Testing for high-dimensional white noise using maximum cross correlations

JINYUAN CHANG
School of Statistics, Southwestern University of Finance and Economics, Chengdu
Sichuan 611130, China (jinyuan.chang@unimelb.edu.au)

QIWEI YAO
Department of Statistics, London School of Economics and Political Science
WC2A 2AE, UK (q.yao@lse.ac.uk)

WEN ZHOU
Department of Statistics, Colorado State University, Fort Collins
Colorado 80523, USA (riczw@stat.colostate.edu)

6 November 2016

Abstract

We propose a new omnibus test for vector white noise using the maximum absolute autocorrelations and cross-correlations of the component series. Based on an approximation by the L_∞-norm of a normal random vector, the critical value of the test can be evaluated by bootstrapping from a multivariate normal distribution. In contrast to the conventional white noise test, the new method is proved to be valid for testing the departure from white noise that is not independent and identically distributed. We illustrate the accuracy and the power of the proposed test by simulation, which also shows that the new test outperforms several commonly used methods including, for example, the Lagrange multiplier test and the multivariate Box–Pierce portmanteau tests, especially when the dimension of time series is high in relation to the sample size. The numerical results also indicate that the performance of the new test can be further enhanced when it is applied to pre-transformed data obtained via the time series principal component analysis proposed by Chang, Guo and Yao (arXiv:1410.2323). The proposed procedures have been implemented in an `R` package.

Keywords: Autocorrelation; Normal approximation; Parametric bootstrap; Portmanteau test; Time series principal component analysis; Vector white noise.

1 Introduction

Testing for white noise or serial correlation is a fundamental problem in statistical inference, as many testing problems in linear modelling can be transformed into a white noise test. Testing for white noise is often pursued in two different manners: (i) the departure from white noise is specified as an alternative hypothesis in the form of an explicit parametric family such as an autoregressive moving average model, and (ii) the alternative hypothesis is unspecified. With an explicitly specified alternative, a likelihood ratio test can be applied. Likelihood-based tests typically have more power to detect a specific form of the departure than omnibus tests which try to detect arbitrary departure from white noise. The likelihood approach has been taken further in the nonparametric context using the generalized likelihood ratio test initiated by Fan et al. (2001); see Section 7.4.2 of Fan & Yao (2003) and also Fan & Zhang (2004). Nevertheless many applications including model diagnosis do not lead to a natural alternative model. Therefore various omnibus tests, especially the celebrated Box–Pierce test and its variants, remain popular. Those portmanteau tests are proved to be asymptotically χ^2-distributed under the null hypothesis, which makes their application extremely easy. See Section 3.1 of Li (2004) and Section 4.4 of Lütkepohl (2005) for further information on those portmanteau tests.

While portmanteau tests are designed for testing white noise, their asymptotic χ^2-distributions are established under the assumption that observations under the null hypothesis are independent and identically distributed. However, empirical evidence, including that in Section 4 below, suggests that this may represent another case in which the theory is more restrictive than the method itself. Asymptotic theory of portmanteau tests for white noise that is not independent and identically distributed has attracted a lot of attention. One of the most popular approaches is to establish the asymptotic normality of a normalized portmanteau test statistic. An incomplete list in this endeavour includes Durlauf (1991), Romano & Thombs (1996), Deo (2000), Lobato (2001), Francq et al. (2005), Escanciano & Lobato (2009) and Shao (2011). However, the convergence is typically slow. Horowitz et al. (2006) proposed a double blockwise bootstrap method to test for white noise that is not independent and identically distributed.

In this paper we propose a new omnibus test for vector white noise. Instead of using a portmanteau-type statistic, the new test is based on the maximum absolute auto- and cross-correlation of all component time series. This avoids the impact of small correlations. When most auto- and cross-correlations are small, the Box–Pierce tests have too many degrees of freedom in their asymptotic distributions. In contrast the new test performs well when there is at least one large absolute auto- or cross-correlation at a non-zero lag. The null distribution of the maximum correlation test statistic can be approximated asymptotically by that of $|G|_\infty$, where G is a Gaussian random vector, and $|u|_\infty = \max_{1\le i\le s}|u_i|$ denotes the L_∞-norm of a vector $u = (u_1, \ldots, u_s)^{\mathrm{T}}$. Its critical values can therefore be evaluated by bootstrapping from a multivariate normal distribution.

An added advantage of the new test is its ability to handle high-dimensional series, in the sense that the number of series is as large as, or even larger than, their length. Nowadays, it is common to model and forecast many time series at once, which has direct applications in, among others, finance, economics, environmental and medical studies. The current literature on high-dimensional time series focuses on estimation and dimension-reduction aspects. See, for example, Basu & Michailidis (2015), and Guo et al. (2016) and the references within for high-dimensional vector autoregressive models, and Bai & Ng (2002), Forni et al. (2005), Lam & Yao (2012) and Chang et al. (2015) for high-dimensional time series factor models. The model diagnostics has largely been untouched, as far as we are aware. The test proposed in this paper represents an effort to fill in this gap.

We compare the performance of the new test with those of the three Box–Pierce types of portmanteau tests, the Lagrange multiplier test and a likelihood ratio test in simulation, which shows that the new test attains the nominal significance levels more accurately and is also more powerful when the dimension of time series is large or moderately large. Its performance can be further enhanced by first applying time series principal component analysis, proposed by Chang, Guo and Yao (arXiv:1410.2323).

Let $\otimes$ and vec denote, respectively, the Kronecker product and the vectorization for matrices, I_s be the $s \times s$ identity matrix, and $|A|_\infty = \max_{1\le i\le \ell, 1\le j\le m}|a_{ij}|$ for an $\ell \times m$ matrix $A \equiv (a_{i,j})$. Denote by $\lceil x \rceil$ and $\lfloor x \rfloor$, respectively, the smallest integer not less than x and the largest integer not greater than x.

2 Methodology

2.1 Tests

Let $\{\varepsilon_t\}$ be a p-dimensional weakly stationary time series with mean zero. Denote by $\Sigma(k) = \mathrm{cov}(\varepsilon_{t+k}, \varepsilon_t)$ and $\Gamma(k) = \mathrm{diag}\{\Sigma(0)\}^{-1/2}\Sigma(k)\mathrm{diag}\{\Sigma(0)\}^{-1/2}$, respectively, the autocovariance and the autocorrelation of ε_t at lag k, where $\mathrm{diag}(\Sigma)$ denotes the diagonal matrix consisting of the diagonal elements of Σ only. When $\Sigma(k) \equiv 0$ for all $k \neq 0$, $\{\varepsilon_t\}$ is white noise.

With the available observations $\varepsilon_1, \ldots, \varepsilon_n$, let

$$\widehat{\Gamma}(k) \equiv \{\widehat{\rho}_{ij}(k)\}_{1\le i,j\le p} = \mathrm{diag}\{\widehat{\Sigma}(0)\}^{-1/2}\widehat{\Sigma}(k)\mathrm{diag}\{\widehat{\Sigma}(0)\}^{-1/2} \tag{1}$$

be the sample autocorrelation matrix at lag k, where

$$\widehat{\Sigma}(k) = \frac{1}{n}\sum_{t=1}^{n-k} \varepsilon_{t+k}\varepsilon_t^{\mathrm{T}} \tag{2}$$

is the sample autocovariance matrix.

Consider the hypothesis testing problem

$$H_0 : \{\varepsilon_t\} \text{ is white noise} \quad \text{versus} \quad H_1 : \{\varepsilon_t\} \text{ is not white noise.} \tag{3}$$

Since $\Gamma(k) \equiv 0$ for any $k \ge 1$ under H_0, our test statistic T_n is defined as

$$T_n = \max_{1\le k\le K} T_{n,k}, \tag{4}$$

where $T_{n,k} = \max_{1\le i,j\le p} n^{1/2}|\widehat{\rho}_{ij}(k)|$ and $K \ge 1$ is a prescribed integer. We reject H_0 if $T_n > \mathrm{cv}_\alpha$, where $\mathrm{cv}_\alpha > 0$ is the critical value determined by

$$\mathrm{pr}(T_n > \mathrm{cv}_\alpha) = \alpha \tag{5}$$

under H_0, and $\alpha \in (0, 1)$ is the significance level of the test.

To determine cv_α, we need to derive the distribution of T_n under H_0. Proposition 1 below shows that the Kolmogorov distance between this distribution and that of the L_∞-norm of a $N(0, \Xi_n)$ random vector converges to zero, even when p diverges at an exponential rate of n, where

$$\Xi_n = (I_K \otimes W)E(\xi_n\xi_n^{\mathrm{T}})(I_K \otimes W), \tag{6}$$

$$\xi_n = n^{1/2}(\mathrm{vec}\{\widehat{\Sigma}(1)\}^{\mathrm{T}}, \ldots, \mathrm{vec}\{\widehat{\Sigma}(K)\}^{\mathrm{T}})^{\mathrm{T}}, \quad W = \mathrm{diag}\{\Sigma(0)\}^{-1/2} \otimes \mathrm{diag}\{\Sigma(0)\}^{-1/2}.$$

This paves the way to evaluate cv_α simply by drawing a bootstrap sample from $N(0, \widehat{\Xi}_n)$, where $\widehat{\Xi}_n$ is an appropriate estimator for Ξ_n.

