The Annals of Statistics
2019, Vol. 47, No. 6, 3382–3412
https://doi.org/10.1214/18-AOS1782

ON TESTING FOR HIGH-DIMENSIONAL WHITE NOISE

BY ZENG LI[†,1], CLIFFORD LAM[*], JIANFENG YAO[‡,2] AND QIWEI YAO[*]

London School of Economics[*], *Pennsylvania State University*[†] *and The University of Hong Kong*[‡]

Testing for white noise is a classical yet important problem in statistics, especially for diagnostic checks in time series modeling and linear regression. For high-dimensional time series in the sense that the dimension p is large in relation to the sample size T, the popular omnibus tests including the multivariate Hosking and Li–McLeod tests are extremely conservative, leading to substantial power loss. To develop more relevant tests for high-dimensional cases, we propose a portmanteau-type test statistic which is the sum of squared singular values of the first q lagged sample autocovariance matrices. It, therefore, encapsulates all the serial correlations (up to the time lag q) within and across all component series. Using the tools from random matrix theory and assuming both p and T diverge to infinity, we derive the asymptotic normality of the test statistic under both the null and a specific VMA(1) alternative hypothesis. As the actual implementation of the test requires the knowledge of three characteristic constants of the population cross-sectional covariance matrix and the value of the fourth moment of the standardized innovations, nontrivial estimations are proposed for these parameters and their integration leads to a practically usable test. Extensive simulation confirms the excellent finite-sample performance of the new test with accurate size and satisfactory power for a large range of finite (p, T) combinations, therefore, ensuring wide applicability in practice. In particular, the new tests are consistently superior to the traditional Hosking and Li–McLeod tests.

1. Introduction. Testing for white noise is an important problem in statistics. It is indispensable in diagnostic checking for linear regression and linear time series modeling in particular. The surge of recent interests in modeling high-dimensional time series adds a further challenge: diagnostic checking demands the testing for high-dimensional white noise in the sense that the dimension of time series is comparable to or even greater than the sample size (i.e., the observed length of the time series). One prominent example showing the need for

Received June 2017; revised September 2018.

[1]Supported by NIDA, NIH Grants P50 DA039838, NSF Grant DMS 1512422 and National Nature Science Foundation of China (NNSFC), 11690015.

[2]Supported by HKSAR RGC General Research Fund #17306918.

MSC2010 subject classifications. Primary 62M10, 62H15; secondary 15A52.

Key words and phrases. Large autocovariance matrix, Hosking test, Li–McLeod test, high-dimensional time series, random matrix theory.

diagnostic checking in high-dimensional time series concerns the vector autoregressive model, which has a large literature. When the dimension is large, most existing works regularize the fitted models by Lasso (Basu and Michailidis (2015), Haufe et al. (2009), Hsu, Hung and Chang (2008), Shojaie and Michailidis (2010)), Dantzig penalization (Han, Lu and Liu (2015)), banded autocovariances (Bickel and Gel (2011)) or banded auto-coefficient matrices (Guo, Wang and Yao (2016)). However, none of them have developed any residual-based diagnostic tools. Another popular approach is to represent high-dimensional time series by lower-dimensional factors; see, for example, Stock and Watson (1989, 1998, 1999), Forni et al. (2000, 2005), Bai and Ng (2002), Lam and Yao (2012) and Chang, Guo and Yao (2015). Again, there is a pertinent need to develop appropriate tools for checking the validity of the fitted factor models through careful examination of the residuals.

There are several well-established white noise tests for univariate time series (Li (2004)). Some of them have been extended for testing vector time series (Hosking (1980), Li and McLeod (1981), Lütkepohl (2005)). However, these methods are designed for the cases where the dimension of the time series is small or relatively small compared to the sample size. For the purpose of model diagnostic checking, the so-called omnibus tests are often adopted which are designed to detect any forms of departure from white noise. The celebrated Box–Pierce portmanteau test and its variations are the most popular omnibus tests. The fact that the Box–Pierce test and its variations are asymptotically distribution-free and χ^2-distributed under the null hypothesis makes them particularly easy to use in practice. However, it is well known in the literature that the slow convergence to their asymptotic null distributions is particularly pronounced in multivariate cases. On the other hand, testing for high-dimensional time series is still in an infancy stage. To our best knowledge, the only available methods are Chang, Yao and Zhou (2017) and Tsay (2017).

To appreciate the challenge in testing for a high-dimensional white noise, we refer to an example reported in Section 3.1 below where, say we have to check the residuals from a fitted multivariate volatility for a portfolio containing $p = 50$ stocks using their daily returns over a period of one semester. The length of the returns time series is then approximately $T = 100$. Table 1 shows that the two variants of the multivariate portmanteau test, namely the Hosking and Li–McLeod tests, all have actual sizes around 0.1%, instead of the nominal level of 5%. These omnibus tests are thus extremely conservative and they will not be able to detect an eventual misfitting of the volatility model.

The above example illustrates the following fact, which is now better understood: many popular tools in multivariate statistics are severely challenged by the emergence of high-dimensional data, and they need to be *reexamined* or *corrected*. Recent advances in high-dimensional statistics demonstrate that feasible and quality solutions to these high-dimensional challenges can be obtained by exploiting

TABLE 1
Empirical sizes for our tests G_q and $G_{q,1}$, the Hosking test $\widetilde{Q}_q$ and the Li–McLeod test Q_q^*

			G_q		$G_{q,1}$		$\widetilde{Q}_q$		Q_q^*	
p	T	p/T	$q=1$	$q=3$	$q=1$	$q=3$	$q=1$	$q=3$	$q=1$	$q=3$
5	1000	0.005	0.0630	0.0615	0.0610	0.0645	0.0490	0.0478	0.0488	0.0476
10	2000	0.005	0.0630	0.0580	0.0615	0.0575	0.0492	0.0440	0.0492	0.0436
25	5000	0.005	0.0520	0.0470	0.0575	0.0535	0.0498	0.0528	0.0498	0.0528
40	8000	0.005	0.0565	0.0395	0.0540	0.0430	0.0508	0.0520	0.0508	0.0520
10	1000	0.01	0.0740	0.0565	0.0675	0.0570	0.0472	0.0468	0.0470	0.0464
20	2000	0.01	0.0500	0.0555	0.0540	0.0540	0.0502	0.0530	0.0502	0.0530
50	5000	0.01	0.0455	0.0555	0.0450	0.0580	0.0488	0.0498	0.0488	0.0498
80	8000	0.01	0.0500	0.0490	0.0510	0.0520	0.0464	0.0406	0.0464	0.0404
50	1000	0.05	0.0375	0.0495	0.0410	0.0475	0.0408	0.0466	0.0408	0.0466
100	2000	0.05	0.0570	0.0525	0.0560	0.0515	0.0432	0.0414	0.0432	0.0414
250	5000	0.05	0.0500	0.0480	0.0495	0.0500	0.0456	0.0436	0.0456	0.0434
400	8000	0.05	0.0410	0.0480	0.0455	0.0505	0.0418	0.0410	0.0418	0.0410
10	100	0.1	0.0570	0.0555	0.0555	0.0570	0.0300	0.0400	0.0280	0.0362
40	400	0.1	0.0560	0.0590	0.0575	0.0525	0.0362	0.0342	0.0358	0.0338
60	600	0.1	0.0465	0.0585	0.0550	0.0595	0.0340	0.0340	0.0340	0.0338
100	1000	0.1	0.0515	0.0500	0.0435	0.0480	0.0370	0.0268	0.0366	0.0264
50	100	0.5	0.0520	0.0465	0.0480	0.0520	0.0006	0.0018	0.0006	0.0018
200	400	0.5	0.0400	0.0415	0.0505	0.0545	0.0010	0.0004	0.0010	0.0004
300	600	0.5	0.0390	0.0480	0.0455	0.0480	0.0002	0.0008	0.0002	0.0008
500	1000	0.5	0.0470	0.0470	0.0430	0.0545	0	0	0	0
90	100	0.9	0.0555	0.0580	0.0460	0.0455	0	0	0	0
360	400	0.9	0.0475	0.0520	0.0535	0.0405	0	0	0	0
540	600	0.9	0.0535	0.0550	0.0550	0.0540	0	0	0	0
900	1000	0.9	0.0495	0.0505	0.0545	0.0515	0	0	0	0

tools of random matrix theory via a precise spectral analysis of large sample covariance or sample autocovariance matrices. For a review on such progress, we refer to Johnstone (2007), Paul and Aue (2014) and monograph Yao, Zheng and Bai (2015). In particular, asymptotic results found in this context using random matrix theory exhibit fast convergence rates, and hence provide satisfactory finite sample approximation for data analysis.

This paper proposes a new method for testing high-dimensional white noise. The test statistic encapsulates the serial correlations within and across all component series. Precisely, the statistic is the sum of the squared singular values of several lagged sample autocovariance matrices. Using random matrix theory, asymptotic normality for the test statistics under the null is established under the Marčenko–Pastur asymptotic regime where p and T are large and comparable.

Next, original methods are proposed for estimation of a few parameters in the limiting distribution in order to get a fully implementable version of the test. The asymptotic power of the test under a specific alternative of first-order vector moving average process (VMA(1)) has also been derived. Extensive simulation demonstrates excellent behavior of the proposed tests for a wide array of combinations of (p, T), with accurate size and satisfactory power. In this paper, we also explore the reasons why the popular multivariate Hosking and Li–McLeod tests are no longer reliable when the dimension is large in relation to the sample size.

The rest of the paper is organized as follows. Section 2 presents the main contributions of the paper. A new high-dimensional test for white noise is introduced; its asymptotic distributions under both the null and the VMA(1) alternative hypothesis are established. Section 3 reports extensive Monte Carlo experiments which assess the finite sample behavior of the tests. Whenever possible, comparison is made with the popular Hosking and Li–McLeod tests. Numerical evidence also indicates that the new test is more powerful than that of Chang, Yao and Zhou (2017). Section 4 collects all the technical proofs.

2. A test for high-dimensional white noise.

Let $\mathbf{x}_1, \dots, \mathbf{x}_T$ be observations from a $p \times 1$ complex-valued linear process of the form

$$\mathbf{x}_t = \sum_{l \geq 0} A_l \mathbf{z}_{t-l},$$

where A_l are $p \times p$ coefficient matrices, $\{\mathbf{z}_t\}$ is a sequence of p-dimensional random vectors such that, if the coordinates of $\mathbf{z}_t$ are $\{z_{it}\}$, then the two-dimensional array $\{z_{it} : 1 \leq i \leq p, t \geq 1\}$ of variables are i.i.d. satisfying the moment conditions $\mathbb{E}z_{it} = 0$, $\mathbb{E}|z_{it}|^2 = 1$ and $\mathbb{E}|z_{it}|^4 = \nu_4 < \infty$. Hence $\mathbb{E}\mathbf{x}_t = \mathbf{0}$, and $\Sigma_\tau \equiv \mathrm{Cov}(\mathbf{x}_{t+\tau}, \mathbf{x}_t)$ depends on τ only. Note that $\Sigma_0 = \mathrm{var}(\mathbf{x}_t)$ is the population covariance matrix of the time series. The goal is to test the null hypothesis

$$H_0 : \mathbf{x}_t = A_0 \mathbf{z}_t, \tag{2.1}$$

where A_0 is unknown. This in fact tests the independence instead of linear independence (i.e., $\Sigma_\tau = 0$ for all $\tau \neq 0$), which is however a common practice in the literature of white noise tests. Throughout the paper, the complex adjoint of a matrix (or vector) A is denoted by A^*. For $\tau \geq 1$, let $\widehat{\Sigma}_\tau$ be the *lag* τ *sample autocovariance matrix*

$$\widehat{\Sigma}_\tau = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_{t-\tau}^*,$$

where by convention $\mathbf{x}_t = \mathbf{x}_{T+t}$ when $t \leq 0$. Under the null hypothesis, $\mathbb{E}(\widehat{\Sigma}_\tau) = 0$ for $\tau \neq 0$, and a natural test statistic is the sum of squared singular values of the first q lagged sample autocovariance matrices:

$$G_q = \sum_{\tau=1}^{q} \mathrm{Tr}(\widehat{\Sigma}_\tau^* \widehat{\Sigma}_\tau) = \sum_{\tau=1}^{q} \sum_j \alpha_{\tau,j}^2,$$

where $\{\alpha_{\tau,j}\}$ are the singular values of $\widehat{\Sigma}_\tau$, and Tr denotes the trace operation for square matrices. We reject the null hypothesis H_0 for large values of G_q.

Notice that the setting here allows for complex-valued observations: this is important for applications in areas such as signal processing where signal time series are usually complex-valued. However, for the sake of presentation, we mostly focus on the real-valued case in the subsequent sections. Directions on how the tests can be extended to accommodate complex-valued observations will be given in the last Section 2.4.

