Statistica Sinica Preprint No: SS-2024-0037	
Title	Testing High-dimensional White Noise Based On Modified Portmanteau Tests
Manuscript ID	SS-2024-0037
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0037
Complete List of Authors	Zeren Zhou and Min Chen
Corresponding Authors	Min Chen
E-mails	mchen@amss.ac.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Testing High-dimensional White Noise Based On Modified Portmanteau Tests

Zeren Zhou[1] and Min Chen[2]

[1]*Capital University of Economics and Business,*

[2]*Shanxi University,* [2]*Chinese Academy of Sciences*

Abstract: For high-dimensional time series, we propose a new test to detect white noise that is not necessarily assumed to be independent and identically distributed. The test can be viewed as a modified portmanteau test in high dimensions, and the critical value of the test statistic is approximated by a multiplier bootstrap method. We provide asymptotic properties of our test under the null hypothesis. The usefulness of our tests is demonstrated by simulations and one real example, particularly for detecting dense alternatives.

Key words and phrases: high-dimensional time series, white noise, bootstrap, hypothesis testing.

Corresponding author: Min Chen, E-mail: mchen@amss.ac.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

1. Introduction

The high-dimensional time series model has received much attention due to its wide application in economics, finance, biology, environmental studies, etc. The large p/small N problem becomes common with the extensively available dataset due to advances in information technology. The white noise testing is one of the most fundamental statistical problems in time series analysis. For univariate time series, the Box and Pierce portmanteau test (see Box and Pierce (1970)) and the Ljung and Box portmanteau test (see Ljung and Box (1978)) are popular choices for testing white noise, and they are designed for testing whether the first L autocorrelations of a time series are zero. Under certain regularity conditions, the null distribution of their test statistic is χ^2_L. For multivariate analysis, several extensions of the portmanteau test exist, such as the multivariate Box and Pierce portmanteau test in Chitturi (1974), Hosking portmanteau test in Hosking (1980), and Li and McLeod portmanteau test in Li and McLeod (1981). They checked whether the first L autocorrelation matrices are zero, and their null distributions are chi-square under some regularity conditions. Two reasons prevent the application of these methods in testing high-dimensional white noise. The first reason is that the chi-square asymptotic distribution is derived under the assumption that p is fixed, and the second reason

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

is that portmanteau tests require data to be independent and identically distributed (i.i.d.). Data can easily violate those conditions for complicated high-dimensional time series, and the portmanteau test almost loses all power.

Assuming data are i.i.d., several pioneer studies have provided different methods for testing white noise in high-dimensional settings. Li et al. (2019) proposed a test based on the Frobenius norm of the first L lagged sample autocovariance matrixes and established the asymptotic normality of the test statistic under the assumption that $p/N \to c \in (0, +\infty)$. Their method can be viewed as a generalization of the portmanteau test method to the high-dimensional setting. Tsay (2020) proposed a ℓ_∞ type statistic based on the largest value of the Spearman rank autocorrelation matrix and derived its asymptotic distribution under the null hypothesis. Ling et al. (2021) proposed two portmanteau tests on the norm.

The above hypothesis testing methods require a strong assumption that the data are i.i.d. to derive the null distributions. For testing white noise that is not necessary to be i.i.d, Chang et al. (2017) proposed a l_∞-type statistic based on the largest absolute value of autocorrelations matrix and obtained critical value based on a bootstrap method. Wang and Shao (2020) proposed a self-normalized test statistic based on a recursive subsampled

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

U-statistic and derived asymptotic distribution under the null hypothesis. Wang et al. (2022) proposed statistics based on the average of the largest s absolute values of autocorrelation matrices and obtained critical value based on a bootstrap method.

Both Chang et al. (2017) and Wang et al. (2022) proposed l_∞-type statistics to test whether the first L autocorrelation matrices are zero matrices. It is statistical folklore that l_∞-type statistics have high power against sparse alternatives with only a few strong signals and relatively low power against alternatives whose signals are spread out over a large number of coordinates(see Fan et al. (2015);Wang et al. (2015);He et al. (2021), etc). One of the most popular methods for modeling high-dimensional time series is the factor method, which assumes that high-dimensional time series are driven by a small number of factors. There is a large body of literature that discusses the factor model in high-dimensional settings; see Lam and Yao (2012);Daniel Peña and Yohai (2019);Fan et al. (2021);Baltagi et al. (2021) etc. In the factor model, the autocovariance matrices have small but dense elements that spread out over a large number of coordinates, and our simulation results show that hypothesis testing methods based on the ℓ_∞ statistics are less effective.

In this paper, we propose a new method to test white noise that is

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

unnecessary to be i.i.d.. We construct a U-statistic to perform hypothesis testing and obtain critical values based on the bootstrap method. Our method can be viewed as a modified portmanteau test in high-dimension settings. Moreover, Our method is particularly useful for detecting non-white noise under dense alternatives, where signals are spread out over a large number of coordinates. It is worth noting that the bootstrap method has been widely used to validate the portmanteau tests for time series models in both univariate and multivariate cases, see Zhu and Li (2015), Zhu (2016, 2019), Mukherjee (2020), Zhu et al. (2020), Li and Zhang (2022) and many others.

The remainder of this article is organized as follows. Section 2 provides the test statistic and new bootstrap method to approximate the distribution of test statistics. Section 3 investigates asymptotic properties of the new bootstrap method. Section 4 reports the simulation results, and Section 5 considers a real example. Section 6 concludes the paper.

2. Methodolgy

Let $\{X_t : t = 1, \ldots, N\}$ be a weak stationary p-dimensional time series with mean zero and autocovariance matrix at lag l given by $\mathrm{Cov}(X_t, X_{t-l}) = \Sigma_l$. We are interested in testing whether $\{X_t : t = 1, \ldots, N\}$ is white noise.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

$\{X_t : t = 1, \ldots, N\}$ being white noise is equivalent to $\Sigma_l = 0$ for $l \neq 0$. Therefore, for a given L, our hypothesis is

$$H_0 : \Sigma_1 = \cdots = \Sigma_L = 0 \text{ v.s. } H_1 : \text{there exist } l \text{ such that } \Sigma_l \neq 0 \qquad (2.1)$$

2.1 Hypothesis testing

For testing the hypothesis (2.1), the multivariate Box and Pierce portmanteau test in Chitturi (1974) considers statistic $N \sum_{l=1}^{L} \operatorname{tr}\left(\hat{\Sigma}_l^\top \hat{\Sigma}_0^{-1} \hat{\Sigma}_l \hat{\Sigma}_0^{-1}\right)$, where $\hat{\Sigma}_l = \frac{1}{N} \sum_{i=1}^{N-l} X_t X_{t+l}^\top$ is sample autocovariance matrix at lag l. When $p > N$, this statistic is unavailable since $\hat{\Sigma}_0$ is not invertible. Li et al. (2019) proposed a statistic $\sum_{l=1}^{L} \operatorname{tr}\left(\hat{\Sigma}_l^\top \hat{\Sigma}_l\right)$ that can be used when $p > N$; this statistic can be viewed as an extension of portmanteau test in high-dimensional settings. Note that

$$\sum_{l=1}^{L} \operatorname{tr}\left(\hat{\Sigma}_l^\top \hat{\Sigma}_l\right) = \frac{1}{N^2} \sum_{l=1}^{L} \left|\operatorname{vec}\left(\hat{\Sigma}_l\right)\right|^2 = \frac{1}{N^2} \sum_{l=1}^{L} \sum_{i=1} \sum_{j=1} \operatorname{vec}(X_i X_{i+l}^\top)^\top \operatorname{vec}(X_j X_{j+l}^\top),$$

let $Y_{t,l} = \operatorname{vec}\left(X_t X_{t+l}^\top\right)$, then we obtain

$$\sum_{l=1}^{L} \operatorname{tr}\left(\hat{\Sigma}_l^\top \hat{\Sigma}_l\right) = \frac{1}{N^2} \sum_{l=1}^{L} \sum_{i=1} \sum_{j=1} Y_{i,l}^\top Y_{j,l}.$$

In Li et al. (2019), the asymptotic normality was established when the data is i.i.d.. When the data is white noise but not i.i.d., obtaining the asymptotic distribution of this statistic becomes challenging; this challenge

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

mainly arises due to the difficulty in obtaining the asymptotic distribution of the diagonal part $\sum_{l=1}^{L}\sum_{i=1} Y_{i,l}^{\top}Y_{i,l}$.

Inspired by above analysis, we remove the diagonal part of $\sum_{l=1}^{L}\sum_{i=1}\sum_{j=1} Y_{i,l}^{\top}Y_{j,l}$ and propose the following statistics:

$$T = \frac{1}{N}\sum_{l=1}^{L} w_l \sum_{i\neq j} Y_{i,l}^{\top}Y_{j,l}, \tag{2.2}$$

where $w_l > 0$ are pre-selected weights. Let q_α be α quantile of T under H_0, $\mathbb{P}\left(T < q_\alpha \mid H_0\right) = \alpha$, given significant level α, we reject null hypothesis H_0 if $T < q_{\frac{\alpha}{2}}$ or $T > q_{1-\frac{\alpha}{2}}$.

If we set $w_l = 1$, our statistics T can be viewed as a diagonal-removed version of statistic $\sum_{l=1}^{L} \mathrm{tr}\left(\hat{\Sigma}_l^{\top}\hat{\Sigma}_l\right)$. By removing the diagonal part, we can establish the asymptotic distribution of our statistics when data is white noise but not i.i.d.. A similar technique has been employed in Xu et al. (2019). Our statistics can be viewed as modified version of portmanteau test in high-dimensional settings.