Proposition 1. *Let Conditions* 1–4 *in Section* 3 *below hold and* $G \sim N(0, \Xi_n)$. *There exists a positive constant* δ_1 *depending only on the constants appeared in Conditions* 1–4 *for which* $\log p \le Cn^{\delta_1}$ *for some constant* $C > 0$. *Then it holds under* H_0 *that*

$$\sup_{s\ge 0} \big|\mathrm{pr}(T_n > s) - \mathrm{pr}(|G|_\infty > s)\big| \to 0, \qquad \text{as } n \to \infty.$$

By replacing Ξ_n in (6) by $\widehat{\Xi}_n$, where $\widehat{\Xi}_n$ is defined in Section 2.2 below, the critical value cv_α in (5) can be replaced by $\widehat{\mathrm{cv}}_\alpha$ which is determined by

$$\mathrm{pr}(|G|_\infty > \widehat{\mathrm{cv}}_\alpha) = \alpha, \tag{7}$$

where $G \sim N(0, \widehat{\Xi}_n)$. In practice, we can draw $G_1, \ldots, G_B$ independently from $N(0, \widehat{\Xi}_n)$ for a large integer B. The $\lfloor B\alpha \rfloor$-th largest value among $|G_1|_\infty, \ldots, |G_B|_\infty$ is taken as the critical value $\widehat{\mathrm{cv}}_\alpha$. We then reject H_0 whenever $T_n > \widehat{\mathrm{cv}}_\alpha$.

Remark 1. *When p is large or moderately large, it is advantageous to apply the time series principal component analysis proposed in arXiv:1410.2323 to the data first. We denote by T_n^* the resulted test. More precisely, we compute an invertible transformation matrix Q using the* R *function* `segmentTS` *in the package* `PCA4TS` *available at* `CRAN`*. Then T_n^* is defined in the same manner as T_n in (4) with $\{\varepsilon_1, \ldots, \varepsilon_n\}$ replaced by $\{\varepsilon_1^*, \ldots, \varepsilon_n^*\}$, where $\varepsilon_t^* = Q\varepsilon_t$. As Q does not depend on t, $\{\varepsilon_t,\, t \geq 1\}$ is white noise if and only if $\{\varepsilon_t^*,\, t \geq 1\}$ is white noise. The time series principal component analysis makes the component autocorrelations as large as possible by suppressing the cross-correlations among different components at all time lags. This makes the maximum correlation greater, and therefore the test more powerful. See also the simulation results in Section 4.*

2.2 Estimation of Ξ_n

By Lemma 3.1 of Chernozhukov et al. (2013), the proposed test in Section 2.1 is valid if the estimator $\widehat{\Xi}_n$ satisfies $|\widehat{\Xi}_n - \Xi_n|_\infty = o_p(1)$. We construct such an estimator now even when the dimension of time series is ultra-high, i.e. $p \gg n$. Let $\tilde{n} = n - K$ and

$$f_t = \{\mathrm{vec}(\varepsilon_{t+1}\varepsilon_t^{\mathrm{T}}), \ldots, \mathrm{vec}(\varepsilon_{t+K}\varepsilon_t^{\mathrm{T}})\}^{\mathrm{T}} \quad (t = 1, \ldots, \tilde{n}). \tag{8}$$

The second factor $E(\xi_n\xi_n^{\mathrm{T}})$ on the right-hand side of (6) is closely related to $\mathrm{var}(\tilde{n}^{-1/2}\sum_{t=1}^{\tilde{n}} f_t)$, the long-run covariance of $\{f_t\}_{t=1}^{\tilde{n}}$. The long-run covariance plays an important role in the inference with dependent data. There exist various estimation methods for long-run covariances, including the kernel-type estimators (Andrews, 1991), and the estimators utilizing the moving block bootstraps (Lahiri, 2003). See also Den Haan & Levin (1997) and Kiefer et al. (2000). We adopt the kernel-type estimator for the long-run covariance of $\{f_t\}_{t=1}^{\tilde{n}}$

$$\widehat{J}_n = \sum_{j=-\tilde{n}+1}^{\tilde{n}-1} \mathcal{K}\left(\frac{j}{b_n}\right) \widehat{H}(j), \tag{9}$$

where $\widehat{H}(j) = \tilde{n}^{-1}\sum_{t=j+1}^{\tilde{n}} f_t f_{t-j}^{\mathrm{T}}$ if $j \geq 0$ and $\widehat{H}(j) = \tilde{n}^{-1}\sum_{t=-j+1}^{\tilde{n}} f_{t+j} f_t^{\mathrm{T}}$ otherwise, $\mathcal{K}(\cdot)$ is a symmetric kernel function that is continuous at 0 with $\mathcal{K}(0) = 1$, and b_n is the bandwidth diverging with n. Among a variety of kernel functions that guarantee the positive definiteness of the long-run covariance estimators, Andrews (1991) derived an optimal kernel, i.e. the quadratic spectral kernel

$$\mathcal{K}_{QS}(x) = \frac{25}{12\pi^2x^2}\left\{\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right\} \tag{10}$$

by minimizing the asymptotic truncated mean square error of the estimator. For the numerical study in Section 4, we always use this kernel function with an explicitly specified bandwidth selection procedure. The theoretical results in Section 3 apply to general kernel functions. As now $\widehat{J}_n$ in (9) provides an estimator for $E(\xi_n\xi_n^{\mathrm{T}})$, Ξ_n in (6) can be estimated by

$$\widehat{\Xi}_n = (I_K \otimes \widehat{W})\widehat{J}_n(I_K \otimes \widehat{W}),$$

where $\widehat{W} = \text{diag}\{\widehat{\Sigma}(0)\}^{-1/2} \otimes \text{diag}\{\widehat{\Sigma}(0)\}^{-1/2}$ for $\widehat{\Sigma}(0)$ defined in (2). Simulation results show that the proposed test with this estimator performs very well.

2.3 Computational issues

To draw a random vector $G \sim N(0, \widehat{\Xi}_n)$, the standard approach consists of three steps: (i) perform the Cholesky decomposition for the $p^2K \times p^2K$ matrix $\widehat{\Xi}_n = L^{\mathrm{T}}L$, (ii) generate p^2K independent $N(0,1)$ random variables $z = (z_1, \ldots, z_{p^2K})^{\mathrm{T}}$, (iii) perform transformation $G = L^{\mathrm{T}}z$. Computationally this is an $(np^4K^2 + p^6K^3)$-hard problem requiring a large storage space for $\{f_t\}_{t=1}^{\tilde{n}}$ and matrix $\widehat{\Xi}_n$. To circumvent the high computing cost with large p and/or K, we propose a method below which requires to generate random variables from an $\tilde{n}$-variate normal distribution instead.

Let Θ be an $\tilde{n} \times \tilde{n}$ matrix with the (i,j)-th element $\mathcal{K}\{(i-j)/b_n\}$. Let $\eta = (\eta_1, \ldots, \eta_{\tilde{n}})^{\mathrm{T}} \sim N(0, \Theta)$ be a random vector independent of $\{\varepsilon_1, \ldots, \varepsilon_n\}$. Then it is easy to see that conditionally on $\{\varepsilon_1, \ldots, \varepsilon_n\}$,

$$G = (I_K \otimes \widehat{W}) \left(\frac{1}{\sqrt{\tilde{n}}} \sum_{t=1}^{\tilde{n}} \eta_t f_t \right) \sim N(0, \widehat{\Xi}_n). \quad (11)$$

Thus a random sample from $N(0, \widehat{\Xi}_n)$ can be obtained from a random sample from $N(0, \Theta)$ via (11). The computational complexity of the new method is only $O(n^3)$, independent of p and K. The required storage space is also much smaller.

3 Theoretical properties

Write $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{p,t})^{\mathrm{T}}$ for each $t = 1, \ldots, n$. To investigate the theoretical properties of the proposed testing procedure, we need the following regularity conditions.

Condition 1. *There exists a constant $C_1 > 0$ independent of p such that var$(\varepsilon_{i,t}) \geq C_1$ uniformly holds for any $i = 1, \ldots, p$.*

Condition 2. *There exist three constants $C_2, C_3 > 0$ and $r_1 \in (0, 2]$ independent of p such that $\sup_t \sup_{1 \leq i \leq p} \text{pr}(|\varepsilon_{i,t}| > x) \leq C_2 \exp(-C_3 x^{r_1})$ for any $x > 0$.*

Condition 3. *Assume that $\{\varepsilon_t\}$ is β-mixing in the sense that $\beta_k \equiv \sup_t E\{\sup_{B \in \mathcal{F}_{t+k}^{\infty}} |\text{pr}(B \mid \mathcal{F}_{-\infty}^{t}) - \text{pr}(B)|\} \to 0$ as $k \to \infty$, where $\mathcal{F}_{-\infty}^{u}$ and $\mathcal{F}_{u+k}^{\infty}$ are the σ-fields generated respectively by $\{\varepsilon_t\}_{t \leq u}$ and $\{\varepsilon_t\}_{t \geq u+k}$. Furthermore there exist two constants $C_4 > 0$ and $r_2 \in (0, 1]$ independent of p such that $\beta_k \leq \exp(-C_4 k^{r_2})$ for all $k \geq 1$.*

Condition 4. *There exists a constant $C_5 > 0$ and $\iota > 0$ independent of p such that*

$$\begin{aligned} C_5^{-1} &< \liminf_{q \to \infty} \inf_{m \geq 0} E\left(\left| \frac{1}{q^{1/2}} \sum_{t=m+1}^{m+q} \varepsilon_{i,t+k}\varepsilon_{j,t} \right|^{2+\iota} \right) \\ &\leq \limsup_{q \to \infty} \sup_{m \geq 0} E\left(\left| \frac{1}{q^{1/2}} \sum_{t=m+1}^{m+q} \varepsilon_{i,t+k}\varepsilon_{j,t} \right|^{2+\iota} \right) < C_5, \quad (i, j = 1, \ldots, p; k = 1, \ldots, K). \end{aligned}$$

Condition 1 ensures that all component series are not degenerate. Condition 2 is a common assumption in the literature on ultra high-dimensional data analysis. It ensures exponential-type upper bounds for the tail probabilities of the statistics concerned. The β-mixing assumption in Condition 3 is mild. Causal autoregressive moving average processes with continuous innovation distributions are β-mixing with exponentially decaying β_k. So are the stationary Markov chains satisfying certain conditions. See Section 2.6.1 of Fan & Yao (2003) and the references within. In fact stationary generalized autoregressive conditional heteroskedasticity models with finite second moments and continuous innovation distributions are also β-mixing with exponentially decaying β_k; see Proposition 12 of Carrasco & Chen (2002). If we only require $\sup_t \sup_{1\le i\le p} \mathrm{pr}(|\varepsilon_{i,t}| > x) = O\{x^{-2(\nu+\epsilon)}\}$ for any $x > 0$ in Condition 2 and $\beta_k = O\{k^{-\nu(\nu+\epsilon)/(2\epsilon)}\}$ in Condition 3 for some $\nu > 2$ and $\epsilon > 0$, we can apply Fuk–Nagaev type inequalities to construct the upper bounds for the tail probabilities of the statistics for which our testing procedure still works for p diverging at some polynomial rate of n. We refer to Section 3.2 of arXiv:1410.2323 for the implementation of Fuk–Nagaev type inequalities in such a scenario. The β-mixing condition can be replaced by the α-mixing condition under which we can justify the proposed method for p diverging at some polynomial rate of n by using Fuk–Nagaev type inequalities. However, it remains open to establish the relevant properties under α-mixing for p diverging at some exponential rate of n. Condition 4 is a technical assumption for the validity of the Gaussian approximation for dependent data.