2.1. *High-dimensional asymptotics.* We adopt the so-called *Marčenko–Pastur regime* for asymptotic analysis, that is, we assume $c_p = p/T \to c > 0$ when $p, T \to \infty$. This asymptotic framework has been widely employed in the literature on high-dimensional statistics; see Johnstone (2007), Paul and Aue (2014), also monograph Yao, Zheng and Bai (2015) and the references therein. Most of the results in this area concern sample covariance matrices only. However, our test statistic G_q is based on the sample autocovariance matrices, which is much less studied; see Liu, Aue and Paul (2015) and Bhattacharjee and Bose (2016).

As a main contribution of the paper, we characterize the asymptotic distribution of G_q in this high-dimensional setting when the observations are real-valued. We introduce the following limits whenever they exist: for $\ell \geq 1$,

$$s_\ell = \lim_{p\to\infty} \frac{1}{p}\operatorname{Tr}(\Sigma_0^\ell), \qquad s_{d,\ell} = \lim_{p\to\infty} \frac{1}{p}\operatorname{Tr}(D^\ell(\Sigma_0)), \tag{2.2}$$

where $D(A)$ denotes the diagonal matrix consisting of the main diagonal elements of A (here the d in the index is a reminder of this diagonal structure).

THEOREM 2.1. *Let $q \geq 1$ be a fixed integer, and the following assertions hold*:

1. $\{\mathbf{z}_t\}$ *is a sequence of real-valued independent $p \times 1$ random vectors with independent components $\mathbf{z}_t = (z_{it})$ satisfying $\mathbb{E}z_{it} = 0$, $\mathbb{E}z_{it}^2 = 1$ and $\mathbb{E}z_{it}^4 = \nu_4 < \infty$*;
2. $\{\Sigma_0\}$ *is a sequence of $p \times p$ semipositive definite matrices with bounded spectral norm such that the limits $\{s_1, s_2\}$ and $\{s_{d,2}\}$ exist*;
3. (*Marčenko–Pastur regime*). *The dimension p and the sample size T grow to infinity in a related way such that $c_p := p/T \to c > 0$.*

Then when $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$, the limiting distribution of the test statistic G_q is

$$G_q - qTc_p^2 s_1^2 \xrightarrow{d} \mathcal{N}(0, \sigma^2(c)), \tag{2.3}$$

where

$$\sigma^2(c) = 2qc^2 s_2^2 + 4q^2c^3(\nu_4 - 3)s_1^2 s_{d,2} + 8q^2c^3 s_1^2 s_2. \tag{2.4}$$

The proof of this theorem is given in Section 4. It is worth mentioning here that in Bhattacharjee and Bose (2016), they considered a simpler case when $\Sigma_0 = \mathbf{I}_p$, $q = 1$ and $p = T$ with Gaussian population distribution, which is consistent with the results above.

Let Z_α be the upper-α quantile of the standard normal distribution at level α. Based on Theorem 2.1, we obtain a procedure for testing the null hypothesis in (2.1) as follows:

$$\textit{Reject}\quad H_0 \quad \textit{if}\quad \{G_q - qTc_p^2 s_1^2 > Z_\alpha \sigma(c)\}. \tag{2.5}$$

The illustration in Section 3 indicates that the test above is much more powerful than some classical alternatives, especially when the dimension p is growing linearly with the sample size T. The power of this test is gained from gathering together the serial correlations from the first q lags within and across all p component series; see the definition of G_q. Also note that the asymptotic mean of G_q is $qTc_p^2 s_1^2$, which grows linearly with T (and p), while its asymptotic variance $\sigma^2(c)$ is a constant. This implies that even for moderately large T, departure from white noise in the first q lags of the autocovariance matrices is likely to result in a large and different mean, which will be a large multiple standard deviation away from $qTc_p^2 s_1^2$ since the standard deviation $\sigma(c)$ is constant.

However, the test G_q in (2.5) is not yet practically usable as it depends on (i) three characteristic constants, s_1, s_2 and $s_{d,2}$ of the (population) cross-sectional covariance matrix Σ_0 and (ii) the fourth moment ν_4 of the innovations $\{\mathbf{z}_t\}$. These issues are addressed below.

2.2. *Estimation of the covariance characteristics* s_1 *and* s_2. If the cross-sectional covariance matrix Σ_0 is known, reasonable approximations of these characteristics are readily calculated from Σ_0. By Slutsky's theorem, these estimates can substitute for the true ones in the asymptotic variance $\sigma^2(c)$ *and* the centering term $qTc_p^2 s_1^2$. The test (2.5) still applies.

However, the population covariance matrix Σ_0 is in general unknown and the situation becomes challenging as estimating a general Σ_0 is somehow out of reach without specific assumptions on its structure. Luckily, as observed previously, we only need consistent estimates of the three characteristics. First of all, in the setting of Theorem 2.1 and under the null, it is not difficult to find *consistent* estimators for these characteristics, thus a consistent estimator of the limiting variance $\sigma^2(c)$. The situation is much more intricate for the centering term $qTc_p^2 s_1^2$. Suppose $\hat{s}_1^2$ is a consistent estimator of s_1^2. Plugging it into the centering term leads to

$$G_{q,1} := G_q - qTc_p^2 \hat{s}_1^2 = \{G_q - qTc_p^2 s_1^2\} + qTc_p^2\{s_1^2 - \hat{s}_1^2\}\,. \tag{2.6}$$

Because of the multiplication by T here, the asymptotic distribution would remain the same only if the estimation error $\{\hat{s}_1^2 - s_1^2\}$ is of order $o_P(1/T)$. This is however not the case and in general the error is exactly of the order $O_p(1/T)$ and $T\{\hat{s}_1^2 - s_1^2\}$ converges to some other normal distribution.

Our method is as follows. First, we establish the joint asymptotic distribution of $G_q - qTc_p^2 s_1^2$ and $p\{\hat{s}_1^2 - s_1^2\}$ for a natural estimator $\hat{s}_1^2$. This result extends Theorem 2.1 which addresses the statistic $G_q - qTc_p^2 s_1^2$ only. Next, the asymptotic null distribution of the "feasible" test statistic $G_{q,1}$ is readily obtained as a simple consequence.

Precisely, consider the sample covariance matrix $\widehat{\Sigma}_0 = \frac{1}{T}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t^*$ and define the natural estimators of s_1 and s_2 as

$$\hat{s}_1 = \frac{1}{p}\operatorname{Tr}(\widehat{\Sigma}_0), \qquad \hat{s}_2 = \frac{1}{p}\operatorname{Tr}(\widehat{\Sigma}_0^2).$$

THEOREM 2.2. *Assume the same conditions as in Theorem 2.1, then when $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$, we have*

$$\begin{pmatrix} p(\hat{s}_1^2 - s_1^2) \\ G_q - qTc_p^2 s_1^2 \end{pmatrix} \xrightarrow{d} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4c(\nu_4 - 3)s_1^2 s_{d,2} + 8cs_1^2 s_2 & 4qc^2(\nu_4 - 3)s_1^2 s_{d,2} + 8qc^2 s_1^2 s_2 \\ 4qc^2(\nu_4 - 3)s_1^2 s_{d,2} + 8qc^2 s_1^2 s_2 & \sigma^2(c) \end{pmatrix}\right),$$

where the variance $\sigma^2(c)$ is given in (2.4).

The proof of this theorem is relegated to Section 4.

Applying Theorem 2.2 to the decomposition (2.6), the following proposition establishes the asymptotic null distribution of the feasible statistic $G_{q,1}$. Second-order terms of the mean and variance of $G_{q,1}$ are also provided to improve finite sample performance.

PROPOSITION 2.1. *Assume the same conditions as in Theorem 2.2 and the observations are real-valued, we have*

$$G_{q,1} = G_q - qTc_p^2 \hat{s}_1^2 \xrightarrow{d} \mathcal{N}(0, \xi^2(c)), \tag{2.7}$$

where $\xi^2(c) = 2qc^2 s_2^2$. Meanwhile,

$$\mathbb{E}(G_{q,1}) = -\frac{q}{T^2}\big(2\operatorname{Tr}(\Sigma_0^2) + (\nu_4 - 3)\operatorname{Tr}(D^2(\Sigma_0))\big), \qquad \mathbb{E}(\hat{s}_1) = \frac{1}{p}\operatorname{Tr}(\Sigma_0),$$

$$\operatorname{Var}(G_{q,1}) = \frac{2q}{T^2}\operatorname{Tr}^2(\Sigma_0^2) + \frac{q}{T^3}\big(2\operatorname{Tr}(\Sigma_0^2) + (\nu_4 - 3)\operatorname{Tr}(D^2(\Sigma_0))\big)^2 + o\left(\frac{1}{T}\right),$$

$$\mathbb{E}(\hat{s}_2) = \frac{1}{p}\operatorname{Tr}(\Sigma_0^2) + \frac{1}{pT}\operatorname{Tr}^2(\Sigma_0) + \frac{1}{pT}\big(\operatorname{Tr}(\Sigma_0^2) + (\nu_4 - 3)\operatorname{Tr}(D^2(\Sigma_0))\big).$$

Now we aim at consistent estimates for the unknown quantity s_2 in the asymptotic variance $\xi^2(c)$. It is well known that almost surely (Bai, Chen and Yao (2010)),

$$\hat{s}_1 \to s_1, \qquad \hat{s}_2 \to s_2 + c s_1^2.$$

Therefore, $\tilde{s}_2 = \hat{s}_2 - c_p \hat{s}_1^2$ is a strongly consistent estimator of s_2.

In summary, when Σ_0 is unknown, we obtain a procedure for testing the null hypothesis of white noise (2.1) as follows:

$$\text{(2.8)} \qquad \textit{Reject } H_0 \textit{ if } \{G_q - qTc_p^2\hat{s}_1^2 > Z_\alpha \tilde{\xi}\},$$

where $\tilde{\xi}^2 = 2qc_p^2\tilde{s}_2^2$.

2.3. *Finite sample correction and estimation for non-Gaussian innovations.* Although the test procedure (2.8) is already practically usable, it can be further improved by finite sample corrections provided in Proposition 2.1 which are especially useful for non-Gaussian population where $\nu_4 \neq 3$. To this goal, it remains to obtain a consistent estimate for (i) the covariance characteristic

$$s_{d,2} = \frac{1}{p}\sum_{i=1}^{p} d_i^2 = \frac{1}{p}\mathrm{Tr}\big(D^2(\Sigma_0)\big),$$

where $d_i = \Sigma_{0,ii}$ is the ith diagonal element of Σ_0, and (ii) the fourth moment ν_4 of the innovations.

(i) *Estimation of* $s_{d,2}$. By its very definition, d_i can be consistently estimated by its sample counterpart

$$\tilde{d}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}^2.$$

It follows that a consistent estimator for $s_{d,2}$ is simply $\tilde{s}_{d,2} = p^{-1}\sum_{i=1}^{p}\tilde{d}_i^2$.

(ii) *Estimation of* ν_4. This is again a nontrivial problem which has not been touched yet in the literature (to our best knowledge). In order to get rid of the role of the unknown cross-sectional covariance matrix Σ_0, we adopt the following splitting strategy: the original data $\{\mathbf{x}_t, t = 1, \dots, T\}$ are split into two halves of length T_1 and T_2, respectively ($T = T_1 + T_2$). Define the two corresponding sample cross-sectional covariance matrices

$$\text{(2.9)} \qquad S_{n,1} = \frac{1}{T_1}\sum_{t=1}^{T_1} \mathbf{x}_t\mathbf{x}_t^*, \qquad S_{n,2} = \frac{1}{T_2}\sum_{t=1}^{T_2} \mathbf{x}_{t+T_1}\mathbf{x}_{t+T_1}^*.$$

This yields the corresponding F-ratio, or Fisher matrix, $F_n = S_{n,1}^{-1}S_{n,2}$. Observe that this matrix does not depend on the value of the cross-sectional covariance Σ_0 so that in what follows we can assume $\Sigma_0 = I$.

Let $(\lambda_j)_{1\le j\le p}$ be the eigenvalues of F_n. Define K test functions $f_k(x) = \log(a_k + b_k x)$ where $(a_k, b_k)_{1\le k\le K}$ are some positive constants. For each k, we have an eigenvalue statistic of the Fisher matrix

$$X_{T,k} = f_k(\lambda_1) + \cdots + f_k(\lambda_p) - p \int f_k(x)\, dF_{c_{p,1},c_{p,2}}(x),$$

where $c_{p,i} = p/T_i$ $(i = 1, 2)$ and $F_{c,c'}$ is the limiting Wachter distribution with index (c, c'); see the formula (3.1) in Zheng (2012). It is proved on page 452 of the reference, when p, T_1, T_2 grow proportionally to infinity,

$$X_{T,k} = u_{T,k} + v_{T,k}\nu_4 + \varepsilon_{T,k}, \tag{2.10}$$

where $\{u_{T,k}, v_{T,k}\}$ are constants depending on $\{c_{p_i}\}$ and (a_k, b_k), and $\{\varepsilon_{T,k}\}$ is a centered and asymptotically Gaussian error. Then the least squares estimator of ν_4 using the above regression model leads to a consistent estimate, say $\hat{\nu}_4$ for the unknown parameter.

