Estimation of the Kronecker Covariance Model by Partial Means and Quadratic Form*

Oliver B. Linton† Haihan Tang‡
University of Cambridge Fudan University

June 24, 2019

Abstract

We propose two new estimators of the Kronecker product model of the covariance matrix. We show that these estimators have good properties in the large dimensional case where \( n \) is large relative to \( T \). In particular, the partial means estimator is consistent in a relative Frobenius norm sense provided \( \log^{3} n/T \to 0 \), while the quadratic form estimator is consistent in a relative Frobenius norm sense provided \( \log^{3} n/(nT) \to 0 \). We obtain the limiting distribution of a Lagrange multiplier (LM) test of the hypothesis of zero mean vector. We show that this test performs well in finite sample situations both when the Kronecker product model is true, but also in some cases where it is not true.

Some key words: Covariance matrix; Kronecker product; Lagrange Multiplier test; Partial means; Quadratic form; Wald statistic

1 Introduction

Covariance matrices are of great importance in many fields. In finance, they are a key element in portfolio choice and risk management (Markowitz (1952)). In psychology, scholars have long assumed that some observed variables are related to the key unobserved traits through a factor model, and then use the covariance matrix of the observed variables to deduce properties of the latent traits. In econometrics, they often appear in test statistics representing the generalized uncertainty about a vector of parameter estimates. Anderson (1984) is a classic statistical reference that studies the estimation of covariance matrices and hypotheses testing about them in the low dimensional case (i.e., the dimension of the covariance matrix, \( n \), is small compared with the sample size \( T \)). We consider the problem of estimating a large covariance matrix where we impose a model structure. In particular, we consider the Kronecker product model.

The Kronecker product model has been previously considered in the psychometric literature (Campbell and O’Connell (1967), Swain (1975), Cudeck (1988), Verhees and Wansbeek (1990) etc.). There is also considerable recent work on multiarray data (Hoff (2011), Hoff (2015), and Hoff (2016)). In the spatial literature, there are a number of studies that consider a Kronecker product model for the correlation matrix of a random field (Loh and Lam (2000)). Robinson (1998) and Hidalgo and Schafgans (2017) exploited separable error covariance matrix structures to develop inference methods without the need for smoothing.

*We thank participants at the Celebration of Peter Phillips’ Forty Years at Yale held at Yale University in October 19th-20th, 2018, and at the CUHK Workshop on Econometrics for helpful comments. We thank Liang Jiang, Chen Wang for useful discussions.

†Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk. Thanks to the Cambridge INET and the Keynes Fund for financial support.

‡Corresponding author. Fanhai International School of Finance and School of Economics, Fudan University, 220 Handan Road, Yangpu District, Shanghai, 200433, China. Email: hhtang@fudan.edu.cn.
These literatures have focussed on the classical case where the dimension $n$ is fixed while the sample size $T$ increases. In fact, the Kronecker product model has a big advantage in the large dimensional case because it leads to substantial dimension reduction even though it need not be sparse in the sense of (2.1) of Fan, Liao, and Liu (2016). Hafner, Linton, and Tang (2018) developed methodology for the large dimensional case using the joint asymptotic framework developed by Phillips and Moon (1999). They showed that the matrix logarithm of the covariance or correlation matrix is a sparse matrix (with $O(\log n)$ unknown quantities) and the logarithm operator converts the multiplicative Kronecker product structure into an additive one. Therefore, the logarithm of the covariance or correlation matrix is a linear function of a much smaller vector of unknown quantities. They used this to develop a closed-form estimator. They established consistency and provided a central limit theorem (CLT). However, their results required strong, albeit sufficient not necessary, conditions; in particular they obtained Frobenius norm consistency of the estimator under the condition that at least $n/T \to 0$, which is very restrictive.

There are many approaches to covariance and precision matrix estimation in the large dimensional case where $n$ can be larger than $T$; see, e.g., Ledoit and Wolf (2003), Bickel and Levina (2008), Fan, Fan, and Lv (2008), Ledoit and Wolf (2012) Fan, Liao, and Mincheva (2013), and Ledoit and Wolf (2015). Fan et al. (2016) give an excellent account of the recent developments in the theory and practice of estimating large dimensional covariance matrices. The usual approach is either to impose some sparsity on the covariance matrix, meaning that many elements of the covariance matrix are assumed to be zero or small, thereby reducing the number of parameters to be estimated, or to “shrink” towards a sparse matrix, or to use some model such as the factor model, to reduce dimension. Most of this literature assumes i.i.d. data. Typically, these methods achieve the average Frobenius norm consistency provided $s \log n/T \to 0$, where $s$ is some sparsity index (e.g., see Bickel and Levina (2008) Theorem 2 with $q = 0$).\footnote{The average Frobenius norm means dividing a Frobenius norm by $\sqrt{n}$, while the relative Frobenius norm means dividing a Frobenius norm by the Frobenius norm of the target matrix, say, the unknown covariance matrix. These two concepts are similar, but not exactly the same.}

In this article, we also consider the large dimensional case where $n$ is possibly larger than the sample size $T$ and the Kronecker product structure holds. We propose two new classes of covariance matrix estimators called the partial means class and the quadratic form class based on the Kronecker product model. Our estimators are averages of elements of the sample covariance matrix, so we obtain rate improvements by the averaging procedure. In particular, we show that the partial means estimator is consistent in the relative Frobenius norm provided $\log^3 n/T \to 0$, which is comparable to the results in the literature achieved based on sparsity assumptions or shrinkage methods. The quadratic form class of estimators can achieve even faster convergence: Under a cross-sectional weak dependence condition, they can achieve the relative Frobenius norm consistency provided $\log^3 n/(nT) \to 0$. Both these methods automatically produce symmetric and positive definite covariance matrix estimators unlike some of the sparsifying methods considered in Fan et al. (2016).

We apply our methodology to a concrete testing problem; we consider the hypothesis of whether a large dimensional mean vector is zero or not. We define the Lagrange multiplier (LM) test of the hypothesis based on our estimated inverse covariance matrix and establish its asymptotic distribution under the null hypothesis. We compare our estimation and testing methods with the Ledoit and Wolf (2004)’s linear shrinkage estimator and the Ledoit and Wolf (2017)’s direct nonlinear shrinkage estimator in simulation experiments. We find that our methods perform very well in moderate sized samples. In fact, they also work well even in situations where the Kronecker product model does not hold.\footnote{Hafner et al. (2018) established a closest approximation property of the Kronecker product model, and some of the results in this article continue to hold even when the Kronecker product model is not true, provided that the target matrix is taken as the Kronecker product matrix which is closest to truth.}

Peter Phillips has made immense contributions to multivariate analysis. His correction of the
textbook presentation of the characteristic function of the F-statistic (Phillips (1982)) lead to a fundamental reappraisal of the properties of this everywhere used test. Maasoumi and Phillips (1982) likewise corrected some fundamental errors in instrumental variable theory. Phillips (1985) and Phillips (1986) respectively, the exact distribution of the SUR estimator and the Wald statistic for testing linear restrictions in a multivariate linear model under Gaussianity. To obtain these results, he developed the theory of fractional matrix calculus, which in itself is a fundamental contribution. Unfortunately we are unable to rely on these results since he required $n < T$. Peter Phillips later worked a lot on multivariate nonstationary time series, and transformed that literature too. There is now a developing literature that is extending that theory to the large dimensional case (Onatski and Wang (2018)).

The rest of the article is structured as follows. In Section 2 we discuss the model and identification. In Section 3.1 we define the partial means class of estimators, while in Section 3.2 we define the quadratic form class of estimators. In Section 4 we give the large sample properties of the estimators. In Section 5 we define test statistics including a Lagrange multiplier statistic. Section 6 conducts a small simulation study comparing our approach with Ledoit and Wolf estimators. Section 7 concludes. The major proofs are put in Appendix while auxiliary lemmas and theorems are in Section B.

1.1 Notation

Let $A$ be an $m \times n$ matrix. Let vec $A$ denote the vector obtained by stacking the columns of $A$ one underneath the other. The commutation matrix $K_{m,n}$ is an $mn \times mn$ orthogonal matrix which translates vec $A$ to vec$(A^t)$, i.e., vec$(A^t) = K_{m,n}$ vec$(A)$. If $A$ is a symmetric $n \times n$ matrix, its $(n-1)/2$ supradiagonal elements are redundant in the sense that they can be deduced from symmetry. If we eliminate these redundant elements from vec $A$, we obtain a new $n(n+1)/2 \times 1$ vector, denoted vech $A$. They are related by the full-column-rank, $n^2 \times n(n+1)/2$ duplication matrix $D_n$: vec $A = D_n$ vech $A$. Conversely, vec $A = D_n^+$ vec $A$, where $D_n^+$ is $n(n+1)/2 \times n^2$ and the Moore-Penrose generalized inverse of $D_n$. In particular, $D_n^+ = (D_n^tD_n)^{-1}D_n^t$ because $D_n$ is full-column-rank.

For $x \in \mathbb{R}^n$, let $\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2}$ and $\|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$ denote the Euclidean ($\ell_2$) norm and the element-wise maximum ($\ell_\infty$) norm, respectively. Let $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denote the maximum and minimum eigenvalues of some real symmetric matrix, respectively. For any real $m \times n$ matrix $A = (a_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$, let $\|A\|_F := \|\text{vec}(A^tA)\|^{1/2} = \|\text{tr}(A^tA)\|^{1/2} = \|\text{vec}A\|_2$, $\|A\|_1 := \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|$, $\|A\|_{\ell_2} := \max_{\|x\|_2 = 1} \|Ax\|_2 = \sqrt{\lambda_{\max}(A^tA)}$, $\|A\|_{\ell_1} := \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{i,j}|$, and $\|A\|_{\ell_\infty} := \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{i,j}|$ denote the Frobenius ($\ell_2$) norm, $\ell_1$ norm, and spectral norm ($\ell_2$-operator norm), maximum column sum matrix norm ($\ell_\infty$-operator norm) of $A$, respectively. Note that $\| \cdot \|_\infty$ can also be applied to matrix $A$, i.e., $\|A\|_\infty = \max_{1 \leq i \leq m, 1 \leq j \leq n} |a_{i,j}|$; however $\| \cdot \|_\infty$ is not a matrix norm so it does not have the submultiplicative property of a matrix norm.

Landau (order) notation in this article, unless otherwise stated, should be interpreted in the sense that $n,T \to \infty$ simultaneously.

2 The Model and Identification

Assume that an $n$-dimensional random vector $y_t$ satisfies $\mu := \mathbb{E}y_t$ and $\Sigma := \mathbb{E}[(y_t - \mu)(y_t - \mu)^t]$. Let $n = n_1 \times \cdots \times n_v$, where $n_j \in \mathbb{Z}$ and $n_j \geq 2$ for $j = 1, \ldots, v$. We suppose that

$$\Sigma = \sigma^2 \times \Sigma_1 \otimes \cdots \otimes \Sigma_v, \quad (2.1)$$

where $\Sigma_j$ is an $n_j \times n_j$ unknown covariance matrix that satisfies $\text{tr}(\Sigma_j) = n_j$ for $j = 1, \ldots, v$, and $0 < \sigma^2 < \infty$ is a scalar parameter. For each $j$, $\Sigma_j$ contains $n_j(n_j+1)/2 - 1$ (unrestricted) parameters. In total, model (2.1) contains $\sum_{j=1}^{v} n_j(n_j+1)/2 - (v-1)$ parameters. This model
is essentially the same as considered in Hafner et al. (2018) except that we make a different identifying restriction. The implied form for $\Sigma^{-1}$ is $\Sigma^{-1} = \sigma^{-2} \times \Sigma_1^{-1} \otimes \cdots \otimes \Sigma_v^{-1}$.

We show that model (2.1) is indeed identified. First, the parameter $\sigma$ is identified because

$$\text{tr}(\Sigma) = \sigma^2 \times \text{tr}(\Sigma_1 \otimes \cdots \otimes \Sigma_v) = \sigma^2 \times \text{tr}(\Sigma_1) \times \cdots \times \text{tr}(\Sigma_v) = \sigma^2 n,$$

whence we have $\sigma^2 = \text{tr}(\Sigma)/n$. We next consider the identification of the remaining parameters. For each $j = 1, \ldots, v$, define the $n_j \times n_j$ matrix

$$A_j := c_{j}^T \otimes \cdots \otimes c_{j-1}^T \otimes I_{j} \otimes c_{j+1}^T \otimes \cdots \otimes c_{n}^T,$$

where $c_j$ is an $n_j \times 1$ vector for $j = 1, \ldots, v$ such that $\{A_j\}_{j=1}^v$ are of full rank. Define the $n_j \times n_j$ matrix

$$\Omega_j := E \left[ A_j(y_t - \mu)(y_t - \mu)^T A_j^T \right] = A_j \Sigma A_j^T = \sigma^2 A_j (\Sigma_1 \otimes \cdots \otimes \Sigma_v) A_j^T = \sigma^2 \left( c_1^T \otimes \cdots \otimes c_{j-1}^T \otimes I_{j} \otimes c_{j+1}^T \otimes \cdots \otimes c_{n}^T \right) \left( \Sigma_1 \otimes \cdots \otimes \Sigma_v \right) \left( c_1 \otimes \cdots \otimes c_{j-1} \otimes I_{j} \otimes c_{j+1} \otimes \cdots \otimes c_{n} \right)$$

$$= \sigma^2 \times \omega_1 \times \cdots \times \omega_{j-1} \times \omega_{j+1} \times \cdots \times \omega_v \times \Sigma_j =: \sigma_{j}^2 \times \Sigma_j,$$

where $\omega_{\ell} := c_{\ell}^T \Sigma c_{\ell}$ for $\ell = 1, \ldots, v$, and $\sigma_{j}^2 := \sigma^2 \times \omega_1 \times \cdots \times \omega_{j-1} \times \omega_{j+1} \times \cdots \times \omega_v$. Given the $n_j \times n_j$ matrix $\Omega_j$ we identify $\Sigma_j$ using the trace restriction, specifically, $\sigma_{j}^2 = \text{tr}(\Omega_j)/n_j$ and

$$\Sigma_j = \frac{\Omega_j}{\text{tr}(\Omega_j)/n_j}.$$ 

We call $A_j : \mathbb{R}^n \to \mathbb{R}^{n_j}$ the \textit{partial means operator matrix}, because it averages out the directions not of interest, similar to Newey (1994) and Linton and Nielsen (1995).

We consider a second identification strategy based on the \textit{partial trace operator} (Filipiak, Klein, and Vojtkova (2018)). Suppose that an $n \times n$ matrix $A$ can be written in terms of $n_1 \times n_1$ blocks of $n_1 \times n_1$ dimensional matrices $A_{-1;i,j}$, where $n_{-1} := n/n_1$; that is, we have

$$A = \begin{pmatrix}
A_{-1;1,1} & \cdots & A_{-1;1,n_1} \\
\vdots & \ddots & \vdots \\
A_{-1;n_1,1} & \cdots & A_{-1;n_1,n_1}
\end{pmatrix}.$$

Then the partial trace operator $\text{PTr}_{n_1} : \mathbb{R}^{n \times n} \to \mathbb{R}^{n_1 \times n_1}$ is defined as follows:

$$\text{PTr}_{n_1}(A) = \begin{pmatrix}
\text{tr}(A_{-1;1,1}) & \cdots & \text{tr}(A_{-1;1,n_1}) \\
\vdots & \ddots & \vdots \\
\text{tr}(A_{-1;n_1,1}) & \cdots & \text{tr}(A_{-1;n_1,n_1})
\end{pmatrix}.$$

Consider model (2.1), and let $\Sigma_{-1} := \Sigma_2 \otimes \cdots \otimes \Sigma_v$. In this case

$$\text{PTr}_{n_1}(\Sigma) = \sigma^2 \text{tr}(\Sigma_{-1}) \times \Sigma_1.$$

Define the $n_1 \times n_1$ matrix $d^{(1)} := \text{PTr}_{n_1}(\Sigma)$. Then

$$\Sigma_1 = \frac{d^{(1)}}{\text{tr}(d^{(1)})/n_1}.$$

According to Definition 1.1(ii) of Filipiak et al. (2018), $\text{PTr}_{n_1}(\Sigma) = \sum_{\ell=1}^{n_{-1}} (I_{n_1} \otimes c_{\ell,n_{-1}}) \Sigma (I_{n_1} \otimes e_{\ell,n_{-1}})$, where $e_{\ell,n_{-1}}$ is the $n_{-1} \times 1$ elementary vector with one in position $\ell$ and zero elsewhere. In this sense, $d^{(1)}$ is a quadratic form of $\Sigma$. 

4
We next consider the remaining components $\Sigma_h$, $h = 2, \ldots, v$. Write

$$
\Sigma_{-h} := \Sigma_{h+1} \otimes \cdots \otimes \Sigma_v \otimes \Sigma_1 \otimes \cdots \otimes \Sigma_{h-1},
$$

for $h = 2, \ldots, v$. Note that $\Sigma_{-h}$ is $n_{-h} \times n_{-h}$ dimensional, where $n_{-h} := n/n_h$. Recalling the identity $B \otimes A = K_{p,m}(A \otimes B)K_{m,p}$ for $A (m \times m)$ and $B (p \times p)$ (Magnus and Neudecker (1986) Lemma 4), we write

$$
\Sigma^{(h)} := K_{n_h \times \cdots \times n_v, n_1 \times \cdots \times n_{h-1}}(\sigma^2 \otimes \Sigma_1 \otimes \cdots \otimes \Sigma_v)K_{n_1 \times \cdots \times n_{h-1}, n_h \times \cdots \times n_v} = \sigma^2 \times \Sigma_h \otimes \Sigma_{h+1} \otimes \cdots \otimes \Sigma_v \otimes \Sigma_1 \otimes \cdots \otimes \Sigma_{h-1} = \sigma^2 \times \Sigma_h \otimes \Sigma_{-h}.
$$

In this case

$$
PTr_{n_h}(\Sigma^{(h)}) = \sigma^2 \text{tr}(\Sigma_{-h}) \times \Sigma_h.
$$

Define the $n_h \times n_h$ matrix $d^{(h)} := PTr_{n_h}(\Sigma^{(h)})$. Then

$$
\Sigma_h = \frac{d^{(h)}}{\text{tr}(d^{(h)})/n_h}.
$$

3 Estimation

We observe an $n$-dimensional weakly stationary time series vector $\{y_t\}_{t=1}^T$ with mean $\mu$ and covariance matrix $\Sigma$. For simplicity, we assume that $\mu = 0$.

3.1 Partial Means Method

In this section we define the partial means estimator for model (2.1). Define the following sample quantities:

$$
M_T := \frac{1}{T} \sum_{t=1}^T y_t y_t^\top,
$$

$$
\hat{\Omega}_j := A_j \left( \frac{1}{T} \sum_{t=1}^T y_t y_t^\top \right) A_j^\top = A_j M_T A_j^\top, \quad (3.1)
$$

$$
\hat{\Sigma}_j := \frac{\hat{\Omega}_j}{\text{tr}(\hat{\Omega}_j)/n_j}
$$

Then the partial means estimator $\hat{\Sigma}$ for $\Sigma$ is:

$$
\hat{\Sigma} = \hat{\sigma}^2 \times \hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v,
$$

$$
\hat{\sigma}^2 := \frac{\text{tr}(M_T)}{n}. \quad (3.2)
$$

This estimator is by construction symmetric and positive semidefinite; it will be positive definite with probability one provided that $\max_{1 \leq j \leq v} n_j < T$. The partial means estimator $\hat{\Sigma}^{-1}$ for $\Sigma^{-1}$ is $\hat{\Sigma}^{-1} = \hat{\sigma}^{-2} \times \hat{\Sigma}_1^{-1} \otimes \cdots \otimes \hat{\Sigma}_v^{-1}$. Note that $\Sigma^{-1}$ exists even if $n > T$.

3.2 Quadratic Form Method

Define $\hat{d}^{(1)} := PTr_{n_1}(M_T)$. Then

$$
\hat{\Sigma}_1 = \frac{\hat{d}^{(1)}}{\text{tr}(\hat{d}^{(1)})/n_1}.
$$
Likewise define the "permuted" sample covariance matrix

\[ M_T^{(h)} := K_{n_h \times \cdots \times n_v, n_1 \times \cdots \times n_{h-1}} M_T K_{n_1 \times \cdots \times n_{h-1}, n_h \times \cdots \times n_v}, \]

for \( h = 2, \ldots, v \). Define \( \hat{d}^{(h)} := \text{PTr}_{n_h}(M_T^{(h)}) \) for \( h = 2, \ldots, v \). Then

\[ \hat{\Sigma}_h = \frac{\hat{d}^{(h)}}{\text{tr}(\hat{d}^{(h)})/n_h}, \]

for \( h = 1, \ldots, v \).

The quadratic form estimator \( \hat{\Sigma} \) for \( \Sigma \) is:

\[ \hat{\Sigma} = \sigma^2 \times \hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v, \]

where \( \sigma^2 \) is defined in (3.2). By Lemma 2.4 of Filipiak et al. (2018), if \( M_T \) is symmetric and positive definite, then so are \( \{\hat{\Sigma}_j\}_{j=1}^v \) and \( \hat{\Sigma} \). However, even for positive semidefinite \( M_T \), \( \{\hat{\Sigma}_j\}_{j=1}^v \) and \( \hat{\Sigma} \) will be positive definite.

The quadratic form estimator \( \hat{\Sigma}^{-1} \) for \( \Sigma^{-1} \) is \( \hat{\Sigma}^{-1} = \sigma^{-2} \times \hat{\Sigma}_1^{-1} \otimes \cdots \otimes \hat{\Sigma}_v^{-1} \). Note that \( \hat{\Sigma}^{-1} \) exists even if \( n > T \). The quadratic form estimators are closely related to the quasi-maximum likelihood estimation (QMLE), but have the particular advantage in large dimensions in the sense that they are in closed form.

In general we expect each element of \( M_T \) to be \( \sqrt{T} \) consistent, but here we are averaging over a large number of such elements. Under a cross-sectional weak dependence condition, like Assumption 4.4, we should have a rate improvement. The typical convergence rate for \( \hat{\sigma}^2 = \text{tr}(M_T)/n \) is \( \sqrt{T/\log n} \) (see Lemma A.2). Under the same weak dependence condition, we should also have a rate improvement. We formally establish these results in Section 4.2.

4 Asymptotic Properties

We make the following assumptions:

**Assumption 4.1.**

(i) The sample \( \{y_t\}_{t=1}^T \) are independent over \( t \).

(ii)

\[ \max_{1 \leq i \leq n} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[|y_{t,i}|^m] \leq A^m, \quad m = 2, 3, \ldots, \]

for some absolute positive constant \( A \).

(iii) Consider a normal random vector \( z_t \) which has the same mean vector and covariance matrix as those of \( y_t \). The \( n^2 \times n^2 \) kurtosis matrix of \( y_t \) satisfies

\[ \text{var}(y_t \otimes y_t) \leq C \text{var}(z_t \otimes z_t), \]

for some absolute positive constant \( C \) for \( 1 \leq t \leq T \), where \( \leq \) is to be interpreted componentwise.