Our main asymptotic results indicate that the critical value $\widehat{\mathrm{cv}}_\alpha$ defined in (7) by the normal approximation is asymptotically valid, and, furthermore, the proposed test is consistent.

Theorem 1. *Let Conditions 1–4 hold, $|\mathcal{K}(x)| \asymp |x|^{-\tau}$ as $|x| \to \infty$ for some $\tau > 1$, and $b_n \asymp n^{\rho}$ for some $0 < \rho < \min\{(\tau-1)/(3\tau), r_2/(2r_2+1)\}$. Let $\log p \le Cn^{\delta_2}$ for some positive constants δ, C, and δ depend on the constants in Conditions 1–4 only. Then it holds under H_0 that*

$$\mathrm{pr}(T_n > \widehat{\mathrm{cv}}_\alpha) \to \alpha, \quad n \to \infty.$$

Theorem 2. *Assume that the conditions of Theorem 1 hold. Let ϱ be the largest element in the main diagonal of Ξ_n, and $\lambda(p,\alpha) = \{2\log(p^2K)\}^{1/2} + \{2\log(1/\alpha)\}^{1/2}$. Suppose that*

$$\max_{1\le k\le K} \max_{1\le i,j\le p} |\rho_{i,j}(k)| \ge \varrho^{1/2}(1+\epsilon_n)n^{-1/2}\lambda(p,\alpha)$$

for some positive ϵ_n satisfying $\epsilon_n \to 0$ and $\epsilon_n^2 \log p \to \infty$. Then it holds under H_1 that

$$\mathrm{pr}(T_n > \widehat{\mathrm{cv}}_\alpha) \to 1, \quad n \to \infty.$$

4 Numerical properties

4.1 Preliminary

In this section, we illustrate the finite sample properties of the proposed test T_n by simulation. Also included is the test T_n^* based on the pre-transformed data as stated in Remark 1 in Section 2.1. We always use the quadratic spectral kernel $\mathcal{K}_{QS}(x)$ specified in (10). In addition, we always use the data-driven bandwidth $b_n = 1.3221\{\widehat{a}(2)\tilde{n}\}^{1/5}$ suggested in Section 6 of Andrews (1991), where $\widehat{a}(2) = \{\sum_{\ell=1}^{p^2K} 4\widehat{\rho}_\ell^2\widehat{\sigma}_\ell^4(1-\widehat{\rho}_\ell)^{-8}\}\{\sum_{\ell=1}^{p^2K}\widehat{\sigma}_\ell^4(1-\widehat{\rho}_\ell)^{-4}\}^{-1}$ with $\widehat{\rho}_\ell$ and $\widehat{\sigma}_\ell^2$ being, respectively, the estimated autoregressive coefficient and innovation variance from fitting an AR(1) model to time series $\{f_{\ell,t}\}_{t=1}^{\tilde{n}}$, where $f_{\ell,t}$ is the ℓ-th component of f_t defined in (8). We draw $G_1,\ldots,G_B$ independently from $N(0,\widehat{\Xi}_n)$ with $B = 2000$ based on (11) and take the $\lfloor B\alpha \rfloor$-th largest value among $|G_1|_\infty, \ldots, |G_B|_\infty$ as the critical value $\widehat{\mathrm{cv}}_\alpha$. We set the nominal significance level at $\alpha = 0.05$, $n = 300$, $p = 3, 15, 50, 150$, and $K = 2, 4, 6, 8, 10$. For each setting, we replicate the experiment 500 times.

We compare the new tests T_n and T_n^* with three multivariate portmanteau tests with test statistics: $Q_1 = n\sum_{k=1}^{K} \mathrm{tr}\{\widehat{\Gamma}(k)^{\mathrm{T}}\widehat{\Gamma}(k)\}$ (Box & Pierce, 1970), $Q_2 = n^2\sum_{k=1}^{K} \mathrm{tr}\{\widehat{\Gamma}(k)^{\mathrm{T}}\widehat{\Gamma}(k)\}/(n-k)$ (Hosking, 1980), and $Q_3 = n\sum_{k=1}^{K} \mathrm{tr}\{\widehat{\Gamma}(k)^{\mathrm{T}}\widehat{\Gamma}(k)\} + p^2K(K+1)/(2n)$ (Li & Mcleod, 1981), where $\widehat{\Gamma}(k)$ is the sample correlation matrix (1). Also, we compare T_n and T_n^* with the Lagrange multiplier test (Lütkepohl, 2005), as well as a likelihood ratio test proposed by Tiao & Box (1981). The test of Tiao & Box (1981) is designed for testing for a vector autoregressive model of order r against that of order $r + 1$ and is therefore applicable for testing (3) with $r = 0$. In particular, different from all the other tests included in the comparison, the test of Tiao & Box (1981) does not involve the lag parameter K. For those tests relying on the asymptotic χ^2-approximation, it is known that the χ^2-approximation is poor when the degree of freedom is large. In our simulation, we perform those tests based on the normal approximation instead when $p > 10$. Further discussions on those tests are referred to Section 3.1 of Li (2004) and Section 4.4 of Lütkepohl (2005). The new tests T_n and T_n^*, together with the aforementioned other tests, have been implemented in an `R` package `HDtest` currently available online at `CRAN`.

4.2 Empirical sizes

To examine the approximations for significance levels of the tests, we generate data from the white noise model $\varepsilon_t = Az_t$, where $\{z_t\}$ is a $p \times 1$ white noise. We consider three different loading matrices for A as following.

Model 1: Let $S = (s_{k\ell})_{1\le k,\ell\le p}$ for $s_{k\ell} = 0.995^{|k-\ell|}$, then let $A = S^{1/2}$.

Model 2: Let $r = \lceil p/2.5 \rceil$, $S = (s_{k\ell})_{1\le k,\ell\le p}$ where $s_{kk} = 1$, $s_{k\ell} = 0.8$ for $r(q-1)+1 \le k \ne \ell \le rq$ for $q = 1, \ldots, \lfloor p/r \rfloor$, and $s_{k\ell} = 0$ otherwise. Let $A = S^{1/2}$ which is a block diagonal matrix.

Model 3: Let $A = (a_{k\ell})_{1\le k,\ell\le p}$ with $a_{k\ell}$'s being independently generated from $U(-1, 1)$.

We consider the two types of white noise: (i) z_t, $t \ge 1$, are independent and $N(0, I_p)$, and (ii) z_t consists of p independent autoregressive conditionally heteroscedastic processes, i.e. each component process is of the form $u_t = \sigma_t e_t$, where e_t are independent and $N(0, 1)$, and $\sigma_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2$ with γ_0 and γ_1 generated from, respectively, $U(0.25, 0.5)$ and $U(0, 0.5)$ independently for different component processes. Experiments with more complex white noise processes are reported in the Supplementary Material.

Tables 1–2 report the empirical sizes of tests T_n and T_n^*, along with those of the three portmanteau tests, the Lagrange multiplier test, and the test of Tiao & Box (1981). As Tiao & Box' test does not involve the lag parameter K, we only report its empirical size once for each p in the tables. Also the Lagrange multiplier test is only applicable when $pK < n$, as the testing statistic is calculated from a multivariate regression.

Tables 1–2 indicate that T_n and T_n^* perform about the same as the other five tests when the dimension p is small, such as $p = 3$. The portmanteau, Lagrange multiplier and Tiao & Box's tests, however, fail badly to attain the nominal significance level as the dimension p increases, as the empirical sizes severely underestimate the nominal level when, for example, $p = 50$. In fact the empirical sizes for the portmanteau tests and Tiao & Box's test are almost 0 under all the settings with $p = 150$, while the Lagrange multiplier test, not available when $p = 150$, deviates quickly from the nominal level when pK is close to n. In contrast, the new test T_n performs much better, though still underestimates the nominal level when p is relatively large, particularly for Model 3. Noticeably, T_n^*, the procedure combining the new test with the time series principal

Table 1: The empirical sizes (%) of the tests T_n, T_n^*, Q_1, Q_2, Q_3, Lagrange multiplier test (LM) and Tiao & Box' test (TB) for testing white noise $\varepsilon_t = Az_t$ at the 5% nominal level, where z_t, $t \geq 1$, are independent and $N(0, I_p)$.

