Journal of Econometrics 194 (2016) 283–297

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing super-diagonal structure in high dimensional covariance matrices[☆]

Jing He[a,c], Song Xi Chen[a,b,*]

[a] Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China

[b] The Center for Statistical Science, Peking University, Beijing 100871, China

[c] School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China

ARTICLE INFO

Article history:
Available online 1 June 2016

JEL classification:
C12
C14
G10

Keywords:
Bandable covariance
High dimensionality
Sparse covariance matrix
Multiple testing

ABSTRACT

The covariance matrices are essential quantities in econometric and statistical applications including portfolio allocation, asset pricing and factor analysis. Testing the entire covariance under high dimensionality endures large variability and causes a dilution of the signal-to-noise ratio and hence a reduction in the power. We consider a more powerful test procedure that focuses on testing along the super-diagonals of the high dimensional covariance matrix, which can infer more accurately on the structure of the covariance. We show that the test is powerful in detecting sparse signals and parametric structures in the covariance. The properties of the test are demonstrated by theoretical analyses, simulation and empirical studies.

1. Introduction

The covariance matrix of a random vector or a multivariate estimating function is a basic ingredient in multivariate analysis and econometrics in gaining information on the dependence between the components of the random vectors and the estimating functions. The celebrated Markowitz theory for optimal portfolio selection (Markowitz, 1952) is based on consistent estimation of the covariance matrix whose dimension is the number of assets of the portfolio. The sample variance is actively employed in an array of multivariate procedures such as the principal component analysis (PCA), the discrimination analysis and the factor analysis. In econometrics, the generalized method of moment (GMM) requires inversion of the covariance matrix of the multivariate moments as the weighting matrix. When the dimension of the data vector or the moments is fixed, the sample/empirical covariance is known to be consistent to the underlying covariance matrix.

Data with dimensions comparable to or larger than the sample size are increasingly encountered in econometric and statistical analyses. They include analyses of large panels of financial portfolios, on-line prices of consumer goods, macro-economic data that measure a large number of features of an economy; see Stock and Watson (2005), Bai and Ng (2002), Lam et al. (2011), Bai and Li (2012), Lam and Yao (2012) and Chang et al. (2015) for over-views and specific results. Fan et al. (2008) considered a covariance matrix estimator for a multi-factor model where the number of factors is allowed to grow with dimension p when p tends to infinity as the sample size n increases.

Extensive research in obtaining consistent estimators of high dimensional covariance matrix has been made. Bickel and Levina (2008a,b) proposed, respectively, the banding and the thresholding estimator of the covariance matrix by either banding or thresholding the sample covariance matrix. Wu and Pourahmadi (2003) and Rothman et al. (2010) studied methods based on the Cholesky decomposition. Cai et al. (2010) proposed a tapering estimator. The banding and tapering estimators are operational when the underlying covariance matrix $\Sigma = (\sigma_{i,j})_{p\times p}$ belongs to the so-called bandable class, which prescribes that $\sigma_{i,j}$ diminishes to zero at certain rates as either j or i increases. There is a set of high dimensionality tests on the covariance Σ. Testing for the identity or sphericity hypotheses of Σ has been considered in Ledoit and

☆ We would like to thank the editor of the Special Issue, Professor Rong Chen, and two referees for helpful comments and suggestions which have substantially improved the presentation of the paper. The research was partially supported by Natural Science Foundation of China grants 11131002, 71371016 and 71532001, National Key Basic Research Program of China Grant 2015CB856000, Center for Statistical Science and LMEQF at Peking University.

* Corresponding author at: Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China.

E-mail addresses: hejing1790@pku.edu.cn (J. He), csx@gsm.pku.edu.cn (S.X. Chen).

http://dx.doi.org/10.1016/j.jeconom.2016.05.007
0304-4076/

Wolf (2002) and Chen et al. (2010). Cai and Jiang (2011) and Qiu and Chen (2012) proposed tests for the bandedness of a covariance matrix. See also Schott (2005) and Srivastava (2005) for other formulations.

We propose a test regarding the super-diagonals of Σ, which has much smaller scale than the existing tests, and targets on global features of Σ, for instance the bandedness or specific parametric structure. The smaller scale of the super-diagonal as compared to the entire Σ does pose theoretical challenges when establishing the asymptotic properties of the test statistic. This is because the variation of the test statistic is much smaller, which requires finer derivations in the asymptotic analysis. The benefits of working with a test statistic being a smaller magnitude is a reduced variance and an increased signal-to-noise ratio of the test, which can produce more power than those targeting on the entire covariance matrix Σ. Tests for overall structures of Σ can be made by multiple testing on the super-diagonals in conjunction with the false discovery rate or the Bonferroni procedure. We demonstrate in the paper that the proposed test is useful to the inference of spatial econometrical and statistical models on covariance, which tends to be written in terms of the super-diagonals (Kapoor et al., 2007; Baltagi et al., 2003; Lee and Yu, 2010; Rodríguez and Bárdossy, 2014).

The paper is organized as follows. We outline the framework of the testing problem, including the hypotheses, assumptions and the proposed test statistics in Section 2. Section 3 provides the theoretical properties of the test statistics and the multiple testing procedure. In Sections 4 and 5, we discuss tests for bandedness and parameter structures of Σ, respectively. Results of simulation studies are provided in Section 6. An empirical analysis is reported in Section 7. All technical details are given in Appendix.

2. Preliminaries

Consider a p-dimensional generic random vector $X = (X_1, X_2, \dots, X_p)^T$, which has mean $\mu = (\mu_1, \mu_2, \dots, \mu_p)^T$ and covariance matrix $\Sigma = (\sigma_{i,j})_{p\times p}$. The observed data $X_i = (X_{i,1}, \dots, X_{i,p})^T$, $i = 1, \dots, n$, are independent copies of X. For $q = 0, 1, \dots, p-1$, let $D_q = \sum_{l=1}^{p-q} \sigma_{l,l+q}^2$ be the sum of the $\sigma_{i,j}^2$ along the qth super-diagonal, where D_0 represents that on the main diagonal.

We consider testing covariance structures with respect to the super- or sub-diagonals of Σ via D_q. We have two specific covariance structures in mind. One is the nonparametric banded structure in that $\sigma_{i,j} = 0$ for any $|i-j| > k$ for an integer k. The smallest such k is called the bandwidth of Σ. And the other is an isomorphic parametric structure where $\sigma_{i,j} = \sigma(|i-j|; \boldsymbol{\theta})$ for a finite dimensional parameter $\boldsymbol{\theta}$, which is a popular form in spatial econometrics.

The banding structure can be produced by a moving average structure such that, for $i = 1, \dots, n$,

$$X_{i,l} = \mu_l + \sum_{j=0}^{k} \gamma_j Z_{i,l-j},$$

where for each given i, μ_l is the mean of $X_{i,l}$ and $\{Z_{i,1}, Z_{i,2}, \dots\}$ is a sequence of independent white noise with zero mean and unit variance, $Z_{i,j} = 0$ for $j \le 0$, and $\gamma_0 = 1$. The integer k is the bandwidth of Σ.

More generally, we consider testing certain parametric model regarding the super-diagonal structure of Σ:

$$H_{0,q} : D_q = D_q(\boldsymbol{\theta}) \quad \text{vs } H_{1,q} : D_q \ne D_q(\boldsymbol{\theta})$$

where $D_q(\boldsymbol{\theta})$ is a parametric form, for $q = 1, 2, \dots, p-1$, and $\boldsymbol{\theta}$ is a finite dimensional parameter. For bandedness test, $D_q(\boldsymbol{\theta}) \equiv 0$ for $q > k$. A motivation for such model comes from the spatial econometrics or statistics where the X_i consists of recordings at p locations. If $\{X_{i,j}\}_{j=1}^p$ is weakly stationary, $\sigma_{j,j+h} = \mathrm{Cov}(X_{i,j}, X_{i,j+h}) = C(h)$ defines a covariance function $C(\cdot)$. Let $\boldsymbol{\theta} = (\sigma^2, \phi)^T$ and commonly used spatial models for $C(\cdot)$ include the spherical model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \left(1 - 1.5(h/\phi) + 0.5(h/\phi)^3\right), \phi > 0, h < \phi;$$

the wave model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \phi \sin(h/\phi)/h;$$

the exponential model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \exp(-h/\phi) \quad \text{and}$$

the Gaussian model

$$C(h; \boldsymbol{\theta}) = \sigma^2 \exp\left(-h^2/\phi\right).$$

See Cressie (1993), Kapoor et al. (2007), Baltagi et al. (2003), Lee and Yu (2010) and Rodríguez and Bárdossy (2014) for more details.

The proposed test statistics for super-diagonals are based on an unbiased estimator of D_q:

$$\hat{D}_q = \sum_{l=1}^{p-q} \Bigg\{ \frac{1}{P_n^2} \sum_{i,j}^{*} (X_{i,l}X_{i,l+q})(X_{j,l}X_{j,l+q}) - \frac{2}{P_n^3} \sum_{i,j,k}^{*} X_{i,l}X_{k,l+q}(X_{j,l}X_{j,l+q}) + \frac{1}{P_n^4} \sum_{i,j,k,m}^{*} X_{i,l}X_{j,l+q}X_{k,l}X_{m,l+q} \Bigg\},$$

where $\sum^*$ denotes summation over mutually different subscripts, and $P_n^b = n!/(n-b)!$. It is clear that $\hat{D}_q$ is a linear combinations of U-statistics. Without loss of generality, we assume $\mu = 0$ since $\hat{D}_q$ is invariant to the location shift.

To quantify the dependence among components of the data vector, we invoke the notion of α-mixing. The α-mixing coefficient of the generic $X = (X_1, \dots, X_p)^T$ is defined as

$$\alpha_X(k) = \sup_{m\in\mathbb{Z}} \alpha(\mathcal{G}_1^m, \mathcal{G}_{m+k}^p), \tag{2.1}$$

where $\alpha(\mathcal{G}_1^m, \mathcal{G}_{m+k}^p) = \sup\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{G}_1^m, B \in \mathcal{G}_{m+k}^p\}$, $\mathcal{G}_1^m$ and $\mathcal{G}_{m+k}^p$ are the σ-fields generated by $\{X_1, \dots, X_m\}$ and $\{X_{m+k}, \dots, X_p\}$, respectively. If $\lim_{k\to\infty} \alpha_X(k) = 0$, the sequence of components in X is said to be α-mixing. Furthermore, we denote the eigenvalues of Σ as $\lambda_{max}(\Sigma) = \lambda_1(\Sigma) \ge \lambda_2(\Sigma) \ge \dots \ge \lambda_p(\Sigma) = \lambda_{min}(\Sigma)$.

Our test procedure does not require any explicit relationship between the sample size n and the dimension p other than that they both diverges to infinity. It allows p to be much larger than n, that is, the "large p, small n" situation. We assume the following conditions in our analysis.

A1 There are positive constants c and $a \in (0, 1)$ such that $\alpha_X(k) \le ca^k$.

A2 The eighth moment of X_ℓ is uniformly bounded, i.e. $\sup_{1\le\ell\le p} \mathrm{E}|X_\ell|^8 \le M$, for a positive constant M. There exists a positive constant ϵ_0, such that $\lambda_{min}(\Sigma) \ge \epsilon_0 > 0$.

A3 Data vectors X_i are generated by $X_i = \Gamma Z_i$ for $i = 1, 2, \cdots, n$, where $\Gamma = \left(\Gamma_{i,j}\right)_{p\times m}$ is a $p \times m$ constant matrix, satisfying $\Gamma\Gamma' = \Sigma$ and $m \ge p$, and $Z_1, Z_2, \cdots, Z_n$ are independently and identically distributed (IID) m-dimensional random vectors such that $\mathrm{E}(Z_i) = 0$ and $\mathrm{Var}(Z_i) = I_m$. Write $Z_i = (Z_{i,1}, \dots, Z_{i,m})^T$. We assume $Z_{i,j}$ have uniformly bounded 8th moment, and there exists a finite constant Δ such that $\mathrm{E}(Z_{i,j}^4) = 3 + \Delta$ for $j = 1, \dots, m$, and $\mathrm{E}(Z_{i,j_1}^{\ell_1} Z_{i,j_2}^{\ell_2} \cdots Z_{i,j_q}^{\ell_q}) = \mathrm{E}(Z_{i,j_1}^{\ell_1})\mathrm{E}(Z_{i,j_2}^{\ell_2}) \cdots \mathrm{E}(Z_{i,j_q}^{\ell_q})$ for any integers $\ell_\nu \ge 0$ with $\sum_{\nu=1}^{q} \ell_\nu \le 8$ and distinct subscripts $j_1, \dots, j_q$.

The α-mixing coefficient in Assumption A1 can be relaxed to be polynomial decay such that $\alpha_X(k) \le ck^{-\beta}$ for positive constants

c and $\beta \in (1, \infty)$ without altering the main conclusion of the paper. The technical details would be more involved though. From A2, by Lyapunov's inequality, we can infer that the rth moment of X_ℓ is also uniformly bounded for $1 \le r < 8$. Assumptions A1 and A2, and the α-mixing of the sequence $\{X_\ell\}_{\ell=1}^p$ together with the Davydov's inequality imply that

$$|\sigma_{i,j}| \le 12\|X_i\|_4\|X_j\|_4\,(\alpha_X(|i-j|))^{\frac{1}{2}} \asymp ca^{\frac{|i-j|}{2}}, \tag{2.2}$$

where $\|X_i\|_r = (\mathrm{E}|X_i|^r)^{1/r}$. Throughout the paper, we define $a_n \asymp b_n$ if and only if $a_n = O(b_n)$ and $b_n = O(a_n)$ for two nonrandom sequences $\{a_n\}$ and $\{b_n\}$. By Gershgorin's Theorem, the eigenvalues of Σ satisfy, for a positive constant C,

$$\lambda_{max}(\Sigma) \le \max_i \sum_{j=1}^p |\sigma_{i,j}| < C.$$

Let $h(q) \triangleq \frac{1}{p-q}D_q = \frac{1}{p-q}\sum_{l=1}^{p-q}\sigma_{l,l+q}^2$ be the average signal strength on the qth super-diagonal. Then, (2.2) implies that

$$h(q) \le ca^q \to 0 \quad \text{as } q \to \infty, \quad \text{and}$$
$$\sum_{q>k} h(q) \le c\sum_{q>k} a^q \to 0 \quad \text{as } k \to \infty. \tag{2.3}$$

Assumption A3 is widely assumed in the studies of high dimensional multivariate analysis, for instance in Bai and Saranadasa (1996), Qiu and Chen (2012) and Li and Chen (2012). It includes the Gaussian family and members of the elliptically contoured distributions as special cases. It leads to trackable expressions of higher order cross moments of X_i.

Our analysis on the covariance are more intimately related to

$$Y_t^{l_1,l_2} = X_{t,l_1}X_{t,l_2} - \sigma_{l_1,l_2} \quad \text{and} \quad \omega_{l_1,l_2} = \mathrm{Cov}(Y_t^{l_1,l_1+q}, Y_t^{l_2,l_2+q})$$

for $l_1, l_2 = 1, \ldots, p$. Moreover, for a given $q = 0, 1, \ldots, p-1$, let

$$\vec{Y}_t(q) = \left(Y_t^{1,1+q}, Y_t^{2,2+q}, \ldots, Y_t^{p-q,p}\right)^T,$$

whose covariance is $W_q = \left(\omega_{l_1,l_2}\right)_{(p-q)\times(p-q)}$. Since $\{X_i\}_{i=1}^n$ are IID, $\{\vec{Y}_t(q)\}_{t=1}^n$ are also IID for each given q. Let $\alpha_Y(k)$ be the α-mixing coefficient of the sequence $\left\{Y_t^{l,l+q}\right\}_{l=1}^{p-q}$, which can be similarly defined as $\alpha_X(k)$ in (2.1). It can be inferred that for each given q, $\alpha_Y(k) \le \alpha_X(k-q)$ for $k > q$.

According to Lemma 1 given in the Appendix, it can be readily shown that

$$\mathrm{tr}(W_q^2) = \sum_{l=1}^{p-q}\lambda_l^2(W_q) \le M'^2(p-q) = O(p-q).$$

To facilitate the analysis on the asymptotic properties of $\hat{D}_q$ and estimating the variance of $\hat{D}_q$ in Section 3, we assume more specifically the following

A4 $\mathrm{tr}(W_q^2) = \sum_{i=1}^{p-q}\lambda_i^2(W_q) \asymp p-q$.

