On testing a high-dimensional white noise

Zeng Li and **Jianfeng Yao**

Department of Statistics and Actuarial Science
The University of Hong Kong
e-mail: u3001205@hku.hk; jeffyao@hku.hk

Clifford Lam and **Qiwei Yao**

Department of Statistics
London School of Economics and Political Science
e-mail: c.lam2@lse.ac.uk; q.yao@lse.ac.uk

Abstract: Testing for white noise is a classical yet important problem in statistics, especially for diagnostic checks in time series modeling. For vector time series where the dimension is large compared to the sample size, this paper demonstrates that popular omnibus portmanteau tests such as the multivariate Hosking and Li-McLeod tests become extremely conservative, losing their size and power dramatically. There is thus an urgent need to develop new tests for testing a high-dimensional white noise. Several new tests are proposed to fill in this gap. One is a new portmanteau test with a scalar test statistic which encapsulates the serial correlations within and across all components. Precisely, the statistic equals to the sum of squares of the eigenvalues in a symmetrized sample auto-covariance matrix at a certain lag. Other multiple-lags based tests are also proposed to complement the single-lag based one. We develop adequate limiting distributions for these test statistics using tools from random matrix theory. Asymptotic normality for the test statistics is derived under different asymptotic regimes when both the dimension p and the sample size T are diverging to infinity. We prove that such high-dimensional limits are valid for a significant range of finite (p, T) combinations, therefore ensuring a wide range of applications in practice. Extensive simulation experiments confirm an excellent behavior of these high-dimensional tests in finite samples with accurate size and satisfactory power. In particular, the new tests are consistently more powerful than the Hosking and Li-McLeod tests even when the latter two have been size-adjusted.

AMS 2000 subject classifications: Primary 62M10, 62H15; secondary 15A52.
Keywords and phrases: large auto-covariance matrix, Hosking's test, Li-McLeod test, high-dimensional time series, random matrix theory.

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 the concerned vector time series is comparable to or even larger than the sample size. One prominent example showing the need for diagnostic checking in high-dimensional time series concerns the vector autoregressive model, which has a large literature. When dimension is large, most existing works regularize the fitted models by Lasso (Hsu et al., 2008; Haufe et al., 2009; Shojaie and Michailidis, 2010; Basu and Michailidis, 2015), Dantzig penalization (Han and Liu, 2013), banded auto-covariances (Bickel and Gel, 2011), or banded auto-coefficient matrices (Guo et al., 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, Bai and Ng (2002), Forni *et al.* (2005), Lam and Yao (2012) and Chang *et al.* (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 et al., 1981; Lütkepohl, 2005). However, those methods are designed for the cases where the dimension of time series is small or relatively small compared to the sample size (i.e., the observed length of the time series). For the purpose of model diagnostic checking, the so-called omnibus tests are often adopted as the goal is 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 widely known in the literature that the slow convergence to their asymptotic null distributions is particularly pronounced in multivariate cases.

To understand better the challenge of testing for a high-dimensional white noise addressed in this paper, let us consider an example where some multivariate volatility model is to be fit to 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 example above is just one more illustration of the following fact which is now better understood in the statistical literature: many popular tools in multivariate statistics are severely challenged by the emergence of high-dimensional data, and they need to be *re-examined* or *corrected*. Recent advances in high-dimensional statistics demonstrate that random matrix theory provides powerful inference tools via a precise spectral analysis of large sample covariance or sample auto-covariance matrices. For a review on such progress, we refer to the review papers Johnstone (2007), Paul and Aue (2014) and the recent monograph Yao et al. (2015). In particular, asymptotic results found in this context using random matrix theory have quite a fast convergence rate, and hence provide satisfactory approximation for data analysis in finite sample situations.

This paper proposes several new tests for testing high-dimensional white noise. One such test is a scalar which encapsulates the serial correlations within and across all components. Precisely, the statistic equals to the sum of squares of the eigenvalues in the symmetrized sample auto-covariance matrix at a certain lag. Using random matrix theory, asymptotic normality for the test statistic is derived under different asymptotic regimes when both p and T can be large. While this proposed test is extremely powerful with very accurate size for a wide array of combinations of (p, T), it can only tests for one lag at a time. To complement this test, we propose other tests which can assimilate information from different lags in the time series. Extensive simulation experiments confirm an excellent behavior of these high-dimensional tests in finite samples with very accurate sizes and satisfactory powers. In particular, the new tests are consistently more powerful than the Hosking and Li-McLeod tests even when the latter two have been adjusted in a way such that their empirical sizes coincide with the nominal level; See Table 3.

The rest of the paper is organized as follows. Section 2 and 3 present the main contributions of the paper. Our high-dimensional tests for white noise are introduced and their asymptotic normality established. The proofs of these results are postponed to Section 7. Section 4 reports on 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, with detailed explanations in why these two multivariate tests fail when applied to high-dimensional data. In Section 5, we provide an in-depth discussion on the extremely challenging situation for testing a high-dimensional white noise when the covariance matrix of the noise is completely arbitrary. Some open questions and a conjecture are also introduced. Section 6 concludes while Section 7 collects all the technical proofs of the paper.

2. Single-lag based tests

Let $\mathbf{x}_1, \cdots, \mathbf{x}_T$ be observations from a p-dimensional weakly stationary time series satisfying

$$\mathbf{x}_t = B^{1/2} \sum_{l \geq 0} A_l \mathbf{z}_{t-l}, \tag{2.1}$$

where $\{\mathbf{z}_t\}$ is a sequence of independent $p-$dimensional random vectors with independent components $\mathbf{z}_t = (z_{it})$ satisfying $\mathbb{E}z_{it} = 0$, $\mathbb{E}|z_{it}|^2 = 1$, $\mathbb{E}|z_{it}|^4 < \infty$. Hence $\{\mathbf{x}_t\}$ has $\mathbb{E}\mathbf{x}_t = \mathbf{0}$, and its lag-$\tau$ auto-covariance matrix $\Sigma_\tau = \mathrm{Cov}(\mathbf{x}_{t+\tau}, \mathbf{x}_t)$ depends on τ only. In particular, $\Sigma_0 = \mathrm{var}(\mathbf{x}_t)$ denotes the population covariance matrix of the series. The goal is to test whether $\mathbf{x}_t$ is a white noise, and for this purpose we test the hypothesis

$$H_0: \quad \mathrm{Cov}(\mathbf{x}_{t+\tau}, \mathbf{x}_t) = \mathbf{0}, \ \tau = 1, \cdots, q, \tag{2.2}$$

where $q \geq 1$ is a prescribed constant integer.

Throughout the paper, the complex adjoint of a matrix (or vector) A is denoted by A^*. Let $\widehat{\Sigma}_\tau$ be the *lag*$-\tau$ *sample auto-covariance matrix*

$$\widehat{\Sigma}_\tau = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_{t-\tau}^*, \tag{2.3}$$

which is the sample counterpart of Σ_τ. Here for convenience, we set $\mathbf{x}_t = \mathbf{x}_{T+t}$ when $t \leq 0$. Since $\widehat{\Sigma}_\tau$ is not symmetric, and in a high-dimensional setting where the dimension p is large, its spectral property is better understood by considering the *symmetrized lag*$-\tau$ *sample auto-covariance matrix*

$$\widetilde{M}_\tau = \frac{1}{2}\left(\widehat{\Sigma}_\tau + \widehat{\Sigma}_\tau^*\right) = \frac{1}{2T} \sum_{t=1}^{T} \left(\mathbf{x}_t \mathbf{x}_{t-\tau}^* + \mathbf{x}_{t-\tau} \mathbf{x}_t^*\right). \tag{2.4}$$

Under the null hypothesis, $\mathbb{E}\widetilde{M}_\tau = 0$ for $1 \le \tau \le q$, and a sensible test statistic is its squared Frobenius norm

$$\widetilde{L}_\tau = \sum_{j=1}^{p} \lambda_{j,\tau}^2 = \mathrm{Tr}(\widetilde{M}_\tau^* \widetilde{M}_\tau), \tag{2.5}$$

where $\{\lambda_{j,\tau},\ j = 1, \cdots, p\}$ are the eigenvalues of $\widetilde{M}_\tau$. Define the scaled statistic

$$\phi_\tau = \frac{T}{p}\widetilde{L}_\tau - \frac{p}{2}. \tag{2.6}$$

The null hypothesis will be rejected for large values of ϕ_τ, for some $1 \le \tau \le q$.

2.1. High dimensional asymptotics when $\mathbf{\Sigma_0} = \mathbf{I}_p$

First we consider high-dimensional situations where the dimension p is large compared to the sample size T. Here we assume the so-called *Marčenko-Pastur regime* for asymptotic analysis, which is $c_p = p/T \to c > 0$ when $p, T \to \infty$. However, most of the results in this area concern sample covariance matrices while our test statistic ϕ_τ is based on the sample auto-covariance matrices, which are much less studied. Only a few related papers have appeared in the last few years. See Johnstone (2007), Paul and Aue (2014) and the recent monograph Yao et al. (2015).

As a main contribution of the paper, we characterize the asymptotic distribution of ϕ_τ in this high-dimensional setting.

Theorem 2.1. *Let $\tau \ge 1$ be a fixed integer, and assume that*

1. *$\{z_{it},\ i = 1, \cdots, p,\ t = 1, \cdots, T\}$ are all independently distributed satisfying $\mathbb{E}z_{it} = 0$, $\mathbb{E}z_{it}^2 = 1$, $\mathbb{E}z_{it}^4 = \nu_4 < \infty$;*
2. *(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 in the simplest setting when $\mathbf{x}_t = \mathbf{z}_t$, the limiting distribution of the test statistic $\widetilde{L}_\tau$ is

$$\phi_\tau \xrightarrow{d} \mathcal{N}\left(\frac{1}{2},\ 1 + \frac{3(\nu_4 - 1)}{2}c\right). \tag{2.7}$$

The proof of this theorem is given in Sections 7.1 and 7.2.

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 (2.2) as follows.

$$\textit{Single Lag-}\tau\textit{ test:}\quad \textit{Reject } H_0 \ \ \textit{if}\ \ \phi_\tau - \frac{1}{2} > Z_\alpha\left(1 + \frac{3(\nu_4 - 1)}{2}c_p\right)^{1/2}. \tag{2.8}$$

As it will be demonstrated in Section 4.2, the test above is much more powerful compared to some classical alternatives, especially in the high dimensional setting where $p/T \to c > 0$. The power of this test comes from gathering information from the eigenvalues in the definition of $\widetilde{L}_\tau$, and is realized from the fact that the asymptotic mean of $\widetilde{L}_\tau$ is $c(Tc+1)/2$ under the high dimensional setting, which grows linearly with T (and p), while the asymptotic variance of the statistic is $c^2\left(1+\frac{3(\nu_4-1)}{2}c\right)$ which is just a constant. It means that when T is large, departure from white noise in the τ-th lag of auto-covariance matrix will likely results in a very large and different mean, which will be a lot of standard deviations away from $c(Tc+1)/2$ since the standard deviation is just a constant.

2.2. *Low dimensional asymptotics when* $\mathbf{\Sigma_0} = \mathbf{I}_p$

Formally, the Marčenko-Pastur regime from the previous section where $p/T \to c > 0,\ p, T \to \infty$ does not apply to the case of $c = 0$, that is, both p, T tend to infinity with $p/T \to 0$. From a practical point of view, such an asymptotic regime will be useful when the dimension p is much smaller than the sample size T. Hereafter, this will be referred to as the *low-dimensional* situation. The result below establishes the asymptotic distribution of the test statistic ϕ_τ in this setting.

Theorem 2.2. *Let $\tau \geq 1$ be a fixed integer, and assume that*

1. *$\{z_{it},\ i = 1, \cdots, p,\ t = 1, \cdots, T\}$ are all independently distributed satisfying $\mathbb{E}z_{it} = 0$, $\mathbb{E}z_{it}^2 = 1$, $\mathbb{E}z_{it}^4 = \nu_4 < \infty$;*
2. *Both the dimension p and the sample size T tend to infinity in a related way such that as $p, T \to \infty$, $p/T \to 0$, $p^3/T = O(1)$.*

Then in the simplest setting when $\mathbf{x}_t = \mathbf{z}_t$, the limiting distribution of the test statistic $\widetilde{L}_\tau$ is

$$\phi_\tau \xrightarrow{d} \mathcal{N}\left(\frac{1}{2},\ 1\right). \tag{2.9}$$

This theorem is proved in Section 7.3. It is worth noting that technically, the proof under this low-dimensional setting is very different from the proof of Theorem 2.1 under the Marčenko-Pastur regime. Indeed, new results from random matrix theory are needed to establish this low-dimensional asymptotics. The proof is also different from the classical *large sample* asymptotics where the limiting results are derived by tending T to infinity while keeping the dimension p fixed.

