UNIVERSITY OF CAMBRIDGE

Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 2050

ESTIMATION OF THE KRONECKER COVARIANCE MODEL BY QUADRATIC FORM

Oliver Linton

Haihan Tang

3 June 2020

We propose a new estimator, the quadratic form estimator, of the Kronecker product model for covariance matrices. We show that this estimator has good properties in the large dimensional case (i.e., the cross-sectional dimension n is large relative to the sample size T). In particular, the quadratic form estimator is consistent in a relative Frobenius norm sense provided $\log^3 n/T \rightarrow 0$. We obtain the limiting distributions of Lagrange multiplier (LM) and Wald tests under both the null and local alternatives concerning the mean vector μ.Testing linear restrictions of _ is also investigated. Finally, our methodology performs well in the finite-sample situations both when the Kronecker product model is true, and when it is not true.

Estimation of the Kronecker Covariance Model by Quadratic Form*

Oliver B. Linton†
University of Cambridge

Haihan Tang‡
Fudan University

May 28, 2020

Abstract

We propose a new estimator, the quadratic form estimator, of the Kronecker product model for covariance matrices. We show that this estimator has good properties in the large dimensional case (i.e., the cross-sectional dimension n is large relative to the sample size T). In particular, the quadratic form estimator is consistent in a relative Frobenius norm sense provided $\log^3 n/T \to 0$. We obtain the limiting distributions of Lagrange multiplier (LM) and Wald tests under both the null and local alternatives concerning the mean vector μ. Testing linear restrictions of μ is also investigated. Finally, our methodology performs well in the finite-sample situations both when the Kronecker product model is true, and when it is not true.

Some key words: Covariance matrix; Kronecker product; Quadratic form; Lagrange multiplier test; Wald test

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 certain latent traits through a factor model, and then use the covariance matrix of the observed variables to deduce properties of the latent traits. In econometrics, covariance matrices often appear in test statistics representing the sampling variability of a vector of parameter estimates. Anderson (1984) is a classic statistical reference that studies estimation of and hypothesis testing about covariance matrices in the low dimensional case (i.e., the dimension of the covariance matrix, n, is small compared with the sample size T).

There are many new methodological approaches to covariance and precision matrix estimation in the large dimensional case (i.e., n is large compared with T);[1] see, e.g., Ledoit and Wolf

*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 2019 for helpful comments. We thank Liang Jiang, Chen Wang, Tengyao Wang for useful discussions. Detailed comments from the Co-Editor and two anonymous referees also greatly improved the article. Any remaining errors are our own.

†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.Tang gratefully acknowledges the financial support from the National Natural Science Foundation of China (grant number 71903034).

[1]Some studies have made a distinction between the large dimensional case and the high dimensional case (Hafner, Linton, and Tang (2019)). We no longer make this distinction in this article. As long as n is large relative to T, regardless of n exceeding T, we call it the large dimensional case.

(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, Liao, and Liu (2016) gave an excellent account of the recent developments in theory and practice of estimating large dimensional covariance matrices. The usual approaches include: 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 at least to "shrink" towards a sparse matrix, or to use a factor model which reduces the dimensionality of the parameter space. Most of this literature assumes i.i.d. data.

We consider the problem of estimating a large covariance matrix Σ. We impose a model structure that reduces the effective dimensionality. In particular, we consider the Kronecker product model. 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

$$\Sigma = \sigma^2 \times \Sigma_1 \otimes \cdots \otimes \Sigma_v, \tag{1.1}$$

where Σ_j is an $n_j \times n_j$ unknown covariance matrix satisfying $\mathrm{tr}(\Sigma_j) = n_j$ for $j = 1, \ldots, v$, and $0 < \sigma^2 < \infty$ is a scalar parameter.

Kronecker product models arise naturally from *multiway data* (c.f. Kroonenberg (2008)). Multiway data are a generalization of two-way or three-way data that are widely encountered in social science. For example, the scores on 3 subjects (mathematics, English and music) of 50 students observed over 10 years are three-way data, the "ways" being subjects, students and years. Let $w_{i,j,t}$ denote the score of subject i of student j in year t. To model $w_{i,j,t}$, one could use an interactive effects model similar to Bai (2009):

$$w_{i,j,t} = \mu_{i,j} + \gamma_{i,t} f_{j,t}, \qquad i = 1, 2, 3, \quad j = 1, \ldots, 50, \quad t = 1, \ldots, 10$$

where $\mu_{i,j}$ is the subject-student specific mean, while $\gamma_{i,t}$ and $f_{j,t}$ are the subject-time specific and student-time specific effects, respectively. Stacking all the observations $\{w_{i,j,t}\}$ of year t into a 150×1 column vector y_t, we have $y_t = \mu + \gamma_t \otimes f_t$, where μ is the 150×1 mean vector containing stacked $\{\mu_{i,j}\}$, $\gamma_t = (\gamma_{1,t}, \gamma_{2,t}, \gamma_{3,t})^\intercal$, and $f_t = (f_{1,t}, \ldots, f_{50,t})^\intercal$. Suppose that γ_t is a random vector independent of f_t, and that both are mean-zero and stationary in time. Then,

$$\mathbb{E}[(y_t - \mu)(y_t - \mu)^\intercal] = \mathbb{E}[\gamma_t \gamma_t^\intercal] \otimes \mathbb{E}[f_t f_t^\intercal].$$

In this case the covariance matrix of y_t is a Kronecker product of two sub-matrices, which describe the subject specific and individual specific dependencies.