The properties of the test statistic $\hat{D}_q$ are given in the following proposition.

Proposition 1. *Under Assumptions* A1 – A4, *for* $q = 0, 1, \ldots, p-1$,

$$\mathrm{E}(\hat{D}_q) = D_q \quad \textit{and}$$
$$\mathrm{Var}(\hat{D}_q) = \left(\frac{4}{n}\sum_{l_1,l_2=1}^{p-q}\sigma_{l_1,l_1+q}\sigma_{l_2,l_2+q}\omega_{l_1,l_2} + \frac{2}{n(n-1)}\mathrm{tr}(W_q^2)\right) \times \{1+o(1)\}.$$

Hence, if Σ is banded with bandwidth k, then for any $q > k$,

$$\mathrm{Var}(\hat{D}_q) = \frac{2}{n(n-1)}\mathrm{tr}(W_q^2).$$

3. Asymptotic results

In this section, we first establish the asymptotic distribution of $\hat{D}_q$. Let

$$V_{p,q,1} = \sum_{l_1,l_2=1}^{p-q}\sigma_{l_1,l_1+q}\sigma_{l_2,l_2+q}\omega_{l_1,l_2} \quad \text{and}$$
$$V_{p,q,2} = \sum_{l_1,l_2=1}^{p-q}\omega_{l_1,l_2}^2 = \mathrm{tr}(W_q^2).$$

The asymptotic variance $\mathrm{Var}(\hat{D}_q)$ has different leading orders for different q. From Assumption A1 and the proof of Proposition 1,

$$\frac{4}{n}V_{p,q,1} \asymp \frac{(p-q)a^q}{n} \quad \text{and} \quad \frac{2}{n(n-1)}V_{p,q,2} \asymp \frac{p-q}{n^2}.$$

These imply the asymptotic variance of $\hat{D}_q$ under three regimes: (i) the smaller q case such that $q = o\{\log(n)\}$, (ii) the "median" q case such that $a^q \asymp 1/n$ and (iii) the larger q case such that $\log(n) = o(q)$ and $q = o(p)$, respectively. The three regimes of q induce different orders for $\mathrm{Var}(\hat{D}_q)$. In the smaller q regime (i), the leading term of $\mathrm{Var}(\hat{D}_q)$ is $\frac{4}{n}V_{p,q,1}$; and in the median q regime (ii), $\frac{4}{n}V_{p,q,1}$ and $\frac{2}{n(n-1)}V_{p,q,2}$ are the joint leading order terms, whereas for the larger q in regime (iii), the leading order term of $\mathrm{Var}(\hat{D}_q)$ becomes $\frac{2}{n(n-1)}V_{p,q,2}$, which is of smaller order than the variance under the smaller q regime (i). These differential orders in the variance reflect two facts. One is that $\hat{D}_q$ involves far more terms when q is smaller, and hence has larger variation relative to the larger q case. The other is that, according to (2.3), the magnitude of $h(q)$ declines as q increases. The latter indicates that $\hat{D}_q$ not only has far less terms for larger q, but also those terms have much smaller magnitude, and hence less variation.

Define $\sigma_{\hat{D}_q}^2 = \frac{4}{n}V_{p,q,1} + \frac{2}{n(n-1)}V_{p,q,2}$. The following theorem establishes the asymptotic normality of $\hat{D}_q$ for $q = o(p)$.

Theorem 1. *Under Assumptions* A1–A4, *for* $q = o(p)$,

$$\sigma_{\hat{D}_q}^{-1}\left(\hat{D}_q - D_q\right) \xrightarrow{d} N(0,1) \quad \textit{as } n \to \infty \quad \textit{and} \quad p \to \infty.$$

We note in passing that for the large q case such that

$$(p-q)/p \to c \in (0,1) \tag{3.1}$$

the variance expression for the regime (iii) of q is still applicable, and specifically $\mathrm{Var}(\hat{D}_q) \asymp \frac{p-q}{n^2}$. However, the analogues of the two key conditions (A.3) and (A.4) in establishing the Martingale Central Limit Theorem (CLT) (Hall and Heyde, 1980) are no longer assured. It is uncertain if we could still have the asymptotic normality under (3.1).

It can be shown that Theorem 1 remains valid if the exponentially decay α-mixing coefficient of X is relaxed to be polynomially decay such that $\alpha_X(k) \le ck^{-\beta}$ for positive constants c and $\beta \in (1, \infty)$, except that the order that divides the smaller and larger q becomes $n^{1/\beta}$, which is larger than the order for the exponentially decay case.

In order to establish a test procedure based on the asymptotic normality, it is necessary to estimate $V_{p,q,1}$ and $V_{p,q,2}$, respectively. We estimate $V_{p,q,1}$ and $V_{p,q,2}$ by U-statistics. Specifically, $V_{p,q,1}$ is estimated by

$$\hat{V}_{p,q,1} = \sum_{l_1,l_2=1}^{p-q}\sum_{i,j,k}^{*}\frac{1}{P_n^3}(\tilde{X}_{i,l_1}\tilde{X}_{i,l_1+q})(\tilde{X}_{j,l_1}\tilde{X}_{j,l_1+q}) \times \left(\tilde{X}_{k,l_1}\tilde{X}_{k,l_1+q} - \hat{\sigma}_{l_1,l_1+q}^{(i,j,k)}\right)\left(\tilde{X}_{k,l_2}\tilde{X}_{k,l_2+q} - \hat{\sigma}_{l_2,l_2+q}^{(i,j,k)}\right), \tag{3.2}$$

where $\hat{\sigma}_{l,l+q}^{(i,j,k)}$ is the sample covariance of X_l and X_{l+q} by avoiding the ith, jth and kth observation and $\tilde{X}_{i,l} = (X_{i,l} - \bar{X}_l^{(i,j,k)})$, $\bar{X}_l^{(i,j,k)} = \frac{1}{n-3}\sum_{s\neq i,j,k} X_{s,l}$. The rationale for using $\hat{\sigma}_{l,l+q}^{(i,j,k)}$ instead of $\hat{\sigma}_{l,l+q}$ in the two subtractions in (3.2) is to have less bias in the estimation of $V_{p,q,1}$, as showed in the proof of Proposition 2 in the appendix.

We use the following estimator to estimate $V_{p,q,2}$, motivated by Chen et al. (2010) and Li and Chen (2012),

$$\hat{V}_{p,q,2} = \frac{1}{P_n^2}\sum_{i,j}^{*}\left(\sum_{l=1}^{p-q}(\tilde{X}_{i,l}\tilde{X}_{i,l+q} - \hat{\sigma}_{l,l+q}^{(i,j)})(\tilde{X}_{j,l}\tilde{X}_{j,l+q} - \hat{\sigma}_{l,l+q}^{(i,j)})\right)^2, \tag{3.3}$$

where $\hat{\sigma}_{l,l+q}^{(i,j)}$ is similarly defined as $\hat{\sigma}_{l_1,l_1+q}^{(i,j,k)}$ in (3.2). A computationally more efficient estimator is

$$\tilde{V}_{p,q,2} = \frac{1}{P_n^2}\sum_{i,j}^{*}\left(\vec{Y}_i(q)^T\vec{Y}_j(q)\right)^2 \tag{3.4}$$

which uses the sample covariance $\hat{\sigma}_{l,l+q}$ instead of the computationally more expensive $\hat{\sigma}_{l,l+q}^{(i,j)}$, where

$\vec{Y}_i(q) = \left(\hat{Y}_i^{1,1+q}, \ldots, \hat{Y}_i^{p-q,p}\right)^T$ and $\hat{Y}_i^{l,l+q} = X_{i,l}X_{i,l+q} - \hat{\sigma}_{l,l+q}$.

The following proposition establishes the consistency of the above estimators.

Proposition 2. *Under Assumptions* A1–A4, *for* $q = o(p)$, *as* $n \to \infty$ *and* $p \to \infty$, $\hat{V}_{p,q,1}/V_{p,q,1} \xrightarrow{p} 1$, $\hat{V}_{p,q,2}/V_{p,q,2} \xrightarrow{p} 1$ *and* $\tilde{V}_{p,q,2}/V_{p,q,2} \xrightarrow{p} 1$.

Let $\hat{\sigma}_{\hat{D}_q}^2 = \frac{4}{n}\hat{V}_{p,q,1} + \frac{2}{n(n-1)}\hat{V}_{p,q,2}$. Proposition 2 together with the asymptotic normality established in Theorem 1 implies, via Slutsky Theorem, that under Assumptions A1–A4, for $q = o(p)$,

$$\hat{\sigma}_{\hat{D}_q}^{-1}\left(\hat{D}_q - D_q\right) \xrightarrow{d} N(0,1) \quad \text{as } n\to\infty \quad \text{and} \quad p\to\infty. \tag{3.5}$$

4. Tests for bandedness

A test for $H_{0,q}: D_q = 0$ versus $H_{1,q}: D_q > 0$ is facilitated by (3.5). Note that the test should be a one-sided test since D_q is always greater than or equal to zero. It is worth mentioning that under $H_{0,q}: D_q = 0$, $V_{p,q,1} = 0$ so that we only need to estimate $V_{p,q,2}$. Thus, a test with a nominal α level of significance rejects $H_{0,q}: D_q = 0$ for $q = o(p)$ if

$$\hat{D}_q > z_{1-\alpha}\sqrt{\frac{2\hat{V}_{p,q,2}}{n(n-1)}},$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of $N(0,1)$.

The asymptotic normality of $\hat{D}_q$ in Theorem 1 allows a power evaluation of the above test. Let $\delta_{np,q} = D_q/\sigma_{\hat{D}_q}$, which may be viewed as a signal to noise ratio for the testing problem. Denote the power of the test by $\beta_{q,\alpha}$.

Theorem 2. *Under Assumptions* A1–A4, *and the alternative* $H_{1,q}$: $D_q > 0$, *then, for* $q = o(p)$, $\liminf_{n,p\to\infty}\beta_{q,\alpha} \geq 1 - \Phi\left(z_{1-\alpha} - \liminf_{n,p\to\infty}\delta_{np,q}\right)$.

For $q = o(\log n)$, $\delta_{np,q} \asymp \sqrt{n(p-q)}h(q)$. Hence, if $\sqrt{n(p-q)}\,h(q) \to \infty$ for $q = o(\log n)$, the power $\beta_{q,\alpha} \to 1$ as $n, p \to \infty$. For large q such that $\log(n) = o(q)$ and $q = o(p)$, if $\delta_{np,q} \asymp n\sqrt{p-q}h(q) \to \infty$, $\beta_{q,\alpha} \to 1$ as $n, p \to \infty$. These results imply that despite $h(q)$ may become smaller as q gets larger, the multiplication by $\sqrt{n(p-q)}$ and $n\sqrt{p-q}$ for the smaller q and the large q cases offsets the declining $h(q)$ to attain the consistency of the test. In a sense, for a fixed q, a larger p is in fact beneficial to the power, as confirmed in the simulation study.

The above individual test for $D_q = 0$ may be combined via a multiple testing procedure to form a test for the bandedness of Σ. Let

$B_k(\Sigma) = \left(\sigma_{i,j}\mathrm{I}\{|i-j| \leq k\}\right)_{p\times p}$

be the banding operator. Then, the hypothesis $H_0 : \Sigma = B_k(\Sigma)$ can be tested by conducting multiple testing for $H_{0,q} : D_q = 0$ for $q = k+1, \ldots, k+K$ via implementing either the Bonferroni correction or the False Discovery Rate (FDR) control procedure (Benjamini and Hochberg, 1995), whenever $k+K = o(p)$, where K is the number of super-diagonals being covered. For using the FDR based multiple testing, denote the p-values for testing $H_{0,k+1}, H_{0,k+2}, \ldots, H_{0,k+K}$ as $P_1, P_2, \ldots, P_K$, respectively. Let $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(K)}$ be the ordered p-values and $H_{(0,k+i)}$ be the null hypothesis corresponding to $P_{(i)}$. If $P_{(K)} < \alpha$, all the hypotheses are rejected. If $P_{(K)} \geq \alpha$, let ℓ be the largest i such that $P_{(i)} \leq \frac{i}{K}\alpha$, then we reject all $H_{(0,k+i)}$ for $i = 1, 2, \ldots, \ell$. The hypothesis $H_0 : \Sigma = B_k(\Sigma)$ is rejected if $\ell > 1$. For using the Bonferroni multiple testing, $H_{0,k+i}$ is rejected if $P_i \leq \alpha/K$. And the joint hypothesis $H_0 : \Sigma = B_k(\Sigma)$ is rejected if there is any $H_{0,k+i}$ being rejected.

The power of the multiple test is determined by the power evaluation of individual tests in Theorem 2. The empirical power of the test is evaluated in simulation studies section.

5. Tests for parametric structures

We consider testing for parametric form $D_q(\boldsymbol{\theta})$ for a parameter $\boldsymbol{\theta} \in \Theta$ which is a compact d-dimensional parameter space and d is fixed. Verifying the parametric form $D_q(\boldsymbol{\theta})$ by data provides confirmation or otherwise on the super-diagonal structure of Σ and facilitates more efficient estimation for Σ. Specifically, we consider testing

$$H_0 : D_q = D_q(\boldsymbol{\theta}),\ q = 1, \ldots, N \quad \text{for a } \boldsymbol{\theta} \in \Theta \tag{5.1}$$

versus H_1 : $D_q \neq D_q(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta$ and some $q \in \{1, \ldots, N\}$, where N is an integer that is $N = o(p)$. It implies that Σ has a parametric structure along the super-diagonals. It is worth mentioning that the amount of data information only allows us to test up to N super-diagonals, while N diverges to infinity more slowly than p as $n \to \infty$. As $\boldsymbol{\theta}$ is unknown, we need to first construct a consistent estimator $\hat{\boldsymbol{\theta}}$ using $\hat{D}_1, \ldots, \hat{D}_r$. As $\boldsymbol{\theta}$'s dimension is fixed, we consider r being fixed and $r \geq d$. We then establish the asymptotic normality for $\hat{D}_q - D_q(\hat{\boldsymbol{\theta}})$ for $q = 1, \ldots, N$. Finally we apply the multiple testing procedure by controlling the FDR for H_0 in (5.1).

Let $g(\boldsymbol{\theta}; q) = \left(\hat{D}_q - D_q(\boldsymbol{\theta})\right)/(p-q)$ and an $r \times 1$ vector

$$G_r(\boldsymbol{\theta}) = \begin{pmatrix} g(\boldsymbol{\theta};1) \\ \vdots \\ g(\boldsymbol{\theta};r)\end{pmatrix} = \begin{pmatrix} (\hat{D}_1 - D_1(\boldsymbol{\theta}))/(p-1) \\ \vdots \\ (\hat{D}_r - D_r(\boldsymbol{\theta}))/(p-r)\end{pmatrix}. \tag{5.2}$$

We estimate $\boldsymbol{\theta}$ by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} G_r(\boldsymbol{\theta})^T G_r(\boldsymbol{\theta}). \tag{5.3}$$

Define the infinity norm $\|(\theta_1, \ldots, \theta_d)\|_\infty = \max_{1\leq i\leq d}|\theta_i|$. The following assumptions are needed to ensure the consistency of $\hat{\boldsymbol{\theta}}$.

C1 The parametric space Θ is a compact subset of $\mathbb{R}^d$ and $\boldsymbol{\theta}_0$ is a fixed value in the interior of Θ. Under H_0, $\mathrm{E}(g(\boldsymbol{\theta}; q)) = 0$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for $q = 1, \ldots, p-1$.

C2 (i) For any q, the function $D_q(\boldsymbol{\theta})$ is continuous differentiable on Θ and satisfies the Lipschitz condition that there exists a constant l_q such that for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$, $|D_q(\boldsymbol{\theta}) - D_q(\boldsymbol{\theta}')| \leq l_q\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_\infty$, (ii) $\sum_{i=1}^r \nabla_{\boldsymbol{\theta}} D_i(\boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}} D_i^T(\boldsymbol{\theta}_0)/(p-i)^2$ is invertible, and (iii) $\frac{\partial D_q(\boldsymbol{\theta})}{\partial \theta_j}/(p-q) \asymp 1$, $\frac{\partial^2 D_q(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k}/(p-q) \asymp 1$ and $\frac{\partial^3 D_q(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l}/(p-q) \asymp 1$ for each $q = 1, \ldots, r, j, k, l = 1, \ldots, d$ and any $\boldsymbol{\theta} \in \Theta$.