Remark1. Our statistics T have a similar form to the statistics considered in Wang and Shao (2020), which was $T_s = \frac{1}{N}\sum_{l=1}^{L}\sum_{|i-j|\geq d} Y_{i,l}^{\top}Y_{j,l}$. In Wang and Shao (2020), a diagonal block of length d was removed from $\sum_{l=1}^{L}\sum_{i=1}\sum_{j=1} Y_{i,l}^{\top}Y_{j,l}$. Wang and Shao (2020) did not obtain the critical value of T_s; consequently, they adopted the self-normalization method to obtain the critical value. To ensure the effectiveness of the self-normalization

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

method, Wang and Shao (2020) required $d \to +\infty$. This reduced the effective sample size and led to a loss of power. Our statistics T preserve the largest effective sample size and are more powerful than the hypothesis test proposed by Wang and Shao (2020), as demonstrated by our simulations.

Remark2. Like univariate portmanteau tests, a proper weight sequence $\{w_l\}$ can improve the finite sample performance of our test; for choice of weight sequence $\{w_l\}$, see the discussion in Section 3.2. Gallagher and Fisher (2015) also provided a useful discussion on the choice of weight sequence.

2.2 Estimate the critical value by bootstrap

Approximating the distribution of T is equivalent to estimating the α quantiles of T under the null hypothesis. Note that statistics T are U-statistics; we adopt the following multiplier bootstrap method to estimate α quantiles.

Let $\{e_t\}_{t=1}^N$ be a sequence of i.i.d. standard normal random variables, i.e $e_t \overset{iid}{\sim} N(0,1)$, which is independent of $\{X_t\}$. Define the bootstrap statistics as follows:

$$T^* = \frac{1}{N}\sum_{l=1}^{L} w_l \sum_{i\neq j} e_i Y_{i,l}^\top Y_{j,l} e_j. \tag{2.3}$$

We use the quantile of T^* as an estimation of critical value.

Our hypothesis testing method is as follows:

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Step 1. Given $L \in \mathbb{N}^+$ and pre-selected weight sequence $\{w_l\}_{l=1}^L$, obtain statistics T based on data $\{X_t : t = 1, \ldots, N\}$ and (2.2).

Step 2. Generate a sequence of i.i.d. random weights $\{e_1, \cdots, e_N\}$, with $e_t \overset{iid}{\sim} N(0,1)$ and that are independent of the sequence $\{X_t\}$. Calculate

$$T^* = \frac{1}{N} \sum_{l=1}^{L} w_l \sum_{i \neq j} e_i Y_{i,l}^\top Y_{j,l} e_j.$$

Step 3. Repeat Step 2 B times to obtain $\{T^{*1}, T^{*2}, \ldots, T^{*B}\}$. Given α, calculate empirical $\frac{\alpha}{2}$ quantile and empirical $1 - \frac{\alpha}{2}$ quantile based on $\{T^{*1}, T^{*2}, \ldots, T^{*B}\}$, denoted as $\hat{q}_{\frac{\alpha}{2}}$ and $\hat{q}_{1-\frac{\alpha}{2}}$.

Step 4. Reject null hypothesis if $T < \hat{q}_{\frac{\alpha}{2}}$ or $T > \hat{q}_{1-\frac{\alpha}{2}}$.

3. Technical Assumptions and Theoretical Results

In this section, we establish the theoretical properties of our hypothesis tests when applied to white noise tests. We demonstrate that, under certain regularity conditions, our hypothesis test can control the probability of type I errors under the null hypothesis. We also provide a detailed discussion regarding the selection of weight sequence $\{w_l\}$ in (2.2).

3.1 Theoretical Results

Assuming the time series $\{X_t\}$ has following form:

$$X_t = f\left(\varepsilon_t, \varepsilon_{t-1}, \ldots\right),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

where f is a measurable function, and $\{\varepsilon_t\}_{t=-\infty}^{+\infty}$ are i.i.d. random elements in some measurable space. The above structure is referred to as time series with physical dependence.

Denote $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ as the σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$. Define $\mathcal{F}_{t,\{k\}} = \sigma(\varepsilon_t, \ldots, \varepsilon_{k+1}, \varepsilon'_k, \varepsilon_{k-1}, \ldots)$, where ε'_k is an i.i.d. copy of ε_k. Let $X_{t,\{k\}} = f\left(\mathcal{F}_{t,\{k\}}\right)$. For a random variable x, denote $\|x\|_q = (\mathbb{E}|x|^q)^{1/q}$. Following Zhang and Cheng (2018), we define the total functional dependence measure for $\{X_t\}$:

$$u_{t,q} = \sup_{1\le l\le L} \sup_{1\le i\le p, 1\le j\le p} \left\|X_{t,i}X_{t+l,j} - X_{t,i,\{0\}}X_{t+l,j,\{0\}}\right\|_q,$$

and

$$U_{m,q} = \sum_{t=m}^{\infty} u_{t,q}.$$

To investigate the theoretical results of our proposed tests, the following regularity conditions are required.

Condition 1. *$U_{1,2} < +\infty$, where $U_{m,q} = \sum_{t=m}^{\infty} u_{t,q}$ and $u_{t,q} = \sup_{1\le l\le L} \sup_{1\le i\le p, 1\le j\le p} \left\|X_{t,i}X_{t+l,j} - X_{t,i,\{0\}}X_{t+l,j,\{0\}}\right\|_q$.*

Condition 2. *Set $\Sigma_0 = \mathrm{Var}(X_t)$, and $\sigma_0 = \mathrm{tr}(\Sigma_0^2)$, there exist a $q = 2+\delta \in (2, 3]$, such that when $N \to +\infty$*

$$\frac{(\sum_{m=1}^{+\infty} \min(m, N)^{\frac{1}{2}-\frac{1}{q}} u_{m,q})^q}{N^\delta \sigma_0} \to 0.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Condition 3. *There exists a constant C such that*

$$\sup_{1\le l\le L}\sup_{1\le i\le p,1\le j\le p}\mathbb{E}\left[|X_{t,i}X_{t,j}X_{t+l,i}X_{t+l,j}|^{q}\right]<C,$$

where q is the same as in condition 2.

Remark3. Condition 1 implies a short-range dependence of X_tX_{t+l}. Condition 2 is closely related to the Uniform Geometric Moment Contraction condition proposed by Wang and Shao (2020). We say that $\{X_t\}$ has Uniform Geometric Moment Contraction (UGUC(k)) property if there exists some positive number k such that

$$\sup_{1\le l\le L}\sup_{1\le i\le p,1\le j\le p}\mathbb{E}\left(|X_{t,i}X_{t+l,j}|^{k}\right)<C<\infty$$

and

$$\sup_{1\le l\le L}\sup_{1\le i\le p,1\le j\le p}\mathbb{E}\left(\left|X_{t,i}X_{t+lj}-X'_{t,i}X'_{t+l,j}\right|^{k}\right)\le C\rho^{t},\quad t\ge 1$$

where $\rho\in(0,1)$ and $X'_{t,j}=f_j\left(\varepsilon_t,\ldots,\varepsilon_1,\varepsilon'_0,\varepsilon'_{-1},\ldots\right)$. If there exists $q\in(2,3]$ such that $\{X_t\}$ is UGUC(q), then it's easy to verify that $u_{m,q}\le C\rho^m$, hence $\sum_{m=1}^{+\infty}\min(m,N)^{\frac{1}{2}-\frac{1}{q}}u_{m,q}<+\infty$, so the condition 2 holds automatically.

The following theorems demonstrate that under the null hypothesis H_0, our hypothesis test can control the probability of type I errors:

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Theorem 1. *Assume Conditions 1-3 hold and* $w_l > 0$ *for* $l \in \mathbb{N}^+$. *Then under* H_0, *and* $\frac{Lp^2}{N^\delta \sigma_0} \to 0$, *there is*

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(T \le t) - \mathbb{P}(T^* \le t)| \to 0 \tag{3.1}$$

Theorem 2. *Assume Conditions 1-3 hold and* $w_l > 0$ *for* $l \in \mathbb{N}^+$. *Then under* H_0, *and* $\frac{Lp^2}{N^\delta \sigma_0} \to 0$, *there is*

$$\mathbb{P}\left(\hat{q}_{\frac{\alpha}{2}} < T < \hat{q}_{1-\frac{\alpha}{2}}\right) \to 1-\alpha \tag{3.2}$$

Remark4. If all the eigenvalues of Σ_0 are bounded, we have $\sigma_0 = O(p)$, then the Theorem 1 and Theorem 2 hold when $Lp = O(N^{\delta-\epsilon})$ for any $\epsilon > 0$. This implies that Theorem 1 and Theorem 2 remain valid as p tends to infinity. Also, since $Lp = O(N^{\delta-\epsilon})$, the maximum lag L is allowed to increase as N increases. For instance, we can set $L = O(lnN)$ and $p = O(\frac{N^{\delta-\epsilon}}{lnN})$, and Theorem 1 and Theorem 2 still hold. This allows us to further explore how the choice of weight sequence $\{w_l\}$ affects the power of our proposed test.

We then look into the power of the tests when an alternative hypothesis H_1 is specified. Let $\{z_t\}$ with $z_t = (z_{t1}, \cdots, z_{tp})^\top$ be p-dimensional i.i.d. random vectors with $\mathbb{E}z_{ti} = 0, \mathbb{E}z_{ti}^2 = 1$ and $\mathbb{E}z_{ti}^8 < \infty$. We assume that under H_1, the observations $\{X_t : t = 1, \ldots, N\}$ are a p-dimensional first-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

order vector moving average process of the form

$$H_1 : X_t = A_0 z_t + A_1 z_{t-1} \tag{3.3}$$

where A_0 and A_1 are $p \times p$ coefficient matrices.

