

# Testing for high-dimensional white noise using maximum cross-correlations

By JINYUAN CHANG

*School of Statistics, Southwestern University of Finance and Economics, Chengdu,  
Sichuan 611130, China*  
changjinyuan@swufe.edu.cn

QIWEI YAO

*Department of Statistics, London School of Economics and Political Science,  
Houghton Street, London, WC2A 2AE, U.K.*  
q.yao@lse.ac.uk

AND WEN ZHOU

*Department of Statistics, Colorado State University, 208 Statistics Building,  
Fort Collins, Colorado 80523, U.S.A.*  
riczw@stat.colostate.edu

## SUMMARY

We propose a new omnibus test for vector white noise using the maximum absolute auto-correlations and cross-correlations of the component series. Based on an approximation by the  $L_\infty$ -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 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 the Lagrange multiplier test and the multivariate Box–Pierce portmanteau tests, especially when the dimension of the 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 J. Chang, B. Guo and Q. Yao (arXiv:1410.2323). The proposed procedures have been implemented in an R package.

*Some key words:* 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: 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, or 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 § 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  $\chi^2$ -distributed under the null hypothesis, which makes their application extremely easy. See § 3.1 of [Li \(2004\)](#) and § 4.4 of [Lütkepohl \(2005\)](#) for further information on those portmanteau tests.

While portmanteau tests are designed for testing white noise, their asymptotic  $\chi^2$ -distributions are established under the assumption that observations under the null hypothesis are independent and identically distributed. However, empirical evidence, including that in § 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 of works 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-correlations 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 nonzero 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  $|u|_\infty = \max_{1 \leq i \leq s} |u_i|$  denotes the  $L_\infty$ -norm of a vector  $u = (u_1, \dots, u_s)^T$ . 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, with direct applications in finance, economics, and environmental and medical studies. The current literature on high-dimensional time series focuses on estimation and dimension reduction. See, for example, [Basu & Michailidis \(2015\)](#), [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. Model diagnostics has largely been unexplored, as far as we are aware. The test proposed in this paper represents an effort to fill 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 and find the new test attains the nominal significance levels more accurately and is 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 ([Chang et al., arXiv:1410.2323](#)).

Let  $\otimes$  and  $\text{vec}$  denote, respectively, the Kronecker product and the vectorization for matrices, let  $I_s$  be the  $s \times s$  identity matrix, and set  $|A|_\infty = \max_{1 \leq i \leq \ell, 1 \leq j \leq m} |a_{ij}|$  for an  $\ell \times m$  matrix  $A \equiv (a_{ij})$ . Denote by  $\lceil x \rceil$  and  $\lfloor x \rfloor$  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  $\Sigma(k) = \text{cov}(\varepsilon_{t+k}, \varepsilon_t)$  and  $\Gamma(k) = \text{diag}\{\Sigma(0)\}^{-1/2} \Sigma(k) \text{diag}\{\Sigma(0)\}^{-1/2}$ , respectively, the autocovariance and the autocorrelation of  $\varepsilon_t$  at lag  $k$ , where  $\text{diag}(\Sigma)$  denotes the diagonal matrix consisting of the diagonal elements of  $\Sigma$  only. When  $\Sigma(k) \equiv 0$  for all  $k \neq 0$ ,  $\{\varepsilon_t\}$  is white noise.

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

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

be the sample autocorrelation matrix at lag  $k$ , where

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

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  $k \geq 1$  under  $H_0$ , our test statistic  $T_n$  is defined as

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

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

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

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

To determine  $\text{cv}_\alpha$ , 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  $L_\infty$ -norm of a  $N(0, \Xi_n)$  random vector converges to zero, even when  $p$  diverges at an exponential rate of  $n$ , where

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

with

$$\xi_n = n^{1/2} (\text{vec}\{\hat{\Sigma}(1)\}^\top, \dots, \text{vec}\{\hat{\Sigma}(K)\}^\top)^\top, \quad W = \text{diag}\{\Sigma(0)\}^{-1/2} \otimes \text{diag}\{\Sigma(0)\}^{-1/2}.$$

This paves the way to evaluating  $\text{cv}_\alpha$  simply by drawing a bootstrap sample from  $N(0, \hat{\Xi}_n)$ , where  $\hat{\Xi}_n$  is an appropriate estimator for  $\Xi_n$ .

PROPOSITION 1. *Suppose that Conditions 1–4 in §3 hold and  $G \sim N(0, \Xi_n)$ . There exists a positive constant  $\delta_1$  depending only on the constants in Conditions 1–4 such that  $\log p \leq Cn^{\delta_1}$  for some constant  $C > 0$ . Then under  $H_0$ ,*

$$\sup_{s \geq 0} |\text{pr}(T_n > s) - \text{pr}(|G|_\infty > s)| \rightarrow 0, \quad n \rightarrow \infty.$$

Upon replacing  $\Xi_n$  in (6) by  $\hat{\Xi}_n$ , where  $\hat{\Xi}_n$  is defined in §2.2 below, the critical value  $\text{cv}_\alpha$  in (5) can be replaced by  $\hat{\text{cv}}_\alpha$  which is determined by

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

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

*Remark 1.* When  $p$  is large or moderately large, it is advantageous to apply the time series principal component analysis proposed by Chang et al. (arXiv:1410.2323) to the data first. We denote the resulting statistic by  $T_n^*$ . More precisely, we compute an invertible transformation matrix  $Q$  using the R function `segmentTS` in the package `PCA4TS` available at CRAN (R Development Core Team, 2017). Then  $T_n^*$  is defined in the same manner as  $T_n$  in (4) with  $\{\varepsilon_1, \dots, \varepsilon_n\}$  replaced by  $\{\varepsilon_1^*, \dots, \varepsilon_n^*\}$ , 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. 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 is more powerful. See also the simulation results in §4.

