k - \mathbf{I}_{2d}$ and λ be an eigenvalue of $\tilde{\mathbf{R}}_k^0$ with corresponding eigenvector $(\mathbf{e}_1, \mathbf{e}_2)^\top$. Straightforward algebra shows $-\lambda$ is an eigenvalue with associated eigenvector $(\mathbf{e}_1, -\mathbf{e}_2)^\top$. It follows that the eigenvalues of $\tilde{\mathbf{R}}_k$ are of the form $1 \pm \lambda$. □

To measure the amount of serial correlation at lag k, we propose:

$$\tilde{r}_k = -\log\det\tilde{\mathbf{R}}_k, \tag{2.4}$$

where $\det\mathbf{A}$ indicates the determinant of a matrix $\mathbf{A}$. Note that we can write $\tilde{\mathbf{R}}_k = \tilde{\mathbf{L}}^\top\tilde{\boldsymbol{\Gamma}}_k\tilde{\mathbf{L}}$, where $\tilde{\mathbf{L}}$ is a $2d \times 2d$ block-diagonal matrix with $\hat{\mathbf{L}}(\hat{\boldsymbol{\epsilon}}_t)$

along the diagonal, and $\tilde{\mathbf{\Gamma}}_k$ is a $2d \times 2d$ matrix with $\hat{\mathbf{\Gamma}}_0$ on the diagonal and $\hat{\mathbf{\Gamma}}_k$ ($\hat{\mathbf{\Gamma}}_k^\top$) on the upper-right (lower-left) diagonal. Using this representation, calculations show that $\tilde{\mathbf{R}}_k$ is positive definite in practice. This observation and Lemma 1 yield $0 < \det \tilde{\mathbf{R}}_k \leq 1$, and therefore $\tilde{r}_k$ exists and is nonnegative. The large sample properties of $\tilde{r}_k$ are illustrated as follows.

Theorem 1. *Under H_0 of no serial correlation, $n\tilde{r}_k$ and $n\tilde{h}_k$ from (2.2) are equivalent asymptotically. Specifically, $n(\tilde{r}_k - \tilde{h}_k) = O_p(n^{-1})$. Furthermore, $n\tilde{\mathbf{r}}_m = n(\tilde{r}_1, \ldots, \tilde{r}_m)^\top$ and $n\tilde{\mathbf{h}}_m = n(\tilde{h}_1, \ldots, \tilde{h}_m)^\top$, for $m \in \mathbb{Z}^+$, share the same asymptotic joint distribution.*

Proof. Let $\lambda_1, \lambda_2, \ldots, \lambda_{2d}$ be the eigenvalues of $\tilde{\mathbf{R}}_k$. Note

$$\begin{aligned} n\sum_{i=1}^{2d}(\lambda_i - 1)^2 &= n\mathrm{tr}(\tilde{\mathbf{R}}_k^\top \tilde{\mathbf{R}}_k) - 2dn \\ &= 2dn + 2n\mathrm{tr}(\hat{\mathbf{R}}_k^\top \hat{\mathbf{R}}_k) - 2dn \\ &= 2n\tilde{h}_k \end{aligned}$$

where the last line follows from the results in Hosking (1981). Next,

$$\begin{aligned} n\tilde{r}_k &= -n\log\det\tilde{\mathbf{R}}_k = -n\log\prod_{i=1}^{2d}\lambda_i \\ &= n\sum_{i=1}^{2d}\left[(\lambda_i - 1)^2/2 + (\lambda_i - 1)^4/4 + (\lambda_i - 1)^6/6 + \cdots\right] \\ &= n\tilde{h}_k + O_p(n^{-1}), \end{aligned} \tag{2.5}$$

where the second line holds since all odd powers are zero by Lemma 1. Following Eaton and Tyler (1991), each λ_k will consistently approximate unity under H_0, and the rate of convergence in the third line comes from the $\sqrt{n}$-consistency of parameter estimates. The above argument holds for all $k = 1, \ldots, m$, whence $n\tilde{\mathbf{r}}_m$ and $n\tilde{\mathbf{h}}_m$ are asymptotically equivalent. $\square$

Large values of the correlation measure $\tilde{h}_k$ indicate the presence of nonzero lag-k serial correlation. By illustrating that $\tilde{r}_k$ is at least as large as $\tilde{h}_k$ and is, in fact, divergent from it under H_1 (as stated formally below), we infer that goodness-of-fit statistics which utilize our measure are more powerful asymptotically than those that use the BP-type measure. Note that we focus on fixed alternative hypothesis models; that is, we do not consider local alternatives (although these are briefly discussed in Section 5), wherein the departure of the true model from the null hypothesis specification vanishes as n increases. Therefore, we can assume that $\lambda_j - 1 \xrightarrow{p} c$ where $c \neq 0$ for some j and where $\xrightarrow{p}$ denotes convergence in probability.

Theorem 2. *The measure $n\tilde{r}_k$ is more powerful than $n\tilde{h}_k$ at detecting serial correlation at lag k, given that critical values are obtained using the same asymptotic approximations. Furthermore, the discrepancy between the measures diverge at the rate of n.*

Proof. Let $A_k = n(\tilde{r}_k - \tilde{h}_k)$ which will consist of terms of the form $n\sum(\lambda_i - 1)^l$ for even values of $l \geq 4$, and therefore, $A_k > 0$. Under the alternative hypothesis, $\lambda_j - 1 \xrightarrow{p} c \neq 0$ for some j, it follows that $A_k = O_p(n)$. □

Although the LB-type measure $\tilde{h}_k^*$ is designed to offer improved power over $\tilde{h}_k$, we see an analog of Theorem 2 holds when we compare $\tilde{r}_k$ to $\tilde{h}_k^*$.

Corollary 1. *The measure $n\tilde{r}_k$ is more powerful than $n\tilde{h}_k^*$ at detecting serial correlation at lag k.*

Proof. Recall the LB-type measure $\tilde{h}_k^*$ and note that $B_k = n(\tilde{h}_k^* - \tilde{h}_k) = nk/(n-k)\tilde{h}_k$ and under H_1, $\tilde{h}_k = O_p(1)$, whence $B_k = O_p(1)$. Further, $n(\tilde{r}_k - \tilde{h}_k^*) = O_p(n)$ while $P(\tilde{r}_k > \tilde{h}_k^*) \to 1$. □

We note from the above that the discrepancy between the LB-type and BP-type measures is bounded whereas the discrepancy between our measure and the LB-type measure is unbounded. This implies that, asymptotically, our measure $\tilde{r}_k$ offers improvement in detection capability not offered by the LB-type. However, small sample performance could deviate.