We investigate the asymptotic behavior of our statistics with an arbitrarily given L and weight sequence $\{w_l > 0 : l \in \mathbb{N}^+\}$. We have the following theorem:

Theorem 3. *Under H_1 in (3.3) with an arbitrarily given L and weight sequence $\{w_l > 0 : l \in \mathbb{N}^+\}$ and $p = o(N^{\frac{3}{4}})$, we have*

$$\left(\frac{1}{N}T - w_1\mu_S\right) / w_1^2\sigma_{S1} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\mu_S = \operatorname{tr}\left(\tilde{\Sigma}_0\tilde{\Sigma}_1\right) + \frac{2}{N}\operatorname{tr}^2\left(\tilde{\Sigma}_{01}\right)$, *and*

$$\begin{aligned}
\sigma_{S1}^2 =& \frac{2}{N^2}\operatorname{tr}^2\left(\tilde{\Sigma}_0^2 + \tilde{\Sigma}_1^2\right) + \frac{6}{N^2}\operatorname{tr}^2\left(\tilde{\Sigma}_0\tilde{\Sigma}_1\right) \\
&+ \frac{4}{N}\left[2\operatorname{tr}\left(\tilde{\Sigma}_0\tilde{\Sigma}_1\right)^2 + (\nu_4 - 3)\operatorname{tr}\left\{D^2\left(\tilde{\Sigma}_0\tilde{\Sigma}_1\right)\right\}\right] \\
&+ \frac{8}{N^2}\operatorname{tr}\left(\tilde{\Sigma}_{01}\tilde{\Sigma}_{01}^\top\right)\operatorname{tr}\left(\tilde{\Sigma}_0^2 + \tilde{\Sigma}_1^2\right) + \frac{16}{N^2}\operatorname{tr}\left(\tilde{\Sigma}_{01}\tilde{\Sigma}_1\right)\operatorname{tr}\left(\tilde{\Sigma}_{01}\tilde{\Sigma}_0\right) \\
&+ \frac{16}{N^2}\operatorname{tr}\left(\tilde{\Sigma}_0 + \tilde{\Sigma}_1\right)\left\{\operatorname{tr}\left(\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_{01}\tilde{\Sigma}_0\right) + \operatorname{tr}\left(\tilde{\Sigma}_{01}\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_1\right)\right\} \\
&+ \frac{16}{N^2}\operatorname{tr}\left(\tilde{\Sigma}_{01}\right)\left\{\operatorname{tr}\left(\tilde{\Sigma}_0^2\tilde{\Sigma}_{01}^\top\right) + \operatorname{tr}\left(\tilde{\Sigma}_1^2\tilde{\Sigma}_{01}\right) + 2\operatorname{tr}\left(\tilde{\Sigma}_1\tilde{\Sigma}_{01}\tilde{\Sigma}_0\right)\right\} \\
&+ \frac{4}{N}\operatorname{tr}\left(\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_{01}\tilde{\Sigma}_0^2 + \tilde{\Sigma}_{01}\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_1^2 + 2\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_1\tilde{\Sigma}_{01}\tilde{\Sigma}_0\right) \\
&+ \frac{4}{N}\operatorname{tr}\left(\tilde{\Sigma}_{01}\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_{01}\right) + \frac{12}{N^2}\operatorname{tr}^2\left(\tilde{\Sigma}_{01}\tilde{\Sigma}_{01}^\top\right) + \frac{16}{N^2}\operatorname{tr}\left(\tilde{\Sigma}_{01}\right)\operatorname{tr}\left(\tilde{\Sigma}_{01}\tilde{\Sigma}_{01}^\top\tilde{\Sigma}_{01}^\top\right) \\
&+ \frac{4}{N^2}\operatorname{tr}^2\left(\tilde{\Sigma}_0\tilde{\Sigma}_{01}\right) + \frac{4}{N^2}\operatorname{tr}^2\left(\tilde{\Sigma}_1\tilde{\Sigma}_{01}\right) + R,
\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

where $R = o(\sigma_{S1}^2)$, $D\left(\tilde{\Sigma}_0\tilde{\Sigma}_1\right)$ *denotes the diagonal matrix consisting of the main diagonal elements of* $\tilde{\Sigma}_0\tilde{\Sigma}_1$ *and* $\tilde{\Sigma}_0 = A_0^\top A_0, \tilde{\Sigma}_1 = A_1^\top A_1, \tilde{\Sigma}_{01} = A_0^\top A_1$, $\nu_4 = \mathbb{E}z_{ti}^4$.

Theorem 3 demonstrates that under local alternative H_1 in 3.3, for our statistics T, $\frac{1}{N}T$ are asymptotically normal distributed with non-zero mean. Therefore, $T \to \infty$ when $N \to \infty$, while the T^* does not go to ∞. Hence our test statistics T have power under local alternative H_1.

We further investigate the power of the proposed test under a more general class of alternatives. Notice that under the alternative hypothesis H_1, there exist some values of l such that $\Sigma_l \neq 0$. Let $Y_{t,l} = \text{vec}\left(X_t X_{t+l}^\top\right)$ and $\mathcal{Y}_t = (\sqrt{w_1}Y_{t,1}^\top, \ldots, \sqrt{w_L}Y_{t,L}^\top)^\top$. Our proposed statistic T can be expressed as $T = \frac{1}{N}\sum_{i\neq j}\mathcal{Y}_i^\top\mathcal{Y}_j$. Let $S_k = \text{Cov}(\mathcal{Y}_{i+k}, \mathcal{Y}_i)$ denote the auto-covariance matrix of $\mathcal{Y}_t$ and $S = \sum_{k=-\infty}^{\infty} S_k$ denote the long-run covariance matrix of $\mathcal{Y}_t$. Under the alternative hypothesis H_1, assuming that there exists some value of L such that $\mathbb{E}\mathcal{Y}_t = \mu \neq 0$, we have the following theorem:

Theorem 4. *Assume Conditions 1-3 hold and* $\frac{\sum_{h=0}^{\infty}\|S_h\|}{\|S\|_F} = o(1)$, *where* $\|\cdot\|$ *denotes the Frobenius norm. Let* $\hat{q}_\alpha$ *denote the critical value obtained in section 2.2. Under the alternative hypothesis* H_1 *with* $\mathbb{E}\mathcal{Y}_t = \mu \neq 0$, *we have the following results:*

1. If $\frac{N\|\mu\|^2}{\|S\|_F} \to c \in (0, \infty)$, *then* $\mathbb{P}(T < \hat{q}_{\frac{\alpha}{2}}$ *or* $T > \hat{q}_{1-\frac{\alpha}{2}}) \to \beta \in (\alpha, 1)$. *That*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

is, our test exhibits nontrivial asymptotic power.

2. If $\frac{N\|\mu\|^2}{\|S\|_F} \to \infty$, *then* $\mathbb{P}(T < \hat{q}_{\frac{\alpha}{2}} \text{ or } T > \hat{q}_{1-\frac{\alpha}{2}}) \to 1$. *Hence, the asymptotic power of our test is 1.*

Theorem 4 indicates that the asymptotic power of our proposed test depends on $\frac{N\|\mu\|^2}{\|S\|_F}$. Our proposed test can distinguish between the null hypothesis H_0 and alternative hypothesis H_1 as long as $\frac{N\|\mu\|^2}{\|S\|_F} \to c > 0$. It is important to note that Theorem 4 only requires that $\mathbb{E}\mathcal{Y}_t = \mu \neq 0$ under the alternative hypothesis H_1. Hence, we can use Theorem 4 to analyze the power of our proposed test under a more general class of alternatives.

3.2 Some Weighting Schemes

We now provide some schemes for selecting weight sequence $\{w_l\}$. Similar to Gallagher and Fisher (2015), we consider two scenarios: one where the maximum lag L is fixed, and another where L increases as N increases.

We first analyze the relationship between the asymptotic behavior of our proposed tests under the null hypothesis and the maximum lag L. The following condition is considered:

Condition 4. *Under* H_0, *let* $X_t = A^{1/2}Z_t$, *where* $Z_t = (z_{t1}, \cdots, z_{tp})^\top$ *is a sequence of p-dimensional i.i.d. random vectors, and each component* z_{ti} *satisfy* $\mathbb{E}z_{ti} = 0, \mathbb{E}z_{ti}^2 = 1$ *and* $\mathbb{E}z_{ti}^8 < +\infty$. *All the eigenvalues of A are*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

bounded.

For our statistics $T = \frac{1}{N}\sum_{l=1}^{L} w_l \sum_{i\neq j} Y_{i,l}^\top Y_{j,l}$, where $Y_{t,l} = \text{vec}\left(X_t X_{t+l}^\top\right)$, we have following theorem:

Theorem 5. *Suppose condition 4 holds, then we have*

$$\frac{T}{2\|\Gamma\|_F} \xrightarrow{d} \mathcal{N}(0, \sum_{l=1}^{L} w_l^2), \tag{3.4}$$

where $\Gamma = \text{Cov}\left(\text{vec}\left(X_t X_{t+1}^\top\right)\right)$.

Theorem 5 demonstrates that under certain conditions, the limiting distribution of T is a normal distribution $\mathcal{N}(0, (\sum_{l=1}^{L} w_l^2)4\|\Gamma\|_F^2)$; therefore, the limiting distribution of bootstrap statistics T^* is also a normal distribution with mean zero and variance equal to $(\sum_{l=1}^{L} w_l^2)4\|\Gamma\|_F^2$. For the asymptotic result of T under alternative hypothesis, Theorem 3 indicates that under alternative hypothesis, if $\{X_t\}$ is a VMA(1) process, then the asymptotic results of T depend only on the first weight w_1, and $T \to \infty$ under H_1. Hence, if the maximum lag L increases as the sample size increases, $\sum_{l=1}^{L} w_l^2$ would affect the power of our test.