### 2.2. Estimation of $\Xi_n$

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

$$f_t = \{\text{vec}(\varepsilon_{t+1}\varepsilon_t^\top), \dots, \text{vec}(\varepsilon_{t+K}\varepsilon_t^\top)\}^\top \quad (t = 1, \dots, \tilde{n}). \tag{8}$$

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

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

where  $\hat{H}(j) = \tilde{n}^{-1} \sum_{t=j+1}^{\tilde{n}} f_t f_{t-j}^\top$  if  $j \geq 0$  and  $\hat{H}(j) = \tilde{n}^{-1} \sum_{t=-j+1}^{\tilde{n}} f_{t+j} f_t^\top$  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, which diverges 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

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

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

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

where  $\hat{W} = \text{diag}\{\hat{\Sigma}(0)\}^{-1/2} \otimes \text{diag}\{\hat{\Sigma}(0)\}^{-1/2}$  for  $\hat{\Sigma}(0)$  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  $G \sim N(0, \hat{\Xi}_n)$ , the standard approach consists of three steps: perform the Cholesky decomposition for the  $p^2 K \times p^2 K$  matrix  $\hat{\Xi}_n = L^\top L$ ; generate  $p^2 K$  independent  $N(0, 1)$  random variables  $z = (z_1, \dots, z_{p^2 K})^\top$ ; perform the transformation  $G = L^\top z$ . Computationally this is an  $(np^4 K^2 + p^6 K^3)$ -hard problem requiring a large storage space for  $\{f_t\}_{t=1}^{\tilde{n}}$  and the matrix  $\hat{\Xi}_n$ . To circumvent the high computing cost with large  $p$  and/or  $K$ , we propose a method below which involves generating random variables from an  $\tilde{n}$ -variate normal distribution instead.

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

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

Thus a random sample from  $N(0, \hat{\Xi}_n)$  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  $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{p,t})^\top$  for each  $t = 1, \dots, n$ . 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  $\text{var}(\varepsilon_{i,t}) \geq C_1$  holds uniformly for any  $i = 1, \dots, p$ .

*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  $\beta$ -mixing in the sense that  $\beta_k \equiv \sup_t E\{\sup_{B \in \mathcal{F}_{t+k}^\infty} |\text{pr}(B | \mathcal{F}_{-\infty}^t) - \text{pr}(B)|\} \rightarrow 0$  as  $k \rightarrow \infty$ , where  $\mathcal{F}_{-\infty}^u$  and  $\mathcal{F}_{u+k}^\infty$  are the  $\sigma$ -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 constants  $C_5 > 0$  and  $\iota > 0$  independent of  $p$  such that

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

Condition 1 ensures that all component series are not degenerate. Condition 2 is a common assumption in the literature on ultrahigh-dimensional data analysis. It ensures exponential-type upper bounds for the tail probabilities of the statistics concerned. The  $\beta$ -mixing assumption in Condition 3 is mild. Causal autoregressive moving average processes with continuous innovation distributions are  $\beta$ -mixing with exponentially decaying  $\beta_k$ . So are stationary Markov chains satisfying certain conditions. See § 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  $\beta$ -mixing with exponentially decaying  $\beta_k$ ; see Proposition 12 of Carrasco & Chen (2002). If we only require  $\sup_t \sup_{1 \leq i \leq p} \text{pr}(|\varepsilon_{i,t}| > x) = O\{x^{-2(\nu+\epsilon)}\}$  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 § 3.2 of Chang et al. (arXiv:1410.2323) for the implementation of Fuk–Nagaev-type inequalities in such a scenario. The  $\beta$ -mixing condition can be replaced by the  $\alpha$ -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 an open problem to establish the relevant properties under  $\alpha$ -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  $\hat{c}_\alpha$  defined in (7) by the normal approximation is asymptotically valid and, furthermore, the proposed test is consistent.

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

$$\text{pr}(T_n > \hat{c}_\alpha) \rightarrow \alpha, \quad n \rightarrow \infty.$$

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

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

for some positive  $\epsilon_n$  satisfying  $\epsilon_n \rightarrow 0$  and  $\epsilon_n^2 \log p \rightarrow \infty$ . Then under  $H_1$ ,

$$\text{pr}(T_n > \hat{c}\hat{v}_\alpha) \rightarrow 1, \quad n \rightarrow \infty.$$