2.3. *A unified test procedure when* $\mathbf{\Sigma_0} = \mathbf{I}_p$

As mentioned earlier, the asymptotic distributions for the test statistic ϕ_τ are derived in Theorem 2.1 and Theorem 2.2 under two different asymptotic regimes and using completely different technical tools. Yet it is striking to observe that these two asymptotic distributions are *self-consistent* in the following sense. Recall that in Theorem 2.1 under the high-dimensional scheme where $p, T \to \infty$ and $c_p = p/T \to c > 0$, it has been found that

$$\phi_\tau \xrightarrow{d} \mathcal{N}\left(\frac{1}{2},\ 1 + \frac{3(\nu_4-1)}{2}c\right).$$

In the whole derivation of this result, it is required that the limiting ratio c should be positive. Indeed, the case with $c = 0$ corresponds to the low-dimensional limit which is derived in Theorem 2.2 using quite a different technique. However, if we let $c = 0$ in the high-dimensional limit above, we found easily that

$$\phi_\tau \xrightarrow{d} \mathcal{N}\left(\frac{1}{2},\ 1\right),$$

which is exactly the low-dimensional result derived in Theorem 2.2.

In other words, both theorems are compatible with each other and express the same type of limiting distribution, a property we qualify as *self-consistency*. As a consequence, we can combine them in a unified result as follows.

Theorem 2.3. *Let $\tau \geq 1$ be a fixed integer, assume that*

1. *$\{z_{it},\ i = 1, \cdots, p,\ t = 1, \cdots, T\}$ are all independently distributed satisfying $\mathbb{E}z_{it} = 0,\ \mathbb{E}z_{it}^2 = 1,\ \mathbb{E}z_{it}^4 = \nu_4 < \infty$;*
2. *Either "$p, T \to \infty,\ c_p := p/T \to c > 0$", or "$p, T \to \infty,\ p/T \to 0,\ p^3/T = O(1)$".*

Then in the simplest setting when $\mathbf{x}_t = \mathbf{z}_t$, we have

$$\phi_\tau \xrightarrow{d} \mathcal{N}\left(\frac{1}{2},\ 1 + \frac{3(\nu_4 - 1)}{2}c\right).$$

This self-consistency has an important consequence in practice. In real data analysis, an analyst knows only the values of p and T in a data set, say for example $p = 50$ and $T = 500$. *Is this a high-dimensional situation where p/T tends to a constant $c = 0.1$, and hence the analyst can proceed with the limiting distribution in Theorem 2.1, or rather a low-dimensional situation where the sample size $T = 500$ can be considered large enough so that $p/T = 50/500 = 0.1$ could be assimilated to zero, and thus the analysis can rely on the limiting distribution in Theorem 2.2?* Clearly, this is a very hard question to answer. Without the self-consistency established in Theorem 2.3, one may be led to quite different decisions regarding the white noise test depending on the chosen limiting regime. This consistency property releases the analyst from such a dilemma: the unified result in Theorem 2.3 implies that the approximation

$$\phi_\tau \simeq \mathcal{N}\left(\frac{1}{2},\ 1 + \frac{3(\hat{\nu}_4 - 1)}{2} \cdot \frac{p}{T}\right),$$

is most likely accurate enough for a wide range of dimension-sample size combinations (p, T) in applications. Meanwhile, when ν_4 is unknown, which is usually the case in practice, we can use its sample counterpart, i.e, $\hat{\nu}_4 = \frac{1}{pT}\sum_{i=1}^{p}\sum_{t=1}^{T} x_{it}^4$, to replace it.

Remark 2.1. *When $\Sigma_0 = \sigma^2 I_p$, the single lag test statistic $\widetilde{L}_\tau$ can also be adopted for white noise test. Suppose $\{z_{it},\ i = 1, \cdots, p,\ t = 1, \cdots, T\}$ satisfies the conditions in Theorem 2.3. If $\mathbf{x}_t = \sigma^2 \mathbf{z}_t$, then the limiting distribution of the test statistic $\widetilde{L}_\tau$ becomes*

$$\frac{1}{\sigma^4} \cdot \frac{T}{p}\widetilde{L}_\tau - \frac{p}{2} \xrightarrow{d} \mathcal{N}\left(\frac{1}{2},\ 1 + \frac{3(\nu_4 - 1)}{2} \cdot \frac{p}{T}\right).$$

Note that σ^2 can be easily estimated from sample data, i.e. $\hat{\sigma}^2 = \frac{1}{pT}\sum_{t=1}^{T}\sum_{i=1}^{p} x_{it}^2$. Since $\hat{\sigma}^4 = \sigma^4 + O\left(\frac{1}{\sqrt{pT}}\right)$, substituting $\hat{\sigma}^4$ for σ^4 will not effect the limiting distribution. Therefore, we reject the null hypothesis for large values of $\left(\frac{1}{\hat{\sigma}^4}\cdot\frac{T}{p}\widetilde{L}_\tau - \frac{p}{2}\right)$.

2.4. Test procedure when Σ_0 is diagonal

Previously in Theorems 2.1 and 2.2, and their combination Theorem 2.3, the asymptotic normality of the test statistic ϕ_τ uses a crucial assumption: the time series $\mathbf{x}_t = \mathbf{z}_t$ has independent and identically distributed components. We now consider the case $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$ where the (unknown) covariance matrix Σ_0 is *diagonal*, say $\Sigma_0 = \operatorname{diag}\left(\sigma_1^2, \cdots, \sigma_p^2\right)$. We have

$$\widetilde{M}_{\tau,0} = \frac{1}{2T}\sum_{t=1}^{T}(\mathbf{z}_t\mathbf{z}_{t-\tau}^* + \mathbf{z}_{t-\tau}\mathbf{z}_t^*) = \frac{1}{2T}\sum_{t=1}^{T}\Sigma_0^{-1/2}(\mathbf{x}_t\mathbf{x}_{t-\tau}^* + \mathbf{x}_{t-\tau}\mathbf{x}_t^*)\Sigma_0^{-1/2} = \Sigma_0^{-1/2}\widetilde{M}_\tau\Sigma_0^{-1/2}.$$

This leads to the statistic

$$\widetilde{L}_{\tau,0} = \operatorname{Tr}(\widetilde{M}_{\tau,0}^*\widetilde{M}_{\tau,0}) = \operatorname{Tr}(\Sigma_0^{-1/2}\widetilde{M}_\tau^*\Sigma_0^{-1}\widetilde{M}_\tau\Sigma_0^{-1/2}) = \operatorname{Tr}(\Sigma_0^{-1}\widetilde{M}_\tau)^2,$$

which is equivalent to the statistic $\widetilde{L}_\tau$ defined in (2.5) for the previous case of $\Sigma_0 = I_p$.

Since Σ_0 is diagonal, $\mathbf{x}_t$ has p independent coordinates. Therefore σ_j^2 can be estimated respectively with the corresponding coordinates of $\mathbf{x}_t$, i.e.,

$$\hat{\sigma}_j^2 = \frac{1}{T}\sum_{t=1}^{T} x_{jt}^2, \quad \widehat{\Sigma}_0 = \operatorname{diag}\left(\widehat{\sigma}_1^2, \cdots, \hat{\sigma}_p^2\right).$$

Calculating the statistic ϕ_τ in (2.6) using the transformed data leads to the statistic

$$\widehat{\phi}_\tau = \frac{T}{p}\widehat{L}_{\tau,0} - \frac{p}{2}, \quad \text{where} \quad \widehat{L}_{\tau,0} = \operatorname{Tr}\left(\widehat{\Sigma}_0^{-1}\widetilde{M}_\tau\right)^2. \tag{2.10}$$

The null hypothesis will be rejected if

$$\widehat{\phi}_\tau > \frac{1}{2} + Z_\alpha\left(1 + \frac{3(\nu_4 - 1)}{2}c_p\right)^{1/2}.$$

In other words, we can first standardize the original data $\mathbf{x}_t$ coordinate-wise, and then apply the previous procedures on the transformed data.

3. Multiple-lags based tests

The test statistic ϕ_τ from previous sections is based on a fixed single lag τ, which can only detect serial dependence in a single lag each time. To capture a multi-lag dependence structure, we propose in this section multi-lag based test statistics to complement the single-lag based one.

Let $q \geq 1$ be a fixed integer, define the $p(q+1)$ dimensional vector $\mathbf{y}_j = \begin{pmatrix} \mathbf{x}_{j(q+1)-q} \\ \vdots \\ \mathbf{x}_{j(q+1)} \end{pmatrix}$, $j = 1, \cdots, N$, $N = \left[\frac{T}{q+1}\right]$. Since $\mathbb{E}\mathbf{x}_t = 0$ and $\Sigma_\tau = Cov(\mathbf{x}_{t+\tau}, \mathbf{x}_t)$, we have

$$\mathrm{Cov}(\mathbf{y}_j) = \begin{pmatrix} \Sigma_0 & \Sigma_1 & \cdots & \Sigma_q \\ \Sigma_1 & \Sigma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Sigma_1 \\ \Sigma_q & \cdots & \Sigma_1 & \Sigma_0 \end{pmatrix}_{(q+1)p\times(q+1)p}.$$

The null hypothesis H_0 : $\mathrm{Cov}(\mathbf{x}_{t+k}, \mathbf{x}_t) = \mathbf{0}$, $k = 1, \cdots, q$ becomes $H_0 : \Sigma_1 = \cdots = \Sigma_q = \mathbf{0}$, a test for a block diagonal covariance structure of the stacked sequence $\{\mathbf{y}_j\}$.

3.1. Test procedure when $\Sigma_0 = \sigma^2 I_p$

When $\Sigma_0 = \sigma^2 I_p$, the white noise test of $\{\mathbf{x}_t\}$ reduces to a sphericity test of $\{\mathbf{y}_j\}$. The well known John's test statistic U_q can be adopted for this purpose. In our case, the corresponding John's test statistic U_q is defined as

$$U_q = \frac{\frac{1}{p(q+1)} \sum_{i=1}^{p(q+1)} \left(l_{i,q} - \bar{l}_q\right)^2}{\bar{l}_q^{\,2}},$$

where $\{l_{i,q},\ i = 1, \cdots, p(q+1)\}$ are the eigenvalues of $S_q = \frac{1}{N} \sum_{j=1}^{N} \mathbf{y}_j \mathbf{y}_j^*$, the sample counterpart of $\mathrm{Cov}(\mathbf{y}_j)$, and $\bar{l}_q$ is the mean value of all $l_{i,q}$'s.

Actually, it has been proven in Li and Yao (2015) that the John's test possesses the powerful *dimension-proof* property, which keeps exactly the same limiting distribution under the null with any (n, p)-asymptotics, even regardless of normality. Specifically, we have the following.

Theorem 3.1. *Let $q \geq 1$ be a fixed integer, assume that*

1. $\{z_{it},\ i = 1, \cdots, p,\ t = 1, \cdots, T\}$ *are all independently distributed satisfying* $\mathbb{E}z_{it} = 0$, $\mathbb{E}z_{it}^2 = 1$, $\mathbb{E}z_{it}^4 = \nu_4$;
2. $p, T \to \infty$, $c_p := p/T \to c \in [0, \infty]$,

Then in the simplest setting when $\mathbf{x}_t = \mathbf{z}_t$, we have

$$NU_q - p(q+1) \xrightarrow{d} \mathcal{N}(\nu_4 - 2, 4), \tag{3.1}$$

where $N = \left[\frac{T}{q+1}\right]$, the integer part of fraction $\frac{T}{q+1}$.

Notice however that the use of blocks above reduces the sample size T to the number of blocks $N = \left[\frac{T}{q+1}\right]$. This may result in certain loss of power for the test based on (3.1). In order to limit such loss of power, we adopt the Simes method for multiple hypothesis testing (Simes, 1986).

Note that $\mathbf{y}_j = \left(\mathbf{x}'_{j(q+1)-q}, \cdots, \mathbf{x}'_{j(q+1)}\right)'$, $j = 1, \cdots, \left[\frac{T}{q+1}\right]$. To make full use of the data, $\mathbf{y}_j$ can also be defined as

$$\mathbf{y}_j = \begin{pmatrix} \mathbf{x}_{j(q+1)-q+k} \\ \vdots \\ \mathbf{x}_{j(q+1)+k} \end{pmatrix},$$

where $k = 0, 1, \cdots, q$, $j = 1, \cdots, \left[\frac{T-k}{q+1}\right]$. Then the John's test statistic U_q can be calculated based on $q+1$ different sets of $\mathbf{y}'_j s$ and thus results in $q+1$ different test statistics $U_q^{(k)}$.

Moreover, let P_k, $0 \le k \le q$ denotes the (asymptotic) P-value for the John's test with the $k-$th set of $\mathbf{y}'_j s$, i.e,

$$P_k = 1 - \Phi\left((NU_q^{(k)} - p(q+1) - \nu_4 + 2)/2\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Let $P_{(1)} \le \cdots \le P_{(q+1)}$ be a permutation of $P_0, \cdots, P_q$. Then by the Simes method, we reject H_0 if $P_{(k)} \le \frac{k}{q+1}\alpha$ at least for one $1 \le k \le q+1$ for the nominal level α.