Since the maximum lag L is allowed to increase as the sample size increases, similar to Gallagher and Fisher (2015), the weighting schemes can be classified into two categories. The first category involves choosing weights such that $\lim_{L\to\infty} \sum_{l=1}^{L} w_l^2 < +\infty$; the second category involves

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

choosing weights such that $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 = \infty$. The commonly employed weights in portmanteau tests, such as $w_l = \frac{N+2}{N-l}$ in Ljung and Box (1978) satisfy $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 = \infty$. If weight sequence meets the condition that $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 = \infty$, and assuming L is large compared to sample size N, then $\sum_{l=1}^{L} w_l^2$ would be relatively large. This results in a loss of power for our tests. If weight sequence satisfies $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 < \infty$, $\sum_{l=1}^{L} w_l^2$ would be bounded when L is large, making our test less sensitive to the choice of L. Hence, if L is large compared to N, we recommend choosing a weight sequence such that $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 < \infty$.

Remark5. We can also use Theorem 4 to demonstrate how the condition that the weight sequences $\{w_l\}$ are squared summable affects the asymptotic results of our proposed tests. This demonstration is carried out under a more general class of alternatives than VMA(q). Consider a VAR(1) model $X_t = \rho X_{t-1} + e_t$, where $|\rho| < 1$ and $e_t \overset{iid}{\sim} N(0, I_p)$. With some simple calculations, we can obtain $\frac{N\|\mu\|^2}{\|S\|_F} \geq \frac{N\left(\sum_{l=1}^{L} w_l\rho^{2l}\right)}{p\|A\|_F\left(\sum_{l=1}^{L} w_l^2\right)^{\frac{1}{2}}}$, where $A = \sum_{l=-\infty}^{+\infty}\mathrm{Cov}(\mathcal{Z}_t, \mathcal{Z}_{t+l})$, $\mathcal{Z}_t = (Y_{t,1}^\top, \ldots, Y_{t,L}^\top)^\top$, and $Y_{t,l}$ is defined in Theorem 4. Notice that A does not depend on $\{w_l\}$. Assume that L increases as the sample size N increases. If N, p and A satisfy $\frac{N}{p\|A\|_F} \to \infty$, then the condition $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 \leq \infty$ ensures that $\frac{N\left(\sum_{l=1}^{L} w_l\rho^{2l}\right)}{p^{\frac{1}{2}}\|A\|_F\left(\sum_{l=1}^{L} w_l^2\right)^{\frac{1}{2}}} \to \infty$. Based on Theorem 4, the asymptotic power of our test is 1. Compared with the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

condition $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 \leq \infty$, the condition $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 = \infty$ may lead to a loss of power of our test, since it can cause $\frac{N\left(\sum_{l=1}^{L} w_l\rho^{2l}\right)}{p\|A\|_F\left(\sum_{l=1}^{L} w_l^2\right)^{\frac{1}{2}}} < \infty$. If N, p and A satisfy $\frac{N}{p\|A\|_F} \to C$, then the condition $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 \leq \infty$ ensures that $\frac{N\left(\sum_{l=1}^{L} w_l\rho^{2l}\right)}{p\|A\|_F\left(\sum_{l=1}^{L} w_l^2\right)^{\frac{1}{2}}} \to C$, and our test has nontrivial asymptotic power. The condition $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 = \infty$ may result in loss of power of our test, since it may cause $\frac{N\left(\sum_{l=1}^{L} w_l\rho^{2l}\right)}{p\|A\|_F\left(\sum_{l=1}^{L} w_l^2\right)^{\frac{1}{2}}} \to 0$. Hence if we allow L to increase as the sample size N increases, the condition $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 = \infty$ may result in loss of power of our tests. Therefore, we recommend choosing a weight sequence such that $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 < \infty$ when L is large compared to N.

If the maximum lag L is fixed and relatively small compared to the sample size, then $\sum_{l=1}^{L} w_l^2$ would be small, which implies that our tests have relatively large power. Consequently, both weighting schemes can be employed in this scenario. For commonly used models such as the VAR model and the dynamic factor model, the autocovariance decays exponentially as the lag increases. Hence, under the alternative hypothesis, the autocovariance at larger lags would be relatively small; therefore, the weight w_l should be relatively large when l is small.

Inspired by the above discussion, we consider two weighting schemes. The first scheme sets $w_l = \frac{N+2}{N-l}\kappa(\frac{l}{L})^2$, where $\kappa(z)$ is a kernel function. Hong

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

(1996) used a similar weighting scheme for testing white noise in univariate time series. This scheme assigns relatively large weight to lags with small order l and satisfies $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 = \infty$. We expect that this weighting scheme will have relatively good power when the maximum lag L is fixed and relatively small. The second scheme sets $w_l = a^l$ with $a \in (0,1)$; this scheme also assigns relatively large weight to lags with small order l. Furthermore, this weight sequence satisfies $\lim_{L\to\infty}\sum_{l=1}^{L} w_l^2 < \infty$. Hence, we expect that this weighting scheme will have relatively good power when max lag L is relatively large compared to the sample size.

4. Simulation studies

In the simulation study, we examine the finite sample performance of our proposed method in comparison with several existing testing methods. To illustrate how different choices of weight sequence $\{w_l\}$ affect the power of our hypothesis test, we consider three choices of $\{w_l\}$. Let T_1 be statistic in (2.2) with $w_l \equiv 1$, i.e,

$$T_1 = \frac{1}{N}\sum_{l=1}^{L}\sum_{i\neq j} Y_{i,l}^{\top} Y_{j,l}.$$

Let T_2 be statistic in (2.2) with $w_l = \frac{N+2}{N-l}\kappa(l/L)^2$, i.e,

$$T_2 = \frac{1}{N}\sum_{l=1}^{L}\frac{N+2}{N-l}\kappa(\frac{l}{L})^2\sum_{i\neq j} Y_{i,l}^{\top} Y_{j,l},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

where $\kappa(z) = \begin{cases} \frac{\sin(\sqrt{3}\pi z)}{\sqrt{3}\pi z} & : |z| < 1 \\ 0 & : |z| \geq 1 \end{cases}$. The weight sequence in T_2 has been considered in Hong (1996), and it exhibits certain optimality properties within the method proposed by Hong (1996). Let T_3 be statistic in (2.2) with $w_l = 0.9^l$, i.e,

$$T_3 = \frac{1}{N} \sum_{l=1}^{L} 0.9^l \sum_{i \neq j} Y_{i,l}^\top Y_{j,l}.$$

The weight sequences $\{w_l\}$ in T_1 and T_2 satisfy $\lim_{L\to\infty} \sum_{l=1}^{L} w_l^2 = \infty$, while weight sequence $\{w_l\}$ in T_3 satisfies $\lim_{L\to\infty} \sum_{l=1}^{L} w_l^2 < \infty$. The hypothesis test based on T_1, T_2 and T_3 is conducted by the method we describe in Section 2. We set the bootstrap number to be $B = 1000$.

For comparison, we consider the following five tests:

(1). T_{SN} denotes the white noise test statistic proposed by Wang and Shao (2020).

$$T_{SN} = \frac{T_s(1)^2}{W_n^2},$$

where $T_s(r) = \sum_{t=1}^{[nr]} \sum_{s=1}^{t} \mathcal{Y}_{t+d}^\top \mathcal{Y}_s$, $r \in [0,1]$, $n = N - L - d$, $W_n^2 = \frac{1}{n} \sum_{k=1}^{n} \left(T_s(k/n) - \frac{k(k+1)}{n(n+1)} T_s(1) \right)^2$. Wang and Shao (2020) showed that under H_0,

$$T_{SN} \xrightarrow{d} \frac{\mathcal{B}(1)^2}{\int_0^1 \left(\mathcal{B}(u^2) - u^2 \mathcal{B}(1) \right)^2 du},$$

where $\mathcal{B}(r), r \in [0,1]$ is the standard Brownian motion. Following simulation setup in Wang and Shao (2020), we set $d = 10$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

(2). T_C denotes the white noise test statistic proposed by Chang et al. (2017).

$$T_C = \sqrt{N} \max_{1\le l\le L} \max_{1\le i,j\le p} |\hat{\rho}_{ij}(l)|,$$

where $\hat{\rho}_{i,j}(l)$ is the (i,j)th element of sample autocorrelation matrix at lag l, $\widehat{\Gamma}_l = \text{diag}\{\hat{\Sigma}_0\}^{-1/2}\hat{\Sigma}_l \,\text{diag}\{\hat{\Sigma}_0\}^{-1/2}$, the critical value of this test is obtained by bootstrap method.

(3). T_{W1} and T_{W1} denote two white noise test statistics proposed by Wang et al. (2022).

$$T_{W1} = \max_{l=1,\ldots,L} w_l A_s\left(\widehat{\Gamma}_l\right),$$
$$T_{W2} = \sum_{l=1}^{L} w_l A_s\left(\widehat{\Gamma}_l\right),$$

where $\{w_l\}_{l=1}^L$ are pre-selected weight. $A_s\left(\widehat{\Gamma}_l\right)$ calculate average of the largest s absolute value of $\widehat{\Gamma}_l$, sample autocorrelation matrix at lag l. The critical value of this test is obtained by a bootstrap method. Following Wang et al. (2022)'s simulation setup, we set $s = p$, $w_l = \frac{1}{N-l}$.

(4). T_{Li} denotes white noise test statistic proposed by Li et al. (2019).

$$T_{Li} = \frac{\sum_{l=1}^{L} \text{tr}(\hat{\Sigma}_l^\top \hat{\Sigma}_l) - Lnc_{p,N}^2 \hat{s}_1^2}{\sqrt{2L} c_{p,N} \left(\hat{s}_2 - c_{p,N}\hat{s}_1^2\right)},$$

where $c_{p,N} = p/N$, $\hat{s}_1 = p^{-1}\,\text{tr}(\hat{\Sigma})$, $\hat{s}_2 = p^{-1}\,\text{tr}(\hat{\Sigma}_l^\top \hat{\Sigma}_l)$, Li et al. (2019) showed that under H_0 and i.i.d assumption,

$$T_{Li} \xrightarrow{d} N(0,1).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

It is worth noting that hypothesis testing proposed by Li et al. (2019) was a one-side test; we reject the null hypothesis when $T_{Li} > z_{1-\alpha}$, where $z_{1-\alpha}$ is $1-\alpha$ quantile of standard normal distribution.

