A Simple Asymptotically F-Distributed Portmanteau Test for Diagnostic Checking of Time Series Models with Uncorrelated Innovations*

Xuexin Wang
School of Economics and WISE
Xiamen University, China

Yixiao Sun
Department of Economics
UC San Diego, USA

Abstract

We propose a simple asymptotically F-distributed portmanteau test for diagnostically checking whether the innovations in a parametric time series model are uncorrelated while allowing them to exhibit higher-order dependence of unknown forms. A transform of sample residual autocovariances removing the influence of parameter estimation uncertainty makes the test simple. Further, by employing the orthonormal series variance estimator, a special sample autocovariances estimator that is asymptotically invariant to parameter estimation uncertainty, we show that the proposed test statistic is asymptotically F-distributed under fixed-smoothing asymptotics. The asymptotic F theory accounts for the estimation error of the variance estimator that the asymptotic chi-squared theory ignores. Moreover, an extensive Monte Carlo study demonstrates that the F test has more accurate finite sample size than existing tests with virtually no power loss. An application to S&P 500 returns illustrates the merits of the proposed methodology.

Keywords: F distribution; Fixed-smoothing Asymptotics; Model Diagnostics; Orthonormal Series Variance Estimator; Parameter Estimation Uncertainty; Uncorrelated Innovations

*We would like to thank a coeditor, an anonymous associate editor, and two anonymous reviewers for detailed and constructive comments.

1 INTRODUCTION

In time series modeling, it is standard practice to use diagnostic tests to check whether the innovations from a parametric model are uncorrelated, since a good time series model should capture the dependence structure adequately, with no autocorrelation left in the innovations. One such prominent diagnostic test is the Q test of Box and Pierce (1970) and Ljung and Box (1978) which aims at checking whether the innovations in an ARMA model are uncorrelated. It has become a routine to report the p-value of the Q test, based on a chi-squared distribution, in empirical applications of ARMA models; see, for example, Baillie and Bollerslev (1989) and Rosenberg and Engle (2002). However, this practice could lead to misleading conclusions, because the chi-squared approximation of the Q statistic relies on a specific linear form of ARMA models (cf. Mcleod, 1978) and the restrictive assumption that the innovations are independently and identically distributed (iid). Francq, Roy, and Zakoïan (2005, FRZ hereafter) point out that if the innovations are merely uncorrelated, the high-order dependence of innovations and the parameter estimation uncertainty need to be considered when deriving the asymptotic distribution of the residual autocovariance or autocorrelation[1].

In this paper, we propose a simple asymptotically F-distributed portmanteau test for diagnostic checking in parametric time series models with uncorrelated innovations. We introduce a family of transforms and apply it to sample residual autocovariances to remove the influence of parameter estimation uncertainty. Subject to this transformation, the sample residual and innovation autocovariances are asymptotically equivalent. To estimate the variance of the sample autocovariances, we employ an orthonormal series (OS) variance estimator, which is not affected asymptotically by parameter estimation uncertainty. Using this variance estimator to studentize the transformed sample autocovariances, we show that the resulting test statistic converges weakly to a convenient F distribution under fixed-smoothing asymptotics. The asymptotic F theory accounts for the estimation error in the underlying variance estimator that the conventional asymptotic chi-squared theory ignores. Monte Carlo simulations reveal that the proposed F test has very accurate size and competitive power in a number of finite sample settings. An application to S&P 500 returns illustrates

[1]The residuals are estimated versions of innovations.

the merits of our methodology.

Our approach is an extension of Wang and Sun (2020) who propose an asymptotically F-distributed portmanteau test for an observed time series. This extension is nontrivial, as parameter estimation uncertainty has a nonnegligible effect on the behavior of the sample residual autocovariances. The commonly used estimators of the variance of the autocovariance estimators are also affected by the parameter estimation uncertainty, even in large samples. We need to find a new way to construct the test statistic and employ a different proof to establish the asymptotic F theory.

Our paper contributes to a large body of the literature on testing the lack of innovation autocorrelation in time series models. For an ARMA model with uncorrelated innovations, FRZ (2005) derive the asymptotic variance of the sample residual autocorrelations when the model is estimated by the ordinary least squares. They show that the asymptotic distribution of the Q test statistic is a weighted sum of independent chi-squared distributions, where the weights are determined by the eigenvalues of the variance matrix of the sample residual autocorrelations. Hence, its asymptotic distribution is not pivotal. To avoid estimating the complicated variance matrix of the sample residual autocorrelations, Zhu (2016) proposes wild bootstrapping procedures for the Q test. On the other hand, Delgado and Velasco (2011, hereafter DV) employ a distribution-free transform of the sample residual autocorrelations and construct a Box-Pierce type test based on the transformed sample residual autocorrelations. They show that the Box-Pierce type statistic follows a chi-squared distribution asymptotically.

Both FRZ (2005) and DV (2011) rely on consistent variance matrix estimators. The consistency requirement amounts to ignoring the error in the variance estimator completely in large sample approximations. This may adversely affect the quality of the large sample approximations, especially when the sample size is small or moderate. The bootstrap approximation proposed by Zhu (2016) may be more accurate, but it often involves a user-chosen tuning parameter, which is difficult to pin down in finite samples[2], and the computational cost is high. To address these problems, Mainassara and Saussereau (2018, hereafter MS) adopt the idea of self-normalization (see, e.g., Shao, 2010, 2015) and propose a new test in a VARMA setting. Their test is an extension of Lobato (2001)

[2]The cases that involve no tuning parameter are quite restrictive, cf. Assumption 4 on pp. 469 in Zhu (2016).

and belongs to the class of the fixed-b tests developed by Kiefer, Vogelsang and Bunzel (2000). The distribution of the test statistic is pivotal but nonstandard, requiring simulation to obtain the critical values. In essence, their test statistic is a quadratic form of the residual autocorrelations studentized by a nonparametric variance estimator with the bandwidth equal to the sample size. However, as pointed out by FRZ (2005), this nonparametric variance estimator is likely noninvertible in (V)ARMA models in finite samples. For example, this is the case when an AR(1) model is fitted to a strong white noise process (see Corollary 1 in Sec. 4 of FRZ, 2005). Further, the Monte Carlo simulations in Section 4 show that this test is quite conservative and not so powerful in finite samples.

Our approach differs from MS (2018) in two aspects. First, we employ a transformation to remove the parameter estimation uncertainty. As a result, the asymptotic distribution of our test statistic is not affected by the asymptotic distribution of the parameter estimator. Under stronger conditions, any consistent parameter estimator leads to the same asymptotic distribution of our test statistic. In contrast, the asymptotic theory in MS (2018) relies crucially on the $\sqrt{T}$-consistency and asymptotic normality of the parameter estimator. The transformation we employ also removes the potential nonsingularity problem pointed out by FRZ (2005). Second, we use an orthonormal series variance estimator, while the approach of MS (2018) amounts to using a kernel variance estimator based on the Bartlett kernel. The convenient F asymptotic theory we develop here is possible only when an orthonormal series variance estimator is used.

The rest of this paper is organized as follows. Section 2 lays out the preliminaries. Section 3 introduces the transformation and the new portmanteau test statistic and establishes its asymptotic properties. Section 4 reports the Monte Carlo evidence. An application to S&P 500 returns is provided in Section 5. Section 6 concludes the paper. All proofs are given in the appendix.

We use the following notation throughout the paper: $\mathbb{I}_m$ is the $m \times m$ identity matrix. For an $m \times d$ matrix A, $\mathbb{P}_A = A\left(A'A\right)^{-1}A'$ and $\mathbb{M}_A = \mathbb{I}_m - \mathbb{P}_A$. "$\to^p$" indicates convergence in probability, and "$\Rightarrow$" indicates convergence in distribution. We use "O_{sp}" to represent a $s \times p$ matrix of zeros. We often omit the dimension and use "O" to represent a matrix of zeros whose dimension(s) may be different across different occurrences. For a symmetric matrix $\Omega > 0$, $\Omega^{1/2}$ is a symmetric square root of Ω such that $\Omega^{1/2}\left(\Omega^{1/2}\right)' = \Omega$.

2 PRELIMINARIES

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a time series process in $\mathbb{R}^k$. For any $\theta \in \Theta \subset \mathbb{R}^p$, consider the filtered process

$$\varepsilon_{\theta,t} = \mathbb{U}\left(\ldots, X_{t-1}, X_t, \ldots, \varepsilon_{\theta,t-2}, \varepsilon_{\theta,t-1}; \theta\right) \in \mathbb{R} \text{ for } t \in \mathbb{Z}, \tag{1}$$

where $\mathbb{U}(\cdot;\theta)$ is a measurable function. Different time series models correspond to different forms of $\mathbb{U}$. As part of the definition, a time series model typically assumes that for some $\theta_0 \in \Theta$, $\varepsilon_{\theta_0,t}$ is an iid sequence, a martingale difference (MD) sequence, or a white noise process. In all three cases, we refer to $\{\varepsilon_{\theta_0,t}\}$ as the innovations.

A prominent example of our general setup is the ARMA model:

Example 1 *Consider an ARMA*$(\tilde{p},\tilde{q})$ *process:*

$$X_t = \sum_{i=1}^{\tilde{p}} a_{i0} X_{t-i} + \varepsilon_t + \sum_{j=1}^{\tilde{q}} b_{j0}\varepsilon_{t-j} \text{ for } \varepsilon_t \sim (0, \sigma_0^2),$$

where $\phi(z) = 1 - a_{10}z - \cdots - a_{\tilde{p}0}z^{\tilde{p}}$ *and* $\psi(z) = 1 + b_{10}z + \cdots + b_{\tilde{q}0}z^{\tilde{q}}$ *have no common root and all roots lying outside the unit circle. Let* $\theta = \left(a_1, \ldots, a_{\tilde{p}}, b_1, \ldots, b_{\tilde{q}}, \sigma^2\right)$ *and define*

$$\varepsilon_{\theta,t} = X_t - \sum_{i=1}^{\tilde{p}} a_i(\theta) X_{t-i} - \sum_{j=1}^{\tilde{q}} b_j(\theta)\varepsilon_{\theta,t-j},$$

where $a_i(\theta) = a_i$ *and* $b_j(\theta) = b_j$*. Then* $\varepsilon_t = \varepsilon_{\theta_0,t}$ *for* $\theta_0 = \left(a_{10}, \ldots, a_{\tilde{p}0}, b_{10}, \ldots, b_{\tilde{q}0}, \sigma_0^2\right)$*. We say that* $\{X_t\}_{t\in\mathbb{Z}}$ *follows a weak ARMA process when* $\{\varepsilon_{\theta_0,t}\}_{t\in\mathbb{Z}}$ *is a white noise process, a semistrong ARMA process when* $\{\varepsilon_{\theta_0,t}\}_{t\in\mathbb{Z}}$ *is an MD sequence, and a strong ARMA process when* $\{\varepsilon_{\theta_0,t}\}_{t\in\mathbb{Z}}$ *is an iid sequence.*

We note that there are a few commonly used processes that are white noises but not martingale differences. Examples include all-pass ARMA processes[3]*, certain forms of bilinear processes, and some nonlinear moving average (NLMA) processes.*

[3] All-pass ARMA processes are ARMA processes where the roots of the autoregressive polynomial are reciprocals of roots of the moving average polynomial and vice versa.

The autocovariance and autocorrelation functions of $\{\varepsilon_{\theta,t}\}_{t\in\mathbb{Z}}$ are

$$\begin{aligned}\gamma_\theta(j) &= Cov(\varepsilon_{\theta,t}, \varepsilon_{\theta,t-j}), j\in\mathbb{Z},\\ \rho_\theta(j) &= \gamma_\theta(j)/\gamma_\theta(0), j\in\mathbb{Z}.\end{aligned}$$

As in Box and Pierce (1970) and Ljung and Box (1978), we are interested in testing whether the first s autocovariances are zero. Therefore, the null hypothesis is

$$H_0^{(s)}: \gamma^{(s)}(\theta_0) = 0 \text{ for some } \theta_0\in\Theta,$$

where $\gamma^{(s)}(\theta) = (\gamma_\theta(1), \ldots, \gamma_\theta(s))'$.

Given the value of θ_0, we assume that we can recover $\{\varepsilon_{\theta_0,t},\ t=1,\ldots,T\}$ from $\{X_t, t=1,\ldots,T\}$[4]. Based on $\{\varepsilon_{\theta_0,t},\ t=1,\ldots,T\}$, we can estimate the innovation autocovariance and autocorrelation functions by $\hat{\gamma}_{\theta_0}(j)$ and $\hat{\rho}_{\theta_0}(j)$, $j=1,...,s$, such that

$$\hat{\gamma}_\theta(j) = \frac{1}{T}\sum_{t=j+1}^{T}(\varepsilon_{\theta,t}-\bar{\varepsilon}_\theta)(\varepsilon_{\theta,t-j}-\bar{\varepsilon}_\theta), \hat{\rho}_\theta(j) = \frac{\hat{\gamma}_\theta(j)}{\hat{\gamma}_\theta(0)},$$

where $\bar{\varepsilon}_\theta = T^{-1}\sum_{t=1}^{T}\varepsilon_{\theta,t}$. The above estimators are not feasible because we do not know θ_0. Given an estimator $\hat{\theta}_T$ of θ_0, we can construct the residuals $\{\varepsilon_{\hat{\theta}_T,t}, t=1,\ldots,T\}$ and estimate the innovation autocovariance and autocorrelation functions by the sample residual autocovariance and autocorrelation functions, namely, $\hat{\gamma}_{\hat{\theta}_T}(j)$ and $\hat{\rho}_{\hat{\theta}_T}(j)$.