Assumption C1 is the identification condition for $\boldsymbol{\theta}_0$. In order to ensure the consistency of $\hat{\boldsymbol{\theta}}$ under H_0, regularity conditions for $D_q(\boldsymbol{\theta})$ are needed in Assumption C2. Specifically, Assumption C2 (i) assures that $D_q(\boldsymbol{\theta})$ satisfies the Lipschitz condition for any q, which is required for establishing uniform convergence in probability of $G_r(\boldsymbol{\theta})$ to $\mathrm{E}G_r(\boldsymbol{\theta})$. The latter is needed in obtaining the consistency of $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$ under H_0. Assumption C2 (ii) is assumed to exclude cases that $\nabla_{\boldsymbol{\theta}} G_r(\boldsymbol{\theta}_0) = 0$. And Assumption C2 (iii) is used for controlling the remainder term of a Taylor expansion in establishing the convergence rate of $\hat{\boldsymbol{\theta}}$. For the exponential and polynomial covariance models given in (6.2) and (6.3), C2 (i)–(iii) are satisfied. It is worth mentioning that if $\{X_{i,l}\}_{l\geq 1}$ follows a mean zero Gaussian AR(1) model such that

$$X_{i,l} = \theta X_{i,l-1} + Z_{i,l} \tag{5.4}$$

with IID Gaussian white noise $\{Z_{i,l}\}_{l\geq 1}$, then Assumption C2 rules out the case of $\theta = 0$. Specifically, we consider $\Theta = [\epsilon - 1, -\epsilon] \bigcup [\epsilon, 1 - \epsilon]$ for a sufficiently small constant $\epsilon > 0$. As a matter of fact, if $\theta_0 = 0$ in Model (5.4), $D_q(\theta_0) = 0$ for $q \geq 1$ and Σ is a diagonal matrix. We are still able to implement the test outlined in Section 4, though we cannot obtain a consistent estimator for θ_0 by (5.3).

The consistency with the convergence rate of $\hat{\boldsymbol{\theta}}$ under the null hypothesis H_0 is given in the following theorem.

Theorem 3. *Suppose that Assumptions* A1–A4 *and* C1–C2 *hold. Then, under H_0, as $n \to \infty$ and $p \to \infty$,* $\sqrt{np}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = O_p(1)$.

We only consider estimating $\hat{\boldsymbol{\theta}}$ using the first r super-diagonals $\hat{D}_1, \ldots, \hat{D}_r$ for the sake of easy implementation. It can be shown that the consistency of $\hat{\boldsymbol{\theta}}$ is still valid if $\hat{\boldsymbol{\theta}}$ is estimated by $\hat{D}_{q_1}, \ldots, \hat{D}_{q_r}$ and $q_1, \ldots, q_r$ are chosen such that q_1 is fixed and there is a fixed proportion of $q_i = o(\log(n))$ whereas the rest of q_i are of larger order of $\log(n)$ and $q_i = o(p)$.

It can be shown that, except for different asymptotic variance, the asymptotic normality in Theorem 1 is still valid by plugging in $\hat{\boldsymbol{\theta}}$. We denote the asymptotic variance of $\hat{D}_q - D_q(\hat{\boldsymbol{\theta}})$ by $\sigma_{q,r}^2$, which is defined separately in the following two regimes:

(i) For $q \in \{1, \ldots, r\}$, $\sigma_{q,r}^2 = \mathbf{u}_{q,r,1}(\boldsymbol{\theta}_0)\mathrm{Var}(G_r(\boldsymbol{\theta}_0))\mathbf{u}_{q,r,1}^T(\boldsymbol{\theta}_0)$, where the $1 \times r$ vector

$$\mathbf{u}_{q,r,1}(\boldsymbol{\theta}_0) = (p-q)e_q^T + \nabla_{\boldsymbol{\theta}} D_q^T(\boldsymbol{\theta}_0) \times \left[\nabla_{\boldsymbol{\theta}} G_r^T(\boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}} G_r(\boldsymbol{\theta}_0)\right]^{-1} \nabla_{\boldsymbol{\theta}} G_r^T(\boldsymbol{\theta}_0) \tag{5.5}$$

and e_i is the ith unit vector.

(ii) For $q > r$ and $q = o(p)$, by defining an $(r+1) \times 1$ vector $F_{q,r}(\boldsymbol{\theta}_0) = \left(G_r^T(\boldsymbol{\theta}_0), g(\boldsymbol{\theta}_0; q)\right)^T$, $\sigma_{q,r}^2 = \mathbf{u}_{q,r,2}(\boldsymbol{\theta}_0)\mathrm{Var}(F_{q,r}(\boldsymbol{\theta}_0))$ $\mathbf{u}_{q,r,2}^T(\boldsymbol{\theta}_0)$, where

$$\mathbf{u}_{q,r,2}(\boldsymbol{\theta}_0) = \left(\nabla_{\boldsymbol{\theta}} D_q^T(\boldsymbol{\theta}_0)\left[\nabla_{\boldsymbol{\theta}} G_r^T(\boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}} G_r(\boldsymbol{\theta}_0)\right]^{-1} \times \nabla_{\boldsymbol{\theta}} G_r^T(\boldsymbol{\theta}_0), (p-q)\right) \tag{5.6}$$

is a $1 \times (r+1)$ vector.

Compared with the asymptotic variance $\sigma_{\hat{D}_q}^2$ in Theorem 1, $\sigma_{q,r}^2$ involves with the variations generated by $\hat{D}_1, \ldots, \hat{D}_r$ which are used to estimate $\boldsymbol{\theta}$. The asymptotic normality of $\hat{D}_q - D_q(\hat{\boldsymbol{\theta}})$ is given in the following theorem.

Theorem 4. *Suppose that the conditions in Theorem 3 hold. Then, under H_0, as $n \to \infty$ and $p \to \infty$, for $q = o(p)$,* $\sigma_{q,r}^{-1}\left(\hat{D}_q - D_q(\hat{\boldsymbol{\theta}})\right) \xrightarrow{d} N(0, 1)$.

We need to estimate $\sigma_{q,r}^2$ in order to develop tests based on the asymptotic normality. An estimator $\hat{\sigma}_{q,r}^2$ is defined in the Appendix as well as the proof of the following proposition.

Proposition 3. *Under Assumptions* A1–A4, *for $q = o(p)$, as $n \to \infty$ and $p \to \infty$,* $\hat{\sigma}_{q,r}^2/\sigma_{q,r}^2 \xrightarrow{p} 1$.

Based on Theorem 4 and Proposition 3, $H_{0,q} : D_q = D_q(\boldsymbol{\theta})$ can be tested which rejects $H_{0,q}$ if $|\hat{D}_q - D_q(\hat{\boldsymbol{\theta}})| > z_{1-\alpha/2}\hat{\sigma}_{q,r}$. The FDR multiple testing procedure can be applied for testing H_0: $D_q = D_q(\boldsymbol{\theta})$ for $q = 1 \ldots, N$.

We conclude this section by considering the power property of the proposed test, which requires first considering estimation of the parameter estimator $\hat{\boldsymbol{\theta}}$ under mis-specified super-diagonal models. Under H_1, define

$$\boldsymbol{\theta}_* = \arg\min_{\boldsymbol{\theta}\in\Theta} \bar{G}_r(\boldsymbol{\theta})^T \bar{G}_r(\boldsymbol{\theta})$$

where $\bar{G}_r(\boldsymbol{\theta}) = ((D_1 - D_1(\boldsymbol{\theta}))/(p-1), \ldots, (D_r - D_r(\boldsymbol{\theta}))/(p-r))^T$ is the expectation of $G_r(\boldsymbol{\theta})$. The following identification condition parallels to C1.

C3 Under H_1, $\bar{G}_r^T(\boldsymbol{\theta})\bar{G}_r(\boldsymbol{\theta})$ has a unique minimum at a fixed $\boldsymbol{\theta}_*$ which is in the interior of Θ and does not change as p increases.

Theorem 5 shows that $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_*$ in probability under H_1 at the rate of $\sqrt{np}$.

Theorem 5. *Suppose that Assumptions* A1–A4, C2 *and* C3 *hold. Then, under H_1, as $n \to \infty$ and $p \to \infty$,* $\sqrt{np}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*\right) = O_p(1)$.

The following theorem evaluates the probability of rejecting $H_{0,q} : D_q = D_q(\boldsymbol{\theta})$ under H_1 by the test that is based on Theorem 4. Let $\delta_{np,q}^* = |D_q - D_q(\boldsymbol{\theta}_*)|/\sigma_{q,r}$, which may be viewed as the signal-to-noise ratio of the test. Let $\beta_{q,\alpha}$, again, be the power of the test for $H_{0,q} : D_q = D_q(\boldsymbol{\theta})$.

Theorem 6. *Suppose that the conditions in Theorem 5 hold. Then, under $H_{1,q} : D_q \neq D_q(\boldsymbol{\theta})$, if $\liminf_{n,p\to\infty} \delta_{np,q}^* \to \infty$, $\beta_{q,\alpha} \to 1$ as $n, p \to \infty$.*

The theorem shows that the proposed test for the parametric model for the qth super-diagonal is consistent provided the signal-to-noise ratio $\delta_{np,q}^*$ diverges.

6. Simulation results

In this section, we report results from simulation experiments which are designed to investigate the numerical performance of the proposed tests.

We first considered testing for the banded structure of Σ using the proposed test on the super-diagonals in conjunction with the FDR control procedure. We also compared with the tests for the banded Σ proposed by Cai and Jiang (2011) and Qiu and Chen (2012). We generated $X_i = (X_{i,1}, \ldots, X_{i,p})$ independently according to the following moving average model:

$$X_{i,j} = Z_{i,j} + 0.4Z_{i,j-1} + 0.4Z_{i,j-2} + 0.4Z_{i,j-3} + 0.4Z_{i,j-4} + 0.4Z_{i,j-5}, \tag{6.1}$$

where $\{Z_{i,j}\}_{j=1}^p$ were i.i.d. random variables with zero mean and unit variance. Two distributions for $Z_{i,j}$ were considered. One was the $N(0, 1)$; and the other was the standardized Gamma(1, 0.5)

Table 1
Empirical sizes of the proposed multiple test, QC's test and CJ's test for testing $H_0 : \Sigma = B_5(\Sigma)$ when data are generated from model Eq. (6.1). The proposed multiple testing is conducted with FDR controlled at 5% and QC's test is implemented at the nominal significance level $\alpha = 0.05$. Comparable empirical sizes are reported for CJ's test and inside the parentheses are the nominal significance levels of CJ's test.

p	Normal			Gamma		
	n			n		
	50	100	200	50	100	200
	Proposed multiple test					
50	0.045	0.046	0.053	0.041	0.061	0.044
100	0.059	0.055	0.034	0.039	0.048	0.039
200	0.048	0.037	0.044	0.050	0.050	0.062
400	0.036	0.049	0.045	0.034	0.038	0.033
600	0.032	0.033	0.036	0.035	0.040	0.046
1000	0.016	0.038	0.032	0.014	0.032	0.044
	QC's test					
50	0.053	0.032	0.047	0.042	0.042	0.041
100	0.049	0.056	0.033	0.054	0.037	0.043
200	0.051	0.053	0.038	0.050	0.058	0.044
400	0.068	0.048	0.066	0.056	0.052	0.054
600	0.053	0.051	0.048	0.060	0.055	0.058
1000	0.049	0.053	0.054	0.048	0.054	0.056
	CJ's test					
50	0.050(0.28)	0.054(0.18)	0.049(0.18)	0.050(0.11)	0.053(0.07)	0.049(0.08)
100	0.053(0.30)	0.051(0.22)	0.054(0.17)	0.046(0.12)	0.055(0.06)	0.049(0.06)
200	0.047(0.40)	0.052(0.22)	0.053(0.14)	0.052(0.14)	0.055(0.07)	0.050(0.04)
400	0.055(0.61)	0.050(0.24)	0.053(0.14)	0.050(0.19)	0.054(0.07)	0.053(0.05)
600	0.049(0.65)	0.051(0.30)	0.050(0.14)	0.050(0.19)	0.048(0.05)	0.064(0.04)
1000	0.047(0.80)	0.050(0.36)	0.051(0.17)	0.052(0.25)	0.048(0.05)	0.053(0.04)

distribution so that it has zero mean and unit variance. The covariance of model (6.1) is banded with the bandwidth $k = 5$. When evaluating the empirical sizes, we tested $H_0 : \Sigma = B_5(\Sigma)$ via testing multiple nulls $H_{0,q} : D_q = 0$ from $q = 6$ to $q = \lfloor 3p/\log(p) \rfloor$ coupled with the FDR procedure. For power evaluation, we tested $H_0 : \Sigma = B_4(\Sigma)$ and conducted the multiple testing on $H_{0,q} : D_q = 0$ from $q = 5$ to $q = \lfloor 3p/\log(p) \rfloor$. The sample size was $n = 50, 100, 200$ and the dimension $p = 50, 100, 200, 400, 600, 1000$, respectively, which created situations of the "large p, small n".

The empirical sizes and powers of the proposed test with the FDR controlled at 5% and those of Qiu and Chen (QC)'s test and Cai and Jiang (CJ)'s test at 5% significance level under the same data generating settings were reported in Tables 1 and 2, respectively. Table 1 indicates that there were some size distortion with CJ's test for bandedness. To ensure fairer comparison of the power, we used larger nominal size, given inside the parentheses, for CJ's proposal such that empirical sizes of their test were closer to 5%. We observed from Table 1 that there was some size deflation of the proposed multiple testing procedure, owing to the conservative nature of the FDR implementation. Despite lower empirical sizes, Table 2 shows that the proposed test on the super-diagonals had significantly better power than those of QC's and CJ's test, especially when the sample size was small. The reason for QC's and CJ's tests having lower power than the proposed test based on the super-diagonals was due to the statistics of these two tests were global involving much more entries of the sample covariance, which caused a much larger variance of the test statistics. Furthermore, CJ's test was based on an extreme value type statistic which converges slowly and they required Gaussian assumption. Indeed, the powers of the two tests were not restored until the sample size reaches 200 in the simulation.

We then considered testing for a parametric model of the super-diagonals $H_{0,q} : D_q = D_q(\boldsymbol{\theta}), q = 1, \ldots, N$ and $N = o(p)$ for a $\boldsymbol{\theta} \in \Theta$. We generated p-dimensional X_i IID from $N(0, \Sigma)$, $i = 1, \ldots, n$, where $\Sigma = (\sigma_{i,j})_{p \times p}$. Two forms of $\sigma_{i,j}$ were experimented. One was the exponentially decay covariances

$$\sigma_{i,j} = \mathrm{I}(i = j) + \theta_1 \exp(-|i - j|/\theta_2) \tag{6.2}$$

Table 2
Empirical powers of the proposed multiple test, QC's test and CJ's test for testing $H_0 : \Sigma = B_4(\Sigma)$ when data are generated from model (6.1). The proposed multiple testing is conducted with FDR controlled at 5% and QC's test is implemented at the nominal significance level $\alpha = 0.05$. The empirical testing powers of CJ's test are corresponding to the empirical sizes in Table 1, where the nominal significance levels of their test are reported inside the parentheses.

p	Normal			Gamma		
	n			n		
	50	100	200	50	100	200
	Proposed Multiple Test					
50	0.664	0.986	1.000	0.589	0.985	1.000
100	0.973	1.000	1.000	0.962	1.000	1.000
200	1.000	1.000	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000	1.000	1.000
600	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000
	QC's Test					
50	0.215	0.516	0.984	0.207	0.537	0.993
100	0.224	0.569	0.982	0.227	0.564	0.980
200	0.236	0.558	0.980	0.229	0.565	0.976
400	0.252	0.561	0.983	0.236	0.567	0.978
600	0.266	0.582	0.984	0.244	0.561	0.980
1000	0.262	0.548	0.976	0.223	0.551	0.984
	CJ's Test					
50	0.371	0.907	1.000	0.370	0.827	1.000
100	0.280	0.903	1.000	0.286	0.808	1.000
200	0.216	0.819	1.000	0.219	0.792	0.999
400	0.191	0.725	1.000	0.186	0.741	1.000
600	0.161	0.733	1.000	0.140	0.640	1.000
1000	0.134	0.714	1.000	0.119	0.577	1.000

with $\boldsymbol{\theta}_0 = (\theta_1, \theta_2)^T = (1, 20)^T$. Under model (6.2), $D_q(\theta) = (p-q)\theta_1^2 \exp(-2q/\theta_2)$. And the other was the polynomially decay covariances

$$\sigma_{i,j} = \mathrm{I}(i = j) + \theta_1 |i - j|^{-\theta_2} \tag{6.3}$$

with $\boldsymbol{\theta}_0 = (\theta_1, \theta_2)^T = (1, 0.4)^T$. Under model (6.3), $D_q(\theta) = (p - q)\theta_1^2 q^{-2\theta_2}$.