Assumption 4.1(i) facilitates our technical analysis, but is perhaps not necessary. Assumption 4.1(ii) assumes the existence of an infinite number of moments of \( y_t \), which allows one to invoke some concentration inequality such as a version of the Bernstein’s inequality. Normal random vectors or random vectors which exhibit some exponential-type tail probability (e.g., subgaussianity, subexponentiality, semiexponentiality etc) satisfy this. Assumption 4.1(iii) assumes that the kurtosis matrix of \( y_t \) is of the same order of magnitude as if a normal random vector were used. We impose this restriction on the kurtosis matrix of \( y_t \) because not much research has touched the unrestricted kurtosis matrix in the large dimensional case.
Assumption 4.2.

(i) Assume $n_j$ fixed for $j = 1, \ldots, v$, but $v \to \infty$ as $n, T \to \infty$.

(ii) $\min_{1 \leq j \leq v} \lambda_{\min}(\Sigma_j)$ is bounded away from zero by an absolute positive constant.

Assumption 4.2(i) assumes that the dimensions of sub-matrices are fixed while the number of sub-matrices tends to infinity. Note that Assumption 4.2(ii) does not necessarily imply that $\lambda_{\min}(\Sigma)$ is bounded away from zero by an absolute positive constant. This is because $\lambda_{\min}(\Sigma) = \sigma^2 \times \prod_{j=1}^v \lambda_{\min}(\Sigma_j)$ and $v \to \infty$.

Lemma 4.1. Suppose Assumption 4.2(i) hold. We have

(i) $v = O(\log n)$.

(ii) $\max_{1 \leq j \leq v} \lambda_{\max}(\Sigma_j)$ is bounded from the above by an absolute positive constant.

Note that Lemma 4.1(ii) does not necessarily imply that $\lambda_{\max}(\Sigma)$ is bounded from the above by an absolute positive constant. This is because $\lambda_{\max}(\Sigma) = \sigma^2 \times \prod_{j=1}^v \lambda_{\max}(\Sigma_j)$ and $v \to \infty$.

4.1 Partial Means

The following theorem gives a rate of convergence for $\|\hat{\Sigma} - \Sigma\|_F$.

Theorem 4.1. Suppose Assumptions 4.1 and 4.2 hold. Then we have

(i) $\frac{\|\hat{\Sigma} - \Sigma\|_F}{\|\Sigma\|_F} = O_p\left(\sqrt{\frac{\log^3 n}{T}}\right)$.

(ii) $\frac{\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_F}{\|\Sigma^{-1}\|_F} = O_p\left(\sqrt{\frac{\log^3 n}{T}}\right)$.

The reason that we divide the Frobenius norm of the estimation error, say, $\|\hat{\Sigma} - \Sigma\|_F$, by the Frobenius norm of the target, i.e., $\|\Sigma\|_F$, is to define a proper notion of "consistency". This is necessary because the dimension of the matrix, $n$, is growing to infinity. In particular, even if every element of a matrix valued estimator is converging in probability to the corresponding element of its target matrix, there is no guarantee that its overall estimation error will converge to zero in probability when $n, T \to \infty$. The rescaling of the Frobenius norm of the estimation error is standard in the large dimensional case, but in the literature scholars tend to divide the Frobenius norm of the estimator error by $\sqrt{n}$ (e.g., see Bickel and Levina (2008) Theorem 2, Fan, Liao, and Mincheva (2011) p3330, Ledoit and Wolf (2004) Definition 1 etc).

Theorem 4.1 is comparable to the convergence rates of other existent estimators in the large dimensional case. Take part (i) of Theorem 4.1 as an illustration: We have $\|\hat{\Sigma} - \Sigma\|_F = O_p\left(\|\Sigma\|_F(\log^3 n/T)^{1/2}\right)$. A typical threshold estimator $\hat{\Sigma}_{\text{thres}}$ has $\|\hat{\Sigma}_{\text{thres}} - \Sigma\|_F = O_p\left((sn \log n/T)^{1/2}\right)$, where $s$ is some sparsity index (see Bickel and Levina (2008) Theorem 2 with $q = 0$). According to Bickel and Levina (2008), $s$ is the upper bound of non-zero elements for every row, so $\|\Sigma\|_F = O(\sqrt{sn})$ under the sparsity model. Hence we can write $\|\hat{\Sigma}_{\text{thres}} - \Sigma\|_F = O_p\left(\|\Sigma\|_F(\log n/T)^{1/2}\right)$. Thus the rates only differ by a logarithmic factor.

We also establish a result for the rate of convergence for $\|\hat{\Sigma} - \Sigma\|_{\ell_2}$. We replace Assumption 4.1(ii) with the following
Assumption 4.3.

\[
\max_{1 \leq j \leq v} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} |a^T w_{j,t}|^{2m} \leq A^m, \quad m = 2, 3, \ldots,
\]

for some absolute positive constant \(A\), where \(w_{j,t} := A_j y_t / \sqrt{\sigma_j^2}\) and \(a \in \mathbb{R}^n\) is arbitrary with \(\|a\|_2 = 1\).

Note that the covariance matrix of \(w_{j,t}\) is \(\Sigma_j\). Assumption 4.3 in essence says that the random vector \(w_{j,t}\) has an infinite number of moments.

Theorem 4.2. Suppose Assumptions 4.1(i), 4.2 and 4.3 hold. Then we have

(i) \[
\frac{\|\hat{\Sigma} - \Sigma\|_{\ell_2}}{\|\Sigma\|_{\ell_2}} = O_p \left( \sqrt{\frac{\log^3 n}{T}} \right).
\]

(ii) \[
\frac{\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_{\ell_2}}{\|\Sigma^{-1}\|_{\ell_2}} = O_p \left( \sqrt{\frac{\log^3 n}{T}} \right).
\]

Theorem 4.2 is comparable to the convergence rates of other existent estimators in the large dimensional case. Take part (i) of Theorem 4.2 as an illustration: We have \(\|\hat{\Sigma} - \Sigma\|_{\ell_2} = O_p \left( \frac{1}{\|\Sigma\|_{\ell_2}} (\log^3 n/T)^{1/2} \right)\). A typical threshold estimator \(\hat{\Sigma}_{\text{thres}}\) has \(\|\hat{\Sigma}_{\text{thres}} - \Sigma\|_{\ell_2} = O_p \left( \frac{1}{\|\Sigma\|_{\ell_2}} (\log n/T)^{1/2} \right)\). Again the rates only differ by a logarithmic factor.

4.2 Quadratic Form

We first give a cross-sectional weak dependence condition.

Assumption 4.4.

\[
\lim_{n \to \infty} \frac{1}{n} \left( \frac{1}{\sigma^4} \|\Sigma\|_F^2 \right) = \lim_{n \to \infty} \frac{1}{n} \left( \prod_{j=1}^{n} \|\Sigma_j\|_F^2 \right) = \omega < \infty.
\]

Assumption 4.4 characterizes the cross-sectional dependence of \(\{y_t\}_{t=1}^{T}\). Recall that the largest eigenvalue of \(\Sigma\) is a summary measure of cross-sectional dependence. In other words, Assumption 4.4 restricts the rate at which \(\lambda_{\max}(\Sigma)\) could grow. In particular, we have

\[
\lambda_{\max}(\Sigma) = \|\Sigma\|_{\ell_2} \leq \|\Sigma\|_F = O(\sqrt{n}).
\]

In this case, according to Chudik and Pesaran (2013), \(\{y_t\}_{t=1}^{T}\) is said to be cross-sectionally weakly dependent. One sufficient condition for Assumption 4.4 is that \(\Sigma/\sigma^2\) has bounded maximum column sum matrix norm (i.e., \(\|\Sigma\|_{\ell_1}/\sigma^2 = O(1)\)) or bounded maximum row sum matrix norm (i.e., \(\|\Sigma\|_{\ell_\infty}/\sigma^2 = O(1)\)). To see this

\[
\frac{1}{n} \left( \frac{1}{\sigma^4} \|\Sigma\|_F^2 \right) \leq \frac{1}{n} \left( \frac{1}{\sigma^4} n \|\Sigma\|_{\ell_1}^2 \right) = O(1)
\]

\[
\frac{1}{n} \left( \frac{1}{\sigma^4} \|\Sigma\|_F^2 \right) \leq \frac{1}{n} \left( \frac{1}{\sigma^4} n \|\Sigma\|_{\ell_\infty}^2 \right) = O(1).
\]

\(^3\)Strictly speaking \(a\) depends on \(j\) but we suppress this dependence.
The assumption of bounded maximum column/row sum matrix norm has been used by Fan, Liao, and Yao (2015) (their Assumption 4.1(i)) and Pesaran and Yamagata (2012) (their Assumption 3). We remark that Assumption 4.4 is only a sufficient condition. Weaker versions of it are possible but the rate of convergence of the quadratic form estimator, say in Theorem 4.3, will be slower. Then we have

**Theorem 4.3.** Suppose Assumptions 4.1, 4.2 and 4.4 hold. Then

(i) \[ \frac{\|\hat{\Sigma} - \Sigma\|_F}{\|\Sigma\|_F} = O_p\left(\sqrt{\frac{\log^3 n}{nT}}\right). \]

(ii) \[ \frac{\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_F}{\|\Sigma^{-1}\|_F} = O_p\left(\sqrt{\frac{\log^3 n}{nT}}\right). \]

(iii) \[ \frac{\|\hat{\Sigma} - \Sigma\|_1}{\|\Sigma\|_1} = O_p\left(\sqrt{\frac{\log^3 n}{nT}}\right). \]

(iv) \[ \frac{\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_1}{\|\Sigma^{-1}\|_1} = O_p\left(\sqrt{\frac{\log^3 n}{nT}}\right). \]

Comparing Theorem 4.3 with Theorem 4.1, we see that an additional factor \(\sqrt{n}\) has appeared in the denominator. Because of the cross-sectional weak dependence condition (Assumption 4.4), the quadratic form estimator is able to achieve a much faster rate of convergence than the partial means estimator does.

### 4.3 Central Limit Theorems

In this section we investigate the asymptotic distribution of the partial means estimator \(\hat{\Sigma}_j\). Recalling that the partial means estimator \(\hat{\Sigma}_j = \hat{\Omega}_j/\{\text{tr}(\hat{\Omega}_j)/n\}_j\) and \(\Sigma_j = \Omega_j/\{\text{tr}(\Omega_j)/n\}_j = \Omega_j/\sigma^2_j\), we write

\[
\sqrt{T} \text{vech}(\hat{\Sigma}_j - \Sigma_j) = \sqrt{T} \text{vech} \left( \frac{\hat{\Omega}_j}{\text{tr}(\hat{\Omega}_j)/n_j} - \frac{\hat{\Omega}_j}{\sigma^2_j} \right) + \sqrt{T} \text{vech} \left( \frac{\Omega_j}{\sigma^2_j} - \frac{\Omega_j}{\sigma^2_j} \right). \tag{4.1}
\]

Theorem A.2 in Appendix shows that the second term on the right side of (4.1) converges weakly to a normal distribution under some conditions. The first term on the right side of (4.1) is not \(o_p(1)\) though, because the convergence rate of \(\text{tr}(\Omega_j)/n_j\) to \(\sigma^2_j\) is only \(\sqrt{T}\), not fast enough. To get around this, we propose a hybrid estimator \(\hat{\Sigma}_j\):

\[
\hat{\Sigma}_j := \hat{\Omega}_j/\sigma^2_j,
\]

\[
\tilde{\sigma}^2_j := \hat{\sigma}^2 \times \tilde{\omega}_1 \times \cdots \times \tilde{\omega}_{j-1} \times \tilde{\omega}_{j+1} \times \cdots \times \tilde{\omega}_v,
\]

\[
\tilde{\omega}_\ell := c_\ell^T \hat{\Sigma}_\ell c_\ell, \quad \ell = 1, \ldots, v,
\]

where \(\hat{\Omega}_j\), \(\hat{\sigma}^2\) and \(\hat{\Sigma}_\ell\) are defined in (3.1), (3.2) and (3.3), respectively. We call \(\hat{\Sigma}_j\) the hybrid estimator of \(\Sigma_j\) because its numerator and denominator are based on the partial means estimator and the quadratic form estimator, respectively. We make the following assumption.
Assumption 4.5.

(i) Suppose \( c_\ell = \iota_\ell \) for \( \ell = 1, \ldots, v \), where \( \iota_\ell \) is a \( \ell \times 1 \) vector of ones.

(ii) Recall that \( w_{j,t} := A_j y_t / \sqrt{\sigma_{2*}^2} \). Consider an \( n_j \times 1 \) normal random vector \( z_{j,t} \) which has the same mean vector \( \theta \) and covariance matrix \( \Sigma_j \) as those of \( w_{j,t} \). Suppose that

\[
\lambda_{\min} \left( D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^T \text{var} (w_{j,t} \otimes w_{j,t}) \right] D_{n_j}^{+\top} \right) \geq C \lambda_{\min} \left( D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^T \text{var} (z_{j,t} \otimes z_{j,t}) \right] D_{n_j}^{+\top} \right)
\]

for some absolute positive constant \( C \) for \( j = 1, \ldots, v \).

Assumption 4.5(i) is for illustration only; other choices of \( c_\ell \) should be possible. Assumption 4.5(ii) assumes that the minimum eigenvalue of some variant of the kurtosis matrix of \( w_{j,t} \) is of the same order of magnitude as if a normal random vector were used. We make this assumption because the kurtosis matrix of a general case has a rather complicated form, but it is reasonable to assume that its minimum eigenvalue is of the same order of magnitude as that of a normal kurtosis matrix.

Theorem 4.4. Suppose Assumptions 4.1(i), 4.2, 4.3, 4.4 and 4.5 hold. Then as \( n, T \to \infty \), for \( j, k = 1, \ldots, v \) (\( j \neq k \)),

\[
\sqrt{T} \text{vech}(\hat{\Sigma}_j - \Sigma_j) \overset{d}{\to} N \left( 0, D_{n_j}^+ \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \text{var}(w_{j,t} \otimes w_{j,t}) \right] D_{n_j}^{+\top} \right) =: N(0, V_{j,j}),
\]

where \( w_{j,t} \) is defined in Assumption 4.5(ii).

\[
\sqrt{T} \begin{bmatrix}
\text{vech}(\hat{\Sigma}_j - \Sigma_j) \\
\text{vech}(\hat{\Sigma}_k - \Sigma_k)
\end{bmatrix} \overset{d}{\to} N(0, V),
\]

where

\[
V := \begin{bmatrix}
V_{j,j} & V_{j,k} \\
V_{k,j} & V_{k,k}
\end{bmatrix}
\]

\[
V_{j,k} := D_{n_j}^+ \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T \text{cov} \left\{ \text{vec} \left( w_{j,t} w_{j,t}^\top - E[w_{j,t} w_{j,t}^\top] \right), \text{vec} \left( w_{k,t} w_{k,t}^\top - E[w_{k,t} w_{k,t}^\top] \right) \right\} \right] D_{n_k}^{+\top},
\]

In the case of normality, we have the following simplified formulas for \( V_{j,j} \) and \( V_{j,k} \).

Corollary 4.1. If the sample \( \{y_t\}_{t=1}^T \) are normally distributed, we have

\[
V_{j,j} = 2D_{n_j}^+ (\Sigma_j \otimes \Sigma_j) D_{n_j}^{+\top} \quad V_{j,k} = 2D_{n_j}^+ \left( \frac{A_j \Sigma A_j^\top}{\sigma_{2*}} \otimes \frac{A_k \Sigma A_k^\top}{\sigma_k^2} \right) D_{n_k}^{+\top}
\]

for \( j, k = 1, \ldots, v \).
5 Test Statistics

We apply our methodology to the testing issue. We consider the problem of testing the null hypothesis $H_0 : \mu = 0$ against the alternative $H_1 : \mu \neq 0$. The classical Wald statistic (based on the sample covariance matrix) is not defined when $n \geq T$; there is a large literature that proposes alternative test statistics. Bai and Saramadasa (1996) proposed a statistic based on $\| \hat{\mu} \|_2^2$, where $\hat{\mu} := \frac{1}{n} \sum_{t=1}^T y_t$, thereby avoiding the inversion of the large sample covariance matrix, and established its asymptotic normality. Pesaran and Yamagata (2012) extended this approach to the Capital Asset Pricing Model (CAPM) regression setting and proposed several test statistics. One of the test statistics is based on $\| t \|_2^2$, where $t$ is a vector of individual $t$-statistics; Pesaran and Yamagata (2012) derived its limiting normal distribution under cross-sectional weak dependence conditions. Fan et al. (2015) considered the Wald statistic for testing the CAPM restrictions inside a linear regression in the large dimensional case. They regularized the estimated error covariance matrix by imposing a sparsity assumption, and used that to form the quadratic form. They established the null limiting distribution of their test statistic (they also proposed a novel power enhancement procedure, which we do not study here).

We can define all the classical Gaussian likelihood based statistics under our model structure. First, the Lagrange multiplier statistic is

$$LM_{n,T} = T \hat{\mu}^\top \Sigma^{-1} \hat{\mu}.$$  

The likelihood ratio statistic is

$$\lambda_{n,T} = T \left( \log \det \Sigma - \log \det \Sigma_{\hat{\mu}} \right),$$

where $\Sigma_{\hat{\mu}}$ is the unrestricted covariance matrix estimator under $H_1 : \mu \neq 0$. The Wald statistic is

$$W_{n,T} = T \hat{\mu}^\top \Sigma_{\hat{\mu}}^{-1} \hat{\mu},$$

which is the Hotelling $T^2$-statistic based on our covariance matrix estimator.

We next present the large sample properties of our test statistic $LM_{n,T}$. We make one more cross-sectional weak dependence assumption.

**Assumption 5.1.**

$$\lim_{n \to \infty} \frac{1}{n} \sigma^2 \| \Sigma^{-1} \|_1 = \lim_{n \to \infty} \frac{1}{n} \left( \prod_{j=1}^v \| \Sigma_j^{-1} \|_1 \right) = \omega' < \infty.$$

Assumption 5.1 restricts the rate at which $\lambda_{\min}(\Sigma)$ could drift to zero. To see this,

$$\frac{1}{\lambda_{\min}(\Sigma)} = \lambda_{\max}(\Sigma^{-1}) = \| \Sigma^{-1} \|_{\ell_2} \leq \| \Sigma^{-1} \|_1 = O(n).$$

One sufficient condition for Assumption 5.1 is that $\sigma^2 \Sigma^{-1}$ has bounded maximum column sum matrix norm (i.e., $\sigma^2 \| \Sigma^{-1} \|_{\ell_1} = O(1)$) or bounded maximum row sum matrix norm (i.e., $\sigma^2 \| \Sigma^{-1} \|_{\ell_\infty} = O(1)$). To see this

$$\frac{1}{n} \sigma^2 \| \Sigma^{-1} \|_1 = \frac{1}{n} \sigma^2 \sum_{i=1}^n \sum_{j=1}^n |(\Sigma^{-1})_{i,j}| \leq \sigma^2 \max_{1 \leq i \leq n} \sum_{j=1}^n |(\Sigma^{-1})_{i,j}| = \sigma^2 \| \Sigma^{-1} \|_{\ell_\infty} = O(1)$$

$$\frac{1}{n} \sigma^2 \| \Sigma^{-1} \|_1 = \frac{1}{n} \sigma^2 \sum_{i=1}^n \sum_{j=1}^n |(\Sigma^{-1})_{i,j}| \leq \sigma^2 \max_{1 \leq j \leq n} \sum_{i=1}^n |(\Sigma^{-1})_{i,j}| = \sigma^2 \| \Sigma^{-1} \|_{\ell_1} = O(1)$$

The assumption of bounded maximum column/row sum matrix norm has been used by Fan et al. (2015) (their Assumption 4.1(i)) and Pesaran and Yamagata (2012) (their Assumption 3). We
remark that Assumption 5.1 is only a sufficient condition for Theorem 5.1. Weaker versions of Assumption 5.1 are possible but one needs a tighter growth rate of $n$ relative to $T$ than $\frac{\log^5 n}{T} = o(1)$ required by Theorem 5.1.

**Theorem 5.1.** Suppose Assumptions 4.1(i), 4.2, 4.4, and 5.1 hold. Assume $\frac{\log^5 n}{T} = o(1)$. In addition, assume that either of the following assumption holds.

(a) The sample $\{y_t\}_{t=1}^T$ are normally distributed.

(b) Consider the Cholesky decomposition of $\Sigma$, i.e., $\Sigma = LL^\top$, where $L$ is a nonsingular lower triangular matrix $L$ with positive diagonal elements. Assume that $x_t := L^{-1}y_t$ is cross-sectionally independent, and

$$\limsup_{n,T \to \infty} \frac{1}{nT} \sum_{t=1}^n \sum_{i=1}^T \mathbb{E}[x^4_{t,i}] < \infty.$$  

Then under $H_0 : \mu = 0$, as $n,T \to \infty$,

$$\frac{LM_{n,T} - n}{\sqrt{2n}} \rightarrow N(0,1).$$

(5.1)

Note that the additional assumption (either (a) or (b)) is standard in the literature. Fan et al. (2015) maintains (a) (their Assumption 4.1(i)) while Pesaran and Yamagata (2012) maintains (b) (their Assumption 2a). Under sequential asymptotics ($T \to \infty$ and then $n \to \infty$), $\sqrt{T}\hat{\mu}$ is approximately normally distributed and so the limiting properties could be calculated for the non-normal case as if assumption (a) held. For assumption (b), note that $\text{var}(x_t) = \text{I}_n$, so strengthening from cross-sectional uncorrelatedness to cross-sectional independence is rather innocuous. In addition, we assume that the grand average (over both $t$ and $i$) of the fourth moment of $x_{t,i}$ is finite for $n,T$ sufficiently large, which is also a weak assumption.

Consider a generic estimator $\hat{\Sigma}_G^{-1}$ for $\Sigma^{-1}$ whose rate of convergence is $\|\hat{\Sigma}^{-1}_G - \Sigma^{-1}\|_{\ell_2} = O_p\left(\sqrt{\log n/T}\right)$ (see, e.g., Fan et al. (2011)). Then (5.1) requires, as both Pesaran and Yamagata (2012) and Fan et al. (2015) have pointed out, $n \log n/T = o(1)$, which is essentially a low-dimensional scenario. Pesaran and Yamagata (2012) and Fan et al. (2015) have hence come up with their own ingenious ways to relax the condition $n \log n/T = o(1)$ and established results similar to (5.1) for their Wald statistics in the CAPM context. Likewise Theorem 5.1 only requires $\log^5 n/T = o(1)$ because of the fast rate of convergence of $\hat{\Sigma}^{-1}_{1} : \|\Sigma^{-1} - \Sigma^{-1}\|_1 = O_p\left(\log^3 n / nT\right)$, the rate of convergence of the partial means estimator $\hat{\Sigma}^{-1}_{1}$,

$$\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_{\ell_2} = O_p\left(\log^3 n / nT\right),$$

is perhaps not sufficient for (5.1) to hold.