## 4. NUMERICAL PROPERTIES

### 4.1. Preliminaries

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 § 2.1. We always use the quadratic spectral kernel  $\mathcal{K}_{QS}(x)$  specified in (10) and the data-driven bandwidth  $b_n = 1.3221\{\hat{a}(2)\tilde{n}\}^{1/5}$  suggested in § 6 of Andrews (1991), where  $\hat{a}(2) = \{\sum_{\ell=1}^{p^2K} 4\hat{\rho}_\ell^2\hat{\sigma}_\ell^4(1 - \hat{\rho}_\ell)^{-8}\}\{\sum_{\ell=1}^{p^2K} \hat{\sigma}_\ell^4(1 - \hat{\rho}_\ell)^{-4}\}^{-1}$  with  $\hat{\rho}_\ell$  and  $\hat{\sigma}_\ell^2$  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  $\ell$ th component of  $f_t$  defined in (8). We draw  $G_1, \dots, G_B$  independently from  $N(0, \hat{\Sigma}_n)$ , with  $B = 2000$  based on (11), and take the  $\lfloor B\alpha \rfloor$ th largest value among  $|G_1|_\infty, \dots, |G_B|_\infty$  as the critical value  $\hat{c}\hat{v}_\alpha$ . We set the nominal significance level at  $\alpha = 0.05$ , and take  $n = 300$ ,  $p = 3, 15, 50, 150$ , 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  $Q_1 = n \sum_{k=1}^K \text{tr}\{\hat{\Gamma}(k)^\top \hat{\Gamma}(k)\}$  (Box & Pierce, 1970),  $Q_2 = n^2 \sum_{k=1}^K \text{tr}\{\hat{\Gamma}(k)^\top \hat{\Gamma}(k)\}/(n - k)$  (Hosking, 1980), and  $Q_3 = n \sum_{k=1}^K \text{tr}\{\hat{\Gamma}(k)^\top \hat{\Gamma}(k)\} + p^2K(K + 1)/(2n)$  (Li & McLeod, 1981), where  $\hat{\Gamma}(k)$  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 last test is designed to testing for a vector autoregressive model of order  $r$  against that of order  $r + 1$  and is therefore applicable to testing (3) with  $r = 0$ . In particular, unlike all the other tests included in the comparison, it does not involve the lag parameter  $K$ . For those tests relying on the asymptotic  $\chi^2$ -approximation, it is known that the  $\chi^2$ -approximation is poor when the degrees of freedom is large. In our simulation, we perform those tests based on the normal approximation instead when  $p > 10$ . For further discussions on those tests, see § 3.1 of Li (2004) and § 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 (R Development Core Team, 2017) 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 follows.

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

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

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

We consider two types of white noise: (i)  $z_t$ ,  $t \geq 1$ , are independent and  $N(0, I_p)$ , and (ii)  $z_t$  consists of  $p$  independent autoregressive conditionally heteroscedastic processes, i.e., each

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

$p$	$K$	Model 1						Model 2						Model 3								
		$T_n$	$T_n^*$	$Q_1$	$Q_2$	$Q_3$	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		

LM, Lagrange multiplier test; TB, Tiao & Box's test.

component process is of the form  $u_t = \sigma_t e_t$ , where the  $e_t$  are independent and  $N(0, 1)$ , and  $\sigma_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2$  with  $\gamma_0$  and  $\gamma_1$  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 and 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's 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 and 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 it still underestimates the nominal level when  $p$  is relatively large, particularly for Model 3. Noticeably,  $T_n^*$ , the procedure combining the new test with time series principal 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.

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

$p$	$K$	Model 1							Model 2							Model 3						
		$T_n$	$T_n^*$	$Q_1$	$Q_2$	$Q_3$	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						

The portmanteau tests  $Q_2$  and  $Q_3$  perform similarly, and outperform  $Q_1$  when  $p$  is large, 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 and 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,  $T_n^*$  performs reasonably well across all the settings, including Model 3.

In summary, the proposed tests, especially  $T_n^*$ , attain the nominal level much more accurately 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  $\varepsilon_t$  are not white noise.