In testing the overall covariance structure along the super-diagonals, let $N = \lfloor 3p/\log(p) \rfloor$ and a multiple testing for $H_{0,q}$:

Table 3
Empirical sizes and powers for testing $H_0 : D_q = (p-q)\theta_1^2 \exp(-2q/\theta_2)$ at 5% significance level. The empirical powers are evaluated for data generated from the polynomial model (6.3).

p	Empirical sizes			Empirical powers		
	n			n		
	50	100	200	50	100	200
50	0.045	0.053	0.055	0.124	0.173	0.138
100	0.049	0.058	0.057	0.365	0.687	1.000
200	0.042	0.050	0.055	0.969	1.000	1.000
400	0.039	0.056	0.057	1.000	1.000	1.000
600	0.032	0.049	0.048	1.000	1.000	1.000
1000	0.033	0.044	0.046	1.000	1.000	1.000

Table 4
Empirical sizes and powers for testing $H_0 : D_q = (p-q)\theta_1^2 q^{-2\theta_2}$ at 5% significance level. The empirical powers are evaluated for data generated from the exponential model (6.2).

p	Empirical sizes			Empirical powers		
	n			n		
	50	100	200	50	100	200
50	0.059	0.061	0.065	0.099	0.114	0.122
100	0.052	0.060	0.059	0.142	0.776	1.000
200	0.050	0.055	0.052	0.873	0.986	1.000
400	0.049	0.057	0.058	1.000	1.000	1.000
600	0.044	0.048	0.054	1.000	1.000	1.000
1000	0.039	0.044	0.047	1.000	1.000	1.000

$D_q = D_q(\boldsymbol{\theta})$, $q = 1, \ldots, N$ was carried out in conjunction with the FDR procedure. We chose $r = 5$ and $\boldsymbol{\theta}$ was estimated by (5.3) using the first five super-diagonals. We tested $H_0 : D_q = (p-q)\theta_1^2 \exp(-2q/\theta_2)$ for data generated from model (6.2) and model (6.3), in order to evaluate the empirical sizes and empirical powers, respectively. The empirical powers of the test were given in Table 3. Then, testing $D_q(\theta) = (p-q)\theta_1^2 q^{-2\theta_2}$ for data generated from model (6.2) and model (6.3) was performed with the empirical sizes and powers reported in Table 4. Average estimates of $\hat{\boldsymbol{\theta}}$ and standard deviations were reported in Table 5, which shows the consistency of $\hat{\boldsymbol{\theta}}$.

Despite relatively lower empirical sizes, Tables 3 and 4 show that the proposed multiple testing maintained high powers for detecting the parametric mis-specification of D_q. It was noted that the multiple tests for the polynomial covariance models when data were generated from the exponential model (6.2) attained higher power than those for testing the exponential covariance model when data were generated from the polynomial model (6.3). This was because the signal to noise ratio $\delta^*_{np,q} = |D_q - D_q(\boldsymbol{\theta}_*)|/\sigma_{q,r}$ for data generated from the exponential model (6.2) was larger than that for model (6.3), since the variance of $\hat{D}_q$ was smaller in model (6.2) than that in model (6.3) due to a quicker decay in $h(q)$ under the exponential model.

7. Empirical study

In this section, we considered modeling the covariance structure among A-shares in Shanghai and Shenzhen Stock Exchanges from January 8, 2010 to October 21, 2014, which consisted of 246 weekly returns. This period was relatively calm after the upheaval of the Great Financial Crisis in 2008 and its after-shocks. We first excluded stocks which had periods of trade suspensions. Then, we selected from the remaining stocks such that its sum of absolute correlation with the other stocks is less than 400, which is the 2/3-quantile of the row sums of the sample correlation matrix of the remaining stocks. This was to produce a more sparse covariance. There were 678 stocks being selected under these criteria. To remove the degree of temporal

Table 5
Average estimates of $\hat{\boldsymbol{\theta}}$ by (5.3) for data generated from Model (6.2) and Model (6.3), respectively. The true parameter values in Model (6.2) and Model (6.3) are, respectively, $(\theta_1, \theta_2) = (1, 20)$ and $(\theta_1, \theta_2) = (1, 0.4)$. Standard deviations are provided inside the parentheses.

n	p	Model (6.2)		Model (6.3)	
		$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
50	50	0.89(0.49)	20.40(7.14)	0.94(0.31)	0.45(0.15)
50	100	0.92(0.44)	19.78(4.32)	0.96(0.27)	0.42(0.09)
50	200	0.96(0.31)	19.88(2.95)	0.99(0.17)	0.41(0.05)
50	400	0.96(0.27)	19.92(2.04)	1.00(0.07)	0.40(0.04)
50	600	0.97(0.27)	20.05(1.67)	1.00(0.05)	0.40(0.03)
50	1000	0.98(0.20)	19.98(1.22)	1.00(0.04)	0.41(0.03)
100	50	0.94(0.35)	20.22(4.83)	0.99(0.16)	0.42(0.12)
100	100	0.94(0.34)	20.05(2.87)	1.00(0.15)	0.41(0.09)
100	200	0.96(0.29)	20.03(1.97)	1.00(0.08)	0.40(0.07)
100	400	0.99(0.11)	20.04(1.32)	1.00(0.07)	0.40(0.04)
100	600	0.98(0.10)	19.97(1.06)	1.00(0.04)	0.40(0.02)
100	1000	0.99(0.09)	19.99(0.84)	1.00(0.03)	0.40(0.02)
200	50	0.96(0.29)	19.96(2.32)	1.00(0.07)	0.41(0.05)
200	100	0.96(0.21)	19.91(1.95)	1.00(0.05)	0.41(0.03)
200	200	0.97(0.20)	19.96(1.14)	1.00(0.04)	0.40(0.02)
200	400	0.99(0.13)	20.03(0.84)	1.00(0.03)	0.40(0.02)
200	600	0.98(0.12)	19.98(0.71)	1.00(0.02)	0.40(0.02)
200	1000	0.99(0.09)	20.02(0.50)	1.00(0.02)	0.40(0.01)

dependence, we considered weekly returns of each stock. We used X_i to denote the standardized weakly returns of the stocks for the ith week, $i = 1, \ldots, 246$. The standardization was done by dividing the weekly returns of stocks by their respective standard deviations.

To achieve a decay of the covariance along the super-diagonals so that (2.2) is satisfied, we arranged the stocks such that more correlated stocks are aligned adjacently using the "Corrgram" algorithm of Friendly (2002). Fig. 1 presents the heatmap of the sample covariance matrix for the weekly returns of the selected 678 stocks. It also plots $\hat{h}(q) = \hat{D}_q/(p-q)$ against q, which shows a decline as q increases.

We first considered testing $H_0 : \Sigma = B_k(\Sigma)$ for Σ being banded. We also implemented Qiu and Chen (QC)'s and Cai and Jiang (CJ)'s bandedness tests to compare with our proposed test. Fig. 2 gives the test statistics and p-values for the two bandedness tests for different bandwidths k. QC's test suggested that Σ was banded with bandwidth $k = 396$, while CJ's test suggested that Σ was banded at the bandwidth $k = 634$. We implemented the proposed multiple testing based on the p-values of individual test for $H_{0,q}$ with the FDR controlled at 5%. Specifically, we tested $H_0 : \Sigma = B_k(\Sigma)$ by testing $H_{0,q}$, $q = k+1, \ldots, 6p/\log(p)$. Our test suggested that Σ was a banded matrix with the bandwidth $k = 660$.

Since the suggested bandwidth for Σ was very large, which indicated that most of the super-diagonals of Σ are not zero, we wanted to gain more knowledge about the specific non-zero super-diagonal structure of the covariance matrix. In this consideration, it was beneficial for us to find out parametric models for Σ. Inspired by the exponentially decay pattern we mentioned before, we tested for a parametric exponential super-diagonal structure:

$$D_q(\boldsymbol{\theta}) = (p-q)\theta_1^2 \exp(-2q/\theta_2). \tag{7.1}$$

The two-dimensional parameter $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ is estimated by the estimator given in (5.3). We chose $r = 5$ and used the first five super-diagonals to estimate $\hat{\boldsymbol{\theta}}$. The estimate was $\hat{\boldsymbol{\theta}} = (0.35, 367.5)^T$. We tested $H_{0,q} : D_q = (p-q)\theta_1^2 \exp(-2q/\theta_2)$, $q = 1, \ldots, 6p/\log(p)$ by implementing the Bonferroni and FDR type multiple testing procedures with the family-wise error rate and the FDR controlled at 5%. Both the Bonferroni and FDR type tests suggested that for q from 1 to 487, there was sufficient statistical support for the exponential decay model (7.1).

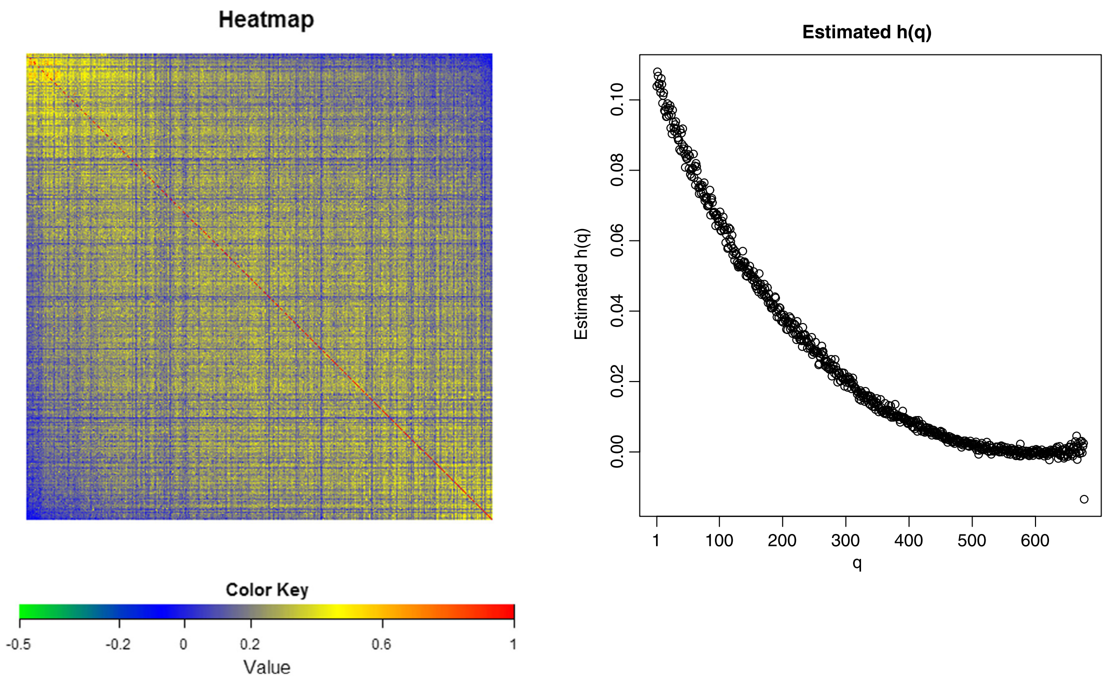


Fig. 1. Heatmaps (left panel) of the covariance matrix of the weekly returns of the 678 stocks, and estimated $\hat{h}(q)$ (right panel) for $q \geq 1$.

We then tested for a polynomial super-diagonal structure such that

$$D_q(\boldsymbol{\theta}) = (p-q)\theta_1^2 q^{-2\theta_2} \tag{7.2}$$

where $\theta_2 \in (0, 1)$. Following the same route for testing covariance model (7.1), the estimated two-dimensional parameter was $\hat{\boldsymbol{\theta}} = (0.46, 0.147)^T$. However, neither the Bonferroni type test nor the FDR test suggested the polynomial decay model (7.2) was adequate to the data. Fig. 3 shows that the exponential covariance model (7.1) can better describe the pattern of $\hat{h}(q)$. In this context, the evidence supporting model (7.1) was quite strong.

The parametric model for the super-diagonals may be used to form an estimator for Σ, say $\Sigma(\hat{\boldsymbol{\theta}})$ whose super-diagonals follow the parametric model. An attraction of using $\Sigma(\hat{\boldsymbol{\theta}})$ is that it is easier to be inverted, whereas the sample covariance matrix is not invertible, due to the fact that $p = 678 > n = 246$. Hence, $\Sigma(\hat{\boldsymbol{\theta}})$ can be used in the selection of optimal portfolio via Markowitz theory. Furthermore, we may test the residual covariance matrix after data prewhitening using statistical methods such as the factor model. All we need to do is to calculate the estimated residuals and then apply our method.

A natural extension of the study would be to consider if the covariance of the stocks is block-diagonal rather than obeying a decay along the super-diagonals. The broad economic sectors of the stocks' main operation may be used to determine the grouping of the stocks, as did in Fan et al. (2015) and Aït-Sahalia and Xiu (2015) when modeling the covariance of residuals. This would require developing a test procedure for block-diagonal covariances, which may be a future research topic.

Appendix A. Technical details

We provide the proofs for the main theorem as well as some lemmas needed.

Lemma 1. *Under Assumptions A1 and A2, for each given q, there exists a positive constant M′ such that* $\max_i \lambda_i(W_q) \leq M'$, *where* $\lambda_i(W_q), i = 1, \dots, p-q$, *are the eigenvalues of* W_q.

Lemma 2. *Under Assumptions* A1–A4, *for* $q = o(p)$ *and any* $i = 1, \dots, n$, *we have,* $\sum_{l_1,l_2,l_3,l_4=1}^{p-q} \mathrm{E}(Y_i^{l_1,l_1+q} Y_i^{l_2,l_2+q} Y_i^{l_3,l_3+q} Y_i^{l_4,l_4+q}) = O\left((p-q)^2\right)$.

The proof for Lemmas 1 and 2 can be found in He and Chen (2016).