In the low-dimensional case ($n$ fixed, $T \to \infty$), the Lagrange multiplier statistic $LM_{n,T}$, the likelihood ratio statistic $\lambda_{n,T}$, and the Wald statistic $W_{n,T}$ are asymptotically equivalent in the sense that they all converge in distribution to $\chi_n^2$. In the large dimensional case ($n,T \to \infty$), we expect that similar results as Theorem 5.1 would hold for $\lambda_{n,T}$ and $W_{n,T}$. However, perhaps stronger conditions on the relative growth of $n$ and $T$ are needed. This is because for the Lagrange multiplier test we estimate $\Sigma^{-1}$ under $H_0 : \mu = 0$ whereas for the Wald or likelihood ratio test, one needs to estimate $\Sigma^{-1}$ under $H_1 : \mu \neq 0$. Definitely there is a cost of estimating $\hat{\Sigma}^{-1}$.

\footnote{The finite sample performance of these statistics is known to vary. Park and Phillips (1988) established higher order approximations for the Wald test of nonlinear restrictions in the finite dimensional case, and showed how to improve the performance of the test statistic. It may be possible to apply their methodology to the large dimensional case.}
the large dimensional mean vector $\mu$, so we expect that the Lagrange multiplier test works better than the other two in the large dimensional case.

In the simulation study below we compare our test with the test statistic that uses the Ledoit and Wolf procedures to regularize the sample covariance matrix estimator.

6 Simulation Study

6.1 The Correctly Specified Case

We suppose that $y_t \sim N(\mu, \Sigma)$ with $\Sigma = \Sigma_1 \otimes \cdots \otimes \Sigma_v$, where

$$\Sigma_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix} \quad |\rho_j| < 1, \quad j = 1, \ldots, v,$$

so in this case $\Sigma$ is also the correlation matrix. Given $\|\Sigma_j\|^2_F = 2(1 + \rho_j^2)$, we have

$$\frac{1}{n} \|\Sigma\|^2_F = \frac{1}{n} \prod_{j=1}^v 2(1 + \rho_j^2) = \prod_{j=1}^v (1 + \rho_j^2),$$

which converges to a finite, non-zero limit as $v \to \infty$ if and only if $\sum_{j=1}^v \rho_j^2$ converges (Knopp (1947) Theorem 28.3). This is to say that in this setting $\Sigma$ satisfies Assumption 4.4 if and only if $\sum_{j=1}^\infty \rho_j^2 < \infty$.

In this case, we have

$$\|\Sigma_j^{-1}\|_1 = \frac{2(1 + |\rho_j|)}{1 - \rho_j^2} = \frac{2}{1 - |\rho_j|} \quad j = 1, \ldots, v,$$

so that via Lemma B.4

$$\frac{1}{n} \|\Sigma^{-1}\|_1 = \frac{1}{n} \prod_{j=1}^v \|\Sigma_j^{-1}\|_1 = \frac{1}{n} \prod_{j=1}^v \frac{1}{1 - |\rho_j|} = \frac{1}{\prod_{j=1}^v (1 - |\rho_j|)}.$$

The denominator of the right side of the preceding display converges to a finite, non-zero limit as $v \to \infty$ if and only if $\sum_{j=1}^v |\rho_j|$ converges (Knopp (1947) Theorem 28.4). This is to say that in this setting $\Sigma$ satisfies Assumption 5.1 if and only if $\sum_{j=1}^\infty |\rho_j| < \infty$. Furthermore, the largest eigenvalue of $\Sigma$ is $\prod_{j=1}^v (1 + |\rho_j|)$, which converges as $v \to \infty$ if and only if $\sum_{j=1}^\infty |\rho_j|$ converges (Knopp (1947) Theorem 28.3).

We consider $\mu = 0$, $T = 252$ and $v = 10$ so that $n = 2^v = 1024$. We set $\rho_j = \rho^j$ for $j = 1, \ldots, v$ with $\rho = 0.5, 0.7, 0.85$. The number of Monte Carlo simulations is 1000. We compare our partial means and quadratic form estimators with Ledoit and Wolf (2004)’s linear shrinkage estimator (the LW04 estimator hereafter) and Ledoit and Wolf (2017)’s direct nonlinear shrinkage estimator (the LW17 estimator hereafter).\footnote{The Matlab code for the LW04 and LW17 estimators is downloaded from the website of Professor Michael Wolf from the Department of Economics at the University of Zurich. We are grateful for this.}

Given a generic estimator $\hat{\Sigma}_G$ of the covariance matrix $\Sigma$ and in each simulation, we compute the relative estimation error

$$\frac{\|\hat{\Sigma}_G - \Sigma\|_F^2}{\|\Sigma\|_F^2}.$$
an estimator of the precision matrix $\Sigma^{-1}$ is of more interest than that of $\Sigma$ itself, so we also compute RE in terms of $\Sigma^{-1}$; that is, we compute the mean of

$$\frac{\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_F^2}{\|\Sigma^{-1}\|_F^2}$$

across simulations. Note that this requires invertibility of the generic estimator $\hat{\Sigma}_G$ and therefore cannot be calculated for the sample covariance matrix $M_T$ when $n > T$.

We also calculate

$$1 - \frac{\mathbb{E}\|\hat{\Sigma} - \Sigma\|_F^2}{\mathbb{E}\|M_T - \Sigma\|_F^2},$$

where the expectation operator is taken with respect to all the simulations. The preceding display is called the simulated percentage relative improvement in average loss (PRIAL) criterion in terms of $\Sigma$ by Ledoit and Wolf (2004). The PRIAL measures the performance of the generic estimator $\hat{\Sigma}_G$ with respect to the sample covariance estimator $M_T$. Note that PRIAL $\in (-\infty, 1]$: A negative value means $\hat{\Sigma}_G$ performs worse than $M_T$ while a positive value means otherwise. Likewise we also compute

$$1 - \frac{\mathbb{E}\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_F^2}{\mathbb{E}\|M_T^{-1} - \Sigma^{-1}\|_F^2}.$$

Note that this requires invertibility of the sample covariance matrix $M_T$ and therefore can only be calculated for $n < T$.

Finally, we also compute the size of the LM test for $H_0 : \mu = 0$ (Theorem 5.1). The significance level is 5%. This also requires invertibility of $\hat{\Sigma}_G$.

The results are reported in Table 1. First, consider the top panel ($\rho = 0.5$). For the RE in terms of $\Sigma$ (i.e., RE-1), all the estimators beat the sample covariance matrix $M_T$ by a large margin. Both the partial means estimator $\hat{\Sigma}$ and quadratic form estimator $\tilde{\Sigma}$ also outperformed the LW04 and LW17 estimators considerably. For the RE in terms of $\Sigma^{-1}$ (i.e., RE-2), a similar pattern exists. Note that the RE-2 cannot be computed for $M_T$ because $M_T$ is not invertible when $n > T$. For the PRIAL in terms of $\Sigma$ (i.e., PRIAL-1), again both $\hat{\Sigma}$ and $\tilde{\Sigma}$ are better than the LW04 and LW17 estimators. The sample covariance matrix $M_T$ has zero PRIAL-1 by definition. The superiority of $\hat{\Sigma}$ and $\tilde{\Sigma}$ in this experiment is expected because the true covariance matrix is indeed a Kronecker product.

Considering the size of the LM test, we realize that the quadratic form estimator $\tilde{\Sigma}$ has the correct size while the LW04 estimator is slightly over-sized. Both the partial means estimator $\hat{\Sigma}$ and the LW17 estimators have massive size distortions. The massive size distortion of $\hat{\Sigma}$ is perhaps not surprising. Theorem 5.1 has been only established for $\tilde{\Sigma}$ so far. The proof of Theorem 5.1 utilized the fast rate of convergence of the quadratic form estimator $\left(\log^3 n/(nT)\right)^{1/2}$. The rate of convergence of the partial means estimator $\left(\log^3 n/(nT)\right)^{1/2}$ is perhaps not sufficient for a result like Theorem 5.1 to hold. Note that the LM test is not defined for $M_T$ because $M_T$ is not invertible. Undoubtedly, the quadratic form estimator $\tilde{\Sigma}$ is the best performing estimator.

As we increase the "mother" correlation parameter $\rho$ from 0.5 to 0.85, the performance of $\hat{\Sigma}$ remains unchanged across all four criteria. In terms of RE-1, the performances of $M_T$ and $\hat{\Sigma}$ improve while the performances of LW04 and LW17 estimators initially worsen and then improve. In terms of RE-2, PRIAL-1 and size of LM, the performances of all three estimators - $\hat{\Sigma}$, LW04 and LW17 - worsen, with $\tilde{\Sigma}$ faring slightly better. Again the quadratic form estimator $\tilde{\Sigma}$ is the best performing estimator.

### 6.2 The Misspecified Case

To gauge how well the Kronecker product model performs when the true covariance matrix does not have a Kronecker product form, we consider the Monte Carlo setting used in Ledoit and Wolf.
Table 1: \( MT, \hat{\Sigma}, \tilde{\Sigma}, \text{LW04} \) and \( \text{LW17} \) stand for the sample covariance matrix, the partial mean estimator (\( \hat{\ell}_\ell = \nu_\ell \) for \( \ell = 1, \ldots, v \)), the quadratic form estimator, the Ledoit and Wolf (2004)’s linear shrinkage estimator, and the Ledoit and Wolf (2017)’s direct nonlinear shrinkage estimator, respectively. RE-1 and RE-2 are the RE in terms of \( \Sigma \) and \( \Sigma^{-1} \), respectively. PRIAL-1 is the PRIAL in terms of \( \Sigma \).

\[
\begin{array}{cccccc}
M_T & \hat{\Sigma} & \tilde{\Sigma} & \text{LW04} & \text{LW17} \\
\rho = 0.5 & & & & \\
RE-1 & 3.001 & 0.076 & 0.000 & 0.242 & 0.243 \\
RE-2 & NA & 0.127 & 0.000 & 0.311 & 0.308 \\
PRIAL-1 & 0 & 0.975 & 1.000 & 0.919 & 0.919 \\
size of LM & NA & 0.427 & 0.051 & 0.085 & 1.000 \\
\rho = 0.7 & & & & \\
RE-1 & 1.767 & 0.071 & 0.000 & 0.429 & 0.429 \\
RE-2 & NA & 0.224 & 0.000 & 0.722 & 0.715 \\
PRIAL-1 & 0 & 0.960 & 1.000 & 0.757 & 0.757 \\
size of LM & NA & 0.494 & 0.050 & 0.158 & 1.000 \\
\rho = 0.85 & & & & \\
RE-1 & 0.503 & 0.059 & 0.001 & 0.320 & 0.315 \\
RE-2 & NA & 0.664 & 0.002 & 0.980 & 0.979 \\
PRIAL-1 & 0 & 0.883 & 0.998 & 0.363 & 0.374 \\
size of LM & NA & 0.705 & 0.051 & 0.334 & 1.000 \\
\end{array}
\]

Ledoit and Wolf (2004). We still assume that \( y_t \sim N(0, \Sigma) \). The true covariance matrix \( \Sigma \) is diagonal without loss of generality. The diagonal entries \( \Sigma_{ii} \) (i.e., the eigenvalues of \( \Sigma \)) are log normally distributed: \( \log \Sigma_{ii} \sim N(\mu_\omega, \sigma^2) \). Ledoit and Wolf (2004) defined the grand mean \( \mu \) of and cross-sectional dispersion \( \alpha^2 \) of the eigenvalues of \( \Sigma \) as, respectively,

\[
\mu := \frac{1}{n} \sum_{i=1}^{n} \Sigma_{ii} \quad \alpha^2 := \frac{1}{n} \sum_{i=1}^{n} (\Sigma_{ii} - \mu)^2.
\]

In the Monte Carlo simulations, we re-define \( \mu \) and \( \alpha^2 \) as the corresponding population counterparts:

\[
\mu = \mathbb{E} \Sigma_{ii} = e^{\mu_\omega + \sigma^2/2} \quad \alpha^2 = \text{var} \Sigma_{ii} = e^{2(\mu_\omega + \sigma^2)} - e^{2\mu_\omega + \sigma^2}.
\]

Ledoit and Wolf (2004) set \( \mu = 1 \), so we can solve \( \mu_\omega = -\log(1 + \alpha^2)/2 \) and \( \sigma^2 = \log(1 + \alpha^2) \), whence we have

\[
\log \Sigma_{ii} \sim N\left(-\frac{\log(1 + \alpha^2)}{2}, \log(1 + \alpha^2)\right).
\]

Note that in this data generating process, there are two sources of randomness: one from the normal distribution of \( y_t \) and the other from the log normal distribution of \( \Sigma_{ii} \). Also note that a diagonal covariance matrix need not have a Kronecker product structure unless, say, the diagonal elements are all equal. The number of Monte Carlo simulations is again set at 1000. In the baseline setting of Ledoit and Wolf (2004), \( n = 20, T = 40 \) and \( \alpha^2 = 0.5 \).

There are a few Kronecker products which we can consider to approximate \( \Sigma \) (see Hafner et al. (2018) for more discussions of model selection). The possible Kronecker factorizations are \( 5 \times 2 \times 2, 4 \times 5, 2 \times 10 \). Within each Kronecker factorization, we can further permute the Kronecker submatrices to obtain different Kronecker models. We experiment all the Kronecker products and compare with the LW04 and LW17 estimators. We only consider the quadratic form estimator given its better small sample properties exhibited in the correctly specified case.
Table 2: $M_T$, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, the quadratic form estimator (factorisations given in parentheses), the Ledoit and Wolf (2004)’s linear shrinkage estimator, and the Ledoit and Wolf (2017)’s direct nonlinear shrinkage estimator, respectively. RE-1 and RE-2 are the RE in terms of $\Sigma$ and $\Sigma^{-1}$, respectively. PRIAL-1 and PRIAL-2 are the PRIAL in terms of $\Sigma$ and $\Sigma^{-1}$, respectively.

|       | $M_T$ | $\tilde{\Sigma}$ | $\tilde{\Sigma}$ | $\tilde{\Sigma}$ | $\Sigma$ |
|-------|-------|-------------------|-------------------|-------------------|-----------|
|       | (5 $\times$ 2 $\times$ 2) | (2 $\times$ 5 $\times$ 2) | (2 $\times$ 2 $\times$ 5) | (4 $\times$ 5) |
| RE-1  | 0.456 | 0.137            | 0.137            | 0.138            | 0.141     |
| RE-2  | 5.768 | 0.152            | 0.151            | 0.152            | 0.159     |
| PRIAL-1 | 0     | 0.690            | 0.691            | 0.688            | 0.681     |
| PRIAL-2 | 0     | 0.973            | 0.973            | 0.973            | 0.972     |
| size of LM | 0.004 | 0.043            | 0.050            | 0.038            | 0.038     |

|       | $\Sigma$ | $\tilde{\Sigma}$ | $\tilde{\Sigma}$ | LW04 | LW17 |
|-------|----------|-------------------|-------------------|------|------|
|       | (5 $\times$ 4) | (2 $\times$ 10) | (10 $\times$ 2) |      |      |
| RE-1  | 0.140    | 0.191            | 0.191            | 0.113| 0.127|
| RE-2  | 0.159    | 0.268            | 0.264            | 0.122| 0.145|
| PRIAL-1 | 0.684  | 0.575            | 0.575            | 0.744| 0.713|
| PRIAL-2 | 0.972  | 0.953            | 0.954            | 0.978| 0.974|
| size of LM | 0.041 | 0.035            | 0.028            | 0.074| 0.015|

The results are reported in Table 2. The first observation is that the performance of the quadratic form estimator $\tilde{\Sigma}$ is relatively robust to the Kronecker product factorization; the best performing one is 2 $\times$ 5 $\times$ 2. All the candidate estimators beat the sample covariance matrix $M_T$. In terms of RE-1 and RE-2, the LW04 and LW17 estimators are only slightly better than $\tilde{\Sigma}$ (2 $\times$ 5 $\times$ 2). In terms of PRIAL-1 and PRIAL-2, $\tilde{\Sigma}$ (2 $\times$ 5 $\times$ 2) is almost as good as the LW04 and LW17 estimators. In terms of the size of the LM test, $\tilde{\Sigma}$ (2 $\times$ 5 $\times$ 2) has the correct size while the LW04 estimator is slightly over-sized and the LW17 estimator is under-sized.

We next vary $\alpha^2$. We base the comparisons on the 2 $\times$ 5 $\times$ 2 Kronecker product factorization. The results are reported in Table 3. As $\alpha^2$ increases, the performance of $M_T$ actually improves in terms of RE-1 and RE-2. On the other hand, the performances of $\Sigma$, the LW04 and LW17 estimators worsen in terms of RE-1, RE-2, PRIAL-1 and PRIAL-2. The worsening performance of $\tilde{\Sigma}$ is not surprising because $\alpha^2$ can be interpreted as the distance of $\Sigma$ from a Kronecker product model. The worsening performance of the LW04 estimator has also been documented by Ledoit and Wolf (2004). The size of the LM test increases monotonically with $\alpha^2$ for both $\tilde{\Sigma}$ and the LW04 estimator. For the LW17 estimator, the size of the LM test first increases and then decreases with $\alpha^2$.

We next vary the ratio $n/T$. In the baseline setting we have $n/T = 0.5$. Here we consider two variations. The first variation is $n = 16, T = 50$ with a ratio of $n/T = 0.32$. The second variation is $n = 40, T = 20$ with a ratio of $n/T = 2$. For the first variation, we identify the Kronecker product factorizations: 2 $\times$ 2 $\times$ 2 $\times$ 2, 4 $\times$ 4, 4 $\times$ 2 $\times$ 2 and 2 $\times$ 8. For the second variation, we use the Kronecker product factorizations: 5 $\times$ 2 $\times$ 2 $\times$ 2, 5 $\times$ 2 $\times$ 4, 5 $\times$ 8 and 10 $\times$ 2 $\times$ 2. We also considered the permutations of submatrices for each factorization, but the performances remained almost unchanged, so we do not report them in the interest of space. The results are reported in Table 4.

Consider the top panel of Table 4 first. All the candidate estimators beat the sample covariance matrix $M_T$. The performance of the quadratic form estimator $\tilde{\Sigma}$ is relatively robust to the Kronecker product factorizations (2 $\times$ 2 $\times$ 2, 4 $\times$ 4 and 4 $\times$ 2 $\times$ 2); the best performing
Table 3: $M_T$, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, the quadratic form estimator (factorisations given in parentheses), the Ledoit and Wolf (2004)’s linear shrinkage estimator, and the Ledoit and Wolf (2017)’s direct nonlinear shrinkage estimator, respectively. RE-1 and RE-2 are the RE in terms of $\Sigma$ and $\Sigma^{-1}$, respectively. PRIAL-1 and PRIAL-2 are the PRIAL in terms of $\Sigma$ and $\Sigma^{-1}$, respectively.

One is $4 \times 2 \times 2$. In terms of RE-1, RE-2, PRIAL-1 and PRIAL-2, the quadratic form estimator $\tilde{\Sigma}$ ($4 \times 2 \times 2$) is only slightly worse than the LW04 and LW17 estimators. In terms of the size of the LM test, $\tilde{\Sigma}$ ($4 \times 2 \times 2$) has the correct size while the LW04 estimator is slightly over-sized and the LW17 estimator is under-sized.

Next consider the bottom panel of Table 4. All the candidate estimators beat the sample covariance matrix $M_T$ again. The best performing quadratic form estimator has a factorization ($5 \times 2 \times 2 \times 2$). In terms of RE-1, RE-2 and PRIAL-1, $\tilde{\Sigma}$ ($5 \times 2 \times 2 \times 2$) is comparable to the LW04 and LW17 estimators. In terms of the size of the LM test, $\tilde{\Sigma}$ ($5 \times 2 \times 2 \times 2$) has the correct size while the LW04 and LW17 estimators are over-sized.

By looking at Tables 2 and 4 together, we observe that as $n/T$ increases, PRIAL-1 increases monotonically for the best performing quadratic form estimator as well as the LW04 and LW17 estimators. Such a pattern is consistent with Ledoit and Wolf (2004). In terms of RE-1 and RE-2, the performances of the best performing quadratic form estimator as well as the LW04 and LW17 estimators worsen as $n/T$ increases. In terms of the size of the LM test, the best performing quadratic form estimator always has the correct size while that of the LW04 estimator increases monotonically with $n/T$.

7 Concluding Remarks

We remark on a number of possible extensions. One can generalize to the weakly dependent time series case, perhaps where the spectral density matrix is Kronecker product factored, but the limiting distributions become more complicated, see Hafner et al. (2018) for some work in this direction. We may consider the two sample case where $\Sigma_1 = \mathbb{E}(y_{1t} - \mu_1)(y_{1t} - \mu_1)^T$ and $\Sigma_2 = \mathbb{E}(y_{2t} - \mu_2)(y_{2t} - \mu_2)^T$, where $\mu_1 = \mathbb{E}(y_{1t})$ and $\mu_2 = \mathbb{E}(y_{2t})$. Cho and Phillips (2018) show that the hypothesis that $\Sigma_1 = \Sigma_2$ can be tested based on $\text{tr}(\Sigma_1 \Sigma_2^{-1}) = n$; if both matrices have
Table 4: \( M_T, \tilde{\Sigma}, \text{LW04} \text{ and LW17 stand for the sample covariance matrix, the quadratic form estimator (factorisations given in parentheses), the Ledoit and Wolf (2004)’s linear shrinkage estimator, and the Ledoit and Wolf (2017)’s direct nonlinear shrinkage estimator, respectively. RE-1 and RE-2 are the RE in terms of \( \Sigma \text{ and } \Sigma^{-1}, \text{ respectively. PRIAL-1 and PRIAL-2 are the PRIAL in terms of } \Sigma \text{ and } \Sigma^{-1}, \text{ respectively.} \)

A conformable Kronecker structure this simplifies to \( \text{tr}(\Sigma_{11}\Sigma_{21}^{-1}) \times \cdots \times \text{tr}(\Sigma_{1v}\Sigma_{2v}^{-1}) = n. \)

A Appendix

A.1 Proof of Lemma 4.1

Proof. For part (i), since \( \prod_{j=1}^{v} n_j = n, \) we have \( \left( \min_{1 \leq j \leq v} n_j \right)^{v} \leq n. \) Thus

\[ v \leq \log n / \log \left( \min_{1 \leq j \leq v} n_j \right) = O(\log n). \]

For part (ii):

\[ \max_{1 \leq j \leq v} \lambda_{\text{max}}(\Sigma_j) \leq \max_{1 \leq j \leq v} \text{tr}(\Sigma_j) = \max_{1 \leq j \leq v} n_j < \infty. \]

\[ \square \]

A.2 Proof of Theorem 4.1

We first give a few auxiliary lemmas leading to the proof of Theorem 4.1.