To examine the finite sample behavior, we consider several scenarios with combinations of $p = 20, 50, 80, 120$, $N = 100, 200$, and $L = 5, 10$. The level of significance is always set at $\alpha = 5\%$. For each experiment, we have 500 Monte Carlo replicates.

4.1 Empirical size

To compare the empirical sizes, consider the following four models,

Model 1. $X_t = Ae_t$, where $e_t \overset{iid}{\sim} N(0, I_p)$, $A = S^{1/2}$, and $S = (s_{kl})_{1\le k,l\le p}$ with $s_{kl} = 0.995^{|k-l|}$.

Model 2. $X_t = e_t$, where for $i = 1, \dots p$, the ith component of e_t, denoted as $e_{t,i}$, has the following form: $e_{t,i} = h_{t,i}^{1/2}\varepsilon_{t,i}$, where $\varepsilon_{t,i} \overset{iid}{\sim} N(0,1)$, $h_{t,i} = 0.01 + \alpha_i e_{t-1,i}^2 + \beta_i h_{t-1,i}$, $\beta_i = 0.98 - \alpha_i$, $\alpha_i = 0.05 + 0.9u_i$ and $u_i \overset{\text{iid}}{\sim} \text{unif}(0,1)$.

Model 3. $X_t = e_t \odot e_{t-1} \odot e_{t-2}$, where $e_t \overset{iid}{\sim} N(0, I_p)$, and $\odot$ denotes the Hadamard product.

Model 4. $X_t = \delta_t \times e_t + 3\,(1-\delta_t) \times e_t'$, where e_t and e_t' are independent normal

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

random vectors from $N(0, S)$ with $S = (s_{ij})_{1\leqslant i,j\leqslant p}$ and $s_{ij} = 0.5^{|i-j|}$. δ_t is Bernoulli random value with $\mathbb{P}(\delta_t = 1) = \mathbb{P}(\delta_t = 0) = 0.5$. $\{\delta_t\}$ are independent with $\{e_t\}$ and $\{e'_t\}$

Model 1 is i.i.d. white noise, which is the setting considered in Chang et al. (2017) and Wang and Shao (2020). Model 2 is a multivariate generalized autoregressive conditional heteroskedasticity sequence; Wang et al. (2022) considered the same setting. Model 3 is a non-i.i.d white noise sequence. Model 4 is an i.i.d white noise sequence.

Set significance level at $\alpha = 5\%$, and empirical sizes for different models are reported in Table 1 and Table 2. Our tests, denoted as T_1, T_2, and T_3, and the test proposed by Wang and Shao (2020), denoted as T_{SN}, exhibit accurate and stable empirical sizes for Model 1 to 4. The test proposed by Chang et al. (2017), denoted as T_C, exhibit accurate empirical size for Model 1 but not for other models. For Model 2 and Model 4, T_C tends to underreject the null frequently. For Model 3, T_C over-rejects the null when N is small and p is large (for instance, $N = 100, L = 5$ and $p = 120$); however, the size appears quite accurate when $N = 200$ and $p = 50$. This indicates that T_C can not control type I error for this model. Two tests from Wang et al. (2022), denoted as T_{W1} and T_{W2}, exhibit accurate and stable empirical sizes for Model 1, 2, and 4. For Model 3, T_{W2} over-rejects

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Table 1: Empirical sizes (in %) of different test statistics at $\alpha = 5\%$ significant level for Model 1 and Model 2.

Model 1																
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}
				N=100, L=5								N=100, L=10				
20	5.2	4.2	4.4	4.4	3.4	6.0	5.6	3.2	5.4	5.8	5.0	4.2	3.4	6.0	5.8	4.4
50	4.6	5.0	4.8	6.0	5.0	5.2	5.6	3.8	2.8	4.0	4.2	4.8	2.4	4.4	3.8	3.6
80	6.8	5.8	5.8	4.6	3.4	7.0	7.4	3.2	4.6	5.4	5.2	3.2	2.6	4.4	5.2	4.0
120	3.6	4.0	4.6	4.2	3.2	4.4	4.2	4.4	6.2	5.4	6.0	6.6	2.2	5.2	6.2	4.4
				N=200, L=5								N=200, L=10				
20	3.8	4.4	5.0	4.0	2.9	6.2	5.6	4.8	5.4	5.2	5.4	4.0	4.2	6.8	3.8	3.6
50	4.6	5.0	4.0	6.0	5.0	5.2	5.6	3.8	3.0	4.0	3.8	4.8	2.2	4.4	3.8	3.6
80	5.0	6.2	5.0	4.4	4.4	6.0	6.8	4.0	3.0	4.6	4.8	5.8	3.8	4.4	4.6	3.8
120	4.6	5.8	6.0	4.6	5.4	6.2	6.0	4.4	5.2	5.2	4.8	4.6	4.8	6.2	5.8	4.6
Model 2																
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}
				N=100, L=5								N=100, L=10				
20	4.4	2.8	4.6	5.2	0	6.0	5.4	4.0	5.8	4.4	4.6	4.6	0.2	6.4	6.0	0.6
50	4.6	3.8	4.0	4.0	0.2	4.2	5.6	0.6	5.2	4.8	4.8	4.6	0	5.4	5.2	0
80	6.2	6.0	5.6	6.2	0.6	5.6	5.8	0.2	5.4	4.0	4.2	6.2	0.4	6.6	6.2	0
120	5.6	4.6	4.8	6.2	1.6	5.8	4.6	0	5.6	4.4	4.2	5.4	0.8	6.4	3.8	0
				N=200, L=5								N=200, L=10				
20	5.4	3.8	4.8	6.6	0.2	5.0	4.4	10.6†	4.6	5.4	5.0	5.2	0	4.4	4.8	3.6
50	4.6	5.4	4.6	4.8	0.2	5.4	4.8	5.0	4.2	4.6	4.4	6.2	0	6.4	5.8	0
80	5.8	5.6	5.2	4.4	0.4	3.0	7.0	0	5.2	5.2	5.6	6.0	0	6.2	5.2	0
120	5.2	5.0	4.6	4.4	0.2	5.0	4.8	0	6.0	5.2	5.0	5.4	0.2	4.8	4.6	0

†: Type-I errors are out of control

the null when $p = 80, 120$, T_{W1} also over-rejects the null in some cases (for instance, $N = 200$ and $p = 80$). The test from Li et al. (2019), indicated as T_{Li}, exhibits accurate empirical size for Model 1 but not for other models. For Model 2, when $N = 200, L = 5$, and $p = 20$, T_{Li} over-rejects the null; when $N = 200, L = 5$, and $p = 50$, the size is quite accurate; however, when $N = 200, L = 5$, and $p = 80$, T_{Li} under-rejects the null. The same situations also occurs in Model 3 and Model 4, suggesting that the empirical

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Table 2: Empirical sizes (in %) of different test statistics at $\alpha = 5\%$ significant level for Model 3 and Model 4.

Model 3															
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}
	N=100, L=5								N=100, L=10							
20	5.4	4.0	4.4	3.2	6.4	7.2	5.6	33.6†	5.6	5.6	5.0	3.8	11.2†	7.6	10.8†	14.5†
50	5.4	5.0	4.8	4.4	25.4†	9.4†	14.0†	25.6†	4.4	3.8	4.4	2.4	44.2†	5.6	13.4†	4.2
80	5.6	3.8	4.2	2.6	44.6†	8.8	14.4†	19.3†	6.0	5.0	5.2	2.4	72.0†	8.4	18.2†	2.3
120	6.0	5.0	5.8	2.4	72.0†	8.4	18.2†	12.2†	6.0	5.2	5.2	3.6	90.6†	9.2†	33.0†	1.0
	N=200, L=5								N=200, L=10							
20	5.8	4.4	4.6	4.8	0.4	5.6	5.6	45.5†	5.0	4.2	4.6	3.4	1.6	7.8	7.2	26.8†
50	4.4	4.4	3.8	2.4	5.0	7.2	6.0	40.4†	4.6	4.8	5.0	3.2	8.8	5.4	6.6	15.4†
80	6.2	4.8	4.4	5.6	10.6†	5.8	11.8†	36.8†	6.2	4.8	5.4	4.8	17.0†	9.4†	11.7†	7.4
120	3.8	6.4	6.0	3.4	20.0†	7.2	11.2†	30.2†	4.4	5.2	5.8	3.8	36.8†	7.6	14.8†	1.9
Model 4																
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_{Li}
	N=100, L=5								N=100, L=10							
20	3.2	6.4	4.8	6.2	0	3.8	4.8	2.6	5.8	3.6	5.8	5.8	0	3.0	5.2	1.9
50	6.0	5.8	5.2	6.2	0	6.6	6.0	5.4	4.8	4.0	4.4	5.0	0	6.0	3.8	1.6
80	5.8	4.6	5.0	5.2	0	5.6	5.4	5.6	5.8	4.0	4.2	4.0	0.4	6.4	4.6	2.6
120	5.8	3.0	3.8	3.8	0.6	6.6	6.0	10.6†	5.2	3.2	4.2	6.0	0.8	4.4	3.8	2.3
	N=200, L=5								N=200, L=10							
20	3.6	3.8	4.0	4.0	0	5.0	3.4	4.6	3.6	5.8	4.8	6.4	0	6.8	4.6	2.4
50	3.6	6.6	6.2	5.0	0	5.8	3.4	6.4	6.2	4.0	6.0	4.6	0	4.4	6.0	2.2
80	3.6	5.2	4.6	4.2	0	5.2	3.4	10.0†	4.8	4.8	5.2	5.8	0.2	5.8	4.8	2.4
120	5.0	6.6	5.8	6.0	0	4.6	5.0	11.3†	3.4	5.2	4.2	4.0	0	6.2	6.4	3.4

†: Type-I errors are out of control

size of T_{Li} is not stable for Model 2, 3, and 4. Our proposed tests have fairly accurate empirical sizes for Model 1, 2, 3, and 4.