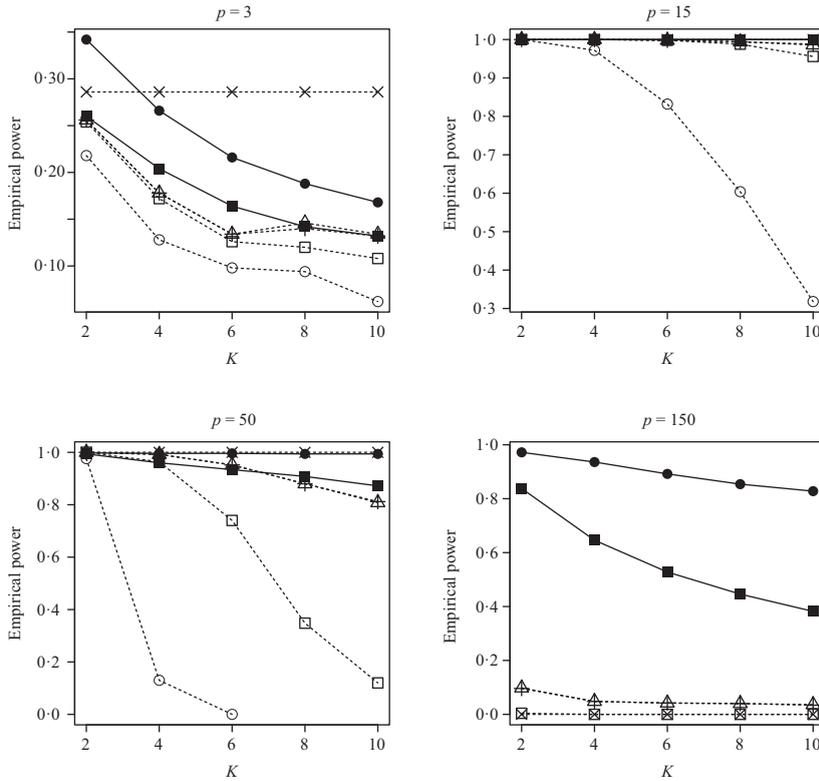


Fig. 1. Plots of empirical power against lag  $K$  for the new tests  $T_n$  (solid and  $\blacksquare$  lines) and  $T_n^*$  (solid and  $\bullet$  lines), the portmanteau tests  $Q_1$  (dashed and  $\triangle$  lines),  $Q_2$  (dashed and  $+$  lines) and  $Q_3$  (dashed and  $\square$  lines), the Lagrange multiplier test (dashed and  $\circ$  lines), and Tiao and Box's test (dashed and  $\times$ ). The data are generated from Model 4 with sample size  $n = 300$ . The nominal level is  $\alpha = 5\%$ .

Model 5:  $\varepsilon_t = Az_t$ , where  $z_t = (z_{1,t}, \dots, z_{p,t})^T$ . For  $1 \leq k \leq k_0$ ,  $(z_{k,1}, \dots, z_{k,n})^T \sim N(0, \Sigma)$ , where  $\Sigma$  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  $k > k_0$ ,  $z_{k,1}, \dots, z_{k,n}$  are independent  $t_8$  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  $1 \leq k \leq p$ .

Figures 1 and 2 display the empirical power curves of the seven tests under consideration against the lag parameter  $K$ . As Tiao & Box's 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  $p = 50$  with  $K = 2, 4, 6$ . 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 Material 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.

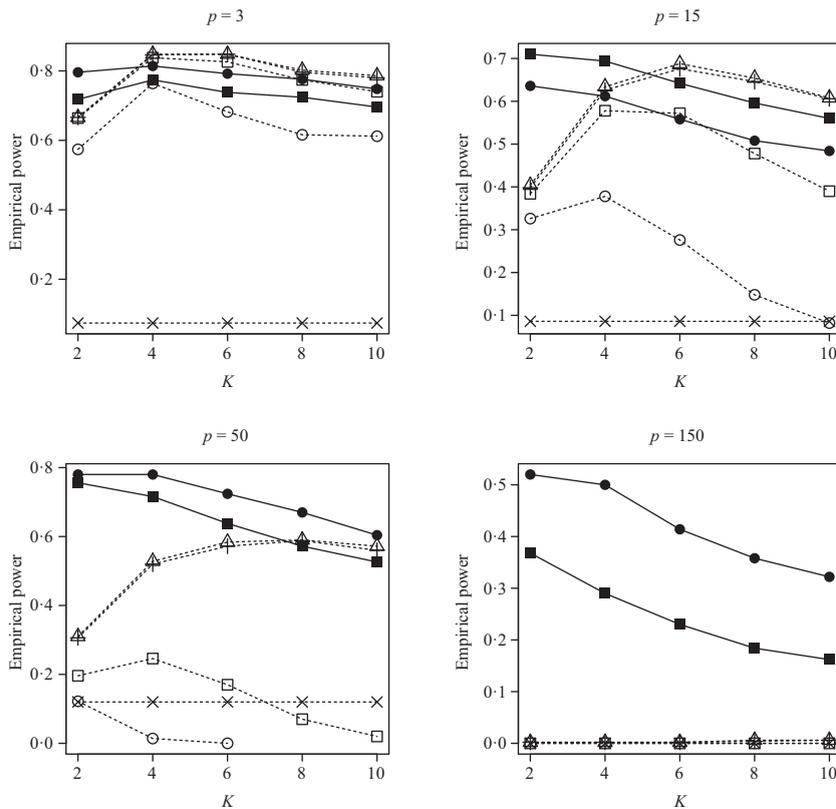


Fig. 2. Plots of empirical power against lag  $K$  for the new tests  $T_n$  (solid and ■ lines) and  $T_n^*$  (solid and ● lines), the portmanteau tests  $Q_1$  (dashed and Δ lines),  $Q_2$  (dashed and + lines) and  $Q_3$  (dashed and □ lines), the Lagrange multiplier test (dashed and ○ lines), and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

This pattern is also observed in a more extensive comparison reported in the Supplementary Material.

### 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

$$\hat{\varepsilon}_t \equiv y_t - g(u_t; \hat{\theta}) \quad (t = 1, \dots, n), \tag{13}$$

where  $\hat{\theta}$  is an appropriate estimator for  $\theta_0$ .

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  $g(\cdot; \cdot) = \{g_1(\cdot; \cdot), \dots, g_p(\cdot; \cdot)\}^T$ , and let  $\mathcal{U}$  be the domain of  $u_t$ . Let

the true value  $\theta_0$  of model (12) be an inner point of  $\Theta$ . We assume that the link function  $g(\cdot; \cdot)$  satisfies the following condition.