A.2.1 Lemma A.1

Lemma A.1. Suppose Assumptions 4.1 and 4.2 hold. Then we have

(i)

\[ \max_{1 \leq j \leq v} \frac{1}{\sigma_j^{2v}} \| \tilde{\Omega}_j - \Omega_j \|_F = O_p \left( \sqrt{\frac{\log n}{T}} \right). \]

(ii)

\[ \max_{1 \leq j \leq v} \left| \frac{\text{tr}(\tilde{\Omega}_j)}{n_j \sigma_j^{2v}} - \frac{\text{tr}(\Omega_j)}{n_j \sigma_j^{2v}} \right| = O_p \left( \sqrt{\frac{\log n}{T}} \right). \]
(iii) \( \min_{1 \leq j \leq v} |\text{tr}(\Omega_j) / (n_j \sigma_j^2)| = 1 \) and \( \min_{1 \leq j \leq v} |\text{tr}(\hat{\Omega}_j) / (n_j \sigma_j^2)| \) is bounded away from zero by an absolute positive constant with probability approaching 1 as \( n, T \to \infty \).

(iv) 
\[
\max_{1 \leq j \leq v} \| \Sigma_j \|_F \leq \max_{1 \leq j \leq v} \left[ \sqrt{n_j \lambda_{\max}(\Sigma_j)} \right] = O(1), \quad \max_{1 \leq j \leq v} \| \Sigma_j^{-1} \|_F = O(1).
\]
Moreover both \( \min_{1 \leq j \leq v} \| \Sigma_j \|_F \) and \( \min_{1 \leq j \leq v} \| \Sigma_j^{-1} \|_F \) are bounded away from zero by an absolute positive constant.

(v) 
\[
\max_{1 \leq j \leq v} \| \Sigma_j - \Sigma_j \|_F = O_p \left( \sqrt{\log n / T} \right).
\]

(vi) 
\[
\max_{1 \leq j \leq v} \| \Sigma_j^{-1} - \Sigma_j^{-1} \|_F = O_p \left( \sqrt{\log n / T} \right).
\]

**Proof of Lemma A.1.** For part (i),
\[
P \left( \max_{1 \leq j \leq v} \frac{\sqrt{\log n}}{\sigma_j^2} \frac{\| \hat{\Omega}_j - \Omega_j \|_F > M}{\sqrt{n_j \lambda_{\max}(\Sigma_j)}} \right) = \mathbb{P} \left( \bigcup_{1 \leq j \leq v} \left\{ \sqrt{n_j \lambda_{\max}(\Sigma_j)} \bigg\| \sqrt{\log n} \frac{\| \hat{\Omega}_j - \Omega_j \|_F > M}{\sqrt{n_j \lambda_{\max}(\Sigma_j)}} \bigg\} \right)
\]
\[
\leq \sum_{j=1}^v \mathbb{P} \left( \frac{\sqrt{\log n}}{\sigma_j^2} \frac{\| \hat{\Omega}_j - \Omega_j \|_F > M}{\sqrt{n_j \lambda_{\max}(\Sigma_j)}} \right) = \sum_{j=1}^v \frac{2T \mathbb{E} \| \text{vech}(\hat{\Omega}_j - \Omega_j) \|_F^2}{M^2 (\sigma_j^2)^2 \log n} = \sum_{j=1}^v \frac{2T \mathbb{E} \| \text{vech}(\hat{\Omega}_j - \Omega_j)^T \text{vech}(\hat{\Omega}_j - \Omega_j) \|_F^2}{M^2 (\sigma_j^2)^2 \log n}
\]
\[
\leq \sum_{j=1}^v \frac{2T \text{tr} \left( \frac{2G D_{n_j}^+ (\Omega_j \otimes \Omega_j) D_{n_j}^{+T}}{M^2 (\sigma_j^2)^2 \log n} \right)}{M^2 (\sigma_j^2)^2 \log n} = \sum_{j=1}^v \frac{4C \text{tr} \left( D_{n_j}^+ (\Sigma_j \otimes \Sigma_j) D_{n_j}^{+T} \right)}{M^2 \log n}
\]
\[
\leq \sum_{j=1}^v \frac{2C n_j (n_j + 1) \lambda_{\max}(\Sigma_j)}{M^2 \log n} \leq \frac{2C n_j (n_j + 1) \lambda_{\max}(\Sigma_j)}{M^2 \log n} \leq \frac{2C n_j (n_j + 1)}{M^2 \log n} \rightarrow 0
\]
as \( M, n, T \to \infty \), where the second inequality is due to Markov’s inequality, the fourth inequality is due to Lemma B.1, the sixth inequality is due to Lemma B.2, the seventh equality is due to Lemma B.3, and the convergence to zero is due to Lemma 4.1.

For part (ii),
\[
\max_{1 \leq j \leq v} \frac{| \text{tr}(\Omega_j) - n_j \sigma_j^2 |}{n_j \sigma_j^2} \leq \max_{1 \leq j \leq v} \frac{1}{n_j} \sum_{i=1}^{n_j} \left| \frac{1}{\sigma_j^2} (\hat{\Omega}_{j,ii} - \Omega_{j,ii}) \right| \leq \max_{1 \leq j \leq v} \max_{1 \leq i \leq n_j} \frac{1}{\sigma_j^2} | \hat{\Omega}_{j,ii} - \Omega_{j,ii} |
\]
\[
\leq \max_{1 \leq j \leq v} \frac{1}{\sigma_j^2} \| \hat{\Omega} - \Omega \|_\infty \leq \max_{1 \leq j \leq v} \frac{1}{\sigma_j^2} \| \hat{\Omega} - \Omega \|_F = O_p \left( \sqrt{\log n / T} \right),
\]
The first statement of part (iii) is trivial by recognising that \( \text{tr}(\Omega_j)/(n_j\sigma_j^{2s}) = 1 \) for any \( j = 1, \ldots, v \). Next

\[
\min_{1 \leq j \leq v} \left| \frac{\text{tr}(\hat{\Omega}_j)}{n_j\sigma_j^{2s}} \right| \geq \min_{1 \leq j \leq v} \left( \frac{\text{tr}(\hat{\Omega}_j)}{n_j\sigma_j^{2s}} - \frac{\text{tr}(\hat{\Omega}_j)}{n_j\sigma_j^{2s}} \right) \geq \min_{1 \leq j \leq v} \left| \frac{\text{tr}(\hat{\Omega}_j)}{n_j\sigma_j^{2s}} \right| - \max_{1 \leq j \leq v} \left| \frac{\text{tr}(\hat{\Omega}_j)}{n_j\sigma_j^{2s}} - \frac{\text{tr}(\hat{\Omega}_j)}{n_j\sigma_j^{2s}} \right| = 1 - O_p \left( \sqrt{\frac{\log n}{T}} \right).
\]

For part (iv),

\[
\lambda_{\min}(\Sigma_j) \leq \lambda_{\max}(\Sigma_j) \leq \|\Sigma_j\|_F \leq \sqrt{n_j} \lambda_{\max}(\Sigma_j)
\]

whence we deduce that \( \max_{1 \leq j \leq v} \|\Sigma_j\|_F \) is bounded from the above by an absolute positive constant and \( \min_{1 \leq j \leq v} \|\Sigma_j\|_F \) is bounded away from zero by an absolute positive constant via Assumption 4.2. Similarly, we have

\[
\frac{1}{\lambda_{\max}(\Sigma_j)} = \lambda_{\min}(\Sigma_j^{-1}) \leq \lambda_{\max}(\Sigma_j^{-1}) \leq \|\Sigma_j^{-1}\|_F \leq \sqrt{n_j} \lambda_{\max}(\Sigma_j^{-1}) = \sqrt{n_j} \frac{1}{\lambda_{\min}(\Sigma_j)},
\]

whence we deduce that \( \max_{1 \leq j \leq v} \|\Sigma_j^{-1}\|_F \) is bounded from the above by an absolute positive constant and \( \min_{1 \leq j \leq v} \|\Sigma_j^{-1}\|_F \) is bounded away from zero by an absolute positive constant via Assumption 4.2.

For part (v),

\[
\max_{1 \leq j \leq v} \|\Sigma_j - \Sigma\|_F = \max_{1 \leq j \leq v} \left\| \frac{\hat{\Omega}_j}{\sigma_j^{2s}} - \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F
\]

\[
= \max_{1 \leq j \leq v} \left\| \frac{\hat{\Omega}_j}{\sigma_j^{2s}} - \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} + \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} - \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F
\]

\[
\leq \max_{1 \leq j \leq v} \left\| \frac{\hat{\Omega}_j}{\sigma_j^{2s}} - \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F + \max_{1 \leq j \leq v} \left\| \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} - \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F
\]

We consider the first term of (A.1).

\[
\max_{1 \leq j \leq v} \left\| \frac{\hat{\Omega}_j}{\sigma_j^{2s}} - \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F = O_p(1) \max_{1 \leq j \leq v} \left\| \frac{\text{tr}(\hat{\Omega}_j)}{\sigma_j^{2s}} - \frac{\text{tr}(\hat{\Omega}_j)}{\sigma_j^{2s}} \right\|_F = O_p \left( \sqrt{\frac{\log n}{T}} \right)
\]

where the first equality is due to part (iii)-(iv), and the last equality is due to part (i). We consider the second term of (A.1).

\[
\max_{1 \leq j \leq v} \left\| \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F = \max_{1 \leq j \leq v} \left\| \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F
\]

\[
= \max_{1 \leq j \leq v} \left\| \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F = O_p(1) \max_{1 \leq j \leq v} \left\| \frac{\text{tr}(\hat{\Omega}_j)}{\text{tr}(\Omega_j)/(n_j\sigma_j^{2s})} \right\|_F = O_p \left( \sqrt{\frac{\log n}{T}} \right)
\]

where the third equality is due to part (iii), and the last equality is due to part (ii). Part (v) hence follows.

For part (vi), invoke Lemma B.5 and use that \( \max_{1 \leq j \leq v} \|\Sigma_j^{-1}\|_F = O(1) \).

\[\square\]
A.2.2 Lemma A.2

Lemma A.2. Suppose Assumption 4.1 hold. Then we have

\[ |\hat{\sigma}^2 - \sigma^2| = O_p\left(\sqrt{\frac{\log n}{T}}\right), \]

where \( \hat{\sigma}^2 := \text{tr}(M_T)/n \).

Proof of Lemma A.2. Write

\[ |\hat{\sigma}^2 - \sigma^2| = \frac{1}{n} \text{tr}(M_T - \Sigma) = \left| \frac{1}{n} \sum_{i=1}^{n} (M_{T,i,i} - \Sigma_{i,i}) \right| \leq \frac{1}{n} \sum_{i=1}^{n} |M_{T,i,i} - \Sigma_{i,i}| \]

\[
= \max_{1 \leq i \leq n} \left| \frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{t,i}^2 - \mathbb{E}y_{t,i}^2) \right|.
\]

Note that for \( i = 1, \ldots, n, m = 2, 3, \ldots \),

\[
= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}|y_{t,i}^2 - \mathbb{E}y_{t,i}^2|^m \leq \frac{1}{T} \sum_{t=1}^{T} 2^{m-1} (\mathbb{E}|y_{t,i}^2|^m + \mathbb{E}|\hat{y}_{t,i}^2|^m) \leq \frac{1}{T} \sum_{t=1}^{T} 2^{m-1} (\mathbb{E}|y_{t,i}^2|^m + \mathbb{E}|\hat{y}_{t,i}^2|^m) = 2^m \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}|y_{t,i}^2|^m \leq 2^m A^m \leq 2 \cdot m! A^m = \frac{m!}{2} A^{m-2}(A^2 4)
\]

where the first inequality is due to Loeve’s \( c_r \) inequality, the third inequality is due to Assumption 4.1(ii). Now invoke the Bernstein’s inequality in Section B with \( \sigma_0^2 = (4A^2) \): For all \( \epsilon > 0 \)

\[
\mathbb{P}\left( \frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{t,i}^2 - \mathbb{E}y_{t,i}^2) \right) \geq \sigma_0^2 \left[ A\epsilon + \sqrt{2\epsilon} \right] \leq 2e^{-T\sigma_0^2}.\]

Invoking Corollary B.1 in Section B, we have

\[
\max_{1 \leq i \leq n} \left| \frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{t,i}^2 - \mathbb{E}y_{t,i}^2) \right| = O_p\left(\sqrt{\frac{\log n}{T}}\right).
\]

\[ \square \]

A.2.3 Proof of Theorem 4.1

Proof of Theorem 4.1. For part (i),

\[
\|\hat{\Sigma} - \Sigma\|_F / \|\Sigma\|_F = \left\| \delta^2 \times \hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \sigma^2 \times \Sigma_1 \otimes \cdots \otimes \Sigma_v \right\|_F / \|\Sigma\|_F = \left\| \delta^2 \times \hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \sigma^2 \times \Sigma_1 \otimes \cdots \otimes \Sigma_v + \delta^2 \times \Sigma_1 \otimes \cdots \otimes \Sigma_v - \sigma^2 \times \Sigma_1 \otimes \cdots \otimes \Sigma_v \right\|_F / \|\Sigma\|_F \leq \delta^2 \|\hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \Sigma_1 \otimes \cdots \otimes \Sigma_v\|_F / \|\Sigma\|_F + |\delta^2 - \sigma^2| \|\Sigma_1 \otimes \cdots \otimes \Sigma_v\|_F / \|\Sigma\|_F \tag{A.2}\]

We consider the first term in (A.2). By inserting terms like \( \Sigma_1 \otimes \hat{\Sigma}_2 \otimes \cdots \otimes \hat{\Sigma}_v \) and the triangular inequality, we have

\[
\|\hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \Sigma_1 \otimes \cdots \otimes \Sigma_v\|_F \leq \left\| \hat{\Sigma}_1 - \Sigma_1 \right\|_F \prod_{\ell=2}^{v} \|\hat{\Sigma}_\ell\|_F + \sum_{j=2}^{v-1} \left( \prod_{k=1}^{j-1} \|\Sigma_k\|_F \right) \left\| \hat{\Sigma}_j - \Sigma_j \right\|_F \left( \prod_{\ell=j+1}^{v} \|\hat{\Sigma}_\ell\|_F \right) + \sum_{k=1}^{v-1} \|\Sigma_k\|_F \left\| \hat{\Sigma}_v - \Sigma_v \right\|_F. \tag{A.3}\]
We first divide the first term of (A.3) by $\prod_{\ell=1}^v \|\Sigma_{\ell}\|_F$. We have

$$\frac{\|\hat{\Sigma}_1 - \Sigma_1\|_F \prod_{\ell=2}^v \|\hat{\Sigma}_\ell\|_F}{\prod_{\ell=1}^v \|\Sigma_{\ell}\|_F} = \frac{\|\hat{\Sigma}_1 - \Sigma_1\|_F \prod_{\ell=2}^v \|\hat{\Sigma}_\ell\|_F}{\|\Sigma_1\|_F \prod_{\ell=2}^v \|\Sigma_{\ell}\|_F} \leq \left(1 + \frac{\max_{1 \leq k \leq v} \|\hat{\Sigma}_k - \Sigma_k\|_F}{\min_{1 \leq k \leq v} \|\Sigma_k\|_F} \right)^{v-1}. \quad \text{(A.4)}$$

We next divide the summand of the second term of (A.3) by $\prod_{\ell=1}^v \|\Sigma_{\ell}\|_F$. We have for $j = 2, \ldots, v - 1$

$$\frac{\prod_{k=1}^{j-1} \|\Sigma_k\|_F}{\prod_{\ell=1}^v \|\Sigma_{\ell}\|_F} \frac{\|\hat{\Sigma}_j - \Sigma_j\|_F}{\|\Sigma_j\|_F} \prod_{\ell=j+1}^v \|\hat{\Sigma}_{\ell}\|_F \leq \frac{\|\hat{\Sigma}_j - \Sigma_j\|_F}{\|\Sigma_j\|_F} \prod_{\ell=j+1}^v \|\hat{\Sigma}_{\ell}\|_F \left(1 + \frac{\max_{1 \leq k \leq v} \|\hat{\Sigma}_k - \Sigma_k\|_F}{\min_{1 \leq k \leq v} \|\Sigma_k\|_F} \right)^{v-j}. \quad \text{(A.5)}$$

We finally divide the third term of (A.3) by $\prod_{\ell=1}^v \|\Sigma_{\ell}\|_F$. We have

$$\frac{\prod_{k=1}^{v-1} \|\Sigma_k\|_F}{\prod_{\ell=1}^v \|\Sigma_{\ell}\|_F} \frac{\|\hat{\Sigma}_v - \Sigma_v\|_F}{\|\Sigma_v\|_F} = \|\hat{\Sigma}_v - \Sigma_v\|_F. \quad \text{(A.6)}$$

Thus we have

$$\hat{\sigma}^2 \|\hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \Sigma_1 \otimes \cdots \otimes \Sigma_v\|_F / \|\Sigma\|_F \leq \hat{\sigma}^2 \sum_{j=1}^v \frac{\|\hat{\Sigma}_j - \Sigma_j\|_F}{\|\Sigma_j\|_F} \left(1 + \frac{\max_{1 \leq k \leq v} \|\hat{\Sigma}_k - \Sigma_k\|_F}{\min_{1 \leq k \leq v} \|\Sigma_k\|_F} \right)^{v-j} \leq \frac{\hat{\sigma}^2}{\sigma^2} \left(1 + \frac{\max_{1 \leq k \leq v} \|\hat{\Sigma}_k - \Sigma_k\|_F}{\min_{1 \leq k \leq v} \|\Sigma_k\|_F} \right)^{v-1} \sum_{j=1}^v \frac{\|\hat{\Sigma}_j - \Sigma_j\|_F}{\|\Sigma_j\|_F} = \frac{\hat{\sigma}^2}{\sigma^2} O_p(1) \sum_{j=1}^v \frac{\|\hat{\Sigma}_j - \Sigma_j\|_F}{\|\Sigma_j\|_F} = O_p(1) \sum_{j=1}^v \frac{\|\hat{\Sigma}_j - \Sigma_j\|_F}{\|\Sigma_j\|_F} = v O_p \left(\sqrt{\frac{\log n}{T}}\right) = O_p \left(\sqrt{\frac{\log^3 n}{T}}\right)$$

where the first inequality is due to that $\|\Sigma\|_F = \sigma^2 \prod_{j=1}^v \|\Sigma_j\|_F$ via Lemma B.4, (A.4), (A.5) and (A.6), the first equality is due to Lemma A.1(iv) and (v), the second equality is due to Lemmas A.2 and A.1(iv), and the third equality is due to Lemma A.1(v). We now consider the second term in (A.2).

$$|\hat{\sigma}^2 - \sigma^2| \|\hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v\|_F / \|\Sigma\|_F = \frac{|\hat{\sigma}^2 - \sigma^2|}{\sigma^2} O_p \left(\sqrt{\frac{\log n}{T}}\right)$$

where the last equality is due to Lemma A.2. Part (ii) could be established in a similar manner, so we omit the details. \(\square\)

### A.3 Proof of Theorem 4.2

We first give an auxiliary lemma leading to the proof of Theorem 4.2.

#### A.3.1 Lemma A.3

**Lemma A.3.** Suppose Assumptions 4.1(i), 4.2 and 4.3 hold. Then we have, for $j = 1, \ldots, v$, 22
\[(i)\]
\[
\max_{1 \leq j \leq v} \frac{1}{\sigma_j^2} \| \bar{\Omega}_j - \Omega_j \|_{\ell_2} = O_p \left( \sqrt{\frac{\log \log n}{T}} \right).
\]

\[(ii)\]
\[
\max_{1 \leq j \leq v} \| \bar{\Sigma}_j - \Sigma_j \|_{\ell_2} = O_p \left( \sqrt{\frac{\log n}{T}} \right).
\]

\[(iii)\]
\[
\max_{1 \leq j \leq v} \| \bar{\Sigma}_j^{-1} - \Sigma_j^{-1} \|_{\ell_2} = O_p \left( \sqrt{\frac{\log n}{T}} \right).
\]

**Proof of Lemma A.3.** For part (i), we have

\[
\max_{1 \leq j \leq v} \frac{1}{\sigma_j^2} \| \bar{\Omega}_j - \Omega_j \|_{\ell_2} = \max_{1 \leq j \leq v} \frac{1}{\sigma_j^2} \left\| \frac{1}{T} \sum_{t=1}^T \left( A_j y_t y_t^T A_j^T - A_j \mathbb{E}[y_t y_t^T] A_j^T \right) \right\|_{\ell_2}
\]

\[
\leq \max_{1 \leq j \leq v} \frac{1}{\sigma_j^2} 2 \max_{a \in \mathcal{N}_{1/4}} \left| a^T \left[ \frac{1}{T} \sum_{t=1}^T \left( A_j y_t y_t^T A_j^T - A_j \mathbb{E}[y_t y_t^T] A_j^T \right) \right] a \right|
\]

\[
= \max_{1 \leq j \leq v} 2 \max_{a \in \mathcal{N}_{1/4}} \left| \frac{1}{T} \sum_{t=1}^T \left( a^T w_{j,t} w_{j,t}^T a - \mathbb{E}[a^T w_{j,t} w_{j,t}^T] a \right) \right|
\]

where the first inequality is due to Lemma B.7 with \( \varepsilon = 1/4 \), and \( w_{j,t} := A_j y_t / \sqrt{\sigma_j^2} \), where \( a \in \mathbb{R}^{n_j} \) with \( \|a\|_2 = 1 \). (Strictly speaking, both \( a \) and \( \mathcal{N}_{1/4} \) depend on \( j \) but we suppress this dependence.)