Proof of Proposition 1. For $q = 1, \dots, p-1$, define

$$B_{1,q} = \frac{1}{P_n^2} \sum_{l=1}^{p-q} \sum_{i,j}^{*} (X_{i,l}X_{i,l+q})(X_{j,l}X_{j,l+q}),$$

$$B_{2,q} = \frac{1}{P_n^3} \sum_{l=1}^{p-q} \sum_{i,j,k}^{*} X_{i,l}X_{k,l+q}(X_{j,l}X_{j,l+q}), \quad \text{and}$$

$$B_{3,q} = \frac{1}{P_n^4} \sum_{l=1}^{p-q} \sum_{i,j,k,m}^{*} X_{i,l}X_{j,l+q}X_{k,l}X_{m,l+q},$$

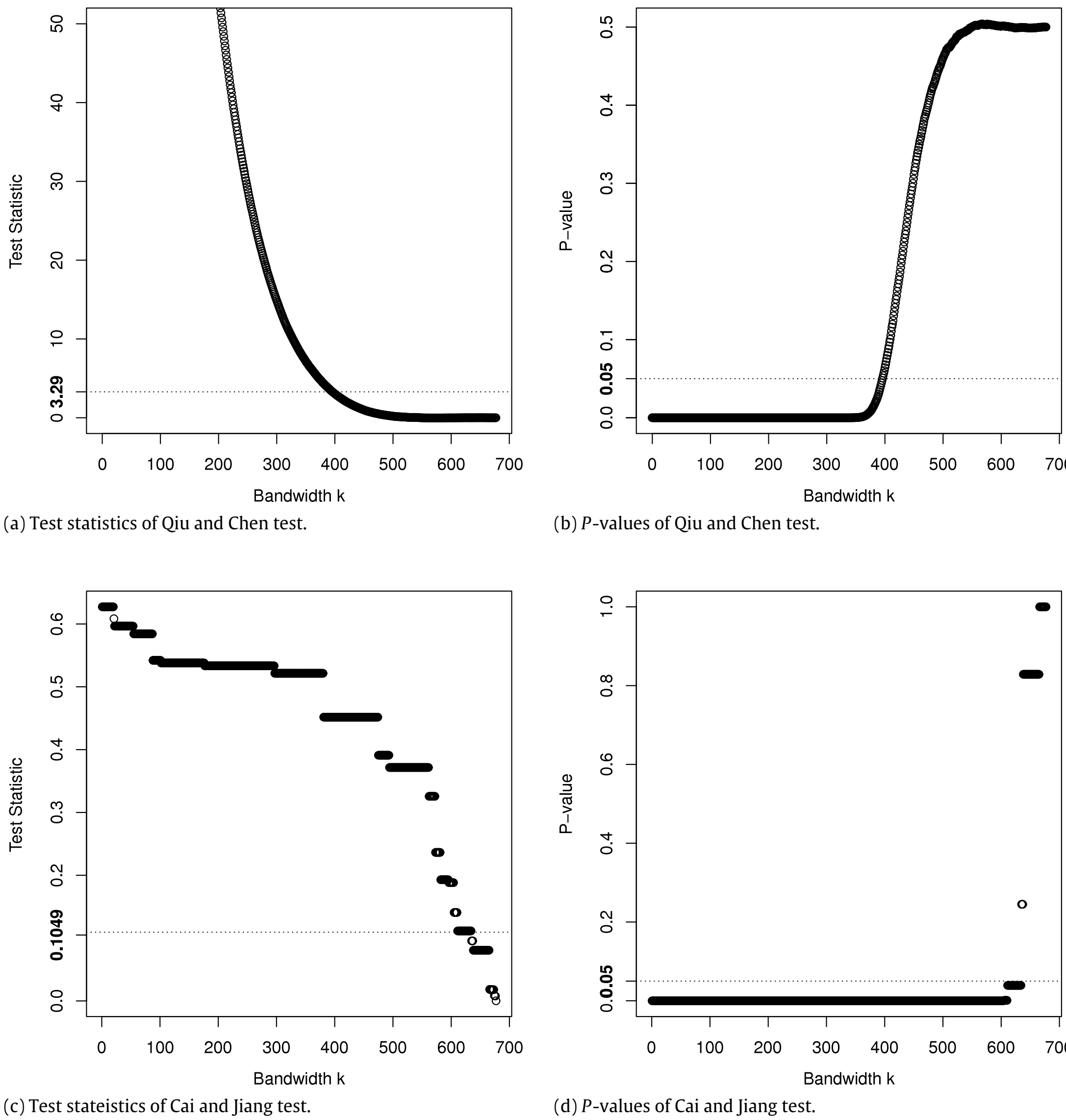
which are respectively U-statistics. Thus $\hat{D}_q = B_{1,q} - 2B_{2,q} + B_{3,q}$. As $\mu = 0$ via the invariance argument, standard derivations show that $\mathrm{E}(B_{1,q}) = \sum_{l=1}^{p-q} \sigma_{l,l+q}^2$, $\mathrm{E}(B_{2,q}) = 0$ and $\mathrm{E}(B_{3,q}) = 0$. Hence, $\mathrm{E}(\hat{D}_q) = D_q$.

In the following, we derive the variance of $B_{l,q}$ for $l = 1, 2, 3$. As $B_{1,q}$ is a U-statistic, we derive the variance of $B_{1,q}$ using the Hoeffding decompositions (Hoeffding, 1948). Zhong and Chen (2011) have proved that the Hoeffding decomposition is still applicable for high-dimensional data when the dimension p diverges. Let $U_i(q) = \left(X_{i,1}X_{i,1+q}, \dots, X_{i,p-q}X_{i,p}\right)^T$, $i = 1, \dots, n$, which are IID random vectors on $\mathbb{R}^{p-q}$, and $H(U_i(q), U_j(q)) = U_i(q)^T U_j(q)$ be a function from $\mathbb{R}^{p-q} \times \mathbb{R}^{p-q}$ to $\mathbb{R}$. Hence, $B_{1,q}$ can be written as

$$B_{1,q} = \frac{1}{\binom{n}{2}} \sum_{C_{n,2}} H(U_{i_1}(q), U_{i_2}(q)),$$

where $\binom{n}{s} = \frac{n!}{(n-s)!s!}$ and $C_{n,s}$ is the set of all distinct combinations of $\{i_1, \dots, i_s\}$ from $\{1, 2, \dots, n\}$.

Let $h_1(u_1) = \mathrm{E}(H(u_1, U_2))$ and $h_2(u_1, u_2) = H(u_1, u_2)$ be the projection of H to lower-dimensional sample spaces and $h =$



(a) Test statistics of Qiu and Chen test.

(b) *P*-values of Qiu and Chen test.

(c) Test stateistics of Cai and Jiang test.

(d) *P*-values of Cai and Jiang test.

Fig. 2. The test statistics (left panels) and P-values (right panels) of QC's test and CJ's test at different bandwidths.

$\mathrm{E}\left(H(U_1, U_2)\right) = D_q$. Denote the mean vector of $U_i(q)$ by $J_q = \left(\sigma_{1,1+q}, \ldots, \sigma_{p-q,p}\right)^T$. Then it is obtained that $h_1(U_i(q)) = J_q^T U_i(q)$ and $h_2(U_i(q), U_j(q)) = U_i(q)^T U_j(q)$. Then,

$$\zeta_1 \hat{=} \mathrm{Var}(h_1(U_i(q))) = J_q^T W_q J_q = \sum_{l_1,l_2=1}^{p-q} \sigma_{l_1,l_1+q}\sigma_{l_2,l_2+q}\omega_{l_1,l_2},$$

where W_q is the covariance matrix of $\vec{Y}_i(q)$ and $\vec{Y}_i(q)$ is defined in Section 2.

Similarly, it can be shown that

$$\zeta_2 \hat{=} \mathrm{Var}(h_2(U_i(q), U_j(q))) = \mathrm{Var}(U_i(q)^T U_j(q))$$
$$= \sum_{l_1,l_2=1}^{p-q} \omega_{l_1,l_2}^2 + 2\sum_{l_1,l_2=1}^{p-q} \sigma_{l_1,l_1+q}\sigma_{l_2,l_2+q}\omega_{l_1,l_2}.$$

According to Hoeffding decomposition,

$$\mathrm{Var}(B_{1,q}) = \binom{n}{2}^{-1} \sum_{c=1}^{2} \binom{2}{c}\binom{n-2}{2-c}\zeta_c$$
$$= \frac{4}{n}\sum_{l_1,l_2=1}^{p-q} \sigma_{l_1,l_1+q}\sigma_{l_2,l_2+q}\omega_{l_1,l_2} + \frac{2}{n(n-1)}\sum_{l_1,l_2=1}^{p-q} \omega_{l_1,l_2}^2. \quad \text{(A.1)}$$

In (A.1), $\mathrm{Var}(B_{1,q})$ consists of two terms. If $D_q = 0$, the first term equals zero and $\mathrm{Var}(B_{1,q}) = \frac{2}{n(n-1)}\sum_{l_1,l_2=1}^{p-q}\omega_{l_1,l_2}^2$. In the general case that $h(q) \asymp a^q$, it can be shown that $\sum_{l_1,l_2=1}^{p-q} \sigma_{l_1,l_1+q}\sigma_{l_2,l_2+q}\omega_{l_1,l_2} \asymp (p-q)a^q$. For the second term of $\mathrm{Var}(B_{1,q})$, we have $\sum_{l_1,l_2=1}^{p-q}\omega_{l_1,l_2}^2 \asymp (p-q)$ according to Lemma 1 and Assumption A4.

In order to obtain the order of $\mathrm{Var}(B_{1,q})$, which depends on q, we consider three different regimes for q:

(i) If $q = o(\log(n))$, the first term of $\mathrm{Var}(B_{1,q})$ is the leading term;
(ii) If $a^q \asymp 1/n$, both terms of $\mathrm{Var}(B_{1,q})$ are of the same order;
(iii) If $\log(n) = o(q)$ and $q = o(p)$, the second term of $\mathrm{Var}(B_{1,q})$ is the leading term.

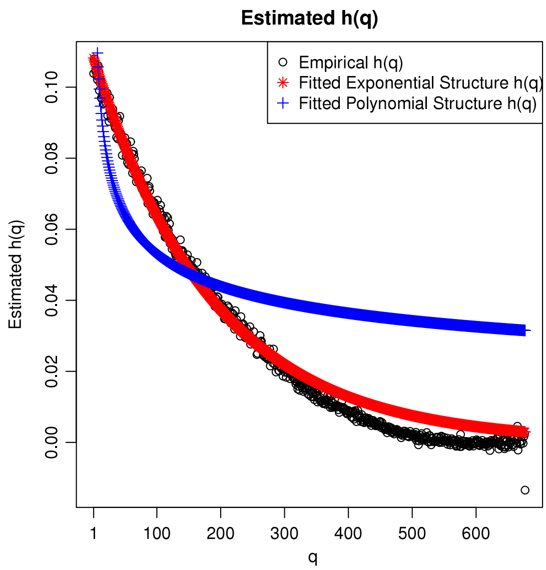


Fig. 3. Estimated average signal $h(q)$ on the super-diagonals for the stocks. The black circles are $\hat{h}(q) = \hat{D}_q/(p-q)$; the red stars are the fitted exponential super-diagonal structure $h(q;\hat{\boldsymbol{\theta}}) = 0.35^2 \exp(-2q/367.5)$ and the blue plus is the fitted polynomial super-diagonal structure $h(q;\hat{\boldsymbol{\theta}}) = 0.46^2 q^{-2\times 0.147}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Hence, we have

$$\mathrm{Var}(B_{1,q}) \asymp \begin{cases} \dfrac{(p-q)a^q}{n}, & \text{for } q = o(\log(n)), \\ \dfrac{(p-q)a^q}{n} \asymp \dfrac{p-q}{n^2} & \text{for } q \text{ such that } a^q \asymp 1/n, \\ \dfrac{p-q}{n^2}, & \text{for } q \text{ such that } \log(n) = o(q) \text{ and } q = o(p). \end{cases} \tag{A.2}$$

Similarly, by the Hoeffding decomposition, we can calculate the variance of $B_{2,q}$ and $B_{3,q}$, respectively. It can be shown that both $\mathrm{Var}(B_{2,q})$ and $\mathrm{Var}(B_{3,q})$ are at a smaller order of $\mathrm{Var}(B_{1,q})$ for $q = o(p)$. See He and Chen (2016) for detailed derivation. By Cauchy–Schwarz inequality, the covariances between $B_{1,q}, B_{2,q}$ and $B_{3,q}$ can be all ignored relative to $\mathrm{Var}(B_{1,q})$. Thus, the leading term of $\mathrm{Var}(\hat{D}_q)$ is given by $\mathrm{Var}(B_{1,q})$ in (A.1). □

Proof of Theorem 1. As $B_{1,q}$ is the leading order term of $\hat{D}_q$, in order to prove the asymptotic normality of $\hat{D}_q$, it is sufficient to prove that, as n and $p \to \infty$,

$$\frac{B_{1,q} - \mathrm{E}(B_{1,q})}{\sqrt{\mathrm{Var}(B_{1,q})}} \xrightarrow{d} N(0,1).$$

We establish it using the martingale central limit theorem (Hall and Heyde, 1980). Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t = \sigma\{X_1, \dots, X_t\}$ be the σ-field generated by $\{X_1, \dots, X_t\}$, and $\mathrm{E}_t(\cdot)$ denote the conditional expectation with respect to $\mathcal{F}_t$ and $\mathrm{E}_0(\cdot) = \mathrm{E}(\cdot)$. Write $L_{q,t} = \mathrm{E}_t(B_{1,q}) - \mathrm{E}_{t-1}(B_{1,q})$ and $\nu_{q,t}^2 = \mathrm{E}_{t-1}(L_{q,t}^2)$. Then $B_{1,q} - \mathrm{E}(B_{1,q}) = \sum_{t=1}^n L_{q,t}$. It can be shown that for every n, $\{(L_{q,t}, \mathcal{F}_t) : t = 1, \dots, n\}$ forms a martingale difference array.

According to Hall and Heyde (1980), it suffices to show that as n and $p \to \infty$,

$$\frac{\sum_{t=1}^n \nu_{q,t}^2}{\mathrm{Var}(B_{1,q})} \xrightarrow{p} 1 \tag{A.3}$$

and

$$\frac{\sum_{t=1}^n \mathrm{E}(L_{q,t}^4)}{\mathrm{Var}^2(B_{1,q})} \to 0. \tag{A.4}$$

To establish (A.3) and (A.4), we first express $L_{q,t}$ and $\nu_{q,t}^2$ explicitly. Specifically,

$$\begin{aligned} L_{q,t} &= (\mathrm{E}_t - \mathrm{E}_{t-1})\, B_{1,q} \\ &= \frac{1}{P_n^2} \sum_{l=1}^{p-q} \sum_{i,j}^{*} (\mathrm{E}_t - \mathrm{E}_{t-1})\,(X_{i,l}X_{i,l+q})(X_{j,l}X_{j,l+q}) \\ &= \frac{2}{n(n-1)} \sum_{i\neq t} \sum_{l=1}^{p-q} (\mathrm{E}_t - \mathrm{E}_{t-1})\,(X_{i,l}X_{i,l+q})(X_{t,l}X_{t,l+q}). \end{aligned}$$

Since $(\mathrm{E}_t - \mathrm{E}_{t-1})\,(X_{i,l}X_{i,l+q})(X_{j,l}X_{j,l+q}) = 0$ for all $j \neq i$ except when $j = t$ or $i = t$.

Note that for $i > t$,

$$\begin{aligned} &\sum_{l=1}^{p-q} (\mathrm{E}_t - \mathrm{E}_{t-1})\,(X_{i,l}X_{i,l+q})(X_{t,l}X_{t,l+q}) \\ &\quad = \sum_{l=1}^{p-q} \sigma_{l,l+q}\left(X_{t,l}X_{t,l+q} - \sigma_{l,l+q}\right), \end{aligned}$$

and for $i \leq t$,

$$\begin{aligned} &\sum_{l=1}^{p-q} (\mathrm{E}_t - \mathrm{E}_{t-1})\,(X_{i,l}X_{i,l+q})(X_{t,l}X_{t,l+q}) \\ &\quad = \sum_{l=1}^{p-q} X_{i,l}X_{i,l+q}\left(X_{t,l}X_{t,l+q} - \sigma_{l,l+q}\right). \end{aligned}$$

Utilizing these two facts and invoking the notation $Y_t^{l_1,l_2} = X_{t,l_1}X_{t,l_2} - \sigma_{l_1,l_2}$, we have

$$\begin{aligned} L_{q,t} &= \frac{1}{P_n^2}\left[\sum_{i\neq t}\left(\sum_{l=1}^{p-q} \sigma_{l,l+q}\left(X_{t,l}X_{t,l+q} - \sigma_{l,l+q}\right)\right)\right. \\ &\quad + \sum_{i=1}^{t-1}\left(\sum_{l=1}^{p-q} X_{i,l}X_{i,l+q}\left(X_{t,l}X_{t,l+q} - \sigma_{l,l+q}\right)\right) \\ &\quad \left. - \sum_{i=1}^{t-1}\left(\sum_{l=1}^{p-q} \sigma_{l,l+q}\left(X_{t,l}X_{t,l+q} - \sigma_{l,l+q}\right)\right)\right] \\ &= \frac{2}{n}\sum_{l=1}^{p-q} \sigma_{l,l+q} Y_t^{l,l+q} + \frac{2}{n(n-1)} \sum_{l=1}^{p-q} Y_t^{l,l+q} Q_{t-1}^{l,l+q}. \end{aligned} \tag{A.5}$$

by rearranging terms in the last two terms of (A.5), where $Q_{t-1}^{l_1,l_2} = \sum_{i=1}^{t-1} Y_i^{l_1,l_2}$. From now on, we will conduct the asymptotic analysis based on $Y_t^{l_1,l_2}$, $l_1, l_2 = 1, \dots, p$, instead of $X_{t,l}$, $l = 1, \dots, p$.