To test the null $H_0^{(s)}$ in an ARMA model, Box and Pierce (1970) and Ljung and Box (1978)

[4]In order to recover $\{\varepsilon_{\theta_0,t}\}$, we may need the initial values of X_t and $\varepsilon_{\theta_0,t}$ for $t\in[-k,0]$ for some $k\in\mathbb{Z}$. For example, in the ARMA setting, we have

$$\varepsilon_{\theta_0,t} = X_t - \sum_{i=1}^{\tilde{p}} a_i(\theta_0)X_{t-i} - \sum_{j=1}^{\tilde{q}} b_j(\theta_0)\varepsilon_{\theta_0,t-j}.$$

With the initial values $X_0,...X_{1-\tilde{p}}$ and $\varepsilon_{\theta_0,0},...,\varepsilon_{\theta_0,1-\tilde{q}}$, we can recover $\{\varepsilon_{\theta_0,t}\}$ recursively. In practice, we do not observe these values, and we can set them to zero. This will introduce some initialization error, but the error can be ignored in large samples when the ARMA process is causal and invertible. For simplicity and clarity, we abstract away the initialization error in this section.

propose the so-called Q test. The Q statistic of Ljung and Box (1978) takes the form

$$Q_T^{(s)}(\hat{\theta}_T) = T\,(T+2)\sum_{j=1}^{s}(T-j)^{-1}\,\hat{\rho}_{\hat{\theta}_T}^2\,(j)\,.$$

Let

$$\hat{\gamma}_T^{(s)}\,(\theta) = (\hat{\gamma}_\theta\,(1)\,,\ldots,\hat{\gamma}_\theta\,(s))'\,,$$

and

$$\hat{\rho}_T^{(s)}\,(\theta) = (\hat{\rho}_\theta\,(1)\,,\ldots,\hat{\rho}_\theta\,(s))'\,.$$

Under some regularity conditions and the assumption that $\varepsilon_{\theta_0,t}$ is iid, we can show that

$$\sqrt{T}\hat{\gamma}_T^{(s)}\,(\theta_0) \Rightarrow N\left(0,\gamma_{\theta_0}^2\,(0)\,\mathbb{I}_s\right) \text{ and } \sqrt{T}\hat{\rho}_T^{(s)}\,(\theta_0) \Rightarrow N\,(0,\mathbb{I}_s)\,.$$

Define $Q_T^{(s)}\,(\theta_0)$ in the same way as $Q_T^{(s)}(\hat{\theta}_T)$ is defined. It then follows that $Q_T^{(s)}\,(\theta_0)$ converges in distribution to χ_s^2, the chi-squared distribution with s degrees of freedom.

The above result relies on the assumption that $\varepsilon_{\theta_0,t}$ is iid. In the absence of any further restriction on the dependence structure of the innovations apart from $H_0^{(s)}$, we can only obtain that

$$\sqrt{T}\hat{\gamma}_T^{(s)}\,(\theta_0) \Rightarrow N\,(0,\Omega_0) \text{ and } \sqrt{T}\hat{\rho}_T^{(s)}\,(\theta_0) \Rightarrow N\left(0,\frac{\Omega_0}{\gamma_{\theta_0}^2\,(0)}\right),$$

where Ω_0 is a $s\times s$ matrix given by $\Omega_0 := \Omega\,(\theta_0)$,

$$\Omega\,(\theta) = \left[\omega_\theta^{(i,j)}\right]_{i,j=1}^{s} \text{ for } \omega_\theta^{(i,j)} = \sum_{\ell=-\infty}^{\infty}\mathbb{E}\left[\varepsilon_{\theta,t}\varepsilon_{\theta,t+i}\varepsilon_{\theta,t+\ell}\varepsilon_{\theta,t+\ell+j}\right],\, i,j=1,\ldots,s.$$

In general, because $\Omega\,(\theta_0)$ is not a diagonal matrix, $Q_T^{(s)}\,(\theta_0)$ is not asymptotically chi-squared.

The estimation error in $\hat{\theta}_T$ complicates the asymptotic distribution of $Q_T^{(s)}(\hat{\theta}_T)$ even more. By the first-order Taylor expansion, for a $\sqrt{T}$-consistent estimator $\hat{\theta}_T$ of θ_0, we can show that

$$\sqrt{T}\hat{\gamma}_T^{(s)}(\hat{\theta}_T) = \sqrt{T}\hat{\gamma}_T^{(s)}\,(\theta_0) + \Gamma_0\sqrt{T}\left(\hat{\theta}_T-\theta_0\right) + o_p\,(1)\,, \tag{2}$$

where

$$\Gamma_0 := \Gamma(\theta_0) \text{ for } \Gamma(\theta) = \operatorname{plim} \partial \hat{\gamma}_T^{(s)}(\theta) / \partial \theta'.$$

Thus, the asymptotic variance matrix of $\sqrt{T}\hat{\gamma}_T^{(s)}(\hat{\theta}_T)$, and hence that of $\sqrt{T}\hat{\rho}_T^{(s)}(\hat{\theta}_T)$, depend on the model characteristics, the form of the estimator $\hat{\theta}_T$, and perhaps the unknown parameter value θ_0. By accounting for the estimation error in an ARMA setting, FRZ (2005) show that $Q_T^{(s)}(\hat{\theta}_T)$ converges in distribution to a weighted sum of independent chi-squared distributions with the weights depending on the eigenvalues of the asymptotic variance matrix of $\sqrt{T}\hat{\rho}_T^{(s)}(\hat{\theta}_T)$. Hence, its asymptotic distribution is nonpivotal.

FRZ (2005) do not consider the quadratic form of $\sqrt{T}\hat{\rho}_T^{(s)}(\hat{\theta}_T)$ weighted by the inverse of a consistent covariance matrix estimator of $\sqrt{T}\hat{\rho}_T^{(s)}(\hat{\theta}_T)$, because this covariance matrix estimator tends to be singular in ARMA models. Later, Duchesne and Francq (2008) suggest constructing portmanteau tests using a generalized inverse of a consistent nonparametric estimator of the asymptotic variance matrix. Subsequently, DV (2011) employ a transform to obtain an asymptotically pivotal test statistic. Their transform allows for high-order dependence and accounts for the parameter estimation error. All these papers require a consistent nonparametric estimator of an asymptotic variance matrix. By invoking the consistency argument, these papers effectively approximate the distribution of a nonparametric variance estimator by a degenerate distribution concentrated at the true variance matrix. Therefore, the estimation error in the variance estimator has been completely ignored. Even with a delicate choice of underlying smoothing parameter, the estimation error in the nonparametric variance estimator can be substantial in finite samples.

Recently, the literature has introduced alternative asymptotics to address the aforementioned problem. Unlike the conventional increasing smoothing asymptotics, the alternative asymptotics hold the amount of nonparametric smoothing fixed, and are hence called the fixed-smoothing asymptotics. There is ample numerical evidence, along with theoretical results, on the higher accuracy of the fixed-smoothing asymptotic approximations relative to the conventional asymptotic approximations, see, e.g., Sun, Phillips, and Jin (2008), and Zhang and Shao (2013) for location models, and Sun (2014a, 2014b) for the generalized method of moments framework. Lobato (2001) is among the first to consider this alternative asymptotics in testing serial uncorrelatedness. Later, MS (2018) extend this idea to test the lack of autocorrelation in VARMA models. Although the asymptotic distributions of the Lobato (2001) and MS (2018) test statistics are pivotal under fixed-smoothing asymptotics, they are not standard, and critical values have to be tabulated by Monte Carlo simulations.

3 MAIN RESULTS

3.1 The Test Statistic

To motivate our test, we rewrite (2) as

$$\sqrt{T}\hat{\gamma}_T^{(s)}(\hat{\theta}_T) \approx \sqrt{T}\gamma^{(s)}(\theta_0) + \Gamma_0\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + \sqrt{T}\left[\hat{\gamma}_T^{(s)}(\theta_0) - \gamma^{(s)}(\theta_0)\right]. \tag{3}$$

For any $s \times p$ matrix Γ with column-rank p, we let $A(\Gamma)$ be an $s \times s$ matrix with rank $(s - p)$ such that $A(\Gamma)'\Gamma = O_{sp} \in \mathbb{R}^{s \times p}$. Assume that the column rank of Γ_0 is p and let $A_0 := A(\Gamma_0)$. Premultiplying (3) by A_0' yields

$$A_0'\sqrt{T}\hat{\gamma}_T^{(s)}(\hat{\theta}_T) \approx \sqrt{T}A_0'\gamma^{(s)}(\theta_0) + \sqrt{T}A_0'\left[\hat{\gamma}_T^{(s)}(\theta_0) - \gamma^{(s)}(\theta_0)\right].$$

The pre-multiplication has removed the effect of the estimation error of $\hat{\theta}_T$. The first term in the above approximation contains the main signal about the hypotheses of interest, and the second term contains the noise. The asymptotic variance of the noise term is $A_0'\Omega_0 A_0$.

To test the null of $\gamma^{(s)}(\theta_0) = 0$, we consider a quadratic form in $A_0'\sqrt{T}\hat{\gamma}_T^{(s)}(\hat{\theta}_T)$ using the inverse of $A_0'\Omega_0 A_0$ as the weighting matrix. This leads to

$$\begin{aligned} Q_0 &= T\left[A_0'\hat{\gamma}_t^{(s)}(\hat{\theta}_T)\right]'\left[A_0'\Omega_0 A_0\right]^{+}\left[A_0'\hat{\gamma}_t^{(s)}(\hat{\theta}_T)\right] \\ &= T\hat{\gamma}_t^{(s)}(\hat{\theta}_T)'\mathbb{G}(A_0, \Omega_0)\hat{\gamma}_t^{(s)}(\hat{\theta}_T), \end{aligned}$$

where

$$\mathbb{G}(A, \Omega) = A\left[A'\Omega A\right]^{+} A'$$

is the weighting matrix and $[A'\Omega A]^{+}$ is the Moore–Penrose inverse of $A'\Omega A$.

Since θ_0 is not known, $\Omega_0 = \Omega(\theta_0)$ and $A_0 := A(\Gamma(\theta_0))$ are not feasible, and therefore, $\mathbb{G}(A_0, \Omega_0)$ and hence Q_0 are not feasible. To construct a test statistic, we need to estimate both Ω_0 and A_0. In this paper, we adopt the OS approach to estimating Ω_0. Let $\{\Phi_\ell(\cdot), \ell = 1, \ldots, K\}$ be a sequence of basis functions in $L^2[0,1]$. Denote

$$f_t(\hat{\theta}_T) = \left(f_{1t}(\hat{\theta}_T), \ldots, f_{st}(\hat{\theta}_T)\right)',$$

with the j-th element given by $f_{jt}(\hat{\theta}_T) = (\varepsilon_{\hat{\theta}_T,t} - \bar{\varepsilon}_{\hat{\theta}_T})(\varepsilon_{\hat{\theta}_T,t-j} - \bar{\varepsilon}_{\hat{\theta}_T})$. Define

$$\Lambda_\ell(\hat{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_\ell\left(\frac{t}{T}\right) f_t(\hat{\theta}_T).$$

The OS variance estimator is given by

$$\hat{\Omega}(\hat{\theta}_T) = \frac{1}{K} \sum_{\ell=1}^{K} \Lambda_\ell(\hat{\theta}_T)\Lambda_\ell(\hat{\theta}_T)'.$$

To ensure that $\hat{\Omega}(\hat{\theta}_T)$ is positive semidefinite, we assume that $K \geq s$.

In the econometrics literature, the OS variance estimator has recently been used by, for example, Phillips (2005), Müller (2007), Sun (2011, 2013, 2014a,b,c), Liu and Sun (2019), and Lazarus, Lewis, Stock, and Watson (2016, 2018). This estimator can be regarded as originating from the multiple-window estimator developed by Thomson (1982) with $\{\Phi_\ell(\cdot)\}$ as the window functions. It also belongs to the class of filter-bank estimators, and $\hat{\Omega}(\hat{\theta}_T)$ is a simple average of the individual filter-bank estimators $\Lambda_\ell(\hat{\theta}_T)\Lambda_\ell(\hat{\theta}_T)'$; see Stoica and Moses (2005, Ch. 5). It can also be regarded as the sample variance of the projection coefficients; see Sun (2011) for more detailed discussions. By construction, the OS variance estimator is automatically positive semidefinite, a desirable property for practical use.

The simplest and most familiar example of the OS variance estimator is the average periodogram estimator, which takes a simple average of the first few periodograms. More specifically, let

$$\Phi_{\ell}(r) = \begin{cases} \sqrt{2}\cos(\pi \ell r), & \text{if } \ell \text{ is even,} \\ \sqrt{2}\sin(\pi(\ell+1)r), & \text{if } \ell \text{ is odd.} \end{cases}$$

Assuming that K is even, we can write the resulting OS variance estimator as

$$\hat{\Omega}(\hat{\theta}_T) = \frac{2}{K}\sum_{\ell=1}^{K/2} \mathrm{Re}\left(\Upsilon_{\ell}(\hat{\theta}_T)\Upsilon_{\ell}^{*}(\hat{\theta}_T)\right),$$

where Υ_{ℓ}^{*} denotes the transpose and complex conjugate of Υ_{ℓ}, and

$$\Upsilon_{\ell}(\hat{\theta}_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \exp\left(-\frac{2\pi \ell t}{T}\cdot\sqrt{-1}\right) f_t(\hat{\theta}_T).$$

Note that $\Upsilon_{\ell}(\hat{\theta}_T)$ is the finite Fourier transform of $f_t(\hat{\theta}_T)$. The computation of $\hat{\Omega}(\hat{\theta}_T)$ is very fast and convenient in popular programming environments such as Matlab and R. We will use the above OS variance estimator in the simulation study.