Our measure can be motivated using likelihood ratio principles. To illustrate, first define the vector $\boldsymbol{\Xi}_t = (\boldsymbol{\epsilon}_t^\top, \boldsymbol{\epsilon}_{t+1}^\top, \ldots, \boldsymbol{\epsilon}_{t+k}^\top)^\top$. When H_0 is true, the covariance matrix of $\boldsymbol{\Xi}_t$ can be approximated via $\hat{\boldsymbol{\mathcal{G}}}_k^*$, where $\hat{\boldsymbol{\mathcal{G}}}_k^*$ is a $d(k+1) \times d(k+1)$ block diagonal matrix where the diagonal blocks are each set as $\hat{\mathbf{\Gamma}}_0$. Consider an alternative hypothesis that allows $\mathbf{\Gamma}_k(\boldsymbol{\epsilon}_t) \neq \mathbf{0}$ while enforcing $\mathbf{\Gamma}_{k'}(\boldsymbol{\epsilon}_t) = \mathbf{0}$ for $k' \neq k$. Therein, the covariance matrix of $\mathbf{\Xi}_t$ is estimated using $\hat{\boldsymbol{\mathcal{G}}}_k$, which is identical to $\hat{\boldsymbol{\mathcal{G}}}_k^*$ with the exception that the upper-right $d \times d$ block is set to $\hat{\mathbf{\Gamma}}_k$ and likewise the lower-left block is set as $\hat{\mathbf{\Gamma}}_k^\top$. Lastly, let $\hat{\boldsymbol{\mathcal{L}}}_k^*$ denote the lower triangular Cholesky decomposition of $(\hat{\boldsymbol{\mathcal{G}}}_k^*)^{-1}$; it follows that $\hat{\boldsymbol{\mathcal{L}}}_k^*$ is block-diagonal where each diagonal block is given by $\hat{\mathbf{L}}(\hat{\boldsymbol{\epsilon}}_t)$. Gaussian likelihood ratio statistics for multivariate data are frequently set as the ratio of the determinant of a covariance matrix calculated under an alternative hypothesis and the determinant of a covariance matrix calculated under the corresponding null hypothesis. We observe

$$\frac{\det \hat{\boldsymbol{\mathcal{G}}}_k}{\det \hat{\boldsymbol{\mathcal{G}}}_k^*} = \det((\hat{\boldsymbol{\mathcal{L}}}_k^*)^\top \hat{\boldsymbol{\mathcal{G}}}_k \hat{\boldsymbol{\mathcal{L}}}_k^*) = \det \tilde{\boldsymbol{\mathcal{R}}}_k,$$

where $\tilde{\boldsymbol{\mathcal{R}}}_k$ is equivalent to a $d(k+1) \times d(k+1)$ identity matrix with the top-right and lower-left blocks replaced with $\hat{\mathbf{R}}_k$ and $\hat{\mathbf{R}}_k^\top$, respectively. Using well known formulas for the determinant of a block-partitioned matrix,

$$\det \tilde{\boldsymbol{\mathcal{R}}}_k = \det(\mathbf{I}_d - \hat{\mathbf{R}}_k^\top \hat{\mathbf{R}}_k) = \det \tilde{\mathbf{R}}_k.$$

Therefore, we describe $\tilde{r}_k$ as a likelihood ratio-type statistic.

We also note that autocorrelation matrices in multivariate time series have been defined within the literature via expressions differing from (2.1). For instance, Chitturi (1974) defines residual autocorrelation via $\hat{\mathbf{R}}_k^{(\dagger)} = \hat{\mathbf{\Gamma}}_k \hat{\mathbf{\Gamma}}_0^{-1}$. We use arguments posited by Mahdi and McLeod (2012) to illustrate that if we define $\tilde{\mathbf{R}}_k^{(\dagger)}$, an analogue of (2.3), by setting the top-left block equal to $\hat{\mathbf{R}}_{-k}^{(\dagger)}$ (note that $\hat{\mathbf{\Gamma}}_{-k} = \hat{\mathbf{\Gamma}}_k^{\top}$) and the bottom-right block equal to $\hat{\mathbf{R}}_k^{(\dagger)}$, it holds that $\det \tilde{\mathbf{R}}_k^{(\dagger)} = \det \tilde{\mathbf{R}}_k$. Therefore, $\tilde{r}_k$ may be equivalently calculated by using $\tilde{\mathbf{R}}_k^{(\dagger)}$ in place of $\tilde{\mathbf{R}}_k$. However, if we calculate residual autocorrelation by using (2.1) with $\hat{\mathbf{L}}$ replaced by a diagonal matrix that has the inverse of the square root of the diagonal elements of $\hat{\mathbf{\Gamma}}_0$ along its diagonal (this gives the traditional definition of correlation), we cannot use the calculations that yield $\tilde{r}_k$ to extract a useful measure.

Lastly, note that the Ljung-Box correction can be used in conjunction with our measure of correlation. For instance, define $\tilde{r}_k^* = n\tilde{r}_k/(n-k)$. This measure is asymptotically equivalent to and more powerful than each of $\tilde{h}_k$, $\tilde{h}_k^*$, and $\tilde{r}_k$. However, goodness-of-fit statistics based on $\tilde{r}_k^*$ tend to have a slightly liberal type I error in finite samples, and as such, further discussion of this measure is withheld till Section 5.

3. Portmanteau Statistics

In practice, correlation at a single lag is rarely considered when assessing the adequacy of a fitted time series model. Instead, practitioners look at the serial correlation at a multitude of lags; this leads to the so-called portmanteau test. In the ensuing subsections a wide variety of portmanteau test statistics are illustrated for use in settings involving independent innovations as well as innovations that are uncorrelated but dependent.

Each of the statistics outlined below, as originally described in the literature, is constructed using the BP-type or LB-type measure of correlation. We propose revised versions that substitute our measure of correlation. As a consequence of Theorem 1, the new statistics have the same asymptotic distribution as their respective BP-type and LB-type versions. From Theorem 2 and Corollary 1 we see that statistics which employ our measure are more powerful asymptotically than those that use the BP-type or LB-type measures. The model assumptions required by each statistic that is defined using our proposed measure are the same as those required by its BP- or LB-type analogue; this follows from (2.5).