4.2 Empirical power

To study the empirical power, we consider the following four models:

Model 5. $X_t = 0.15X_{t-1} + e_t$, where $\{e_t\}$ has same data generation process as $\{e_t\}$ in Model 2, i.e. $\{e_t\}$ is a multivariate generalized autoregressive

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

conditional heteroskedasticity sequence.

Model 6. $X_t = AX_{t-1} + e_t$, where e_t is i.i.d. $N(0, I_p)$, and

$$A = \begin{bmatrix} A_0 & 0 \\ 0 & 0 \end{bmatrix}_{p\times p},$$

A_0 is a $k_0 \times k_0$ matrix with $A_0(i,j) \sim U(-0.25, 0.25)$, and $k_0 = \min\{[p/5], 12\}$, $[\cdot]$ stands for floor function. The first k_0 elements of X_t are not white noises.

Model 7. $X_t = AX_{t-1} + e_t$, where e_t is i.i.d. $N(0, I_p)$, and $A = (a_{ij})$, $a_{ij} = 0.9^{|i-j|}$, then we normalize A so that $\|A\|_2 = 0.7$.

Model 8. $X_t = BY_t + E_t$, where E_t is i.i.d. $N(0, I_p)$, $B \in \mathbb{R}^{p\times 4}$ is a $p \times 4$ matrix, $B = (b_{ij})$, b_{ij} is first generate independently from uniform distribution $U(-1, 1)$, then be divided by $p^{0.25}$, $Y_t \in \mathbb{R}^4$ with $Y_t = AY_{t-1} + e_t$, $A \in \mathbb{R}^{4\times 4}$ is a 4 dimensional diagonal matrix with diagonal element set to be $(-0.3, 0.35, 0.25, -0.4)$. $e_t \overset{iid}{\sim} N(0, I_4)$ and are independent with $\{E_t\}$.

Model 5 is an example of a sparse high-dimensional VAR model; a similar setting has been considered in Wang et al. (2022). Model 6 is also an example of a sparse high-dimensional VAR model, in which only the first k_0 elements of X_t are not white noises; the same setting has been considered

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

in Chang et al. (2017) and Wang and Shao (2020). Model 7 and Model 8 are added to examine the behavior of our test in the case of dense alternatives. Model 7 is similar to the setting in Wang and Shao (2020). Model 8 is a dynamic factor model.

Since the empirical size of T_{Li} is not stable for Model 2, 3, and 4, we do not consider the empirical power of T_{Li} and only compare empirical power of T_1, T_2, T_{SN}, T_C, T_{W1} and T_{W2}. Since the empirical size of T_C, T_{W1}, and T_{W2} is largely distorted in Model 3, we report the size-adjusted power of these tests.

Table 3 and Table 4 present results on empirical powers for different testing methods at the 5% significance level. Our proposed tests exhibit nontrivial power for all four models. Under sparse alternatives (Model 5 and Model 6), as N increases, the powers of T_1, T_2, and T_3 quickly rise to around 1. The tests proposed by Wang et al. (2022) are strong competitors to our test. T_{SN} and T_C exhibit low empirical powers for Model 5 and Model 6. In Model 5, T_1, T_2, and T_3 outperform the rest of the tests, and there is no definitive conclusion regarding how T_1, T_2, and T_3 compare to each other. The empirical power of T_C for Model 5 is low when $N = 100$, and decreases when p increases, indicating that T_C does not have satisfactory empirical powers. In Model 6, T_2 outperforms the rest of the tests, T_3 is

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Table 3: Empirical power (in %) of different test statistics at $\alpha = 5\%$ significant level for Model 5 and Model 6.

Model 5														
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}
	N=100, L=5							N=100, L=10						
20	52.0	**63.8**	60.0	7.8	0.4	43.2	16.8	56.8	**59.2**	58.0	6.4	0	43.6	21.8
50	91.2	89.4	**94.8**	8.8	0.2	88.8	48.8	**98.8**	93.6	98.4	5.6	0	89.4	67.2
80	**99.8**	99.0	99.6	9.0	0.2	98.0	71.6	**100**	99.2	99.8	5.4	0.6	99.0	87.6
120	**100**	**100**	**100**	10.4	0.6	**100**	90.2	**100**	**100**	**100**	6.0	1.4	**100**	98.6
	N=200, L=5							N=200, L=10						
20	73.6	**91.4**	88.6	18.6	0.2	51.0	29.8	75.0	**86.6**	85.8	9.4	0	51.6	30.6
50	98.6	**99.0**	98.6	21.2	0	97.0	83.0	**100**	99.6	**100**	13.6	0	96.2	87.6
80	**100**	**100**	**100**	16.8	0	99.6	95.6	**100**	**100**	**100**	11.6	0	99.6	99.4
120	**100**	**100**	**100**	19.4	0	**100**	99.6	**100**	**100**	**100**	12.8	0.2	**100**	**100**
Model 6														
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}
	N=100, L=5							N=100, L=10						
20	10.6	**26.8**	19.6	7.0	0	11.4	3.4	10.0	**24.0**	12.6	5.6	0	7.2	1.6
50	49.4	**91.0**	71.6	26.2	0	19.4	18.6	26.8	**77.0**	65.6	15.8	0	11.6	8.4
80	61.4	**95.0**	82.6	31.0	0	47.6	26.4	37.2	**89.0**	74.4	16.0	0	31.4	13.6
120	37.8	**79.6**	56.0	17.0	0	35.8	11.6	26.8	**65.8**	51.0	11.4	0	22.6	6.2
	N=200, L=5							N=200, L=10						
20	29.2	**70.0**	44.2	16.6	5.6	16.0	12.0	16.2	**49.8**	41.2	7.2	2.8	10.2	6.0
50	94.8	**100**	99.6	68.6	7.8	72.0	92.8	74.6	**99.8**	99.6	48.2	5.0	36.2	64.2
80	97.2	**100**	**100**	80.4	10.0	95.8	97.8	89.2	**100**	**100**	59.4	5.2	69.4	82.8
120	86.2	**100**	97.8	61.8	7.4	83.4	85.8	66.6	**99.2**	97.4	51.6	3.4	54.2	57.0

usually the second-best test, T_1, T_{W1}, and T_{W2} have satisfactory power, and T_{SN} has second worst performance. The empirical powers of T_C for Model 5 are low when $N = 100$, and increase dramatically when $N = 200$; this indicates that the powers of T_C are low when sample size N is small.

For dense alternatives (Model 7 and Model 8), T_2 outperforms the rest of the tests, and T_3 is usually the second-best test; this shows that our tests have high power against dense alternatives. T_{SN} exhibit relatively low power, partly because the effective sample size for T_{SN} is small. T_C

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Table 4: Empirical power (in %) of different test statistics at $\alpha = 5\%$ significant level for Model 7 and Model 8.

Model 7														
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}
	N=100, L=5							N=100, L=10						
20	82.2	**92.4**	88.6	31.4	0	63.2	44.8	70.2	**88.6**	84.2	20.4	0	52.4	34.4
50	64.0	**75.8**	75.2	21.6	0	37.6	16.8	51.0	**74.0**	69.0	13.4	0	28.0	13.0
80	59.4	**74.2**	69.8	14.0	0	38.0	12.0	49.4	**68.4**	67.8	9.2	0	24.0	9.0
120	61.4	**78.4**	71.8	14.2	0	37.6	9.8	50.2	**71.6**	70.8	9.8	0	30.2	7.4
	N=200, L=5							N=200, L=10						
20	98.4	**99.8**	99.6	67.8	0.4	36.2	19.0	95.6	**100**	**100**	51.6	0	22.0	12.6
50	93.6	**99.6**	97.0	51.2	0	70.0	45.0	86.6	**98.4**	96.6	39.0	0	58.0	36.4
80	88.8	**98.2**	96.6	43.8	0	63.2	30.8	85.6	**98.8**	96.2	33.4	0	51.8	25.6
120	90.8	**100**	96.8	48.6	0	63.8	26.4	88.4	**100**	96.4	37.8	0	55.0	19.0
Model 8														
p	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}	T_1	T_2	T_3	T_{SN}	T_C	T_{W1}	T_{W2}
	N=100, L=5							N=100, L=10						
20	17.0	**38.8**	25.6	7.4	0	10.4	4.0	13.6	**26.2**	21.2	6.0	0	6.4	1.4
50	26.6	**54.2**	31.6	13.6	0	12.2	4.8	12.8	**34.4**	29.2	6.8	0	7.8	2.6
80	47.2	**60.6**	50.6	23.6	0	8.8	4.8	21.8	**58.4**	46.0	12.6	0	9.0	2.4
120	37.8	**79.6**	40.8	17.0	0	11.8	2.8	26.8	**65.8**	37.2	11.4	0	8.8	1.6
	N=200, L=5							N=200, L=10						
20	44.2	**80.8**	59.0	21.8	0	4.0	1.0	25.4	**69.4**	57.2	16.8	0	4.2	0.6
50	55.6	**92.0**	77.8	29.4	0	23.2	16.4	32.4	**80.6**	70.8	18.2	0	14.2	7.8
80	61.0	**94.4**	82.0	38.6	0	25.4	17.6	42.6	**89.2**	76.8	23.8	0	19.4	9.8
120	67.8	**89.4**	76.8	35.8	0	23.8	16.8	46.4	**76.4**	72.4	24.0	0	11.6	6.2

can hardly detect serial correlations in dense alternatives. T_{W1} and T_{W2} perform relatively well compared to T_C. As demonstrated by Wang et al. (2022), this is mainly because T_{W1} and T_{W2} are based on the average of the largest s absolute values of autocorrelation matrix, making them better at picking up dense signals compared to T_C.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

4.3 Numerical analysis of weight w_l

We study how weight sequence $\{w_l\}$ affects the power of our proposed tests as the maximum lag L increases. We consider the following two models:

Model 9. $X_t = e_t exp(\sigma_t)$, $\sigma_t = 0.25\sigma_{t-1} + 0.05u_t$, $e_t \overset{iid}{\sim} N(0, S_e)$ and $u_t \overset{iid}{\sim} N(0, S_u)$, where $S_e = (s_{e,ij})_{p\times p}$ and $S_u = (s_{u,ij})_{p\times p}$ with $s_{e,ij} = I(i = j) + 0.4I(i \neq j)$ and $s_{u,ij} = 0.9^{|i-j|}$.