3.2. Test procedure with general Σ_0

Previously, the white noise test of $\{\mathbf{x}_t\}$ reduces to a sphericity test of $\{\mathbf{y}_j\}$ when $\Sigma_0 = \sigma^2 I_p$. Now if $\Sigma_0 \neq \sigma^2 I_p$, the white noise test becomes a test of block diagonal structure of the covariance matrix of $\{\mathbf{y}_j\}_{1\le j\le N}$. Actually, Srivastava (2005) has derived the limiting distribution of the John's test statistic with a general population covariance matrix and *normally* distributed samples.

More specifically, if $\mathbf{y}_1, \cdots, \mathbf{y}_N$ are samples from $\mathcal{N}_{p(q+1)}(\mu, \Sigma_y)$, denote $a_i = \frac{1}{p(q+1)}\mathrm{Tr}(\Sigma_y^i)$ and let

$$\widehat{U}_q = \frac{\frac{1}{p(q+1)}\mathrm{Tr}\left(\widehat{S}_q^2\right)}{\left(\frac{1}{p(q+1)}\mathrm{Tr}\left(\widehat{S}_q\right)\right)^2},$$

where $\widehat{S}_q = \frac{1}{N-1}\sum_{j=1}^N(\mathbf{y}_j - \overline{\mathbf{y}})(\mathbf{y}_j - \overline{\mathbf{y}})^*$, $\overline{\mathbf{y}} = \frac{1}{N}\sum_{j=1}^N \mathbf{y}_j$. Then according to Theorem 3.1 in Srivastava (2005), we have the following.

Proposition 3.1. *Let $q \ge 1$ be a fixed integer. Assume that*

(1) As $p \to \infty$, $a_i \to a_i^0$, $0 < a_i^0 < \infty$, $i = 1, \cdots, 8$;
(2) p, $N \to \infty$, $N = O(p^\delta)$, $0 < \delta \le 1$.

Then

$$\frac{(N-1)^3}{(N-2)(N+1)}\widehat{U}_q - \frac{p(q+1)(N-1)^2}{(N-2)(N+1)} - \frac{a_2}{a_1^2}(N-1) \xrightarrow{d} \mathcal{N}\left(0, 4\tau_1^2\right), \tag{3.2}$$

where $\tau_1^2 = \frac{2N\left(a_4a_1^2 - 2a_1a_2a_3 + a_2^3\right)}{p(q+1)a_1^6} + \frac{a_2^2}{a_1^4}$.

Note first that if $\Sigma_y = \sigma^2 I_{p(q+1)}$, then equation (3.2) is asymptotically equivalent to equation (3.1) for normal samples ($\nu_4 = 3$).

Moreover, under H_0,

$$\Sigma_y = \mathrm{Cov}(\mathbf{y}_j) = \begin{pmatrix} \Sigma_0 & 0 & \cdots & 0 \\ 0 & \Sigma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_0 \end{pmatrix}_{(q+1)p\times(q+1)p},$$

then the a_i's in (3.2) can be written as $a_i = \frac{1}{p}\mathrm{Tr}(\Sigma_0^i)$, where $\Sigma_0 = \mathrm{Cov}(\mathbf{x}_t)$, $t = 1, \cdots, T$. Since the a_i's are usually unknown in practice, we can use their sample counterparts to derive consistent estimators for them.

In particular, denote $b_i = \frac{1}{p}\mathrm{Tr}(S_x^i)$, where $S_x = \frac{1}{T}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t^*$, and define the estimators $(\hat{a}_i)_{1\le i\le 4}$ as the solutions to the system in Lemma 2.16 in Yao et al. (2015), we then have

$$\begin{cases} \hat{a}_1 = b_1, \\ \hat{a}_2 = b_2 - c_p b_1^2, \\ \hat{a}_3 = b_3 - 3c_p b_1 b_2 + 2c_p^2 b_1^3, \\ \hat{a}_4 = b_4 - 4c_p b_1 b_3 - 2c_p b_2^2 + 10c_p^2 b_1^2 b_2 - 5c_p^3 b_1^4, \\ \hat{\tau}_1^2 = \dfrac{2N\left(\hat{a}_4\hat{a}_1^2 - 2\hat{a}_1\hat{a}_2\hat{a}_3 + \hat{a}_2^3\right)}{p(q+1)\hat{a}_1^6} + \dfrac{\hat{a}_2^2}{\hat{a}_1^4}. \end{cases}$$

Here $c_p = \frac{p}{T}$. Substituting these estimators $(\hat{a}_i)$ for their population counterparts (a_i) will not change the limiting distribution in Proposition 3.1. Therefore for a fixed integer $q \ge 1$, when $\mathbf{x}_t \sim \mathcal{N}_p(0, \Sigma_0)$, we should reject H_0 for large values of $\widehat{U}_q$, i.e.,

$$\textit{Reject } H_0 \quad \textit{if} \quad \frac{(N-1)^3}{(N-2)(N+1)}\widehat{U}_q > \frac{p(q+1)(N-1)^2}{(N-2)(N+1)} + \frac{\hat{a}_2}{\hat{a}_1^2}(N-1) + 2\hat{\tau}_1 Z_\alpha.$$

Similarly the Simes' method can also be adopted here to enhance the power of the test.

4. Simulation experiments

Most of the experiments of this section are designed in order to compare the test procedure in (2.8) based on the statistic ϕ_τ, and the procedure based on the statistic U_q with the Simes method implemented as described at the end of Section 3.1, with two well known classical white noise tests, namely the Hosking test (Hosking, 1980) and the Li-McLeod test (Li et al., 1981). At the end of the section, experiments are conducted to assess the performance of the test statistic $\widehat{\phi}_\tau$ in (2.10) using standardized data when the population covariance matrix is diagonal.

To introduce the Hosking and Li-McLeod tests and using their notations, 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 auto-covariance matrices $\hat{C}_\tau$ of the residuals $\{\hat{\mathbf{a}}_t\}$:

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

Hosking (1980) proposed the portmanteau statistic

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

while Li et al. (1981) recommended the use of the statistic

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

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 Single Lag-τ test statistic ϕ_τ and multiple lags John's test with Simes method, we set $u = v = 0$. All tests use 5% significance level and the critical regions of the three tests are as follows:

(i) Single lag-τ test: $\left\{\phi_\tau > \frac{1}{2} + Z_{0.95}(1 + \frac{3(\nu_4-1)}{2} c_p)^{1/2}\right\}$;
(ii) Multi-lag-q test: $\left\{\text{at least for one } 1 \le k \le q+1, P_{(k)} \le \frac{k}{q+1} 0.05\right\}$;
(iii) Hosking's test: $\left\{\widetilde{Q}_q > \chi^2_{0.95,\ qp^2}\right\}$;
(iv) Li-McLeod test: $\left\{Q_q^* > \chi^2_{0.95,\ qp^2}\right\}$.

Here $Z_{0.95}$ and $\chi^2_{0.95,\ m}$ denote the 95 percentile of the standard normal distribution and the chi-squared distribution with degrees of freedom m, respectively. Empirical statistics are obtained using 5000 independent replicates.

4.1. Empirical sizes

The data is generated as $\mathbf{x}_t = \mathbf{z}_t$, with $\mathbf{z}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ being independent and identically distributed, $t = 1, \cdots, T$. Table 1 compares the sizes of the four tests for two different q. 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's 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 ϕ_τ and U_q have reasonable sizes when compared to the 5% nominal level across all the tested (p, T) combinations.

4.2. Empirical powers and adjusted powers

In this section, we compare the empirical powers of the tests by assuming that $\mathbf{x}_t$ follows a vector autoregressive process of order 1,

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{z}_t,$$

where $A = \alpha\mathbf{I}_p$, $\mathbf{z}_t \sim N_p(\mathbf{0}, \mathbf{I}_p)$ being independent of each other for $t = 1, \cdots, T$. Here we assign $\alpha = 0.1$ and apply the three test procedures to get the power values as in Table 2.

From Table 1 we know that the two classic tests become seriously biased when the dimension p is larger compared to the sample size T. Their sizes approach zero when p/T becomes larger. From Table 2, we see that due to the biased critical values used in $\widetilde{Q}_q$ and Q^*_q as shown in Table 1, their powers are driven downward. This is particularly severe when the ratio p/T is larger than 0.5.

In Table 3, we compare the *intrinsic powers* of the four procedures. Namely, we empirically find the 95 percentiles of $\widetilde{Q}_q$ and Q^*_q and use these values as the corrected critical values for the power comparison. It is interesting to observe that after such correction, both $\widetilde{Q}_q$ and Q^*_q show very reasonable powers which all increase to 1 when the dimension and the sample size increase. However, even with such empirically adjusted critical values, our single-lag based test still dominates these two tests by displaying a generally much higher power in all the tested (p, T) combinations. Table 4 demonstrates the feasibility of our test statistics when the dimension p is larger than the sample size T where the other two tests are not even applicable. Comparison with the Hosking's and the Li-McLeod tests sheds new light on the superiority of our test statistics in both low and high dimensional cases.

4.3. Why both the Hosking's and the Li-Mcleod tests fail in high dimension?

The experiments here are designed to explore the reasons behind the failure of the Hosking's 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}). Statistics for ϕ_τ are given in Table 5. We observe a very good agreement between the empirical and theoretical values in all tested (p, T)-combinations. 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 $(\theta_{theo} - \theta_{emp})/\theta_{emp}$ (in percentage). This is done in Table 6. 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 percentiles which, although in a lesser proportion, is enough to create a high down-bias in the empirical sizes of these two classical tests with high-dimensional data; See Table 1.

4.4. Case of a diagonal Σ_0

Simulations have also been carried out to attest the finite-sample performance of the new test statistic $\widehat{\phi}_\tau$ in (2.10). We fix $\{\sigma_j^2,\ j=1,\cdots,p\}$ to be an arithmetic sequence running from $\sigma_1^2 = 0.5$ to $\sigma_p^2 = 3$. The $\mathbf{z}_t$'s are independent p-variate standard normal, $\mathbf{z}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p),\ t=1,\cdots,T$. The data is generated by letting $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$. First we check the characteristics of the test statistic $\widehat{\phi}_\tau$ as opposed to those of ϕ_τ. Note that if Σ_0 were known, the standardization $\Sigma_0^{-1/2}\mathbf{x}_t$ leads to the statistic $\widetilde{L}_{\tau,0} = \mathrm{Tr}(\Sigma_0^{-1}\widetilde{M}_\tau)^2$, which is equivalent to the test statistic ϕ_τ studied previously. This procedure will be referred as *the oracle procedure* for comparison. Empirical means, variances and 95th percentiles for $\widehat{\phi}_\tau$ (in fact a scaled version $c_p^{-2}\widehat{\phi}_\tau$) are given in Table 7 where the corresponding benchmark values from ϕ_τ are given in bold for comparison. It can be seen from the table that the empirical means of $\widehat{\phi}_\tau$ match very well to those of ϕ_τ while certain discrepancy exists between the empirical variances, thus the empirical 95th percentiles of $\widehat{\phi}_\tau$ and their benchmark values. It is also observed that such discrepancy becomes more severe with large values of the ratio $c_p = p/T$. This will lead to biased empirical sizes and powers of the test based on $\widehat{\phi}_\tau$ as shown in Table 8. Here for the evaluation of the power of the test, the sequence $(\mathbf{z}_t)$ is chosen to follow a vector autoregressive process of order 1,

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + \varepsilon_t,$$

where $A = \alpha\mathbf{I}_p$, $\alpha = 0.1$, $\varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ which is independent of each other for $t = 1,\cdots,T$. Similarly, $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$ for the simulated sequence. Multi-lag-q test procedure with general Σ_0, i.e. $\widehat{U}_q$ in Srivastava (2005) combined with Simes' method, is also adopted here for comparison.

It is striking to observe that although the empirical sizes for the test based on $\widehat{\phi}_\tau$ with standardized data are clearly down biased, its empirical powers remain reasonably high in almost all the tested (p,T) combinations. In other words, the very conservative trend of the statistic $\widehat{\phi}_\tau$ in term of test size has not annihilated all its power. Therefore, in the case of an unknown diagonal cross-sectional covariance matrix, the white noise test based on $\widehat{\phi}_\tau$ remains recommendable with satisfactory power and a low Type I error.

5. Case of a general covariance matrix Σ_0

When the population covariance matrix Σ_0 is general without any particular structure, the testing problem becomes even more intricate in high dimensions. To fix the idea, assume again the data vectors are of the form $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$ where the $\mathbf{z}_t$'s have standardized i.i.d. components. So what about the data standardization procedure advocated in Section 2.4 for diagonal Σ_0's, that is by first finding an estimator $\widehat{\Sigma}_0$ of Σ_0, and then applying the theory developed previously when $\Sigma_0 = \mathbf{I}_p$ with the statistic ϕ_τ? Unfortunately enough, this "natural" approach ends up fruitless here due to the lack of an efficient estimator of Σ_0 when the dimension is high. As far as we know, no consistent estimator is available for a general high dimensional covariance matrix Σ_0 without a particular structure such as diagonal, banded or being sparse. As a consequence, the standardized observations $\mathbf{y}_t = \widehat{\Sigma}_0^{-1/2}\mathbf{x}_t$ will have a covariance matrix far away from the identity matrix and applying the test statistic ϕ_τ will lead to dramatic errors.