3.1 Independent Innovations

In the seminal work of Box and Pierce (1970), the portmanteau test for time series goodness-of-fit testing in the univariate setting is introduced. Therein, the asymptotic distribution of the autocorrelation function is derived for the residuals from a fitted ARMA model with iid innovations. The goodness-of-fit test statistic of Box and Pierce (1970) is defined as the sum of first m (where m is the maximum lag considered) squared residual autocorrelations. As alluded to previously, Hosking (1980) extends the findings of Box and Pierce (1970) to the multivariate setting. Therein, the foundational BP-type and LB-type portmanteau test statistics are written

$$Q_m = n\sum_{k=1}^{m}\tilde{h}_k \quad \text{and} \quad Q_m^* = n\sum_{k=1}^{m}\tilde{h}_k^*, \tag{3.1}$$

respectively, where $\tilde{h}_k$ is as defined in (2.2) and $\tilde{h}_k^* = n\tilde{h}_k/(n-k)$. Both Q_m and Q_m^* follow a $\chi^2_{d^2(m-p-q)}$ distribution for large n (Hosking, 1980).

A version of Q_m that utilizes our measure of correlation is expressed

$$\tilde{Q}_m = n\sum_{k=1}^{m}\tilde{r}_k, \tag{3.2}$$

where $\tilde{r}_k$ was defined in (2.4). As noted above, from Theorem 1 it follows that $\tilde{Q}_m$ will have the same limit behavior as Q_m and Q_m^* under H_0. Likewise, the improvement in power offered by $\tilde{Q}_m$ over Q_m and Q_m^* follows from Theorem 2 and Corollary 1.

3.2 Uncorrelated but Dependent Innovations

Over the past three decades, there has been growing interest in non-linear time series models, particularly those that model heteroskedasticity such as the GARCH model and the Stochastic Volatility model of Taylor

(1986). Therein, the error series is uncorrelated but not independent. Recall that time series which satisfy (1.1) with a uncorrelated but dependent error structure are said to have a *weak* VARMA representation. As shown in Romano and Thombs (1996) and Francq et al. (2005), the methods of Box and Pierce (1970) can be quite poor under the assumption of uncorrelated innovations rather than the stronger assumption of independence.

Many authors have explored this problem by developing methods for uncorrelated innovations. Shao (2011) showed weighting the Box–Pierce test will provide some robustness to the uncorrelated error problem if the maximum lag m grows with the sample size. Lobato (2001) provides a statistic for a *weak* ARMA fit whose asymptotic null distribution is not standard. A robust version of the Box–Pierce measure that includes second moment information of the residuals is discussed in Lobato et al. (2001).

In Lobato et al. (2002) and Francq et al. (2005), the asymptotic distribution of Q_m is found under some weak assumptions that allows for dependent innovations such as a GARCH process. However, in those settings, unlike in Box and Pierce (1970), the covariance matrix of $\hat{\boldsymbol{\rho}}_m = (\hat{\rho}_1, \ldots, \hat{\rho}_m)^\top$ does not have a simple form. Those authors present methods to consistently estimate the covariance matrix and provide an alternative distribution to the BP-type test when the innovations are uncorrelated. In Francq and Raïssi

(2007), these results are generalized to the multivariate setting wherein one fits a VAR model. In such settings, Q_m and Q_m^* from (3.1) and $\tilde{Q}_m$ from (3.2) are asymptotically distributed as a linear combination of iid χ_1^2 variates where the coefficients are the eigenvalues of

$$\mathbf{\Sigma}_{Q_m} = \left(\mathbf{I}_m \otimes \mathbf{\Sigma}_{\boldsymbol{\epsilon}}^{-1/2} \otimes \mathbf{\Sigma}_{\boldsymbol{\epsilon}}^{-1/2}\right) \mathbf{\Sigma}_{\boldsymbol{\gamma}} \left(\mathbf{I}_m \otimes \mathbf{\Sigma}_{\boldsymbol{\epsilon}}^{-1/2} \otimes \mathbf{\Sigma}_{\boldsymbol{\epsilon}}^{-1/2}\right).$$

In the above expression, recall from (1.1) that $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$ is the covariance of the time series innovations. Further, $\mathbf{\Sigma}_{\boldsymbol{\gamma}}$ is the covariance matrix of $\boldsymbol{\gamma} = \left(\{\text{vec}\mathbf{\Gamma}_1(\hat{\boldsymbol{\epsilon}}_t)\}^\top, \ldots, \{\text{vec}\mathbf{\Gamma}_m(\hat{\boldsymbol{\epsilon}}_t)\}^\top\right)^\top$ and models nuisance parameters in the covariance of Q_m. This result follows from Francq and Raïssi (2007) and Theorem 2. Francq and Raïssi (2007) provide an algorithm for a consistent estimator of $\mathbf{\Sigma}_{Q_m}$ based on $\hat{\mathbf{\Sigma}}_{\boldsymbol{\epsilon}}$ and an autoregressive spectral estimator (see den Haan and Levin, 1997) for determining $\boldsymbol{\gamma}$. The distribution of Q_m (Q_m^* and $\tilde{Q}_m$) can be determined numerically via the algorithm of Imhof (1961) or by a gamma approximation from Box (1954) (used in our simulations).

3.3 Weighted Methods

In practice, residual autocorrelation in ill-fit models of stationary processes tends to gravitate towards lower lags. Therefore, weighted portmanteau tests, wherein the option of emphasizing certain lags over others, are gaining in popularity (see Hong, 1996a; Fisher and Gallagher, 2012; Mahdi and McLeod, 2012; Gallagher and Fisher, 2015, for example).

Most of the published work discussing general schemes for weighting portmanteau tests considers univariate data only (see Gallagher and Fisher, 2015, for example). However, multivariate analogues of these techniques can be developed by applying the weighting mechanisms discussed in these references to the statistic of Hosking (1980) (although we are unaware of any published results demonstrating their utility). Specifically, consider weighted versions of (3.1) and (3.2):

$$Q_m^w = n\sum_{k=1}^{m} w_k \tilde{h}_k, \quad Q_m^{w*} = n\sum_{k=1}^{m} w_k \tilde{h}_k^*, \quad \text{and} \quad \tilde{Q}_m^w = n\sum_{k=1}^{m} w_k \tilde{r}_k,$$

where the $\{w_k\}$ are a sequence of positive lag-based weights. $\tilde{Q}_m^w$ has the same limit distribution as Q_m^w and Q_m^{w*} under H_0 but will have more power under H_1. For a finite m, Q_m^w (and therefore Q_m^{w*} and $\tilde{Q}_m^w$) are asymptotically distributed as a linear combination of d^2m iid χ_1^2 random variables; see Hosking (1980) and Gallagher and Fisher (2015) for details on the asymptotic distribution and its approximations.