*Condition 5.* Denote by  $\Theta_0$  a small neighbourhood of  $\theta_0$ . For some given metric  $|\cdot|_*$  defined on  $\Theta$ , we have  $|g_i(u; \theta^*) - g_i(u; \theta^{**})| \leq M_i(u)|\theta^* - \theta^{**}|_* + R_i(u; \theta^*, \theta^{**})$  for any  $\theta^*, \theta^{**} \in \Theta_0$ ,  $u \in \mathcal{U}$  and  $i = 1, \dots, p$ , where  $\{M_i(\cdot)\}_{i=1}^p$  and  $\{R_i(\cdot; \cdot, \cdot)\}_{i=1}^p$  are two sets of nonnegative functions that satisfy  $\sup_{1 \leq i \leq p} n^{-1} \sum_{t=1}^n M_i^2(u_t) = O_p(\varphi_{1,n})$  and  $\sup_{1 \leq i \leq p} \sup_{\theta^*, \theta^{**} \in \Theta_0} n^{-1} \sum_{t=1}^n R_i^2(u_t; \theta^*, \theta^{**}) = O_p(\varphi_{2,n})$  for some  $\varphi_{1,n} > 0$ , which may diverge, and  $\varphi_{2,n} \rightarrow 0$  as  $n \rightarrow \infty$ .

In fact, the first part of Condition 5 can be replaced by the Lipschitz continuity condition  $|g_i(u; \theta^*) - g_i(u; \theta^{**})| \leq M_i(u)|\theta^* - \theta^{**}|_*^\phi + R_i(u; \theta^*, \theta^{**})$  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 models with an infinite-dimensional parameter  $\theta_0$ . When  $\theta_0$  has a 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  $\theta$ , it follows from a Taylor expansion that  $|g_i(u; \theta^*) - g_i(u; \theta^{**})| \leq |\nabla_\theta g_i(u; \bar{\theta})|_2 |\theta^* - \theta^{**}|_2$  for some  $\bar{\theta}$  between  $\theta^*$  and  $\theta^{**}$ . If there exists an envelope function  $M_i(\cdot)$  satisfying  $\sup_{\theta \in \Theta} |\nabla_\theta g_i(u; \bar{\theta})|_2 \leq M_i(u)$  for any  $u \in \mathcal{U}$ , the first part of Condition 5 holds with  $R_i(u; \theta^*, \theta^{**}) \equiv 0$ . When  $\theta_0$  is an infinite-dimensional parameter, we can select  $|\cdot|_*$  as the vector  $L_1$ -norm. Put  $\theta = (\theta_1, \theta_2, \dots)^\top$ . If  $\partial g_i(u; \theta) / \partial \theta_j$  exists for any  $j = 1, 2, \dots$ , 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  $\bar{\theta}$  between  $\theta^*$  and  $\theta^{**}$ . For some given diverging  $d$ , letting  $M_i(u) = \sup_{1 \leq j \leq d} \sup_{\theta \in \Theta} |\partial g_i(u; \theta) / \partial \theta_j|$  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

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

**THEOREM 3.** *Suppose that Condition 5 and the conditions of Theorems 1 and 2 hold. Let  $|\hat{\theta} - \theta_0|_* = O_p(\zeta_n)$  for some  $\zeta_n \rightarrow 0$ . Assume that  $\zeta_n^2 \varphi_{1,n} \rightarrow 0$  as  $n \rightarrow \infty$ . Then Theorems 1 and 2 still hold if  $\{\varepsilon_1, \dots, \varepsilon_n\}$  is replaced by  $\{\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n\}$  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 and the Joint Lab of Data Science and Business Intelligence 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 comparisons by simulation of the seven tests employed in § 4.

## APPENDIX

## Technical lemmas

Let  $\hat{\mu} = [\text{vec}\{\hat{\Gamma}(1)\}^\top, \dots, \text{vec}\{\hat{\Gamma}(K)\}^\top]^\top$ ,  $\hat{W} = \text{diag}\{\hat{\Sigma}(0)\}^{-1/2} \otimes \text{diag}\{\hat{\Sigma}(0)\}^{-1/2}$ . Then the testing statistic is  $T_n = n^{1/2}|\hat{\mu}|_\infty$ . It follows from (1) that  $\hat{\mu} \equiv (\hat{\mu}_1, \dots, \hat{\mu}_{p^2K})^\top = (I_K \otimes \hat{W})[\text{vec}\{\hat{\Sigma}(1)\}^\top, \dots, \text{vec}\{\hat{\Sigma}(K)\}^\top]^\top$ .

Let  $\mu \equiv (\mu_1, \dots, \mu_{p^2K})^\top = (I_K \otimes W)[\text{vec}\{\hat{\Sigma}(1)\}^\top, \dots, \text{vec}\{\hat{\Sigma}(K)\}^\top]^\top$ ,

$$\hat{Z} = n^{1/2} \max_{1 \leq \ell \leq p^2K} \hat{\mu}_\ell, \quad Z = n^{1/2} \max_{1 \leq \ell \leq p^2K} \mu_\ell, \quad V = \max_{1 \leq \ell \leq p^2K} G_\ell,$$

where  $G = (G_1, \dots, G_{p^2K})^\top \sim N(0, \Xi_n)$  with  $\Xi_n$  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 A1. Assume that Conditions 1–3 hold. Let  $\gamma$  satisfy  $\gamma^{-1} = 2r_1^{-1} + r_2^{-1}$ , and assume  $\log p = o\{n^{\gamma/(2-\gamma)}\}$ . Then  $|\hat{W} - W|_\infty \leq Cn^{-1/2}(\log p)^{1/2}$  with probability at least  $1 - Cp^{-1}$ .