Note that for \( j = 1, \ldots, v, m = 2, 3, \ldots, \)

\[
\frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[ \left| a^T w_{j,t} \right|^2 - \mathbb{E}[a^T w_{j,t}]^2 \right] \leq \frac{1}{T} \sum_{t=1}^T 2^{m-1} \left( \mathbb{E}[a^T w_{j,t}]^2 \right)^m + \mathbb{E}[a^T w_{j,t}]^2 \right) \leq 2^m \frac{1}{T} \sum_{t=1}^T \mathbb{E}[a^T w_{j,t}]^2 \leq 2^m A^m \leq 2 \cdot m! A^m
\]

where the first inequality is due to Loeve’s \( c \) inequality, the third inequality is due to Assumption 4.1(ii). Now invoke the Bernstein’s inequality in Section B with \( \sigma_0^2 = (4A^2) \): For all \( \varepsilon > 0 \)

\[
\mathbb{P} \left( \frac{1}{T} \sum_{t=1}^T \left( a^T w_{j,t} - \mathbb{E}[a^T w_{j,t}] \right)^2 \right) \geq \sigma_0^2 \left[ A \varepsilon + \sqrt{2 \varepsilon} \right] \leq 2e^{-T \sigma_0^2 \varepsilon}.
\]

Invoking Corollary B.1 in Appendix B, we have

\[
\max_{1 \leq j \leq v} 2 \max_{a \in \mathcal{N}_{1/4}} \left| \frac{1}{T} \sum_{t=1}^T \left( a^T w_{j,t} - \mathbb{E}[a^T w_{j,t}] \right)^2 \right| = O_p \left( \frac{\log [\mathcal{N}_{1/4} : v]}{T} \right) \vee \sqrt{\frac{\log [\mathcal{N}_{1/4} : v]}{T} \vee \log \frac{\max_{1 \leq j \leq v} n_j + \log \log n}{T}}.
\]

Invoking Lemma B.6, we have \( |\mathcal{N}_{1/4}| \leq 9^{n_j} \). Thus we have

\[
\max_{1 \leq j \leq v} \frac{1}{\sigma_j^2} \| \bar{\Omega}_j - \Omega_j \|_{\ell_2} \leq \max_{1 \leq j \leq v} 2 \max_{a \in \mathcal{N}_{1/4}} \left| \frac{1}{T} \sum_{t=1}^T \left( a^T w_{j,t} - \mathbb{E}[a^T w_{j,t}] \right)^2 \right|\]

\[
= O_p \left( \frac{\max_{1 \leq j \leq v} n_j + \log \log n}{T} \left( \max_{1 \leq j \leq v} n_j + \log \log n \right) \right) = O_p \left( \sqrt{\frac{\log \log n}{T}} \right).
\]
For part (ii),
\[
\max_{1 \leq j \leq v} \left\| \hat{\Sigma}_j - \Sigma_j \right\|_{\ell_2} = \max_{1 \leq j \leq v} \left\| \frac{\hat{\Omega}_j / \sigma_j^{2*}}{\text{tr}(\hat{\Omega}_j)/[n_j \sigma_j^{2*}]} - \frac{\Omega_j / \sigma_j^{2*}}{\text{tr}(\Omega_j)/[n_j \sigma_j^{2*}]} \right\|_{\ell_2} = \max_{1 \leq j \leq v} \left\| \frac{\hat{\Omega}_j / \sigma_j^{2*}}{\text{tr}(\hat{\Omega}_j)/[n_j \sigma_j^{2*}]} - \frac{\Omega_j / \sigma_j^{2*}}{\text{tr}(\Omega_j)/[n_j \sigma_j^{2*}]} \right\|_{\ell_2} + \max_{1 \leq j \leq v} \left\| \frac{\Omega_j / \sigma_j^{2*}}{\text{tr}(\Omega_j)/[n_j \sigma_j^{2*}]} - \frac{\hat{\Omega}_j / \sigma_j^{2*}}{\text{tr}(\hat{\Omega}_j)/[n_j \sigma_j^{2*}]} \right\|_{\ell_2} \tag{A.7}
\]
We consider the first term of (A.7),
\[
\max_{1 \leq j \leq v} \left\| \frac{\hat{\Omega}_j / \sigma_j^{2*}}{\text{tr}(\hat{\Omega}_j)/[n_j \sigma_j^{2*}]} - \frac{\Omega_j / \sigma_j^{2*}}{\text{tr}(\Omega_j)/[n_j \sigma_j^{2*}]} \right\|_{\ell_2} = O_p(1) \max_{1 \leq j \leq v} \frac{1}{\sigma_j^{2*}} \left\| \hat{\Omega}_j - \Omega_j \right\|_{\ell_2} = O_p \left( \sqrt{\frac{\log \log n}{T}} \right)
\]
where the first equality is due to Lemma A.1(iii) and the last equality is due to part (i). We consider the second term of (A.7),
\[
\max_{1 \leq j \leq v} \left\| \frac{\Omega_j / \sigma_j^{2*}}{\text{tr}(\Omega_j)/[n_j \sigma_j^{2*}]} - \frac{\hat{\Omega}_j / \sigma_j^{2*}}{\text{tr}(\hat{\Omega}_j)/[n_j \sigma_j^{2*}]} \right\|_{\ell_2} = O_p(1) \max_{1 \leq j \leq v} \frac{1}{n_j \sigma_j^{2*}} \left\| \text{tr}(\Omega_j) - \text{tr}(\hat{\Omega}_j) \right\|_{\ell_2} = O_p \left( \sqrt{\frac{\log n}{T}} \right)
\]
where the third equality is due to Lemma A.1 part (iii) and Lemma 4.1(ii), and the last equality is due to Lemma A.1 part (ii). Part (ii) hence follows.

For part (iii), invoke Lemma B.5 and use that \(\max_{1 \leq j \leq v} \| \Sigma_j^{-1} \|_{\ell_2} = \max_{1 \leq j \leq v} \lambda_{\min}(\Sigma_j) = O(1)\). \(\square\)

A.3.2 Proof of Theorem 4.2
Proof of Theorem 4.2. For part (i), we can repeat the same argument used in the proof of Theorem 4.1 to write
\[
\left\| \hat{\Sigma} - \Sigma \right\|_{\ell_2} / \|\Sigma\|_{\ell_2} \leq \hat{\sigma}^2 \left\| \hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \Sigma_1 \otimes \cdots \otimes \Sigma_v \right\|_{\ell_2} / \|\Sigma\|_{\ell_2} + |\hat{\sigma}^2 - \sigma^2| \left\| \Sigma_1 \otimes \cdots \otimes \Sigma_v \right\|_{\ell_2} / \|\Sigma\|_{\ell_2} \tag{A.8}
\]
We consider the first term in (A.8). By inserting terms like \(\Sigma_1 \otimes \hat{\Sigma}_2 \otimes \cdots \otimes \hat{\Sigma}_v\) and the triangular inequality, we have
\[
\left\| \hat{\Sigma}_1 \otimes \cdots \otimes \hat{\Sigma}_v - \Sigma_1 \otimes \cdots \otimes \Sigma_v \right\|_{\ell_2} \leq \\
\left\| \hat{\Sigma}_1 - \Sigma_1 \right\|_{\ell_2} \prod_{\ell=2}^v \|\hat{\Sigma}_\ell\|_{\ell_2} + \sum_{j=2}^{v-1} \left( \prod_{k=1}^{j-1} \|\Sigma_k\|_{\ell_2} \right) \|\hat{\Sigma}_j - \Sigma_j\|_{\ell_2} \left( \prod_{\ell=j+1}^v \|\hat{\Sigma}_\ell\|_{\ell_2} \right) + \left[ \prod_{k=1}^{v-1} \|\Sigma_k\|_{\ell_2} \right] \|\Sigma_v - \Sigma_v\|_{\ell_2}.
\]
Thus
\[
\hat{\sigma}^2 \|\Sigma_1 \otimes \cdots \otimes \hat{\Sigma}_v - \Sigma_1 \otimes \cdots \otimes \Sigma_v\|_{\ell_2}/\|\Sigma\|_{\ell_2} \leq \frac{\hat{\sigma}^2}{\sigma^2} O_P(1) \sum_{j=1}^v \frac{\|\hat{\Sigma}_j - \Sigma_j\|_{\ell_2}}{\|\Sigma_j\|_{\ell_2}}
\]
\[= O_P(1) \sum_{j=1}^v \frac{\|\hat{\Sigma}_j - \Sigma_j\|_{\ell_2}}{\|\Sigma_j\|_{\ell_2}} = v O_P \left( \sqrt{\frac{\log n}{nT}} \right) \]
where the first inequality follows similarly from that in the proof of Theorem 4.1, the first equality is due to Lemma A.2 and Assumption 4.2(ii), and the second equality is due to Lemma A.3(ii). We now consider the second term in (A.2).

\[
|\hat{\sigma}^2 - \sigma^2| \|\Sigma_1 \otimes \cdots \otimes \Sigma_v\|_{\ell_2}/\|\Sigma\|_{\ell_2} = \hat{\sigma}^2/\sigma^2 = O_P \left( \sqrt{\frac{\log n}{nT}} \right)
\]
where the last equality is due to Lemma A.2. Backing up, we have
\[\frac{\|\hat{\Sigma} - \Sigma\|_{\ell_2}}{\|\Sigma\|_{\ell_2}} = O_P \left( \sqrt{\frac{\log^2 n}{nT}} \right) .\]
Part (ii) could be established in a similar manner, so we omit the details. \(\square\)

### A.4 Proof of Theorem 4.3

We first give an auxiliary theorem leading to the proof of Theorem 4.3.

#### A.4.1 Theorem A.1

**Theorem A.1.** Suppose Assumptions 4.1, 4.2 and 4.4 hold. Then

(i) \[
\max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{h-n}} |\hat{d}_{i,j}^{(h)} - d_{i,j}^{(h)}| = O_P \left( \sqrt{\frac{\log n}{nT}} \right) .
\]

(ii) We have \(\text{tr}(d^{(h)})/(n_h n_{-h}) = \sigma^2 > 0\) for \(h = 1, \ldots, v\). Also,
\[
\max_{1 \leq h \leq v} \frac{1}{n_h n_{-h}} \frac{1}{n_{h-n}} \left| \text{tr}(d^{(h)}) - \text{tr}(\hat{d}^{(h)}) \right| = O_P \left( \sqrt{\frac{\log n}{nT}} \right) .
\]
As a result, \(\min_{1 \leq h \leq v} \text{tr}(\hat{d}^{(h)})/(n_h n_{-h})\) is bounded away from zero by an absolute positive constant in probability.

(iii) \[
\max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \left| \hat{\sigma}_{h;i,j} - \sigma_{h;i,j} \right| = O_P \left( \frac{\log n}{nT} \right) ,
\]
where \(\hat{\Sigma}_h =: (\hat{\sigma}_{h;i,j})\) and \(\Sigma_h =: (\sigma_{h;i,j})\).

(iv) \[
|\sigma^2 - \sigma^2| = O_P \left( \frac{1}{\sqrt{nT}} \right) .
\]

(v) \[
\max_{1 \leq h \leq v} \|\hat{\Sigma}_h - \Sigma_h\|_F = O_P \left( \frac{\log n}{nT} \right) .
\]
\begin{align}
\text{(vi)} & \\
& \max_{1 \leq h \leq v} \|\hat{\Sigma}_h^{-1} - \Sigma_h^{-1}\|_F = O_p\left(\sqrt{\frac{\log n}{nT}}\right). \\
\text{(vii)} & \\
& \max_{1 \leq h \leq v} \|\hat{\Sigma}_h - \Sigma_h\|_1 = O_p\left(\sqrt{\frac{\log n}{nT}}\right). \\
\text{(viii)} & \\
& \max_{1 \leq h \leq v} \|\hat{\Sigma}_h^{-1} - \Sigma_h^{-1}\|_1 = O_p\left(\sqrt{\frac{\log n}{nT}}\right).
\end{align}

Proof. For part (i), note that $d_{i,j}^{(h)} = \sigma^2 \sigma_{h(i,j)} \text{tr}(\Sigma_{-h})$ and $d_{i,j}^{(h)} = \text{tr}(M_{T,h(i,j)})$, where $M_{T,h(i,j)}$ is the $(i,j)$th block of $M_T^{(h)}$ (each block is $n_{-h} \times n_{-h}$ dimensional). Note that $E[d_{i,j}^{(h)}] = d_{i,j}^{(h)}$.

Write for some $M > 0$

$$
P\left(\max_{1 \leq h \leq v} \max_{1 \leq i,j \leq n_h} \left| \frac{nT}{\log n n_{-h}} \left[ d_{i,j}^{(h)} - d_{i,j} \right] \right| > M \right) = P\left(\bigcup_{1 \leq h \leq v} \bigcup_{1 \leq i,j \leq n_h} \left\{ \left| \frac{nT}{\log n n_{-h}} \left[ d_{i,j}^{(h)} - d_{i,j} \right] \right| > M \right\} \right) \leq \sum_{h=1}^v \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} P\left(\left| \frac{nT}{\log n n_{-h}} \left[ d_{i,j}^{(h)} - d_{i,j} \right] \right| > M \right) \leq \frac{vn^2 nT \max_{1 \leq h \leq v} \max_{1 \leq i,j \leq n_h} \text{var}(d_{i,j}^{(h)} / n_{-h})}{\log n \cdot M^2}
$$

where the second inequality is due to Chebyshev’s inequality. Thus part (i) would follow if we show that

$$
\max_{1 \leq h \leq v} \max_{1 \leq i,j \leq n_h} \text{var}(d_{i,j}^{(h)} / n_{-h}) = O\left(\frac{1}{nT}\right).
$$

We now show this. For arbitrary $i, j = 1, \ldots, n_h$

$$
\text{var}(d_{i,j}^{(h)} / n_{-h}) = \frac{1}{n_{-h}} \text{var} \left( \sum_{k=1}^{n_h} \left[ M_{T,h(i,j)} \right]_{kk} \right) = \frac{1}{n_{-h}} \text{var} \left( \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^{n_{-h}} y_{t(i-1)n_{-h}+k} y_{t(j-1)n_{-h}+k} \right) \\
= \frac{1}{n_{-h}^2 T} \sum_{k=1}^{n_{-h}} \sum_{t=1}^T \text{cov} \left( y_{t(i-1)n_{-h}+k} y_{t(j-1)n_{-h}+k}, y_{t(i-1)n_{-h}+\ell} y_{t(j-1)n_{-h}+\ell} \right) \\
\leq \frac{C}{n_{-h}^2 T} \sum_{k=1}^{n_{-h}} \sum_{t=1}^T \text{cov} \left( z_{t(i-1)n_{-h}+k} z_{t(j-1)n_{-h}+k}, z_{t(i-1)n_{-h}+\ell} z_{t(j-1)n_{-h}+\ell} \right),
$$

(A.9)

where $y_t^{(h)} := K_{n_h \times \cdots \times n_{-h} + 1} y_t$ such that $E[y_t^{(h)} y_t^{(h)\top}] = \Sigma^{(h)}$ and $z_t^{(h)}$ is to be interpreted similarly. The third equality is due to independence over $t$ of $y_t$ in Assumption 4.1(i), and the first inequality is due to Assumption 4.1(iii). Using Lemma 9 of Magnus and Neudecker (1986), we have

$$
\text{var} \left( \text{vec}(z_t^{(h)} z_t^{(h)\top}) \right) = \text{var} \left( z_t^{(h)} \otimes z_t^{(h)} \right) = 2D_n D_n^\top (\Sigma^{(h)} \otimes \Sigma^{(h)}) = (I_n^2 + K_{n,n}) (\Sigma^{(h)} \otimes \Sigma^{(h)}),
$$

where the last equality is due to (33) of Magnus and Neudecker (1986). Thus we recognise that the summand on the right side of (A.9) is some element of $(I_n^2 + K_{n,n}) (\Sigma^{(h)} \otimes \Sigma^{(h)})$. We need to determine the exact position of the summand on the right side of (A.9) in $(I_n^2 + K_{n,n}) (\Sigma^{(h)} \otimes \Sigma^{(h)})$. We consider $\Sigma^{(h)} \otimes \Sigma^{(h)}$ and $K_{n,n} (\Sigma^{(h)} \otimes \Sigma^{(h)})$ separately. Consider $\Sigma^{(h)} \otimes \Sigma^{(h)}$ first. We now introduce a new way to locate an element in a matrix. Divide the $n^2 \times n^2$ matrix $\Sigma^{(h)} \otimes \Sigma^{(h)}$.
into \( n \times n \) blocks of matrices, each of which is \( n \times n \) dimensional. Then \( (\Sigma^{(h)} \otimes \Sigma^{(h)})\{[x,w],[p,q]\} \) refers to the \([p,q]th\) element of the \([x,w]th\) block matrix of \(\Sigma^{(h)} \otimes \Sigma^{(h)}\), where \(x,w,p,q = 1, \ldots , n\). It is not difficult to see that

\[
\text{cov}\left(z_{x, w}^{(h)}z_{t, (i-1)n - h + k}^{(h)}z_{t, (j-1)n - h + k}^{(h)}z_{t, (i-1)n - h + \ell}^{(h)}z_{t, (j-1)n - h + \ell}^{(h)}\right)
\]

corresponds to

\[
(\Sigma^{(h)} \otimes \Sigma^{(h)})\{ ([i-1]n - h + k, [i-1]n - h + \ell) , ([j-1]n - h + k, [j-1]n - h + \ell) \}.
\]  

(A.10)

We now consider \(K_{n,n}(\Sigma^{(h)} \otimes \Sigma^{(h)})\). It is important to recognise that \(K_{n,n}\) is a permutation matrix. Left multiplication of \(\Sigma^{(h)} \otimes \Sigma^{(h)}\) by \(K_{n,n}\) permutes the rows of \(\Sigma^{(h)} \otimes \Sigma^{(h)}\). Since \(K_{n,n}\) is \(n \times n\), we can also divide \(K_{n,n}\) into \(n \times n\) blocks of matrices, each of which is \(n \times n\) dimensional. Since \(K_{n,n}\) is also a permutation matrix, its elements can only be either 0 or 1. It is not difficult to see that, for arbitrary \(x, w = 1, \ldots , n\), the \([q,p]th\) element of the \([p,q]th\) block matrix of \(K_{n,n}\) is 1 for \(p, q = 1, \ldots , n\); all other elements of \(K_{n,n}\) are 0. Switch back to the traditional way to locate an element in a matrix. For \(p, q = 1, \ldots , n\), \([K_{n,n}]_{[p-1]n+q, [q-1]n+p} = 1\). This implies that the \(((p-1)n + q)th\) row of \(K_{n,n}(\Sigma^{(h)} \otimes \Sigma^{(h)})\) is actually the \(((q-1)n + p)th\) row of \(\Sigma^{(h)} \otimes \Sigma^{(h)}\). Switch back to the new way to locate an element in a matrix. This says that, for arbitrary \(x, w = 1, \ldots , n\), the \([q,x]th\) element of the \([p,w]th\) block matrix of \(K_{n,n}(\Sigma^{(h)} \otimes \Sigma^{(h)})\) is the \([p,x]th\) element of the \([q,w]th\) block matrix of \(\Sigma^{(h)} \otimes \Sigma^{(h)}\). Thus

\[
\text{cov}\left(z_{x, w}^{(h)}z_{t, (i-1)n - h + k}^{(h)}z_{t, (j-1)n - h + k}^{(h)}z_{t, (i-1)n - h + \ell}^{(h)}z_{t, (j-1)n - h + \ell}^{(h)}\right)
\]

corresponds to

\[
[K_{n,n}(\Sigma^{(h)} \otimes \Sigma^{(h)})]\{ ([i-1]n - h + k, [i-1]n - h + \ell) , ([j-1]n - h + k, [j-1]n - h + \ell) \} = (\Sigma^{(h)} \otimes \Sigma^{(h)})\{ ([j-1]n - h + k, [i-1]n - h + \ell) , ([i-1]n - h + k, [j-1]n - h + \ell) \}.
\]  

(A.11)
Using (A.10) and (A.11), we have

\[
\max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \text{var}(\hat{d}_{i,j}^{(h)}/n_{-h})
\]

\[
= \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} \text{cov} \left( y_{t,i}^{(h)} y_{t,j}^{(h)} y_{t,(i-1)n_{-h}+k}^{(h)} y_{t,(j-1)n_{-h}+\ell}^{(h)} \right)
\]

\[
\leq \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{C}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} \text{cov} \left( z_{t,i}^{(h)} z_{t,j}^{(h)} z_{t,(i-1)n_{-h}+k}^{(h)} z_{t,(j-1)n_{-h}+\ell}^{(h)} \right)
\]

\[
= \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} \left[ (\Sigma^{(h) \otimes \Sigma^{(h)}}) \right] \left\{ \left[ (i-1)n_{-h}+k, (j-1)n_{-h}+\ell \right] \right\}
\]

+ \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} \left[ (\Sigma^{(h) \otimes \Sigma^{(h)}}) \right] \left\{ \left[ (j-1)n_{-h}+k, (i-1)n_{-h}+\ell \right] \right\}
\]

\[
= \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} \left[ (\Sigma^{(h)}) \right] \left\{ \left[ (i-1)n_{-h}+k, (i-1)n_{-h}+\ell \right] \right\}
\]

+ \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} \left[ (\Sigma^{(h)}) \right] \left\{ \left[ (j-1)n_{-h}+k, (j-1)n_{-h}+\ell \right] \right\}
\]

\[
= \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} \left[ \sigma_{h,(i,i)} \cdot [\Sigma_{h}]_{k,\ell} + \sigma_{h,(j,j)} \cdot [\Sigma_{h}]_{k,\ell} + \sigma_{h,(i,j)} \cdot [\Sigma_{h}]_{k,\ell} \cdot [\Sigma_{h}]_{k,\ell} \right]
\]

\[
= \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \left( \sigma_{h,(i,i)} \cdot [\Sigma_{h}]_{k,\ell} + \sigma_{h,(j,j)} \cdot [\Sigma_{h}]_{k,\ell} \right) \frac{1}{n_{-h}^T} \sum_{k=1}^{n_h} \sum_{\ell=1}^{n_{-h}} [\Sigma_{h}]_{k,\ell}^2
\]

\[
\leq \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \left( \sigma_{h,(i,i)} \cdot [\Sigma_{h}]_{k,\ell} + \sigma_{h,(j,j)} \cdot [\Sigma_{h}]_{k,\ell} \right) \frac{1}{n_{-h}^T} \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{2\sigma_{h,(i,i)} \sigma_{h,(j,j)} [\Sigma_{h}]_{k,\ell} [\Sigma_{h}]_{k,\ell}}{n_{-h}^T} \leq 1
\]

\[
\leq \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{O(1)}{n_{-h}^T} \|\Sigma_{h}\|_F^2 = \max_{1 \leq h \leq v} \frac{O(1)}{n_{-h}^T} \frac{1}{n} \left( \frac{n}{n_{-h}^T} \|\Sigma_{h}\|_F^2 \|\Sigma_{h}\|_F^2 \right) = O \left( \frac{1}{n_{-h}^T} \right)
\]

\[
= O \left( \frac{1}{n_{-h}^T} \right),
\]

where the first inequality is due to Cauchy-Schwarz inequality \(\sigma_{h,(i,i)} \leq \sqrt{\sigma_{h,(i,i)}^2} \sqrt{\sigma_{h,(j,j)}^2} \) using the fact that \(\Sigma_{h}\) is a covariance matrix, the second inequality uses the fact that \(\max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \sigma_{h,(i,i)} \leq \max_{1 \leq h \leq v} \lambda_{\max}(\Sigma_{h}) < \infty\), the second last equality is due to Lemma A.1(iv), and the last equality is due to Assumption 4.4.