Model 10. $X_t = AX_{t-1} + e_t$, where e_t is i.i.d. $N(0, I_p)$, and $A = (a_{ij})$, $a_{ij} = 0.9^{|i-j|}$, then we normalize A so that $\|A\|_2 = 0.6$.

Model 9 is a stochastic volatility model; we study the empirical sizes of T_1, T_2, and T_3. Model 10 is a dense VAR model similar to Model 7; we study the empirical powers of T_1, T_2, and T_3. We set $N = 150$, $p = 20$ and $L = 5, 10, 15, 20, 25, 30, 35, 40$, the significance level is set at $\alpha = 5\%$.

The empirical sizes of T_1, T_2, and T_3 at $\alpha = 5\%$ significant level in Model 9 are presented in Figure 1; our proposed tests can control type I errors when the maximum lag L is large. The empirical powers of T_1, T_2 and T_3 at $\alpha = 5\%$ significant level in Model 10 are presented in Figure 2. The weight sequences $\{w_l\}$ in T_1 and T_2 satisfy $\lim_{L\to\infty} \sum_{l=1}^{L} w_l^2 = \infty$. The empirical powers of T_1 and T_2 are gradually decreasing as the maximum lag L increases, which is consistent with our analyses in Section 3. Figure

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

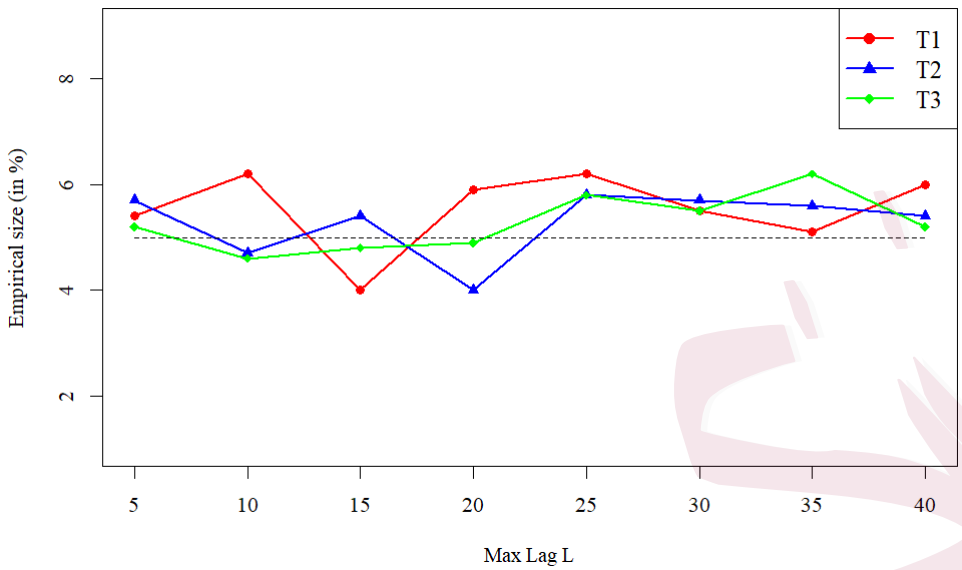

Figure 1: Empirical size of T_1, T_2 and T_3 for Model 9

2 shows that the empirical power of T_3 does not decrease significantly as the maximum lag L increases. Note that the weight sequence $\{w_l\}$ in T_3 satisfies $\lim_{L\to\infty} \sum_{l=1}^{L} w_l^2 < \infty$; as analyzed in Section 3, the power of T_3 is less sensitive to the choice of L.

5. Real Data Example

In this section, we analyze the U.S. stock market using our proposed tests. The dynamic factor model is commonly used for analyzing the stock market, and we are interested in testing whether the use of dynamic factor model is reasonable. The data contains daily returns of 120 securities of the S&P

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

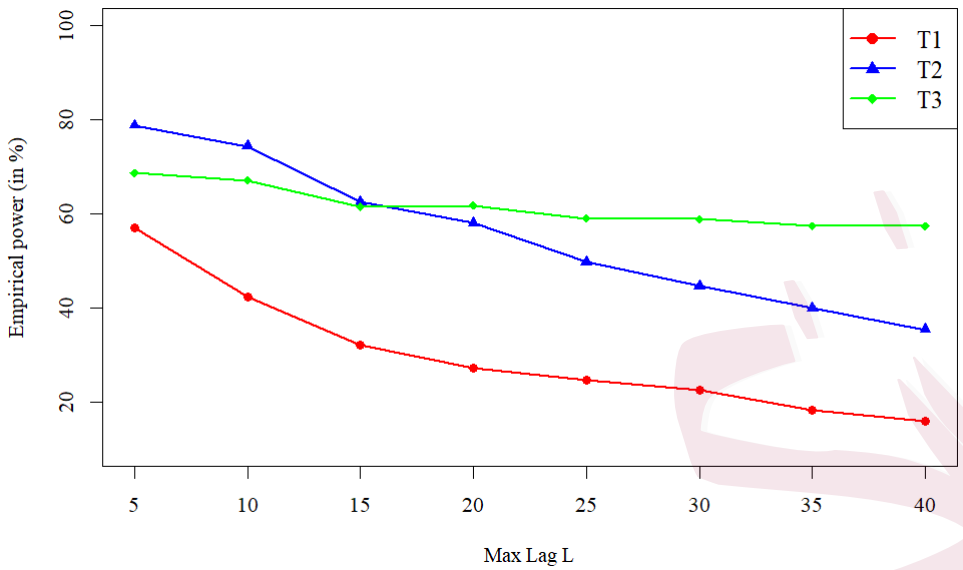

Figure 2: Empirical power of T_1, T_2 and T_3 for Model 10

500 index from January 2001 to April 2002. Denote our data as $\{X_t\}$; the dimension of X_t is 120, and sample size is 366.

We first use our proposed tests to test whether the data is white noise. Let T_1 be statistic in (2.2) when $w_l \equiv 1$, T_2 be statistic in (2.2) when $w_l = \frac{N+2}{N-l}\kappa(l/L)^2$, where $\kappa(z) = \begin{cases} \frac{\sin(\sqrt{3}\pi z)}{\sqrt{3}\pi z} & : |z| < 1 \\ 0 & : |z| \geq 1 \end{cases}$, and T_3 be statistic in (2.2) when $w_l = 0.9^l$. For comparison, we consider the white noise test proposed by Chang et al. (2017)(denoted as T_C) and two white noise tests proposed by Wang et al. (2022)(denoted as T_{W1} and T_{W2}), which have been described in Section 4. We set $L = 2, 4, 6, 8, 10$ and the bootstrap number

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Table 5: p-values(in %) of $T_1, T_2, T_3, T_C, T_{W1}, T_{W2}$ for testing $\{X_t\}$ is white noise

	L				
p.value	2	4	6	8	10
T_1	11.7	51.1	89.7	94.0	91.2
T_2	3.3	2.8	3.7	10.8	19.2
T_3	2.1	4.6	2.0	8.6	17.1
T_C	79.3	89.9	93.6	94.7	96.3
T_{W1}	5.0	19.0	25.3	23.6	27.0
T_{W2}	3.4	14.4	31.0	22.4	36.0

to be 1000. We calculate the p-values of these white noise tests, and the results are shown in Table 5.

If we consider a significance level of 5%, the p-values of T_2 and T_3 are lower than 5% when $L = 2, 4, 6$, and the p-values of T_{W1} and T_{W2} are lower than 5% when $L = 2$. Based on our simulation, we believe the data is not white noise; T_1 and T_C fail to detect series dependent.