Under the null hypothesis, the observations are independent. We may repeat this estimation procedure, say B times, by taking B random splits of the initial sample. The final estimate of ν_4 is then taken to be the average of the estimates $\{\hat{\nu}_{4,b}\}_{1\le b\le B}$.

Finally, we can implement the following test procedure with finite sample correction for the null hypothesis of white noise (2.1):

Reject H_0 *if*

$$\left\{G^*_{q,1} = G_q - qTc_p^2\hat{s}_1^2 + \frac{1}{T}\cdot qc_p\bigl(2\tilde{s}_2 + (\hat{\nu}_4 - 3)\tilde{s}_{d,2}\bigr) > Z_\alpha\hat{\xi}\right\}, \tag{2.11}$$

where

$$\hat{\xi}^2 = 2qc_p^2\tilde{s}_2^2 + \frac{1}{T}\cdot qc_p^2\bigl(2\tilde{s}_2 + (\hat{\nu}_4 - 3)\tilde{s}_{d,2}\bigr)^2$$

with the above estimator $\hat{\nu}_4$ for the fourth moment. Note that the estimation procedure proposed above for ν_4 is only feasible when $p < T$, thus we can only implement test (2.11) when $p < T$. However, our primary test statistic is $G_{q,1}$ in (2.8) which does not require estimation of ν_4. In fact, simulation results in Section 3.4 and 3.5 show that the statistic $G_{q,1}$ already performs well. Therefore, we can directly use $G_{q,1}$ when $p > T$.

2.4. *Tests when the observations are complex-valued.* To proceed, we first define $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$ where $\mathbf{z}_t$ is a *proper* complex random vector, and $\Sigma_0^{1/2}$ is such that $\Sigma_0^{1/2}$ is Hermitian with $\Sigma_0 = \Sigma_0^{1/2}(\Sigma_0^{1/2})^*$ (*Properness* of a complex random vector $\mathbf{z}_t$ means that $\mathbb{E}(\mathbf{z}_t\mathbf{z}_t^T) = 0$). We immediately have

$$0 = \mathbb{E}\bigl(\mathbf{z}_t\mathbf{z}_t^T\bigr) = \mathbb{E}\bigl(z_{it}^2\bigr)I_p,$$

so that $\mathbb{E}(z_{it}^2) = 0$ for all $i = 1, \dots, p$ and $t = 1, \dots, T$. It also implies that $b = |\mathbb{E}(z_{it}^2)|^2 = 0$. Since $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$, we have

$$\mathbb{E}(\mathbf{x}_t\mathbf{x}_t^T) = \mathbb{E}(\Sigma_0^{1/2}\mathbf{z}_t\mathbf{z}_t^T\Sigma_0^{T/2}) = 0,$$

so that we are also assuming an observed vector $\mathbf{x}_t$ is proper.

From Corollary 4.1, since $b = 0$ from the properness of $\mathbf{z}_t$, the asymptotic covariance of G_q is then

$$\operatorname{Var}(G_q) \to qc^2s_2^2 + 4q^2c^3s_1^2[(\nu_4 - 2)s_{d,2} - s_2' + 2s_{r,2}],$$

where $s_2' = \lim_{p\to\infty} \operatorname{Tr}(\Sigma_0\Sigma_0^T)/p$, $s_{r,2} = \lim_{p\to\infty}\operatorname{Tr}(\Re^2(\Sigma_0))/p$, with $\Re(A) = (\Re(a_{ij}))$, the matrix of the real parts of all entries in A.

Using Lemma 1.1 of the Supplementary Material (Li et al. (2019)) defining $\Im(A) = (\Im(a_{ij}))$ to be the matrix of the imaginary parts of all entries in A, we have

$$\begin{aligned} 2\operatorname{Tr}(\Re^2(\Sigma_0)) - \operatorname{Tr}(\Sigma_0\Sigma_0^T) &= 2\operatorname{Tr}(\Sigma_0\Re(\Sigma_0)) - \operatorname{Tr}(\Sigma_0(\Re(\Sigma_0) - i\Im(\Sigma_0))) \\ &= \operatorname{Tr}(\Sigma_0(\Re(\Sigma_0) + i\Im(\Sigma_0))) = \operatorname{Tr}(\Sigma_0^2), \end{aligned}$$

so that $2s_{r,2} - s_2' = s_2$. The asymptotic variance for G_q is then

$$\operatorname{Var}(G_q) \to \sigma^2(c) = qc^2s_2^2 + 4q^2c^3s_1^2[(\nu_4 - 2)s_{d,2} + s_2],$$

which can be estimated consistently using the estimators suggested in Section 2.2.

2.5. *Testing power of $G_{q,1}$.* In this section, we look into the power function of the tests when an alternative hypothesis H_1 is specified. Here we assume that under H_1, the observations $\mathbf{x}_1, \dots, \mathbf{x}_T$ follows from a $p \times 1$ real-valued p-dimensional first-order vector moving average process, VMA(1) in short, of the form

$$H_1: \mathbf{x}_t = A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}, \tag{2.12}$$

where A_0, A_1 are $p \times p$ coefficient matrices. Now we only consider the asymptotic behavior of our test statistic G_q and $G_{q,1}$ when $q = 1$ since higher order autocorrelations of $\mathbf{x}_t$ are null under both H_0 and H_1.

Denote

$$\widetilde{\Sigma}_0 = A_0^*A_0, \qquad \widetilde{\Sigma}_1 = A_1^*A_1, \qquad \widetilde{\Sigma}_{01} = A_0^*A_1,$$

we characterize the joint limiting distribution of $\hat{s}_1^2$ and G_1 under the VMA(1) alternative (2.12) as follows.

THEOREM 2.3. *Assume that*:

1. $\{\mathbf{z}_t\}$ *is a sequence of real-valued independent* $p \times 1$ *random vectors with independent components* $\mathbf{z}_t = (z_{it})$ *satisfying* $\mathbb{E}z_{it} = 0$, $\mathbb{E}z_{it}^2 = 1$ *and* $\mathbb{E}z_{it}^4 = \nu_4 < \infty$;

2. $\widetilde{\Sigma}_0$, $\widetilde{\Sigma}_1$ *and* $\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*$ *all have bounded spectral norm and for integers* $i, j, k \geq 0$, $1 \leq i + j + k \leq 4$, *the limits* $\lim_{T\to\infty} \frac{1}{T}\mathrm{Tr}(\widetilde{\Sigma}_0^i \widetilde{\Sigma}_1^j \widetilde{\Sigma}_{01}^k)$ *exist*;

3. (*Marčenko–Pastur regime*). *The dimension* p *and the sample size* T *grow to infinity in a related way such that* $c_p := p/T \to c > 0$.

Then under the VMA(1) *alternative* (2.12), *the joint limiting distribution of the* G_1 *and* $\hat{s}_1^2$ *is*

$$\begin{pmatrix} \sigma_G^2 & \sigma_{GS} \\ \sigma_{GS} & \sigma_S^2 \end{pmatrix}^{-1/2} \begin{pmatrix} G_1 - \mu_G \\ T c_p^2 \hat{s}_1^2 - \mu_S \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\mathbf{0},\ I_2),$$

where

$$\begin{aligned}
\mu_G &= \frac{1}{T}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1) + \mathrm{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) + \frac{2}{T}\mathrm{Tr}^2(\widetilde{\Sigma}_{01}) \\
&\quad + \frac{1}{T}\big[\mathrm{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) + (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_0)D(\widetilde{\Sigma}_1)\big)\big], \\
\mu_S &= \frac{1}{T}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1) + \frac{4}{T^2}\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) \\
&\quad + \frac{1}{T^2}\big[2\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big], \\
\sigma_S^2 &= \frac{4}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&\quad + \frac{16}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) + R_n
\end{aligned}$$

and

$$\begin{aligned}
\sigma_G^2 &= \frac{4}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&\quad + \frac{8}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}\big(\widetilde{\Sigma}_0\widetilde{\Sigma}_1(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) \\
&\quad + (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)D(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&\quad + \frac{2}{T^2}\mathrm{Tr}^2\big(\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_1^2\big) + \frac{6}{T^2}\mathrm{Tr}^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) \\
&\quad + \frac{4}{T}\big[2\mathrm{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)\big)\big] \\
&\quad + \frac{8}{T^2}\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big)\mathrm{Tr}\big(\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_1^2\big) + \frac{16}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_1)\mathrm{Tr}(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0) \\
&\quad + \frac{16}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[\mathrm{Tr}\big(\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0\big) + \mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1\big)\big]
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{16}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_{01})\big[\mathrm{Tr}\big(\widetilde{\Sigma}_0^2\widetilde{\Sigma}_{01}^*\big) + \mathrm{Tr}\big(\widetilde{\Sigma}_1^2\widetilde{\Sigma}_{01}\big) + 2\mathrm{Tr}(\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0)\big] \\
&+ \frac{4}{T}\mathrm{Tr}\big(\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1^2 + 2\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0\big) \\
&+ \frac{16}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) + \frac{16}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_{01})\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 \\
&+ \frac{32}{T^3}\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\mathrm{Tr}(\widetilde{\Sigma}_{01})\mathrm{Tr}\big(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) \\
&+ \frac{4}{T}\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\big) + \frac{12}{T^2}\mathrm{Tr}^2\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) \\
&+ \frac{16}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_{01})\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}^*\big) \\
&+ \frac{16}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_{01})\big[\mathrm{Tr}(\widetilde{\Sigma}_{01})^2 + 2\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_{01})\big)\big] \\
&+ \frac{8}{T^2}\mathrm{Tr}^2(\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}) \\
&+ \frac{16}{T^3}\mathrm{Tr}(\widetilde{\Sigma}_{01})\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}\big(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) \\
&+ (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_{01})D(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&+ \frac{8}{T^2}\mathrm{Tr}^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_{01}) + \frac{16}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_{01})\big[2\mathrm{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}) \\
&+ (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)D(\widetilde{\Sigma}_{01})\big)\big] + R_n,
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{GS} = {}& \frac{4}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&+ \frac{4}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}\big(\widetilde{\Sigma}_0\widetilde{\Sigma}_1(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) \\
&+ (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)D(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&+ \frac{8}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[\mathrm{Tr}\big(\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0\big) + \mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1\big)\big] \\
&+ \frac{16}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\mathrm{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) \\
&+ \frac{8}{T^3}\mathrm{Tr}(\widetilde{\Sigma}_{01})\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}\big(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) \\
&+ (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_{01})D(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&+ \frac{16}{T^3}\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\mathrm{Tr}(\widetilde{\Sigma}_{01})\mathrm{Tr}\big(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) + R_n.
\end{aligned}
$$

Here the R_n's, possibly different, represent remainders which have smaller orders than the other terms listed in σ_S^2, σ_G^2 and σ_{GS}, respectively.

The proof of this theorem is relegated to Section 4. Similarly, applying Theorem 2.3 to the decomposition (2.6), the following proposition establishes the asymptotic distribution of our test statistic $G_{q,1}$ under the VMA(1) alternative (2.12) when $q = 1$.

PROPOSITION 2.2. *Assume the same conditions as in Theorem 2.3, when $\mathbf{x}_t = A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}$ and the observables are real-valued, we have*

$$\sigma_{G_{1,1}}^{-1}\left(G_1 - Tc_p^2\hat{s}_1^2 - \mu_{G_{1,1}}\right) \xrightarrow{d} \mathcal{N}(0,1), \tag{2.13}$$

where

$$\begin{aligned}
\mu_{G_{1,1}} &= \operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) + \frac{2}{T}\operatorname{Tr}^2(\widetilde{\Sigma}_{01}) + \frac{1}{T}\big[\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) + (\nu_4 - 3)\operatorname{Tr}\big(D(\widetilde{\Sigma}_0)D(\widetilde{\Sigma}_1)\big)\big] \\
&\quad - \frac{4}{T^2}\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) - \frac{1}{T^2}\big[2\operatorname{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\operatorname{Tr}\big(D^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big], \\
\sigma_{G_{1,1}}^2 &= \frac{2}{T^2}\operatorname{Tr}^2\big(\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_1^2\big) + \frac{4}{T}\big[2\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\operatorname{Tr}\big(D^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)\big)\big] \\
&\quad + \frac{6}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) + \frac{8}{T^2}\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big)\operatorname{Tr}\big(\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_1^2\big) \\
&\quad + \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_1)\operatorname{Tr}(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0) \\
&\quad + \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01})\big[\operatorname{Tr}\big(\widetilde{\Sigma}_0^2\widetilde{\Sigma}_{01}^*\big) + \operatorname{Tr}\big(\widetilde{\Sigma}_1^2\widetilde{\Sigma}_{01}\big) + 2\operatorname{Tr}(\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0)\big] \\
&\quad + \frac{4}{T}\operatorname{Tr}\big(\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1^2 + 2\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0\big) \\
&\quad + \frac{16}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_{01})\operatorname{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 \\
&\quad + \frac{4}{T}\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\big) + \frac{12}{T^2}\operatorname{Tr}^2\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) \\
&\quad + \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01})\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}^*\big) \\
&\quad + \frac{16}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_{01})\big[\operatorname{Tr}(\widetilde{\Sigma}_{01})^2 + 2\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) + (\nu_4 - 3)\operatorname{Tr}\big(D^2(\widetilde{\Sigma}_{01})\big)\big] \\
&\quad + \frac{8}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}) + \frac{8}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_{01}) \\
&\quad + \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01})\big[2\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}) + (\nu_4 - 3)\operatorname{Tr}\big(D(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)D(\widetilde{\Sigma}_{01})\big)\big] + R_n.
\end{aligned}$$

Here R_n represents a remainder of smaller order than the other terms listed in $\sigma^2_{G_{1,1}}$.