*Proof.* Put  $\text{diag}\{\hat{\Sigma}(0)\} = \text{diag}\{\hat{\sigma}_1^2, \dots, \hat{\sigma}_p^2\}$  and  $\text{diag}\{\Sigma(0)\} = \text{diag}\{\sigma_1^2, \dots, \sigma_p^2\}$ . By Condition 1,

$$|\hat{W} - W|_\infty = \max_{1 \leq i, j \leq p} |\hat{\sigma}_i^{-1} \hat{\sigma}_j^{-1} - \sigma_i^{-1} \sigma_j^{-1}| \leq \left( \max_{1 \leq i \leq p} |\hat{\sigma}_i^{-1} - \sigma_i^{-1}| \right)^2 + C \max_{1 \leq i \leq p} |\hat{\sigma}_i^{-1} - \sigma_i^{-1}|. \quad (\text{A1})$$

To bound the term on the right-hand side of (A1), we first consider the tail probability of  $\max_{1 \leq i \leq p} |\hat{\sigma}_i - \sigma_i|$ . Following the same arguments as for Lemma 9 in Chang et al. (arXiv:1410.2323), we have

$$\begin{aligned} \text{pr} \left( \max_{1 \leq i \leq p} |\hat{\sigma}_i^2 - \sigma_i^2| > \varepsilon \right) &\leq Cp n \exp(-C\varepsilon^\gamma n^\gamma) + Cp n \exp(-C\varepsilon^{\tilde{\gamma}/2} n^{\tilde{\gamma}}) \\ &\quad + Cp \exp(-C\varepsilon^2 n) + Cp \exp(-C\varepsilon n) \end{aligned}$$

for any  $\varepsilon > 0$  such that  $n\varepsilon \rightarrow \infty$ , where  $\tilde{\gamma}^{-1} = r_1^{-1} + r_2^{-1}$ . Therefore, if  $\log p = o\{n^{\gamma/(2-\gamma)}\}$ , with probability at least  $1 - Cp^{-1}$ ,  $\max_{1 \leq i \leq p} |\hat{\sigma}_i^2 - \sigma_i^2| \leq Cn^{-1/2}(\log p)^{1/2}$ . Since  $\hat{\sigma}_i^2 - \sigma_i^2 = (\hat{\sigma}_i - \sigma_i)^2 + 2\sigma_i(\hat{\sigma}_i - \sigma_i)$ , with probability at least  $1 - Cp^{-1}$  we have that  $\max_{1 \leq i \leq p} |\hat{\sigma}_i - \sigma_i| \leq Cn^{-1/2}(\log p)^{1/2}$ . Finally, it follows from the identity  $\hat{\sigma}_i^{-1} - \sigma_i^{-1} = -(\hat{\sigma}_i - \sigma_i)\hat{\sigma}_i^{-1}\sigma_i^{-1}$  that  $\max_{1 \leq i \leq p} |\hat{\sigma}_i^{-1} - \sigma_i^{-1}| \leq Cn^{-1/2}(\log p)^{1/2}$  holds with probability at least  $1 - Cp^{-1}$ . Now the lemma follows from (A1) immediately.  $\square$

LEMMA A2. 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

$$\begin{aligned} \text{pr} \left[ \max_{1 \leq k \leq K} |\text{vec}\{\hat{\Sigma}(k)\} - \text{vec}\{\Sigma(k)\}|_\infty > s \right] &\leq Cp^2 n \exp(-Cs^\gamma n^\gamma) + Cp^2 n \exp(-Cs^{\tilde{\gamma}/2} n^{\tilde{\gamma}}) \\ &\quad + Cp^2 \exp(-Cs^2 n) + Cp^2 \exp(-Csn) \end{aligned}$$

for any  $s > 0$  and  $ns \rightarrow \infty$ .

*Proof.* Notice that  $|\text{vec}\{\hat{\Sigma}(k)\} - \text{vec}\{\Sigma(k)\}|_\infty = \max_{1 \leq i, j \leq p} |\hat{\sigma}_{ij}(k) - \sigma_{ij}(k)|$ . For given  $k = 1, \dots, K$ , Lemma 9 in Chang et al. (arXiv:1410.2323) implies that

$$\begin{aligned} \text{pr} [|\text{vec}\{\hat{\Sigma}(k)\} - \text{vec}\{\Sigma(k)\}|_\infty > s] &\leq Cp^2 n \exp(-Cs^\gamma n^\gamma) + Cp^2 n \exp(-Cs^{\tilde{\gamma}/2} n^{\tilde{\gamma}}) \\ &\quad + Cp^2 \exp(-Cs^2 n) + Cp^2 \exp(-Csn) \end{aligned}$$

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

LEMMA A3. Assume that Conditions 1–3 hold. Let  $\gamma^{-1} = 2r_1^{-1} + r_2^{-1}$  and  $\log p = o\{n^{\gamma/(2-\gamma)}\}$ . Then under the null hypothesis  $H_0$ ,  $|\hat{Z} - Z| \leq Cn^{-1/2} \log p$  holds with probability at least  $1 - Cp^{-1}$ .

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

LEMMA A4. Assume that Conditions 1–4 hold. Let  $\log p \leq Cn^\delta$  for some  $\delta > 0$ . Then under  $H_0$  we have that  $\sup_{s \in \mathbb{R}} |\text{pr}(Z \leq s) - \text{pr}(V \leq s)| = o(1)$ .