Here for a general Σ_0 we propose another test statistic, namely

$$G_q = \sum_{\tau=1}^{q} Q_\tau, \text{ where } Q_\tau = \mathrm{Tr}\left(\widetilde{\Sigma}(\tau)\widetilde{\Sigma}(\tau)^*\right) \text{ with } \widetilde{\Sigma}(\tau) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{t-\tau}\mathbf{x}_t^*, \tag{5.1}$$

where $\mathbf{x}_t = \mathbf{x}_{T+t}$ for $t \le 0$.

A conjecture about the asymptotic normality of G_q is formulated below under the high dimensional setting $p/T \to c > 0$ (Marčenko-Pastur). Although theoretical proof of the asymptotic normality has not yet been fully established, simulation studies, on the other hand, lend full support to the result as follows.

Conjecture 5.1. *Let $q \ge 1$ be a fixed integer. Then under the assumptions that the components $\{z_{it}, i = 1, \dots, p, t = 1, \dots, T\}$ of $\{\mathbf{z}_t\}$ are all independently distributed satisfying $\mathbb{E}z_{it} = 0, \mathbb{E}z_{it}^2 = 1, \mathbb{E}z_{it}^4 = \nu_4 < \infty$, we have when $p, T \to \infty$ and $p/T \to c > 0$,*

$$G_q - qTc_p^2 s_1^2 \xrightarrow{d} \mathcal{N}\left(0, 2qc^2 s_2^2 + 4q^2c^3(\nu_4 - 3)s_1^2 s_{d,2} + 8q^2c^3 s_1^2 s_2\right),$$

where $s_\ell = \lim_{p\to\infty} \frac{1}{p}\mathrm{Tr}(\Sigma_0^\ell)$, $s_{d,\ell} = \lim_{p\to\infty} \frac{1}{p}\mathrm{Tr}(\mathrm{diag}^\ell(\Sigma_0))$.

When $\Sigma_0 = \mathbf{I}_p$ and the z_{it}'s are normally distributed, we have $s_1 = s_2 = s_{d,2} = 1$ and $\nu_4 = 3$. From this proposition, we can see that in general, when $\Sigma_0 \ne \mathbf{I}_p$, we need to estimate four more quantities for carrying out the white noise test, namely, s_1, s_2, $s_{d,2}$ and ν_4.

Simulations are carried out to check the validity of the result in Conjecture 5.1. In these experiments, the covariance matrix Σ_0 is taken from the following three profiles:

1. $\Sigma_0 = 4\mathbf{I}_p$;
2. $\Sigma_0 = \mathbf{I}_p + Q_0 D Q_0^*$, where Q_0 is an orthogonal matrix generated randomly each time, and D is diagonal with 10% of entries being randomly generated $U(0,2)p^{1/3}$, and the rest are $U(0,1)p^{-1/2}$.
3. $\Sigma_0 = Q_0 D Q_0^*$, where Q_0 is an orthogonal matrix generated randomly each time, and D is diagonal with elements generated randomly as $U(1,6)$.

Once Σ_0 is chosen, we consider two types of $\{\mathbf{z}_t\}$ with either the z_{it}'s being independent and identically distributed (i.i.d.) standard normal, or being i.i.d. $U(-2\sqrt{3}, 2\sqrt{3})$, with $\nu_4 = 1.8$. Thus in total six different scenarios for the data $\mathbf{x}_t = \Sigma_0^{1/2}\mathbf{z}_t$ are thus considered. For each scenario, we calculate the statistic G_5 and standardize it using the result from Conjecture 5.1. Again 5000 independent replications are used in these experiments where $p = 500$ and $T = 600$ have been fixed. Results are shown in Figure 1. In all the six scenarios, the conjecture seems well confirmed.

Despite this empirical confirmation, the practical usage of it is still limited unfortunately. For instance, if we have an estimator $\hat{s}_1$ of s_1, to utilize the result, we compute $G_q - qTc_p^2\hat{s}_1^2$ and hope that this is centered at 0 when the data is truly white noise. Consider

$$\begin{aligned} G_q - qTc_p^2\hat{s}_1^2 &= (G_q - qTc_p^2 s_1^2) + qTc_p^2(s_1^2 - \hat{s}_1^2) \\ &= (G_q - qTc_p^2 s_1^2) + qTc_p^2(s_1 + \hat{s}_1)(s_1 - \hat{s}_1). \end{aligned}$$

Hence for the above to center at 0 asymptotically, we need $s_1 - \hat{s}_1 = o_P(T^{-1})$, so that the second term on the right hand side above will be $o_P(1)$, while the first term goes to 0 by the result of Conjecture 5.1. Unfortunately, we can only prove that $s_1 - \hat{s}_1 = O_P(T^{-1})$ if $\hat{s}_1 = p^{-1}\mathrm{Tr}(S)$ where S is the sample covariance matrix of the data. Even if $\Sigma_0 = \sigma^2\mathbf{I}_p$ does not help since a natural estimator of σ^2 is indeed the very same $\hat{s}_1$, still having $s_1 - \hat{s}_1 = O_P(T^{-1})$ in this particular case. This argument highlights the difficulty in high dimensional testing of white noise.

6. Concluding remarks

In this paper, two types of test statistics are proposed for testing a high dimensional white noise, namely the single-lag-τ serial test statistics ϕ_τ, $\widehat{\phi_\tau}$ and the multi-lag-q serial test statistics U_q, $\widehat{U}_q$ and G_q. In practice, different test statistics should be carefully chosen to fit in different scenarios of observations. For example, the data $\mathbf{x}_t$ can either come from a normal or non-normal populations; the population covariance Σ_0 of $\mathbf{x}_t$ can either be spherical, diagonal or completely arbitrary. A summary of these test statistics is given in the table below.

Summary of different test statistics

	Single-lag-τ	**Multi-lag-q**	
$\Sigma_0 = \sigma^2 I_p$	ϕ_τ	U_q	
Diagonal Σ_0	$\widehat{\phi_\tau}$	$\widehat{U}_q$ (normal population only);	G_q
General Σ_0	–	$\widehat{U}_q$ (normal population only);	G_q

It can be seen from the table that all these test statistics are applicable for general (non-normal) populations except for $\widehat{U}_q$. It is of interest to extend the asymptotic result of $\widehat{U}_q$ to cover general non-normal populations. Alternatively, a rigorous proof of the conjecture for limiting distribution of G_q would also provide a notable progress. On the other hand, given the extraordinarily powerful performance of single-lag-τ test statistic ϕ_τ, its extension to the setting with a general Σ_0 is surely worth further investigation. Unfortunately, due to the complexity of the analysis needed, these investigations are much beyond the scope of this paper. More efforts are still needed for such further exploration.

7. Proofs

7.1. Preliminaries

For any $n \times n$ Hermitian matrix M with real eigenvalues $\lambda_1, \cdots, \lambda_n$, the empirical spectral distribution (ESD for short) of M is defined by $F^M = n^{-1}\sum_{j=1}^n \delta_{\lambda_j}$, where δ_a denotes the Dirac mass at a. The Stieltjes transform of any distribution G is defined as

$$m_G(z) = \int \frac{1}{x-z} dG(x), \ \Im(z) > 0,$$

where $\Im(z)$ stands for the imaginary part of z.

Consider the lag$-\tau$ sample auto-covariance matrix $\widehat{M}_\tau$, Jing et al. (2014) derived the limit of the ESD of $\widehat{M}_\tau$ with finite $(2+\delta)$-th moment restriction under $p/T \to c$ asymptotic. Bai and Wang (2015) further consolidates the results by providing an alternative approach to derive the limiting spectral distribution(LSD). Theorem 1.1 in Bai and Wang (2015) states as follows:

Theorem 7.1. *Assume*

(a) $\tau \geq 1$ *is a fixed integer.*

(b) $\mathbf{x}_t = \left(x_{1t}, \cdots, x_{pt}\right)$, $t = 1, \cdots, T$ *is* p *dimensional vectors with independent components with*

$$\sup_{1\le i\le p, 1\le t\le T} \mathbb{E}|x_{it}|^{2+\delta} \le M < \infty,$$

for some $\delta \in (0, 2]$ *and for any* $\eta > 0$,

$$\frac{1}{\eta^{2+\delta} pT} \sum_{i=1}^{p} \sum_{t=1}^{T} \mathbb{E}\left(|x_{it}|^{2+\delta} I\left(|x_{it}| \geq \eta T^{1/(2+\delta)}\right)\right) = o(1).$$

(c) $p/T \to c \in (0, \infty)$ *as* $p, T \to \infty$.

(d) $\widehat{M}_\tau = \frac{1}{2T} \sum_{t=1+\tau}^{T} (\mathbf{x}_t \mathbf{x}_{t-\tau}^* + \mathbf{x}_{t-\tau} \mathbf{x}_t^*)$.

Then as $p, T \to \infty$, $F^{\widehat{M}_\tau} \xrightarrow{d} F_c$ *a.s. and* F_c *has a density function given by*

$$\phi_c(x) = \frac{1}{2c\pi} \sqrt{\frac{y_0^2}{1+y_0} - \left(\frac{1-c}{|x|} + \frac{1}{\sqrt{1+y_0}}\right)^2}, \ |x| \le a,$$

where

$$a = \begin{cases} \dfrac{(1-c)\sqrt{1+y_1}}{y_1 - 1}, & c \neq 1, \\ 2, & c = 1, \end{cases}$$

y_0 *is the largest real root of equation:* $y^3 - \frac{(1-c)^2 - x^2}{x^2} y^2 - \frac{4}{x^2} y - \frac{4}{x^2} = 0$ *and* y_1 *is the only real root of the equation:* $((1-c)^2 - 1)y^3 + y^2 + y - 1 = 0$ *such that* $y_1 > 1$ *if* $c < 1$ *and* $y_1 \in (0, 1)$ *if* $c > 1$. *Further, if* $c > 1$, *then* F_c *has a point mass* $1 - 1/c$ *at the origin. Meanwhile, the Stieltjes transform* $m(z)$ *of* F_c *satisfies*

$$(1 - c^2 m^2(z))(c + czm(z) - 1)^2 = 1.$$

Bai and Silverstein (2004) and Zheng et al. (2015) studied the central limit theorem for linear spectral statistics of large-dimensional sample covariance matrices. By their notation, let

$B_n = (1/N)T_n^{1/2}X_nX_n^*T_n^{1/2}$, where $X_n = (X_{ij})$ is $n \times N$ with i.i.d. complex standardized entries having finite fourth moment, $T_n^{1/2}$ is a Hermitian square root of the nonnegative definite Hermitian matrix T_n. It has been proven that if for all n, i, j, $X_{i,j}^n$ are independent, with probability 1, $F^{T_n} \xrightarrow{d} H$, a proper cumulative distribution function(c.d.f.) and $n/N \to y > 0$ as $n \to \infty$, then with probability 1, ESD of B_n, F^{B_n} converges in distribution to $F^{y,H}$, a non-random proper c.d.f. If $\underline{B}_n = (1/N)X_n^*T_nX_n$, then its LSD $\underline{F}^{y,H}$ satisfies

$$\underline{F}^{y,H} = (1-y)I_{[0,\infty)} + yF^{y,H},$$

and its Stieltjes transform has inverse

$$z = -\frac{1}{\underline{m}} + y\int \frac{t}{1+t\underline{m}}\,\mathrm{d}H(t). \tag{7.1}$$

Define

$$G_n(x) = n\left[F^{B_n}(x) - F^{y_n,H_n}(x)\right],$$

and $f_1, \cdots, f_k$ be functions on $\mathbb{R}$ analytic, assume $\mathbb{E}(X_{ij}) = 0$, $\mathbb{E}(|X_{ij}|^2) = 1$, $\mathbb{E}(|X_{ij}|^4) = \nu_4 < \infty$, then random vector

$$\left(\int f_1(x)\,\mathrm{d}G_n(x), \cdots, \int f_k(x)\,\mathrm{d}G_n(x)\right), \tag{7.2}$$

forms a tight sequence in n and (7.2) converges weakly to a Gaussian vector $(X_{f_1}, \cdots, X_{f_k})$ with means

$$\mathbb{E}X_f = -\frac{1}{2\pi i}\oint f(z)\frac{y\int \frac{\underline{m}(z)^3t^2}{(1+t\underline{m}(z))^3}\,\mathrm{d}H(t)}{\left(1-y\int \frac{\underline{m}(z)^2t^2}{(1+t\underline{m}(z))^2}\,\mathrm{d}H(t)\right)^2}\mathrm{d}z - \frac{\nu_4-3}{2\pi i}\oint f(z)\frac{y\int \frac{\underline{m}(z)^3t^2}{(1+t\underline{m}(z))^3}\,\mathrm{d}H(t)}{1-y\int \frac{\underline{m}(z)^2t^2}{(1+t\underline{m}(z))^2}\,\mathrm{d}H(t)}\mathrm{d}z, \tag{7.3}$$

and covariance function

$$\begin{aligned} Cov(X_f, X_g) = &-\frac{1}{2\pi^2}\oint\oint \frac{f(z_1)g(z_2)}{\left(\underline{m}(z_1)-\underline{m}(z_2)\right)^2}\frac{d}{dz_1}\underline{m}(z_1)\frac{d}{dz_2}\underline{m}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2 \\ &- \frac{y(\nu_4-3)}{4\pi^2}\oint\oint f(z_1)g(z_2)\left(\int \frac{t}{(t\underline{m}(z_1)+1)^2}\cdot\frac{t}{(t\underline{m}(z_2)+1)^2}\,\mathrm{d}H(t)\right)d\underline{m}(z_1)d\underline{m}(z_2), \end{aligned} \tag{7.4}$$

$(f, g \in \{f_1, \cdots, f_k\})$. The contours in (7.3) and (7.4) (two in (7.4), which we may assume to be non-overlapping) are closed and are taken in the positive direction in the complex plane, each enclosing the support of $F^{y,H}$.