In view of $A_0 = A(\Gamma_0)$, we can estimate A_0 by plugging an estimator of Γ_0 into $A(\cdot)$. We estimate Γ_0 by

$$\hat{\Gamma}(\hat{\theta}_T) = \frac{\partial \hat{\gamma}_T^{(s)}\left(\hat{\theta}_T\right)}{\partial \theta'} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial f_t(\hat{\theta}_T)}{\partial \theta'}$$

and A_0 by $\hat{A}(\hat{\theta}_T) := A(\hat{\Gamma}(\hat{\theta}_T))$.

With the estimators $\hat{\Omega}(\hat{\theta}_T)$ and $\hat{A}(\hat{\theta}_T)$, we can estimate $\mathbb{G}(A_0, \Omega_0)$ by

$$\mathbb{G}(\hat{A}(\hat{\theta}_T), \hat{\Omega}(\hat{\theta}_T)) = \hat{A}(\hat{\theta}_T)\left[\hat{A}(\hat{\theta}_T)'\hat{\Omega}(\hat{\theta}_T)\hat{A}(\hat{\theta}_T)\right]^{+}\hat{A}(\hat{\theta}_T)'.$$

We can then construct our test statistic as follows:

$$\begin{aligned} \tilde{Q}_T^{(s)}(\hat{\theta}_T) &= \frac{K-q+1}{Kq}\cdot T\hat{\gamma}_T^{(s)}(\hat{\theta}_T)'\mathbb{G}(\hat{A}(\hat{\theta}_T), \hat{\Omega}(\hat{\theta}_T))\hat{\gamma}_T^{(s)}(\hat{\theta}_T) \\ &= \frac{K-q+1}{Kq}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t(\hat{\theta}_T)\right]'\mathbb{G}(\hat{A}(\hat{\theta}_T), \hat{\Omega}(\hat{\theta}_T))\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t(\hat{\theta}_T)\right], \qquad (4) \end{aligned}$$

where

$$q = s - p.$$

Note that $(K - q + 1)/K$ is a finite sample correction term; it approaches 1 when K is large.

It remains to pin down $A(\cdot)$. Let $\Delta \in \mathbb{R}^{s \times s}$ be a nonsingular symmetric matrix. With some abuse of notation, for any $s \times p$ matrix Γ with column-rank p, we consider

$$A(\Gamma) := A(\Gamma; \Delta) = \Delta^{-1/2} - \Delta^{-1}\Gamma \left(\Gamma' \Delta^{-1} \Gamma\right)^{-1} \Gamma' \Delta^{-1/2} = \Delta^{-1/2} \mathbb{M}_{\Delta^{-1/2}\Gamma}.$$

Such a choice satisfies the rank condition that $A(\Gamma; \Delta)$ is of column rank q and the orthogonal condition that $A(\Gamma; \Delta)' \Gamma = O_{sp}$.

Let $U \Xi V'$ be a singular value decomposition (SVD) of Γ where

$$\Xi = \begin{pmatrix} D \\ O \end{pmatrix} \in \mathbb{R}^{s \times p},$$

$D \in \mathbb{R}^{p \times p}$ is a diagonal matrix, and $U \in \mathbb{R}^{s \times s}$ and $V \in \mathbb{R}^{p \times p}$ are orthogonal matrices. We write $U = (U_{sp}, U_{sq})$ where $U_{sp} \in \mathbb{R}^{s \times p}$ consists of the left singular vectors of Γ corresponding to the p nonzero singular values and $U_{sq} \in \mathbb{R}^{s \times q}$ consists of the left singular vectors of Γ corresponding to the zero singular value.

The following lemma gives a representation $\mathbb{G}(A(\Gamma; \Delta), \Omega)$.

Lemma 1 *Assume that $\Gamma \in \mathbb{R}^{s \times p}$ has column rank p, and $\Delta \in \mathbb{R}^{s \times s}$ and $\Omega \in \mathbb{R}^{s \times s}$ are nonsingular and symmetric. Then*

$$\mathbb{G}(A(\Gamma; \Delta), \Omega) = U_{sq} \cdot \left[U_{sq}' \cdot \Omega \cdot U_{sq}\right]^{-1} \cdot U_{sq}'. \tag{5}$$

Lemma 1 assumes that Γ, Δ and Ω are deterministic matrices. If Γ, Δ and Ω are random and the rank conditions hold almost surely (or with probability approaching one as $T \to \infty$), it is easy to see that the result of the lemma holds almost surely (or with probability approaching one as $T \to \infty$). We will use stochastic extensions of Lemma 1 without further discussion.

Lemma 1 shows that $\mathbb{G}(A(\Gamma; \Delta), \Omega)$ depends only on U_{sq} and Ω. In particular, it does not depend on Δ. Our test statistic is, therefore, invariant to the choice of Δ. If we set Δ equal to the identity matrix $\mathbb{I}_s$, we have

$$A\left(\Gamma;\mathbb{I}_s\right)=\mathbb{M}_{\Gamma},\ \hat{A}(\hat{\theta}_T)=\mathbb{M}_{\hat{\Gamma}(\hat{\theta}_T)},$$

and

$$\mathbb{G}(\hat{A}(\hat{\theta}_T),\hat{\Omega}(\hat{\theta}_T))=\mathbb{M}_{\hat{\Gamma}(\hat{\theta}_T)}\left(\mathbb{M}_{\hat{\Gamma}(\hat{\theta}_T)}\hat{\Omega}(\hat{\theta}_T)\mathbb{M}_{\hat{\Gamma}(\hat{\theta}_T)}\right)^{+}\mathbb{M}_{\hat{\Gamma}(\hat{\theta}_T)}. \tag{6}$$

If we set Δ equal to Ω_0, we have $A\left(\Gamma;\Omega_0\right)=\Omega_0^{-1/2}\cdot\mathbb{M}_{\Omega_0^{-1/2}\Gamma}$,

$$\hat{A}(\hat{\theta}_T)=\hat{\Omega}(\hat{\theta}_T)^{-1/2}\cdot\mathbb{M}_{\hat{\Omega}(\hat{\theta}_T)^{-1/2}\hat{\Gamma}(\hat{\theta}_T)},$$

and

$$\mathbb{G}(\hat{A}(\hat{\theta}_T),\hat{\Omega}(\hat{\theta}_T))=\hat{\Omega}(\hat{\theta}_T)^{-1/2}\mathbb{M}_{\hat{\Omega}(\hat{\theta}_T)^{-1/2}\hat{\Gamma}(\hat{\theta}_T)}\hat{\Omega}(\hat{\theta}_T)^{-1/2}. \tag{7}$$

By Lemma 1, the representations in (6) and (7) are numerically identical and hence give rise to an identical test statistic. However, the representation in (7) is more convenient for practical implementations, as it does not involve a generalized inverse. Our test statistic can then be computed as in (4) with $\mathbb{G}(\hat{A}(\hat{\theta}_T),\hat{\Omega}(\hat{\theta}_T))$ given in (7).

For our theoretical analysis, we employ $A\left(\Gamma\right):=\mathbb{M}_{\Gamma}$ without loss of generality. We note that the implied transformation is the same as that in Katayama (2008), which considers only the case with iid innovations and hence does not allow for high-order dependence in the innovations. We could also employ $A\left(\Gamma\right):=A\left(\Gamma;\Omega_0\right)=\Omega_0^{-1/2}\cdot\mathbb{M}_{\Omega_0^{-1/2}\Gamma}$ in our theoretical analysis. In fact, this was done in an earlier version of this paper. Consistent with Lemma 1, the theoretical results are identical.

We note that the right-hand side of (5) takes the same form as $\mathbb{G}(A,\Omega)$. The only difference is that $A\in\mathbb{R}^{s\times s}$, but $U_{sq}\in\mathbb{R}^{s\times q}$. To reflect the dimensionality difference, we may write the right-hand side of (5) as $\tilde{\mathbb{G}}(U_{sq},\Omega)$. Note that the left null space of Γ is of dimension q when the column space of Γ is of dimension p. It is clear that the columns of A are linearly dependent while those of U_{sq} are not. To remove the linear dependence, we could replace it by the $s\times q$ matrix $\tilde{A}\left(\Gamma\right)$ that consists of the left singular vectors of Γ corresponding to the zero singular value. The advantage of this approach is that $\tilde{A}\left(\Gamma\right)'\Omega\tilde{A}\left(\Gamma\right)$ is not singular if Ω is not, and we do not need to take a generalized inverse. The disadvantage is that the test statistic involves a single value decomposition, which may not be empirically appealing. We opt for a square A with a reduced rank in this paper, but exactly the same results can be obtained using a non-square A with full column rank — the only difference lies in technical details.

3.2 Fixed-Smoothing Asymptotics under the Null

To establish the asymptotic distribution of the generalized Q statistic $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ when K is fixed as $T \to \infty$, we need to make the following high-level assumptions, all of which are standard in the literature on fixed-smoothing asymptotics:

Assumption 1 *For $\ell = 1, 2, \ldots, K$, the basis functions $\Phi_\ell(\cdot)$ are piecewise monotonic, continuously differentiable, and orthonormal in $L^2[0,1]$, and satisfy $\int_0^1 \Phi_\ell(x)\, dx = 0$.*

Assumption 2 *For $r \in (0,1]$, $T^{-1/2}\sum_{t=1}^{[Tr]} f_t(\theta_0) \Rightarrow \Omega_0^{1/2} W_f(r)$, where $W_f(r)$ is a standard s-dimensional Brownian motion process and Ω_0 has full rank s.*

Assumption 3 *For any $\tilde{\theta}_T = \theta_0 + O_p\left(T^{-1/2}\right)$,*

$$T^{-1}\sum_{t=1}^{[Tr]} \frac{\partial f_t\left(\tilde{\theta}_T\right)}{\partial \theta'} \xrightarrow{p} r\Gamma_0$$

uniformly in $r \in (0,1]$ where $\Gamma_0 = \Gamma(\theta_0)$ for $\Gamma(\theta) = \mathbb{E}\frac{\partial f_t(\theta)}{\partial \theta'}$. In addition, Γ_0 has a full rank p.

Assumption 4 *$\sqrt{T}(\hat{\theta}_T - \theta_0) \Rightarrow \Psi_0$ for some distribution Ψ_0.*

Assumptions 2 and 3 are high-level assumptions. For vector ARMA models, sufficient conditions for the FCLT in Assumption 2 and its proof can be found in MS (2018, the proof of Theorem 3 in Section A.2). These sufficient conditions include a moment condition on $\{\varepsilon_{\theta_0,t}\}$ and a strong mixing condition; see Assumption A7 in MS (2018). For other time series models, we may impose similar conditions and follow the same arguments to establish the FCLT.

Assumption 3 is a form of the uniform law of large numbers (ULLN). Let Θ_0 be a closed ball with center θ_0 and radius $T^{-1/2+\delta}$ for some $\delta > 0$. Assumption 3 holds if the pointwise LLN holds:

$$T^{-1}\sum_{t=1}^{[Tr]}\frac{\partial f_t(\theta)}{\partial\theta'}\xrightarrow{p} r\Gamma(\theta) \text{ for each } \theta\in\Theta_0 \text{ and } r\in[0,1],$$

and the empirical process

$$H_T(\theta,r)=T^{-1}\sum_{t=1}^{T}\left(\frac{\partial f_t(\theta)}{\partial\theta'}-\Gamma(\theta)\right)1\{t\leq[Tr]\}$$

is stochastically equicontinuous on $\Theta_0\times[0,1]$. See Andrews (1992, Theorem 1). Like the FCLT in Assumption 2, the above sufficient conditions hold under some moment and mixing conditions.

Assumption 4 requires only a $\sqrt{T}$-consistent estimator. The form of the asymptotic distribution Ψ_0 does not matter, as it will not affect the asymptotic distribution of the test statistic. Non-normal Ψ_0 is allowed, even though Ψ_0 is typically a normal distribution. As we discuss later, the convergence rate requirement on $\hat{\theta}_T$ can be relaxed.

Under Assumption 3, $\hat{A}(\hat{\theta}_T)=\mathbb{M}_{\hat{\Gamma}(\hat{\theta}_T)}\to^p\mathbb{M}_{\Gamma_0}=A(\theta_0):=A_0$. Then, by the first-order Taylor expansion, we have

$$\begin{aligned}A(\hat{\theta}_T)'\Lambda_\ell(\hat{\theta}_T)&=A(\hat{\theta}_T)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)f_t(\hat{\theta}_T)\\&=A(\theta_0)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)f_t(\theta_0)(1+o_p(1))\\&\quad+\left[\frac{1}{T}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)A(\check{\theta}_T)'\frac{\partial f_t(\check{\theta}_T)}{\partial\theta'}\right]\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)(1+o_p(1)),\end{aligned}$$

where $\check{\theta}_T$ is a value between $\hat{\theta}_T$ and θ_0. Using Assumptions 1, 3 and 4, we can show that

$$\left[\frac{1}{T}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)A(\check{\theta}_T)'\frac{\partial f_t(\check{\theta}_T)}{\partial\theta'}\right]\cdot\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)=o_p(1). \tag{8}$$

See the proof of Lemma 1(b) in Sun (2014b) for details. Therefore,

$$A(\hat{\theta}_T)'\Lambda_\ell(\hat{\theta}_T) = A\left(\theta_0\right)' \Lambda_\ell(\theta_0) + o_p\left(1\right).$$

A key assumption underlying this result is that $\int_0^1 \Phi_\ell\left(x\right)dx = 0$. It then follows that

$$A(\hat{\theta}_T)'\hat{\Omega}(\hat{\theta}_T)A(\hat{\theta}_T) = A\left(\theta_0\right)' \hat{\Omega}(\theta_0)A\left(\theta_0\right) + o_p\left(1\right).$$