		Model 1							Model 2							Model 3						
p	K	T_n	T_n^*	Q_1	Q_2	Q_3	LM	TB	T_n	T_n^*	Q_1	Q_2	Q_3	LM	TB	T_n	T_n^*	Q_1	Q_2	Q_3	LM	TB
3	2	5.2	5.8	5.2	5.6	5.6	5.2	5.2	3.2	6.6	3.8	3.8	3.8	3.8	4.8	4.0	6.4	4.0	4.0	4.0	5.2	3.8
	4	4.6	7.4	3.6	4.4	4.2	4.4		4.0	7.4	3.2	3.4	3.4	3.6		3.8	5.4	4.8	5.0	5.0	5.4	
	6	5.6	8.6	4.4	5.2	5.0	5.4		2.8	7.2	3.2	3.6	3.4	3.0		4.0	5.4	6.0	6.4	6.2	5.2	
	8	4.4	8.4	3.6	5.0	4.4	3.0		3.8	6.2	2.6	3.0	2.8	3.2		3.8	6.4	5.0	6.8	6.2	4.6	
	10	4.2	7.8	3.6	4.4	4.2	4.0		3.0	6.0	1.4	3.0	2.4	2.4		3.6	5.6	5.4	7.4	7.2	4.6	
15	2	3.8	5.2	4.2	4.8	4.8	5.0	4.8	2.8	4.4	4.2	5.0	5.0	5.4	7.6	3.0	3.8	3.4	4.0	4.0	3.8	5.2
	4	4.0	5.4	2.8	5.0	5.0	3.8		2.6	4.2	2.8	4.6	4.6	3.6		2.4	4.8	2.2	3.0	3.0	3.2	
	6	3.6	6.2	3.2	5.2	5.2	3.8		2.2	5.2	3.4	5.2	5.0	3.4		2.0	5.8	1.6	3.2	3.2	2.4	
	8	3.6	6.6	2.0	5.2	5.0	1.0		2.4	6.0	0.8	5.0	4.6	2.0		2.2	7.2	0.8	2.8	2.8	1.4	
	10	3.0	7.0	1.4	5.6	5.2	0.4		2.2	6.2	1.0	5.0	4.8	1.6		2.6	6.6	1.0	4.0	3.8	0.8	
50	2	2.4	4.0	1.6	2.4	2.4	1.2	8.8	3.0	4.2	1.4	2.4	2.4	1.4	7.8	1.8	4.8	1.6	2.8	2.8	1.2	7.8
	4	4.0	4.4	0.6	3.0	2.8	0.0		2.6	4.6	0.6	2.2	2.2	0.0		2.2	5.2	0.8	2.6	2.6	0.0	
	6	3.6	4.8	0.0	3.8	3.6			1.8	5.2	0.2	2.8	2.6			2.0	6.4	0.2	2.2	2.2		
	8	3.8	4.4	0.0	3.8	3.6			2.0	5.4	0.0	2.2	2.2			1.6	7.2	0.0	2.8	2.4		
	10	4.6	4.8	0.0	3.0	3.0			1.4	5.4	0.0	2.8	2.2			1.4	6.2	0.0	2.0	1.8		
150	2	3.0	4.4	0.0	0.0	0.0		0.0	3.0	3.8	0.0	0.2	0.0		0.0	1.4	3.6	0.0	0.2	0.2		0.0
	4	1.4	4.2	0.0	0.0	0.0			2.0	4.2	0.0	0.0	0.0			1.4	3.4	0.0	0.0	0.0		
	6	1.8	2.8	0.0	0.0	0.0			2.4	3.2	0.0	0.0	0.0			1.2	4.2	0.0	0.0	0.0		
	8	2.2	3.8	0.0	0.0	0.0			1.8	3.2	0.0	0.2	0.2			0.6	4.8	0.0	0.0	0.0		
	10	3.2	4.6	0.0	0.2	0.0			1.6	4.2	0.0	0.0	0.0			0.4	5.4	0.0	0.0	0.0		

Table 2: The empirical sizes (%) of the tests T_n, T_n^*, Q_1, Q_2, Q_3, Lagrange multiplier test (LM) and Tiao & Box' test (TB) for testing white noise $\varepsilon_t = Az_t$ at the 5% nominal level, where z_t consists of p independent autoregressive conditionally heteroscedastic processes.

		Model 1							Model 2							Model 3						
p	K	T_n	T_n^*	Q_1	Q_2	Q_3	LM	TB	T_n	T_n^*	Q_1	Q_2	Q_3	LM	TB	T_n	T_n^*	Q_1	Q_2	Q_3	LM	TB
3	2	4.0	5.4	3.0	3.0	4.2	4.0	4.4	6.4	8.6	7.0	7.0	9.4	8.4	6.4	3.8	7.2	5.4	5.6	8	6.6	7.2
	4	4.4	7.6	4.4	4.6	5.2	5.0		5.6	8.2	5.2	6.0	7.4	6.2		5.0	7.8	5.4	6.0	8.2	7.2	
	6	3.0	6.6	4.8	5.6	6.4	4.6		5.0	6.6	5.8	6.2	7.2	4.8		4.8	6.8	4.0	4.4	6.6	4.8	
	8	3.2	6.4	4.4	5.8	7.4	5.6		4.6	6.8	5.8	7.0	7.8	6.4		4.8	6.6	4.2	5.0	5.4	3.4	
	10	3.6	6.0	5.0	5.8	7.8	5.6		4.4	6.2	5.4	6.4	7.2	4.2		4.8	5.8	4.6	5.0	5.4	4.0	
15	2	4.2	5.6	4.0	3.8	4.6	5.0	7.0	4.8	5.0	3.2	3.2	3.6	3.4	5.6	2.4	4.8	5.0	5.4	6.6	5.2	6.8
	4	4.0	5.8	5.0	5.0	5.2	4.0		3.8	5.0	4.0	4.0	4.0	2.2		2.6	7.0	2.8	2.8	2.8	3.2	
	6	4.2	5.0	4.2	4.0	4.2	3.0		2.8	6.2	5.6	5.0	5.6	2.0		2.4	6.8	3.0	3.0	3.2	2.6	
	8	3.8	6.0	4.8	4.8	4.8	1.4		2.2	5.8	4.8	4.8	4.8	2.0		2.8	6.2	1.8	2.8	3.8	2.2	
	10	4.6	4.8	5.6	5.4	5.4	1.2		3.4	5.4	4.0	3.8	4.4	1.4		2.4	8.2	0.8	4.2	4.4	0.8	
50	2	4.4	4.2	2.2	2.2	2.2	0.6	6.2	3.2	4.0	1.4	2.4	2.8	1.0	8.2	2.2	3.2	2.2	2.0	2.0	0.4	7.6
	4	3.8	4.6	2.8	2.8	3.0	0.0		2.2	5.4	2.0	2.0	2.0	0.0		2.2	4.0	2.0	1.8	1.8	0.0	
	6	4.6	6.2	1.4	1.4	1.8			3.6	5.2	1.8	2.8	1.8			1.2	4.8	2.0	2.0	2.0		
	8	3.6	7.2	3.0	3.0	3.0			2.4	6.0	1.4	1.2	1.6			1.6	5.8	0.0	0.0	1.6		
	10	3.6	5.8	3.2	3.2	2.8			2.2	5.6	1.8	1.8	1.8			1.4	6.6	0.0	0.0	1.6		
150	2	4.8	3.6	0.0	0.0	0.0		0.0	1.2	2.8	0.0	0.0	0.0		0.2	1.6	2.8	0.0	0.0	0.0		0.0
	4	2.8	3.2	0.0	0.0	0.0			2.2	3.4	0.0	0.0	0.0			1.0	3.2	0.2	0.0	0.2		
	6	2.0	4.2	0.0	0.0	0.0			2.6	3.6	0.0	0.0	0.0			1.0	3.2	0.0	0.0	0.0		
	8	1.6	5.0	0.0	0.0	0.0			1.6	4.4	0.0	0.0	0.0			0.8	3.4	0.2	0.0	0.2		
	10	2.6	4.8	0.2	0.2	0.2			2.0	5.0	0.4	0.2	0.4			1.0	4.6	0.0	0.0	0.0		

component analysis, produces empirical sizes much closer to the nominal level than all other tests across almost all the settings with $p = 50$ and 150.

We also observe that both the portmanteau tests Q_2 and Q_3 perform similarly, and outperform Q_1 when p is large. This is in line with the fact that the asymptotic approximations for Q_2 and Q_3 are more accurate than that for Q_1. In addition, Tables 1–2, as well as the results in the Supplementary Material, indicate that the proposed tests are more robust with respect to the choice of the prescribed lag parameter K. The test T_n, and the portmanteau tests, perform better under Models 1 and 2 than under Model 3 when p is large. As the entries in the loading matrix A in Model 3 can be both positive and negative, the signals z_t may be weakened due to possible cancellations. Nevertheless, with the aid of time series principal component analysis, T_n^* perform reasonably well across all the settings including Model 3.

In summary, the proposed tests, especially T_n^*, attain the nominal level much more accurate than existing tests when p is large. For small p, all the tests are about equally accurate in attaining the nominal significance level.

4.3 Empirical power

To conduct the power comparison among the different tests, we consider two non-white noise models. Put $k_0 = \min(\lceil p/5 \rceil, 12)$.

Model 4: $\varepsilon_t = A\varepsilon_{t-1} + e_t$, where e_t, $t \geq 1$, are independent, each e_t consists of p independent t_8 random variables, and the coefficient matrix $A \equiv (a_{k\ell})$ is generated as follows: $a_{k\ell} \sim U(-0.25, 0.25)$ independently for $1 \leq k, \ell \leq k_0$, and $a_{k\ell} = 0$ otherwise. Thus only the first k_0 components of ε_t are not white noise.

Model 5: $\varepsilon_t = Az_t$, where $z_t = (z_{1,t}, \ldots, z_{p,t})^{\mathrm{T}}$. For $1 \leq k \leq k_0$, $(z_{k,1}, \ldots, z_{k,n})^{\mathrm{T}} \sim N(0, \Sigma)$, where Σ is an $n \times n$ matrix with 1 as the main diagonal elements, $0.5|i-j|^{-0.6}$ as the (i,j)-th element for $1 \leq |i-j| \leq 7$, and 0 as all the other elements. For $k > k_0$, $z_{k,1}, \ldots, z_{k,n}$ are independent and t_8 random variables. The coefficient matrix $A \equiv (a_{k\ell})$ is generated as follows: $a_{k\ell} \sim U(-1, 1)$ with probability $1/3$ and $a_{k\ell} = 0$ with probability $2/3$ independently for $1 \leq k \neq \ell \leq p$, and $a_{kk} = 0.8$ for $1 \leq k \leq p$.