Various choices of $\{w_k\}$ have been suggested. These schemes can be segmented into two groupings: divergent and convergent sequences of weights. Hong (1996a) proposes the weights be determined by the square of a kernel function and the Daniel kernel is shown to be *optimal* under a certain class of kernels. Shao (2011) demonstrates this approach provides a level of robustness in *weak* ARMA models. Weights that are convergent were suggested in Gallagher and Fisher (2015) and have similar properties. They suggest that by utilizing weights that decrease sufficiently fast it alleviates the need for a practitioner to select a maximum lag m. Our measure of correlation has the utility to be used in either of these large m situations.

Goodness-of-fit statistics based on the log of the determinant of a single Toeplitz matrix (as constructed using several lags of autocorrelations) have been proposed previously (Peña and Rodríguez, 2006; Mahdi and McLeod, 2012). The statistic of Mahdi and McLeod (2012) with maximum lag $m = 1$ is equivalent to $\tilde{Q}_m$. In general, these statistics are asymptotically equivalent to a version of Q_m^w described in Fisher and Gallagher (2012). Unlike these extant matrix-based methods, our proposed measure enables the flexibility to be used in conjunction with any weighting scheme. Although Peña and Rodríguez (2002, 2006) and Mahdi and McLeod (2012) demonstrate their matrix-based tests can improve power over competing methods, their matrix does not obey a property akin to Lemma 1 herein. Therefore, their test is more powerful than the asymptotically equivalent method of Fisher and Gallagher (2012) in some circumstances and not in others. Interestingly, the statistic from Peña and Rodríguez (2006) uses a version of $\tilde{r}_k$

constructed with the partial autocorrelation function for univariate data and as a consequence is more powerful than the weighted Monti (1994) statistic from Fisher and Gallagher (2012).

Lastly, we note that the $\tilde{Q}_m$ statistic can be motivated as a data-weighted statistic in the vein of Gallagher and Fisher (2015). In the univariate setting, our measure obeys $n\tilde{r}_k = n\,(1 + \hat{\rho}_k^2/2 + \hat{\rho}_k^4/3 + \cdots)\,\hat{\rho}_k^2$. Since each $\tilde{h}_k = \hat{\rho}_k^2$ is multiplied by the term $(1 + \hat{\rho}_k^2/2 + \hat{\rho}_k^4/3 + \cdots)$, our proposed statistic places greater emphasis on lags that observe higher residual autocorrelations. Nonetheless, portmanteau tests that employ deterministic weighting schemes are more common in the literature than data-driven weights. Therefore, the utility of our measure when used in conjunction with deterministic weights is explored in our simulations.

3.4 Other methods

Even though the weighted statistics above can assuage the impact the maximum lag m has on the performance of the portmanteau test, in practice a user must choose a maximum lag or set some acceptable criterion for its growth. Recent work in the literature has attempted to alleviate this issue.

Consider the work of Escanciano and Lobato (2009) for univariate time series and the extension to multivariate time series given by Escanciano et al. (2013). They propose a method that automatically selects the maximum lag for Q_m based on a penalty term that relates to the well-known AIC and BIC criteria. Under the null hypothesis of an adequately fitted model, the asymptotic distributed is found based off the observation that $\tilde{m} \xrightarrow{p} 1$ under H_0 (see Escanciano et al., 2013, for details). Simulations in Escanciano et al. (2013) demonstrate the automatic lag selected test tends to have slightly inflated type I errors. Our simulations (as seen in Section 4) found that for moderate m, methods based on our measure have type I errors that are comparable to those seen in analogous LB-type methods. Therefore, we anticipate that the procedure based on automatic lag selection using our measure will also have slightly inflated type I errors.

McLeod and Li (1983) propose the use of transformations, such as squaring of the residual series, to determine if a nonlinear process such as that in Section 3.2 is present within an observed time series. This concept was later used for multivariate time series in Mahdi and McLeod (2012). Specifically, they consider methods based on autocorrelation matrices $\hat{\mathbf{R}}_k(\hat{\boldsymbol{\epsilon}}_t^2)$, where $\hat{\mathbf{R}}_k(\cdot)$ is as defined in (2.1) and for the d-dimensional fitted residuals, $\boldsymbol{\epsilon}_t^2 = (\epsilon_1^2, \ldots, \epsilon_d^2)_t^\top$. Once established that a time series has a nonlinear structure, modeling may be performed using a (multivariate) GARCH or some similar model. We are unaware of any extant goodness-of-fit techniques for multivariate GARCH so we briefly highlight the univariate work of Li and Mak (1994). Under the null hypothesis of an adequately fitted GARCH model, Li and Mak (1994) show the vector of autocorrelations constructed from the autocorrelations of the standardized residuals follows a quadratic form asymptotically. A statistic constructed with our modified measure, $\tilde{r}_k$, will provide more power than that of Li and Mak (1994)

4. Simulation Studies

We study the improvement provided by our proposed methods over those in the literature via simulation. For brevity, we limit our study to the cases of iid and uncorrelated innovations. We exclude a large study on different weighting techniques, methods using automatic lag selection, and diagnostics for nonlinear models. We encourage the interested reader to consult the relevant references and reiterate that our method applies in those settings.

4.1 Goodness-of-fit in IID Data

Consider a bivariate centered VAR(2) process satisfying (1.1) with parameters

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}, \ \mathbf{\Phi}_2 = \begin{bmatrix} 0 & 0 \\ 0 & -\delta \end{bmatrix} \ \text{ and } \ \mathbf{\Sigma}_{\boldsymbol{\epsilon}} = \begin{bmatrix} 1 & 0.71 \\ 0.71 & 2 \end{bmatrix}. \quad (4.1)$$

Given that the introduced measure clearly will be more powerful under the alternative hypothesis, the primary concern about (3.2) is the finite sample performance under the null hypothesis; i.e., are the extra terms A_k from Theorem 2 collectively negligible in practice. A series of size $n = 80$ is generated for $\delta = 0$ and fit as a VAR(1), the goodness-of-fit tests are found for maximum lags $m = 4$ and 7 at significance levels $\alpha = 5\%$, 1% and 0.1%. The process is repeated for sample size $n = 160$ with maximum lags $m = 5$ and 8 where the maximum lag values were chosen based on rates in Hong (1996a), $[\log(n)]$ and $[3n^{0.2}]$. Results are shown in Table 1

Table 1: Rate of rejections, out of 10,000 replications, under the null hypothesis ($\delta = 0$) at two sample sizes n, two lags m, and three significance levels for data generated as VAR(2) in (4.1) and fit as a VAR(1).