Note that the rank $A(\hat{\theta}_T)'\hat{\Omega}(\hat{\theta}_T)A(\hat{\theta}_T)$ remains the same as that of $A\left(\theta_0\right)' \hat{\Omega}(\theta_0)A\left(\theta_0\right)$ with probability approaching 1 as T increases. Using Stewart (1969) and the continuous mapping theorem, we have

$$\left[A(\hat{\theta}_T)'\hat{\Omega}(\hat{\theta}_T)A(\hat{\theta}_T)\right]^+ = \left[A\left(\theta_0\right)' \hat{\Omega}(\theta_0)A\left(\theta_0\right)\right]^+ + o_p\left(1\right). \tag{9}$$

Next, by the first-order Taylor expansion, we have

$$\begin{aligned} A(\hat{\theta}_T)'\frac{1}{\sqrt{T}}\sum_{t=1}^T f_t(\hat{\theta}_T) &= A(\theta_0)'\frac{1}{\sqrt{T}}\sum_{t=1}^T f_t\left(\theta_0\right) \\ &\quad + \left[\frac{1}{T}\sum_{t=1}^T A(\check{\theta}_T)'\frac{\partial f_t\left(\check{\theta}_T\right)}{\partial\theta'}\right]\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)\left(1 + o_p\left(1\right)\right). \end{aligned}$$

Invoking Assumptions 3 and 4, we have

$$\left[\frac{1}{T}\sum_{t=1}^T A(\check{\theta}_T)'\frac{\partial f_t\left(\check{\theta}_T\right)}{\partial\theta'}\right]\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) = A_0'\Gamma_0\left(1 + o_p\left(1\right)\right) = o_p\left(1\right), \tag{10}$$

where we have used $A_0'\Gamma_0 = 0$, which is built into the construction of our test statistic. As a result,

$$A(\hat{\theta}_T)'\frac{1}{\sqrt{T}}\sum_{t=1}^T f_t(\hat{\theta}_T) = A(\theta_0)'\frac{1}{\sqrt{T}}\sum_{t=1}^T f_t\left(\theta_0\right) + o_p\left(1\right). \tag{11}$$

Combining (9) and (11), we obtain

$$\begin{aligned}
\tilde{Q}_T^{(s)}(\hat{\theta}_T) &= \frac{K-q+1}{Kq}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t\left(\theta_0\right)\right]' \mathbb{G}(A(\theta_0),\hat{\Omega}(\theta_0))\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t\left(\theta_0\right)\right]\left(1+o_p\left(1\right)\right) \\
&= \tilde{Q}_T^{(s)}(\theta_0)\left(1+o_p\left(1\right)\right).
\end{aligned}$$

We present this result formally as a lemma.

Lemma 2 *Let Assumptions 1–4 hold. Then, under H_0,*

$$\tilde{Q}_T^{(s)}(\hat{\theta}_T) = \tilde{Q}_T^{(s)}(\theta_0)\left(1+o_p\left(1\right)\right).$$

The lemma shows that $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ and $\tilde{Q}_T^{(s)}(\theta_0)$ are asymptotically equivalent: $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ and $\tilde{Q}_T^{(s)}(\theta_0)$ will converge weakly to the same distribution. Hence, the weak limit of $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ does not depend on $\hat{\theta}_T$. Every $\sqrt{T}$-consistent estimator $\hat{\theta}_T$ will lead to the same asymptotic distribution for $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$; all we need is the $\sqrt{T}$ convergence rate of $\hat{\theta}_T$. The estimator $\hat{\theta}_T$ can exploit the moment conditions that $\mathbb{E}\left[f_t\left(\theta_0\right)\right]=0$, but does not have to. Intuitively, our test statistic $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ examines only the directions orthogonal to those that the estimation error can have an asymptotic effect. This is a consequence of using matrix $\hat{A}(\hat{\theta}_T)$ to transform the residual sample autocovariances.

The keys to the asymptotic equivalence result and hence the invariance of the asymptotic distribution of $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ to $\hat{\theta}_T$ are (8) and (10). If we assume that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} A(\tilde{\theta}_T)' \frac{\partial f_t\left(\tilde{\theta}_T\right)}{\partial \theta'} = O_p\left(1\right) \tag{12}$$

uniformly over r and for any $\tilde{\theta}_T = \theta_0 + o_p\left(1\right)$, then Lemma 2 holds for any consistent estimator $\hat{\theta}_T$. We do not even require a rate of convergence for $\hat{\theta}_T$. Such a result resembles the result in a two-step estimation framework where the first-step estimation error has no effect on the asymptotic distribution of the second-step estimator when an orthogonality condition holds. A sufficient condition for (12) is that the empirical process $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} A(\theta)' \frac{\partial f_t(\theta)}{\partial \theta'}$ converges to a tight and continuous process with zero mean at $\theta=\theta_0$ for any $r\in[0,1]$.

We proceed to use Lemma 2 to establish the asymptotic distribution of $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$. Under Assumptions 1 and 2, using summation and integration by part and the continuous mapping theorem, we can obtain

$$A\left(\theta_0\right)' \Lambda_\ell(\theta_0) := A_0' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_\ell\left(\frac{t}{T}\right) f_t\left(\theta_0\right) \Rightarrow A_0' \Omega_0^{1/2} \int_0^1 \Phi_\ell\left(r\right) dW_f\left(r\right) := A_0' \Omega_0^{1/2} \eta_{f,\ell}, \tag{13}$$

where

$$\eta_{f,\ell} = \int_0^1 \Phi_\ell\left(r\right) dW_f\left(r\right) \sim iidN(0, \mathbb{I}_s) \text{ over } \ell = 1, \ldots, K.$$

The fact that $\eta_{f,\ell}$ is iid over $\ell = 1, \ldots, K$ follows from the assumption that $\{\Phi_\ell\left(\cdot\right)\}$ are orthonormal in $L^2[0,1]$.

Combining the above with

$$A(\theta_0)' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t\left(\theta_0\right) \Rightarrow A_0' \Omega_0^{1/2} W_f\left(1\right),$$

Lemma 2, and the continuous mapping theorem yields:

$$\begin{aligned} \tilde{Q}_T^{(s)}(\hat{\theta}_T) &\Rightarrow \frac{K-q+1}{Kq} \left[A_0' \Omega_0^{1/2} W_f\left(1\right)\right]' \left[A_0' \Omega_0^{1/2} \Pi_0 \Omega_0^{1/2} A_0\right]^{+} \left[A_0' \Omega_0^{1/2} W_f\left(1\right)\right] \\ &= \frac{K-q+1}{Kq} \left[\Omega_0^{1/2} W_f\left(1\right)\right]' \mathbb{G}\left(A_0, \Omega_0^{1/2} \Pi_0 \Omega_0^{1/2}\right) \left[\Omega_0^{1/2} W_f\left(1\right)\right], \end{aligned}$$

where

$$\Pi_0 = \frac{1}{K} \sum_{\ell=1}^{K} \eta_{f,\ell} \eta_{f,\ell}'.$$

The asymptotic distribution of $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ looks complicated. The following theorem shows that the asymptotic distribution is actually a standard F distribution. This is an intriguing and convenient result.

Theorem 2 *Let Assumptions 1–4 hold. Then, under H_0,*

$$\tilde{Q}_T^{(s)}(\hat{\theta}_T) \Rightarrow F\left(q, K-q+1\right),$$

where $F(q, K-q+1)$ is the standard F distribution with q and $K-q+1$ degrees of freedom.

This result is similar to Theorem 1 in Wang and Sun (2020). Here, the degrees of freedom of the F distribution are adjusted according to the number of parameters estimated. The proof, however, is more challenging than that in Wang and Sun (2020). We first use Lemma 1 to simplify $\mathbb{G}(A_0, \Omega_0^{1/2}\Pi_0\Omega_0^{1/2})$ and then use the rotational invariance of the standard normal and Wishart distributions and singular value decompositions to reduce the asymptotic distribution to the F distribution.

Theorem 2 allows us to perform the test using F critical values. For easy reference, we call the test the generalized F test, or simply the F test, when there is no confusion. The generalized F test is as easy to use as the conventional Q tests based on chi-squared approximations. It is also more convenient to use than the MS (2018) test, which is based on a nonstandard reference distribution.

To implement the generalized F test, we need to choose the smoothing parameter K. We can follow Sun, Phillips and Jin (2008) and Sun (2014a) to choose a testing-optimal K that minimizes a weighted average of the type I and type II errors of the proposed test. However, this approach requires high-order expansions that have to be established under each specific data generating process. Given that we maintain a general setting and the high-order expansions are highly technical, we opt for a traditional approach here. Following Phillips (2005), we choose K to minimize the mean square error of $\hat{\Omega}(\hat{\theta}_T)$. Although the MSE-based approach is not best suited for hypothesis testing, Monte Carlo simulations in the next section reveal that this choice of K delivers good finite sample performances for the generalized F test.

3.3 Fixed-Smoothing Asymptotics under the Local Alternative

We consider the following local alternatives

$$H_{1T}^{(s)} : \gamma_{\theta_0}^{(s)} = \frac{\delta}{\sqrt{T}}.$$

Under the local alternatives, Assumptions 1, 3, and 4 can still be valid. Assumption 2 has to be replaced by the following assumption.

Assumption 5 *For* $r \in (0,1]$, $T^{-1/2}\sum_{t=1}^{[Tr]} f_t(\theta_0) \Rightarrow r\delta + \Omega_0^{1/2} W_f(r)$.

Theorem 3 *Let Assumptions 1, 3, 4, and 5 hold. Then, under* $H_{1T}^{(s)}$,

$$\tilde{Q}_T^{(s)}(\hat{\theta}_T) \Rightarrow F_\lambda(q, K-q+1),$$

where

$$\begin{aligned}\lambda &= \delta' A(\Gamma_0;\Omega_0)\delta \\ &= \delta'\left[\Omega_0^{-1} - \Omega_0^{-1}\Gamma_0\left(\Gamma_0'\Omega_0^{-1}\Gamma_0\right)^{-1}\Gamma_0'\Omega_0^{-1}\right]\delta\end{aligned}$$

and $F_\lambda(q, K-q+1)$ *is the noncentral F distribution with noncentrality parameter* λ *and* q *and* $K-q+1$ *degrees of freedom.*

Note that the noncentrality parameter is not equal to $\delta'\Omega_0^{-1}\delta$. If θ_0 is known, we can show that the noncentrality parameter will be $\delta'\Omega_0^{-1}\delta$. Thus, the generalized F test has power against any local departure of order $1/\sqrt{T}$ from the null. When θ_0 is not known and has to be estimated, the generalized F test can have trivial power in certain directions, that is, the directions with $\delta \neq 0$ but $\lambda = 0$. More specifically, if δ is in the space spanned by the columns of Γ_0 such that $\delta = \Gamma_0 B$ for any $p \times p$ matrix B, then

$$\lambda = B\Gamma_0'\left[\Omega_0^{-1} - \Omega_0^{-1}\Gamma_0\left(\Gamma_0'\Omega_0^{-1}\Gamma_0\right)^{-1}\Gamma_0'\Omega_0^{-1}\right]\Gamma_0 B = 0,$$

and the test has only trivial power. Note that the effect of the estimation error in $\hat{\theta}_T$ on the residual autocovariances is $\Gamma_0(\hat{\theta}_T - \theta_0)$. If the local departure aligns perfectly with the effect of the estimation error, our approach will not be able to detect this departure.

In general, we have $\lambda \leq \delta'\Omega_0^{-1}\delta$, and so the generalized F test becomes less powerful when the model parameters have to be estimated. Thus, there is a cost of estimating the model parameters. However, this is typical and similar to the problem in any over-identification test.

4 MONTE CARLO EVIDENCE

4.1 ARMA Models

In this section, we investigate the finite sample performances of the generalized F test, that is, the asymptotic F test based on the generalized Q statistic $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$.

We consider ARMA models with possible weak innovations. The null is an AR(1) model such that

$$X_t = 0.9X_{t-1} + \varepsilon_t.$$

For $\eta_t \sim iid\ N(0,1)$, we consider the following error specifications:

M1: iid normal process: $\{\varepsilon_t\}$ is a sequence of iid $N(0,1)$ random variables.

M2: GARCH (1,1) process: $\varepsilon_t = h_t\eta_t$, where $h_t^2 = 0.1 + 0.09\varepsilon_{t-1}^2 + 0.9h_{t-1}^2$.

M3: GJR-GARCH process: $\varepsilon_t = h_t\eta_t$, where

$$h_t^2 = 0.01 + 0.7h_{t-1}^2 + 0.1\varepsilon_{t-1}^2 + 0.03\varepsilon_{t-1}^2 1(\varepsilon_{t-1} < 0) + 0.01\varepsilon_{t-3}^2 1(\varepsilon_{t-3} < 0).$$

M4: all-pass ARMA(1,1) process: $\varepsilon_t = 0.8\varepsilon_{t-1} + \eta_t - (1/0.8)\,\eta_{t-1}$.

M5: bilinear process: $\varepsilon_t = \eta_t + 0.5\eta_{t-1}\varepsilon_{t-2}$.

M6: heteroskedastic bilinear process: $\varepsilon_t = \nu_t + 0.5\nu_{t-1}\varepsilon_{t-2}$, where $\nu_t = h_t\eta_t$ and $h_t^2 = 0.1 + 0.09\nu_{t-1}^2 + 0.9h_{t-1}^2$.

M7: non-MD-1 process: $\varepsilon_t = \eta_t^2\eta_{t-1}$.

M8: NLMA process: $\varepsilon_t = \eta_{t-2}\eta_{t-1}\left(\eta_{t-2} + \eta_t + 1\right)$.

In error specifications M1–M3, $\{\varepsilon_t\}$ are MD sequences. The iid specification in M1 serves as a basic benchmark. The GARCH specifications in M2 and M3 are empirically relevant in the financial literature. M4–M8 are non-MD processes but have zero autocorrelations. M4 and M5 are examined in Lobato et al. (2002). M6 allows for heteroskedasticity in the bilinear process and exhibits stronger dependence than M5. M7 is examined in MS (2018), while M8 is examined in Horowitz et al. (2006).