For part (ii), note that for \(h = 1, \ldots, v\)

\[
\text{tr}(d_{i,j}^{(h)}/(n_h n_{-h})) = \frac{1}{n_{-h}} \sigma^2 \text{tr}(\Sigma_{-h}) = \frac{1}{n_{-h}} \sigma^2 \text{tr}(\Sigma_{h+1}) \times \cdots \times \text{tr}(\Sigma_v) \times \text{tr}(\Sigma_1) \times \cdots \times \text{tr}(\Sigma_{h-1}) = \sigma^2 > 0.
\]

Now write

\[
\max_{1 \leq h \leq v} \frac{1}{n_{-h}} \left| \text{tr}(d_{i,j}^{(h)}) - \text{tr}(d_{i,j}^{(h)}) \right| = \max_{1 \leq h \leq v} \frac{1}{n_{-h}} \sum_{i=1}^{n_h} \left| d_{i,i}^{(h)} - d_{i,i}^{(h)} \right| \leq \max_{1 \leq h \leq v} \frac{1}{n_{-h}} \sum_{i=1}^{n_h} \left| d_{i,i}^{(h)} - d_{i,i}^{(h)} \right|
\]

\[
\leq \max_{1 \leq h \leq v} \frac{1}{n_{-h}} \left| d_{i,i}^{(h)} - d_{i,i}^{(h)} \right| = O_p \left( \frac{\log n}{nT} \right),
\]

where the last equality is due to part (i).
For part (iii), write
\[
\max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} |\hat{\sigma}_{i,j}^{(h)} - \sigma_{i,j}^{(h)}| \leq \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \left| \frac{\hat{d}_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} - \frac{d_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} \right| + \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \left| \frac{d_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} - \frac{d_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} \right|
\]
(A.12)

Consider the first term on the right side of (A.12).
\[
\max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \left| \frac{\hat{d}_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} - \frac{d_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} \right| = \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{\text{tr}(d^{(h)})/(n_h n_{-h})} \frac{1}{n_{-h}} |\hat{d}_{i,j}^{(h)} - d_{i,j}^{(h)}| = O_p(1) \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{1}{n_{-h}} |\hat{d}_{i,j}^{(h)} - d_{i,j}^{(h)}| = O_p \left( \sqrt{\frac{\log n}{nT}} \right),
\]
where the second equality is due to part (ii) and the last equality is due to part (i). Consider the second term on the right side of (A.12).
\[
\max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \left| \frac{d_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} - \frac{d_{i,j}^{(h)}}{\text{tr}(d^{(h)})/n_h} \right| = \max_{1 \leq h \leq v} \max_{1 \leq i, j \leq n_h} \frac{\text{tr}(d^{(h)})}{n_h n_{-h}} \frac{\text{tr}(d^{(h)})}{n_h n_{-h}} \frac{1}{n_{-h}} |\sigma_{i,j}^{(h)}| = O_p \left( \sqrt{\frac{\log n}{nT}} \right) \lambda_{\max}(\Sigma_h) = O_p \left( \sqrt{\frac{\log n}{nT}} \right),
\]
where the second equality is due to part (ii), and the last equality is due to Lemma 4.1(ii). Part (iii) hence follows.

For part (iv), note that \( \hat{\sigma}^2 = \text{tr}(M_T)/n \) and \( \sigma^2 = \text{tr}(\Sigma)/n \) so \( \mathbb{E}[\hat{\sigma}^2] = \sigma^2 \). Consider arbitrary \( \varepsilon > 0 \),
\[
P \left( |\hat{\sigma}^2 - \sigma^2| > \varepsilon \right) = P \left( \left| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{t,i}^2 - \mathbb{E} y_{t,i}^2 \right| > \varepsilon \right) \leq \text{var} \left( \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{t,i}^2 \right) \varepsilon^2 \text{var} \left( \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{t,i}^2 \right) = O(1/nT).
\]
Part (iv) would follow if we show \( \text{var} \left( \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{t,i}^2 \right) = O(1/nT) \). We now show this
\[
\text{var} \left( \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{t,i}^2 \right) = \frac{1}{T} \text{var} \left( \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{t,i}^2 \right) = \frac{1}{T n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov} \left( y_{t,i} y_{t,j} \right) \leq \frac{C}{T n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov} \left( z_{t,i} z_{t,j} \right) = \frac{C}{T n^2} \sum_{i=1}^n \sum_{j=1}^n \left( (\Sigma \otimes \Sigma)_{i,j} + (K_{n,n} (\Sigma \otimes \Sigma))_{i,j} \right) = 2 \frac{C}{T n^2} \sum_{i=1}^n \sum_{j=1}^n \Sigma_{i,j} = 2 \frac{C}{T n^2} \|\Sigma\|_F^2 = O \left( \frac{1}{T n} \right)
\]
where the first equality is due to independence over \( t \) of Assumption 4.1, the third and fourth equalities are due to the similar arguments which we used in part (i), and the last equality is due to Assumption 4.4.
For part (v), we have
\[
\max_{1 \leq h \leq v} \| \tilde{\Sigma}_h - \Sigma_h \|_F = \max_{1 \leq h \leq v} \left( \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} [\tilde{\sigma}_{h(i,j)} - \sigma_{h(i,j)}]^2 \right) \leq \max_{1 \leq h \leq v} \sqrt{n_h^2 \max_{1 \leq i,j \leq n_h} [\tilde{\sigma}_{h(i,j)} - \sigma_{h(i,j)}]^2}
\]
\[
= \max_{1 \leq h \leq v} \max_{1 \leq i,j \leq n_h} n_h k |\tilde{\sigma}_{h(i,j)} - \sigma_{h(i,j)}| = O_p \left( \sqrt{\frac{\log n}{nT}} \right)
\]
where the last equality is due to part (iii).

For part (vi), invoke Lemma B.5 and use that \(\max_{1 \leq h \leq v} \| \Sigma_h^{-1} \|_F = O(1)\) in Lemma A.1(iii).

For part (vii), we have
\[
\max_{1 \leq h \leq v} \| \tilde{\Sigma}_h - \Sigma_h \|_1 \leq \max_{1 \leq h \leq v} n_h \| \tilde{\Sigma}_h - \Sigma_h \|_F = O_p \left( \sqrt{\frac{\log n}{nT}} \right).
\]

For part (viii), we have
\[
\max_{1 \leq h \leq v} \| \tilde{\Sigma}_h^{-1} - \Sigma_h^{-1} \|_1 \leq \max_{1 \leq h \leq v} n_h \| \tilde{\Sigma}_h^{-1} - \Sigma_h^{-1} \|_F = O_p \left( \sqrt{\frac{\log n}{nT}} \right).
\]

\[\square\]

A.4.2 Proof of Theorem 4.3

Proof of Theorem 4.3. The proofs for parts (i)-(ii) are exactly the same as that of Theorem 4.1 using Theorem A.1 (v) and (vi). The proofs for parts (iii)-(iv) are also exactly the same as that of Theorem 4.1 using Theorem A.1 (vii) and (viii) and Lemma B.4(iii).

\[\square\]

A.5 Proof of Theorem 4.4

We first give an auxiliary theorem and an auxiliary lemma leading to the proof of Theorem 4.4.

A.5.1 Theorem A.2

Theorem A.2. Suppose Assumptions 4.1(i), 4.2, 4.3 and 4.5(ii) hold. Then for every \(n\) and \(j, k = 1, \ldots, v\) \((j \neq k)\), as \(T \to \infty\),

(i)
\[
\sqrt{T} \frac{1}{\sigma_j^2} \text{vech}(\tilde{\Omega}_j - \Omega_j) \overset{d}{\to} N \left( 0, D_{n_j}^+ \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{var}(w_{j,t} \otimes w_{j,t}) \right] D_{n_j}^{++} \right) =: N(0, V_{j,j}),
\]

where \(w_{j,t}\) is defined in Assumption 4.3.

(ii)
\[
\sqrt{T} \left[ \frac{1}{\sigma_j^2} \text{vech}(\tilde{\Omega}_j - \Omega_j) \right] \overset{d}{\to} N(0, V),
\]

where \(V\) is defined in Theorem 4.4.
Proof of Theorem A.2. For part (i), consider for any $n_j(n_j + 1)/2 \times 1$ non-zero vector $a$ with $\|a\|_2 = 1$,\(^6\)

\[
\frac{\sqrt{T}}{\sigma_{j,2}} a^\intercal \text{vech}(\hat{\Lambda}_j - \Omega_j) = \frac{\sqrt{T}}{\sigma_{j,2}} a^\intercal \text{vech} (A_j(M_T - \Sigma)A_j^\intercal) = \sum_{t=1}^{T} \frac{T^{-1/2}}{\sigma_{j,2}} a^\intercal \text{vech} (A_j(y_t y_t^\intercal - E[y_t y_t^\intercal])A_j^\intercal)
\]

\[
= \sum_{t=1}^{T} T^{-1/2} a^\intercal (w_{j,t} w_{j,t}^\intercal - E[w_{j,t} w_{j,t}^\intercal]) =: \sum_{t=1}^{T} U_{n,T,t},
\]

where $w_{j,t} := \frac{1}{\sqrt{\sigma_{j,2}^2}} A_j y_t$ with $E[w_{j,t}] = 0$ and $\text{var}(w_{j,t}) = \Sigma_j$. It is easy to see that $E[U_{n,T,t}] = 0$ and 

\[
\sum_{t=1}^{T} \text{var}(U_{n,T,t}) = \sum_{t=1}^{T} \text{var}(U_{n,T,t}) = \sum_{t=1}^{T} \text{var} \left( T^{-1/2} a^\intercal (w_{j,t} w_{j,t}^\intercal - E[w_{j,t} w_{j,t}^\intercal]) \right)
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \text{var} \left( a^\intercal (w_{j,t} w_{j,t}^\intercal) \right) = a^\intercal D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^{T} \text{var}(w_{j,t} w_{j,t}^\intercal) \right] D_{n_j}^{+\intercal} a
\]

\[
= a^\intercal D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^{T} \text{var}(w_{j,t} \otimes w_{j,t}) \right] D_{n_j}^{+\intercal} a =: s_{n,T}^2.
\]

Thus we consider $\sum_{t=1}^{T} \xi_{n,T,t} := \sum_{t=1}^{T} \frac{U_{n,T,t}}{s_{n,T}^2}$. To establish $\sum_{t=1}^{T} \xi_{n,T,t} \overset{d}{\rightarrow} \mathcal{N}(0,1)$, we just need to verify the Lyapounov’s condition in Theorem B.1 part (a) in Appendix B: For some $\delta > 0$,

\[
\lim_{T \to \infty} \sum_{t=1}^{T} \frac{1}{s_{n,T}^{2+\delta}} \text{var}(U_{n,T,t})^{2+\delta} = 0. \quad (A.13)
\]

We first investigate at what rate the denominator $(s_{n,T}^2)^{1+\delta/2}$ goes to zero. Note that

\[
s_{n,T}^2 = a^\intercal D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^{T} \text{var}(w_{j,t} \otimes w_{j,t}) \right] D_{n_j}^{+\intercal} a \geq \lambda_{\min} \left( D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^{T} \text{var}(w_{j,t} \otimes w_{j,t}) \right] D_{n_j}^{+\intercal} \right)
\]

\[
\geq C \lambda_{\min} \left( D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^{T} \text{var}(z_{j,t} \otimes z_{j,t}) \right] D_{n_j}^{+\intercal} \right) = C \lambda_{\min} \left( D_{n_j}^+ \left[ \frac{1}{T} \sum_{t=1}^{T} 2D_{n_j} D_{n_j}^+ (\Sigma_j \otimes \Sigma_j) D_{n_j}^{+\intercal} \right] D_{n_j}^{+\intercal} \right)
\]

\[
= C \lambda_{\min} \left( D_{n_j}^+ 2D_{n_j} D_{n_j}^+ (\Sigma_j \otimes \Sigma_j) D_{n_j}^{+\intercal} \right) = 2C \lambda_{\min} \left( D_{n_j}^+ (\Sigma_j \otimes \Sigma_j) D_{n_j}^{+\intercal} \right)
\]

\[
\geq 2C \lambda_{\min} (D_{n_j}^+ D_{n_j}^{+\intercal}) \lambda_{\min}(\Sigma_j \otimes \Sigma_j) = C \lambda_{\min}(\Sigma_j \otimes \Sigma_j) = C \lambda_{\min}^2(\Sigma_j) > 0,
\]

where the second inequality is due to Assumption 4.5(ii), the first equality is due to Magnus and Neudecker (1986) Lemma 9, the third inequality is due to Lemma B.2, the fourth equality is due to Lemma B.3, and the last inequality is due to Assumption 4.2(ii). We conclude that the denominator $(s_{n,T}^2)^{1+\delta/2}$ is bounded away from zero by an absolute positive constant.

\(^6\)Strictly speaking $a$ depends on $j$ but we suppress this dependence.
To verify (A.13), it suffices to prove $\lim_{T \to \infty} \sum_{t=1}^{T} \mathbb{E}|U_{n,T,t}|^{2+\delta} = 0$. Write

$$
\sum_{t=1}^{T} \mathbb{E}|U_{n,T,t}|^{2+\delta} = \sum_{t=1}^{T} \mathbb{E}|T^{-1/2}a^\top \text{vech} \left( w_{j,t}w_{j,t}^\top - \mathbb{E}[w_{j,t}w_{j,t}^\top]\right)|^{2+\delta} \\
= \sum_{t=1}^{T} T^{-\frac{2+\delta}{2}} \mathbb{E}|a^\top \text{vech} \left( w_{j,t}w_{j,t}^\top\right) - a^\top \text{vech} \left( \mathbb{E}[w_{j,t}w_{j,t}^\top]\right)|^{2+\delta} \\
\leq \sum_{t=1}^{T} T^{-\frac{2+\delta}{2}} 2^{1+\delta} \left( \mathbb{E}|a^\top \text{vech} \left( w_{j,t}w_{j,t}^\top\right)|^{2+\delta} + \mathbb{E}|a^\top \text{vech} \left( \mathbb{E}[w_{j,t}w_{j,t}^\top]\right)|^{2+\delta}\right)
$$

(A.14)

where the inequality is due to Loeve’s $c_r$ inequality. We consider the first term in the parenthesis of (A.14).

$$
\mathbb{E}|a^\top \text{vech} \left( w_{j,t}w_{j,t}^\top\right)|^{2+\delta} \leq ||a||_2 \left\| \text{vech} \left( w_{j,t}w_{j,t}^\top\right) \right\|_2^{2+\delta} = \mathbb{E}||D_{n_j}^+ \text{vech} \left( w_{j,t}w_{j,t}^\top\right)||_2^{2+\delta} \\
\leq \mathbb{E}\left[ \left| \max_{1 \leq k, \ell \leq n_j} \left| w_{j,t,k}w_{j,t,\ell} \right| \right|^{2+\delta} \right] \leq n_j^{2+\delta} \mathbb{E}\left[ \sum_{k=1}^{n_j} \sum_{\ell=1}^{n_j} \left| w_{j,t,k}w_{j,t,\ell} \right|^{2+\delta} \right] \\
= n_j^{2+\delta} \left[ \max_{1 \leq k, \ell \leq n_j} \left( |w_{j,t,k}|^{1+2\delta} \right)^{1/2} \left( |w_{j,t,\ell}|^{1+2\delta} \right)^{1/2} \right] = n_j^{4+\delta} \max_{1 \leq k, \ell \leq n_j} \mathbb{E}|w_{j,t,k}|^{4+2\delta},
$$

(A.15)

where the second equality is due to Lemma B.3, and the last inequality is due to Cauchy-Schwarz inequality.

We consider the second term in the parenthesis of (A.14).

$$
\mathbb{E}|a^\top \text{vech} \left( \mathbb{E}[w_{j,t}w_{j,t}^\top]\right)|^{2+\delta} \leq ||a||_2 \left\| \text{vech} \left( \mathbb{E}[w_{j,t}w_{j,t}^\top]\right) \right\|_2^{2+\delta} = \mathbb{E}||D_{n_j}^+ \text{vech} \left( \mathbb{E}[w_{j,t}w_{j,t}^\top]\right)||_2^{2+\delta} \\
\leq ||D_{n_j}^+ \|_{\ell_2} \left\| \text{vech} \left( \mathbb{E}[w_{j,t}w_{j,t}^\top]\right) \right\|_2^{2+\delta} = \left\| \mathbb{E}[w_{j,t}w_{j,t}^\top]\right\|_F^{2+\delta} = ||\Sigma_j||_F^{2+\delta} \leq K
$$

(A.16)

for some absolute positive constant $K$, where the last inequality is due to Lemma A.1(iv). Substituting (A.15) and (A.16) into (A.14), we have

$$
\sum_{t=1}^{T} \mathbb{E}|U_{n,T,t}|^{2+\delta} \leq \sum_{t=1}^{T} T^{-\frac{2+\delta}{2}} 2^{1+\delta} \max_{1 \leq k \leq n_j} \mathbb{E}|w_{j,t,k}|^{4+2\delta} + \sum_{t=1}^{T} T^{-\frac{2+\delta}{2}} 2^{1+\delta} K
$$

$$
= T^{-\frac{\delta}{2}} 2^{1+\delta} n_j^{4+\delta} \left[ \frac{1}{T} \sum_{t=1}^{T} \max_{1 \leq k \leq n_j} \mathbb{E}|w_{j,t,k}|^{4+2\delta} \right] + T^{-\frac{\delta}{2}} 2^{1+\delta} K \to 0
$$

as $T \to \infty$, where the convergence is due to Assumption 4.3.

Thus we have

$$
\sqrt{T} \frac{1}{\sigma_j^2} a^\top \text{vech}(\Omega_j - \Omega_j) = \sum_{t=1}^{T} U_{n,T,t} \xrightarrow{d} N \left( 0, a^\top D_{n_j}^+ \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{var} \left( w_{j,t} \otimes w_{j,t} \right) \right] D_{n_j}^+ a \right),
$$

whence we have, via Cramer-Wold device,

$$
\sqrt{T} \frac{1}{\sigma_j^2} \text{vech}(\Omega_j - \Omega_j) \xrightarrow{d} N \left( 0, D_{n_j}^+ \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{var} \left( w_{j,t} \otimes w_{j,t} \right) \right] D_{n_j}^+ \right).
$$
For part (ii), consider any \([n_j(n_j+1)/2 + n_k(n_k+1)/2] \times 1\) non-zero vector \(a = (a_1^T, a_2^T)^T\), where \(a_1\) is \(n_j(n_j+1)/2 \times 1\) and \(a_2\) is \(n_k(n_k+1)/2 \times 1\).

\[
\sqrt{T}a^T \begin{bmatrix}
\text{vech}(\hat{\Omega}_j - \Omega_j)/\sigma_j^{2*}
\text{vech}(\hat{\Omega}_k - \Omega_k)/\sigma_k^{2*}
\end{bmatrix}
= \sqrt{T}a^T \begin{bmatrix}
\text{vech} \left( A_j \left( \frac{1}{T} \sum_{t=1}^T [y_{jt}^T - \mathbb{E}[y_{jt}^T]] A_j^T \right) / \sigma_j^{2*} \right)
\text{vech} \left( A_k \left( \frac{1}{T} \sum_{t=1}^T [y_{kt}^T - \mathbb{E}[y_{kt}^T]] A_k^T \right) / \sigma_k^{2*} \right)
\end{bmatrix}
= \sum_{t=1}^T T^{-1/2}a^T \begin{bmatrix}
\text{vech} \left( w_{jt}w_{jt}^T - \mathbb{E}[w_{jt}w_{jt}^T]\right)
\text{vech} \left( w_{kt}w_{kt}^T - \mathbb{E}[w_{kt}w_{kt}^T]\right)
\end{bmatrix}
=: \sum_{t=1}^T U_{n,T,t}.
\]

It is easy to see that \(\mathbb{E}U_{n,T,t} = 0\). We calculate \(\text{var}(U_{n,T,t})\). In particular, we have

\[
\text{cov} \left\{ \left( w_{jt}w_{jt}^T - \mathbb{E}[w_{jt}w_{jt}^T]\right), \left( w_{kt}w_{kt}^T - \mathbb{E}[w_{kt}w_{kt}^T]\right) \right\}
= D_{nj}^+ \text{cov} \left\{ \left( w_{jt}w_{jt}^T - \mathbb{E}[w_{jt}w_{jt}^T]\right), \left( w_{kt}w_{kt}^T - \mathbb{E}[w_{kt}w_{kt}^T]\right) \right\} D_{nj}^+ =: V_{t,j,k}.
\]

Thus, we have

\[
\sum_{t=1}^T \text{var}(U_{n,T,t}) = \frac{1}{T} \sum_{t=1}^T a^T \begin{bmatrix}
V_{t,j,j}
V_{t,k,k}
\end{bmatrix} a =: a^T V_T a,
\]

where

\[
V_{t,j,j} := D_{nj}^+ \text{var}(w_{jt} \otimes w_{jt}) D_{nj}^{++},
\]

for \(j = 1, \ldots, v\).

It is possible that \(V_T\) is positive semi-definite only, but our proof is robust to this. Suppose there exist non-zero \(\{a^*\}\) such that \(\sum_{t=1}^T \text{var}(U_{n,T,t}) = 0\). Then by Chebyshev’s inequality and temporal independence, we have \(\sum_{t=1}^T U_{n,T,t} = 0\) almost surely. Trivially,

\[
\sqrt{T}a^* \begin{bmatrix}
\text{vech}(\hat{\Omega}_j - \Omega_j)/\sigma_j^{2*}
\text{vech}(\hat{\Omega}_k - \Omega_k)/\sigma_k^{2*}
\end{bmatrix}
\xrightarrow{d} 0 = a^* V_T a^*.
\]

For other \(a \in \mathbb{R}^{[n_j(n_j+1)/2 + n_k(n_k+1)/2] \times 1 \setminus \{a^* \cup 0\}}\), we consider

\[
\sum_{t=1}^T \xi_{n,T,t} \xrightarrow{d} N(0, 1),
\]

we just need to verify the Lyapounov’s condition in Theorem B.1 part(a): for some \(\delta > 0\),

\[
\lim_{T \to \infty} \sum_{t=1}^T \frac{1}{(a^T V_T a)^{1+\delta/2}} \mathbb{E}|U_{n,T,t}|^{2+\delta} = 0.
\]

Since we have already ruled out the trivial case \(a^* V_T a^* = 0\),\(^7\) it suffices to prove \(\lim_{T \to \infty} \sum_{t=1}^T \mathbb{E}|U_{n,T,t}|^{2+\delta} = 0\). We can recycle the proof in part (i) to get

\[
\sqrt{T}a^T \begin{bmatrix}
\text{vech}(\hat{\Omega}_j - \Omega_j)/\sigma_j^{2*}
\text{vech}(\hat{\Omega}_k - \Omega_k)/\sigma_k^{2*}
\end{bmatrix}
= \sum_{t=1}^T U_{n,T,t} \xrightarrow{d} N \left(0, a^T \left( \lim_{T \to \infty} V_T \right) a\right),
\]

whence, together with the trivial case, we have via Cramer-Wold device,

\[
\sqrt{T} \begin{bmatrix}
\text{vech}(\hat{\Omega}_j - \Omega_j)/\sigma_j^{2*}
\text{vech}(\hat{\Omega}_k - \Omega_k)/\sigma_k^{2*}
\end{bmatrix}
\xrightarrow{d} N(0, V),
\]

where \(V := \lim_{T \to \infty} V_T\).