Under the notation $Y_t^{l_1,l_2}$, we have

$$\begin{aligned} \sum_{t=1}^n \nu_{q,t}^2 &= \frac{4}{n^2} \sum_{t=1}^n \sum_{l_1,l_2=1}^{p-q} \sigma_{l_1,l_1+q}\sigma_{l_2,l_2+q}\omega_{l_1,l_2} \\ &\quad + \frac{4}{n^2(n-1)^2} \sum_{t=1}^n \sum_{l_1,l_2=1}^{p-q} \omega_{l_1,l_2} Q_{t-1}^{l_1,l_1+q} Q_{t-1}^{l_2,l_2+q} \\ &\quad + \frac{8}{n^2(n-1)} \sum_{t=1}^n \sum_{l_1,l_2=1}^{p-q} \sigma_{l_1,l_1+q}\omega_{l_1,l_2} Q_{t-1}^{l_2,l_2+q}. \end{aligned} \tag{A.6}$$

It is easy to check that $\mathrm{E}\left(\sum_{t=1}^{n} v_{q,t}^2\right) = \mathrm{Var}(B_{1,q})$. In order to derive (A.3), it is enough to show that $\mathrm{Var}\left(\sum_{t=1}^{n} v_{q,t}^2\right) = o(\mathrm{Var}^2(B_{1,q}))$. Specifically, denote the three parts of $v_{q,t}^2$ in (A.6) by I_1, I_2 and I_3, respectively. We first show that the variances of I_2 and I_3 are at a smaller order of $\mathrm{Var}^2(B_{1,q})$.

Let $V_{i,j} = \sum_{l_1,l_2=1}^{p-q} \omega_{l_1,l_2}\left(Y_i^{l_1,l_1+q}Y_i^{l_2,l_2+q} - \omega_{l_1,l_2}\delta_{i,j}\right)$, where $\delta_{i,j}$ is the indicator function taking the value 1 for $i=j$ and the value 0 otherwise. Since $\mathrm{E}(I_2) = \frac{2}{n(n-1)}\sum_{l_1,l_2=1}^{p-q}\omega_{l_1,l_2}^2$, then,

$$I_2 - \mathrm{E}(I_2) = \sum_{i,j=1}^{n-1} \frac{4(n-\max\{i,j\})}{n^2(n-1)^2} V_{i,j}.$$

Thus,

$$\begin{aligned}\mathrm{Var}\,(I_2) &= \frac{8(2n-1)}{n^3(n-1)^3}\mathrm{Var}(V_{i,i}) \\ &\quad + \frac{8(n-2)}{n^3(n-1)^2}\left(\mathrm{Var}(V_{i,j}) + \mathrm{Cov}(V_{i,j}, V_{j,i})\right) \\ &= O\left(\frac{(p-q)^2}{n^5}\right) + O\left(\frac{(p-q)}{n^4}\right)\end{aligned}$$

since for $i = 1, \ldots, n$, according to Assumption A2 and Lemma 1,

$$\begin{aligned}\mathrm{Var}(V_{i,i}) &= \sum_{l_1,l_2=1}^{p-q}\sum_{l_3,l_4=1}^{p-q}\omega_{l_1,l_2}\omega_{l_3,l_4}\mathrm{E}\left(Y_i^{l_1,l_1+q}Y_i^{l_2,l_2+q}Y_i^{l_3,l_3+q}Y_i^{l_4,l_4+q}\right) \\ &\quad - \sum_{l_1,l_2=1}^{p-q}\sum_{l_3,l_4=1}^{p-q}\omega_{l_1,l_2}\omega_{l_3,l_4}\omega_{l_1,l_2}\omega_{l_3,l_4} \\ &= O\left((p-q)^2\right)\end{aligned}$$

and for $i \neq j$,

$$\mathrm{Var}(V_{i,j}) = \mathrm{Cov}(V_{i,j}, V_{j,i}) = \mathrm{tr}(W_q^4) = O(p-q).$$

Considering the order of $\mathrm{Var}(B_{1,q})$ which is specified in (A.2) in different regimes, we have $\mathrm{Var}\,(I_2) = o\left(\mathrm{Var}^2(B_{1,q})\right)$ for q such that (i) $q = o(\log(n))$, (ii) $a^q \asymp 1/n$ and (iii) $\log(n) = o(q)$ and $q = o(p)$.

Similarly, let $S_i = \sum_{l_1,l_2=1}^{p-q}\sigma_{l_1,l_1+q}\omega_{l_1,l_2}Y_i^{l_2,l_2+q}$. Then, I_3 can be written as

$$I_3 = \sum_{i=1}^{n-1}\frac{8(n-i)}{n^2(n-1)}S_i.$$

Thus, we have $\mathrm{Var}(I_3) = \frac{32(2n-1)}{3n^3(n-1)}\mathrm{Var}(S_i)$, where

$$\begin{aligned}\mathrm{Var}(S_i) &= \frac{1}{(p-q)^2}\sum_{l_1,l_2=1}^{p-q}\sum_{l_3,l_4=1}^{p-q}\sigma_{l_1,l_1+q}\sigma_{l_3,l_3+q}\omega_{l_1,l_2}\omega_{l_3,l_4}\omega_{l_2,l_4}, \\ &= \frac{1}{(p-q)^2}J_q^T W_q^3 J_q = \frac{1}{(p-q)^2}\mathrm{tr}\left(W_q^3 J_q J_q^T\right) \\ &\leq \frac{1}{p-q}\mathrm{tr}(W_q^3)h(q) \leq M^3 a^q,\end{aligned}$$

and the last inequality holds since $\mathrm{tr}(W_q^3) = \sum_{l=1}^{p-q}\lambda_l^3(W_q) \leq M^3(p-q)$.

Therefore, since for q such that (i) $q = o(\log(n))$ and (ii) $a^q \asymp 1/n$, $\mathrm{Var}(B_{1,q}) \asymp (p-q)a^q/n$, we have $\mathrm{Var}(I_3) = o\left(\mathrm{Var}^2(B_{1,q})\right)$. On the other hand, for q such that $\log(n) = o(q)$ and $q = o(p)$, we have $\mathrm{Var}(B_{1,q}) \asymp (p-q)/n^2$ and

$$\frac{\mathrm{Var}\,(I_3)}{\mathrm{Var}^2(B_{1,q})} \leq na^q \to 0.$$

According to Cauchy–Schwarz's inequality, since both $\mathrm{Var}(I_2)$ and $\mathrm{Var}(I_3)$ are smaller order of $\mathrm{Var}^2(B_{1,q})$ for $q = o(p)$, the covariance between I_2 and I_3 is also smaller order of $\mathrm{Var}^2(B_{1,q})$. Thus, this completes the proof of (A.3).

To establish (A.4), write

$$\begin{aligned}L_{q,t} &= \frac{2}{n}\sum_{l=1}^{p-q}\sigma_{l,l+q}Y_t^{l,l+q} \\ &\quad + \frac{2}{n(n-1)}\sum_{l=1}^{p-q}Y_t^{l,l+q}Q_{t-1}^{l,l+q} \hat{=} A_t + B_t, \quad \text{say.}\end{aligned}$$

Notice that

$$\begin{aligned}\sum_{t=1}^{n}\mathrm{E}(L_{q,t}^4) &= \sum_{t=1}^{n}\mathrm{E}(A_t^4) + 4\sum_{t=1}^{n}\mathrm{E}(A_t^3 B_t) \\ &\quad + 6\sum_{t=1}^{n}\mathrm{E}A_t^2B_t^2 + 4\sum_{t=1}^{n}\mathrm{E}(A_tB_t^3) + \sum_{t=1}^{n}\mathrm{E}(B_t^4)\end{aligned} \tag{A.7}$$

According to Lemma 1 and Assumption A4, it can be shown that each term in (A.7) is of smaller order of $\mathrm{Var}^2(B_{1,q})$. Please see He and Chen (2016) for more details. This completes the proof of (A.4). □

Proof of Proposition 2. The ratio consistency of $\hat{V}_{p,q,1}$ and $\hat{V}_{p,q,2}$ is established by calculating their means and variances, respectively. Please see He and Chen (2016) for detailed proof. □

Proof of Theorem 2. Under the alternative $H_{1,q}$, the power of the test is

$$\begin{aligned}\beta_{q,\alpha} &= \Pr\left\{\frac{\hat{D}_q}{\sqrt{2\hat{V}_{p,q,2}/n(n-1)}} > z_{1-\alpha}\,\middle|\, D_q \neq 0\right\} \\ &= \Pr\left(\frac{\hat{D}_q - D_q}{\sigma_{\hat{D}_q}} > z_{1-\alpha}\frac{\sqrt{2\hat{V}_{p,q,2}/n(n-1)}}{\sigma_{\hat{D}_q}} - \delta_{np,q}\,\middle|\, D_q \neq 0\right) \\ &\geq \Pr\left(\frac{\hat{D}_q - D_q}{\sigma_{\hat{D}_q}} > z_{1-\alpha}\sqrt{\frac{\hat{V}_{p,q,2}}{V_{p,q,2}}} - \delta_{np,q}\,\middle|\, D_q \neq 0\right),\end{aligned} \tag{A.8}$$

since $V_{p,q,1} \geq 0$ due to the fact that W_q is nonnegative definite.

According to Proposition 2, for any η, $\Pr(B_\eta) \to 1$ where $B_\eta = \{\hat{V}_{p,q,2} < V_{p,q,2}\,(1+\eta)\}$ for $q = o(p)$, as n and p go to infinity. That is, for any $\varepsilon > 0$, there exist positive integers n_1 and p_1, such that for all $n > n_1$ and $p > p_1$, $\Pr(B_\eta) > 1 - \varepsilon$. Hence, from (A.8), we have

$$\begin{aligned}\beta_{q,\alpha} &\geq \Pr\left(\frac{\hat{D}_q - D_q}{\sigma_{\hat{D}_q}} > z_{1-\alpha}\sqrt{\frac{\hat{V}_{p,q,2}}{V_{p,q,2}}} - \delta_{np,q}, B_\eta\right) \\ &\geq \Pr\left(\frac{\hat{D}_q - D_q}{\sigma_{\hat{D}_q,2}} > z_{1-\alpha}(1+\eta) - \delta_{np,q}\right) - \Pr(B_\eta^c).\end{aligned}$$

Then, from Theorem 1,

$$\begin{aligned}\liminf_{n,p\to\infty}\beta_{q,\alpha} &\geq \liminf_{n,p\to\infty}\Pr\left(\frac{\hat{D}_q - D_q}{\sigma_{\hat{D}_q}} > z_{1-\alpha}(1+\eta) - \delta_{np,q}\right) \\ &\quad - \limsup_{n,p\to\infty}\Pr(B_\eta^c) \\ &\geq 1 - \Phi\left(z_{1-\alpha}(1+\eta) - \liminf_{n,p\to\infty}\delta_{np,q}\right) - \varepsilon.\end{aligned}$$

By letting ε and $\eta \to 0$, we have

$$\liminf_{n,p\to\infty}\beta_{q,\alpha} \geq 1 - \Phi\left(z_{1-\alpha} - \liminf_{n,p\to\infty}\delta_{np,q}\right). \quad \square$$

Proof of Theorem 3. The convergence rate of $\hat{\boldsymbol\theta}$ is established in two steps. The first step in the proof shows that $\hat{\boldsymbol\theta} \xrightarrow{p} \boldsymbol\theta_0$. And the second step takes the Taylor expansion of the first-order condition for $\hat{\boldsymbol\theta}$ and shows that $\sqrt{np}(\hat{\boldsymbol\theta} - \boldsymbol\theta_0) = O_p(1)$.

Step 1(Consistency): From Proposition 1, it is shown that for fixed $\boldsymbol\theta$,

$$\mathrm{E}G_r(\boldsymbol\theta) = \Big((D_1(\boldsymbol\theta_0) - D_1(\boldsymbol\theta))/(p-1), \ldots, (D_r(\boldsymbol\theta_0) - D_r(\boldsymbol\theta))/(p-r)\Big)^T,$$

and $\mathrm{Var}(G_r(\boldsymbol\theta))$ is an $r \times r$ matrix with the (i, j)th element being $\mathrm{Cov}(\hat{D}_i, \hat{D}_j)/(p-i)(p-j)$ defined in (A.20).

Analogous derivation to Proposition 1 shows that for $i, j = 1, \ldots, r$,

$$\mathrm{Cov}(\hat{D}_i, \hat{D}_j) \asymp (p - \max\{i, j\})/n.$$

Thus, we have $\mathrm{Var}(G_r(\boldsymbol\theta)) = O(1/np)$ for any fixed $\boldsymbol\theta$. It is shown that, for any $\boldsymbol\theta \in \Theta$, by Chebyshev's inequality,

$$\begin{aligned}\Pr\left(\|G_r(\boldsymbol\theta) - \mathrm{E}G_r(\boldsymbol\theta)\|_2 > \varepsilon\right) &\le \frac{\mathrm{E}\|G_r(\boldsymbol\theta) - \mathrm{E}G_r(\boldsymbol\theta)\|_2^2}{\varepsilon^2} \\ &= \frac{1}{\varepsilon^2}\mathrm{E}\left(\mathrm{tr}((G_r(\boldsymbol\theta) - \mathrm{E}G_r(\boldsymbol\theta))(G_r(\boldsymbol\theta) - \mathrm{E}G_r(\boldsymbol\theta))^T)\right) \\ &= \frac{1}{\varepsilon^2}\mathrm{tr}\left(\mathrm{Var}(G_r(\boldsymbol\theta))\right) = O\left((np)^{-1}\right)\end{aligned}$$

Thus, $\Pr\left(\|G_r(\boldsymbol\theta) - \mathrm{E}G_r(\boldsymbol\theta)\|_2 > \varepsilon\right) \to 0$ as $n, p \to \infty$, and for any $\boldsymbol\theta \in \Theta$, $G_r(\boldsymbol\theta) \xrightarrow{p} \mathrm{E}G_r(\boldsymbol\theta)$.

It can be shown that the convergence in probability of $G_r(\boldsymbol\theta)$ to $\mathrm{E}G_r(\boldsymbol\theta)$ is uniform in $\boldsymbol\theta$ using analogous derivation to Bosq (1996), see He and Chen (2016) for more details. Hence, for any $\varepsilon > 0$, we have $\Pr\left\{|G_r^T(\boldsymbol\theta_0)G_r(\boldsymbol\theta_0)| > \varepsilon/2\right\} \to 0$ and $\Pr\left\{|G_r^T(\hat{\boldsymbol\theta})G_r(\hat{\boldsymbol\theta}) - \mathrm{E}G_r^T(\hat{\boldsymbol\theta})\mathrm{E}G_r(\hat{\boldsymbol\theta})| > \varepsilon/2\right\} \to 0$. Thus

$$\Pr\left\{|G_r^T(\hat{\boldsymbol\theta})G_r(\hat{\boldsymbol\theta}) - \mathrm{E}G_r^T(\hat{\boldsymbol\theta})\mathrm{E}G_r(\hat{\boldsymbol\theta})| + |G_r^T(\boldsymbol\theta_0)G_r(\boldsymbol\theta_0)| > \varepsilon\right\} \to 0. \tag{A.9}$$

Note that

$$\begin{aligned}&|\mathrm{E}G_r^T(\hat{\boldsymbol\theta})\mathrm{E}G_r(\hat{\boldsymbol\theta})| \\ &\le |G_r^T(\hat{\boldsymbol\theta})G_r(\hat{\boldsymbol\theta}) - \mathrm{E}G_r^T(\hat{\boldsymbol\theta})\mathrm{E}G_r(\hat{\boldsymbol\theta})| + |G_r^T(\hat{\boldsymbol\theta})G_r(\hat{\boldsymbol\theta})| \\ &\le |G_r^T(\hat{\boldsymbol\theta})G_r(\hat{\boldsymbol\theta}) - \mathrm{E}G_r^T(\hat{\boldsymbol\theta})\mathrm{E}G_r(\hat{\boldsymbol\theta})| + |G_r^T(\boldsymbol\theta_0)G_r(\boldsymbol\theta_0)|,\end{aligned} \tag{A.10}$$

where the second inequality holds since $G_r^T(\hat{\boldsymbol\theta})G_r(\hat{\boldsymbol\theta}) \le G_r^T(\boldsymbol\theta_0)G_r(\boldsymbol\theta_0)$. Hence, (A.9) and (A.10) imply that $\Pr\left\{|\mathrm{E}G_r^T(\hat{\boldsymbol\theta})\mathrm{E}G_r(\hat{\boldsymbol\theta})| > \varepsilon\right\} \to 0$.

Under Assumption C1, for any small neighborhood $B(\boldsymbol\theta_0)$ of $\boldsymbol\theta_0$, there exists a $\delta > 0$, such that $|\mathrm{E}G_r^T(\boldsymbol\theta)\mathrm{E}G_r(\boldsymbol\theta)| > \delta$ for any $\boldsymbol\theta \in B^c(\boldsymbol\theta_0)$. Therefore, $\Pr(\hat{\boldsymbol\theta} \in B(\boldsymbol\theta_0)) \to 1$. Since $B(\boldsymbol\theta_0)$ is an arbitrary neighborhood of $\boldsymbol\theta_0$, we have $\hat{\boldsymbol\theta} \xrightarrow{p} \boldsymbol\theta_0$.