7.2. Proof for Theorem 2.1

Let $\widehat{N}_\tau = \frac{1}{2p}\sum_{t=1+\tau}^{T}(\mathbf{x}_t\mathbf{x}_{t-\tau}^* + \mathbf{x}_{t-\tau}\mathbf{x}_t^*) = \frac{T}{p}\widehat{M}_\tau$. To test H_0, we let $\mathbf{x}_t = \mathbf{z}_t$ and focus on test statistic

$$\widehat{L}_\tau = \sum_{j=1}^{p}\hat{l}_{j,\tau}^2 = \mathrm{Tr}(\widehat{N}_\tau^*\widehat{N}_\tau),$$

where $\{\hat{l}_{j,\tau},\ 1 \le j \le p\}$ are eigenvalues of $\widehat{N}_\tau$.

Note that

$$\begin{aligned}
\widehat{N}_\tau &= \frac{1}{2p}\sum_{t=1+\tau}^{T}(\mathbf{x}_t\mathbf{x}_{t-\tau}^* + \mathbf{x}_{t-\tau}\mathbf{x}_t^*)\\
&= \frac{1}{p}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)\begin{pmatrix} 0 & \cdots & \frac{1}{2} & \cdots & 0\\ \vdots & \ddots & 0 & \frac{1}{2} & \vdots\\ \frac{1}{2} & 0 & \ddots & 0 & \frac{1}{2}\\ \vdots & \frac{1}{2} & 0 & \ddots & \vdots\\ 0 & \cdots & \frac{1}{2} & \cdots & 0\end{pmatrix}\begin{pmatrix}\mathbf{x}_1^*\\ \mathbf{x}_2^*\\ \vdots\\ \mathbf{x}_T^*\end{pmatrix}\\
&= \frac{1}{p}X_T C_{T,\tau} X_T^*
\end{aligned}$$

where $C_{T,\tau}$ is $T \times T$ matrix with two bands of $\frac{1}{2}$ which are $\tau-$distance from main diagonal. According to results in Bai and Wang (2015),

Lemma 7.1. *The $T \times T$ matrix $C_{T,\tau}$ has $\tau - 1$ zero eigenvalues and other $T - \tau + 1$ eigenvalues are*

$$\lambda_k = \cos\frac{k\pi}{T-\tau+2},\ k = 1, 2, \cdots, T-\tau+1.$$

As $T \to \infty$, the empirical spectral distribution(ESD) of $C_{T,\tau}$ tends to H with density function

$$H'(t) = \frac{1}{\pi\sqrt{1-t^2}},\ t \in (-1, 1).$$

Following the theory in Bai and Silverstein (2004), let

$$B_n = \frac{1}{p}C_{T,\tau}X_T^*X_T,\ \underline{B}_n = \frac{1}{p}X_T C_{T,\tau}X_T^*,\ f(x) = x^2,\ \frac{T}{p} \to \frac{1}{c} = y,$$

then

$$\begin{aligned}
\int f(x)\,\mathrm{d}G_n(x) &= \int x^2\,\mathrm{d}n\left[F^{B_n}(x) - F^{y_n,H_n}(x)\right]\\
&= T\int x^2\,\mathrm{d}F^{\frac{1}{p}C_{T,\tau}X_T^*X_T} - T\int x^2\,\mathrm{d}F^{\frac{1}{c},C_{T,\tau}}\\
&= \sum_{j=1}^{p}\hat{l}_{j,\tau}^2 - T\int x^2\,\mathrm{d}H_n(x)\\
&= \sum_{j=1}^{p}\hat{l}_{j,\tau}^2 - \sum_{k=1}^{T-\tau+1}\left(\cos\frac{k\pi}{T-\tau+2}\right)^2 = \sum_{j=1}^{p}\hat{l}_{j,\tau}^2 - \frac{1}{2}(T-\tau)
\end{aligned}$$

where $\{\hat{l}_{j,\tau},\ 1 \leq j \leq p\}$ are eigenvalues of $\widehat{N}_\tau = \frac{1}{p} X_T C_{T,\tau} X_T^*$. The above equation holds because

$$\int x \,\mathrm{d}H(x) = \int_{-1}^{1} \frac{x}{\pi\sqrt{1-x^2}} \,\mathrm{d}x = 0,$$
$$\int x^2 \,\mathrm{d}H(x) = \int_{-1}^{1} \frac{x^2}{\pi\sqrt{1-x^2}} \,\mathrm{d}x = \frac{1}{2}.$$

According to (7.1), the Stieltjes transform of LSD of $\widehat{N}_\tau = \frac{1}{p} X_T C_{T,\tau} X_T^*$ satisfies

$$\begin{aligned} z &= -\frac{1}{\underline{m}} + y \int \frac{t}{1+t\underline{m}} \,\mathrm{d}H(t) \\ &= -\frac{1}{\underline{m}} + \frac{1}{c} \int_{-1}^{1} \frac{t}{1+t\underline{m}} \cdot \frac{1}{\pi\sqrt{1-t^2}} \,\mathrm{d}t, \end{aligned}$$

thus,

$$\begin{aligned} z\underline{m} &= -1 + \frac{1}{c} - \frac{1}{c} \int_{-1}^{1} \frac{1}{1+t\underline{m}} \cdot \frac{1}{\pi\sqrt{1-t^2}} \,\mathrm{d}t \\ &= -1 + \frac{1}{c} - \frac{1}{c\sqrt{1-\underline{m}^2}}. \end{aligned}$$

Taking derivative with respective to z on both side of equation (7.1), we have

$$\frac{d\underline{m}}{\mathrm{d}z} = \frac{\underline{m}^2}{1 - y\int \frac{t^2\underline{m}^2}{(1+t\underline{m})^2} \,\mathrm{d}H(t)},$$

by (7.3), we have, for the first term in $\mathbb{E}X_f$,

$$\begin{aligned} &-\frac{1}{2\pi i} \oint f(z) \frac{y\int \frac{\underline{m}(z)^3 t^2}{(1+t\underline{m}(z))^3} \,\mathrm{d}H(t)}{\left(1 - y\int \frac{\underline{m}(z)^2 t^2}{(1+t\underline{m}(z))^2} \,\mathrm{d}H(t)\right)^2} \mathrm{d}z \\ &= -\frac{1}{2\pi i} \oint z^2 \frac{y\int \frac{\underline{m}(z) t^2}{(1+t\underline{m}(z))^3} \,\mathrm{d}H(t)}{1 - y\int \frac{\underline{m}(z)^2 t^2}{(1+t\underline{m}(z))^2} \,\mathrm{d}H(t)} \cdot \frac{d\underline{m}}{\mathrm{d}z} \mathrm{d}z = \frac{1}{2\pi i} \oint z^2 \frac{\frac{\underline{m}(1+2\underline{m}^2)}{2c\left(\sqrt{1-\underline{m}^2}\right)^5}}{1 - \frac{1}{c} + \frac{1}{c} \cdot \frac{1-2\underline{m}^2}{\left(\sqrt{1-\underline{m}^2}\right)^3}} d\underline{m} \end{aligned}$$

since $z = \frac{1}{\underline{m}}\left(-1 + \frac{1}{c} - \frac{1}{c\sqrt{1-\underline{m}^2}}\right)$, by residue theorem, the first term in $\mathbb{E}X_f$ equals to

$$\frac{1}{2\pi i}\oint \frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}^2}}\right)^2}{\underline{m}^2} \cdot \frac{\underline{m}(1+2\underline{m}^2)}{2(c-1)\left(\sqrt{1-\underline{m}^2}\right)^5 + 2(1-2\underline{m}^2)(1-\underline{m}^2)} d\underline{m}$$
$$= \left.\frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}^2}}\right)^2 (1+2\underline{m}^2)}{2(c-1)\left(\sqrt{1-\underline{m}^2}\right)^5 + 2(1-2\underline{m}^2)(1-\underline{m}^2)}\right|_{\underline{m}=0} = \frac{1}{2c}.$$

Similarly, for the second term in $\mathbb{E}X_f$,

$$-\frac{(\nu_4 - 3)}{2\pi i}\oint f(z)\frac{y\int \frac{\underline{m}(z)^3 t^2}{(1+t\underline{m}(z))^3}\,\mathrm{d}H(t)}{1 - y\int \frac{\underline{m}(z)^2 t^2}{\left(1+t\underline{m}(z)\right)^2}\,\mathrm{d}H(t)}\mathrm{d}z$$
$$= -\frac{(\nu_4 - 3)}{2\pi i}\oint z^2 \cdot \frac{1}{c}\left(\int \frac{\underline{m}(z)t^2}{(1+t\underline{m}(z))^3}\,\mathrm{d}H(t)\right) \cdot \frac{d\underline{m}}{\mathrm{d}z}\,\mathrm{d}z = 0.$$

Therefore, the mean term

$$\mathbb{E}X_f = \frac{1}{2c}.$$

By (7.4), we have, for the first term of $Var\left(X_f\right)$,

$$-\frac{1}{2\pi^2}\oint\!\!\oint \frac{z_1^2 z_2^2}{\left(\underline{m}(z_1) - \underline{m}(z_2)\right)^2} d\underline{m}(z_1) d\underline{m}(z_2)$$
$$= -\frac{1}{2\pi^2}\oint \frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}_2^2}}\right)^2}{\underline{m}_2^2} d\underline{m}_2 \oint \frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2}{\underline{m}_1^2\left(\underline{m}_1 - \underline{m}_2\right)^2} d\underline{m}_1,$$

Similarly, by residue theorem,

$$\frac{1}{2\pi i}\oint \frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2}{\underline{m}_1^2\left(\underline{m}_1 - \underline{m}_2\right)^2} d\underline{m}_1 = \frac{1}{2\pi i}\oint \frac{1}{\underline{m}_1^2} \cdot \frac{\left(1 + (c-1)\sqrt{1-\underline{m}_1^2}\right)^2}{c^2\left(\underline{m}_1 - \underline{m}_2\right)^2\left(1 - \underline{m}_1^2\right)} d\underline{m}_1$$
$$= \left.\left[\frac{\left(1 + (c-1)\sqrt{1-\underline{m}_1^2}\right)^2}{c^2\left(\underline{m}_1 - \underline{m}_2\right)^2\left(1 - \underline{m}_1^2\right)}\right]^{(1)}\right|_{\underline{m}_1=0} = \frac{2}{\underline{m}_2^3},$$

then, the first term of $Var\left(X_f\right)$ equals to

$$
\begin{aligned}
&2 \cdot \frac{1}{2\pi i} \oint \frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}_2^2}}\right)^2}{\underline{m}_2^2} \cdot \frac{2}{\underline{m}_2^3} d\underline{m}_2 \\
&= 4 \cdot \frac{1}{4!} \left[\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}_2^2}}\right)^2\right]^{(4)}\Bigg|_{\underline{m}_2=0} = \frac{1+3c}{c^2}
\end{aligned}
$$