\(^7\)Another pathological case is that \(a^T V_T a^*\) drifts to zero at some rate. This could be handled in a similar manner as the \(a^* V_T a^* = 0\) case.
A.5.2 Lemma A.4

Lemma A.4. Suppose Assumptions 4.1(i), 4.2, 4.4 and 4.5(i) hold. Then as \( n, T \to \infty \),

(i) \[
\max_{1 \leq j \leq n} |\tilde{\omega}_j - \omega_j| = O_p \left( \sqrt{\frac{\log n}{nT}} \right).
\]

(ii) \[
\max_{1 \leq j \leq n} \left| \frac{\sigma^2_j - \tilde{\sigma}^2_j}{\sigma^2_j} \right| = O_p \left( \sqrt{\frac{\log n}{nT}} \right).
\]

Proof. For part (i)

\[
\max_{1 \leq j \leq n} |\tilde{\omega}_j - \omega_j| = \max_{1 \leq j \leq n} \left| t'_{nj} \hat{\Sigma}_j t_{nj} - t'_{nj} \Sigma_j t_{nj} \right| = \max_{1 \leq j \leq n} \left| \sum_{i=1}^{n_j} \sum_{k=1}^{n_j} (\hat{\sigma}_{j;i,k} - \sigma_{j;i,k}) \right|
\]

\[
\leq \max_{1 \leq j \leq n} \sum_{i=1}^{n_j} \sum_{k=1}^{n_j} \|\hat{\sigma}_{j;i,k} - \sigma_{j;i,k}\| \leq \max_{1 \leq j \leq n} \max_{1 \leq i \leq k \leq n_j} n_j^2 |\hat{\sigma}_{j;i,k} - \sigma_{j;i,k}| = O_p \left( \sqrt{\frac{\log n}{nT}} \right),
\]

where the last equality is due to Theorem A.1(iii).

For part (ii)

\[
\left| \frac{\sigma^2_j - \tilde{\sigma}^2_j}{\sigma^2_j} \right| \leq \frac{\hat{\sigma}^2}{|\sigma^2_j|} |\tilde{\omega}_1 \times \cdots \times \tilde{\omega}_{j-1} \times \tilde{\omega}_{j+1} \times \cdots \times \tilde{\omega}_v - \omega_1 \times \cdots \times \omega_{j-1} \times \omega_{j+1} \times \cdots \times \omega_v|
\]

\[
+ \frac{|\hat{\sigma}^2 - \sigma^2|}{|\sigma^2_j|} |\omega_1 \times \cdots \times \omega_{j-1} \times \omega_{j+1} \times \cdots \times \omega_v|.
\]

(A.17)

We consider the first term on the right side of (A.17). For \( 1 < j < v \), by inserting terms like \( \omega_1 \times \tilde{\omega}_2 \times \cdots \times \tilde{\omega}_v \) and the triangular inequality, we have

\[
|\tilde{\omega}_1 \times \cdots \times \tilde{\omega}_{j-1} \times \tilde{\omega}_{j+1} \times \cdots \times \tilde{\omega}_v - \omega_1 \times \cdots \times \omega_{j-1} \times \omega_{j+1} \times \cdots \times \omega_v| \leq
\]

\[
|\tilde{\omega}_1 - \omega_1| \cdot \prod_{\ell=2}^{v} |\tilde{\omega}_\ell| + \sum_{p=2}^{v-1} \left[ \prod_{k=1}^{p-1} |\omega_k| \right] |\tilde{\omega}_p - \omega_p| \left( \prod_{\ell=p+1}^{v} |\tilde{\omega}_\ell| \right) + \left( \prod_{k=1}^{v-1} |\omega_k| \right) |\tilde{\omega}_v - \omega_v|.
\]

(A.18)

For \( j = 1 \), by inserting terms like \( \omega_2 \times \tilde{\omega}_3 \times \cdots \times \tilde{\omega}_v \) and the triangular inequality, we have

\[
|\tilde{\omega}_2 \times \cdots \times \tilde{\omega}_v - \omega_2 \times \cdots \times \omega_v| \leq
\]

\[
|\tilde{\omega}_2 - \omega_2| \cdot \prod_{\ell=3}^{v} |\tilde{\omega}_\ell| + \sum_{p=3}^{v-1} \left[ \prod_{k=2}^{p-1} |\omega_k| \right] |\tilde{\omega}_p - \omega_p| \left( \prod_{\ell=p+1}^{v} |\tilde{\omega}_\ell| \right) + \left( \prod_{k=2}^{v-1} |\omega_k| \right) |\tilde{\omega}_v - \omega_v|.
\]

(A.19)

For \( j = v \), by inserting terms like \( \omega_1 \times \tilde{\omega}_2 \times \cdots \times \tilde{\omega}_{v-1} \) and the triangular inequality, we have

\[
|\tilde{\omega}_1 \times \cdots \times \tilde{\omega}_{v-1} - \omega_1 \times \cdots \times \omega_{v-1}| \leq
\]

\[
|\tilde{\omega}_1 - \omega_1| \cdot \prod_{\ell=2}^{v-1} |\tilde{\omega}_\ell| + \sum_{p=2}^{v-1} \left[ \prod_{k=1}^{p-1} |\omega_k| \right] |\tilde{\omega}_p - \omega_p| \left( \prod_{\ell=p+1}^{v-1} |\tilde{\omega}_\ell| \right) + \left( \prod_{k=1}^{v-2} |\omega_k| \right) |\tilde{\omega}_{v-1} - \omega_{v-1}|.
\]

(A.20)
Thus, by observing (A.18), (A.19) and (A.20), we have
\[
\max_{1 \leq j \leq v} \frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2} |\tilde{\omega}_1 \times \cdots \times \tilde{\omega}_{j-1} \times \tilde{\omega}_{j+1} \times \cdots \times \tilde{\omega}_v - \omega_1 \times \cdots \times \omega_{j-1} \times \omega_{j+1} \times \cdots \times \omega_v| \\
\leq \max_{1 \leq j \leq v} O_p(1) \sum_{k=1 \neq j}^v \frac{|\tilde{\omega}_k - \omega_k|}{|\omega_k|} \leq O_p(1) \max_{1 \leq k \leq v} \frac{|\tilde{\omega}_k - \omega_k|}{|\omega_k|} = O_p\left(\sqrt{\frac{\log n}{nT}}\right)
\]
where the first inequality is due to the trick used in the proof of Theorem 4.1, and the last equality is due to part (i) and the fact \( \min_{1 \leq k \leq v} |\omega_k| > \min_{1 \leq k \leq v} |\omega_k|/n_k \geq \min_{1 \leq k \leq v} \lambda_{\min}(\Sigma_k) > 0 \).

We now consider the second term on the right side of (A.17).
\[
\frac{|\hat{\sigma}^2 - \sigma^2|}{|\hat{\sigma}^2|} |\omega_1 \times \cdots \times \omega_{j-1} \times \omega_{j+1} \times \cdots \times \omega_v| = \frac{|\hat{\sigma}^2 - \sigma^2|}{|\hat{\sigma}^2|} \prod_{k=1 \neq j}^v \frac{\omega_k}{\hat{\omega}_k} = \frac{|\hat{\sigma}^2 - \sigma^2|}{|\hat{\sigma}^2|} O_p(1) \\
= O_p\left(\sqrt{\frac{1}{nT}}\right)
\]
where the third equality is due to Theorem A.1(iv). The result hence follows. \(\square\)

### A.5.3 Proof of Theorem 4.4

**Proof of Theorem 4.4.** For part (i), write
\[
\sqrt{T} \text{vech}(\Sigma_j - \Sigma_j) = \sqrt{T} \text{vech} \left( \frac{\tilde{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right) + \sqrt{T} \text{vech} \left( \frac{\hat{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right). \tag{A.21}
\]

Theorem A.2 shows that the second term on the right side of (A.21) has an asymptotic distribution \( N \left(0, V_{j,j}\right) \). To prove the theorem, we need to show the first term on the right side of (A.21) is \( o_p(1) \). Write
\[
\sqrt{T} \left\| \text{vech} \left( \frac{\tilde{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right) \right\|_F \leq \sqrt{T} \left\| \frac{\tilde{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right\|_F = \left\| \Sigma_j \right\|_F + O_p\left(\sqrt{T} \frac{\log n}{n} \right) = O(1),
\]
where the first equality is due to Lemma A.1(i), and the last equality is due to Lemma A.1(iv).

We now show that \( \sqrt{T} \left| \frac{\sigma^2_j}{\sigma^2_j} - 1 \right| = o_p(1) \).
\[
\sqrt{T} \left| \frac{\sigma^2_j}{\sigma^2_j} - 1 \right| = \sqrt{T} \left| \frac{\hat{\sigma}^2_j - \bar{\sigma}^2_j}{\sigma^2_j} \right| = \sqrt{T} O_p\left(\sqrt{\frac{\log n}{nT}}\right) = o_p(1),
\]
where the second equality is due to Lemma A.4(ii). For part (ii),
\[
\sqrt{T} \left[ \text{vech}(\Sigma_j - \Sigma_j) \right] = \sqrt{T} \text{vech} \left( \frac{\tilde{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right) + \sqrt{T} \text{vech} \left( \frac{\hat{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right) + \sqrt{T} \text{vech} \left( \frac{\hat{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right)
\]
\[
= o_p(1) + \left( \frac{\tilde{\Omega}_j}{\sigma^2_j} - \frac{\bar{\Omega}_j}{\sigma^2_j} \right) + N(0, V),
\]

35
where the second equality is due to that the first term on the right side of (A.21) is \( o_p(1) \), and the last weak convergence is due to Theorem A.2(ii). \( \square \)

### A.5.4 Proof of Corollary 4.1

**Proof of Corollary 4.1.** For \( V_{j,j} \), we have

\[
V_{j,j} = D^+_{n_j} \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{var}(w_{j,t} \otimes w_{j,t}) \right] D^+_{n_j} = D^+_{n_j} \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} 2D_{nj} D^+_{nj} (\Sigma_j \otimes \Sigma_j) \right] D^+_{n_j}
\]

\[
= 2D^+_{nj} (\Sigma_j \otimes \Sigma_j) D^+_{nj}.
\]

For \( V_{j,k} \), we have

\[
V_{j,k} = D^+_{n_j} \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{cov} \left\{ \left( w_{j,t} \otimes w_{j,t} - \mathbb{E}[w_{j,t} \otimes w_{j,t}] \right), \left( w_{k,t} \otimes w_{k,t} - \mathbb{E}[w_{k,t} \otimes w_{k,t}] \right) \right\} \right] D^+_{n_k}
\]

\[
= \frac{1}{\sigma_j^2 \sigma_k^2} D^+_{nj} (A_j \otimes A_j) \left[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \text{var} \left( \text{vec}(y_t) \right) \right] (A_k^T \otimes A_k^T) D^+_{nk}
\]

\[
= \frac{1}{\sigma_j^2 \sigma_k^2} D^+_{nj} (A_j \otimes A_j) 2D_{nj} D^+_{nj} (\Sigma_j \otimes \Sigma_j) (A_k^T \otimes A_k^T) D^+_{nk} = \frac{2}{\sigma_j^2 \sigma_k^2} D^+_{nj} (A_j \otimes A_j) (\Sigma_j \otimes \Sigma_j) (A_k^T \otimes A_k^T) D^+_{nk}
\]

\[
= \frac{2}{\sigma_j^2 \sigma_k^2} D^+_{nj} (A_j \Sigma A_k^T \otimes A_j \Sigma A_k^T) D^+_{nk},
\]

where the second last equality is due to a slight generalization of Lemma 11 of Magnus and Neudecker (1986). \( \square \)

### A.6 Proof of Theorem 5.1

We first give an auxiliary lemma and an auxiliary theorem leading to the proof of Theorem 5.1.

#### A.6.1 Lemma A.5

**Lemma A.5.** Suppose Assumption 4.5(ii) hold. Then we have

\[
\max_{1 \leq i \leq n} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t,i} \right| = O_p(\sqrt{\log n}).
\]

**Proof.** Under Assumption 4.5(ii) and weak stationarity, we have, for \( i = 1, \ldots, n, m = 2, 3, \ldots, \)

\[
\frac{1}{T} \sum_{t=1}^{T} |y_{t,i}|^m \leq A^m \leq \frac{m!}{2} A^{m-2} A^2,
\]

for some absolute positive constant \( A \). Now invoke the Bernstein’s inequality in Section B with \( \sigma_0^2 = A^2 \): For all \( \epsilon > 0 \)

\[
P \left( \frac{1}{T} \sum_{t=1}^{T} |y_{t,i}| \geq \sigma_0^2 \left( A\epsilon + \sqrt{2\epsilon} \right) \right) \leq 2e^{-T \sigma_0^2 \epsilon}.
\]

36
Invoking Corollary B.1 in Section B, we have
\[
\max_{1 \leq i \leq n} \left| \frac{1}{T} \sum_{t=1}^{T} y_{t,i} \right| = O_p \left( \frac{\log n}{T} \vee \sqrt{\frac{\log n}{T}} \right) = O_p \left( \sqrt{\frac{\log n}{T}} \right).
\]
The lemma follows. \(\square\)

A.6.2 Theorem A.3

The following theorem is adapted from Theorem 1 of Kelejian and Prucha (2001).

**Theorem A.3.** Consider \( \{\varepsilon_{n,T,i} : 1 \leq i \leq n, n \geq 1, T \geq 1\} \) and \( Q_{n,T} := \sum_{i=1}^{n} \varepsilon_{n,T,i}^2 \). Suppose that

(i) \( \mathbb{E}[\varepsilon_{n,T,i}] = 0 \) for \( 1 \leq i \leq n, n \geq 1, T \geq 1 \). Furthermore, for each \( n \geq 1, T \geq 1 \), \( \varepsilon_{n,T,1}, \ldots, \varepsilon_{n,T,n} \) are (mutually) independent.

(ii) \( \limsup_{T \to \infty} \sup_{n \geq 1} \sup_{1 \leq i \leq n} \mathbb{E}[|\varepsilon_{n,T,i}|^{4+2\delta}] < \infty \), for some \( \delta > 0 \).

(iii) \( \liminf_{n,T \to \infty} \frac{1}{n} \text{var}(Q_{n,T}) \geq C > 0 \) for some absolute positive constant \( C \).

Then as \( n, T \to \infty \),
\[
\frac{Q_{n,T} - \mathbb{E}[Q_{n,T}]}{\sqrt{\text{var}(Q_{n,T})}} \xrightarrow{d} N(0, 1).
\]

**Proof.** We can calculate that
\[
\mathbb{E}[Q_{n,T}] = \mathbb{E}\left[ \sum_{i=1}^{n} \varepsilon_{n,T,i}^2 \right] = \sum_{i=1}^{n} \mathbb{E}[\varepsilon_{n,T,i}^2] =: \sum_{i=1}^{n} \sigma_{n,T,i}^2
\]
\[
Q_{n,T} - \mathbb{E}[Q_{n,T}] = \sum_{i=1}^{n} (\varepsilon_{n,T,i}^2 - \sigma_{n,T,i}^2) =: \sum_{i=1}^{n} Y_{n,T,i}^2
\]
\[
Y_{n,T,i}^2 = (\varepsilon_{n,T,i}^2 - \sigma_{n,T,i}^2)^2 = \varepsilon_{n,T,i}^4 + \sigma_{n,T,i}^4 - 2\varepsilon_{n,T,i}^2 \sigma_{n,T,i}^2
\]
\[
\mathbb{E}[Y_{n,T,i}^2] = \mathbb{E}[\varepsilon_{n,T,i}^4] - \sigma_{n,T,i}^4
\]
\[
\text{var}(Q_{n,T}) = \mathbb{E}\left[ \sum_{i=1}^{n} Y_{n,T,i}^2 \right] = \mathbb{E}\left[ \sum_{i=1}^{n} \sum_{j=1}^{n} Y_{n,T,i} Y_{n,T,j} \right] = \sum_{i=1}^{n} \mathbb{E}[Y_{n,T,i}^2].
\]

We now show that
\[
\frac{Q_{n,T} - \mathbb{E}[Q_{n,T}]}{\sqrt{\text{var}(Q_{n,T})}} = \sum_{i=1}^{n} \frac{Y_{n,T,i}}{\sqrt{\text{var}(Q_{n,T})}} \xrightarrow{d} N(0, 1)
\]
as \( n, T \to \infty \). This boils down to verifying the Lyapounov’s condition in Theorem B.1 part (b); that is, for some \( \delta > 0 \),
\[
\lim_{n,T \to \infty} \sum_{i=1}^{n} \frac{1}{\text{var}(Q_{n,T})^{1+\delta/2}} \mathbb{E}[|Y_{n,T,i}|^{2+\delta}] = 0.
\]
Let's first find an upper bound for \( \mathbb{E}|Y_{n,T,i}|^{2+\delta} \). We have

\[
\mathbb{E}|Y_{n,T,i}|^{2+\delta} = \mathbb{E}|\varepsilon_{n,T,i} - \sigma_{n,T,i}^{2}|^{2+\delta} \leq 2^{1+\delta} \left( \mathbb{E}|\varepsilon_{n,T,i}|^{2+\delta} + \mathbb{E}|\sigma_{n,T,i}^{2}|^{2+\delta} \right)
\]

for some absolute positive constant \( K \) for sufficiently large \( T \), where the first inequality is due to Loeve's \( c_p \) inequality, and the last inequality is due to the assumption (ii) of the theorem. Then we have

\[
\sum_{i=1}^{n} \frac{\mathbb{E}|Y_{n,T,i}|^{2+\delta}}{\text{var}(Q_{n,T})^{1+\delta/2}} \leq \frac{nK}{|n^{-1}\text{var}(Q_{n,T})|^{1+\delta/2}n^{1+\delta/2}} = \frac{K}{|n^{-1}\text{var}(Q_{n,T})|^{1+\delta/2}n^{1+\delta/2}} \to 0
\]
as \( n,T \to \infty \), where the convergence to 0 relies on the assumption (iii) of the theorem.

### A.6.3 Proof of Theorem 5.1

**Proof of Theorem 5.1.** Write

\[
\frac{LM_n,T - n}{\sqrt{2n}} = \frac{T\mu^T \Sigma^{-1} \mu - n}{\sqrt{2n}} = \frac{T\hat{\mu}^T \Sigma^{-1} \mu - n}{\sqrt{2n}} + \frac{T\hat{\mu}^T (\Sigma^{-1} - \Sigma^{-1}) \hat{\mu}}{\sqrt{2n}}.
\]

We first show that as \( n,T \to \infty \),

\[
\frac{T\hat{\mu}^T \Sigma^{-1} \mu - n}{\sqrt{2n}} \overset{d}{\to} N(0,1).
\] (A.22)

Under the assumption (a) of the theorem (i.e., Assumption 4.5(ii)) and \( H_0 \), using Assumption 4.1(i), we have \( \sum_{t=1}^{T} y_t \sim N(0,T \Sigma) \), whence we have \( \sqrt{T} \hat{\mu} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \sim N(0,\Sigma) \) for any \( T \). Then it is well-known that for any \( T \),

\[
T\hat{\mu}^T \Sigma^{-1} \mu \sim \chi_n^2.
\]

Then for any \( T \)

\[
\frac{T\hat{\mu}^T \Sigma^{-1} \mu - n}{\sqrt{2n}} \overset{d}{\to} N(0,1), \quad \text{as } n \to \infty.
\]

Obviously the result holds for \( n,T \to \infty \). The theorem would follow if we show that

\[
\frac{T\hat{\mu}^T (\Sigma^{-1} - \Sigma^{-1}) \hat{\mu}}{\sqrt{2n}} = o_p(1).
\]

We now show this.

\[
\frac{T|\hat{\mu}^T (\Sigma^{-1} - \Sigma^{-1}) \hat{\mu}|}{\sqrt{2n}} = \frac{1}{\sqrt{2n}} \left| \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t,i} \right) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t,j} \right) (\Sigma_{i,j}^{-1} - \Sigma_{i,j}^{-1}) \right|
\]

\[
\leq \frac{1}{\sqrt{2n}} \left( \max_{1 \leq i \leq n} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t,i} \right| \right)^2 \sum_{i=1}^{n} \sum_{j=1}^{n} |\Sigma_{i,j}^{-1} - \Sigma_{i,j}^{-1}| = \frac{1}{\sqrt{2n}} \left( \max_{1 \leq i \leq n} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{t,i} \right| \right)^2 \|\Sigma^{-1} - \Sigma^{-1}\|_1
\]

\[
= O_p \left( \frac{\log n}{\sqrt{n}} \right) \|\Sigma^{-1}\|_1 O_p \left( \sqrt{\frac{\log^3 n}{nT}} \right) = \frac{1}{n} \|\Sigma^{-1}\|_1 O_p \left( \sqrt{\frac{\log^3 n}{nT}} \right) = O_p \left( \sqrt{\frac{\log^5 n}{T}} \right) = o_p(1)
\]

38
where the fourth equality is due to Lemma A.5 and Theorem 4.3(iv), and the sixth equality is due to Assumption 5.1.