Figs. 1–2 display the empirical power curves of the seven tests under consideration against the lag parameter K. As Tiao & Box' test involves no lag parameter K, its power curves are flat. Also note that the Lagrange multiplier test is only available for $p = 3, 15$ and $p = 50$ while $K = 2, 4, 6$. When $p = 150$, the proposed tests, especially T_n^*, maintain substantial power while all the other five tests are powerless. Under Model 4, where the autocorrelation decays relatively fast, the proposed tests T_n and T_n^* are substantially more powerful than the portmanteau tests and the Lagrange multiplier test even when p is small. In addition, Fig. 1 and the results in the Supplementary Materials also indicate that the existing tests compromise more in power than the new tests when the loading matrix A is relatively sparse. When the autocorrelation is strong, as in Model 5, the portmanteau tests and the Lagrange multiplier test perform well when p is small, e.g., $p = 3$; see Fig. 2. Finally, as expected, T_n^* is more powerful than T_n when p is large, and the improvement is substantial when, for example, $p = 150$. Overall, our proposed tests T_n and T_n^* are more powerful than the traditional tests when the dimension p is large or moderately large. This pattern is also observed in a more extensive comparison reported in the Supplementary Material.

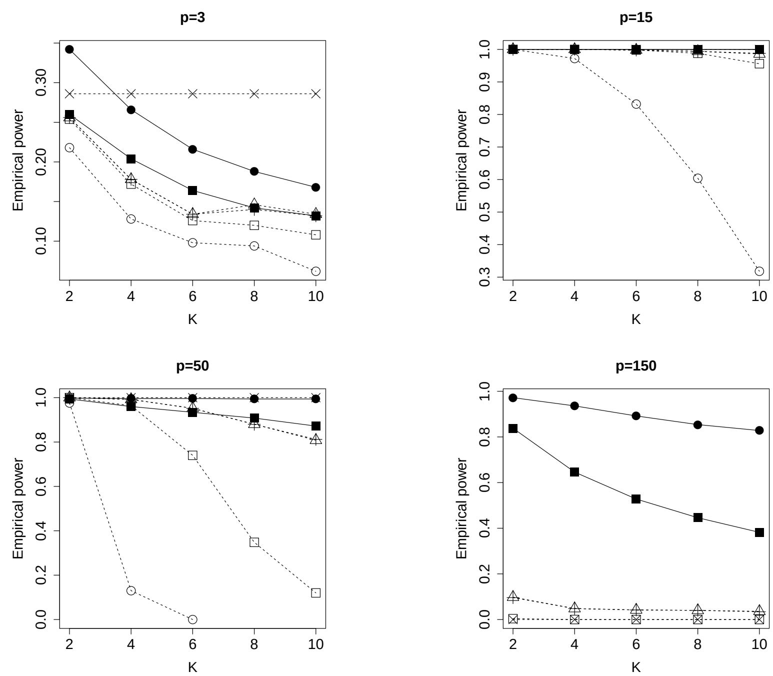

Figure 1: Plots of empirical power against lag K for the new tests T_n (solid and ■ lines) and T_n^* (solid and • lines), the portmanteau tests Q_1 (dashed and △ lines), Q_2 (dashed and + lines) and Q_3 (dashed and □ lines), the Lagrange multiplier test (dashed and ○ lines), and Tiao and Box' test (dashed and ×). The data are generated from Model 4 with sample size $n = 300$. The nominal level is $\alpha = 5\%$.

5 Applications in model diagnosis

Let $\{y_t\}$ and $\{u_t\}$ be observable $p \times 1$ and $q \times 1$ time series, respectively. Let

$$y_t = g(u_t; \theta_0) + \varepsilon_t, \tag{12}$$

where $g(\cdot;\cdot)$ is a known link function, and $\theta_0 \in \Theta$ is an unknown $s \times 1$ parameter vector. One of the most frequently used procedures for model diagnosis is to test if the error process $\{\varepsilon_t\}$ is white noise. Since $\{\varepsilon_t\}$ is unknown, the diagnostic test is instead applied to the residuals

$$\widehat{\varepsilon}_t \equiv y_t - g(u_t; \widehat{\theta}), \qquad t = 1, \cdots, n, \tag{13}$$

where $\widehat{\theta}$ is an appropriate estimator for θ_0.

Model (12) encompasses a large number of frequently used models, including both linear and nonlinear vector autoregressive models with or without exogenous variables. It also includes linear invertible and identifiable vector autoregressive and moving average models by allowing $q = \infty$ and $s = \infty$. Let $g(\cdot;\cdot) = \{g_1(\cdot;\cdot), \ldots, g_p(\cdot;\cdot)\}^{\mathrm{T}}$, and $\mathcal{U}$ be the domain of u_t. Let the true value θ_0 of model (12) be an inner point of Θ. We assume that the link function $g(\cdot;\cdot)$ satisfies the following condition.

Condition 5. *Denote by Θ_0 a small neighborhood of θ_0. For some given metric $|\cdot|_*$ defined on Θ, it holds that $|g_i(u;\theta^*) - g_i(u;\theta^{**})| \leq M_i(u)|\theta^* - \theta^{**}|_* + R_i(u;\theta^*,\theta^{**})$ for any $\theta^*, \theta^{**} \in \Theta_0$, $u \in \mathcal{U}$ and $i = 1, \ldots, p$, where $\{M_i(\cdot)\}_{i=1}^p$ and $\{R_i(\cdot;\cdot,\cdot)\}_{i=1}^p$ are two sets of non-negative functions that*

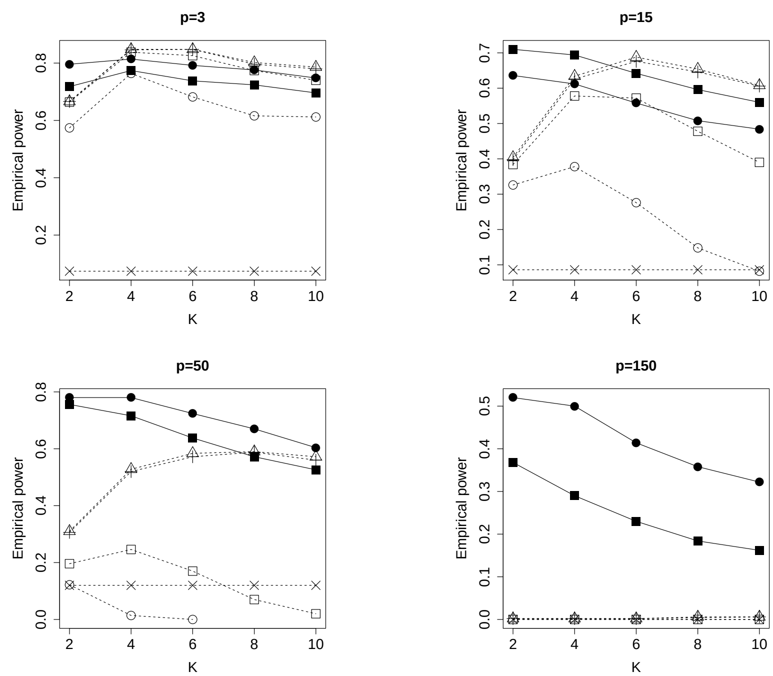

Figure 2: Plots of empirical power against lag K for the new tests T_n (solid and ■ lines) and T_n^* (solid and • lines), the portmanteau tests Q_1 (dashed and △ lines), Q_2 (dashed and + lines) and Q_3 (dashed and □ lines), the Lagrange multiplier test (dashed and ○ lines), and Tiao and Box' test (dashed and ×). The data are generated from Model 5 with sample size $n = 300$. The nominal level is $\alpha = 5\%$.

satisfy $\sup_{1\le i\le p} n^{-1}\sum_{t=1}^n M_i^2(u_t) = O_p(\varphi_{1,n})$ *and* $\sup_{1\le i\le p}\sup_{\theta^*,\theta^{**}\in\Theta_0} n^{-1}\sum_{t=1}^n R_i^2(u_t;\theta^*,\theta^{**}) = O_p(\varphi_{2,n})$ *for some* $\varphi_{1,n}>0$ *(which may diverge) and* $\varphi_{2,n}\to 0$ *as* $n\to\infty$.

In fact, the first part of Condition 5 can be replaced by the Lipschitz continuity $|g_i(u;\theta^*)-g_i(u;\theta^{**})| \le M_i(u)|\theta^*-\theta^{**}|_*^\phi + R_i(u;\theta^*,\theta^{**})$ for some $\phi\in(0,1]$. Since the proofs for Theorem 3 under these two types of continuity are identical, we only state the result for $\phi=1$ explicitly. The remainder term $R_i(\cdot;\cdot,\cdot)$ is employed to accommodate the models with infinite-dimensional parameter θ_0. When θ_0 has finite number of components, we can let $|\cdot|_*$ be the standard L_2-norm. If the link function $g_i(u;\theta)$ is continuously differentiable with respect to θ, it follows from a Taylor expansion that $|g_i(u;\theta^*)-g_i(u;\theta^{**})| \le |\nabla_\theta g_i(u;\bar\theta)|_2|\theta^*-\theta^{**}|_2$ for some $\bar\theta$ lies between θ^* and θ^{**}. If there exists an envelop function $M_i(\cdot)$ satisfying $\sup_{\theta\in\Theta}|\nabla_\theta g_i(u;\bar\theta)|_2\le M_i(u)$ for any $u\in\mathcal{U}$, the first part of Condition 5 holds with $R_i(u;\theta^*,\theta^{**})\equiv 0$. When θ_0 is an infinite dimensional parameter, we can select $|\cdot|_*$ as the vector L_1-norm. Put $\theta=(\theta_1,\theta_2,\ldots)^{\mathrm{T}}$. If $\partial g_i(u;\theta)/\partial\theta_j$ exists for any $j=1,2,\ldots$, it follows from a Taylor expansion that $g_i(u;\theta^*)-g_i(u;\theta^{**}) = \sum_{j=1}^\infty(\theta_j^*-\theta_j^{**})\partial g_i(u;\bar\theta)/\partial\theta_j$ for some $\bar\theta$ lies between θ^* and θ^{**}. For some given diverging d, letting $M_i(u)=\sup_{1\le j\le d}\sup_{\theta\in\Theta}|\partial g_i(u;\theta)/\partial\theta_j|$ and $R_i(u;\theta^*,\theta^{**}) = |\sum_{j=d+1}^\infty(\theta_j^*-\theta_j^{**})\partial g_i(u;\bar\theta)/\partial\theta_j|$, we have

$$\begin{aligned}|g_i(u;\theta^*)-g_i(u;\theta^{**})| &\le \sup_{1\le j\le d}\left|\frac{\partial g_i(u;\bar\theta)}{\partial\theta_j}\right|\sum_{j=1}^d|\theta_j^*-\theta_j^{**}| + \left|\sum_{j=d+1}^\infty(\theta_j^*-\theta_j^{**})\frac{\partial g_i(u;\bar\theta)}{\partial\theta_j}\right| \\ &\le M_i(u)|\theta^*-\theta^{**}|_1 + R_i(u;\theta^*,\theta^{**}).\end{aligned}$$