*Proof.* It follows 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  $\tilde{\beta}_k$  ( $k \geq 1$ ) be the  $\beta$ -mixing coefficients generated by the process  $\{u_t\}$ . Obviously,  $\tilde{\beta}_k \leq \beta_{(k-K)^+}$ . Define  $\tilde{u} = \tilde{n}^{-1} \sum_{t=1}^{\tilde{n}} u_t \equiv (\tilde{u}_1, \dots, \tilde{u}_{p^2K})^\top$  and  $\tilde{Z} = \tilde{n}^{1/2} \max_{1 \leq \ell \leq p^2K} \tilde{u}_\ell$ . In addition, let  $d_n = \sup_{s \in \mathbb{R}} |\text{pr}(Z \leq s) - \text{pr}(V \leq s)|$  and  $\tilde{d}_n = \sup_{s \in \mathbb{R}} |\text{pr}(\tilde{Z} \leq s) - \text{pr}(V \leq s)|$ . We proceed with the proof of  $d_n = o(1)$  in two steps: (i) to show  $d_n \leq \tilde{d}_n + o(1)$ , and (ii) to prove  $\tilde{d}_n = o(1)$ .

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

$$\begin{aligned} \text{pr}(Z \leq s) - \text{pr}(V \leq s) &\leq \text{pr}(\tilde{Z} \leq s + \varepsilon) - \text{pr}(V \leq s + \varepsilon) + \text{pr}(|Z - \tilde{Z}| > \varepsilon) + \text{pr}(s < V \leq s + \varepsilon) \\ &\leq \tilde{d}_n + \text{pr}(|Z - \tilde{Z}| > \varepsilon) + \text{pr}(s < V \leq s + \varepsilon). \end{aligned}$$

Similarly, we can obtain the reverse inequality. Therefore,

$$d_n \leq \tilde{d}_n + \text{pr}(|Z - \tilde{Z}| > \varepsilon) + \sup_{s \in \mathbb{R}} \text{pr}(|V - s| \leq \varepsilon). \tag{A2}$$

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

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

Following the arguments of Lemma 9 of Chang et al. (arXiv:1410.2323), we can show that under  $H_0$ ,

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

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

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

Let  $\varepsilon = Cn^{-1} (\log p)^{1/2}$ . Then (A2) implies that  $d_n \leq \tilde{d}_n + o(1)$ .

The proof of (ii) is the same as that giving  $d_1 = o(1)$  in the proof of Theorem 1 of Chang et al. (arXiv:1603.06663). Therefore, if  $\log p \leq Cn^\delta$  for some  $\delta > 0$ , we have  $\tilde{d}_n = o(1)$ . This completes the proof of Lemma A4.  $\square$

*Proof of Proposition 1*

Following the arguments in the proof of Proposition 1 in the supplementary file to Chang et al. (arXiv:1406.1939), it suffices to show that  $\sup_{s \in \mathbb{R}} |\text{pr}(\hat{Z} > s) - \text{pr}(V > s)| = o(1)$ , where  $\hat{Z}$  and  $V$  are defined in the first paragraph of the Appendix. Recall that  $d_n = \sup_{s \in \mathbb{R}} |\text{pr}(Z \leq s) - \text{pr}(V \leq s)|$ . By similar arguments to those giving (A2), it can be proved that  $\sup_{s \in \mathbb{R}} |\text{pr}(\hat{Z} > s) - \text{pr}(V > s)| \leq d_n + \text{pr}(|\hat{Z} - Z| > \varepsilon) + C\varepsilon\{\log(p/\varepsilon)\}^{1/2}$ . Set  $\varepsilon = Cn^{-1/2} \log p$ ; then Lemmas A3 and A4 yield that  $\sup_{s \in \mathbb{R}} |\text{pr}(\hat{Z} > s) - \text{pr}(V > s)| = o(1)$ . This completes the proof of Proposition 1.

*Proof of Theorem 1*

Based on Lemma 4 of Chang et al. (arXiv:1603.06663) and Proposition 1, we can proceed with the proof in the same manner as the proof of Theorem 2 in arXiv:1603.06663.

*Proof of Theorem 2*

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

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

where  $\hat{\Xi}_1, \dots, \hat{\Xi}_{p^2 K}$  are the elements in the diagonal of  $\hat{\Xi}_n$ . On the other hand,  $\text{pr}\{|G|_\infty \geq E(|G|_\infty | \mathcal{X}_n) + u | \mathcal{X}_n\} \leq \exp\{-u^2 / (2 \max_{1 \leq \ell \leq p^2 K} \hat{\Xi}_\ell)\}$  holds for any  $u > 0$ . Let  $\Xi_1, \dots, \Xi_{p^2 K}$  be the elements in the main diagonal of  $\Xi_n$ . In addition, for any  $v > 0$ , let  $\mathcal{E}_0(v) = \{\max_{1 \leq \ell \leq p^2 K} |\hat{\Xi}_\ell^{1/2} / \Xi_\ell^{1/2} - 1| \leq v\}$ . Restricted to  $\mathcal{E}_0(v)$ , we have that

$$\hat{c}v_\alpha \leq (1 + v) ([1 + \{2 \log(p^2 K)\}^{-1}] \{2 \log(p^2 K)\}^{1/2} + \{2 \log(1/\alpha)\}^{1/2}) \max_{1 \leq \ell \leq p^2 K} \Xi_\ell^{1/2}.$$

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

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

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

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

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

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

Notice that  $u \sim \varepsilon_n$ . Thus  $u \lambda(p, \alpha) \max_{1 \leq \ell \leq p^2 K} \Xi_\ell^{1/2} \rightarrow \infty$ , which implies that  $\text{pr}(T_n > \hat{c}v_\alpha) \rightarrow 1$ . This completes the proof of Theorem 2.