Notice that if $A_1 = \mathbf{0}$, $\widetilde{\Sigma}_1 = 0$ and $\widetilde{\Sigma}_{01} = 0$, then Theorem 2.3 and Proposition 2.2 reduce to Theorem 2.2 and Proposition 2.1, respectively.

Actually, under the VMA(1) alternative (2.12) with $q = 1$, we have almost surely, $\tilde{\xi} = \sqrt{2} c_p \tilde{s}_2 \to \xi_0$ as $p, T \to \infty$, where

$$\xi_0 = \lim_{T\to\infty} \sqrt{2}\left[\frac{1}{T}\operatorname{Tr}(\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_1^2) + \frac{2}{T}\operatorname{Tr}(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*) + \frac{2}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_{01})\right]. \tag{2.14}$$

With Propositions 2.1 and 2.2, the power function of the test (2.8) is then easily derived.

PROPOSITION 2.3. *Assume the same conditions as in Theorem 2.3, then under $H_1 : \mathbf{x}_t = A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}$, as $p, T \to \infty$, the power function*

$$\beta_\alpha = \Pr\big(G_1 - T c_p^2 \hat{s}_1^2 > Z_\alpha \tilde{\xi} \mid H_1\big) \to \Pr\left(Z > Z_\alpha \frac{\xi_0}{\widetilde{\sigma}_{G_{1,1}}} - \frac{\widetilde{\mu}_{G_{1,1}}}{\widetilde{\sigma}_{G_{1,1}}}\right),$$

where Z represents a standard normal random variable, Z_α is the upper-α quantile of the standard normal distribution, $\widetilde{\mu}_{G_{1,1}}$ and $\widetilde{\sigma}_{G_{1,1}}$ are limits of $\mu_{G_{1,1}}$ and $\sigma_{G_{1,1}}$ as $T \to \infty$.

In fact, under H_1, when $\widetilde{\Sigma}_0$ and $\widetilde{\Sigma}_1$ have bounded spectral norm, both $\widetilde{\sigma}_{G_{1,1}}$ and ξ_0 are of order $O(1)$ and $0 < \frac{\xi_0}{\widetilde{\sigma}_{G_{1,1}}} \le 1$, while the leading order term of $\widetilde{\mu}_{G_{1,1}}$ is

$$\lim_{T\to\infty} \operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) = \lim_{T\to\infty} \operatorname{Tr}\big(A_1 A_0^* A_0 A_1^*\big) > 0.$$

Consequently, we have the following:

Case 1. If $\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)$ diverges as $T \to \infty$, then the power function $\beta_\alpha \to 1$;

Case 2. If $\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)$ is of order $\Omega(1)$ (bounded from below and above), then the power function β_α converges to the constant

$$\beta = \Pr\left(Z > Z_\alpha \frac{\xi_0}{\widetilde{\sigma}_{G_{1,1}}} - \frac{\widetilde{\mu}_{G_{1,1}}}{\widetilde{\sigma}_{G_{1,1}}}\right) \quad \text{and} \quad \alpha \le \beta \le 1.$$

Therefore, as expected the asymptotic power of the test (2.8) under the VMA(1) alternative (2.12) depends on the eigenstructure of the coefficient matrix A_1. To illustrate, assume that (i) $A_0 A_0^*$ is of rank $r_{0p} \sim rp$ for some constant $0 < r \le 1$; (ii) $A_0 A_0^*$ is of rank $1 \ll r_{1p} \ll p$, for example, $r_{1p} \sim r' \log p$ for some constant $r' > 0$, and that the nonnull eigenvalues of both matrices are of order $\Omega(1)$. Then $\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) \sim r'' r_{1p} \to \infty$ for some constant $r'' > 0$, and the asymptotic power is equal to 1 (Case 1). If instead, $r_{1p} = \Omega(1)$, then the asymptotic power can be smaller than 1 (Case 2). Both situations correspond to a low-rank alternative for A_1, with exploding ranks in Case 1 and constant order ranks in Case 2.

Finally, as here the alternative is a VMA(1), one would expect that $G_{q,1}$ with $q > 1$ might have smaller power than $G_{1,1}$. This is indeed true because $\widetilde{\mu}_{G_{q,1}}$ remains the same with $\widetilde{\mu}_{G_{1,1}}$ under H_1, while $\widetilde{\sigma}_{G_{q,1}}$ is larger than $\widetilde{\sigma}_{G_{1,1}}$ and ξ_0 increases with q as well.

3. Simulation experiments. Most of the experiments of this section are designed in order to compare the test procedures in (2.5) and (2.8) based on the statistics G_q and $G_{q,1}$, with two well-known classical white noise tests, namely the Hosking test (Hosking (1980)) and the Li–McLeod test (Li and McLeod (1981)).

To introduce the Hosking and Li–McLeod tests and using their notation, consider a p-dimensional VARMA(u, v) process of the form

$$\mathbf{x}_t - \Phi_1\mathbf{x}_{t-1} - \cdots - \Phi_u\mathbf{x}_{t-u} = \mathbf{a}_t - \Theta_1\mathbf{a}_{t-1} - \cdots - \Theta_v\mathbf{a}_{t-v},$$

where $\mathbf{a}_t$ is a p-dimensional white noise with mean zero and variance Σ. Since $\mathbf{x}_t$ is observed, with an initial guess of u and v, by assuming $\mathbf{a}_t$ to be Gaussian, estimation of parameters $\{\Phi, \Theta\}$ is conducted by the method of maximum likelihood. The initial estimates of u and v are further refined at the diagnostic checking stage based on the autocovariance matrices $\hat{C}_\tau$ of the residuals $\{\hat{\mathbf{a}}_t\}$:

$$\hat{C}_\tau = \frac{1}{T}\sum_{t=\tau+1}^{T} \hat{\mathbf{a}}_t\hat{\mathbf{a}}_{t-\tau}^*, \qquad \tau = 0, 1, 2, \ldots.$$

Hosking (1980) proposed the portmanteau statistic

$$\widetilde{Q}_q = T^2\sum_{\tau=1}^{q}\frac{1}{T-\tau}\operatorname{Tr}(\hat{C}_\tau^*\hat{C}_0^{-1}\hat{C}_\tau\hat{C}_0^{-1}),$$

while Li and McLeod (1981) recommended the use of the statistic

$$Q_q^* = T\sum_{\tau=1}^{q}\operatorname{Tr}(\hat{C}_\tau^*\hat{C}_0^{-1}\hat{C}_\tau\hat{C}_0^{-1}) + \frac{p^2q(q+1)}{2T}.$$

When $\{\mathbf{x}_t\}$ follows a VARMA(u, v) model, both $\widetilde{Q}_q$ and Q_q^* converge to $\chi^2(p^2(q-u-v))$ distribution as $T \to \infty$, while the dimension p remains fixed.

To compare with our multilag q test statistics G_q and $G_{q,1}$ when Σ_0 is either known or unknown, we set $u = v = 0$. All tests use 5% significance level and the critical regions of the three tests are as follows:

(i) G_q when all the limiting parameters are known as defined in (2.5) with $\alpha = 5\%$;

(ii) $G_{q,1}$ with estimated limiting parameters as defined in (2.8) with $\alpha = 5\%$;

(iii) Hosking test: $\{\widetilde{Q}_q > \chi^2_{0.05,qp^2}\}$;

(iv) Li–McLeod test: $\{Q_q^* > \chi^2_{0.05,qp^2}\}$.

Here $Z_{0.05}$ and $\chi^2_{0.05,m}$ denote the upper-5% quantile of the standard normal distribution and the chi-squared distribution with degrees of freedom m, respectively. Empirical statistics are obtained using 2000 independent replicates.

3.1. *Empirical sizes.* The data is generated as $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$, with $\mathbf{z}_t$, $t = 1, \dots, T$ being independent and identically distributed. We adopt diverse settings for $\mathbf{z}_t$ and Σ_0, respectively, to compare the sizes of four test statistics.

As for $\mathbf{z}_t$, we use two models to represent different distributions for $\mathbf{z}_t$:

(I) $\mathbf{z}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, i.i.d. $t = 1, \dots, T$;

(II) $\mathbf{z}_t$ with i.i.d. components $z_{it} \sim \text{Gamma}(4, 0.5) - 2$, $i = 1, \dots, p$, $t = 1, \dots, T$, $\mathbb{E}(z_{it}) = 0$, $\text{Var}(z_{it}) = 1$, $\nu_4(z_{it}) = 4.5$.

As for Σ_0, we use two different models as follows:

(III) $\Sigma_0 = \mathbf{I}_p$;

(IV) $\Sigma_0 = \frac{4}{p} A_0 A_0^*$, A_0 is $p \times p$ matrix with entries $a_{ij} \sim U(-1, 1)$ i.i.d.

Table 1 compares the sizes of the four tests for two different q when $\Sigma_0 = \mathbf{I}_p$. Cases when $p > T$ are not considered here since $\widetilde{Q}_q$ and Q_q^* are not applicable then.

The main information from Table 1 is that classical test procedures derived using large sample scheme, namely by letting the sample size $T \to \infty$ while the dimension p remains fixed, are heavily biased when the dimension p is in fact not negligible with respect to the sample size. To be more precise, these biases are clearly present when the dimension-to-sample ratio p/T is not "small" enough, say, greater than 0.1. Such high-dimensional traps for classical procedures have already been reported in other testing problems; see, for example, Bai et al. (2009) and Wang and Yao (2013). Here we observe that the empirical sizes of the Hosking and the Li–McLeod tests quickly degenerate to 0 as the ratio p/T increases from 0.1 to 0.5. In other words, the critical values from their $\chi^2_{qp^2}$ asymptotic limits are seemingly *too large*. On the other hand, the statistics G_q and $G_{q,1}$ have reasonable sizes when compared to the 5% nominal level across all the tested (p, T) combinations. Various (p, T) combinations are accommodated to testify the adaptability of our test statistics, G_q and $G_{q,1}$. Test sizes in both high and low dimension cases are shown in Table 2. It can be seen that both G_q and $G_{q,1}$ attain the nominal level accurately under various scenarios.