Theorem 3. *Let Condition 5 and the conditions of Theorems* 1 *and* 2 *hold. Let* $|\widehat\theta-\theta_0|_* = O_p(\zeta_n)$ *for some* $\zeta_n\to 0$. *Assume that* $\zeta_n^2\varphi_{1,n}\to 0$ *as* $n\to\infty$. *Then Theorems* 1 *and* 2 *still hold if* $\{\varepsilon_1,\cdots,\varepsilon_n\}$ *is replaced by* $\{\widehat\varepsilon_1,\cdots,\widehat\varepsilon_n\}$ *defined in (13).*

Acknowledgement

We are grateful to the Editor, an Associate Editor and two referees for their helpful suggestions. Jinyuan Chang was supported in part by the Fundamental Research Funds for the Central Universities of China, the National Natural Science Foundation of China, and the Center of Statistical Research at Southwestern University of Finance and Economics. Qiwei Yao was supported in part by the U.K. Engineering and Physical Sciences Research Council. Wen Zhou was supported in part by the U.S. National Science Foundation.

Supplementary material

Supplementary material available at Biometrika online contains more extensive comparison by simulation of the seven tests employed in Section 4.

Appendix

A.1 Technical lemmas

Let

$$\widehat\mu = [\mathrm{vec}\{\widehat\Gamma(1)\}^{\mathrm{T}},\ldots,\mathrm{vec}\{\widehat\Gamma(K)\}^{\mathrm{T}}]^{\mathrm{T}}, \quad \widehat W = \mathrm{diag}\{\widehat\Sigma(0)\}^{-1/2}\otimes\mathrm{diag}\{\widehat\Sigma(0)\}^{-1/2}.$$

Then the testing statistic $T_n = n^{1/2}|\widehat{\mu}|_\infty$. It follows from (1) that

$$\widehat{\mu} \equiv (\widehat{\mu}_1, \ldots, \widehat{\mu}_{p^2K})^{\mathrm{T}} = (I_K \otimes \widehat{W})[\mathrm{vec}\{\widehat{\Sigma}(1)\}^{\mathrm{T}}, \ldots, \mathrm{vec}\{\widehat{\Sigma}(K)\}^{\mathrm{T}}]^{\mathrm{T}}.$$

Let

$$\mu \equiv (\mu_1, \ldots, \mu_{p^2K})^{\mathrm{T}} = (I_K \otimes W)[\mathrm{vec}\{\widehat{\Sigma}(1)\}^{\mathrm{T}}, \ldots, \mathrm{vec}\{\widehat{\Sigma}(K)\}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\widehat{Z} = n^{1/2} \max_{1\le \ell \le p^2K} \widehat{\mu}_\ell, \qquad Z = n^{1/2} \max_{1\le \ell \le p^2K} \mu_\ell, \qquad V = \max_{1\le \ell \le p^2K} G_\ell,$$

where $G = (G_1, \ldots, G_{p^2K})^{\mathrm{T}} \sim N(0, \Xi_n)$ with Ξ_n specified in (6). Throughout the Appendix, $C \in (0, \infty)$ denotes a generic constant that does not depend on p and n, and it may be different at different places.

Lemma 1. *Assume that Conditions* 1–3 *hold. Let* γ *satisfy* $\gamma^{-1} = 2r_1^{-1} + r_2^{-1}$, *and* $\log p = o\{n^{\gamma/(2-\gamma)}\}$. *Then* $|\widehat{W} - W|_\infty \le Cn^{-1/2}(\log p)^{1/2}$ *with probability at least* $1 - Cp^{-1}$.

Proof. Put $\mathrm{diag}\{\widehat{\Sigma}(0)\} = \mathrm{diag}(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_p^2)$ and $\mathrm{diag}\{\Sigma(0)\} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. By Condition 1,

$$|\widehat{W} - W|_\infty = \max_{1\le i,j\le p} |\widehat{\sigma}_i^{-1}\widehat{\sigma}_j^{-1} - \sigma_i^{-1}\sigma_j^{-1}| \le \left(\max_{1\le i\le p} |\widehat{\sigma}_i^{-1} - \sigma_i^{-1}|\right)^2 + C\max_{1\le i\le p} |\widehat{\sigma}_i^{-1} - \sigma_i^{-1}|. \tag{14}$$

To bound the term on the right-hand side of (14), we first consider the tail probability of $\max_{1\le i\le p} |\widehat{\sigma}_i - \sigma_i|$. Following the same arguments of Lemma 9 in arXiv:1410.2323, it holds that

$$\mathrm{pr}\left(\max_{1\le i\le p} |\widehat{\sigma}_i^2 - \sigma_i^2| > \varepsilon\right) \le Cpn\exp(-C\varepsilon^\gamma n^\gamma) + Cpn\exp(-C\varepsilon^{\tilde{\gamma}/2}n^{\tilde{\gamma}}) \\ + Cp\exp(-C\varepsilon^2 n) + Cp\exp(-C\varepsilon n)$$

for any $\varepsilon > 0$ such that $n\varepsilon \to \infty$, where $\tilde{\gamma}^{-1} = r_1^{-1} + r_2^{-1}$. Therefore, if $\log p = o\{n^{\gamma/(2-\gamma)}\}$, with probability at least $1 - Cp^{-1}$, $\max_{1\le i\le p} |\widehat{\sigma}_i^2 - \sigma_i^2| \le Cn^{-1/2}(\log p)^{1/2}$. Since $\widehat{\sigma}_i^2 - \sigma_i^2 = (\widehat{\sigma}_i - \sigma_i)^2 + 2\sigma_i(\widehat{\sigma}_i - \sigma_i)$, it holds with probability at least $1 - Cp^{-1}$ that $\max_{1\le i\le p} |\widehat{\sigma}_i - \sigma_i| \le Cn^{-1/2}(\log p)^{1/2}$. Finally, it follows from the identify $\widehat{\sigma}_i^{-1} - \sigma_i^{-1} = -(\widehat{\sigma}_i - \sigma_i)\widehat{\sigma}_i^{-1}\sigma_i^{-1}$ that $\max_{1\le i\le p} |\widehat{\sigma}_i^{-1} - \sigma_i^{-1}| \le Cn^{-1/2}(\log p)^{1/2}$ holds with probability at least $1 - Cp^{-1}$. Now the lemma follows from (14) immediately. □

Lemma 2. *Assume that Conditions* 1–3 *hold. Let* $\gamma^{-1} = 2r_1^{-1} + r_2^{-1}$ *and* $\tilde{\gamma}^{-1} = r_1^{-1} + r_2^{-1}$. *Then*

$$\mathrm{pr}\left[\max_{1\le k\le K} |\mathrm{vec}\{\widehat{\Sigma}(k)\} - \mathrm{vec}\{\Sigma(k)\}|_\infty > s\right] \le Cp^2n\exp(-Cs^\gamma n^\gamma) + Cp^2n\exp(-Cs^{\tilde{\gamma}/2}n^{\tilde{\gamma}}) \\ + Cp^2\exp(-Cs^2n) + Cp^2\exp(-Csn)$$

for any $s > 0$ *and* $ns \to \infty$.

Proof. Notice that $|\mathrm{vec}\{\widehat{\Sigma}(k)\} - \mathrm{vec}\{\Sigma(k)\}|_\infty = \max_{1\le i,j\le p} |\widehat{\sigma}_{i,j}(k) - \sigma_{i,j}(k)|$. For given $k = 1, \ldots, K$, Lemma 9 in arXiv:1410.2323 implies that

$$\mathrm{pr}\big[|\mathrm{vec}\{\widehat{\Sigma}(k)\} - \mathrm{vec}\{\Sigma(k)\}|_\infty > s\big] \le Cp^2n\exp(-Cs^\gamma n^\gamma) + Cp^2n\exp(-Cs^{\tilde{\gamma}/2}n^{\tilde{\gamma}}) \\ + Cp^2\exp(-Cs^2n) + Cp^2\exp(-Csn)$$

for any $s > 0$ and $ns \to \infty$. Consequently, the lemma follows directly from the Bonferroni inequality. □

Lemma 3. *Assume that Conditions* 1–3 *hold. Let* $\gamma^{-1} = 2r_1^{-1} + r_2^{-1}$ *and* $\log p = o\{n^{\gamma/(2-\gamma)}\}$. *Then it holds under null hypothesis* H_0 *that* $|\widehat{Z} - Z| \le Cn^{-1/2}\log p$ *with probability at least* $1 - Cp^{-1}$.