We investigate the size properties for sample sizes $T = 100$ and 200. The number of simulation replications is $10,000$. The ARMA models are estimated by the OLS method. The nominal level of all tests is 5%. For the maximum order of autocovariance s, we consider the values up to 10 and 15 for $T = 100$ and 200, respectively. Figures 1 and 2 report the empirical rejection probabilities of the generalized F test. For comparison, these two figures also report the empirical rejection probabilities of the transformed Box-Pierce test proposed by DV (2011), the Q test with the nonpivotal distribution approximation proposed by FRZ (2005), the nonstandard test proposed by MS (2018) and the bootstrapped Q tests proposed by Zhu (2016).

For the DV test, the first $s + p$ autocorrelations are employed to transform the first s autocorrelations, and the critical values of the resulting Box-Pierce type test are based on a chi-squared distribution with s degrees of freedom. The test is implemented using two different variance matrix estimators of sample innovation autocorrelations. The first estimator is a matrix with diagonal elements equal to

$$\frac{1}{T} \sum_{t=j+1}^{T} \hat{\varepsilon}^2_{\hat{\theta}_T,t} \hat{\varepsilon}^2_{\hat{\theta}_T,t-j} / \hat{\gamma}_{\hat{\theta}_T}(0), j = 1, \dots, s + p.$$

The second estimator is the *vector autoregressive heteroskedasticity and autocorrelation consistent* (VARHAC) variance estimator described in Lobato et al. (2002), with the VAR order selected by the Bayesian information criterion. The maximum VAR order is set at $1.2T^{1/3}$. This procedure differs slightly from Den Haan and Levin (1997): it only conducts a global VAR order search, instead of searching for a unique autoregression order for each series separately[5]. For the FRZ test, we follow FRZ (2005) and employ Imhof's (1961) algorithm to compute the critical values of the nonpivotal asymptotic distribution of $Q_T^{(s)}(\hat{\theta}_T)$. The reference distribution of the MS test is nonstandard, and the critical values are obtained from Table 1 in Lobato (2001). We examine the finite properties of $Q_T^{(s)}(\hat{\theta}_T)$, employing the bootstrapped critical values from both the regular (i.e., non-blocking) and blockwise procedures of Zhu (2016). The bootstrap sample size is $J = 500$, and the random weights are generated from the standard exponential distribution. For the blockwise boostrapping procedure, the block size is set equal to 3.

The main features of the results are as follows:

i. For the iid process, the empirical rejection probabilities of the generalized F test, the DV tests,

[5] In preliminary simulations, we also consider the Newey and West (1994) data-driven procedure, used by Delgado and Velasco (2011), but the results are much worse than the VARHAC approach.

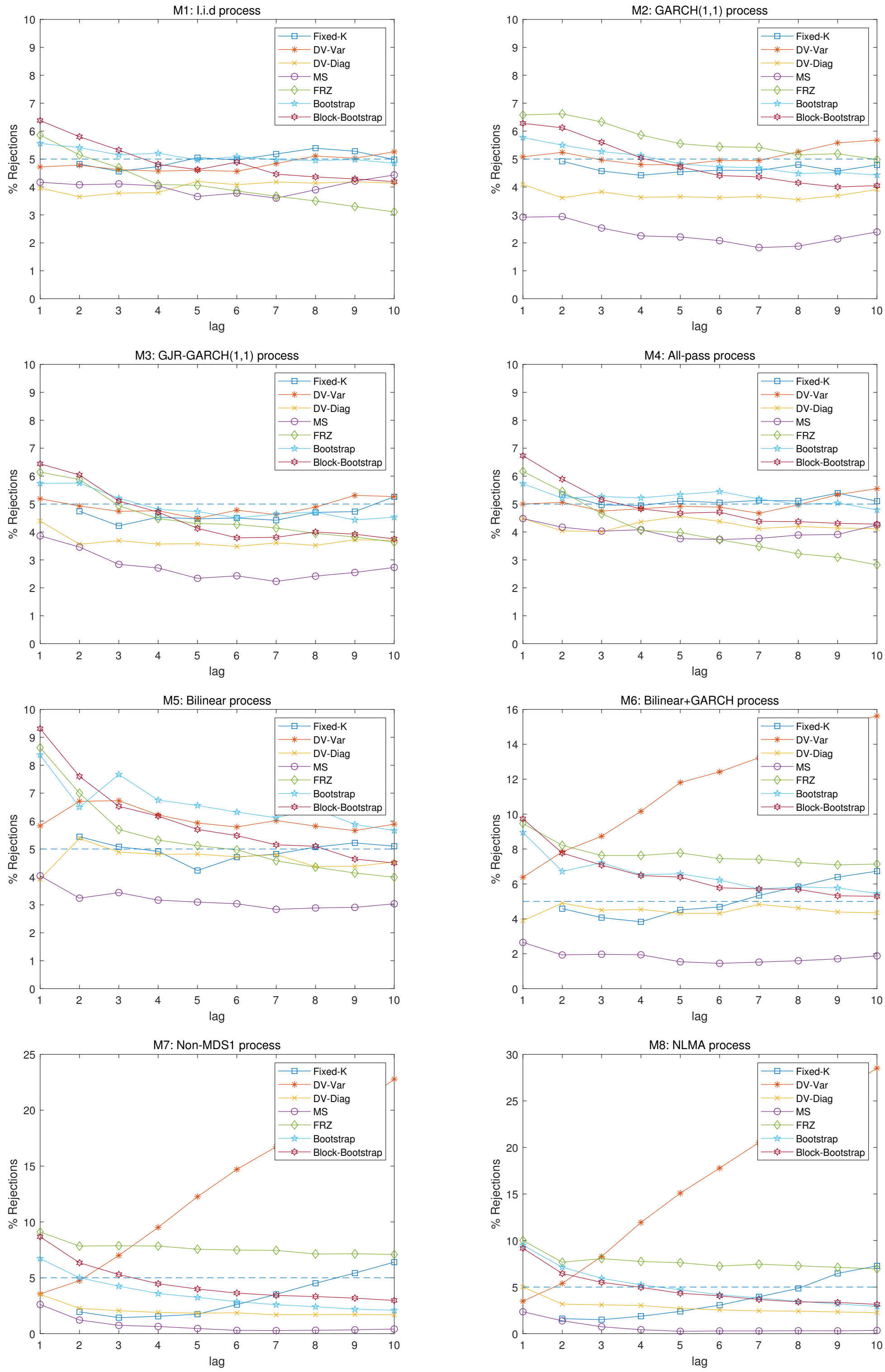

Figure 1: Percentage of rejections of portmanteau tests in terms of lag s under the null with $T = 100$. Fixed-K represents the generalized F test based on the modified statistic $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ (this paper); MS represents the nonstandard MS test; DV-Var represents the DV test with VARHAC variance estimator; DV-Diag represents the DV test with diagonal variance estimator; FRZ represents the Q test with a nonpivotal distribution approximation proposed by FRZ. Bootstrap represents the Q test with a nonblock-bootstrapped critical value of Zhu (2016); Block-Bootstrap represents the Q test with a block-bootstrapped critical value of Zhu (2016).

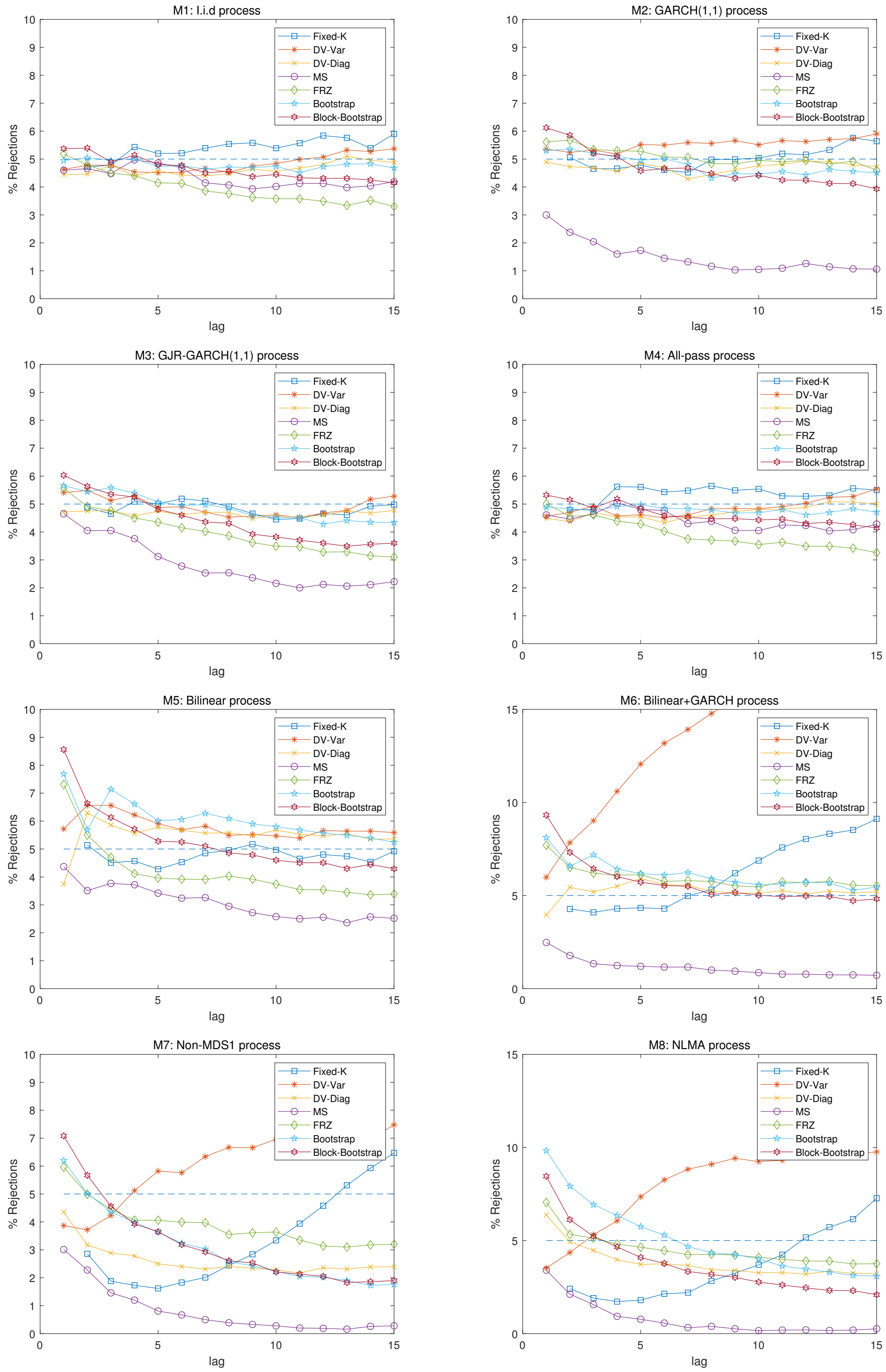

Figure 2: Percentage of rejections of portmanteau tests in terms of lag s under the null with $T = 200$. Fixed-K represents the generalized F test based on the modified statistic $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ (this paper); MS represents the nonstandard MS test; DV-Var represents the DV test with VARHAC variance estimator; DV-Diag represents the DV test with diagonal variance estimator; FRZ represents the Q test with a nonpivotal distribution approximation proposed by FRZ. Bootstrap represents the Q test with a nonblock-bootstrapped critical value of Zhu (2016); Block-Bootstrap represents the Q test with a block-bootstrapped critical value of Zhu (2016).

and the Zhu tests are reasonably close to the nominal level of 5%. However, both the FRZ and MS tests tend to be undersized, especially when s is large.

ii. For the GARCH process, the empirical rejection probabilities of the generalized F, DV, and FRZ tests are close to the nominal level of 5% for all s considered, while the Zhu tests tend to be slightly oversized when s is small. In contrast, the MS test is heavily undersized, especially when s is large.

iii. For the GJR-GARCH and all-pass ARMA (1,1) processes, the empirical size patterns of these tests are similar to those of tests with iid errors.

iv. For bilinear processes, the empirical rejection probabilities of the generalized F test are quite close to the nominal level of 5%. However, those of the DV test with the VARHAC variance estimator tend to exceed the nominal level of 5% and become much worse for the heteroskedastic bilinear process, whereas the DV test with the diagonal variance estimator controls the size sufficiently well. The Zhu tests tend to be oversized when s is small. On the other hand, the FRZ test can be over-sized or under-sized, and the MS test is heavily under-sized.

v. For the non-MD-1 process and the NLMA process, the empirical rejection probabilities of the generalized F test are reasonably close to the nominal level. The FRZ test is over-sized when $T = 100$, but tends to be under-sized when $T = 200$. The DV test with the VARHAC variance estimator is heavily over-sized, and the one with the diagonal variance estimator is under-sized. The MS test is heavily under-sized, especially when s is large. The Zhu (2016) tests tend to oversized, especially when s is small.

Overall, the generalized F test performs remarkably well for the uncorrelated error specifications we consider here. The DV tests perform reasonably well with MD error specifications for both the diagonal and VARHAC variance estimators, but tend to be heavily oversized with non-MD error specifications for the VARHAC procedure. Interestingly, for the non-MD cases, the DV test performs much better with the diagonal matrix procedure than with the VARHAC procedure, although the former estimator is not consistent for the population variance matrix. This could, to a certain degree, indicate that increasing-smoothing asymptotic approximations do not work well in finite samples. The Zhu tests tend to be oversized when s is small. Furthermore, for the MDS processes (M1-M3), the blockwise bootstrapping procedure delivers worse size performances than the regular procedure, this tendency is reversed when it comes to the non-MDS processes (M4-M8). As for other competing tests, the MS test tends to be heavily undersized, while the FRZ test tends to be undersized in many cases.