3.2. *Empirical powers and adjusted powers.* In this section, we compare the empirical powers of the tests by assuming that $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{y}_t$, $\mathbf{y}_t$ follows a vector autoregressive process of order 1,

$$\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{y}_t, \qquad \mathbf{y}_t = A\mathbf{y}_{t-1} + \mathbf{z}_t,$$

where $A = a\mathbf{I}_p$, $\mathbf{z}_t \sim N_p(\mathbf{0}, \mathbf{I}_p)$ being independent of each other for $t = 1, \dots, T$. First, we check the power of two classic test procedures, $\widetilde{Q}_q$ and Q_q^*. Table A

TABLE 2
Test sizes of our tests G_q and $G_{q,1}$

			Gaussian (I)				Non-Gaussian (II)				
			G_q		$G_{q,1}$		G_q		$G_{q,1}$		
p	T	p/T	$q=1$	$q=3$	$q=1$	$q=3$	$q=1$	$q=3$	$q=1$	$q=3$	
5	500	0.01	0.0500	0.0545	0.0465	0.0485	0.0650	0.0655	0.0655	0.0540	
10	1000	0.01	0.0565	0.0420	0.0575	0.0400	0.0515	0.0615	0.0600	0.0575	
20	2000	0.01	0.0545	0.0570	0.0515	0.0525	0.0610	0.0595	0.0600	0.0510	
25	500	0.05	0.0550	0.0570	0.0630	0.0510	0.0570	0.0645	0.0520	0.0565	
50	1000	0.05	0.0520	0.0515	0.0510	0.0455	0.0500	0.0485	0.0495	0.0455	
100	2000	0.05	0.0565	0.0410	0.0545	0.0355	0.0500	0.0595	0.0440	0.0530	(III)
100	100	1	0.0515	0.0545	0.0565	0.0520	0.0515	0.0520	0.0395	0.0420	
200	200	1	0.0540	0.0460	0.0485	0.0395	0.0475	0.0495	0.0450	0.0520	
400	400	1	0.0570	0.0565	0.0505	0.0450	0.0385	0.0420	0.0505	0.0510	
200	100	2	0.0530	0.0480	0.0560	0.0380	0.0560	0.0545	0.0370	0.0420	
400	200	2	0.0480	0.0500	0.0510	0.0420	0.0545	0.0515	0.0470	0.0390	
800	400	2	0.0505	0.0485	0.0480	0.0520	0.0475	0.0470	0.0405	0.0445	
5	500	0.01	0.0630	0.0715	0.0585	0.0665	0.0670	0.0560	0.0650	0.0585	
10	1000	0.01	0.0680	0.0645	0.0695	0.0580	0.0555	0.0540	0.0545	0.0565	
20	2000	0.01	0.0590	0.0545	0.0575	0.0540	0.0655	0.0520	0.0635	0.0560	
25	500	0.05	0.0510	0.0545	0.0505	0.0505	0.0635	0.0590	0.0595	0.0580	
50	1000	0.05	0.0435	0.0425	0.0475	0.0405	0.0550	0.0555	0.0535	0.0465	
100	2000	0.05	0.0480	0.0460	0.0470	0.0420	0.0600	0.0460	0.0595	0.0520	(IV)
100	100	1	0.0500	0.0525	0.0455	0.0455	0.0545	0.0485	0.0595	0.0530	
200	200	1	0.0510	0.0530	0.0530	0.0505	0.0495	0.0460	0.0480	0.0520	
400	400	1	0.0535	0.0495	0.0530	0.0390	0.0450	0.0440	0.0510	0.0520	
200	100	2	0.0550	0.0545	0.0480	0.0605	0.0480	0.0485	0.0415	0.0450	
400	200	2	0.0470	0.0485	0.0540	0.0525	0.0545	0.0525	0.0460	0.0520	
800	400	2	0.0415	0.0505	0.0450	0.0495	0.0480	0.0490	0.0510	0.0495	

in the Supplementary Material (Li et al. (2019)) gives these empirical powers for $a = 0.1$ and various combinations (p, T).

From Table 1, we know that the two classic tests become seriously biased when the dimension p is large compared to the sample size T. Their sizes approach zero when p/T becomes larger. From Table A of Li et al. (2019), we see that due to such biased critical values used in $\widetilde{Q}_q$ and Q_q^*, their powers are driven downward. This is particularly severe when the ratio p/T is larger than 0.5.

To explore more of these two traditional tests, we also examine their *intrinsic powers* when $\Sigma_0 = I_p$. Namely, we empirically find the 95 percentiles of $\widetilde{Q}_q$ and Q_q^* under the null and use these values as the corrected critical value for the power comparison. Empirical values are reported in Table B of the Supplementary Material (Li et al. (2019)). It is interesting to observe that after such correction, both $\widetilde{Q}_q$

TABLE 3
Test power of our tests G_q and $G_{q,1}$ under VAR(1)

				Gaussian (I)				Non-Gaussian (II)				
				G_q		$G_{q,1}$		G_q		$G_{q,1}$		
p	T	p/T	a	$q=1$	$q=3$	$q=1$	$q=3$	$q=1$	$q=3$	$q=1$	$q=3$	
5	500	0.01	0.05	0.2355	0.1535	0.2500	0.1540	0.2485	0.1475	0.2465	0.1505	
10	1000	0.01	0.05	0.5280	0.2770	0.5335	0.2935	0.5135	0.2645	0.5265	0.2930	
20	2000	0.01	0.05	0.9460	0.6620	0.9495	0.6995	0.9355	0.6010	0.9500	0.6670	
25	500	0.05	0.05	0.2260	0.1300	0.2485	0.1770	0.2315	0.1395	0.2585	0.1810	
50	1000	0.05	0.05	0.5410	0.2800	0.5995	0.3785	0.5105	0.2495	0.5960	0.3750	
100	2000	0.05	0.05	0.9580	0.6550	0.9815	0.8275	0.9500	0.5895	0.9805	0.8385	(III)
100	100	1	0.1	0.2615	0.2205	0.6170	0.8190	0.2100	0.1750	0.6165	0.8285	
200	200	1	0.1	0.6010	0.4720	0.9870	0.9995	0.4460	0.3370	0.9865	1	
400	400	1	0.1	0.9745	0.9230	1	1	0.9025	0.7875	1	1	
200	100	2	0.1	0.3275	0.2710	0.9375	0.9980	0.2420	0.2135	0.9390	0.9995	
400	200	2	0.1	0.7415	0.6745	1	1	0.5715	0.4830	1	1	
800	400	2	0.1	0.9995	0.9930	1	1	0.9710	0.9350	1	1	
5	500	0.01	0.05	0.2540	0.1680	0.2590	0.1700	0.2355	0.1505	0.2450	0.1615	
10	1000	0.01	0.05	0.4650	0.2870	0.4730	0.2850	0.4650	0.2885	0.4825	0.2970	
20	2000	0.01	0.05	0.8750	0.6170	0.8815	0.6285	0.8880	0.5980	0.8950	0.6190	
25	500	0.05	0.05	0.2580	0.1630	0.2555	0.1710	0.2475	0.1415	0.2655	0.1750	
50	1000	0.05	0.05	0.5215	0.2650	0.5525	0.3110	0.5165	0.2575	0.5450	0.3270	
100	2000	0.05	0.05	0.9450	0.6500	0.9555	0.7320	0.9345	0.6240	0.9635	0.7405	(IV)
100	100	1	0.1	0.2145	0.1690	0.3700	0.4695	0.1970	0.1470	0.3765	0.4495	
200	200	1	0.1	0.4910	0.3470	0.8335	0.9005	0.4355	0.2935	0.8430	0.9150	
400	400	1	0.1	0.9205	0.7690	1	1	0.8655	0.6735	1	1	
200	100	2	0.1	0.2450	0.2035	0.6255	0.8115	0.2240	0.1745	0.6425	0.8235	
400	200	2	0.1	0.5815	0.4790	0.9915	1	0.5000	0.3770	0.9880	1	
800	400	2	0.1	0.9705	0.9205	1	1	0.9425	0.8525	1	1	

and Q_q^* show very reasonable powers which all increase to 1 when the dimension and the sample size increases. Our test statistics G_q and $G_{q,1}$ also maintain comparably high power in all the tested (p, T) combinations. Table 3 demonstrates the feasibility of our test statistics under both high and low dimension cases. Interestingly enough, $G_{q,1}$ shows slightly better power than G_q under the present AR(1) alternative which is not intuitive. A comparison with the Hosking and the Li–McLeod tests sheds new light on the superiority of our test statistics in both low- and high-dimensional cases.

3.3. *Why both the Hosking and the Li–McLeod tests fail in high dimension.* The experiments here are designed to explore the reasons behind the failure of the Hosking and the Li–McLeod tests in high dimension. For the test statistics $\widetilde{Q}_q$ and Q_q^* as well as our test statistic ϕ_τ, we consider their empirical mean, variance and the 95% quantile, say θ_{emp}, with their theoretical values predicted by their respective asymptotic distributions (denoted as θ_{theo}). As for the two classical tests, we have often observed very large discrepancy between these values so it is more convenient to report the corresponding relative errors $(\theta_{\mathrm{theo}} - \theta_{\mathrm{emp}})/\theta_{\mathrm{emp}}$ (in percentage). Empirical values are reported in Table C of the Supplementary Material (Li et al. (2019)). It clearly appears from this table that for both statistics $\widetilde{Q}_q$ and Q_q^*, the traditional asymptotic theory severely overestimated their variances, that is, their empirical means are close to the degree of freedom $p^2(q-u-v)$ of the asymptotic chi-squared distribution while their empirical variances are much smaller than $2p^2(q-u-v)$ as suggested by the same chi-squared limit. This leads to an inflated 95th percentile which, although in a lesser proportion, is enough to create a high downward-bias in the empirical sizes of these two classical tests with high-dimensional data; see Table 1.

3.4. *Comparison with other test statistics.* In this section, we compare our test statistics with some others in recent literature. Chang, Yao and Zhou (2017) proposed an omnibus test for vector white noise using the maximum absolute autocorrelations and cross-correlations of the component series. Let

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

be the sample autocorrelation matrix at lag k, where $\hat{\Sigma}(k) = \frac{1}{T}\sum_{t=1}^{T-k}\mathbf{x}_{t+k}\mathbf{x}_t^*$. Their test statistic T_n is defined as

$$T_n = \max_{1\le k\le q} T_{n,k},$$

where $T_{n,k} = \max_{1\le i,j\le p} T^{1/2}|\hat{\rho}_{ij}(k)|$. Another test statistic T_n^* is defined in the same manner as T_n, only that the time series principal component analysis proposed by Chang, Guo and Yao (2015) is applied to the data $\{\mathbf{x}_t\}$ first.

Here we fix $p = 20$, $T = 100$ and adopt the spherical AR(1) process for power comparison, that is, $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{y}_t$, $\mathbf{y}_t = A\mathbf{y}_{t-1} + \mathbf{z}_t$, $A = a\mathbf{I}_p$, where $\mathbf{z}_t$ and Σ_0 follow different combinations of settings. Power values of all the five test statistics, that is, G_q, $G_{q,1}$, $G_{q,1}^*$, T_n and T_n^*, are compared when VAR coefficient a grows from 0 to 0.5. Here $G_{q,1}^*$ is our test statistic with finite sample correction as demonstrated in (2.11). Empirical statistics are obtained using 2000 independent replicates. Results are shown in Figure 1. Notice that on these displays, $G_{q,1}$ and $G_{q,1}^*$ coincide almost everywhere showing a high accuracy of the parameter estimates used in $G_{q,1}^*$.

It can be seen that our test statistics show better performance under this spherical AR(1) model setting. Designed via Frobenius norm of sample autocovariance matrices, the strength of our test statistics are fully demonstrated in such VAR(1) settings. While T_n and T_n^* are more adapted to settings where the majority coordinates of the test sequence $\mathbf{x}_t$ or their linear transformations remain to be white

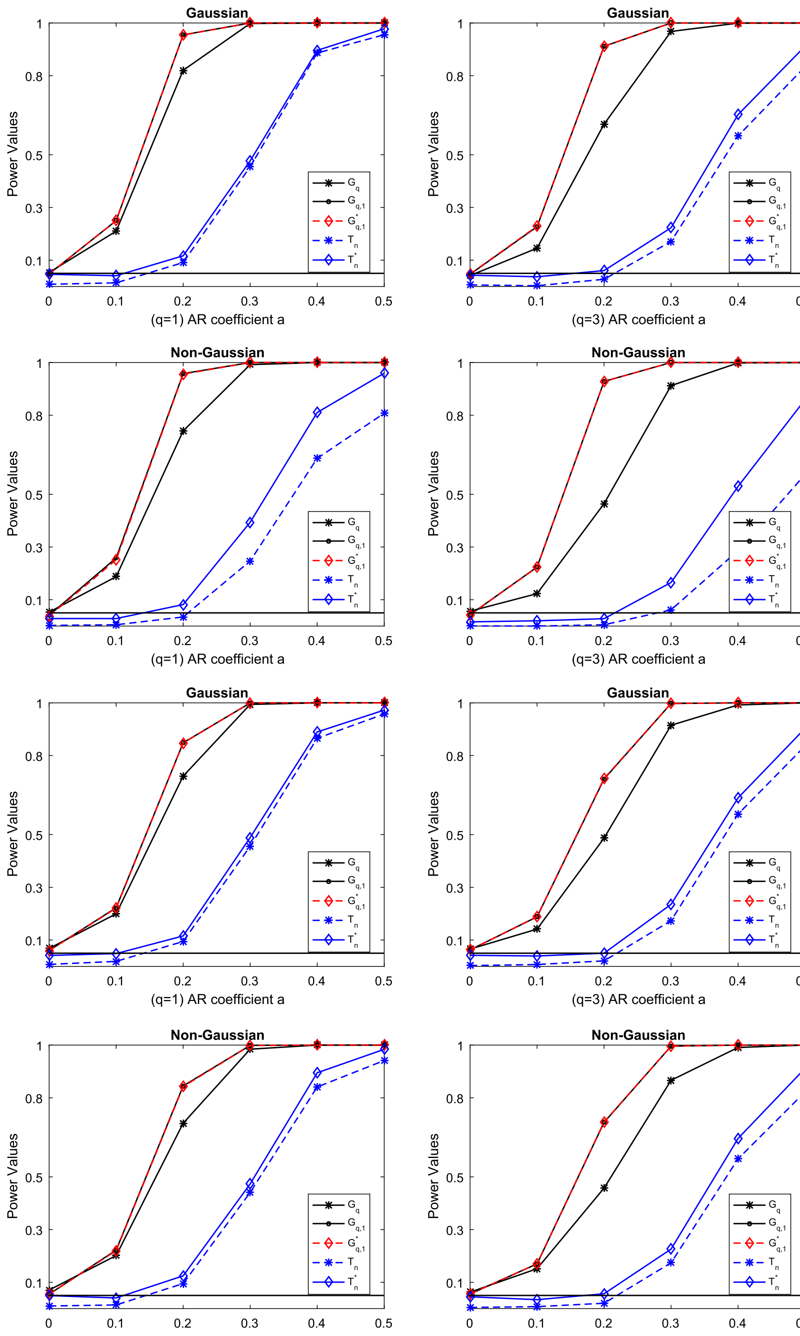

FIG. 1. *Power comparison under* VAR(1) *with* $(p, T) = (20, 100)$. *Left column with* $q = 1$ *and right column with* $q = 3$. *First two rows under alternative model* (III); *last two rows under alternative model* (IV).

noise; see the model settings in Chang, Yao and Zhou (2017). Moreover, it can be seen that test size of T_n is a little biased when $p = 20$, $T = 100$. Actually, such bias appears to be more significant when we increase the dimension-to-sample ratio p/T to a relative higher level, say 0.5. On the contrary, our test statistics maintain the nominal level accurately in both low- and high-dimensional settings. T_n^* shows very resilient powerful performance while it is quite time-consuming due to its relatively complicated bootstrap procedures. All in all, our test statistics G_q, $G_{q,1}$ and $G_{q,1}^*$ provide very satisfactory alternatives for high-dimensional diagnostic checking.