Proof. Note that $|\widehat{Z}-Z| \le |\widehat{W}-W|_\infty \max_{1\le k\le K} n^{1/2}|\mathrm{vec}\{\widehat{\Sigma}(k)\}|_\infty$. By Lemma A2, we have $\max_{1\le k\le K} |\mathrm{vec}\{\widehat{\Sigma}(k)\}|_\infty \le Cn^{-1/2}(\log p)^{1/2}$ with probability at least $1 - Cp^{-1}$ under H_0. This, together with Lemma A1, implies the required assertion. □

Lemma 4. *Assume that Conditions* 1–4 *hold. Let* $\log p \le Cn^\delta$ *for some* $\delta > 0$. *Then it holds under* H_0 *that* $\sup_{s\in\mathbb{R}} |\mathrm{pr}(Z \le s) - \mathrm{pr}(V \le s)| = o(1)$.

Proof. It follow from (2) that $\mu = n^{-1}\sum_{t=1}^{\tilde{n}} u_t + R_n$, where $\tilde{n} = n - K$, each element of u_t has the form $x_{i,t+k}x_{j,t}/(\sigma_i\sigma_j)$, and R_n is the remainder term. Let $\widetilde{\beta}_k$ $(k \ge 1)$ be the β-mixing coefficients generated by the process $\{u_t\}$. Obviously, it holds that $\widetilde{\beta}_k \le \beta_{(k-K)^+}$. Define $\bar{u} = \tilde{n}^{-1}\sum_{t=1}^{\tilde{n}} u_t \equiv (\bar{u}_1, \ldots, \bar{u}_{p^2K})^{\mathrm{T}}$ and $\widetilde{Z} = \tilde{n}^{1/2}\max_{1\le\ell\le p^2K} \bar{u}_\ell$. In addition, let $d_n = \sup_{s\in\mathbb{R}} |\mathrm{pr}(Z \le s) - \mathrm{pr}(V \le s)|$ and $\widetilde{d}_n = \sup_{s\in\mathbb{R}} |\mathrm{pr}(\widetilde{Z} \le s) - \mathrm{pr}(V \le s)|$. We proceed the proof for $d_n = o(1)$ in two steps: (i) to show $d_n \le \widetilde{d}_n + o(1)$, and (ii) to prove $\widetilde{d}_n = o(1)$.

To prove (i), note that for any $s \in \mathbb{R}$ and $\varepsilon > 0$,

$$\begin{aligned}\mathrm{pr}(Z \le s) - \mathrm{pr}(V \le s) &\le \mathrm{pr}(\widetilde{Z} \le s+\varepsilon) - \mathrm{pr}(V \le s+\varepsilon) + \mathrm{pr}(|Z-\widetilde{Z}| > \varepsilon) + \mathrm{pr}(s < V \le s+\varepsilon)\\ &\le \widetilde{d}_n + \mathrm{pr}(|Z-\widetilde{Z}| > \varepsilon) + \mathrm{pr}(s < V \le s+\varepsilon).\end{aligned}$$

Similarly, we can obtain the reverse inequality. Therefore,

$$d_n \le \widetilde{d}_n + \mathrm{pr}(|Z-\widetilde{Z}| > \varepsilon) + \sup_{s\in\mathbb{R}} \mathrm{pr}(|V-s| \le \varepsilon). \tag{15}$$

By the anti-concentration inequality of Gaussian random variables, $\sup_{s\in\mathbb{R}} \mathrm{pr}(|V-s| \le \varepsilon) \le C\varepsilon\{\log(p/\varepsilon)\}^{1/2}$. It follows from the triangle inequality and Condition 1 that

$$\begin{aligned}|Z-\widetilde{Z}| &\le (n^{1/2} - \tilde{n}^{1/2}) \max_{1\le\ell\le p^2K} |\mu_\ell| + \tilde{n}^{1/2} \max_{1\le\ell\le p^2K} |\mu_\ell - \bar{u}_\ell|\\ &\le \frac{C}{n^{1/2}} \max_{1\le k\le K} |\mathrm{vec}\{\widehat{\Sigma}(k)\}|_\infty + \frac{C}{n^{1/2}}|\bar{u}|_\infty + n^{1/2}|R_n|_\infty.\end{aligned}$$

Following the arguments of Lemma 9 of arXiv:1410.2323, we can show that under H_0,

$$\begin{aligned}\mathrm{pr}\left(\frac{C}{n^{1/2}}|\bar{u}|_\infty > \frac{\varepsilon}{3}\right) &\le Cp^2n\exp(-C\varepsilon^\gamma n^{3\gamma/2}) + Cp^2n\exp(-C\varepsilon^{\tilde{\gamma}/2}n^{5\tilde{\gamma}/4})\\ &\quad + Cp^2\exp(-C\varepsilon^2n^2) + Cp^2\exp(-C\varepsilon n^{3/2})\,,\end{aligned}$$

provided $n^3\varepsilon^2 \to \infty$. It can also shown in the same manner that under H_0, $\mathrm{pr}(n^{1/2}|R_n|_\infty > \varepsilon/3)$ can be also controlled by the same upper bound specified above. Now by Lemma A2, it holds under H_0 that

$$\begin{aligned}\mathrm{pr}(|Z-\widetilde{Z}| > \varepsilon) &\le Cp^2n\exp(-C\varepsilon^\gamma n^{3\gamma/2}) + Cp^2n\exp(-C\varepsilon^{\tilde{\gamma}/2}n^{5\tilde{\gamma}/4})\\ &\quad + Cp^2\exp(-C\varepsilon^2n^2) + Cp^2\exp(-C\varepsilon n^{3/2}).\end{aligned}$$

Let $\varepsilon = Cn^{-1}(\log p)^{1/2}$. Then (15) implies that $d_n \le \widetilde{d}_n + o(1)$.

The proof of (ii) is the same as that to show $d_1 = o(1)$ in the proof of Theorem 1 of an unpublished technical report of Chang, Qiu, Yao and Zou (arXiv:1603.06663). Therefore, if $\log p \le Cn^\delta$ for some $\delta > 0$, we have $\widetilde{d}_n = o(1)$. This completes the proof of Lemma A4. □

A.2 Proof of Proposition 1

Following the arguments in the proof of Proposition 1 in the supplementary file of an unpublished technical report of Chang, Zhou and Zhou (arXiv:1406.1939), it suffices to show $\sup_{s\in\mathbb{R}} |\mathrm{pr}(\widehat{Z} > s) - \mathrm{pr}(V > s)| = o(1)$, where $\widehat{Z}$ and V are defined in the first paragraph of Appendix. Recall $d_n = \sup_{s\in\mathbb{R}} |\mathrm{pr}(Z \le s) - \mathrm{pr}(V \le s)|$. By the similar arguments of (15), it can be proved that $\sup_{s\in\mathbb{R}} |\mathrm{pr}(\widehat{Z} > s) - \mathrm{pr}(V > s)| \le d_n + \mathrm{pr}(|\widehat{Z} - Z| > \varepsilon) + C\varepsilon\{\log(p/\varepsilon)\}^{1/2}$. Set $\varepsilon = Cn^{-1/2}\log p$, Lemmas A3 and A4 yield that $\sup_{s\in\mathbb{R}} |\mathrm{pr}(\widehat{Z} > s) - \mathrm{pr}(V > s)| = o(1)$. This completes the proof of Theorem 1.

A.3 Proof of Theorem 1

Based on Lemma 4 of arXiv:1603.06663 and Proposition 1, we can proceed the proof in the same manner as the proof for Theorem 2 of arXiv:1603.06663.

A.4 Proof of Theorem 2

Let $\mathcal{X}_n = \{\varepsilon_1, \ldots, \varepsilon_n\}$. Since $G \sim N(0, \widehat{\Xi}_n)$ conditionally on $\mathcal{X}_n$, it holds that

$$E(|G|_\infty \mid \mathcal{X}_n) \le [1 + \{2\log(p^2K)\}^{-1}]\{2\log(p^2K)\}^{1/2} \max_{1\le\ell\le p^2K} \widehat{\Xi}_\ell^{1/2},$$

where $\widehat{\Xi}_1, \ldots, \widehat{\Xi}_{p^2K}$ are the elements in the diagonal of $\widehat{\Xi}_n$. On the other hand, it holds $\mathrm{pr}\{|G|_\infty \ge E(|G|_\infty \mid \mathcal{X}_n) + u \mid \mathcal{X}_n\} \le \exp\{-u^2/(2\max_{1\le\ell\le p^2K} \widehat{\Xi}_\ell)\}$ for any $u > 0$. Let $\Xi_1, \ldots, \Xi_{p^2K}$ be the elements in the main diagonal of Ξ_n. In addition, for any $v > 0$, let $\mathcal{E}_0(v) = \{\max_{1\le\ell\le p^2K} |\widehat{\Xi}_\ell^{1/2}/\Xi_\ell^{1/2} - 1| \le v\}$. Restricted on $\mathcal{E}_0(v)$, it holds that

$$\widehat{\mathrm{cv}}_\alpha \le (1+v)([1 + \{2\log(p^2K)\}^{-1}]\{2\log(p^2K)\}^{1/2} + \{2\log(1/\alpha)\}^{1/2}) \max_{1\le\ell\le p^2K} \Xi_\ell^{1/2}.$$

Let $(i_0, j_0, k_0) = \arg\max_{1\le k\le K} \max_{1\le i,j\le p} |\rho_{i,j}(k)|$. Without loss of generality, we assume $\rho_{i_0,j_0}(k_0) > 0$. Then, restricted on $\mathcal{E}_0(v)$, it holds that

$$T_n \ge n^{1/2}\widehat{\rho}_{i_0,j_0}(k_0) \ge n^{1/2}\widehat{\sigma}_{i_0}^{-1}\widehat{\sigma}_{j_0}^{-1}\{\widehat{\sigma}_{i_0,j_0}(k_0) - \sigma_{i_0,j_0}(k_0)\} + n^{1/2}\rho_{i_0,j_0}(k_0)(1+v)^{-2}.$$

Choose u in such a way that $(1+v)^2[1 + \{\log(p^2K)\}^{-1} + u] = 1 + \varepsilon_n$, for $\varepsilon_n > 0$ satisfying that $\varepsilon_n \to 0$ and $\varepsilon_n(\log p)^{1/2} \to \infty$. Consequently,

$$n^{1/2}\rho_{i_0,j_0}(k_0) \ge (1+v)^2[1 + \{\log(p^2K)\}^{-1} + u]\lambda(p,\alpha) \max_{1\le\ell\le p^2K} \Xi_\ell^{1/2}.$$