We used the information criteria approach of Bai and Ng (2002) to determine the number of factors (denoted as r) to use in the model. We consider the following three information criteria:

$$IC_1(r) = \log\left(V_r(\hat{F}, \hat{\Lambda})\right) + r\left(\frac{N+p}{Np}\right)\log\left(\frac{Np}{N+p}\right),$$
$$IC_2(r) = \log\left(V_r(\hat{F}, \hat{\Lambda})\right) + r\left(\frac{N+p}{Np}\right)\log(\min\{N, p\}),$$
$$IC_3(r) = \log\left(V_r(\hat{F}, \hat{\Lambda})\right) + r\frac{\log(\min\{N, p\})}{\min\{N, p\}}.$$

where $V_r(\hat{F}, \hat{\Lambda}) = \sum_{i=1}^{p}\sum_{t=1}^{N}\mathbb{E}\left[\hat{\epsilon}_{i,t}^2\right]/Np$ and $\hat{\epsilon}_{i,t} = X_{t,i} - \hat{F}_t\hat{\Lambda}_i$. The result is shown in Figure 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

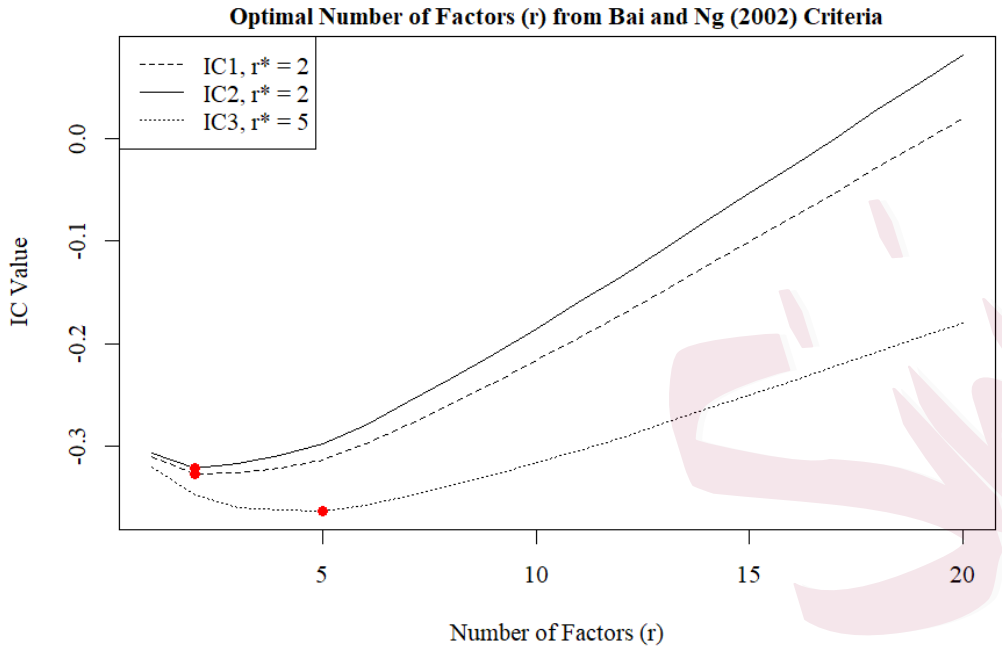

Figure 3: Information Criteria for number of factors

The result indicates that three information criteria yield two different optimal numbers of factors; the results of IC_1 and IC_2 suggest that r is 2, while the result of IC_3 suggests that r is 5. By employing our white noise testing method, we can determine the optimal number of factors by testing whether residuals are white noise for $r = 2$ and $r = 5$. The results for $r = 2$ are presented in Table 6, and the results for $r = 5$ are presented in Table 7.

Given the significance level to be 5%, we calculate the p-values for testing whether the residuals are white noise when the number of factors is 2 and 5, using our white noise tests (T_1, T_2, and T_3), the white noise test

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Table 6: p-values(in %) of $T_1, T_2, T_3, T_C, T_{W1}, T_{W2}$ for testing whether residuals is white noise when the number of factors $r = 2$

	L				
p.value	2	4	6	8	10
T_1	50.0	35.9	53.3	5.8	16.3
T_2	3.8	4.2	13.7	17.9	21.9
T_3	2.9	3.1	8.5	5.2	4.2
T_C	67.5	76.1	83.2	83.4	89.5
T_{W1}	58.2	36.0	32.6	41.0	16.2
T_{W2}	67.4	26.8	16.0	2.2	13.4

Table 7: p-values(in %) of $T_1, T_2, T_3, T_C, T_{W1}, T_{W2}$ for testing whether residuals is white noise when the number of factors $r = 5$

	L				
p.value	2	4	6	8	10
T_1	41.8	47.2	39.0	10.3	19.9
T_2	10.4	11.7	15.0	19.9	23.8
T_3	13.7	20.6	21.0	24.3	25.1
T_C	65.5	76.0	82.9	86.4	90.2
T_{W1}	61.4	12.4	9.8	49.2	42.6
T_{W2}	37.8	25.4	18.0	17.0	19.4

proposed by Chang et al. (2017) (T_C), and two white noise tests proposed by Wang et al. (2022)(T_{W1} and T_{W2}). Table 6 shows p-values of those tests when the number of factors is 2, the p-values of T_2 are less than 5% when $L = 2, 4, 6$, the p-values of T_3 are less than 5% when $L = 2, 4, 10$. This suggests the residuals are not white noise when the number of factors is 2; therefore, using 2 factors is insufficient for modeling our data. Table 7 shows that all p-values are larger than 5% when the number of factors

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

is 5; this indicates using 5 factors is sufficient for modeling our data. In summary, we choose the number of factors to be 5. In conclusion, it is reasonable to employ dynamic factor model with factor numbers set to be 5 for modeling the U.S. stock market.

6. Conclusions

In this paper, we propose a novel high-dimensional white noise testing method, which can be viewed as an extension of portmanteau tests in high-dimension settings. Simulation results indicate that our proposed tests are well-suited for detecting non-white noise, especially when signals are distributed across a large number of coordinates. We apply our method to determine the number of factors in the dynamic factor model for modeling the U.S. stock market, demonstrating the practical value of our proposed method.

There are several important future research directions worth noting. The first is to develop an automatic method for determining the maximum lag L in our tests. The second is to develop a test capable of detecting non-white noise signals in both sparse and dense cases since we typically lack prior knowledge about whether the alternative is sparse or dense. We shall leave these topics for future research.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

Supplementary Material

Additional details and supporting information can be found in the supplementary file online. In Section S1, we provide additional simulation results of the white noise test for fitted residuals. Section S2 contains proofs for the theoretical results outlined in the main text.

Acknowledgments

The authors would like to thank the editor, the associate editor, and two anonymous reviewers for their constructive comments and suggestions that led to significant improvements in the paper. We acknowledge the Key Laboratory of Complex Systems and Data Science of the Ministry of Education for partial support. This research is partially supported by the Youth Academic Innovation Team Construction project of Capital University of Economics and Business (Grant No.QNTD202303).

References

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Baltagi, B. H., C. Kao, and F. Wang (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics 220*(2), 349–365.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

REFERENCES

Box, G. E. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association 65*(332), 1509–1526.

Chang, J., Q. Yao, and W. Zhou (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika 104*(1), 111–127.

Chitturi, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association 69*(348), 928–934.

Daniel Peña, E. S. and V. J. Yohai (2019). Forecasting multiple time series with one-sided dynamic principal components. *Journal of the American Statistical Association 114*(528), 1683–1694.

Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica 83*(4), 1497–1541.

Fan, J., K. Wang, Y. Zhong, and Z. Zhu (2021). Robust High-Dimensional Factor Models with Applications to Statistical Machine Learning. *Statistical Science 36*(2), 303–327.

Gallagher, C. M. and T. J. Fisher (2015). On weighted portmanteau tests for time-series goodness-of-fit. *Journal of Time Series Analysis 36*(1), 67–83.

He, Y., G. Xu, C. Wu, and W. Pan (2021). Asymptotically independent u-statistics in high-dimensional testing. *The Annals of Statistics 49*(1), 154–181.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica 64*(4),

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

837–864.

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

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics 40*(2), 694–726.

Li, M. and Y. Zhang (2022). Bootstrapping multivariate portmanteau tests for vector autoregressive models with weak assumptions on errors. *Computational Statistics & Data Analysis 165*, 107321.

Li, W. and A. McLeod (1981). Distribution of the residual autocorrelations in multivariate arma time series models. *Journal of the Royal Statistical Society: Series B (Methodological) 43*(2), 231–239.

Li, Z., C. Lam, J. Yao, and Q. Yao (2019). On testing for high-dimensional white noise. *The Annals of Statistics 47*(6), 3382–3412.

Ling, S., R. S. Tsay, and Y. Yang (2021). Testing serial correlation and arch effect of high-dimensional time-series data. *Journal of Business & Economic Statistics 39*(1), 136–147.

Ljung, G. M. and G. E. Box (1978, 08). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Mukherjee, K. (2020). Bootstrapping m-estimators in generalized autoregressive conditional heteroscedastic models. *Biometrika 107*(3), 753–760.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

REFERENCES

Tsay, R. S. (2020). Testing serial correlations in high-dimensional time series via extreme value theory. *Journal of Econometrics 216*(1), 106–117.

Wang, L., E. Kong, and Y. Xia (2022). Bootstrap tests for high-dimensional white-noise. *Journal of Business & Economic Statistics 41*(1), 241–254.

Wang, L., B. Peng, and R. Li (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association 110*(512), 1658–1669.

Wang, R. and X. Shao (2020). Hypothesis testing for high-dimensional time series via self-normalization. *The Annals of Statistics 48*(5), 2728–2758.

Xu, M., D. Zhang, and W. B. Wu (2019). Pearson's chi-squared statistics: approximation theory and beyond. *Biometrika 106*(3), 716–723.

Zhang, X. and G. Cheng (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli 24*(4A), 2640–2675.

Zhu, K. (2016). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(2), 463–485.

Zhu, K. (2019). Statistical inference for autoregressive models under heteroscedasticity of unknown form. *The Annals of Statistics 47*(6), 3185–3215.

Zhu, K. and W. K. Li (2015). A bootstrapped spectral test for adequacy in weak arma models. *Journal of Econometrics 187*(1), 113–130.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0037

REFERENCES

Zhu, Q., R. Zeng, and G. Li (2020). Bootstrap inference for garch models by the least absolute deviation estimation. *Journal of Time Series Analysis 41*(1), 21–40.

[1]School of Statistics, Capital University of Economics and Business, Beijing 100070, P.R.China

E-mail: zhouzeren@cueb.edu.cn

[2]School of Mathematics and Statistics, Shanxi University, Taiyuan 030006, P.R.China; Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R.China

E-mail: mchen@amss.ac.cn