As for the second term of $Var\left(X_f\right)$, we have,

$$
\begin{aligned}
&- \frac{y(\nu_4 - 3)}{4\pi^2} \oint\oint f(z_1)g(z_2)\left(\int \frac{t}{(t\underline{m}(z_1)+1)^2} \cdot \frac{t}{(t\underline{m}(z_2)+1)^2} \, \mathrm{d}H(t)\right) d\underline{m}(z_1) d\underline{m}(z_2) \\
&= -\frac{(\nu_4 - 3)}{4\pi^2 c} \oint \frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}_2^2}}\right)^2}{\underline{m}_2^2} d\underline{m}_2 \oint \frac{\left(1 - \frac{1}{c} + \frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2}{\underline{m}_1^2} \left(\int \frac{t^2}{(t\underline{m}_1+1)^2(t\underline{m}_2+1)^2} \, \mathrm{d}H(t)\right) d\underline{m}_1,
\end{aligned}
$$

since

$$
\begin{aligned}
&\int \frac{t^2}{(t\underline{m}(z_1)+1)^2(t\underline{m}(z_2)+1)^2} \, \mathrm{d}H(t) = \int_{-1}^{1} \frac{t^2}{(t\underline{m}_1+1)^2(t\underline{m}_2+1)^2} \cdot \frac{1}{\pi\sqrt{1-t^2}} \, \mathrm{d}t \\
&= \Bigg[\underline{m}_2\underline{m}_1^2\left(-1+2\underline{m}_2^2\right)\sqrt{1-\underline{m}_1^2} + \underline{m}_1\left(\sqrt{1-\underline{m}_1^2} - \left(\sqrt{1-\underline{m}_2^2}\right)^3\right) \\
&- \underline{m}_1^3\left(\sqrt{1-\underline{m}_1^2} - 2\left(\sqrt{1-\underline{m}_2^2}\right)^3\right) + \underline{m}_2\left(\sqrt{1-\underline{m}_1^2} - \sqrt{1-\underline{m}_2^2}\right) \\
&+ \underline{m}_2^3\left(-2\sqrt{1-\underline{m}_1^2} + \sqrt{1-\underline{m}_2^2}\right)\Bigg] \Bigg/ \left[\left(\underline{m}_1 - \underline{m}_2\right)^3 \left(\sqrt{1-\underline{m}_1^2}\right)^3 \left(\sqrt{1-\underline{m}_2^2}\right)^3\right],
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{2\pi i}\oint \frac{\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2}{\underline{m}_1^2}\left(\int \frac{t^2}{(t\underline{m}_1+1)^2(t\underline{m}_2+1)^2}\,\mathrm{d}H(t)\right)d\underline{m}_1 \\
&= \frac{1}{2\pi i}\oint \frac{\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2}{\underline{m}_1^2}\cdot \frac{\underline{m}_1\left(\sqrt{1-\underline{m}_1^2}-\left(\sqrt{1-\underline{m}_2^2}\right)^3\right)}{\left(\underline{m}_1-\underline{m}_2\right)^3\left(\sqrt{1-\underline{m}_1^2}\right)^3\left(\sqrt{1-\underline{m}_2^2}\right)^3}d\underline{m}_1 \\
&+ \frac{1}{2\pi i}\oint \frac{\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2}{\underline{m}_1^2}\cdot \frac{\sqrt{1-\underline{m}_1^2}\left(\underline{m}_2-2\underline{m}_2^3\right)-\underline{m}_2\left(\sqrt{1-\underline{m}_2^2}\right)^3}{\left(\underline{m}_1-\underline{m}_2\right)^3\left(\sqrt{1-\underline{m}_1^2}\right)^3\left(\sqrt{1-\underline{m}_2^2}\right)^3}d\underline{m}_1, \\
&= \left.\frac{\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2\left(\sqrt{1-\underline{m}_1^2}-\left(\sqrt{1-\underline{m}_2^2}\right)^3\right)}{\left(\underline{m}_1-\underline{m}_2\right)^3\left(\sqrt{1-\underline{m}_1^2}\right)^3\left(\sqrt{1-\underline{m}_2^2}\right)^3}\right|_{\underline{m}_1=0} \\
&+ \left.\left[\frac{\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_1^2}}\right)^2\left(\sqrt{1-\underline{m}_1^2}\left(\underline{m}_2-2\underline{m}_2^3\right)-\underline{m}_2\left(\sqrt{1-\underline{m}_2^2}\right)^3\right)}{\left(\underline{m}_1-\underline{m}_2\right)^3\left(\sqrt{1-\underline{m}_1^2}\right)^3\left(\sqrt{1-\underline{m}_2^2}\right)^3}\right]^{(1)}\right|_{\underline{m}_1=0} \\
&= \frac{4}{\underline{m}_2^3}-\frac{4}{\underline{m}_2^3\left(\sqrt{1-\underline{m}_2^2}\right)^3}+\frac{6}{\underline{m}_2\left(\sqrt{1-\underline{m}_2^2}\right)^3},
\end{aligned}
$$

then

$$
\begin{aligned}
&\frac{\nu_4-3}{c}\cdot\frac{1}{2\pi i}\oint \frac{\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_2^2}}\right)^2}{\underline{m}_2^2}\cdot\left(\frac{4}{\underline{m}_2^3}-\frac{4}{\underline{m}_2^3\left(\sqrt{1-\underline{m}_2^2}\right)^3}+\frac{6}{\underline{m}_2\left(\sqrt{1-\underline{m}_2^2}\right)^3}\right)d\underline{m}_2 \\
&= \frac{\nu_4-3}{c}\cdot\frac{4}{4!}\cdot\left.\left[\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_2^2}}\right)^2\left(1-\frac{1}{\left(\sqrt{1-\underline{m}_2^2}\right)^3}\right)\right]^{(4)}\right|_{\underline{m}_2=0} \\
&+ \frac{\nu_4-3}{c}\cdot\frac{6}{2!}\cdot\left.\left[\left(1-\frac{1}{c}+\frac{1}{c\sqrt{1-\underline{m}_2^2}}\right)^2\cdot\frac{1}{\left(\sqrt{1-\underline{m}_2^2}\right)^3}\right]^{(2)}\right|_{\underline{m}_2=0} = \frac{3(\nu_4-3)}{2c}.
\end{aligned}
$$

Hence,

$$Var\left(X_f\right) = \frac{1+3c}{c^2} + \frac{3(\nu_4-3)}{2c} = \frac{1+\frac{3(\nu_4-1)}{2}c}{c^2}.$$

therefore,

$$\widehat{L}_\tau - \frac{T-\tau}{2} \xrightarrow{d} \mathcal{N}\left(\frac{1}{2c}, \frac{1+\frac{3(\nu_4-1)}{2}c}{c^2}\right),$$

thus the high dimensional asymptotic normality in Theorem 2.1 follows, i.e.

$$\frac{T}{p}\widetilde{L}_\tau - \frac{p}{2} \xrightarrow{d} \mathcal{N}\left(\frac{1}{2}, 1+\frac{3(\nu_4-1)}{2}c\right).$$

7.3. *Proof for Theorem 2.2*

In the paper Li and Yao (2015), we consider the re-normalized sample covariance matrix

$$\widetilde{A} = \sqrt{\frac{1}{n}}\left(\frac{1}{\sqrt{tr(\Sigma_p^2)}}Z^*\Sigma_p Z - \frac{tr(\Sigma_p)}{\sqrt{tr(\Sigma_p^2)}}I_n\right), \tag{7.5}$$

where $Z = (z_{ij})_{p\times n}$ and $z_{ij}, i = 1, \cdots, p,\ j = 1, \cdots, n$ are i.i.d. real random variables with mean zero and variance one, I_n is the identity matrix of order n, Σ_p is a sequence of $p\times p$ non-negative definite matrices with bounded spectral norm. Assume the following limit exist,

(a) $\gamma = \lim_{p\to\infty} \frac{1}{p} tr(\Sigma_p)$,
(b) $\theta = \lim_{p\to\infty} \frac{1}{p} tr(\Sigma_p^2)$,
(c) $\omega = \lim_{p\to\infty} \frac{1}{p} \sum_{i=1}^{p} (\Sigma_{ii})^2$,

it can be proved that, under the ultra-dimensional setting ($p/n \to \infty, p, n \to \infty$), with probability one, the ESD of matrix $\widetilde{A}$, $F^{\widetilde{A}}$ converges to the semicircle law F with density

$$F'(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & \text{if } |x| \le 2, \\ 0, & \text{if } |x| > 2. \end{cases}$$

We denote the Stieltjes transform of the semicircle law F by $m(z)$. Let $\mathscr{S}$ denote any open region on the complex plane including $[-2, 2]$, the support of F and $\mathscr{M}$ be the set of functions which are analytic on $\mathscr{S}$. For any $f \in \mathscr{M}$, denote

$$G_n(f) \triangleq n\int_{-\infty}^{+\infty} f(x)d\left(F^{\widetilde{A}}(x) - F(x)\right) - \sqrt{\frac{n^3}{p}}\Phi_3(f)$$

where

$$\Phi_k(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos(\theta))\cos(k\theta)\,\mathrm{d}\theta,$$

the central limit theorem (CLT) of linear functions of eigenvalues of the re-normalized sample covariance matrix $\widetilde{A}$ when the dimension p is much larger than the sample size n is stated as follows.

Theorem 7.2. *Suppose that*

(1) $\mathbf{Z} = (z_{ij})_{p\times n}$ *where* $\{z_{ij} : i = 1, \cdots, p;\ j = 1, \cdots, n\}$ *are i.i.d. real random variables with* $\mathbb{E}z_{ij} = 0$, $\mathbb{E}z_{ij}^2 = 1$ *and* $\nu_4 = \mathbb{E}z_{ij}^4 < \infty$.

(2) Σ_p *is a sequence of* $p \times p$ *non-negative definite matrices with bounded spectral norm. Assume the following limit exist,*

 (a) $\gamma = \lim_{p\to\infty} \frac{1}{p}\, tr(\Sigma_p)$,

 (b) $\theta = \lim_{p\to\infty} \frac{1}{p}\, tr(\Sigma_p^2)$,

 (c) $\omega = \lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p}(\Sigma_{ii})^2$,

(3) $p/n \to \infty$ *as* $n \to \infty$, $n^3/p = O(1)$.

Then, for any $f_1, \cdots, f_k \in \mathscr{M}$*, the finite dimensional random vector* $(G_n(f_1), \cdots, G_n(f_k))$ *converges weakly to a Gaussian vector* $(Y(f_1), \cdots, Y(f_k))$ *with mean function*

$$\mathbb{E}Y(f) = \frac{1}{4}(f(2) + f(-2)) - \frac{1}{2}\Phi_0(f) + \frac{\omega}{\theta}(\nu_4 - 3)\Phi_2(f),$$

and covariance function

$$\begin{aligned} cov\,(Y(f_1), Y(f_2)) &= \frac{\omega}{\theta}(\nu_4 - 3)\Phi_1(f_1)\Phi_1(f_2) + 2\sum_{k=1}^{\infty} k\Phi_k(f_1)\Phi_k(f_2) \\ &= \frac{1}{4\pi^2}\int_{-2}^{2}\int_{-2}^{2} f_1'(x)f_2'(y)H(x,y)\,\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{7.6}$$

where

$$\Phi_k(f) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos\theta)e^{ik\theta}\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos\theta)\cos k\theta\,\mathrm{d}\theta,$$

$$H(x,y) = \frac{\omega}{\theta}(\nu_4 - 3)\sqrt{4-x^2}\sqrt{4-y^2} + 2\log\left(\frac{4 - xy + \sqrt{(4-x^2)(4-y^2)}}{4 - xy - \sqrt{(4-x^2)(4-y^2)}}\right).$$

Another useful lemma in Li and Yao (2015) derived from Theorem 7.2 is as follows:

Lemma 7.2. *Let* $\{\widetilde{\lambda}_i,\ 1 \le i \le n\}$ *be eigenvalues of matrix* $\widetilde{A} = \sqrt{\frac{1}{n}}\left(\frac{1}{\sqrt{tr(\Sigma_p^2)}}Z^*\Sigma_p Z - \frac{tr(\Sigma_p)}{\sqrt{tr(\Sigma_p^2)}}I_n\right)$*, where* Z, Σ_p *satisfies the assumptions in Theorem 7.2, then*

$$\begin{pmatrix} \sum_{i=1}^{n}\widetilde{\lambda}_i^2 - n - \left(\frac{\omega}{\theta}(\nu_4-3)+1\right) \\ \sum_{i=1}^{n}\widetilde{\lambda}_i \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & \frac{\omega}{\theta}(\nu_4-3)+2 \end{pmatrix}\right)$$

as $p/n \to \infty$, $n \to \infty$, $n^3/p = O(1)$.