To investigate the power properties, we consider the following ARMA(1,1) model:

$$X_t = 0.7X_{t-1} + \varepsilon_t + 0.2\varepsilon_{t-1},$$

where the error specifications are again M1 to M8. The null hypothesis (i.e., the weak AR(1)) is tested at the nominal level 5%. We use $1,000$ replications for sample sizes 100 and 200. For power comparisons, the critical values are adjusted to make the empirical rejection probabilities of the tests under the null exactly 5%, except for the Zhu tests. Note that empirical size adjustment is not practically feasible. The empirical power curves are reported in Figures 3 and 4.

The main features of the results are as follows:

i. The generalized F test has substantial power and is comparable to the DV tests and the Zhu tests in many cases.

ii. The FRZ and MS tests suffer from substantial power loss even after empirical size adjustment.

5 AN EMPIRICAL APPLICATION TO S&P500 RETURNS

In this section, we revisit the application to the daily returns of the S&P500 index, which is considered by FRZ (2005) and Zhu (2016). The price index $\{P_t\}$ ranges from January 3, 1979 to December 31, 2001, containing $5,808$ observations. The series of log-returns $\log(P_t/P_{t-1})$ is denoted as $\{y_t\}_1^{5807}$. FRZ (2005) and Zhu (2016) show that, at the 5% level of significance, the strong white noise hypothesis is rejected, whereas the weak white noise hypothesis is not rejected.

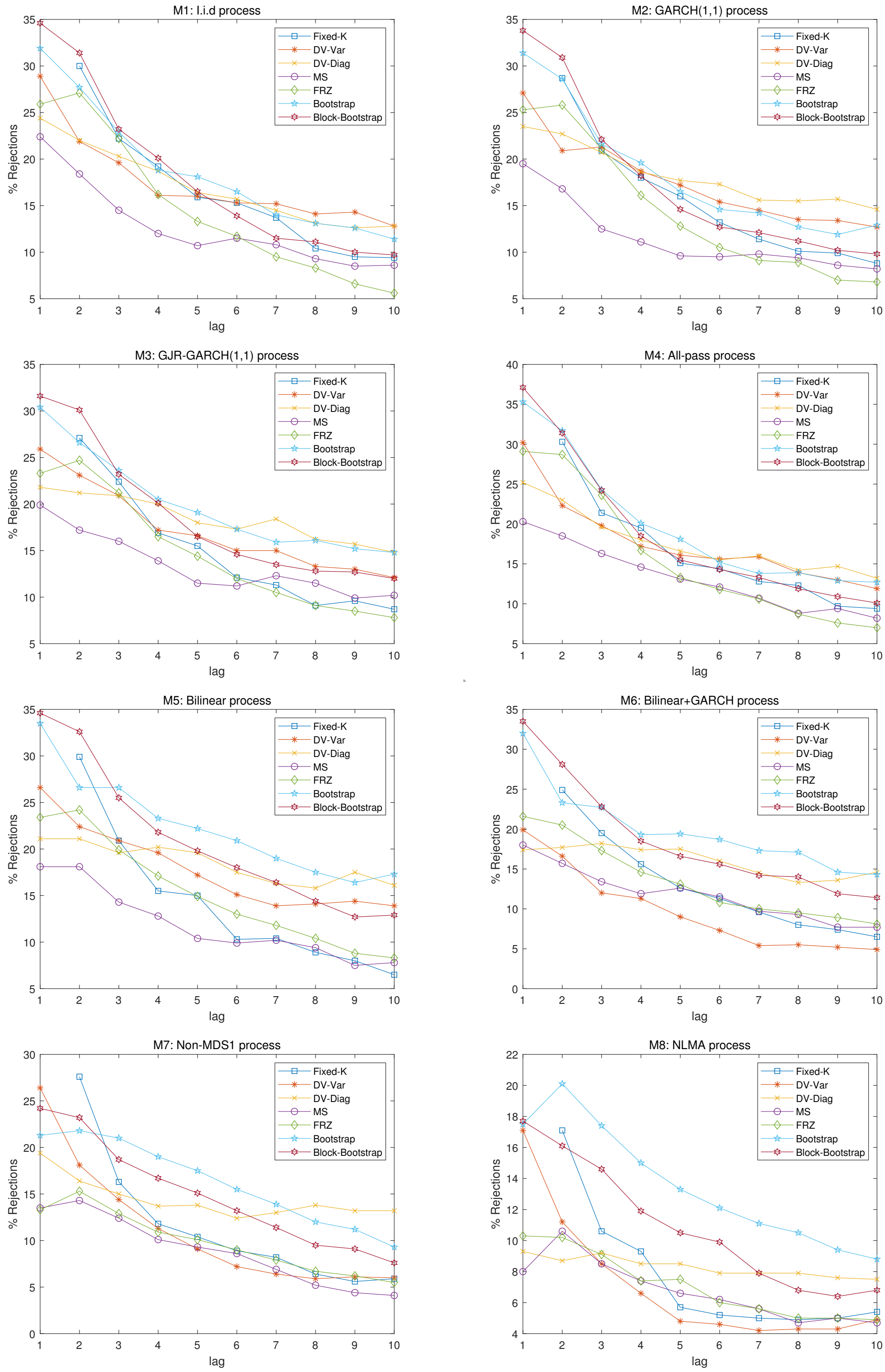

Figure 3: Power of the portmanteau tests in terms of lag s with $T = 100$. Fixed-K represents the generalized F test based on the modified statistic $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ (this paper); MS represents the nonstandard MS test; DV-Var represents the DV test with VARHAC variance estimator; DV-Diag represents the DV test with diagonal variance estimator; and FRZ represents the Q test with a nonpivotal distribution approximation proposed by FRZ. Bootstrap represents the Q test with a nonblock-bootstrapped critical value of Zhu (2016); Block-Bootstrap represents the Q test with a block-bootstrapped critical value of Zhu (2016).

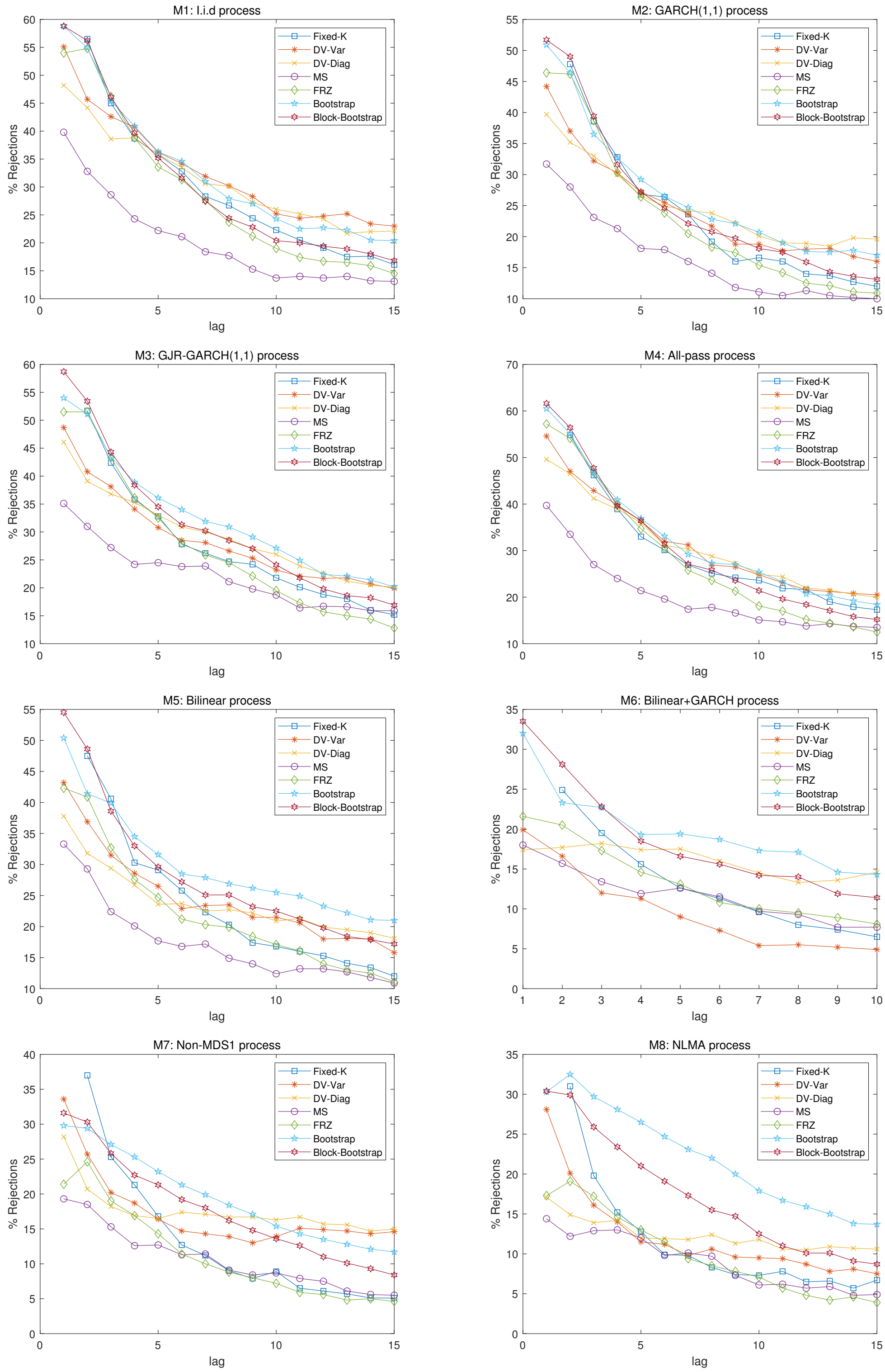

Figure 4: Power of portmanteau tests in terms of lag s with $T = 200$. Fixed-K represents the generalized F test based on the modified statistic $\tilde{Q}_T^{(s)}(\hat{\theta}_T)$ (this paper); MS represents the non-standard MS test; DV-Var represents the DV test with the VARHAC variance estimator; DV-Diag represents the DV test with the diagonal variance estimator; and FRZ represents the Q test with a nonpivotal distribution approximation proposed by FRZ. Bootstrap represents the Q test with a nonblock-bootstrapped critical value of Zhu (2016); Block-Bootstrap represents the Q test with a block-bootstrapped critical value of Zhu (2016).

We check whether a (G)ARCH model is adequate to fit $\{y_t\}$. We consider a GARCH(1,1) model:

$$y_t = h_t \eta_t,$$

$$h_t^2 = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}^2,$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. Under some assumptions that imply the second-order stationarity of y_t^2, it is easy to show that y_t^2 admits an ARMA(1,1) representation such that

$$X_t = (\alpha + \beta) X_{t-1} + v_t - \beta v_{t-1}, \tag{14}$$

where $X_t = y_t^2 - E\left(y_t^2\right)$, $v_t = h_t^2\left(\eta_t^2 - 1\right)$. When $\{\eta_t\}$ is iid, $\{v_t\}$ is an MDS; when $\{\eta_t\}$ is an MDS, $\{v_t\}$ is an uncorrelated sequence. Thus we can check the adequacy of a (G)ARCH model for $\{y_t\}$ by checking whether $\{v_t\}$ is uncorrelated up to some fixed order in (14).

Let $\tilde{X}_t := y_t^2 - \frac{1}{T}\sum_{\tau=1}^{T} y_\tau^2$ be the mean-corrected series. We first test the null that $\left\{\tilde{X}_t\right\}$ satisfies a weak AR(1) model. The fitted AR(1) model is $\tilde{X}_t - 0.113\tilde{X}_{t-1} = v_t$. Table 1 reports the p-values of the generalized F test, the Zhu test with the regular bootstrapping procedure, the FRZ test, and the DV test. From the results of the generalized F test, we find that, up to order 15, the null that $\left\{\tilde{X}_t\right\}$ satisfies a weak AR(1) model can be rejected at the 5% significance level when $s > 3$. However, the Zhu test and the FRZ test would not reject the null at the 5% significance level, up to order 15. The DV test would reject the null at the 5% significance level, when $s > 12$.

As a caveat, we note that all the diagnostic tests under consideration serve to test model adequacy. If we reject the null of zero innovation autocorrelation up to *any* order s, then the model is deemed inadequate as it does not absorb all autocorrelations. On the other hand, if we do not reject the null for a given s, then we can not say that the model is adequate because the test only examines autocorrelations up to the order s. Autocorrelation of order higher than s may be still present. Also, due to the parameter estimation error, we may fail to reject the null even if the innovations are autocorrelated. The choice of s is an empirical question that should reflect the maximum order of innovation autocorrelation that we care about.

Lag	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Fixed-K		0.29	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bootstrap	0.73	0.25	0.27	0.28	0.25	0.25	0.25	0.29	0.25	0.25	0.25	0.25	0.24	0.24	0.26
FRZ	0.25	0.13	0.30	0.34	0.28	0.25	0.25	0.23	0.42	0.37	0.36	0.27	0.27	0.26	0.22
DV-Diag	0.28	0.13	0.24	0.25	0.30	0.39	0.35	0.26	0.26	0.10	0.08	0.06	0.02	0.02	0.02

Table 1: The p-values of the generalized F test (fixed-K), the bootstrapped Portmanteau test of Zhu (2016) (Bootstrap), the FRZ test and the DV test (DV-Diag) for the null of a weak AR(1) model.

We next test the null that $\left\{\tilde{X}_t\right\}$ satisfies a weak ARMA(1,1) model. The fitted ARMA(1,1) model is $\tilde{X}_t - 0.832\tilde{X}_{t-1} = v_t - 0.727v_{t-1}$. Table 2 reports the p-values of our generalized F test, the FRZ test[6] and the Zhu test with the regular bootstrapping procedure. It shows that, up to order 15, a weak ARMA (1,1) model cannot be rejected at the 5% significance level. This is in accordance with the conclusion in FRZ (2005) and Zhu (2016).