3.5. *Performance under* VMA(1) *model.* In this section, we compare performance of the tests when $\mathbf{x}_t$ follows a vector moving average process of order 1, that is,

$$\mathbf{x}_t = A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}.$$

We use different settings for $\mathbf{z}_t$ and A_0, A_1, respectively, to compare our test statistic G_q as defined in (2.5) and $G_{q,1}$ in (2.8) under nominal level $\alpha = 5\%$.

As for $\mathbf{z}_t$, we use the same two models as defined in (I) and (II) in Section 3.1. As for A_0 and A_1, we use two different models as follows:

(V) $A_0 = \mathbf{I}_p$ and $A_1 = a\mathbf{I}_p$, $0 < a < 1$.

(VI) $A_0 = \mathbf{I}_p$ and for $0 < r < 1$, take $d = [pr]$. Here $[\cdot]$ means to take the closest integer to the given value. $A_1 = (\frac{4}{p}E_0E_0^*)^{1/2}$, where E_0 is $p \times d$ matrix with entries $e_{ij} \sim U(-1, 1)$ i.i.d., thus $\operatorname{rank}(A_1) \le d < p$.

To evaluate the performance of our test statistics G_q and $G_{q,1}$ under VMA(1) models, we assign $a = 0.07$ and $r = 0.01$, $d = \max(1, [pr])$, respectively, for Scenario (V) and (VI). Testing power of G_q and $G_{q,1}$ are shown in Table 4 for $q = 1$ under various (p, T) combinations. The asymptotic power $\beta(G_{1,1})$ of the test statistic $G_{1,1}$ derived in Proposition 2.3 are also listed for comparison. All empirical results are obtained using 2000 independent replicates.

Similarly, as in Section 3.4, we further compare our test statistics with others, that is, T_n and T_n^* in Chang, Yao and Zhou (2017) under the VMA(1) settings. Here we fix $p = 20$, $T = 100$ and let $\mathbf{x}_t = A_0\mathbf{z}_t + A_1\mathbf{z}_{t_1}$ where A_1 follows model (V) or (VI) and $\mathbf{z}_t$ is either Gaussian or non-Gaussian. Power values of all the five test statistics, that is, G_q, $G_{q,1}$, $G_{q,1}^*$, T_n and T_n^*, are compared under model (V) and (VI) separately. Figure 2 shows the results under model (V) when coefficient a of A_1 grows from 0 to 0.5 (top rows), and for model (VI) when parameter r varies from 0 to 0.5 (bottom rows). All results are based on 2000 independent experiments.

From Table 4, it can be seen that our test statistics G_1 and $G_{1,1}$ consistently show reasonable powers for various (p, T) combinations under both VMA(1) model settings. Especially $G_{1,1}$ performs surprisingly well under VMA model (VI) even when $d(\operatorname{rank}(A_1))$ is very small. Meanwhile, the empirical power of $G_{1,1}$ is consistent with the asymptotic values $\beta(G_{1,1})$ derived in Proposition 2.3. As for

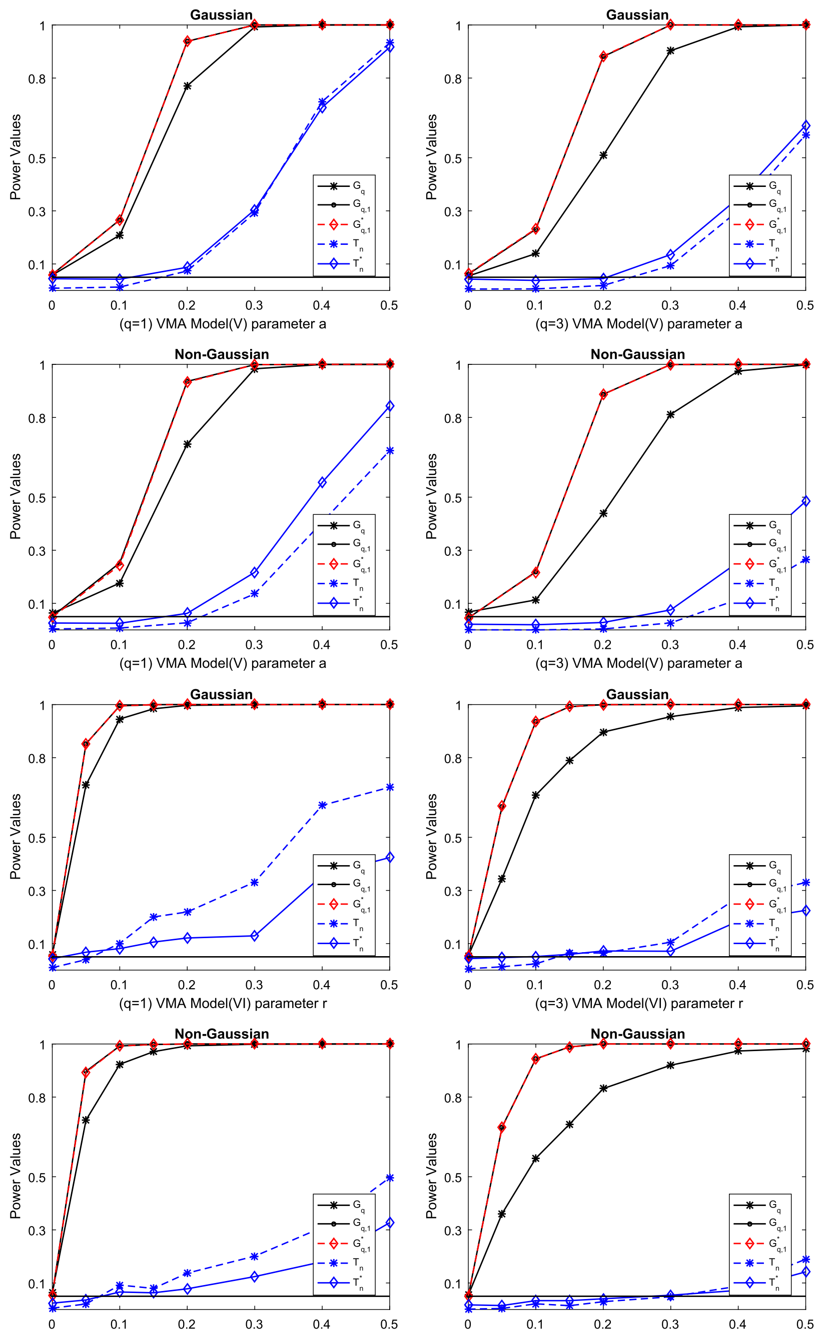

FIG. 2. *Power comparison under* VMA(1) *with* $(p, T) = (20, 100)$. *Left column with* $q = 1$ *and right column with* $q = 3$. *First two rows under alternative model* (V); *last two rows under alternative model* (VI).

TABLE 4
Test power of our tests G_1 and $G_{1,1}$ under VMA(1)

p	T	p/T	a	Gaussian (I) G_1	Gaussian (I) $G_{1,1}$	Gaussian (I) $\beta(G_{1,1})$	Non-Gaussian (II) G_1	Non-Gaussian (II) $G_{1,1}$	Non-Gaussian (II) $\beta(G_{1,1})$	
10	200	0.05	0.07	0.2085	0.2260	0.2144	0.1865	0.1990	0.2159	
20	400	0.05	0.07	0.4135	0.4410	0.4530	0.3805	0.4315	0.4548	
40	800	0.05	0.07	0.8350	0.8985	0.8903	0.7910	0.8885	0.8910	
20	200	0.1	0.07	0.1830	0.2120	0.2235	0.1755	0.2165	0.2250	
40	400	0.1	0.07	0.3915	0.4925	0.5015	0.3605	0.4800	0.5034	
80	800	0.1	0.07	0.8480	0.9395	0.9372	0.7995	0.9485	0.9377	
50	100	0.5	0.07	0.1185	0.1705	0.1790	0.1070	0.1730	0.1804	
100	200	0.5	0.07	0.2070	0.3820	0.3958	0.1600	0.3850	0.3977	(V)
200	400	0.5	0.07	0.4940	0.8395	0.8521	0.3660	0.8400	0.8531	
100	100	1	0.07	0.1305	0.2670	0.2754	0.1120	0.2715	0.2771	
200	200	1	0.07	0.2540	0.6605	0.6485	0.1925	0.6470	0.6502	
400	400	1	0.07	0.5520	0.9900	0.9903	0.4110	0.9895	0.9904	
200	100	2	0.07	0.1510	0.4990	0.5157	0.1225	0.5000	0.5177	
400	200	2	0.07	0.3005	0.9480	0.9500	0.2385	0.9500	0.9504	
800	400	2	0.07	0.7310	1	0.9999	0.5500	1	0.9999	
p	T	p/T	r	G_1	$G_{1,1}$	$\beta(G_{1,1})$	G_1	$G_{1,1}$	$\beta(G_{1,1})$	
10	200	0.05	0.01	0.9955	0.9995	0.9838	1	1	0.9884	
20	400	0.05	0.01	1	1	0.9995	1	1	0.9994	
40	800	0.05	0.01	1	1	0.9999	1	1	0.9999	
20	200	0.1	0.01	0.9700	0.9935	0.9705	0.9875	0.9980	0.9815	
40	400	0.1	0.01	0.9995	1	0.9970	0.9980	1	0.9979	
80	800	0.1	0.01	1	1	0.9999	0.9995	1	0.9999	
50	100	0.5	0.01	0.0530	0.0445	0.0500	0.0615	0.0510	0.0500	
100	200	0.5	0.01	0.3255	0.5855	0.6185	0.2765	0.5925	0.6155	(VI)
200	400	0.5	0.01	0.6080	0.9390	0.9439	0.5150	0.9565	0.9544	
100	100	1	0.01	0.1135	0.1910	0.2132	0.1110	0.2575	0.2759	
200	200	1	0.01	0.2200	0.5665	0.5690	0.1780	0.5450	0.5770	
400	400	1	0.01	0.5110	0.9755	0.9709	0.3810	0.9620	0.9628	
200	100	2	0.01	0.0910	0.2430	0.2553	0.1020	0.2660	0.2772	
400	200	2	0.01	0.1625	0.5785	0.5972	0.1320	0.5575	0.5917	
800	400	2	0.01	0.3695	0.9755	0.9714	0.2615	0.9785	0.9746	

comparison with T_n and T_n^* in Figure 2, our test statistics in general show better performance under VMA(1) model settings. The test sizes of T_n and T_n^* are a little biased when $p = 20$, $T = 100$, especially for non-Gaussian cases, while our test statistics maintain the nominal level accurately and uphold higher detection power even when the signals are relatively weak.

4. Proofs.

4.1. *Proof of Theorem 2.1*. To derive the null distribution of G_q when $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$, we looked into the Free probability and moment method proposed by Bhattacharjee and Bose (2016). In Section 4.2.3 of Bhattacharjee and Bose (2016), they have proved the following result.