Following the same arguments of Lemma A2, we can choose suitable $v \to 0$ such that $\mathrm{pr}\{\mathcal{E}_0(v)^c\} \to 0$. Therefore,

$$\begin{aligned}\mathrm{pr}(T_n > \widehat{\mathrm{cv}}_\alpha) &\ge \mathrm{pr}\bigg(n^{1/2}\widehat{\rho}_{i_0,j_0}(k_0) > [1 + \{\log(p^2K)\}^{-1}]\lambda(p,\alpha) \max_{1\le\ell\le p^2K} \Xi_\ell^{1/2}\bigg)\\ &\ge \mathrm{pr}\bigg[\frac{n^{1/2}\{\widehat{\sigma}_{i_0,j_0}(k_0) - \sigma_{i_0,j_0}(k_0)\}}{\widehat{\sigma}_{i_0}\widehat{\sigma}_{j_0}} > -u\lambda(p,\alpha) \max_{1\le\ell\le p^2K} \Xi_\ell^{1/2},\ \mathcal{E}_0(v) \text{ holds}\bigg]\\ &\ge 1 - \mathrm{pr}\bigg[\frac{n^{1/2}\{\widehat{\sigma}_{i_0,j_0}(k_0) - \sigma_{i_0,j_0}(k_0)\}}{\widehat{\sigma}_{i_0}\widehat{\sigma}_{j_0}} \le -u\lambda(p,\alpha) \max_{1\le\ell\le p^2K} \Xi_\ell^{1/2}\bigg] - \mathrm{pr}\{\mathcal{E}_0(v)^c\}.\end{aligned}$$

Notice that $u \sim \varepsilon_n$. Thus $u\lambda(p,\alpha)\max_{1\le\ell\le p^2K} \Xi_\ell^{1/2} \to \infty$, which implies that $\mathrm{pr}(T_n > \widehat{\mathrm{cv}}_\alpha) \to 1$. This completes the proof of Theorem 2.

A.5 Proof of Theorem 3

Let $\widehat{W}^*$, $\widehat{\Sigma}^*(0)$, $\widehat{J}_n^*$ and $\widehat{\Xi}_n^*$ be, respectively, the analogues of $\widehat{W}$, $\widehat{\Sigma}(0)$, $\widehat{J}_n$ and $\widehat{\Xi}_n$ with ε_t replaced by $\widehat{\varepsilon}_t$. By Lemma 3.1 of Chernozhukov et al. (2013), we only need to show $|\widehat{\Xi}_n^* - \widehat{\Xi}_n|_\infty = o_p(1)$. Recall $\widehat{\Xi}_n = (I_K \otimes \widehat{W})\widehat{J}_n(I_K \otimes \widehat{W})$ and $\widehat{\Xi}_n^* = (I_K \otimes \widehat{W}^*)\widehat{J}_n^*(I_K \otimes \widehat{W}^*)$, it suffices to prove $|\widehat{W}^* - \widehat{W}|_\infty = o_p(1)$ and $|\widehat{J}_n^* - \widehat{J}_n|_\infty = o_p(1)$. Since the proofs for those two assertions are similar, we only present the proof for $|\widehat{W}^* - \widehat{W}|_\infty = o_p(1)$ below. As $\widehat{W} = [\mathrm{diag}\{\widehat{\Sigma}(0)\}]^{-1/2} \otimes [\mathrm{diag}\{\widehat{\Sigma}(0)\}]^{-1/2}$ and $\widehat{W}^* = [\mathrm{diag}\{\widehat{\Sigma}^*(0)\}]^{-1/2} \otimes [\mathrm{diag}\{\widehat{\Sigma}^*(0)\}]^{-1/2}$, it suffices to show $|\widehat{\Sigma}^*(0) - \widehat{\Sigma}(0)|_\infty = o_p(1)$. Put $\widehat{\varepsilon}_t = (\widehat{\varepsilon}_{1,t}, \ldots, \widehat{\varepsilon}_{p,t})^{\mathrm{T}}$ and $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{p,t})^{\mathrm{T}}$. For any i, j, the (i,j)-th element of $\widehat{\Sigma}^*(0) - \widehat{\Sigma}(0)$ is given by $\Delta_{i,j} = n^{-1}\sum_{t=1}^n(\widehat{\varepsilon}_{i,t}\widehat{\varepsilon}_{j,t} - \varepsilon_{i,t}\varepsilon_{j,t})$. Notice that $\widehat{\varepsilon}_{i,t} = y_{i,t} - g_i(u_t; \widehat{\theta})$ and $\varepsilon_{i,t} = y_{i,t} - g_i(u_t; \theta_0)$. It holds that

$$\begin{aligned}\Delta_{i,j} = &\frac{1}{n}\sum_{t=1}^n \{g_i(u_t;\widehat{\theta}) - g_i(u_t;\theta_0)\}\{g_j(u_t;\widehat{\theta}) - g_j(u_t;\theta_0)\} \\ &- \frac{1}{n}\sum_{t=1}^n \{g_i(u_t;\widehat{\theta}) - g_i(u_t;\theta_0)\}\varepsilon_{j,t} - \frac{1}{n}\sum_{t=1}^n \varepsilon_{i,t}\{g_j(u_t;\widehat{\theta}) - g_j(u_t;\theta_0)\}.\end{aligned}$$

It follows from Cauchy–Schwarz inequality that

$$\begin{aligned}\Delta_{i,j}^2 \le\ &3\bigg[\frac{1}{n}\sum_{t=1}^n \{g_i(u_t;\widehat{\theta}) - g_i(u_t;\theta_0)\}^2\bigg]\bigg[\frac{1}{n}\sum_{t=1}^n \{g_j(u_t;\widehat{\theta}) - g_j(u_t;\theta_0)\}^2\bigg] \\ &+ 3\bigg[\frac{1}{n}\sum_{t=1}^n \{g_i(u_t;\widehat{\theta}) - g_i(u_t;\theta_0)\}^2\bigg]\bigg(\frac{1}{n}\sum_{t=1}^n \varepsilon_{j,t}^2\bigg) \\ &+ 3\bigg[\frac{1}{n}\sum_{t=1}^n \{g_j(u_t;\widehat{\theta}) - g_j(u_t;\theta_0)\}^2\bigg]\bigg(\frac{1}{n}\sum_{t=1}^n \varepsilon_{i,t}^2\bigg).\end{aligned} \tag{16}$$

By Condition 5, it holds uniformly for any $i = 1, \ldots, p$ that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n \{g_i(u_t;\widehat{\theta}) - g_i(u_t;\theta_0)\}^2 &\le |\widehat{\theta} - \theta_0|_*^2\bigg\{\frac{2}{n}\sum_{t=1}^n M_i^2(u_t)\bigg\} + \frac{2}{n}\sum_{t=1}^n R_i^2(u_t;\widehat{\theta},\theta_0) \\ &= O_p(\zeta_n^2\varphi_{1,n} + \varphi_{2,n}).\end{aligned}$$

On the other hand, Lemma A2 implies that $\sup_{1\le i\le p} n^{-1}\sum_{t=1}^n \varepsilon_{i,t}^2 = O_p(1)$. This together with (16) imply that $\Delta_{ij}^2 = O_p(\zeta_n^2\varphi_{1,n} + \varphi_{2,n})$ uniformly for any $i, j = 1, \ldots, p$. Thus $|\widehat{\Sigma}^*(0) - \widehat{\Sigma}(0)|_\infty = O_p(\zeta_n\varphi_{1,n}^{1/2} + \varphi_{2,n}^{1/2}) = o_p(1)$. This completes the proof of Theorem 3.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858.

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70**, 191-221.

Basu, S. & Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, **43**, 1535-1567 .

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

Carrasco, M. & Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 17–39.

Chang, J., Guo, B. & Yao, Q. (2015). High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. *Journal of Econometrics*, **189**, 297-312.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41**, 2786–2819.

Den Haan, W. J. & Levin, A. (1997). A Practioner's Guide to Robust Covariance Matrix Estimation, *Handbook of Statistics 15*, Chapter 12, 291–341.

Deo, R. S. (2000). Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *Journal of Econometrics*, **99**, 291–315.

Durlauf, S. N. (1991). Spectral based testing of the martingale hypothesis. *Journal of Econometrics*, **50**, 355–376.

Escanciano, J. C. & Lobato, I. N. (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics*, **151**, 140–149.

Fan, J. & Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

Fan, J., Zhang, C. & Zhang, J. (2001). Generalized likelihood test statistic and Wilks phenomenon. *The Annals of Statistics*, **29**, 153–193.

Fan, J. & Zhang, W. (2004). Generalised likelihood ratio tests for spectral density. *Biometrika*, **91**, 195–209.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting, *Journal of the American Statistical Association*, **100**, 830–840.

Francq, C., Roy, R. & Zakoian, J. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association*, **100**, 532–544.

Guo, S., Wang, Y. & Yao, Q. (2016). High-dimensional and banded autoregrssions. *Biometrika*, in press.

Horowitz, J. L., Lobato, I. N., Nankervis, J. C. & Savin, N. E. (2006). Bootstrapping the Box-Pierce Q test: A robust test of uncorrelatedness. *Journal of Econometrics*, **133**, 841–862.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, **75**, 602–607.

Kiefer, N. M., Vogelsang, T. J. & Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, **68**, 695–714.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data.* Springer, Berlin.

Lam, C. & Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, **40**, 694–726.

Li, W. K. (2004). *Diagnostic Checks in Time Series.* Chapman & Hall/CRC, Boca Raton.

Li, W. K. & McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate time series models. *Journal of the Royal Statistical Society*, **B**, **43**, 231–239.

Lobato I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, **96**, 1066–1076.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer, Berlin.

Romano, J. P. & Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, **91**, 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.

Tiao, G. C. & Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, **76**, 802–816.