	$n = 80$						$n = 160$					
	$m = 4$			$m = 7$			$m = 5$			$m = 8$		
	5%	1%	0.1%	5%	1%	0.1%	5%	1%	0.1%	5%	1%	0.1%
Q_m	3.0	0.4	0.0	2.5	0.3	0.0	3.9	0.6	0.1	3.2	0.6	0.1
Q_m^*	4.1	0.7	0.0	4.1	0.6	0.1	4.5	0.8	0.1	4.4	0.9	0.1
$\tilde{Q}_m$	4.4	0.8	0.1	3.7	0.6	0.1	4.5	0.8	0.1	4.0	0.8	0.1
Q_m^w	4.3	0.8	0.1	3.3	0.5	0.1	4.4	0.8	0.1	3.9	0.8	0.1
Q_m^{w*}	5.1	1.1	0.1	4.7	1.0	0.1	4.9	0.9	0.1	4.8	1.0	0.1
D_m	2.4	0.3	0.1	3.5	0.6	0.0	2.1	0.3	0.0	2.8	0.5	0.1
$\tilde{Q}_m^w$	5.5	1.3	0.1	4.5	1.0	0.1	5.0	1.0	0.1	4.6	0.9	0.1

comparing the proposed portmanteau test $\tilde{Q}_m$ (3.2) with the traditional methods Q_m and Q_m^* from (2.2). To further demonstrate the utility of our measure we implement it in a weighted statistic using the weighting scheme of Fisher and Gallagher (2012), $w_k = (m - k + 1)/m$. Table 1 also reports the empirical type I error rates of Q_m^w, Q_m^{w*} and $\tilde{Q}_m^w$ representing a Weighted BP-type statistic, a weighted LB-type (where the weights are a convolution

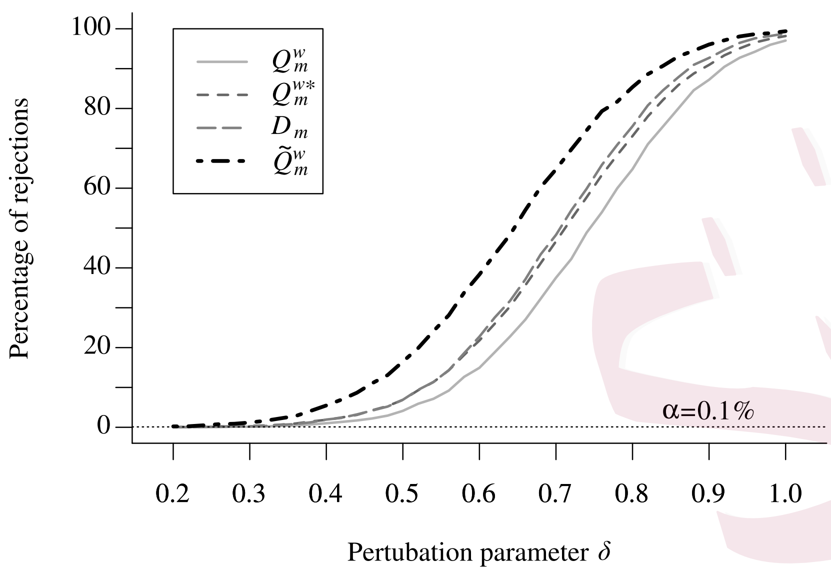

Figure 1: Empirical power, at α =0.1%, of $\tilde{Q}_m^w$, Q_m^{w*}, D_m, and Q_m^w in detecting underfit VAR(2) process at $m = 4$ with $n = 40$ as a function of δ for parameters in (4.1).

of w_k and $n/(n-k)$) and a weighted statistic using our proposed measure, respectively. For further comparison, we also include the statistic from Mahdi and McLeod (2012), D_m, which is asymptotically equivalent to Q_m^w. A Gamma approximation for the asymptotic distribution is utilized where the first two cumulants are adjusted with the fitted degrees of freedom (see Peña and Rodríguez, 2002; Hosking, 1980) for Q_m^w, Q_m^{w*} and $\tilde{Q}_m^w$ while the published χ^2 approximation is used for D_m; see Mahdi and McLeod (2012). Note acceptable-to-conservative type I error performance for all methods.

The potential increase in power is explored as a function of the perturbation parameter δ. A series of length $n = 40$ is generated from the VAR(2) process in (4.1) and an inadequate vector autoregressive of order 1 is fit to the bivariate series. The three weighted goodness-of-fit statistics and the matrix based statistic D_m are calculated with maximum lag 4 ($[\log(n)]$). The rate of rejection is calculated at significance level 0.1% based on 10,000 replications. Figure 1 provides the empirical power of each statistic as a function of the parameter δ. The figure demonstrates the proposed method can provide substantial improvement in terms of power (roughly 27% more power over Q_m^w and 17% over D_m at $\delta = 0.66$) and overall is more powerful while still providing acceptable type I error performance.

To further demonstrate the utility of our method consider a scenario of higher dimension: A $d = 4$ centered VAR(2) is generated with parameters

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 0.87 & 0.55 & 0 \\ -1.5 & -0.07 & 0.46 & 0 \\ 0 & 0 & 0 & 0.35 \end{bmatrix}, \boldsymbol{\Phi}_2 = \delta \begin{bmatrix} 0 & 0 & 0 & 0.04 \\ 0 & 0 & -0.59 & 0 \\ 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.2}$$

and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \boldsymbol{I}_4$, where the parameters are based on the *significant* values from the fitted VAR(2) of monthly real stock returns, interest rates, industrial production growth and the inflation rate in Zivot and Wang (2006).

Figure 2 provides the empirical power of Q_m, Q_m^* and $\tilde{Q}_m$ for $n = 360$,

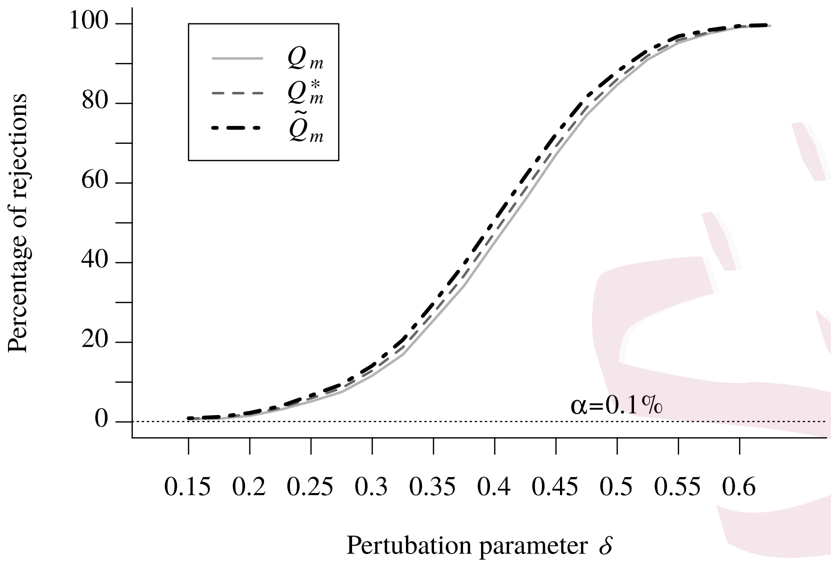

Figure 2: Empirical power, at α =0.1%, of $\tilde{Q}_m$, Q^*_m, and Q_m in detecting underfit VAR(2) process of dimension 4 at $m = 6$ with $n = 360$ as a function of δ for parameters in (4.2).