PROPOSITION 4.1. *Let $\Pi := \Pi(\widehat{\Sigma}_\tau, \widehat{\Sigma}_\tau^* : \tau \geq 0)$ be a symmetric polynomial in $\{\widehat{\Sigma}_\tau, \widehat{\Sigma}_\tau^* : \tau \geq 0\}$,*

$$\sigma_\Pi^2 = \lim \mathbb{E}\big(\mathrm{Tr}(\Pi) - \mathbb{E}(\mathrm{Tr}(\Pi))\big)^2.$$

They have

$$\lim \mathbb{E}\big(\mathrm{Tr}(\Pi) - \mathbb{E}(\mathrm{Tr}(\Pi))\big)^k = \begin{cases} 0 & \text{if } k = 2d-1, \\ \left(\prod_{l=1}^{d}(2d-2l+1)\right)\sigma_\Pi^{2d} & \text{if } k = 2d, \end{cases}$$

therefore, as $p, T \to \infty$, $c_p = p/T \to c \in (0, \infty)$,

$$\mathrm{Tr}(\Pi) - \mathbb{E}\mathrm{Tr}(\Pi) \xrightarrow{d} \mathcal{N}(0, \sigma_\Pi^2).$$

Since G_q is a symmetric polynomial in $\{\widehat{\Sigma}_\tau, \widehat{\Sigma}_\tau^* : \tau \geq 0\}$, its asymptotic normality directly results from the proposition above. It remains to determine its first two moments in order to get the null distribution. This is done in the following corollary which is a direct consequence of moment calculations presented in Section 1 of the Supplementary Material (Li et al. (2019)).

COROLLARY 4.1. *Let the assumptions for $\mathbf{z}_t$ in Theorem 2.1 hold. Under the framework $p/T \to c > 0$, assume that $\|\Sigma_0\| = O(1)$. Then as $p, T \to \infty$,*

$$\begin{aligned} &\mathbb{E}(G_q) \sim qp^2 s_1^2/T, \\ &\mathrm{Var}(G_q) \to qc^2\big(s_2^2 + b^2(s_2')^2\big) \\ &\qquad + 4q^2c^3(\nu_4 - b - 2)s_1^2 s_{d,2} + 8q^2c^3 s_1^2 s_{r,2} + 4q^2c^3(b-1)s_1^2 s_2', \end{aligned}$$

where $s_2' = \lim_{p\to\infty} \mathrm{Tr}(\Sigma_0 \Sigma_0^T)/p$, $s_{r,2} = \lim_{p\to\infty} \mathrm{Tr}(\Re^2(\Sigma_0))/p$.

If the z_{it}'s are real, then Σ_0 is real symmetric and $b = 1$, $s_2' = s_{r,2} = s_2$. The asymptotic formula for $\mathrm{Var}(G_q)$ then reduces to

$$2qc^2 s_2^2 + 4q^2c^3(\nu_4 - 3)s_1^2 s_{d,2} + 8q^2c^3 s_1^2 s_2,$$

which further reduces to $2qc^2s_2^2 + 8q^2c^3s_1^2s_2$ if all the z_{it}'s are Gaussian.

4.2. *Proof of Theorem* 2.2. The proof of Theorem 2.2 is similar to that of Theorem 2.1, while in this proof we only consider the real value cases. Both G_q and $p(\hat{s}_1^2 - s_1^2)$ are symmetric polynomials in $\{\widehat{\Sigma}_\tau, \widehat{\Sigma}_\tau^* : \tau \geq 0\}$, thus the asymptotic normality of any linear combinations of these two statistics have been proven by Proposition 4.1. We can directly calculate the first two moments and covariance of these two statistics to obtain the joint limiting distribution. By directly conducting moment calculations as in Section 1 of the Supplementary Material (Li et al. (2019)), we have the following proposition.

PROPOSITION 4.2. *Let the assumptions for* $\mathbf{z}_t$ *in Theorem* 2.1 *hold. Under the framework* $p/T \to c > 0$, *assume that* $\|\Sigma_0\| = O(1)$. *Then as* $p, T \to \infty$,

$$
\begin{aligned}
\mathbb{E}(p\hat{s}_1^2) &= \frac{1}{p}\mathrm{Tr}^2(\Sigma_0) + \frac{1}{pT}\big(2\mathrm{Tr}(\Sigma_0^2) + (\nu_4 - 3)\mathrm{Tr}(D^2(\Sigma_0))\big),\\
\mathrm{Var}(p\hat{s}_1^2) &= \frac{8}{p^2T}\mathrm{Tr}(\Sigma_0^2)\mathrm{Tr}^2(\Sigma_0) + \frac{4}{p^2T}(\nu_4 - 3)\mathrm{Tr}^2(\Sigma_0)\mathrm{Tr}(D(\Sigma_0))\\
&\quad + o\left(\frac{1}{T}\right),\\
\mathbb{E}(\hat{s}_2) &= \frac{1}{p}\mathrm{Tr}(\Sigma_0^2) + \frac{1}{pT}\mathrm{Tr}^2(\Sigma_0)\\
&\quad + \frac{1}{pT}\big(\mathrm{Tr}(\Sigma_0^2) + (\nu_4 - 3)\mathrm{Tr}(D^2(\Sigma_0))\big),\\
\mathbb{E}(G_q) &= \frac{q}{T}\mathrm{Tr}^2(\Sigma_0),\\
\mathrm{Var}(G_q) &= \frac{4q^2}{T^3}\mathrm{Tr}^2(\Sigma_0)\big(2\mathrm{Tr}(\Sigma_0^2) + (\nu_4 - 3)\mathrm{Tr}(D^2(\Sigma_0))\big) + \frac{2q}{T^2}\mathrm{Tr}^2(\Sigma_0^2)\\
&\quad + \frac{q}{T^3}\big(2\mathrm{Tr}(\Sigma_0^2) + (\nu_4 - 3)\mathrm{Tr}(D^2(\Sigma_0))\big)^2 + o\left(\frac{1}{T}\right),\\
\mathrm{Cov}(G_q,\ p\hat{s}_1^2) &= \frac{4q}{pT^2}\mathrm{Tr}^2(\Sigma_0)\big(2\mathrm{Tr}(\Sigma_0^2) + (\nu_4 - 3)\mathrm{Tr}(D^2(\Sigma_0))\big) + o\left(\frac{1}{T}\right).
\end{aligned}
$$

Results in Theorem 2.2 and Proposition 2.1 naturally follows from Proposition 4.2. The proof of Proposition 4.2 is postponed to Section 2 of the Supplementary Material (Li et al. (2019)).

4.3. *Proof of Theorem* 2.3. The proof of Theorem 2.3 is similar to that of Theorem 2.2 while the calculations are more complicated. When $\mathbf{x}_t = A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}$, both G_q and $p(\hat{s}_1^2 - s_1^2)$ are still symmetric polynomials in $\{\widehat{\Sigma}_\tau, \widehat{\Sigma}_\tau^* : \tau \geq 0\}$, thus the asymptotic normality of any linear combinations of these two statistics have been proven by Proposition 4.1. We can directly calculate the first two moments and covariance of these two statistics to obtain the joint limiting distribution.

To elucidate the calculations of moments, we implement the following decompositions on both G_q and $qTc_p^2\hat{s}_1^2$ when $\mathbf{x}_t = A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}$ for $q=1$. Actually,

$$\begin{aligned}
G_1 &= \frac{1}{T^2}\sum_{s,t=1}^{T}(A_0\mathbf{z}_s + A_1\mathbf{z}_{s-1})^*(A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}) \\
&\quad\times (A_0\mathbf{z}_{t-1} + A_1\mathbf{z}_{t-2})^*(A_0\mathbf{z}_{s-1} + A_1\mathbf{z}_{s-2}) \\
&= G(\mathrm{I}) + G(\mathrm{II}) + G(\mathrm{III}), \\
Tc_p^2\hat{s}_1^2 &= \frac{1}{T^3}\sum_{s,t=1}^{T}(A_0\mathbf{z}_s + A_1\mathbf{z}_{s-1})^*(A_0\mathbf{z}_s + A_1\mathbf{z}_{s-1}) \\
&\quad\times (A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1})^*(A_0\mathbf{z}_t + A_1\mathbf{z}_{t-1}) \\
&= S(\mathrm{I}) + S(\mathrm{II}) + S(\mathrm{III}),
\end{aligned}$$

where

$$\begin{aligned}
G(\mathrm{I}) &= \frac{1}{T^2}\sum_{s,t=1}^{T}\big(\mathbf{z}_s^*A_0^*A_0\mathbf{z}_t\mathbf{z}_{t-1}^*A_0^*A_0\mathbf{z}_{s-1} + \mathbf{z}_{s-1}^*A_1^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-2}^*A_1^*A_1\mathbf{z}_{s-2} \\
&\quad + \mathbf{z}_s^*A_0^*A_0\mathbf{z}_t\mathbf{z}_{t-2}^*A_1^*A_1\mathbf{z}_{s-2} + \mathbf{z}_{s-1}^*A_1^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-1}^*A_0^*A_0\mathbf{z}_{s-1}\big), \\
G(\mathrm{II}) &= \frac{1}{T^2}\sum_{s,t=1}^{T}\big(\mathbf{z}_s^*A_0^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-1}^*A_0^*A_0\mathbf{z}_{s-1} + \mathbf{z}_{s-1}^*A_1^*A_0\mathbf{z}_t\mathbf{z}_{t-1}^*A_0^*A_0\mathbf{z}_{s-1} \\
&\quad + \mathbf{z}_{s-1}^*A_1^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-1}^*A_0^*A_1\mathbf{z}_{s-2} + \mathbf{z}_{s-1}^*A_1^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-2}^*A_1^*A_0\mathbf{z}_{s-1} \\
&\quad + \mathbf{z}_s^*A_0^*A_0\mathbf{z}_t\mathbf{z}_{t-2}^*A_1^*A_0\mathbf{z}_{s-1} + \mathbf{z}_s^*A_0^*A_0\mathbf{z}_t\mathbf{z}_{t-1}^*A_0^*A_1\mathbf{z}_{s-2} \\
&\quad + \mathbf{z}_s^*A_0^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-2}^*A_1^*A_1\mathbf{z}_{s-2} + \mathbf{z}_{s-1}^*A_1^*A_0\mathbf{z}_t\mathbf{z}_{t-2}^*A_1^*A_1\mathbf{z}_{s-2}\big), \\
G(\mathrm{III}) &= \frac{1}{T^2}\sum_{s,t=1}^{T}\big(\mathbf{z}_s^*A_0^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-1}^*A_0^*A_1\mathbf{z}_{s-2} + \mathbf{z}_{s-1}^*A_1^*A_0\mathbf{z}_t\mathbf{z}_{t-2}^*A_1^*A_0\mathbf{z}_{s-1} \\
&\quad + \mathbf{z}_s^*A_0^*A_1\mathbf{z}_{t-1}\mathbf{z}_{t-2}^*A_1^*A_0\mathbf{z}_{s-1} + \mathbf{z}_{s-1}^*A_1^*A_0\mathbf{z}_t\mathbf{z}_{t-1}^*A_0^*A_1\mathbf{z}_{s-2}\big), \\
S(\mathrm{I}) &= \frac{1}{T^3}\sum_{s,t=1}^{T}\big(\mathbf{z}_t^*A_0^*A_0\mathbf{z}_t\mathbf{z}_s^*A_0^*A_0\mathbf{z}_s + \mathbf{z}_{t-1}^*A_1^*A_1\mathbf{z}_{t-1}\mathbf{z}_{s-1}^*A_1^*A_1\mathbf{z}_{s-1} \\
&\quad + \mathbf{z}_t^*A_0^*A_0\mathbf{z}_t\mathbf{z}_{s-1}^*A_1^*A_1\mathbf{z}_{s-1} + \mathbf{z}_{t-1}^*A_1^*A_1\mathbf{z}_{t-1}\mathbf{z}_s^*A_0^*A_0\mathbf{z}_s\big),
\end{aligned}$$

$$
\begin{aligned}
S(\mathrm{II}) = \frac{1}{T^2} \sum_{s,t=1}^{T} \big(& \mathbf{z}_t^* A_0^* A_1 \mathbf{z}_{t-1} \mathbf{z}_s^* A_0^* A_0 \mathbf{z}_s + \mathbf{z}_{t-1}^* A_1^* A_0 \mathbf{z}_t \mathbf{z}_s^* A_0^* A_0 \mathbf{z}_s \\
& + \mathbf{z}_{t-1}^* A_1^* A_1 \mathbf{z}_{t-1} \mathbf{z}_s^* A_0^* A_1 \mathbf{z}_{s-1} + \mathbf{z}_{t-1}^* A_1^* A_1 \mathbf{z}_{t-1} \mathbf{z}_{s-1}^* A_1^* A_0 \mathbf{z}_s \\
& + \mathbf{z}_t^* A_0^* A_0 \mathbf{z}_t \mathbf{z}_{s-1}^* A_1^* A_0 \mathbf{z}_s + \mathbf{z}_t^* A_0^* A_0 \mathbf{z}_t \mathbf{z}_s^* A_0^* A_1 \mathbf{z}_{s-1} \\
& + \mathbf{z}_t^* A_0^* A_1 \mathbf{z}_{t-1} \mathbf{z}_{s-1}^* A_1^* A_1 \mathbf{z}_{s-1} + \mathbf{z}_{t-1}^* A_1^* A_0 \mathbf{z}_t \mathbf{z}_{s-1}^* A_1^* A_1 \mathbf{z}_{s-1} \big), \\
S(\mathrm{III}) = \frac{1}{T^2} \sum_{s,t=1}^{T} \big(& \mathbf{z}_t^* A_0^* A_1 \mathbf{z}_{t-1} \mathbf{z}_s^* A_0^* A_1 \mathbf{z}_{s-1} + \mathbf{z}_{t-1}^* A_1^* A_0 \mathbf{z}_t \mathbf{z}_{s-1}^* A_1^* A_0 \mathbf{z}_s \\
& + \mathbf{z}_t^* A_0^* A_1 \mathbf{z}_{t-1} \mathbf{z}_{s-1}^* A_1^* A_0 \mathbf{z}_s + \mathbf{z}_{t-1}^* A_1^* A_0 \mathbf{z}_t \mathbf{z}_s^* A_0^* A_1 \mathbf{z}_{s-1} \big).
\end{aligned}
$$

By conducting moment calculations similar to Section 1 of the Supplementary Material (Li et al. (2019)), we have the following proposition.