Note that

$$\widetilde{L}_\tau = \mathrm{Tr}(\widetilde{M}_\tau^*\widetilde{M}_\tau),$$

$$\begin{aligned}
\widetilde{M}_\tau &= \frac{1}{2T}\sum_{t=1}^{T}\left(\mathbf{x}_t\mathbf{x}_{t-\tau}^* + \mathbf{x}_{t-\tau}\mathbf{x}_t^*\right),\\
&= \frac{1}{2T}\left(\mathbf{x}_1,\cdots,\mathbf{x}_T\right)\left(D_\tau + D_\tau^*\right)\left(\mathbf{x}_1,\cdots,\mathbf{x}_T\right)^*\\
&= \frac{1}{2T}X_T\left(D_\tau + D_\tau^*\right)X_T^*,
\end{aligned}$$

where permutation matrix

$$D_1 = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}_{T\times T}, \qquad D_\tau = D_1^\tau = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \\ \vdots & 0 & & \ddots & \vdots \\ 1 & & \ddots & & 1 \\ \vdots & \ddots & & 0 & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}_{T\times T}$$

satisfies

$$D_1 D_1^* = D_1^* D_1 = \mathbf{I}_T,$$

then when $\mathbf{x}_t = \mathbf{z}_t$, $\mathcal{M}_q$ conforms to the quadratic form (7.5) and $\mathcal{L}_q$ can be seen as a linear function of eigenvalues of $\mathcal{M}_q$. Therefore results in Theorem 7.2 and Lemma 7.2 can be directly applied to derive the low dimensional asymptotic of our single-lagged test statistic $\widetilde{L}_\tau$. Specifically, (p,T) corresponds to (n,p) in $\widetilde{A} = \sqrt{\frac{1}{n}}\left(\frac{1}{\sqrt{tr(\Sigma_p^2)}}Z^*\Sigma_p Z - \frac{tr(\Sigma_p)}{\sqrt{tr(\Sigma_p^2)}}I_n\right)$ and $\Sigma_p = \frac{1}{2}\left(D_\tau + D_\tau^*\right)$. Henceforth,

$$\begin{aligned}
\gamma &= \lim_{p\to\infty}\frac{1}{p}tr(\Sigma_p) = 0,\\
\omega &= \lim_{p\to\infty}\frac{1}{p}\sum_{i=1}^{p}(\Sigma_{ii})^2 = 0,\\
\theta &= \lim_{p\to\infty}\frac{1}{p}tr(\Sigma_p^2) = \lim_{T\to\infty}\frac{1}{T}tr\left(\frac{1}{4}\left(D_\tau + D_\tau^*\right)^2\right) = \frac{1}{2},
\end{aligned}$$

$$\widetilde{A} = \sqrt{\frac{1}{n}}\left(\frac{1}{\sqrt{tr(\Sigma_p^2)}}Z^*\Sigma_p Z - \frac{tr(\Sigma_p)}{\sqrt{tr(\Sigma_p^2)}}I_n\right) = \frac{1}{\sqrt{p\cdot\frac{T}{2}}}X_T\left(\frac{1}{2}\left(D_\tau + D_\tau^*\right)\right)X_T^* = \sqrt{\frac{2T}{p}}\,\widetilde{M}_\tau,$$

therefore, according to lemma 7.2,

$$\sum_{i=1}^{n}\widetilde{\lambda}_i^2 = tr\left(\widetilde{A}\,\widetilde{A}^*\right) = \frac{2T}{p}\; tr\left(\widetilde{M}_\tau\widetilde{M}_\tau^*\right) = \frac{2T}{p}\widetilde{L}_\tau,$$

since

$$\sum_{i=1}^{n}\widetilde{\lambda}_i^2 - n - \left(\frac{\omega}{\theta}(\nu_4 - 3) + 1\right) \xrightarrow{d} \mathcal{N}(0,4),$$

$$\frac{2T}{p}\widetilde{L}_\tau - p - 1 \xrightarrow{d} \mathcal{N}(0,4),$$

i.e.

$$\frac{T}{p}\widetilde{L}_\tau - \frac{p}{2} \xrightarrow{d} \mathcal{N}\left(\frac{1}{2},\ 1\right).$$

References

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

Bai, Zhidong D., and Jack W. Silverstein.(2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probability* 32.1 (2004): 553.

Zhidong Bai, Dandan Jiang, Jianfeng Yao and Shurong Zheng(2009). Corrections to LRT on Large Dimensional Covariance Matrix by RMT. *Annals of Statistics* 37(6B), 3822-3840

Bai, Zhidong, and Chen Wang.(2015). A note on the limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Statistics & Probability Letters* 96: 333-340.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Forthcoming* in *Ann. Statist.*

Bickel, P. J. and Gel, Y. R. (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society B*, **73**, 549-92.

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

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

Guo, S., Wang, Y. & Yao, Q.(2016). High dimensional and banded vector autoregressions. *A preprint.*

Gray, Robert M. (2006). Toeplitz and Circulant Matrices: A Review. *Now Publishers Inc.*

Grenander, U., Szegö, G.(1958). Toeplitz Forms and Their Applications. In: California Monographs in Mathematical Sciences. *University of California Press*, Berkeley.

Han, F. and Liu, H.(2013). A direct estimation of high dimensional stationary vector autoregressions. *Manuscript.*

Haufe, S., Nolte, G., Mueller, K. R., and Krämer, N. (2009). Sparse causal discovery in multivariate time series. `http://arxiv.org/abs/0901.2234`.

Hosking, J. R. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, **75** (371), 602-608.

Hsu, N. J., Hung, H. L., and Chang, Y. M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics and Data Analysis*, **52**, 3645-3657.

Jin, B., Wang, C., Bai, Z. D., Nair, K. K., Harding, M.(2014). Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *The Annals of Applied Probability*, **24(3)**, 1199-1225.

Johnstone, I. M. (2007). High dimensional statistical inference and random matrices. *International Congress of Mathematicians*, **I**, 307-333. Zürich, Switzerland: European Mathematical Society.

Lam, C. (2015). Nonparametric Eigenvalue-Regularized Precision or Covariance Matrix Estimator. *Annals of Statistics, to appear*.

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

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

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

Li, Z., and Yao, J. (2015). Testing the Sphericity of a covariance matrix when the dimension is much larger than the sample size. *arXiv preprint* arXiv:1508.02498.

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

Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, **150**, 1-29.

Sarkar, S. K. and Chang, C. K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, **92(440)**, 1601-1608.

Shojaie, A. and Michailidis, G. (2010). Discovering graphical Granger causality using the truncating LASSO penalty. *Bioinformatics*, **26**, 517-523.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73(3)**, 751-754.

M.S. Srivastava(2005). Some tests concerning the covariance matrix in high-dimensional data, *Journal of The Japan Statistical Society*, 35,251-272.

Wang, Q. and Yao, J.(2013). On the sphericity test with large-dimensional observations. *Electronic Journalof Statistics* **7**,2164-2192.

Yao, J., Zheng, S. and Bai. Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press, New York and London.

Zheng, S., Bai, Z., and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics*, 43(2), 546-591.

TABLE 1
Empirical sizes for the four test statistics

			ϕ_τ	U_q		$\widetilde{Q}_q$		Q^*_q	
p	T	p/T	$\tau = 1$	$q = 1$	$q = 3$	$q = 1$	$q = 3$	$q = 1$	$q = 3$
5	1000	0.005	0.0756	0.0646	0.0548	0.0490	0.0478	0.0488	0.0476
10	2000	0.005	0.0696	0.0564	0.0438	0.0492	0.0440	0.0492	0.0436
25	5000	0.005	0.0604	0.0568	0.0500	0.0498	0.0528	0.0498	0.0528
40	8000	0.005	0.0568	0.0524	0.0450	0.0508	0.0520	0.0508	0.0520
10	1000	0.01	0.0640	0.0602	0.0536	0.0472	0.0468	0.0470	0.0464
20	2000	0.01	0.0620	0.0494	0.0438	0.0502	0.0530	0.0502	0.0530
50	5000	0.01	0.0512	0.0518	0.0480	0.0488	0.0498	0.0488	0.0498
80	8000	0.01	0.0552	0.0506	0.0450	0.0464	0.0406	0.0464	0.0404
50	1000	0.05	0.0588	0.0510	0.0430	0.0408	0.0466	0.0408	0.0466
100	2000	0.05	0.0560	0.0488	0.0514	0.0432	0.0414	0.0432	0.0414
250	5000	0.05	0.0542	0.0470	0.0406	0.0456	0.0436	0.0456	0.0434
400	8000	0.05	0.0512	0.0490	0.0406	0.0418	0.0410	0.0418	0.0410
10	100	0.1	0.0714	0.0582	0.0518	0.0300	0.0400	0.0280	0.0362
40	400	0.1	0.0574	0.0502	0.0480	0.0362	0.0342	0.0358	0.0338
60	600	0.1	0.0562	0.0486	0.0446	0.0340	0.0340	0.0340	0.0338
100	1000	0.1	0.0564	0.0474	0.0504	0.0370	0.0268	0.0366	0.0264
50	100	0.5	0.0510	0.0580	0.0562	0.0006	0.0018	0.0006	0.0018
200	400	0.5	0.0522	0.0498	0.0462	0.0010	0.0004	0.0010	0.0004
300	600	0.5	0.0520	0.0430	0.0410	0.0002	0.0008	0.0002	0.0008
500	1000	0.5	0.0514	0.0430	0.0438	0	0	0	0
90	100	0.9	0.0508	0.0472	0.0558	0	0	0	0
180	200	0.9	0.0516	0.0470	0.0416	0	0	0	0
540	600	0.9	0.0518	0.0494	0.0462	0	0	0	0
900	1000	0.9	0.0550	0.0454	0.0432	0	0	0	0

TABLE 2
Power comparison for the three test statistics

			ϕ_τ	$\widetilde{Q}_q$		Q^*_q	
p	T	p/T	$\tau = 1$	$q = 1$	$q = 3$	$q = 1$	$q = 3$
10	100	0.1	0.3022	0.0952	0.0952	0.0914	0.0926
20	200	0.1	0.6488	0.2392	0.1994	0.2362	0.1958
40	400	0.1	0.9828	0.6638	0.5410	0.6622	0.5380
60	600	0.1	1	0.9406	0.8452	0.9404	0.8448
100	1000	0.1	1	1	0.9982	1	0.9982
50	100	0.5	0.4094	0.0014	0.0060	0.0014	0.0052
100	200	0.5	0.8548	0.0036	0.0208	0.0030	0.0194
200	400	0.5	0.9998	0.0330	0.2022	0.0328	0.1994
300	600	0.5	1	0.1156	0.6348	0.1138	0.6306
500	1000	0.5	1	0.5816	0.9974	0.5806	0.9972
80	100	0.8	0.4798	0	0	0	0
160	200	0.8	0.9158	0	0	0	0
320	400	0.8	1	0	0	0	0
480	600	0.8	1	0	0	0	0
800	1000	0.8	1	0.0004	0.0038	0.0004	0.0032
90	100	0.9	0.4950	0	0	0	0
180	200	0.9	0.9344	0	0	0	0
360	400	0.9	1	0	0	0	0
540	600	0.9	1	0	0	0	0
900	1000	0.9	1	0	0	0	0

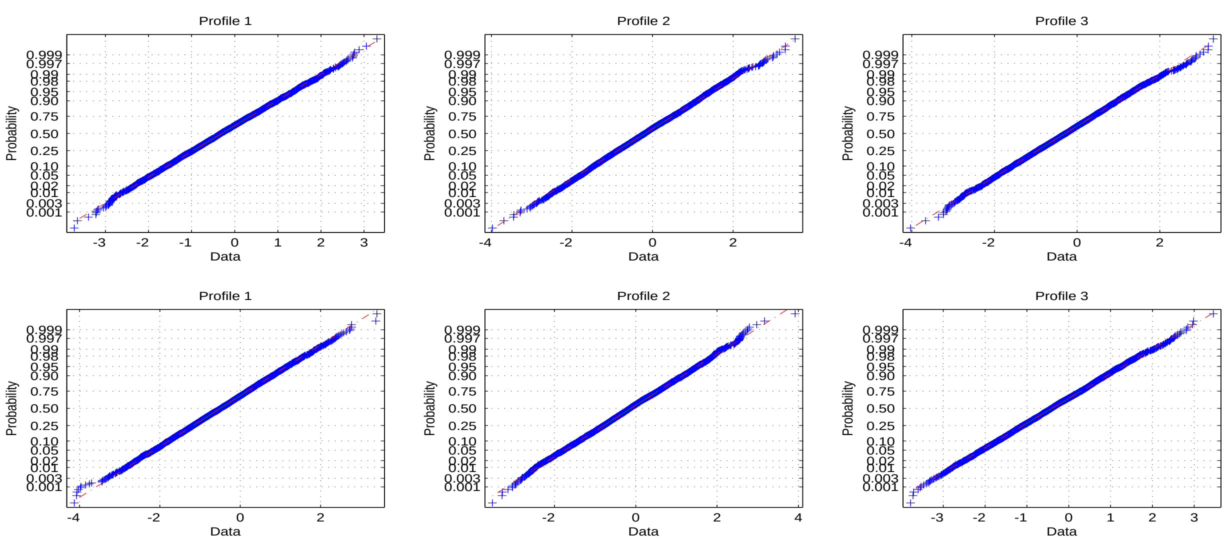

Fig 1: Normal QQ-plot of the statistic G_5 after standardization. Upper row: The z_{it}'s are standard normal. Lower row: The z_{it}'s are $U(-2\sqrt{3}, 2\sqrt{3})$. Left panel: Profile 1. Middle panel: Profile 2. Right panel: Profile 3.