*Proof of Theorem 3*

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

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

It follows from the Cauchy–Schwarz inequality that

$$\begin{aligned} \Delta_{ij}^2 &\leq 3 \left[ \frac{1}{n} \sum_{t=1}^n \{g_i(u_t; \hat{\theta}) - g_i(u_t; \theta_0)\}^2 \right] \left[ \frac{1}{n} \sum_{t=1}^n \{g_j(u_t; \hat{\theta}) - g_j(u_t; \theta_0)\}^2 \right] \\ &\quad + 3 \left[ \frac{1}{n} \sum_{t=1}^n \{g_i(u_t; \hat{\theta}) - g_i(u_t; \theta_0)\}^2 \right] \left( \frac{1}{n} \sum_{t=1}^n \varepsilon_{j,t}^2 \right) \\ &\quad + 3 \left[ \frac{1}{n} \sum_{t=1}^n \{g_j(u_t; \hat{\theta}) - g_j(u_t; \theta_0)\}^2 \right] \left( \frac{1}{n} \sum_{t=1}^n \varepsilon_{i,t}^2 \right). \end{aligned} \tag{A3}$$

By Condition 5, we have that uniformly for any  $i = 1, \dots, p$ ,

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

On the other hand, Lemma A2 implies that  $\sup_{1 \leq i \leq p} n^{-1} \sum_{t=1}^n \varepsilon_{i,t}^2 = O_p(1)$ . This, together with (A3), implies that  $\Delta_{ij}^2 = O_p(\zeta_n^2 \varphi_{1,n} + \varphi_{2,n})$  uniformly for any  $i, j = 1, \dots, p$ . Thus  $|\hat{\Sigma}^*(0) - \hat{\Sigma}(0)|_\infty = O_p(\zeta_n \varphi_{1,n}^{1/2} + \varphi_{2,n}^{1/2}) = o_p(1)$ . This completes the proof of Theorem 3.

REFERENCES

ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–88.

BAI, J. & NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

BASU, S. & MICHAELIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43**, 1535–67.

BOX, G. E. P. & PIERCE, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Statist. Assoc.* **65**, 1509–26.

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

CHANG, J., GUO, B. & YAO, Q. (2015). High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. *J. Economet.* **189**, 297–312.

- CHERNOZHUKOV, V., CHETVERIKOV, D. & KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41**, 2786–819.
- DEN HAAN, W. J. & LEVIN, A. (1997). A Practitioner's Guide to Robust Covariance Matrix Estimation. *Handbook of Statistics*, vol. 15. G. S. Maddala and C. R. Rao, eds. Amsterdam: Elsevier., Chapter 12, pp. 291–341.
- DEO, R. S. (2000). Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *J. Economet.* **99**, 291–315.
- DURLAUF, S. N. (1991). Spectral based testing of the martingale hypothesis. *J. Economet.* **50**, 355–76.
- ESCANCIANO, J. C. & LOBATO, I. N. (2009). An automatic Portmanteau test for serial correlation. *J. Economet.* **151**, 140–9.
- FAN, J. & YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.
- FAN, J., ZHANG, C. & ZHANG, J. (2001). Generalized likelihood test statistic and Wilks phenomenon. *Ann. Statist.* **29**, 153–93.
- 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. *J. Am. Statist. Assoc.* **100**, 830–40.
- FRANCO, C., ROY, R. & ZAKOIAN, J. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *J. Am. Statist. Assoc.* **100**, 532–44.
- GUO, S., WANG, Y. & YAO, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika* **103**, 889–903.
- HOROWITZ, J. L., LOBATO, I. N., NANKERVIS, J. C. & SAVIN, N. E. (2006). Bootstrapping the Box–Pierce Q test: A robust test of uncorrelatedness. *J. Economet.* **133**, 841–62.
- HOSKING, J. R. M. (1980). The multivariate portmanteau statistic. *J. Am. Statist. Assoc.* **75**, 602–7.
- 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*. New York: Springer.
- LAM, C. & YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40**, 694–726.
- LI, W. K. (2004). *Diagnostic Checks in Time Series*. Boca Raton: Chapman & Hall/CRC.
- LI, W. K. & MCLEOD, A. I. (1981). Distribution of the residual autocorrelations in multivariate time series models. *J. R. Statist. Soc. B* **43**, 231–9.
- LOBATO, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Am. Statist. Assoc.* **96**, 1066–76.
- LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York: Springer.
- R DEVELOPMENT CORE TEAM (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, <http://www.R-project.org>.
- ROMANO, J. P. & THOMBS, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Am. Statist. Assoc.* **91**, 590–600.
- SHAO, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Economet. Theory* **27**, 312–43.
- TIAO, G. C. & BOX, G. E. P. (1981). Modeling multiple time series with applications. *J. Am. Statist. Assoc.* **76**, 802–16.

[Received on 28 November 2015. Editorial decision on 5 November 2016]