Step 2(Convergence Rate): The first order condition for $\hat{\boldsymbol\theta}$ is that

$$\nabla_{\boldsymbol\theta} G_r^T(\hat{\boldsymbol\theta})G_r(\hat{\boldsymbol\theta}) = 0.$$

Let $f(\boldsymbol\theta) = \nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta)G_r(\boldsymbol\theta)$, which is a d-dimensional vector valued function. Take the second order Taylor expansion to $f(\hat{\boldsymbol\theta})$ at $\boldsymbol\theta_0$ so that

$$0 = f(\boldsymbol\theta_0) + \nabla_{\boldsymbol\theta} f(\boldsymbol\theta_0)(\hat{\boldsymbol\theta} - \boldsymbol\theta_0) + \frac{1}{2}\left[I_d \otimes (\hat{\boldsymbol\theta} - \boldsymbol\theta_0)^T\right]\nabla_{\boldsymbol\theta}^2 f(\tilde{\boldsymbol\theta})(\hat{\boldsymbol\theta} - \boldsymbol\theta_0), \tag{A.11}$$

where $\tilde{\boldsymbol\theta}$ is on the line segment between $\hat{\boldsymbol\theta}$ and $\boldsymbol\theta_0$ and $\nabla_{\boldsymbol\theta} f(\boldsymbol\theta_0)$ and $\nabla_{\boldsymbol\theta}^2 f(\boldsymbol\theta_0)$ are the first and second order derivatives of $f(\boldsymbol\theta)$ with respect to $\boldsymbol\theta$ evaluated at $\boldsymbol\theta_0$, which can be explicitly expressed as

$$\nabla_{\boldsymbol\theta} f(\boldsymbol\theta_0) = \nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0) + \sum_{i=1}^r g(\boldsymbol\theta_0; i)\nabla_{\boldsymbol\theta}^2 g(\boldsymbol\theta_0; i)$$

and

$$\nabla_{\boldsymbol\theta}^2 f(\tilde{\boldsymbol\theta}) = \begin{pmatrix}\nabla_{\boldsymbol\theta}^2 f_1(\tilde{\boldsymbol\theta}) \\ \vdots \\ \nabla_{\boldsymbol\theta}^2 f_d(\tilde{\boldsymbol\theta})\end{pmatrix}, \tag{A.12}$$

where $\nabla_{\boldsymbol\theta}^2 f_j(\tilde{\boldsymbol\theta})$ is the Hessian matrix of $f_j(\boldsymbol\theta)$ evaluated at $\tilde{\boldsymbol\theta}$ for $j = 1, \ldots, d$.

It is noted that

$$\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0) = \sum_{i=1}^r \nabla_{\boldsymbol\theta} D_i(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} D_i^T(\boldsymbol\theta_0)/(p-i)^2 \asymp 1 \tag{A.13}$$

according to Assumption C2 (iii). For the second term in the expression of $\nabla_{\boldsymbol\theta} f(\boldsymbol\theta_0)$ in (A.12), since $g(\boldsymbol\theta_0; q) = (\hat{D}_q - D_q(\boldsymbol\theta_0))/(p-q)$, we have

$$\sum_{i=1}^r g(\boldsymbol\theta_0; i)\nabla_{\boldsymbol\theta}^2 g(\boldsymbol\theta_0; i) = \sum_{i=1}^r g(\boldsymbol\theta_0; i)\frac{\nabla_{\boldsymbol\theta}^2 D_i(\boldsymbol\theta_0)}{p-i}.$$

Note that $\mathrm{E}(g(\boldsymbol\theta_0; i)) = 0$ and $\mathrm{Var}(g(\boldsymbol\theta_0; i)) \asymp 1/(n(p-i))$ for $i = 1, \ldots, r$. According to Assumption C2 (iii), $\nabla_{\boldsymbol\theta}^2 D_i(\boldsymbol\theta_0)/(p-i)$ is bounded. Since r is fixed, we have $\sum_{i=1}^r g(\boldsymbol\theta_0; i)\nabla_{\boldsymbol\theta}^2 g(\boldsymbol\theta_0; i) = o_p(1)$. Combining these two terms, $\nabla_{\boldsymbol\theta} f(\boldsymbol\theta_0) = \nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)(1 + o_p(1)) = O(1)$. It can be shown that the remainder term in (A.11) is at higher order than the first order term.

Hence, (A.11) can be written as

$$0 = f(\boldsymbol\theta_0) + \left\{\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)\right\}(\hat{\boldsymbol\theta} - \boldsymbol\theta_0)(1 + o_p(1)).$$

Solving for $\hat{\boldsymbol\theta} - \boldsymbol\theta_0$ and multiplying by $\sqrt{np}$, we get

$$\sqrt{np}(\hat{\boldsymbol\theta} - \boldsymbol\theta_0) = -\left[\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)\right]^{-1} \times \nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\sqrt{np}G_r(\boldsymbol\theta_0)(1 + o_p(1)). \tag{A.14}$$

Since $\mathrm{E}(G_r(\boldsymbol\theta_0)) = 0$, the leading order term has zero mean and its variance

$$\begin{aligned}&\mathrm{Var}\left(\left[\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)\right]^{-1}\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\sqrt{np}G_r(\boldsymbol\theta_0)\right) \\ &= \left[\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)\right]^{-1}\left[np\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\mathrm{Var}(G_r(\boldsymbol\theta_0))\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)\right] \\ &\quad\times \left[\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)\right]^{-1}.\end{aligned} \tag{A.15}$$

According to (A.13), since $\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0) \asymp 1$, we only need to work on $\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\mathrm{Var}(G_r(\boldsymbol\theta_0))\nabla_{\boldsymbol\theta} G_r(\boldsymbol\theta_0)$. Using the expression of $\mathrm{Cov}\left(\hat{D}_i, \hat{D}_j\right)$ in (A.20), we have

$$\begin{aligned}&\nabla_{\boldsymbol\theta} G_r^T(\boldsymbol\theta_0)\mathrm{Var}(G_r(\boldsymbol\theta_0))G_r(\boldsymbol\theta_0) \\ &= \sum_{i,j=1}^r \frac{1}{(p-i)(p-j)}\mathrm{Cov}\left(\hat{D}_i, \hat{D}_j\right)\nabla_{\boldsymbol\theta} D_i(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} D_j^T(\boldsymbol\theta_0) \\ &= \frac{4}{n}\sum_{i,j=1}^r \frac{1}{(p-i)(p-j)} \\ &\quad\times \sum_{l_1=1}^{p-i}\sum_{l_2=1}^{p-j}\sigma_{l_1,l_1+i}\sigma_{l_2,l_2+j}\omega(l_1, l_2; i, j)\nabla_{\boldsymbol\theta} D_i(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta} D_j^T(\boldsymbol\theta_0)\end{aligned}$$

$$+\frac{2}{n(n-1)}\sum_{i,j=1}^{r}\frac{1}{(p-i)(p-j)} \times \sum_{l_1=1}^{p-i}\sum_{l_2=1}^{p-j}\omega^2(l_1,l_2;i,j)\nabla_{\boldsymbol\theta}D_i(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta}D_j^T(\boldsymbol\theta_0). \quad (A.16)$$

It is sufficient to establish the order for the two terms in (A.16). It can be shown that $np\nabla_{\boldsymbol\theta}G_r^T(\boldsymbol\theta_0)\mathrm{Var}(G_r(\boldsymbol\theta_0))G_r(\boldsymbol\theta_0)=O(1)$. Detailed derivation is available in He and Chen (2016). From (A.14) and (A.15), we have $\sqrt{np}(\hat{\boldsymbol\theta}-\boldsymbol\theta_0)=O_p(1)$. This completes the proof for Theorem 3. □

Proof of Theorem 4. Note that

$$\hat D_q - D_q(\hat{\boldsymbol\theta}) = (\hat D_q - D_q(\boldsymbol\theta_0)) - (D_q(\hat{\boldsymbol\theta}) - D_q(\boldsymbol\theta_0)). \quad (A.17)$$

Apply the Mean Value Theorem to $D_q(\hat{\boldsymbol\theta})-D_q(\boldsymbol\theta_0)$,

$$D_q(\hat{\boldsymbol\theta})-D_q(\boldsymbol\theta_0)=\nabla_{\boldsymbol\theta}D_q^T(\boldsymbol\theta_1)(\hat{\boldsymbol\theta}-\boldsymbol\theta_0),$$

where $\boldsymbol\theta_1$ is between $\hat{\boldsymbol\theta}$ and $\boldsymbol\theta_0$. As $\hat{\boldsymbol\theta}\xrightarrow{p}\boldsymbol\theta_0$,

$$D_q(\hat{\boldsymbol\theta})-D_q(\boldsymbol\theta_0)=\nabla_{\boldsymbol\theta}D_q^T(\boldsymbol\theta_0)(\hat{\boldsymbol\theta}-\boldsymbol\theta_0)(1+o_p(1)). \quad (A.18)$$

Combine (A.18) with (A.14),

$$D_q(\hat{\boldsymbol\theta})-D_q(\boldsymbol\theta_0) = \nabla_{\boldsymbol\theta}D_q^T(\boldsymbol\theta_0)\left[\nabla_{\boldsymbol\theta}G_r^T(\boldsymbol\theta_0)\nabla_{\boldsymbol\theta}G_r(\boldsymbol\theta_0)\right]^{-1} \times\nabla_{\boldsymbol\theta}G_r^T(\boldsymbol\theta_0)G_r(\boldsymbol\theta_0)(1+o_p(1)). \quad (A.19)$$

Hence, substituting (A.19) into (A.17), we have the following two scenarios:

(i) if $q=1,\dots,r$, $\hat D_q-D_q(\hat{\boldsymbol\theta})=\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)G_r(\boldsymbol\theta_0)(1+o_p(1))$, where $\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)$ is defined in (5.5); and

(ii) if $q>r$ and $q=o(p)$, $\hat D_q-D_q(\hat{\boldsymbol\theta})=\mathbf{u}_{q,r,2}(\boldsymbol\theta_0)F_{q,r}(\boldsymbol\theta_0)(1+o_p(1))$, where $F_{q,r}(\boldsymbol\theta_0)=\left(G_r^T(\boldsymbol\theta_0),g(\boldsymbol\theta_0;q)\right)^T$ and $\mathbf{u}_{q,r,2}(\boldsymbol\theta_0)$ is defined in (5.6).

Therefore, the asymptotic variance of $\hat D_q-D_q(\hat{\boldsymbol\theta})$ is determined by

$$\sigma_{q,r}^2=\begin{cases}\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)\mathrm{Var}\left(G_r(\boldsymbol\theta_0)\right)\mathbf{u}_{q,r,1}^T(\boldsymbol\theta_0), & \text{if } q=1,\dots,r,\\ \mathbf{u}_{q,r,2}(\boldsymbol\theta_0)\mathrm{Var}\left(F_{q,r}(\boldsymbol\theta_0)\right)\mathbf{u}_{q,r,2}^T(\boldsymbol\theta_0), & \text{if } q>r \text{ and } q=o(p).\end{cases}$$

Thus, it is sufficient to establish the asymptotic normality of $\sigma_{q,r}^{-1}(\hat D_q-D_q(\hat{\boldsymbol\theta}))$. Using the same technique in the proof of Theorem 1, for $q=1,\dots,r$, write

$$G_r(\boldsymbol\theta_0)=\sum_{t=1}^{n}\left(\frac{1}{p-1}L_{1,t},\dots,\frac{1}{p-r}L_{r,t}\right)^T \hat{=} \sum_{t=1}^{n}\xi_{q,t},$$

where $L_{q,t}$ is defined in the proof of Theorem 1. Then, we have $\hat D_q-D_q(\hat{\boldsymbol\theta})=\sum_{t=1}^n\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)\xi_{q,t}$. It can be shown that for every n, $(\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)\xi_{q,t},\mathcal{F}_t)$: $t=1,\dots,n$ forms a martingale difference array.

Using the martingale central limit theorem (Hall and Heyde, 1980), it can be shown that as n and $p\to\infty$, the following two conditions hold:

$$\frac{\sum_{t=1}^{n}\mathrm{E}_{t-1}\left((\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)\xi_{q,t})^2\right)}{\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)\mathrm{Var}\left(G_r(\boldsymbol\theta_0)\right)\mathbf{u}_{q,r,1}^T(\boldsymbol\theta_0)}\xrightarrow{p}1,$$

and

$$\frac{\sum_{t=1}^{n}\mathrm{E}\left((\mathbf{u}_{q,r,1}\xi_{q,t})^4\right)}{\left(\mathbf{u}_{q,r,1}(\boldsymbol\theta_0)\mathrm{Var}\left(G_r(\boldsymbol\theta_0)\right)\mathbf{u}_{q,r,1}^T(\boldsymbol\theta_0)\right)^2}\to 0.$$

Hence, $\sigma_{q,r}^{-1}(\hat D_q-D_q(\hat{\boldsymbol\theta}))\xrightarrow{d}N(0,1)$ for $q=1,\dots,r$. The case for $q>r$ and $q=o(p)$ can be established similarly. This completes the proof of Theorem 4. □

Before proceeding to the proof of Proposition 3, we introduce the estimator for $\sigma_{q,r}^2$. According to the definitions of $\sigma_{q,r}^2$, it suffices to estimate $\mathrm{Var}(G_r(\boldsymbol\theta_0))$ and $\mathrm{Var}(F_{q,r}(\boldsymbol\theta_0))$. Since r is fixed, the covariance matrices of $G_r(\boldsymbol\theta_0)$ and $F_{q,r}(\boldsymbol\theta_0)$ can be estimated element-wisely. Taking regime (i) as an example, write $\mathrm{Var}(G_r(\boldsymbol\theta_0))=(\mathrm{Var}(G_r(\boldsymbol\theta_0))(i,j))_{r\times r}$. Then $(\mathrm{Var}(G_r(\boldsymbol\theta_0)))\,(q,q)=\sigma_{\hat D_q}^2/(p-q)^2$, which can be estimated by

$$(\widehat{\mathrm{Var}(G_r(\boldsymbol\theta_0))})(q,q)=\frac{4}{n(p-q)^2}\hat V_{p,q,1}+\frac{2}{n(n-1)(p-q)^2}\hat V_{p,q,2}$$

as outlined in Section 3. Similarly, for $i\neq j$, the (i,j)th element of $\mathrm{Var}(G_r(\boldsymbol\theta_0))$ can be expressed as

$$\begin{aligned}&(\mathrm{Var}(G_r(\boldsymbol\theta_0)))\,(i,j)\\ &=\frac{4}{n(p-i)(p-j)}\sum_{l_1=1}^{p-i}\sum_{l_2=1}^{p-j}\sigma_{l_1,l_1+i}\sigma_{l_2,l_2+j}\omega(l_1,l_2;i,j)\\ &\quad+\frac{2}{n(n-1)(p-i)(p-j)}\sum_{l_1=1}^{p-i}\sum_{l_2=1}^{p-j}\omega^2(l_1,l_2;i,j)\\ &\hat{=}\frac{4}{n(p-i)(p-j)}C_{i,j,1}+\frac{2}{n(n-1)(p-i)(p-j)}C_{i,j,2},\ \text{say,}\end{aligned} \quad (A.20)$$

where $\omega(l_1,l_2;i,j)=\mathrm{Cov}(Y_t^{l_1,l_1+i},Y_t^{l_2,l_2+j})$. Hence, $C_{i,j,1}$ and $C_{i,j,2}$ can be estimated respectively by

$$\hat C_{i,j,1}=\sum_{l_1=1}^{p-i}\sum_{l_2=1}^{p-j}\sum_{s,t,v}^{*}\frac{1}{P_n^3}(\tilde X_{s,l_1}\tilde X_{s,l_1+i})(\tilde X_{t,l_2}\tilde X_{t,l_2+q}) \times\left(\tilde X_{v,l_1}\tilde X_{v,l_1+i}-\hat\sigma_{l_1,l_1+i}^{(s,t,v)}\right)\left(\tilde X_{v,l_2}\tilde X_{v,l_2+j}-\hat\sigma_{l_2,l_2+q}^{(s,t,v)}\right)$$

and

$$\hat C_{i,j,2}=\frac{1}{P_n^2}\sum_{(s,t)}^{*}\sum_{l_1=1}^{p-i}\sum_{l_2=1}^{p-j}(\tilde X_{s,l_1}\tilde X_{s,l_1+q}-\hat\sigma_{l_1,l_1+i}^{(s,t)})(\tilde X_{s,l_2}\tilde X_{s,l_2+j}-\hat\sigma_{l_2,l_2+j}^{(s,t)}) \times(\tilde X_{t,l_1}\tilde X_{t,l_1+q}-\hat\sigma_{l_1,l_1+i}^{(s,t)})(\tilde X_{t,l_2}\tilde X_{t,l_2+j}-\hat\sigma_{l_2,l_2+j}^{(s,t)}).$$