$m = 6$ and $\alpha = 0.1\%$ as a function of perturbation parameter δ. We see $\tilde{Q}_m$ offers upwards of 3.5% more power than $\tilde{Q}_m$ around $\delta = 0.4$. Further, while not reported in Figure 2, the empirical type I error rates of Q_m, Q^*_m and $\tilde{Q}_m$ were 0.03, 0.08 and 0.09, respectively. Lastly we note that higher dimensional time series require bigger n to obtain stable performance of any of the test statistics, and the improvement offered by our method is less noticeable for larger n; this is discussed more in the following subsection.

4.2 Goodness-of-fit in Uncorrelated but Dependent Data

Next we consider a simulation with data from a *weak* VAR process.

Here we report the modified versions of Q_m, Q_m^* and $\tilde{Q}_m$ using the distribution described in Section 3.2. We follow the estimation procedure in Francq and Raïssi (2007) and choose the intermediate autoregressive order, $r \in \{0, 1, 2, 3\}$ in step 6 of their algorithm, via BIC. Here, we only consider a maximum order of 3 as we are working with smaller sample sizes. In the first study, data are generated from a bivariate VAR(2) with parameters

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}, \quad \boldsymbol{\Phi}_2 = -\delta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{4.3}$$

with innovations from

$$\boldsymbol{\epsilon}_t = \begin{pmatrix} \eta_{1t}\eta_{1t-1}\eta_{1t-2} \\ \eta_{2t}\eta_{2t-1}\eta_{2t-2} \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \text{ iid } N_2\left(\mathbf{0}, \mathbf{I}_2\right). \tag{4.4}$$

Note that the residual series is uncorrelated but not serially independent.

The results in Francq and Raïssi (2007) demonstrate the portmanteau test for weak VAR processes can be conservative for large lags relative to the sample size. Here, we consider $n = 160$ and 320 with lags $m = 2$ and 3 and 3 and 4, respectively. The results are in Table 2 and, consistent with Francq and Raïssi (2007), the tests appear to have conservative type I error rates. Although not reported here, the statistics based off the asymptotic

Table 2: Rate of rejections, out of 10,000 replications, under the null hypothesis ($\delta = 0$) at two sample sizes n, two lags m, and three significance levels for data generated as a *weak* VAR(2) in (4.3) with innovations from (4.4) and fit as a VAR(1).

	$n = 160$						$n = 320$					
	$m = 2$			$m = 3$			$m = 3$			$m = 4$		
	5%	1%	0.1%	5%	1%	0.1%	5%	1%	0.1%	5%	1%	0.1%
Q_m	2.5	0.3	0.0	2.1	0.3	0.0	2.4	0.3	0.0	2.2	0.3	0.0
Q_m^*	2.7	0.3	0.0	2.3	0.3	0.0	2.4	0.4	0.0	2.3	0.3	0.0
$\tilde{Q}_m$	2.9	0.3	0.0	2.5	0.3	0.0	2.5	0.4	0.0	2.4	0.3	0.0

chi-square distribution of Hosking (1980) report highly inflated type I errors.

Lastly, in a study analogous to Figures 1 and 2, consider the possible improvement by using our recommended statistic. A series of length $n = 160$ is generated from a *weak* VAR(2) with parameters from (4.3) and the innovations follow the structure outlined in (4.4). Figure 3 provides the power of each statistic at lag $m = 2$ for $\alpha = 1\%$ as a function of δ. We see that the proposed method offers substantially more power than Q_m^* and Q_m for larger values of δ. We also see an interesting phenomenon in that as δ increases, the power of Q_m and Q_m^* appears to level off compared to the proposed method and all methods lose some power as δ approaches 1 (the point at which the process becomes non-stationary).

For further insight into the above observations, Figure 4 provides the median value (of the 10,000 replicates) of the three test statistics and the critical point at $\alpha = 1\%$ (determined from the data) at each perturbation value δ. The figure indicates that Q_m and Q_m^* tend to observe similar values for all δ; however, the $\tilde{Q}_m$ statistic diverges from the other two

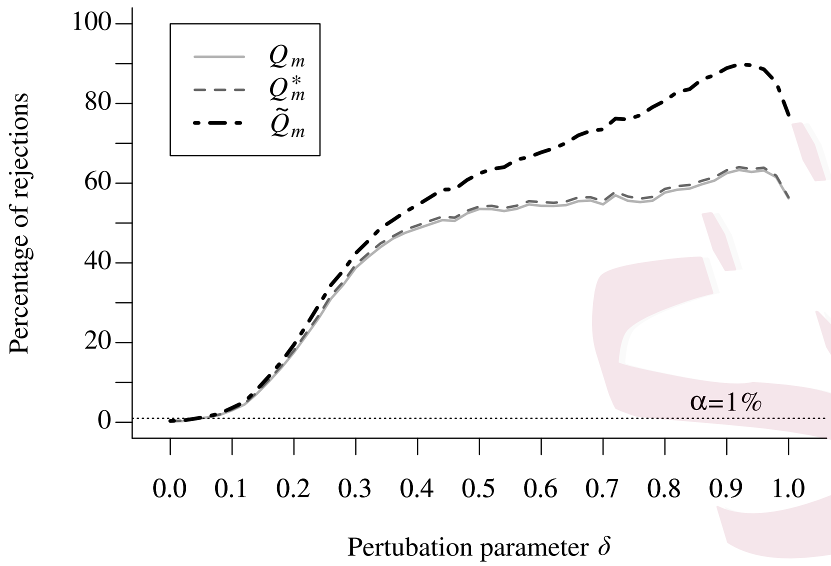

Figure 3: Empirical power, at α =1%, of $\tilde{Q}_m$, Q_m^*, and Q_m in detecting underfit *weak* vector autoregressive process at $m = 2$ with $n = 160$ as a function of δ for parameters in (4.3) with innovations from (4.4).