PROPOSITION 4.3. *Let the assumptions in Theorem 2.3 hold, as* $p, T \to \infty$, $p/T \to c > 0$, *we have*

$$
\begin{aligned}
\mathbb{E}\big(G(I)\big) &= \frac{1}{T}\big[\mathrm{Tr}(\widetilde{\Sigma}_0 \widetilde{\Sigma}_1) + (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_0) D(\widetilde{\Sigma}_1)\big)\big] \\
&\quad + \frac{1}{T}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1) + \mathrm{Tr}(\widetilde{\Sigma}_0 \widetilde{\Sigma}_1), \\
\mathbb{E}\big(G(\mathrm{II})\big) &= 0, \qquad \mathbb{E}\big(G(\mathrm{III})\big) = \frac{2}{T}\mathrm{Tr}^2(\widetilde{\Sigma}_{01}), \\
\mathbb{E}\big(S(\mathrm{I})\big) &= \frac{1}{T^2}\big[2\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&\quad + \frac{1}{T}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1), \\
\mathbb{E}\big(S(\mathrm{II})\big) &= 0, \qquad \mathbb{E}\big(S(\mathrm{III})\big) = \frac{4}{T^2}\mathrm{Tr}\big(\widetilde{\Sigma}_{01} \widetilde{\Sigma}_{01}^*\big)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}\big(G(\mathrm{I})\big) = \frac{4}{T^3}\mathrm{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[& 2\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 \\
& + (\nu_4 - 3)\mathrm{Tr}\big(D^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
& + \frac{8}{T^2}\mathrm{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\mathrm{Tr}\big(\widetilde{\Sigma}_0 \widetilde{\Sigma}_1 (\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) \\
& + (\nu_4 - 3)\mathrm{Tr}\big(D(\widetilde{\Sigma}_0 \widetilde{\Sigma}_1) D(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big]
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{2}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_1^2) + \frac{6}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_1) \\
&+ \frac{4}{T}\big[2\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)^2 + (\nu_4 - 3)\operatorname{Tr}\big(D^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)\big)\big] + R_n,
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Var}\big(G(\mathrm{III})\big) = {}& \frac{4}{T}\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\big) + \frac{12}{T^2}\operatorname{Tr}^2\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) \\
&+ \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01})\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}^*\big) \\
&+ \frac{16}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_{01})\big[\operatorname{Tr}(\widetilde{\Sigma}_{01})^2 + 2\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) \\
&+ (\nu_4 - 3)\operatorname{Tr}\big(D^2(\widetilde{\Sigma}_{01})\big)\big] + R_n,
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Var}\big(G(\mathrm{II})\big) = {}& \frac{8}{T^2}\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big)\operatorname{Tr}\big(\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_1^2\big) + \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_1)\operatorname{Tr}(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0) \\
&+ \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[\operatorname{Tr}\big(\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0\big) + \operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1\big)\big] \\
&+ \frac{16}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01})\big[\operatorname{Tr}\big(\widetilde{\Sigma}_0^2\widetilde{\Sigma}_{01}^*\big) + \operatorname{Tr}\big(\widetilde{\Sigma}_1^2\widetilde{\Sigma}_{01}\big) \\
&+ 2\operatorname{Tr}(\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0)\big] \\
&+ \frac{4}{T}\operatorname{Tr}\big(\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0^2 + \widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1^2 + 2\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0\big) \\
&+ \frac{16}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\operatorname{Tr}\big(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\big) \\
&+ \frac{16}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_{01})\operatorname{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)^2 \\
&+ \frac{32}{T^3}\operatorname{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\operatorname{Tr}(\widetilde{\Sigma}_{01})\operatorname{Tr}\big(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) + R_n,
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Cov}\big(G(\mathrm{I}), G(\mathrm{III})\big) = {}& \frac{4}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_0\widetilde{\Sigma}_{01}) + \frac{4}{T^2}\operatorname{Tr}^2(\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}) \\
&+ \frac{8}{T^3}\operatorname{Tr}(\widetilde{\Sigma}_{01})\operatorname{Tr}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big[2\operatorname{Tr}\big(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big) \\
&+ (\nu_4 - 3)\operatorname{Tr}\big(D(\widetilde{\Sigma}_{01})D(\widetilde{\Sigma}_0 + \widetilde{\Sigma}_1)\big)\big] \\
&+ \frac{8}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_{01})\big[2\operatorname{Tr}(\widetilde{\Sigma}_0\widetilde{\Sigma}_1\widetilde{\Sigma}_{01}) \\
&+ (\nu_4 - 3)\operatorname{Tr}\big(D(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)D(\widetilde{\Sigma}_{01})\big)\big] + R_n,
\end{aligned}
$$

$$
\operatorname{Cov}\big(G(\mathrm{I}), G(\mathrm{II})\big) = o(1), \qquad \operatorname{Cov}\big(G(\mathrm{II}), G(\mathrm{III})\big) = o(1);
$$

$$
\begin{aligned}
\operatorname{Var}\bigl(S(\mathrm{I})\bigr) &= \frac{4}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigl[2\operatorname{Tr}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)^2 \\
&\quad + (\nu_4-3)\operatorname{Tr}\bigl(D^2(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigr)\bigr] + R_n, \\
\operatorname{Var}\bigl(S(\mathrm{II})\bigr) &= \frac{16}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\operatorname{Tr}\bigl(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\bigr) + R_n, \\
\operatorname{Var}\bigl(S(\mathrm{III})\bigr) &= \frac{32}{T^4}\operatorname{Tr}^2\bigl(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\bigr), \qquad \operatorname{Cov}\bigl(S(\mathrm{I}), S(\mathrm{II})\bigr) = o(1), \\
\operatorname{Cov}\bigl(S(\mathrm{I}), S(\mathrm{III})\bigr) &= o(1), \qquad \operatorname{Cov}\bigl(S(\mathrm{II}), S(\mathrm{III})\bigr) = o(1); \\
\operatorname{Cov}\bigl(G(\mathrm{I}), S(\mathrm{I})\bigr) &= \frac{4}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigl[2\operatorname{Tr}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)^2 \\
&\quad + (\nu_4-3)\operatorname{Tr}\bigl(D^2(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigr)\bigr] \\
&\quad + \frac{4}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigl[2\operatorname{Tr}\bigl(\widetilde{\Sigma}_0\widetilde{\Sigma}_1(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigr) \\
&\quad + (\nu_4-3)\operatorname{Tr}\bigl(D(\widetilde{\Sigma}_0\widetilde{\Sigma}_1)D(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigr)\bigr] + R_n, \\
\operatorname{Cov}\bigl(G(\mathrm{II}), S(\mathrm{II})\bigr) &= \frac{8}{T^2}\operatorname{Tr}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigl[\operatorname{Tr}\bigl(\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_{01}\widetilde{\Sigma}_0\bigr) + \operatorname{Tr}\bigl(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\widetilde{\Sigma}_1\bigr)\bigr] \\
&\quad + \frac{16}{T^3}\operatorname{Tr}^2(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\operatorname{Tr}\bigl(\widetilde{\Sigma}_{01}\widetilde{\Sigma}_{01}^*\bigr) \\
&\quad + \frac{16}{T^3}\operatorname{Tr}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\operatorname{Tr}(\widetilde{\Sigma}_{01})\operatorname{Tr}\bigl(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigr) + R_n, \\
\operatorname{Cov}\bigl(G(\mathrm{III}), S(\mathrm{I})\bigr) &= \frac{8}{T^3}\operatorname{Tr}(\widetilde{\Sigma}_{01})\operatorname{Tr}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigl[2\operatorname{Tr}\bigl(\widetilde{\Sigma}_{01}(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigr) \\
&\quad + (\nu_4-3)\operatorname{Tr}\bigl(D(\widetilde{\Sigma}_{01})D(\widetilde{\Sigma}_0+\widetilde{\Sigma}_1)\bigr)\bigr] + R_n, \\
\operatorname{Cov}\bigl(G(\mathrm{III}), S(\mathrm{III})\bigr) &= 0, \qquad \operatorname{Cov}\bigl(G(\mathrm{I}), S(\mathrm{II})\bigr) = o(1), \\
\operatorname{Cov}\bigl(G(\mathrm{I}), S(\mathrm{III})\bigr) &= 0, \\
\operatorname{Cov}\bigl(G(\mathrm{II}), S(\mathrm{I})\bigr) &= o(1), \qquad \operatorname{Cov}\bigl(G(\mathrm{II}), S(\mathrm{III})\bigr) = o(1), \\
\operatorname{Cov}\bigl(G(\mathrm{III}), S(\mathrm{II})\bigr) &= o(1).
\end{aligned}
$$

Here the R_n's are possibly different: they represent remainder terms with smaller orders than the others listed in each variance covariance items.

Theorem 2.3 naturally follows from Proposition 4.3.

SUPPLEMENTARY MATERIAL

Supplement to "On testing for high-dimensional white noise" (DOI: 10.1214/18-AOS1782SUPP; .pdf). This supplemental article contains some technical lemmas, the proof of Proposition 4.2 of the main article and some additional simulation results.

REFERENCES

BAI, Z., CHEN, J. and YAO, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52** 423–437. MR2791528

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870

BHATTACHARJEE, M. and BOSE, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist.* **44** 598–628. MR3476611

BICKEL, P. J. and GEL, Y. R. (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 711–728. MR2867455

CHANG, J., GUO, B. and YAO, Q. (2015). High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. *J. Econometrics* **189** 297–312. MR3414901

CHANG, J., YAO, Q. and ZHOU, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika* **104** 111–127. MR3626482

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100** 830–840. MR2201012

GUO, S., WANG, Y. and YAO, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika* **103** 889–903. MR3620446

HAN, F., LU, H. and LIU, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.* **16** 3115–3150. MR3450535

HAUFE, S., NOLTE, G., MUELLER, K. R. and KRÄMER, N. (2009). Sparse causal discovery in multivariate time series. In *Proceedings of the* 2008*th International Conference on Causality: Objectives and Assessment* **6** 97–106. JMLR.org.

HOSKING, J. R. M. (1980). The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.* **75** 602–608. MR0590689

HSU, N.-J., HUNG, H.-L. and CHANG, Y.-M. (2008). Subset selection for vector autoregressive processes using Lasso. *Comput. Statist. Data Anal.* **52** 3645–3657. MR2427370

JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I* 307–333. Eur. Math. Soc., Zürich. MR2334195

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663

LI, W. K. (2004). *Diagnostic Checks in Time Series*. CRC Press/CRC, Boca Raton.

LI, W. K. and MCLEOD, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. Ser. B* **43** 231–239. MR0626770

LI, Z., LAM, C., YAO, J. and YAO, Q. (2019). Supplement to "On testing for high-dimensional white noise." DOI:10.1214/18-AOS1782SUPP.

LIU, H., AUE, A. and PAUL, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140

LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718

SHOJAIE, A. and MICHAILIDIS, G. (2010). Discovering graphical Granger causality using the truncating LASSO penalty. *Bioinformatics* **26** 517–523.

STOCK, J. H. and WATSON, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroecon. Annu.* **4** 351–394.

STOCK, J. H. and WATSON, M. W. (1998). Diffusion Indexes (No. W6702). National Bureau of Economic Research.

STOCK, J. H. and WATSON, M. W. (1999). Forecasting inflation. *J. Monet. Econ.* **44** 293–335.

TSAY, R. (2017). Testing for serial correlations in high-dimensional time series via extreme value theory. Preprint.

WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916

YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis. Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554

ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263

Z. LI
DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
STATE COLLEGE, PENNSYLVANIA 16802
USA
E-MAIL: zxl278@psu.edu

C. LAM
Q. YAO
DEPARTMENT OF STATISTICS
LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
COLUMBIA HOUSE
HOUGHTON STREET
LONDON WC2A 2AE
UNITED KINGDOM
E-MAIL: c.lam2@lse.ac.uk
q.yao@lse.ac.uk

J. YAO
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
THE UNIVERSITY OF HONG KONG
POKFULAM ROAD
HONG KONG
E-MAIL: jeffyao@hku.hk