We now establish (A.22) using the assumption (b) of the theorem. Write

\[
\frac{T \hat{u}' \Sigma^{-1} \hat{u} - n}{\sqrt{2n}} = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \right)' \left(L^{-1}\right)' L^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \right) - n
\]

where

\[
= \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \right)' \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \right) - n
\]

Note that for each \( n \geq 1, T \geq 1, z_{n,T,1}, \ldots, z_{n,T,n} \) are (mutually) independent under assumption (b) of the theorem and Assumption 4.1(i). Under \( H_0 \),

\[
E[z_{n,T,i}] = E \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t,i} \right] = 0
\]

\[
\text{var}(z_{n,T}) = \text{var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \right) = I_n
\]

\[
E[Q_{n,T}] = E \left[ \sum_{i=1}^{n} \gamma_{n,T,i}^2 \right] = \sum_{i=1}^{n} E \left[ \gamma_{n,T,i}^2 \right] = n
\]

\[
\text{var}(Q_{n,T}) = \text{var} \left( \sum_{i=1}^{n} \gamma_{n,T,i}^2 \right) = \sum_{i=1}^{n} \text{var} \left( \gamma_{n,T,i}^2 \right) = \sum_{i=1}^{n} \left[ E[\gamma_{n,T,i}^4] - \left( E[\gamma_{n,T,i}^2] \right)^2 \right]
\]

\[
= \sum_{i=1}^{n} \left( E[\gamma_{n,T,i}^4] - 1 \right) = \sum_{i=1}^{n} (\gamma_{z,i} + 2)
\]

where \( \gamma_{z,i} \) is the excess kurtosis of \( z_{n,T,i} \):

\[
\gamma_{z,i} := \frac{E[\gamma_{n,T,i}^4]}{|\text{var}(z_{n,T,i})|^2} - 3 = E[\gamma_{n,T,i}^4] - 3.
\]

We next calculate \( E[z_{n,T,i}^4] \) in terms of moments of \( x_{t,i} \).

\[
E[z_{n,T,i}^4] = E \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t,i} \right)^4 \right] = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{k=1}^{T} \sum_{\ell=1}^{T} E \left[ x_{t,i} x_{s,i} x_{k,i} x_{\ell,i} \right]
\]

(A.23)

Note that the summand in (A.23) is non-zero only if \( t = s = k = \ell, t = s \neq k = \ell, t = k \neq s = \ell, \) and \( t = \ell \neq k = s \). First, consider the case \( t = s = k = \ell \). Collecting all the summands in (A.23) satisfying this, we have

\[
\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{k=1}^{T} \sum_{\ell=1}^{T} E \left[ x_{t,i} x_{s,i} x_{k,i} x_{\ell,i} \right] = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} (\gamma_{x,t,i} + 3) = \frac{1}{T^2} \sum_{t=1}^{T} \gamma_{x,t,i} + 3 \frac{T}{T}
\]

(A.24)

where \( \gamma_{x,t,i} \) is the excess kurtosis of \( x_{t,i} \):

\[
\gamma_{x,t,i} := \frac{E \left[ x_{t,i}^4 \right]}{|\text{var}(x_{t,i})|^2} - 3 = E \left[ x_{t,i}^4 \right] - 3.
\]

Second, consider the case \( t = s \neq k = \ell \). Collecting all the summands in (A.23) satisfying this, we have

\[
\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{k=1}^{T} \sum_{\ell \neq t} \sum_{\ell \neq t} E \left[ x_{t,i}^2 \right] E \left[ x_{k,i}^2 \right] E \left[ x_{t,i}^2 \right] = \frac{T(T-1)}{T^2} = 1 - \frac{1}{T}
\]

(A.25)
Likewise for cases $t = k \neq s = \ell$ and $t = \ell \neq k = s$, both sums are $1 - 1/T$. Substituting (A.24) and (A.25) into (A.23), we have

$$\mathbb{E}[z_{n,T,i}^4] = \frac{1}{T^2} \sum_{t=1}^{T} \gamma_{x,t,i} + \frac{3}{T} + 3 \left(1 - \frac{1}{T}\right) = \frac{1}{T^2} \sum_{t=1}^{T} \gamma_{x,t,i} + 3$$

whence we have $\gamma_{z,i} = \mathbb{E}[z_{n,T,i}^4] - 3 = \frac{1}{T^2} \sum_{t=1}^{T} \gamma_{x,t,i}$ and

$$\text{var}(Q_{n,T}) = \sum_{i=1}^{n} (\gamma_{z,i} + 2) = \sum_{i=1}^{n} \left(\frac{1}{T^2} \sum_{t=1}^{T} \gamma_{x,t,i} + 2\right) = 2n \left(1 + \frac{1}{2T} \sum_{i=1}^{n} \sum_{t=1}^{T} \gamma_{x,t,i}\right).$$

It remains to verify condition (ii)-(iii) of Theorem A.3. We have

$$\frac{1}{n} \text{var}(Q_{n,T}) = 2 + \frac{1}{T} \left(\frac{nT}{\sum_{i=1}^{n} \sum_{t=1}^{T} \gamma_{x,t,i}}\right) > 0$$

for large enough $T$ because $\gamma_{x,t,i} > -3$ for all $t$ and $i$ by definition of the excess kurtosis. Hence (iii) of Theorem A.3 is satisfied. Condition (ii) of Theorem A.3 is also satisfied: for some $\delta > 0$

$$\limsup_{T \to \infty} \sup_{n \geq 1} \sup_{1 \leq i \leq n} \mathbb{E} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t,i} \right|^{4+2\delta} < \infty$$

because $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{t,i} \overset{d}{\to} N(0,1)$ for all $1 \leq i \leq n$ and $n \geq 1$ as $T \to \infty$. Thus we have

$$\frac{T \hat{\mu} \Sigma^{-\frac{1}{2}} \hat{\mu} - n}{\sqrt{2n}} = \frac{Q_{n,T} - n}{\sqrt{2n}} = \frac{Q_{n,T} - n}{\sqrt{2n} \left(1 + \frac{1}{4T} \sum_{i=1}^{n} \sum_{t=1}^{T} \gamma_{x,t,i}\right)} + o_p(1) \overset{d}{\to} N(0,1),$$

under $H_0$ as $n, T \to \infty$, where the second equality is due to the assumption $\limsup_{n,T \to \infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E}[x_{t,i}^4] < \infty$, and the weak convergence is due to Theorem A.3.

## B Auxiliary Lemmas

This section contains auxiliary lemmas and theorems which have been used in Appendix.

**Lemma B.1.** Suppose Assumption 4.1 hold. Then we have

$$\text{var} \left(\text{vech}(\hat{\Omega}_j)\right) \leq \frac{2C}{T} D_{n_j}^+ (\Omega_j \otimes \Omega_j) D_{n_j}^{+\top},$$

for some absolute positive constant $C$, where $\leq$ is to be interpreted componentwise.

**Proof.**

$$\text{var} \left(\text{vech}(\hat{\Omega}_j)\right) = \text{var} \left(D_{n_j}^+ (A_j \otimes A_j) \text{vec} M_T\right) = \text{var} \left(D_{n_j}^+ (A_j \otimes A_j) \text{vec} \left(\frac{1}{T} \sum_{t=1}^{T} y_t y_t^\top\right)\right)$$

by independence, (i.e. $y_t$ over $t$), the first inequality is due to Assumption 4.1(iii), the fourth equality is due to Magnus and Neudecker (1986) Lemma 9, and the fifth equality is due to a slight generalization of Lemma 11 of Magnus and Neudecker (1986). \qed
We first give two central limit theorems for double-index \((n, T)\) processes.

**Theorem B.1.**

(a) Suppose \(Y_{n,T,t}\) is a random variable independent across \(1 \leq t \leq T\) for \(n \geq 1\) and \(T \geq 1\).
Assume that
\[
E[Y_{n,T,t}] = 0 \quad E[Y^2_{n,T,t}] = \sigma^2_{n,T,t}. 
\]
Define
\[
s^2_{n,T} := \sum_{t=1}^{T} \sigma^2_{n,T,t} \quad \xi_{n,T,t} := \frac{Y_{n,T,t}}{s_{n,T}}. 
\]
Suppose the following Lyapounov’s condition hold: For some \(\delta > 0\),
\[
\lim_{n,T \to \infty} \sum_{t=1}^{T} \frac{1}{s^2_{n,T}} E|Y_{n,T,t}|^{2+\delta} = 0. 
\]
Then as \(n, T \to \infty\)
\[
\sum_{t=1}^{T} \xi_{n,T,t} \xrightarrow{d} N(0,1). 
\]

(b) Suppose \(Y_{n,T,i}\) is a random variable independent across \(1 \leq i \leq n\) for \(n \geq 1\) and \(T \geq 1\).
Assume that
\[
E[Y_{n,T,i}] = 0 \quad E[Y^2_{n,T,i}] = \sigma^2_{n,T,i}. 
\]
Define
\[
s^2_{n,T} := \sum_{i=1}^{n} \sigma^2_{n,T,i} \quad \xi_{n,T,i} := \frac{Y_{n,T,i}}{s_{n,T}}. 
\]
Suppose the following Lyapounov’s condition hold: For some \(\delta > 0\),
\[
\lim_{n,T \to \infty} \sum_{i=1}^{n} \frac{1}{s^2_{n,T}} E|Y_{n,T,i}|^{2+\delta} = 0. 
\]
Then as \(n, T \to \infty\)
\[
\sum_{i=1}^{n} \xi_{n,T,i} \xrightarrow{d} N(0,1). 
\]

**Proof.** The proofs can be easily adapted from the Lyapounov’s condition for triangular arrays (cf. p362 Billingsley (1995)) \(\Box\)

**Theorem B.2** (Bernstein’s inequality). We let \(Z_1, \ldots, Z_T\) be independent random variables, satisfying for positive constants \(A\) and \(\sigma^2_0\)
\[
E[Z_t] = 0 \quad \forall t, \quad \frac{1}{T} \sum_{t=1}^{T} E|Z_t|^m \leq \frac{m!}{2} A^{m-2} \sigma^2_0, \quad m = 2, 3, \ldots. 
\]
Let \(\epsilon > 0\) be arbitrary. Then
\[
P\left(\left|\frac{1}{T} \sum_{t=1}^{T} Z_t\right| \geq \sigma^2_0 \left[A \epsilon + \sqrt{2\epsilon}\right]\right) \leq 2e^{-T\sigma^2_0}. 
\]

**Proof.** Slightly adapted from Bühlmann and van de Geer (2011) p487. \(\Box\)

41
We can use Bernstein’s inequality to establish a rate for the maximum.

**Corollary B.1.** Suppose via Bernstein’s inequality that we have for $1 \leq i \leq n$

$$\mathbb{P}\left(\left|\frac{1}{T} \sum_{t=1}^{T} Z_{t,i}\right| \geq \sigma_0^2 K \epsilon + \sqrt{2\epsilon}\right) \leq 2e^{-T \sigma_0^2 \epsilon}.$$

Then

$$\max_{1 \leq i \leq n} \left|\frac{1}{T} \sum_{t=1}^{T} Z_{t,i}\right| = O_p\left(\frac{\log n}{T} \vee \sqrt{\frac{\log n}{T}}\right).$$

**Proof.** We need to use joint asymptotics $n, T \to \infty$. We shall use the preceding inequality with $\epsilon = (2 \log n)/(T \sigma_0^2)$. Fix $\epsilon > 0$. These exist $N_\epsilon := 2/\epsilon$, $T_\epsilon$, and $M_\epsilon := \max(4K, 4\sigma_0)$ such that for all $n > N_\epsilon$ and $T > T_\epsilon$ we have

$$\mathbb{P}\left(\max_{1 \leq i \leq n} \left|\frac{1}{T} \sum_{t=1}^{T} Z_{t,i}\right| \geq M_\epsilon \left(\frac{\log n}{T} \vee \sqrt{\frac{\log n}{T}}\right)\right) \leq \frac{n}{T} \sum_{i=1}^{n} \mathbb{P}\left(\left|\frac{1}{T} \sum_{t=1}^{T} Z_{t,i}\right| \geq \sigma_0^2 \left[K \epsilon + \sqrt{2\epsilon}\right]\right) \leq 2e^{\log n - 2\log n} = \frac{2}{n} < \epsilon.
$$

\[\square\]

**Lemma B.2.** Suppose matrix $A$ is real symmetric. Then for any comparable real matrix $B$

$$\lambda_{\min}(A) \lambda_{\min}(BB^\top) \leq \lambda_{\min}(BAB^\top) \leq \lambda_{\max}(BAB^\top) \leq \lambda_{\max}(A) \lambda_{\max}(BB^\top).$$

**Proof.** First, note that $BAB^\top$ is Hermitian. By Rayleigh-Ritz theorem, we have

$$\lambda_{\max}(BAB^\top) = \max_{||c||_2=1} c^\top BAB^\top c \leq \max_{||c||_2=1} \lambda_{\max}(A)||B^\top c||^2 = \lambda_{\max}(A) \max_{||c||_2=1} c^\top BB^\top c = \lambda_{\max}(A) \lambda_{\max}(BB^\top).$$

On the other hand,

$$\lambda_{\min}(BAB^\top) = \min_{||c||_2=1} c^\top BAB^\top c \geq \min_{||c||_2=1} \lambda_{\min}(A)||B^\top c||^2 = \lambda_{\min}(A) \min_{||c||_2=1} c^\top BB^\top c = \lambda_{\min}(A) \lambda_{\min}(BB^\top).$$

\[\square\]

**Lemma B.3.** Given the $n^2 \times (n(n+1)/2)$ duplication matrix $D_n$ and its Moore-Penrose generalized inverse $D_n^+ = (D_n^\top D_n)^{-1} D_n^\top$ (i.e., $D_n$ is full-column rank), we have

(i) \quad \lambda_{\max}(D_n^+ D_n^+\top) = 1 \quad \lambda_{\min}(D_n^+ D_n^+\top) = \frac{1}{2},

(ii) \quad ||D_n^+||_2 = 1.
Proof. First note that $D_n^TD_n$ is a diagonal matrix with diagonal entries either 1 or 2. Using the fact that for any real matrix $A$, $AA^T$ and $A^TA$ have the same non-zero eigenvalues, we have

$$\lambda_{\max}(D_n^TD_n^T) = \lambda_{\max}((D_n^TD_n)^{-1}) = 1$$
$$\lambda_{\min}(D_n^TD_n^T) = \lambda_{\min}((D_n^TD_n)^{-1}) = 1/2$$
$$\|D_n^+\|_2^2 = \lambda_{\max}(D_n^+D_n^+) = \lambda_{\max}(D_n^+D_n^+) = 1$$

\[\square\]

Lemma B.4. For any real matrices $A$ and $B$,

(i) $$\|A \otimes B\|_F = \|A\|_F \times \|B\|_F.$$  

(ii) $$\|A \otimes B\|_{\ell_2} = \|A\|_{\ell_2} \times \|B\|_{\ell_2}.$$  

(iii) $$\|A \otimes B\|_1 = \|A\|_1 \times \|B\|_1.$$  

Proof. For part (i),

$$\|A \otimes B\|_F^2 = \text{tr}[(A^T \otimes B^T)(A \otimes B)] = \text{tr}[A^TA \otimes B^TB] = \text{tr}(A^TA) \text{tr}(B^TB) = \|A\|_F^2 \|B\|_F^2.$$  

For part (ii),

$$\|A \otimes B\|_{\ell_2} = \sqrt{\maxe}[(A \otimes B)^T(A \otimes B)] = \sqrt{\maxe}[A^T \otimes B^T](A \otimes B)]$$

$$= \sqrt{\maxe}[A^TA \otimes B^TB] = \sqrt{\maxe}[A^TA] \maxe[B^TB] = \|A\|_{\ell_2} \|B\|_{\ell_2},$$

where the fourth equality is due to the fact that both $A^TA$ and $B^TB$ are symmetric and positive semidefinite. For part (iii), suppose that $A$ is $m \times n$ and $B$ is $p \times q$.

$$\|A \otimes B\|_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} (|a_{i,j}|\|B\|_1) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(|a_{i,j}| \sum_{k=1}^{p} \sum_{\ell=1}^{q} |b_{i,j}| \right) = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}| \right) \left(\sum_{k=1}^{p} \sum_{\ell=1}^{q} |b_{i,j}| \right)$$

$$= \|A\|_1 \|B\|_1.$$  

\[\square\]

Lemma B.5. Let $\hat{\Omega}_{n,j}$ and $\Omega_{n,j}$ be invertible (both possibly stochastic) $n \times n$ square matrices for $j = 1, \ldots, m$, where both $n$ and $m$ could be growing. Let $T$ be the sample size. For any matrix norm $\| \cdot \|$, suppose that $\max_{1 \leq j \leq m} \|\hat{\Omega}_{n,j}^{-1}\| = O_p(1)$ and $\max_{1 \leq j \leq m} \|\hat{\Omega}_{n,j} - \Omega_{n,j}\| = O_p(a_{m,n,T})$ for some sequence $a_{m,n,T}$ with $a_{m,n,T} \to 0$ as $m, n, T \to \infty$ simultaneously. Then $\max_{1 \leq j \leq m} \|\hat{\Omega}_{n,j}^{-1} - \Omega_{n,j}^{-1}\| = O_p(a_{m,n,T}).$

Proof. The original proof could be found in Saikkonen and Lutkepohl (1996) Lemma A.2.

$$\|\hat{\Omega}_{n,j}^{-1} - \Omega_{n,j}^{-1}\| \leq \|\hat{\Omega}_{n,j}^{-1}\| \|\Omega_{n,j} - \hat{\Omega}_{n,j}\| \|\Omega_{n,j}^{-1}\| \leq (\|\Omega_{n,j}^{-1}\| + \|\hat{\Omega}_{n,j}^{-1} - \Omega_{n,j}^{-1}\|) \|\Omega_{n,j} - \hat{\Omega}_{n,j}\| \|\Omega_{n,j}^{-1}\|.$$  

43
Let \( v_{j,n,T}, z_{j,n,T} \) and \( x_{j,n,T} \) denote \( \| \Omega_{j,n}^{-1} \|, \| \hat{\Omega}_{j,n}^{-1} - \Omega_{j,n}^{-1} \| \) and \( \| \Omega_{j,n} - \hat{\Omega}_{j,n} \| \), respectively. From the preceding equation, we have

\[
w_{j,n,T} := \frac{z_{j,n,T}}{v_{j,n,T} + z_{j,n,T}} \leq x_{j,n,T},
\]
whence we have \( \max_{1 \leq j \leq m} w_{j,n,T} \leq \max_{1 \leq j \leq m} x_{j,n,T} = O_p(a_{m,n,T}) = o_p(1) \). We now solve for \( z_{j,n,T} \):

\[
z_{j,n,T} = \frac{v_{j,n,T}^2 w_{j,n,T}}{1 - v_{j,n,T} w_{j,n,T}}.
\]

Then we have

\[
\max_{1 \leq j \leq m} z_{j,n,T} = \max_{1 \leq j \leq m} \frac{v_{j,n,T}^2 w_{j,n,T}}{1 - v_{j,n,T} w_{j,n,T}} = \frac{\max_{1 \leq j \leq m} v_{j,n,T}^2 \max_{1 \leq j \leq m} w_{j,n,T}}{1 - \max_{1 \leq j \leq m} v_{j,n,T} \max_{1 \leq j \leq m} w_{j,n,T}} = O_p(a_{m,n,T})
\]

where the second equality is due to the fact that \( 0 \leq v_{j,n,T} w_{j,n,T} \leq 1 \) for any \( j \).

We next review definitions of nets and covering numbers.

**Definition B.1** (Nets and covering numbers). Let \((T, d)\) be a metric space and fix \( \varepsilon > 0 \).

(i) A subset \( N_\varepsilon \) of \( T \) is called an \( \varepsilon \)-net of \( T \) if every point \( x \in T \) satisfies \( d(x, y) \leq \varepsilon \) for some \( y \in N_\varepsilon \).

(ii) The minimal cardinality of an \( \varepsilon \)-net of \( T \) is denoted \( |N_\varepsilon| \) and is called the covering number of \( T \) (at scale \( \varepsilon \)). Equivalently, \( |N_\varepsilon| \) is the minimal number of balls of radius \( \varepsilon \) and with centers in \( T \) needed to cover \( T \).

**Lemma B.6.** The unit Euclidean sphere \( \{ x \in \mathbb{R}^n : \| x \|_2 = 1 \} \) equipped with the Euclidean metric satisfies for every \( \varepsilon > 0 \) that

\[
|N_\varepsilon| \leq \left( 1 + \frac{2}{\varepsilon} \right)^n.
\]

**Proof.** See Vershynin (2011) Lemma 5.2 p8. \( \square \)

Recall that for a symmetric \( n \times n \) matrix \( A \), its \( \ell_2 \) spectral norm can be written as: \( \| A \|_{\ell_2} = \max_{\| x \|_2 = 1} | x^T A x | \).

**Lemma B.7.** Let \( A \) be a symmetric \( n \times n \) matrix, and let \( N_\varepsilon \) be an \( \varepsilon \)-net of the unit sphere \( \{ x \in \mathbb{R}^n : \| x \|_2 = 1 \} \) for some \( \varepsilon \in [0, 1) \). Then

\[
\| A \|_{\ell_2} \leq \frac{1}{1 - 2\varepsilon} \max_{x \in N_\varepsilon} | x^T A x |.
\]

**Proof.** See Vershynin (2011) Lemma 5.4 p8. \( \square \)
References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons.

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

Bickel, P. J. and E. Levina (2008). Covariance regularization by thresholding. *The Annals of Statistics* 36(6), 2577–2604.

Billingsley, P. (1995). *Probability and Measure*. Wiley.

Bühlmann, P. and S. van de Geer (2011). *Statistics for High-Dimensional Data*. Springer.

Campbell, D. T. and E. J. O’Connell (1967). Method factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research* 2, 409–426.

Cho, J. S. and P. C. B. Phillips (2018). Pythagorean generalization of testing the equality of two symmetric positive definite matrices. *Journal of Econometrics* 202(1), 45–56.

Chudik, A. and M. H. Pesaran (2013). Large panel data models with cross-sectional dependence: A survey. *Working Paper*.

Cudeck, R. (1988). Multiplicative models and mtmm matrices. *Journal of Educational Statistics* 13(2), 131–147.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147, 186–197.

Fan, J., Y. Liao, and H. Liu (2016). An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal* 19(1), 1–32.

Fan, J., Y. Liao, and M. Mincheva (2011). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics* 39(6), 3320–3356.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(4), 603680.

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

Filipiak, K., D. Klein, and E. Vojtkova (2018). The properties of partial trace and block trace operators of partitioned matrices. *Electronic Journal of Linear Algebra* 33, 3–15.

Hafner, C. M., O. B. Linton, and H. Tang (2018). Estimation of a multiplicative correlation structure in the large dimensional case. *arXiv:1605.04838*.

Hidalgo, J. and M. Schafgans (2017). Inference without smoothing for large panels with cross-sectional and temporal dependence. *Working Paper*.

Hoff, P. D. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis* 6(2), 179–196.

Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *The Annals of Applied Statistics* 9(3), 11691193.
Hoff, P. D. (2016). Equivariant and scale-free tucker decomposition models. *Bayesian Analysis* 11(3), 627-648.

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics* 104(1), 219–257.

Knopp, K. (1947). *Theory and Application of Infinite Series*. Blackie and Son.

Ledoit, O. and M. Wolf (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10, 603–621.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88, 365-411.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large dimensional covariance matrices. *Annals of Statistics* 40(2), 1024–1060.

Ledoit, O. and M. Wolf (2015). Spectrum estimation: a unified framework for covariance matrix estimation and pca in large dimensions. *Journal of Multivariate Analysis* 139(2), 360–384.

Ledoit, O. and M. Wolf (2017). Direct nonlinear shrinkage estimation of large-dimensional covariance matrices. *Working Paper*.

Linton, O. and J. P. Nielson (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.

Loh, W.-L. and T.-K. Lam (2000). Estimating structured correlation matrices in smooth gaussian random field models. *The Annals of Statistics* 28(3), 880–904.

Maasoumi, E. and P. C. B. Phillips (1982). On the behaviour of inconsistent instrumental variable estimators. *Journal of Econometrics* 19, 183–203.

Magnus, J. R. and H. Neudecker (1986). Symmetry, 0-1 matrices and jacobians a review. *Econometric Theory* (2), 157–190.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance* 7(1), 77–91.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10(2), 233–253.

Onatski, A. and C. Wang (2018). Alternative asymptotics for cointegration tests in large vars. *Econometrica* 86(4), 1465–1478.

Park, J. Y. and P. C. B. Phillips (1988). On the formulation of wald tests of nonlinear restrictions. *Econometrica* 56(5), 1065–1083.

Pesaran, M. H. and T. Yamagata (2012). Testing capm with a large number of assets. *Working Paper*.

Phillips, P. C. B. (1982). The true characteristic function of the f distribution. *Biometrika* 69(1), 261–264.

Phillips, P. C. B. (1985). The exact distribution of the sur estimator. *Econometrica* 53(4), 745–756.

Phillips, P. C. B. (1986). The exact distribution of the wald statistic. *Econometrica* 54(4), 881–895.
Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5), 1057–1111.

Robinson, P. M. (1998). Inference-without-smoothing in the presence of nonparametric autocorrelation. *Econometrica* 66(5), 1163–1182.

Saikkonen, P. and H. Lutkepohl (1996). Infinite-order cointegrated vector autoregressive processes. *Econometric Theory* 12, 814–844.

Swain, A. J. (1975). Analysis of parametric structures for variance matrices. *PhD Thesis, University of Adelaide*.

Verhees, J. and T. J. Wansbeek (1990). A multimode direct product model for covariance structure analysis. *British Journal of Mathematical and Statistical Psychology* 43(2), 231240.

Vershynin, R. (2011). Introduction to the non-asymptotic analysis of random matrices. *Chapter 5 of: Compressed Sensing, Theory and Applications*. 

47