with increasing δ—this observation is in accordance with Theorem 2 and Corollary 1. In fact, similar patterns are observed when an analogous graph is made using time series that have iid innovations (not shown). However, the explanation for the marked improvement in power offered by our method in Figure 3 is the fact that the critical value (the same critical value is used for all tests) increases with δ. For iid innovations, the critical value is given by a χ^2 distribution and therefore is invariant of terms like δ. Therefore, when δ is large enough for $\tilde{Q}_m$ to diverge from the other statistics, all statistics have power close to 100% in the iid setting. Since critical

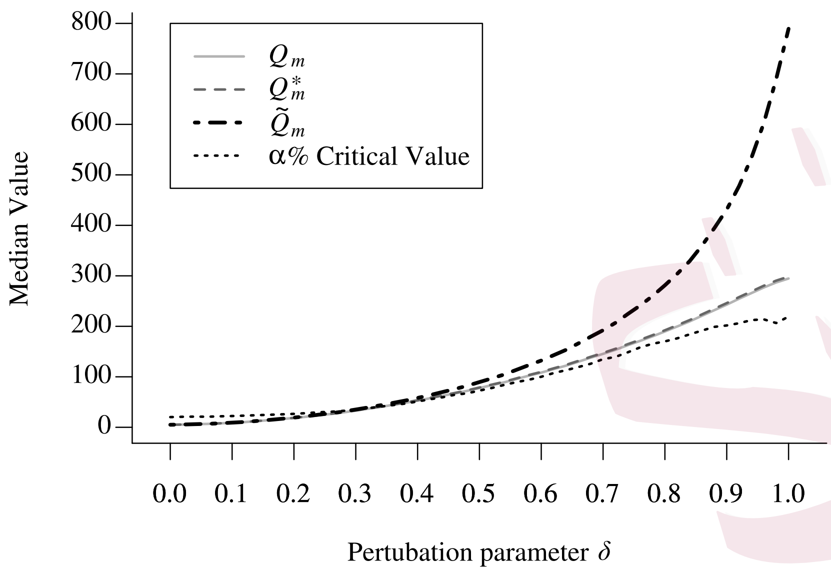

Figure 4: Median values of of $\tilde{Q}_m$, Q_m^*, Q_m, and the 1% critical point for each of 10,000 iterations in detecting underfit *weak* vector autoregressive process at $m = 2$ with $n = 160$ as a function of δ for parameters in (4.3) with innovations from (4.4).

values observed under the uncorrelated setting increase with δ, all methods have lower power, and therefore the improvement offered by our method is more visible. We theorize that the critical value increases with δ in the uncorrelated setting because $\mathbf{\Sigma}_{Q_m}$ is not consistently estimated under the alternative hypothesis. Further, although not visualized here we note the distribution of the critical point appears to be strongly skewed near the point of non-stationarity, which along with the median value in Figure 4 explains the power functions in Figure 3.

Consistent with the theoretical results presented earlier, we expect that the proposed measure will provide more power than McLeod and Li (1983) in detecting nonlinear processes and that of Li and Mak (1994) when used to diagnose the fit of a GARCH process. However, we anticipate that the improvement will be modest since both are designed for univariate data. Likewise, when critical values of the test statistics are determined via bootstrapping (see Lin and McLeod (2006)), we expect that our method will have power that is comparable to the analogous statistic. Overall, we found the proposed method to be most effective in the multivariate setting and, in line with Robbins and Fisher (2015), the "more incorrect" the null hypothesis. For larger sample sizes and significance levels, the differences between the proposed and established tests is minimal. However, for smaller α values, larger deviations from H_0 (measured by δ in our simulations) are needed, and the proposed is method most effective.

5. Discussion

In summary, our finite-sample simulations show the proposed measure performs quite well, reporting acceptable (or conservative) type I error rates consistent with the traditional method, all while improving the power. Weighting the statistic (3.2) in a way similar to Hong (1996a), Peña and Rodríguez (2006), Fisher and Gallagher (2012), and Gallagher and Fisher (2015) can provide additional power compared to the results herein and should outperform those published methods as the underlying measure of correlation in the residual time series is more powerful.

As noted at the end of Section 2, one can define $\tilde{r}_k^* = n\tilde{r}_k/(n-k)$, which is a version of our measure that incorporates the Ljung-Box correction. We define a new statistic $\tilde{Q}_m^*$, which represents $\tilde{Q}_m$ from (3.1) with $\tilde{r}_k$ replaced by $\tilde{r}_k^*$; this statistic will observe a higher rejection rate under H_1 than the standard Ljung–Box method in all settings. Under the setting used to generate the results of Table 1, where we isolate to $\alpha = 1\%$, $n = 80$, and $m = 4$, $\tilde{Q}^*_{,}$ has an estimated type I error of 1.1% (compared to respective value of 0.7% for the standard Ljung–Box technique). As this method may result in liberal type I errors, we recommend $\tilde{Q}_m$ over $\tilde{Q}_m^*$ in practice.

Although theoretical arguments illustrate the improvement in power provided by our method are asymptotic in nature, simulations indicate the improvement is more prominent when there are strong departures from the null hypothesis. Therefore, our method is preferable over existing methods in moderately sized samples (therein, the departure from the null hypothesis may be large while existing methods do not have power close to unity). Similarly, we do not anticipate that our method will perform well (in comparison to extant procedures) under local alternatives. For instance, consider $\lambda_j - 1 = \mathcal{O}_p(n^{-1/\nu})$ for some ν and some j within (2.5) (note $\lambda_j - 1 = \mathcal{O}_p(1)$

for some j under fixed alternatives). If $\nu > 2$, the Hosking quantity $\tilde{h}_k$ will have detection power asymptotically. When $\nu \geq 4$, our statistic $\tilde{r}_k$ will offer asymptotic improvement in power over $\tilde{h}_k$. However, if $2 < \nu < 4$, $\tilde{r}_k$ converges to $\tilde{h}_k$, meaning our method offers no asymptotic improvement.

The results of this article have several logical expansions for further development. One could take the results of Section 4.1 in Robbins and Fisher (2015) and construct a statistic for gauging the cross-correlation between two series using a statistic such as $\tilde{Q}_m$ herein. Following Hong (1996b); Bouhaddioui and Roy (2006); Robbins and Fisher (2015), a weighted variant can further improve power. The results of Peña and Rodríguez (2002) and Mahdi and McLeod (2012) are based on large Toeplitz matrices with the kth off-diagonal populated with an $\hat{\mathbf{R}}_k$ term – one could develop an analogous matrix-based test using the proposed measure of correlation.

Acknowledgements

The authors are grateful to Professors Francq and Raïssi for sharing their Fortran code implementing the algorithm in Francq and Raïssi (2007). The authors have adapted it for use in the R Project and plan to release these methods in an upcoming package implementing an assortment of diagnostic tests for time series.