TABLE 3
Adjusted powers of $\widetilde{Q}_q$ and Q_q^* compared to powers of ϕ_τ and U_q

			ϕ_τ	U_q		$\widetilde{Q}_q$		Q_q^*	
p	T	p/T	$\tau = 1$	$q = 1$	$q = 3$	$q = 1$	$q = 3$	$q = 1$	$q = 3$
5	1000	0.005	0.6634	0.8024	0.6364	0.4764	0.2672	0.4764	0.2680
10	2000	0.005	0.9824	0.9994	0.9878	0.9048	0.6344	0.9048	0.6350
25	5000	0.005	1	1	1	1	0.9996	1.0000	0.9996
40	8000	0.005	1	1	1	1	1	1	1
10	1000	0.01	0.6940	0.8338	0.6538	0.4606	0.2630	0.4606	0.2634
20	2000	0.01	0.9940	1	0.9918	0.9286	0.6314	0.9286	0.6314
50	5000	0.01	1	1	1	1	0.9996	1	0.9996
80	8000	0.01	1	1	1	1	1	1	1
50	1000	0.05	0.8166	0.8492	0.6638	0.4868	0.3010	0.4868	0.3012
100	2000	0.05	0.9992	1	0.9960	0.9326	0.6916	0.9326	0.6926
250	5000	0.05	1	1	1	1	1	1	1
400	8000	0.05	1	1	1	1	1	1	1
10	100	0.1	0.3154	0.0932	0.0848	0.1392	0.1218	0.1392	0.1214
40	400	0.1	0.9816	0.2910	0.1958	0.7082	0.5968	0.7082	0.5988
60	600	0.1	1	0.4948	0.3440	0.9598	0.8808	0.9598	0.8816
100	1000	0.1	1	0.8628	0.6662	1	0.9992	1	0.9992
200	2000	0.1	1	1	0.9944	1	1	1	1
50	100	0.5	0.4164	0.0914	0.0766	0.1004	0.1376	0.1004	0.1380
200	400	0.5	0.9998	0.2970	0.1942	0.4012	0.7708	0.4012	0.7708
300	600	0.5	1	0.5290	0.3388	0.6626	0.9746	0.6626	0.9748
500	1000	0.5	1	0.8812	0.6718	0.9666	1	0.9666	1
1000	2000	0.5	1	1	0.9948	1	1	1	1
90	100	0.9	0.4878	0.0858	0.0706	0.1384	0.0992	0.1384	0.1002
360	400	0.9	1	0.3116	0.2008	0.7138	0.5172	0.7138	0.5176
540	600	0.9	1	0.5304	0.3384	0.9496	0.8368	0.9496	0.8368
900	1000	0.9	1	0.9006	0.6600	0.9998	0.9966	0.9998	0.9966
1800	2000	0.9	1	1	0.9944	1	1	1	1

TABLE 4
Size and power of ϕ_τ and John's test with Simes method when $c > 1$

			Size			**Power**		
p	T	p/T	$\phi_\tau\ (\tau = 1)$	$U_q(q = 1)$	$U_q(q = 3)$	$\phi_\tau\ (\tau = 1)$	$U_q(q = 1)$	$U_q(q = 3)$
150	100	1.5	0.0570	0.0544	0.0532	0.6084	0.0938	0.0798
600	400	1.5	0.0528	0.0512	0.0432	1	0.3250	0.1998
900	600	1.5	0.0518	0.0496	0.0470	1	0.5692	0.3476
1500	1000	1.5	0.0516	0.0498	0.0474	1	0.9206	0.6714
3000	2000	1.5	0.0526	0.0472	0.0398	1	1	0.9966
200	100	2	0.0580	0.0536	0.0544	0.7110	0.0944	0.0814
800	400	2	0.0504	0.0454	0.0464	1	0.3460	0.1918
1200	600	2	0.0526	0.0466	0.0382	1	0.5878	0.3282
2000	1000	2	0.0512	0.0428	0.0444	1	0.9276	0.6608
4000	2000	2	0.0474	0.0470	0.0476	1	1	0.9960
500	100	5	0.0552	0.0576	0.0572	0.9432	0.1112	0.0756
2000	400	5	0.0506	0.0502	0.0490	1	0.4328	0.1998
3000	600	5	0.0494	0.0442	0.0378	1	0.7284	0.3372
5000	1000	5	0.0482	0.0486	0.0432	1	0.9812	0.6670
10000	2000	5	0.0510	0.0466	0.0410	1	1	0.9946

Table 5
Empirical mean, variance and 95 percentile of $\phi_\tau = \frac{T}{p}\widetilde{L}_\tau - \frac{p}{2}$

			Mean		**Variance**		**Quantile**	
p	T	p/T	$\tau = 2$	$\tau = 4$	$\tau = 2$	$\tau = 4$	$\tau = 2$	$\tau = 4$
2	500	0.004	0.49	0.48	1.49	1.51	2.92	2.90
8	2000	0.004	0.48	0.50	1.09	1.17	2.37	2.39
20	5000	0.004	0.51	0.51	1.07	1.12	2.26	2.27
32	8000	0.004	0.52	0.53	1.06	1.08	2.27	2.24
(Theory for $c = 0.004$**)**			**0.50**		**1.00**		**2.14**	
20	500	0.04	0.49	0.50	1.13	1.20	2.31	2.35
80	2000	0.04	0.54	0.48	1.14	1.15	2.31	2.27
200	5000	0.04	0.51	0.51	1.12	1.13	2.25	2.26
320	8000	0.04	0.50	0.50	1.14	1.10	2.26	2.23
(Theory for $c = 0.04$**)**			**0.50**		**1.12**		**2.24**	
50	100	0.5	0.50	0.47	2.67	2.62	3.23	3.24
200	400	0.5	0.49	0.48	2.65	2.57	3.22	3.15
400	800	0.5	0.52	0.51	2.51	2.49	3.10	3.15
500	1000	0.5	0.52	0.49	2.47	2.51	3.09	3.15
(Theory for $c = 0.5$**)**			**0.50**		**2.50**		**3.10**	
100	100	1	0.53	0.55	4.03	4.09	3.83	3.87
400	400	1	0.52	0.54	4.00	4.08	3.84	3.95
800	800	1	0.47	0.45	4.09	3.88	3.81	3.65
1000	1000	1	0.48	0.50	4.05	3.99	3.77	3.82
(Theory for $c = 1$**)**			**0.50**		**4.00**		**3.79**	
200	100	2	0.53	0.49	7.42	7.10	5.05	4.96
800	400	2	0.53	0.55	7.21	7.19	4.89	5.03
1600	800	2	0.51	0.52	7.14	7.12	4.87	5.02
2000	1000	2	0.49	0.52	7.32	7.02	4.93	4.93
(Theory for $c = 2$**)**			**0.50**		**7.00**		**4.85**	
500	100	5	0.45	0.45	15.59	15.60	7.02	7.12
2000	400	5	0.49	0.53	15.22	16.06	6.97	7.01
4000	800	5	0.50	0.50	15.64	15.76	6.94	7.05
5000	1000	5	0.49	0.53	16.38	16.06	7.18	7.15
(Theory for $c = 5$**)**			**0.50**		**16.00**		**7.08**	

TABLE 6
Relative errors for the mean, variance and 95 percentile for Hosking's statistic $\widetilde{Q}_q$ and Li-McLeod statistic Q_q^* (with $q = 3$)

			$\widetilde{Q}_q$			Q_q^*		
p	T	p/T	**Mean**	**Variance**	**95%Quantile**	**Mean**	**Variance**	**95%Quantile**
10	100	0.1	0.234%	19.976%	1.366%	0.234%	24.922%	1.547%
20	200	0.1	0.067%	30.862%	0.993%	0.067%	33.526%	1.049%
40	400	0.1	-0.015%	22.057%	0.253%	-0.015%	23.286%	0.265%
60	600	0.1	0.000%	21.457%	0.162%	0.000%	22.269%	0.162%
100	1000	0.1	0.007%	20.666%	0.125%	0.007%	21.153%	0.125%
50	100	0.5	0.041%	267.179%	1.322%	0.041%	282.546%	1.354%
100	200	0.5	0.007%	284.025%	0.655%	0.007%	291.875%	0.662%
200	400	0.5	0.000%	289.080%	0.330%	0.000%	292.998%	0.330%
300	600	0.5	0.000%	297.059%	0.222%	0.000%	299.734%	0.222%
500	1000	0.5	0.000%	296.364%	0.134%	0.000%	297.941%	0.134%
80	100	0.8	0.010%	1742.257%	1.289%	0.005%	1820.096%	1.300%
160	200	0.8	0.000%	2020.024%	0.655%	0.000%	2063.959%	0.657%
320	400	0.8	0.000%	2214.386%	0.332%	0.000%	2237.811%	0.332%
480	600	0.8	0.001%	2266.151%	0.223%	0.001%	2282.093%	0.223%
800	1000	0.8	0.000%	2348.823%	0.137%	0.000%	2358.701%	0.137%
90	100	0.9	0.004%	5382.234%	1.292%	0.000%	5618.993%	1.297%
180	200	0.9	0.000%	6906.920%	0.657%	0.000%	7053.897%	0.658%
360	400	0.9	0.000%	8110.500%	0.332%	0.000%	8195.108%	0.332%
540	600	0.9	0.000%	8705.234%	0.222%	0.000%	8764.569%	0.222%
900	1000	0.9	0.000%	9170.563%	0.133%	0.000%	9208.205%	0.133%

TABLE 7
Empirical mean, variance and 95 percentile of the test statistic $c_p^{-2}\widehat{\phi}_\tau$ using standardized data compared to the benchmark values from $c_p^{-2}\phi_\tau$ (in bold)

			Mean	**Variance**	**95% Quantile**				**Mean**	**Variance**	**95% Quantile**
p	T	p/T	$\tau = 1$	$\tau = 1$	$\tau = 1$	p	T	p/T	$\tau = 1$	$\tau = 1$	$\tau = 1$
10	100	0.1	4.72	117.02	23.95	150	100	1.5	0.35	1.14	2.15
40	400	0.1	5.08	111.43	22.92	600	400	1.5	0.34	1.12	2.06
80	800	0.1	5.05	109.49	22.30	1200	800	1.5	0.33	1.13	2.10
100	1000	0.1	4.95	111.24	22.62	1500	1000	1.5	0.33	1.08	2.01
(Theory for $c = 0.1$**)**			**5**	**130**	**23.75**	**(Theory for** $c = 1.5$**)**			**0.33**	**2.44**	**2.91**
50	100	0.5	1.03	5.99	5.09	500	100	5	0.13	0.25	0.94
200	400	0.5	1.03	6.14	5.13	2000	400	5	0.10	0.24	0.93
400	800	0.5	1.07	5.96	5.02	4000	800	5	0.11	0.25	0.93
500	1000	0.5	1.01	6.02	5.13	5000	1000	5	0.11	0.23	0.91
(Theory for $c = 0.5$**)**			**1**	**10**	**6**	**(Theory for** $c = 5$**)**			**0.1**	**0.64**	**1.42**

Table 8
Empirical size and power for $\widehat{\phi}_\tau(\tau = 1)$ with standardized data and multi-lag-q test statistic $\widehat{U}_q$ with Simes' method

			Size			**Power**		
p	T	c_p	$\hat{\phi}_\tau\,(\tau = 1)$	$\widehat{U}_q(q = 1)$	$\widehat{U}_q(q = 3)$	$\hat{\phi}_\tau\,(\tau = 1)$	$\widehat{U}_q(q = 1)$	$\widehat{U}_q(q = 3)$
10	100	0.1	0.052	0.054	0.057	0.272	0.077	0.072
20	200	0.1	0.044	0.052	0.048	0.657	0.115	0.109
40	400	0.1	0.040	0.047	0.046	0.990	0.214	0.186
60	600	0.1	0.040	0.052	0.044	1	0.354	0.297
100	1000	0.1	0.042	0.046	0.041	1	0.677	0.591
200	2000	0.1	0.037	0.049	0.042	1	0.995	0.986
50	100	0.5	0.017	0.050	0.054	0.390	0.078	0.072
100	200	0.5	0.016	0.047	0.045	0.908	0.136	0.102
200	400	0.5	0.022	0.049	0.042	1	0.265	0.195
300	600	0.5	0.017	0.045	0.042	1	0.473	0.330
500	1000	0.5	0.019	0.046	0.041	1	0.835	0.645
1000	2000	0.5	0.019	0.050	0.048	1	1	0.992
80	100	0.8	0.018	0.051	0.044	0.499	0.090	0.078
160	200	0.8	0.012	0.049	0.047	0.974	0.130	0.102
320	400	0.8	0.012	0.047	0.045	1	0.286	0.199
480	600	0.8	0.011	0.047	0.045	1	0.489	0.324
800	1000	0.8	0.011	0.047	0.040	1	0.864	0.643
1600	2000	0.8	0.010	0.040	0.040	1	1	0.994
90	100	0.9	0.014	0.049	0.049	0.538	0.083	0.070
180	200	0.9	0.013	0.052	0.049	0.988	0.136	0.105
360	400	0.9	0.015	0.047	0.045	1	0.286	0.186
540	600	0.9	0.011	0.048	0.042	1	0.488	0.328
900	1000	0.9	0.011	0.048	0.041	1	0.870	0.652
1800	2000	0.9	0.010	0.045	0.041	1	1	0.996