Therefore, $(\mathrm{Var}(G_r(\boldsymbol\theta_0)))\,(i,j)$ is estimated by

$$(\widehat{\mathrm{Var}(G_r(\boldsymbol\theta_0))})(i,j)=\frac{4\hat C_{i,j,1}}{n(p-i)(p-j)}+\frac{2\hat C_{i,j,2}}{n(n-1)(p-i)(p-j)}.$$

Define $\widehat{\mathrm{Var}(G_r(\boldsymbol\theta_0))}=(\widehat{\mathrm{Var}(G_r(\boldsymbol\theta_0))})(i,j)$ as the matrix consisting of estimators for each elements of $\mathrm{Var}(G_r(\boldsymbol\theta_0))$. Thus,

$$\hat\sigma_{q,r}^2=\mathbf{u}_{q,r,1}(\hat{\boldsymbol\theta})\widehat{\mathrm{Var}(G_r(\boldsymbol\theta_0))}\mathbf{u}_{q,r,1}^T(\hat{\boldsymbol\theta}).$$

In regime (ii), for $q>r$ and $q=o(p)$, write $\mathrm{Var}(F_{q,r}(\boldsymbol\theta_0))$ as

$$\mathrm{Var}(F_{q,r}(\boldsymbol\theta_0))=\begin{pmatrix}\mathrm{Var}(G_r(\boldsymbol\theta_0)), & \mathrm{Cov}(G_r(\boldsymbol\theta_0),g(\boldsymbol\theta_0;q))\\ \mathrm{Cov}^T(G_r(\boldsymbol\theta_0),g(\boldsymbol\theta_0;q)), & \mathrm{Var}(g(\boldsymbol\theta_0;q))\end{pmatrix},$$

where $\mathrm{Cov}(G_r(\boldsymbol\theta_0),g(\boldsymbol\theta_0;q))$ is an $r\times 1$ vector with its ith element being

$$\mathrm{Cov}(g(\boldsymbol\theta_0;i),g(\boldsymbol\theta_0;q))=\frac{4}{n(p-i)(p-q)}C_{i,q,1}+\frac{2}{n(n-1)(p-i)(p-q)}C_{i,q,2}.$$

Similarly, $\mathrm{Cov}(g(\boldsymbol\theta_0;i),g(\boldsymbol\theta_0;q))$ is estimated by $\widehat{\mathrm{Cov}(g(\boldsymbol\theta_0;i),g(\boldsymbol\theta_0;q))}$, which replaces $C_{i,q,1}$ and $C_{i,q,2}$ with $\hat C_{i,q,1}$

and $\hat{C}_{i,q,2}$, respectively. Define an $r \times 1$ vector $\widehat{\mathrm{Cov}(G_r(\boldsymbol{\theta}_0), g(\boldsymbol{\theta}_0; q))}$ whose ith element is $\widehat{\mathrm{Cov}(g(\boldsymbol{\theta}_0; i), g(\boldsymbol{\theta}_0; q))}$. In addition, $\mathrm{Var}(g(\boldsymbol{\theta}_0; q))$ can be estimated by $\widehat{\mathrm{Var}(g(\boldsymbol{\theta}_0; q))} = \frac{4}{n(p-q)^2}\hat{V}_{p,q,1} + \frac{2}{n(n-1)(p-q)^2}\hat{V}_{p,q,2}$. Thus, the estimator for $\mathrm{Var}(F_{q,r}(\boldsymbol{\theta}_0))$ is defined by

$$\widehat{\mathrm{Var}(F_{q,r}(\boldsymbol{\theta}_0))} = \begin{pmatrix} \widehat{\mathrm{Var}(G_r(\boldsymbol{\theta}_0))}, & \widehat{\mathrm{Cov}(G_r(\boldsymbol{\theta}_0), g(\boldsymbol{\theta}_0; q))} \\ \widehat{\mathrm{Cov}^T(G_r(\boldsymbol{\theta}_0), g(\boldsymbol{\theta}_0; q))}, & \widehat{\mathrm{Var}(g(\boldsymbol{\theta}_0; q))} \end{pmatrix},$$

and then

$$\hat{\sigma}_{q,r}^2 = \mathbf{u}_{q,r,2}(\hat{\boldsymbol{\theta}})\widehat{\mathrm{Var}(F_{q,r}(\boldsymbol{\theta}_0))}\mathbf{u}_{q,r,2}^T(\hat{\boldsymbol{\theta}}).$$

Proof of Proposition 3. The ratio consistency of $\hat{C}_{i,j,1}$ and $\hat{C}_{i,j,2}$ is established via analogous derivation to the proof of Proposition 2 by calculating the means and variances of $\hat{C}_{i,j,1}$ and $\hat{C}_{i,j,2}$, respectively. As r is fixed, $\widehat{\mathrm{Var}(G_r(\boldsymbol{\theta}_0))}$ and $\widehat{\mathrm{Var}(F_{q,r}(\boldsymbol{\theta}_0))}$ are both consistent to their population counterparts. On the other hand, since $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, we have $\mathbf{u}_{q,r,1}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{u}_{q,r,1}(\boldsymbol{\theta}_0)$ and $\mathbf{u}_{q,r,2}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{u}_{q,r,2}(\boldsymbol{\theta}_0)$. Combining these two points, we have $\hat{\sigma}_{q,r}^2 \xrightarrow{p} \sigma_{q,r}^2$. This completes the proof for Proposition 3. □

Proof of Theorem 5. The proof parallels to the proof of Theorem 3. We only show the consistency of $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_*$ under H_1. In the proof of Theorem 3, we have shown that $G_r(\boldsymbol{\theta}) \xrightarrow{p} \mathrm{E}G_r(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}$. Hence, for any $\varepsilon > 0$, we have

$$\Pr\{|G_r^T(\hat{\boldsymbol{\theta}})G_r(\hat{\boldsymbol{\theta}}) - \mathrm{E}G_r^T(\hat{\boldsymbol{\theta}})\mathrm{E}G_r(\hat{\boldsymbol{\theta}})| + |G_r^T(\boldsymbol{\theta}_*)G_r(\boldsymbol{\theta}_*) - \mathrm{E}G_r^T(\boldsymbol{\theta}_*)\mathrm{E}G_r(\boldsymbol{\theta}_*)| > \varepsilon.\}.$$

Note that

$$\begin{aligned} &|\mathrm{E}G_r^T(\hat{\boldsymbol{\theta}})\mathrm{E}G_r(\hat{\boldsymbol{\theta}}) - \mathrm{E}G_r^T(\boldsymbol{\theta}_*)\mathrm{E}G_r(\boldsymbol{\theta}_*)| \\ &\quad \le |G_r^T(\hat{\boldsymbol{\theta}})G_r(\hat{\boldsymbol{\theta}}) - \mathrm{E}G_r^T(\hat{\boldsymbol{\theta}})\mathrm{E}G_r(\hat{\boldsymbol{\theta}})| \\ &\qquad + |G_r^T(\hat{\boldsymbol{\theta}})G_r(\hat{\boldsymbol{\theta}}) - \mathrm{E}G_r^T(\boldsymbol{\theta}_*)\mathrm{E}G_r(\boldsymbol{\theta}_*)| \\ &\quad \le |G_r^T(\hat{\boldsymbol{\theta}})G_r(\hat{\boldsymbol{\theta}}) - \mathrm{E}G_r^T(\hat{\boldsymbol{\theta}})\mathrm{E}G_r(\hat{\boldsymbol{\theta}})| \\ &\qquad + |G_r^T(\boldsymbol{\theta}_*)G_r(\boldsymbol{\theta}_*) - \mathrm{E}G_r^T(\boldsymbol{\theta}_*)\mathrm{E}G_r(\boldsymbol{\theta}_*)|, \end{aligned}$$

where the second inequality holds since $G_r^T(\hat{\boldsymbol{\theta}})G_r(\hat{\boldsymbol{\theta}}) \le G_r^T(\boldsymbol{\theta}_*)G_r(\boldsymbol{\theta}_*)$. Thus, it implies that $\Pr\{|\mathrm{E}G_r^T(\hat{\boldsymbol{\theta}})\mathrm{E}G_r(\hat{\boldsymbol{\theta}}) - \mathrm{E}G_r^T(\boldsymbol{\theta}_*)\mathrm{E}G_r(\boldsymbol{\theta}_*)| > \varepsilon\} \to 0$.

However, under Assumption C3, for any small neighborhood $B(\boldsymbol{\theta}_*)$ of $\boldsymbol{\theta}_*$, there exists a $\delta > 0$, such that $|\mathrm{E}G_r^T(\boldsymbol{\theta})\mathrm{E}G_r(\boldsymbol{\theta}) - \mathrm{E}G_r^T(\boldsymbol{\theta}_*)\mathrm{E}G_r(\boldsymbol{\theta}_*)| > \delta$ for any $\boldsymbol{\theta} \in B^c(\boldsymbol{\theta}_*)$. Therefore, $\Pr(\hat{\boldsymbol{\theta}} \in B(\boldsymbol{\theta}_*)) \to 1$. Since $B(\boldsymbol{\theta}_*)$ is an arbitrary neighborhood of $\boldsymbol{\theta}_*$, we have $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_*$.

The convergence rate is established similarly to Step 2 in the proof of Theorem 3 by substituting $\boldsymbol{\theta}_0$ by $\boldsymbol{\theta}_*$. This completes the proof. □

Proof of Theorem 6. For $q = o(p)$, the rejection probability of the qth individual test under $H_{1,q} : D_q \neq D_q(\boldsymbol{\theta})$ is

$$\begin{aligned} \beta_{q,\alpha} &= \Pr\left(\frac{|\hat{D}_q - D_q(\hat{\boldsymbol{\theta}})|}{\hat{\sigma}_{q,r}} > z_{1-\alpha/2}\middle| H_{1,q}\right) \\ &= \Pr\left(\frac{\hat{D}_q - D_q}{\sigma_{q,r}} > z_{1-\alpha/2}\frac{\hat{\sigma}_{q,r}}{\sigma_{q,r}} - \frac{D_q - D_q(\boldsymbol{\theta}_*)}{\sigma_{q,r}} - \frac{D_q(\boldsymbol{\theta}_*) - D_q(\hat{\boldsymbol{\theta}})}{\sigma_{q,r}}\middle| H_{1,q}\right) \\ &\quad + \Pr\left(\frac{\hat{D}_q - D_q}{\sigma_{q,r}} < z_{\alpha/2}\frac{\hat{\sigma}_{q,r}}{\sigma_{q,r}} - \frac{D_q - D_q(\boldsymbol{\theta}_*)}{\sigma_{q,r}} - \frac{D_q(\boldsymbol{\theta}_*) - D_q(\hat{\boldsymbol{\theta}})}{\sigma_{q,r}}\middle| H_{1,q}\right) \\ &\doteq \beta_{q,\alpha,1} + \beta_{q,\alpha,2}. \end{aligned}$$

Since $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_*$ under H_1, we have $\sigma_{q,r}^{-1}|D_q(\boldsymbol{\theta}_*) - D_q(\hat{\boldsymbol{\theta}})| = o_p(\delta_{np,q}^*)$ by the mapping theorem. Similar to the proof of Theorem 2, when $\delta_{np,q}^* \to \infty$, if $D_q - D_q(\boldsymbol{\theta}_*) > 0$, we have

$$\liminf_{n,p\to\infty} \beta_{q,\alpha,1} \ge 1 - \Phi\left(z_{1-\alpha/2} - \liminf_{n,p\to\infty} \delta_{np,q}^*\right) \to 1$$

and

$$\liminf_{n,p\to\infty} \beta_{q,\alpha,2} \ge \Phi\left(z_{\alpha/2} - \liminf_{n,p\to\infty} \delta_{np,q}^*\right) \to 0.$$

Otherwise, if $D_q - D_q(\boldsymbol{\theta}_*) < 0$ and $\delta_{np,q}^* \to \infty$, then

$$\liminf_{n,p\to\infty} \beta_{q,\alpha,1} \ge 1 - \Phi\left(z_{1-\alpha/2} + \liminf_{n,p\to\infty} \delta_{np,q}^*\right) \to 0$$

and

$$\liminf_{n,p\to\infty} \beta_{q,\alpha,2} \ge \Phi\left(z_{\alpha/2} + \liminf_{n,p\to\infty} \delta_{np,q}^*\right) \to 1.$$

Thus, if $\delta_{np,q}^* \to \infty$, we have $\beta_{q,\alpha} = \beta_{q,\alpha,1} + \beta_{q,\alpha,2} \to 1$. □

References

Aït-Sahalia, Y., Xiu, D., 2015. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. Technical Report. University of Chicago.

Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. Ann. Statist. 40, 436–465.

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

Bai, Z., Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statist. Sinica 6, 311–329.

Baltagi, B.H., Song, S.H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. J. Econometrics 117, 123–150.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 289–300.

Bickel, P.J., Levina, E., 2008a. Covariance regularization by thresholding. Ann. Statist. 36, 2577–2604.

Bickel, P.J., Levina, E., 2008b. Regularized estimation of large covariance matrices. Ann. Statist. 36, 199–227.

Bosq, D., 1996. Nonparametric Statistics for Stochastic Processes. Springer, New York.

Cai, T.T., Jiang, T., 2011. Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. Ann. Statist. 39, 1496–1525.

Cai, T.T., Zhang, C.H., Zhou, H.H., 2010. Optimal rates of convergence for covariance matrix estimation. Ann. Statist. 38, 2118–2144.

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

Chen, S.X., Zhang, L.X., Zhong, P.S., 2010. Tests for high-dimensional covariance matrices. J. Amer. Statist. Assoc. 105, 810–819.

Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.

Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. J. Econometrics 147, 186–197.

Fan, J., Furger, A., Xiu, D., 2015. Incorporating global industrial classification standard into portfolio allocation: a simple factor-based large covariance matrix estimator with high frequency data. J. Bus. & Econom. Statist. (in press).

Friendly, M., 2002. Corrgrams: exploratory displays for correlation matrices. Amer. Statist. 56, 316–324.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. Academic press, New York.

He, J., Chen, S.X., 2016. Testing super-diagonal structure in high dimensional covariance matrices. Technical Report. The Center for Statistical Science, Peking University.

Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. Ann. Math. Stat. 19, 293–325.

Kapoor, M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. J. Econometrics 140, 97–130.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: inference for the number of factors. Ann. Statist. 40, 694–726.
Lam, C., Yao, Q., Bathia, N., 2011. Estimation of latent factors for high-dimensional time series. Biometrika 98, 901–918.
Ledoit, O., Wolf, M., 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann. Statist. 30, 1081–1102.
Lee, L., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. J. Econometrics 154, 165–185.
Li, J., Chen, S.X., 2012. Two sample tests for high-dimensional covariance matrices. Ann. Statist. 40, 908–940.
Markowitz, H., 1952. Portfolio selection. J. Financ. 7, 77–91.
Qiu, Y., Chen, S.X., 2012. Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. Ann. Statist. 40, 1285–1314.
Rodríguez, J., Bárdossy, A., 2014. Multivariate interactions modeling through their manifestations: low dimensional model building via the cumulant generating function, arXiv preprint arXiv:1406.2815.
Rothman, A.J., Levina, E., Zhu, J., 2010. A new approach to cholesky-based covariance regularization in high dimensions. Biometrika 97, 539–550.
Schott, J.R., 2005. Testing for complete independence in high dimensions. Biometrika 92, 951–956.
Srivastava, M.S., 2005. Some tests concerning the covariance matrix in high dimensional data. J. Japan Statist. Soc. 35, 251–272.
Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis. Technical Report No. w11467. National Bureau of Economic Research.
Wu, W.B., Pourahmadi, M., 2003. Nonparametric estimation of large covariance matrices of longitudinal data. Biometrika 90, 831–844.
Zhong, P.S., Chen, S.X., 2011. Tests for high-dimensional regression coefficients with factorial designs. J. Amer. Statist. Assoc. 106, 260–274.