References

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, 31, 307–327.

Bouhaddioui, C. and Roy, R. (2006), "A Generalized Portmanteau Test For Independence Of Two Infinite-Order Vector Autoregressive Series," *J. Time Ser. Anal.*, 27, 505–544.

Box, G. E. P. (1954), "Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification," *Ann. Math. Statistics*, 25, 290–302.

Box, G. E. P. and Pierce, D. A. (1970), "Distribution of residual autocorrelations in autoregressive integrated moving average time series models," *J. Amer. Statist. Assoc.*, 65, 1509–1526.

Chitturi, R. V. (1974), "Distribution of residual autocorrelations in multiple autoregressive schemes," *J. Amer. Statist. Assoc.*, 69, 928–934.

den Haan, W. J. and Levin, A. T. (1997), "A practitioner's guide to robust covariance matrix estimation," in *Robust Inference*, Elsevier, vol. 15 of *Handbook of Statistics*, pp. 299 – 342.

Dunsmuir, W. T. M. and Hannan, E. J. (1976), "Vector linear time series models," *Advances in Appl. Probability*, 8, 339–364.

Eaton, M. L. and Tyler, D. E. (1991), "On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix," *Ann. Statist.*, 19, 260–271.

Engle, R. F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

Escanciano, J. C. and Lobato, I. N. (2009), "An automatic Portmanteau test for serial correlation," *Journal of Econometrics*, 151, 140 – 149, recent Advances in Time Series Analysis: A Volume Honouring Peter M. Robinson.

Escanciano, J. C., Lobato, I. N., and Zhu, L. (2013), "Automatic Specification Testing for Vector Autoregressions and Multivariate Nonlinear Time Series Models," *Journal of Business & Economic Statistics*, 31, 426–437.

Fisher, T. J. and Gallagher, C. M. (2012), "New weighted portmanteau statistics for time series goodness of fit testing," *J. Amer. Statist. Assoc.*, 107, 777–787.

Francq, C. and Raïssi, H. (2007), "Multivariate Portmanteau Test For Autoregressive Models with Uncorrelated but Nonindependent Errors," *Journal of Time Series Analysis*, 28, 454–470.

Francq, C., Roy, R., and Zakoan, J.-M. (2005), "Diagnostic Checking in ARMA Models with Uncorrelated Errors," *Journal of the American Statistical Association*, 100, pp. 532–544.

Gallagher, C. M. and Fisher, T. J. (2015), "On Weighted Portmanteau Tests for Time Series Goodness-of-fit," *J. Time Ser. Anal.*, 36, 67–83.

Hong, Y. (1996a), "Consistent Testing for Serial Correlation of Unknown Form," *Econometrica*, 64, pp. 837–864.

Hong, Y. (1996b), "Testing for independence between two covariance stationary time series," *Biometrika*, 83, 615–625.

Hosking, J. R. M. (1980), "The multivariate portmanteau statistic," *J. Amer. Statist. Assoc.*, 75, 602–608.

Hosking, J. R. M. (1981), "Equivalent forms of the multivariate portmanteau statistic," *J. Roy. Statist. Soc. Ser. B*, 43, 261–262.

Imhof, J. P. (1961), "Computing the distribution of quadratic forms in normal variables," *Biometrika*, 48, 419–426.

Li, W. K. and Mak, T. K. (1994), "On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity," *J. Time Ser. Anal.*, 15, 627–636.

Li, W. K. and McLeod, A. I. (1981), "Distribution of the residual autocorrelations in multivariate ARMA time series models," *J. Roy. Statist. Soc. Ser. B*, 43, 231–239.

Lin, J.-W. and McLeod, A. I. (2006), "Improved Peňa-Rodriguez portmanteau test," *Comput. Statist. Data Anal.*, 51, 1731–1738.

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., Nankervis, J. C., and Savin, N. (2002), "Testing for zero autocorrelation in the presence of statistical dependence," *Econometric Theory*, 18, 730–743.

Lobato, I., Nankervis, J. C., and Savin, N. E. (2001), "Testing for Autocorrelation Using a Modified Box-Pierce Q Test," *International Economic Review*, 42, 187–205.

Lobato, I. N. (2001), "Testing That a Dependent Process Is Uncorrelated," *Journal of the American Statistical Association*, 96, 1066–1076.

Mahdi, E. and McLeod, I. A. (2012), "Improved multivariate portmanteau test," *J. Time Ser. Anal.*, 33, 211–222.

McLeod, A. I. and Li, W. K. (1983), "Diagnostic checking ARMA time series models using squared-residual autocorrelations," *J. Time Ser. Anal.*, 4, 269–273.

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

Peña, D. and Rodríguez, J. (2002), "A powerful portmanteau test of lack of fit for time series," *J. Amer. Statist. Assoc.*, 97, 601–610.

Peña, D. and Rodríguez, J. (2006), "The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series," *J. Statist. Plann. Inference*, 136, 2706–2718.

Robbins, M., Gallagher, C., Lund, R., and Aue, A. (2011), "Mean shift testing in correlated data," *Journal of Time Series Analysis*, 32, 498–511.

Robbins, M. W. and Fisher, T. J. (2015), "Cross-Correlation Matrices for Tests of Independence and Causality between Two Multivariate Time Series," *J. Bus. Econom. Statist.*, 33, 459–473.

Romano, J. P. and Thombs, L. A. (1996), "Inference For Autocorrelations Under Weak Assumptions," *Journal of the American Statistical Association*, 91, pp. 590–600.

Shao, X. (2011), "Testing For White Noise Under Unknown Dependence And Its Applications To Diagnostic Checking For Time Series Models," *Econometric Theory*, 27, 312–343.

Silvennoinen, A. and Teräsvirta, T. (2009), "Multivariate GARCH Models," in *Handbook of Financial Time Series*, eds. Mikosch, T., Kreiß, J.-P., Davis, R. A., and Andersen, T. G., Springer Berlin Heidelberg, pp. 201–229.

Taylor, S. J. (1986), *Modelling Financial Time Series*, New York: Wiley.

Wei, W. W. S. (2006), *Time series analysis*, Addison Wesley/Pearson, Boston, MA, 2nd ed., univariate and multivariate methods.

Zivot, E., and Wang, J. (2006), "Vector Autoregressive Models for Multivariate Time Series," in *Modeling Financial Time Series with S-PLUS*, Springer New York, pp. 385–429.

Department of Statistics, Miami University, 311 Upham Hall, Oxford, OH 45056 U.S.A.

E-mail: (fishert4@miamioh.edu)

RAND Corporation, Pittsburgh, PA 15213 U.S.A.

E-mail: (mrobbins@rand.org)