Lag	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Fixed-K			0.32	0.61	0.69	0.62	0.59	0.78	0.83	0.90	0.87	0.91	0.94	0.79	0.54
FRZ	0.58	0.52	0.57	0.43	0.38	0.38	0.37	0.37	0.37			0.38			
Bootstrap	0.59	0.72	0.65	0.51	0.51	0.52	0.51	0.51	0.51	0.51	0.51	0.52	0.51	0.52	0.53

Table 2: The p-values of the generalized F test (fixed-K) and the bootstrapped Portmanteau test of Zhu (2016) (Bootstrap) for the null of a weak ARMA(1,1) model.

6 CONCLUSION

In this paper, we propose a simple and asymptotically F-distributed portmanteau test for diagnosing a general parametric time series model. In this framework, the parameter estimation error can have non-negligible effects on both the sample covariances of the residuals and the estimator of the asymptotic variance of these sample covariances. We use a transform to remove its effect on the sample covariances and employ the orthonormal series variance estimator, a special asymptotic variance estimator that is not affected by the parameter estimation uncertainty in large samples. It is the combination of the transform and the use of the orthonormal series variance estimator that make the convenient asymptotic F theory possible. In an extensive Monte Carlo simulation study,

[6] As the same data set is used, we just replicate the p-values from Table 5 in FRZ (2005). So the values for $s = 10, 11, 13, 14, 15$ are missing.

we find that the proposed F test is more accurate than existing tests with competitive power in finite samples. It would be interesting to extend this methodology to test the autocorrelations of the residuals from multivariate time series models. We will pursue this extension in future research.

7 APPENDIX: PROOFS OF THE MAIN RESULTS

Proof of Lemma 1. We have

$$\begin{aligned} & A(\Gamma;\Delta)\cdot\left[A(\Gamma;\Delta)'\cdot\Omega\cdot A(\Gamma;\Delta)\right]^{+}A(\Gamma;\Delta)' \\ = \;& \Delta^{-1/2}\mathbb{M}_{\Delta^{-1/2}\Gamma}\cdot\left[\mathbb{M}_{\Delta^{-1/2}\Gamma}\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\mathbb{M}_{\Delta^{-1/2}\Gamma}\right]^{+}\mathbb{M}_{\Delta^{-1/2}\Gamma}\Delta^{-1/2}, \end{aligned}$$

where

$$\mathbb{M}_{\Delta^{-1/2}\Gamma}=\mathbb{I}_s-\Delta^{-1/2}\Gamma\left(\Gamma'\Delta^{-1}\Gamma\right)^{-1}\Gamma'\Delta^{-1/2}.$$

Let

$$\Delta^{-1/2}\Gamma=\tilde{U}\tilde{\Xi}\tilde{V}'$$

be the SVD of $\Delta^{-1/2}\Gamma$ where

$$\tilde{\Xi}_{s\times p}=\begin{pmatrix}\tilde{D}\\ O\end{pmatrix},$$

$\tilde{D}\in\mathbb{R}^{p\times p}$ is a diagonal matrix, and $\tilde{U}$ and $\tilde{V}$ are orthogonal matrices. We write $\tilde{U}=\left(\tilde{U}_{sp},\tilde{U}_{sq}\right)$ for $\tilde{U}_{sp}\in\mathbb{R}^{s\times p}$ and $\tilde{U}_{sq}\in\mathbb{R}^{s\times q}$. Then

$$\begin{aligned} \mathbb{M}_{\Delta^{-1/2}\Gamma} &= \mathbb{I}_s-\tilde{U}\tilde{\Xi}\tilde{V}'\left(\tilde{V}\tilde{\Xi}'\tilde{U}'\tilde{U}\tilde{\Xi}\tilde{V}'\right)^{-1}\tilde{V}\tilde{\Xi}'\tilde{U}'=\mathbb{I}_s-\tilde{U}\tilde{\Xi}\left(\tilde{\Xi}'\tilde{\Xi}\right)^{-1}\tilde{\Xi}'\tilde{U}' \\ &= \tilde{U}\left[\mathbb{I}_s-\tilde{\Xi}\left(\tilde{\Xi}'\tilde{\Xi}\right)^{-1}\tilde{\Xi}'\right]\tilde{U}'=\tilde{U}\begin{pmatrix}O & O\\ O & \mathbb{I}_q\end{pmatrix}\tilde{U}'. \end{aligned}$$

Hence

$$
\begin{aligned}
& \left[A\left(\Gamma;\Delta\right)'\cdot\Omega\cdot A\left(\Gamma;\Delta\right)\right]^{+} \\
= & \left[\mathbb{M}_{\Delta^{-1/2}\Gamma}\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\mathbb{M}_{\Delta^{-1/2}\Gamma}\right]^{+} \\
= & \left[\tilde{U}\begin{pmatrix} O & O \\ O & \mathbb{I}_q \end{pmatrix}\tilde{U}'\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\tilde{U}\begin{pmatrix} O & O \\ O & \mathbb{I}_q \end{pmatrix}\tilde{U}'\right]^{+} \\
= & \tilde{U}\begin{pmatrix} O & O \\ O & \left[\tilde{U}_{sq}'\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\tilde{U}_{sq}\right]^{-1} \end{pmatrix}\tilde{U}',
\end{aligned}
$$

and

$$
\begin{aligned}
& A\left(\Gamma;\Delta\right)\cdot\left[A\left(\Gamma;\Delta\right)'\cdot\Omega\cdot A\left(\Gamma;\Delta\right)\right]^{+}A\left(\Gamma;\Delta\right)' \\
= & \Delta^{-1/2}\mathbb{M}_{\Delta^{-1/2}\Gamma}\cdot\left[\mathbb{M}_{\Delta^{-1/2}\Gamma}\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\mathbb{M}_{\Delta^{-1/2}\Gamma}\right]^{+}\mathbb{M}_{\Delta^{-1/2}\Gamma}\Delta^{-1/2} \\
= & \Delta^{-1/2}\tilde{U}\begin{pmatrix} O & O \\ O & \mathbb{I}_q \end{pmatrix}\begin{pmatrix} O & O \\ O & \left[\tilde{U}_{sq}'\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\tilde{U}_{sq}\right]^{-1} \end{pmatrix}\begin{pmatrix} O & O \\ O & \mathbb{I}_q \end{pmatrix}\tilde{U}'\Delta^{-1/2} \\
= & \begin{pmatrix} O \\ \tilde{U}_{sq}'\Delta^{-1/2} \end{pmatrix}'\begin{pmatrix} O & O \\ O & \left[\tilde{U}_{sq}'\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\tilde{U}_{sq}\right]^{-1} \end{pmatrix}\begin{pmatrix} O \\ \tilde{U}_{sq}'\Delta^{-1/2} \end{pmatrix} \\
= & \Delta^{-1/2}\tilde{U}_{sq}\left[\tilde{U}_{sq}'\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\tilde{U}_{sq}\right]^{-1}\tilde{U}_{sq}'\Delta^{-1/2}.
\end{aligned}
$$

Now, using the SVD's of Γ and $\Delta^{-1/2}\Gamma$, we have

$$
\Delta^{-1/2}\Gamma = \Delta^{-1/2}U\begin{pmatrix} D \\ O \end{pmatrix}V' \text{ and } \Delta^{-1/2}\Gamma = \tilde{U}\begin{pmatrix} \tilde{D} \\ O \end{pmatrix}\tilde{V}',
$$

and so

$$
\Delta^{-1/2}U\begin{pmatrix} D \\ O \end{pmatrix}V' = \tilde{U}\begin{pmatrix} \tilde{D} \\ O \end{pmatrix}\tilde{V}'.
$$

Since both $\tilde{U}$ and V' are orthogonal matrices, we obtain

$$\tilde{U}'\Delta^{-1/2}U\left(\begin{array}{c} D \\ O \end{array}\right) = \left(\begin{array}{c} \tilde{D} \\ O \end{array}\right)\tilde{V}'V.$$

That is,

$$\left(\begin{array}{cc} \tilde{U}_{sp}'\Delta^{-1/2}U_{sp} & \tilde{U}_{sp}'\Delta^{-1/2}U_{sq} \\ \tilde{U}_{sq}'\Delta^{-1/2}U_{sp} & \tilde{U}_{sq}'\Delta^{-1/2}U_{sq} \end{array}\right)\left(\begin{array}{c} D \\ O \end{array}\right) = \left[\begin{array}{c} \tilde{D} \\ O \end{array}\right]\tilde{V}'V.$$

It follows from the second block of the above equation that

$$\left(\Delta^{-1/2}\tilde{U}_{sq}\right)' U_{sp} = O_{qp}.$$

Therefore, $\Delta^{-1/2}\tilde{U}_{sq}$ lies in the orthogonal complement of the space spanned by the columns of U_{sp}. Given that $\{U_{sq}\}$ are complete orthogonal bases for this space and $\Delta^{-1/2}\tilde{U}_{sq}$ has a full column rank q, we have

$$\Delta^{-1/2}\tilde{U}_{sq} = U_{sq}\tilde{C}$$

for some nonsingular $q \times q$ matrix $\tilde{C}$. Using this, we have

$$\begin{aligned}
& A(\Gamma;\Delta)\cdot\left[A(\Gamma;\Delta)'\cdot\Omega\cdot A(\Gamma;\Delta)\right]^{+} A(\Gamma;\Delta)' \\
= \; & \Delta^{-1/2}\tilde{U}_{sq}\left[\tilde{U}_{sq}'\left(\Delta^{-1/2}\Omega\Delta^{-1/2}\right)\tilde{U}_{sq}\right]^{-1}\tilde{U}_{sq}'\Delta^{-1/2} \\
= \; & U_{sq}\tilde{C}\left[\tilde{C}'U_{sq}'\Omega U_{sq}\tilde{C}\right]^{-1}\tilde{C}'U_{sq}' = U_{sq}\left[U_{sq}'\Omega U_{sq}\right]^{-1}U_{sq}',
\end{aligned}$$

as desired. ■

Proof of Theorem 2. Under Assumptions 1–4, we have proven in the main text that

$$\begin{aligned}
& \tilde{Q}_T^{(s)}\left(\hat{\theta}_T\right) \\
\Rightarrow \; & \frac{K-q+1}{Kq}W_f(1)'\,\Omega^{1/2}A_0\left[A_0'\Omega_0^{1/2}\Pi_0\Omega_0^{1/2}A_0\right]^{+}A_0'\Omega_0^{1/2}\left[W_f(1)\right] \\
= \; & \frac{K-q+1}{Kq}W_f(1)'\,\Omega^{1/2}U_{0,sq}\left[U_{0,sq}'\Omega_0^{1/2}\Pi_0\Omega_0^{1/2}U_{0,sq}\right]^{+}U_{0sq}'\Omega_0^{1/2}\left[W_f(1)\right],
\end{aligned}$$

where $U_{0,sq}$ contains the left singular vectors of Γ_0 corresponding to the zero singular value. Here the equality follows from Lemma 1.

Note that $\eta_{f,\ell}$ is normal and

$$\text{cov}\left[\eta_{f,\ell_1}, \eta_{f,\ell_2}\right] = \int_0^1 \Phi_{\ell_1}(r)\,\Phi_{\ell_2}(r)\,dr = 1\left\{\ell_1 = \ell_2\right\},$$

using the orthonormality of $\{\Phi_\ell(\cdot), \ell = 1, \ldots, K\}$ on $L_2[0,1]$. So $\eta_{f,\ell} \sim iidN(0, \mathbb{I}_s)$ over $\ell = 1, \ldots, K$. As a result, $K\Pi_0 \sim \mathbb{W}(\mathbb{I}_s, K)$, a Wishart random variable.

Because $W_f(1)$ and $\eta_{f,\ell}$ are standard normals and for $\ell = 1, \ldots, K$,

$$\begin{aligned}\text{cov}(\eta_{f,\ell}, W_f(1)) &= \mathbb{E}\left[\int_0^1 \Phi_\ell(r)\,dW_f(r)\,W_f(1)'\right] \\ &= \mathbb{E}\left[\int_0^1 \Phi_\ell(r)\,dW_f(r)\int_0^1 dW_f'(v)\right] = \mathbb{I}_s\int_0^1 \Phi_\ell(r)\,dr = 0,\end{aligned}$$

we know that $\left\{\eta_{f,\ell}, \ell = 1, \ldots, K\right\}$ is independent of $W_f(1)$. As a result, $K\Pi_0$ is independent of $W_f(1)$.

We partition Π_0 as

$$\Pi_0 = \frac{1}{K}\sum_{\ell=1}^K \eta_{f,\ell}\eta_{f,\ell}' = \begin{pmatrix} \Pi_{pp} & \Pi_{pq} \\ \Pi_{qp} & \Pi_{qq} \end{pmatrix},$$

where Π_{ij} denotes $i \times j$ matrices. Let $\underset{s\times s}{\tilde{U}_0}\,\underset{s\times q}{\tilde{\Xi}_0}\,\underset{q\times q}{\tilde{V}_0'}$ be a singular value decomposition of $\Omega_0^{1/2}U_{0,sq}$. By definition, $\tilde{U}_0'\tilde{U}_0 = \mathbb{I}_s$, $\tilde{V}_0'\tilde{V}_0 = \mathbb{I}_q$, and

$$\tilde{\Xi}_0 = \begin{bmatrix} O \\ \tilde{D}_0 \end{bmatrix}$$

where $\tilde{D}_0$ is diagonal and all diagonal elements are positive almost surely. Then

$$\begin{aligned} & W_f(1)'\,\Omega_0^{1/2}U_{0,sq}\left[U_{0,sq}'\Omega_0^{1/2}\Pi_0\Omega_0^{1/2}U_{0,sq}\right]^+ U_{0,sq}'\Omega_0^{1/2}W_f(1) \\ = \; & W_f(1)'\,\tilde{U}_0\tilde{\Xi}_0\tilde{V}_0'\left[\tilde{V}_0\tilde{\Xi}_0'\tilde{U}_0'\cdot\Pi_0\cdot\tilde{U}_0\tilde{\Xi}_0\tilde{V}_0'\right]^+\tilde{V}_0\tilde{\Xi}_0'\tilde{U}_0'W_f(1) \\ = \; & \left[\tilde{U}_0'W_f(1)\right]'\tilde{\Xi}_0\left[\tilde{\Xi}_0'\cdot\tilde{U}_0'\Pi_0\tilde{U}_0\cdot\tilde{\Xi}_0\right]^+\tilde{\Xi}_0'\left[\tilde{U}_0'W_f(1)\right] \\ =^d \; & \left[W_f(1)\right]'\tilde{\Xi}_0\left[\tilde{\Xi}_0'\Pi_0\tilde{\Xi}_0\right]^+\tilde{\Xi}_0'\left[W_f(1)\right] \end{aligned}$$

where the distributional equivalence in the last line holds because $\left(\tilde{U}_0'\Pi_0\tilde{U}_0, \tilde{U}_0'W_f(1)\right)$ has the same joint distribution as $(\Pi_0, W_f(1))$. Some simple algebra shows that

$$\begin{aligned}
&\tilde{\Xi}_0\left[\tilde{\Xi}_0'\Pi_0\tilde{\Xi}_0\right]^+\tilde{\Xi}_0' \\
=&\begin{pmatrix} O \\ \tilde{D}_0 \end{pmatrix}\left[\begin{pmatrix} O & \tilde{D}_0 \end{pmatrix}\begin{pmatrix} \Pi_{pp} & \Pi_{pq} \\ \Pi_{qp} & \Pi_{qq} \end{pmatrix}\begin{pmatrix} O \\ \tilde{D}_0 \end{pmatrix}\right]^+\begin{pmatrix} O & \tilde{D}_0 \end{pmatrix} \\
=&\begin{pmatrix} O \\ \tilde{D}_0 \end{pmatrix}\left(\tilde{D}_0\Pi_{qq}\tilde{D}_0\right)^{-1}\begin{pmatrix} O & \tilde{D}_0 \end{pmatrix} = \begin{pmatrix} O & O \\ O & \Pi_{qq}^{-1} \end{pmatrix}.
\end{aligned}$$

Let

$$W_f(1) = \begin{pmatrix} W_{f,p}(1) \\ W_{f,q}(1) \end{pmatrix}.$$

Then

$$\begin{aligned}
&[W_f(1)]'\,\tilde{\Xi}_0\left[\tilde{\Xi}_0'\Pi_0\tilde{\Xi}_0\right]^+\tilde{\Xi}_0'\,[W_f(1)] \\
=&\begin{pmatrix} W_{f,p}(1) \\ W_{f,q}(1) \end{pmatrix}'\begin{pmatrix} O & O \\ O & \Pi_{qq}^{-1} \end{pmatrix}\begin{pmatrix} W_{f,p}(1) \\ W_{f,q}(1) \end{pmatrix} \\
=& W_{f,q}'(1)\,\Pi_{qq}^{-1}W_{f,q}(1).
\end{aligned}$$

So

$$\tilde{Q}_T^{(s)}\left(\hat{\theta}_T\right) \Rightarrow \frac{K-q+1}{Kq}W_{f,q}(1)'\,\Pi_{qq}^{-1}W_{f,q}(1).$$

By Proposition 8.2 in Bilodeau and Brenner (1999), we have

$$W_{f,q}(1)'\,(K\Pi_{qq})^{-1}W_{f,q}(1) \overset{d}{=} F_c(q, K-q+1),$$

where $F_c(\cdot,\cdot)$ is a canonical F distribution. Finally, with some degrees-of-freedom adjustment, we obtain

$$\tilde{Q}_T^{(s)}\left(\hat{\theta}_T\right) \Rightarrow F(q, K-q+1).$$

■

Proof of Theorem 3. By Assumption 5 and the similar arguments in the proof of Theorem 2, we have

$$
\begin{aligned}
& \tilde{Q}_T^{(s)}\left(\hat{\theta}_T\right) \\
\Rightarrow\ & \frac{K-q+1}{Kq}\left[\Omega_0^{1/2}W_f(1)+\delta\right]' A_0\left[A_0'\Omega_0^{1/2}\Pi_0\Omega_0^{1/2}A_0\right]^{+} A_0'\left[\Omega_0^{1/2}W_f(1)+\delta\right] \\
=\ & \frac{K-q+1}{Kq}\left(W_f(1)+\tilde{\delta}\right)'\Omega^{1/2}U_{0,sq}\left[U_{0,sq}'\Omega_0^{1/2}\Pi_0\Omega_0^{1/2}U_{0,sq}\right]^{+}U_{0sq}'\Omega_0^{1/2}\left[W_f(1)+\tilde{\delta}\right] \\
=\ & \frac{K-q+1}{Kq}\left(W_f(1)+\tilde{U}_0'\tilde{\delta}\right)'\begin{pmatrix} O & O \\ O & \Pi_{qq}^{-1}\end{pmatrix}\left[W_f(1)+\tilde{U}_0'\tilde{\delta}\right] \\
=\ & \frac{K-q+1}{Kq}\left(W_{fq}(1)+\left[\tilde{U}_0'\tilde{\delta}\right]_q\right)'\Pi_{qq}^{-1}\left(W_{fq}(1)+\left[\tilde{U}_0'\tilde{\delta}\right]_q\right),
\end{aligned}
$$

where $\tilde{\delta}=\Omega_0^{-1/2}\delta$ and $\left[\tilde{U}_0'\tilde{\delta}\right]_q$ is the last q elements of $\tilde{U}_0'\tilde{\delta}$. Then, by Proposition 8.2 in Bilodeau and Brenner (1999), we have

$$\tilde{Q}_T^{(s)}\left(\hat{\theta}_T\right)\Rightarrow F_\lambda(q,K-q+1),$$

where

$$\lambda=\left\|\left[\tilde{U}_0'\tilde{\delta}\right]_q\right\|^2=\delta' A_0\left[A_0'\Omega_0A_0\right]^{+}A_0'\delta.$$

By Lemma 1, λ will not change if we replace $A_0:=A(\Gamma_0;\mathbb{I}_s)$ by $A(\Gamma_0;\Omega_0)$. But

$$A(\Gamma_0;\Omega_0)=\Omega_0^{-1/2}\mathbb{M}_{\Omega_0^{-1/2}\Gamma_0},$$

and so

$$
\begin{aligned}
\lambda &= \delta'\Omega_0^{-1/2}\mathbb{M}_{\Omega_0^{-1/2}\Gamma_0}\left[\mathbb{M}_{\Omega_0^{-1/2}\Gamma_0}\right]^{+}\mathbb{M}_{\Omega_0^{-1/2}\Gamma}\Omega_0^{-1/2}\delta \\
&= \delta'\Omega_0^{-1/2}\mathbb{M}_{\Omega_0^{-1/2}\Gamma}\Omega_0^{-1/2}\delta \\
&= \delta'\left[\Omega_0^{-1}-\Omega_0^{-1}\Gamma_0\left(\Gamma_0'\Omega_0^{-1}\Gamma_0\right)^{-1}\Gamma_0'\Omega_0^{-1}\right]\delta.
\end{aligned}
$$

■

References

[1] Andrews, Donald W. K. (1992). Generic Uniform Convergence. *Econometric Theory* 18(2), 241–257.

[2] Baillie, R. and Bollerslev, T. (2002). The Message in Daily Exchange Rates. *Journal of Business & Economic Statistics* 20(1), 60–68.

[3] Bilodeau, M., and Brenner, D. (1999). *Theory of Multivariate Statistics.* Springer.

[4] Box, G. E. P. and Pierce, D. A. (1970). Distribution of the Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models. *Journal of American Statistical Association* 65, 1509–26.

[5] Delgado, M. A., and Velasco, C. (2011). An Asymptotically Pivotal Transform of the Residuals Sample Autocorrelations With Application to Model Checking. *Journal of the American Statistical Association* 106, 946–958.

[6] Den Haan, W. J. and Levin, A. (1997). A Practitioner's Guide to Robust Covariance Matrix Estimation. *Handbook of statistics* 15, G. S. Maddala and C. R. Rao, eds., Elsevier, Amsterdam, 291–341.

[7] Duchesne, P., and Francq, C. (2008). On Diagnostic Checking Time Series Models With Portmanteau Test Statistics Based on Generalized Inverses and {2}-Inverses. in COMSTAT 2008, *Proceedings in Computational Statistics*, ed. P. Brito, Heidelberg: Physica-Verlag, pp. 143–154.

[8] Francq, C., and Zakoïan, J.-M. (1998). Estimating Linear Representations of Nonlinear Processes. *Journal of Statistical Planning and Inference*, 68, 145–165.

[9] Francq, C., Roy, R. and Zakoïan, J.-M. (2005). Diagnostic Checking in ARMA Models With Uncorrelated Errors. *Journal of the American Statistical Association* 100, 532–544.

[10] Francq, C., and Zakoïan, J.-M. (2010). GARCH Models: Structure, Statistical Inference and Financial Applications. Wiley. ISBN: 978-0-470-68391-0.

[11] Horowitz, J. L., Lobato, I. N., Nankervis, J. C. and Savin, N. E. (2006). Bootstrapping the Box-Pierce Q test: a Robust Test of Uncorrelatedness. *Journal of Econometrics* 133, 841–862.

[12] Imhof, J. P. (1961). Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika* 48, 419–426.

[13] Katayama, N. (2008). An Improvement of the Portmanteau Statistic. *Journal of Time Series Analysis* 29, 359–370.

[14] Liu, C. and Sun, Y. (2019). A Simple and Trustworthy Asymptotic t Test in Difference-in-Differences Regressions. *Journal of Econometrics*, 210(2), 327–362.

[15] Ljung, G. M. and Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika* 65, 297–303.

[16] Lazarus, E., Lewis, D. J., Stock, J. H. and Watson, M. W. (2016). HAR Inference: Kernel Choice, Size Distortions, and Power Losses. Working paper, Department of Economics, Harvard University.

[17] Lazarus, E., Lewis, D. J., Stock, J. H. and Watson, M. W. (2018). HAR Inference: Recommendations for Practice. *Journal of Business and Economic Statistics* 36(4), 541–559.

[18] Lobato, I. N. (2001). Testing that a Dependent Process is Uncorrelated. *Journal of the American Statistical Association* 96, 1066–1076.

[19] Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2002). Testing for Zero Autocorrelation in the Presence of Statistical Dependence. *Econometric Theory* 18, 730–743.

[20] Mainassara, Y. B. and Saussereau, B. (2018). Diagnostic Checking in Multivariate ARMA Models With Dependent Errors Using Normalized Residual Autocorrelations, *Journal of the American Statistical Association* 113(524), 1813–1827.

[21] Lundbergh, S. and Teräsvirta, T. (2002). Evaluating GARCH Models. *Journal of Econometrics* 110, 417–435.

[22] Mcleod, A. I. (1978). On the Distribution of Residual Autocorrelations in Box-Jenkins Models. *Journal of Royal Statistical Society B* 40, 296–302.

[23] Müller, U. K. (2007): A Theory of Robust Long-run Variance Estimation. *Journal of Econometrics* 141(2), 1331–1352.

[24] Newey, W. K. and West, K. D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* 61, 631–653.

[25] Phillips, P. C. B. (2005). HAC Estimation by Automated Regression. *Econometric Theory* 21, 116–142.

[26] Rosenberg J. and Engle R. (2002). Empirical pricing kernels. *Journal of Financial Economics* 64(3), 341–372.

[27] Shao, X. (2010). A Self-normalized Approach to Confidence Interval Construction in Time Series. *Journal of the Royal Statistical Society: Series B* 72, 343–366.

[28] Shao, X. (2015). Self-normalization for Time Series: a Review of Recent Developments. *Journal of the American Statistical Association* 110, 1797–1817.

[29] Stewart, G. W. (1969). On the Continuity of the Generalized Inverse, *SIAM Journal on Applied Mathematics* 17(1), 33–45.

[30] Stoica, P. and Moses, R. (2005). *Spectral Analysis of Signals.* Pearson Prentice Hall.

[31] Sun, Y. (2011). Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter. *Journal of Econometrics* 164(2), 345–366.

[32] Sun, Y. (2013). A Heteroskedasticity and Autocorrelation Robust F test using Orthonormal Series Variance Estimator. *Econometrics Journal* 16, 1–26.

[33] Sun, Y. (2014a). Let's Fix It: fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. *Journal of Econometrics* 178(3), 659–677.

[34] Sun, Y. (2014b). Fixed-smoothing Asymptotics in a Two-step GMM Framework. *Econometrica* 82(6), 2327–2370.

[35] Sun, Y. (2014c). Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation. *Advances in Econometrics* 33, 23–63.

[36] Sun, Y., Phillips, P. C. B., and Jin, S. (2008). Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing. *Econometrica* 76(1), 175–194.

[37] Thomson, D. J. (1982). Spectrum Estimation and Harmonic Analysis. *IEEE Proceedings* 70, 1055–1096.

[38] Wang, X. and Sun, Y. (2020). An Asymptotic F Test for Uncorrelatedness in the Presence of Time Series Dependence. *Journal of Time Series Analysis* 41, 536–550.

[39] Zhu, K. (2016). Bootstrapping the Portmanteau Tests in Weak Auto-regressive Moving Average Models. *Journal of the Royal Statistical Society: Series B* 78, 463–485.

[40] Zhang, X. and Shao, X. (2013). Fixed-smoothing Asymptotics for Time Series. *Annals of Statistics* 41, 1329–1349.