Extending the idea to multiway data, one might think of a typical equity portfolio constructed by intersections of 5 size quintiles, 5 book-to-market ratio quintiles, and 10 industries, in the spirit of Fama and French (1993), over a number of years, as four-way data: sizes $\times$ B/P ratios $\times$ industries $\times$ years. Situations in which higher-way data are collected are also on the increase. For example, electroencephalography (EEG), a non-invasive way of detecting structural abnormalities such as brain tumours, also provide multiway data, such as EEG bands $\times$ patients $\times$ leads $\times$ doses $\times$ time $\times$ task conditions (Estienne, Matthijs, Massart, Ricoux, and Leibovici (2001)).

Consider $(v+1)$-way data $w_{i_1,i_2,\ldots,i_v,t}$, where $i_j = 1, \ldots, n_j$ for $j = 1, \ldots, v$ and $t = 1, \ldots, T$. We use subscript t to denote the $(v+1)$th way of the data in the hope to *broadly* interpret the $(v+1)$th way as "time", T as the sample size, all other ways as the "cross-section", and $n := n_1 \times \cdots \times n_v$ as the cross-sectional dimension. In other words, the $(v+1)$th way of the data need not correspond to the time dimension, should the multiway data contain such a dimension. In the rest of the article, we shall no longer stress this distinction. Suppose that $w_{i_1,i_2,\ldots,i_v,t} = \mu_{i_1,i_2,\ldots,i_v} + \varepsilon^1_{i_1,t}\varepsilon^2_{i_2,t} \cdots \varepsilon^v_{i_v,t}$, where $i_j = 1, \ldots, n_j$ for $j = 1, \ldots, v$, and $t = 1, \ldots, T$. Equivalently, in the stacked form

$$y_t := (w_{1,1,\ldots,1,t}, \ldots, w_{n_1,n_2,\ldots,n_v,t})^\intercal = \mu + \varepsilon^1_t \otimes \varepsilon^2_t \otimes \cdots \otimes \varepsilon^v_t,$$

where μ is the stacked mean vector, $\varepsilon_t^j := (\varepsilon_{1,t}^j, \ldots, \varepsilon_{n_j,t}^j)^\intercal$ is an $n_j \times 1$ mean-zero random vector with covariance matrix $\mathbb{E}[\varepsilon_t^j \varepsilon_t^{j\intercal}]$ for all t for $j = 1, \ldots, v$. If $\varepsilon_t^1, \ldots, \varepsilon_t^v$ are mutually independent for all t, then

$$\mathbb{E}[(y_t - \mu)(y_t - \mu)^\intercal] = \mathbb{E}[\varepsilon_t^1 \varepsilon_t^{1\intercal}] \otimes \mathbb{E}[\varepsilon_t^2 \varepsilon_t^{2\intercal}] \otimes \cdots \otimes \mathbb{E}[\varepsilon_t^v \varepsilon_t^{v\intercal}].$$

We hence see that the covariance matrix of y_t is a Kronecker product of v sub-matrices.

Recent work of Kronecker product models on multiway data include Hoff (2011), Hoff (2015), Hoff (2016) etc. Kronecker product models have also been considered in the psychometric literature (Campbell and O'Connell (1967), Swain (1975), Cudeck (1988), Verhees and Wansbeek (1990) etc). In the spatial literature, there are a number of studies that consider Kronecker product models 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.

These literatures have all focussed on the low dimensional case. Hafner et al. (2019) were the first to study Kronecker product models in the large dimensional case. The proper framework for studying the large dimensional case is the joint limit setting developed by Phillips and Moon (1999) in which n and T tend to infinity simultaneously.[2] Since n tends to infinity, there are two main cases when considering (1.1): (a) $\{n_j\}_{j=1}^v$ are all fixed while $v \to \infty$; (b) $n_j \to \infty$ for at least some j while v is fixed. Case (a) corresponds to practical situations where the data have a large number of ways but in each way the number of entities is small; case (b) often corresponds to, say, three-way or four-way data in which at least one way has a large number of entities. The methodologies developed in Hafner et al. (2019) and this article are perfectly geared for case (a) in the sense that (1.1) is *correctly* specified for the data.

We do not analyse case (b) theoretically, but our estimation and inference procedures can in principle be applied to case (b) also, but the theory will require more work and stronger restrictions on the relationship between n and T. For example, if $v = 2$ and $n_1 = n_2 = \sqrt{n}$, then the sub-matrices Σ_1, Σ_2 each contain order n unknown quantities. If $n/T \to 0$ fast enough, then we may show some consistency of our estimators of the sub-matrices Σ_1, Σ_2. On the other hand, if this rate condition is not satisfied, one could combine the separable structure (i.e., the Kronecker product) with sparsity restrictions on the sub-matrices. This has been investigated in the literature. Other approaches have been considered in Akdemir and Gupta (2011), Hoff (2011), and Hoff (2015). Henceforth when we say the Kronecker product model (1.1), we implicitly mean case (a).

The Kronecker product model leads to substantial dimension reduction even though it need not be sparse in the sense of (2.1) of Fan et al. (2016). Hafner et al. (2019) showed that the matrix logarithm of a Kronecker product covariance or correlation matrix is a sparse matrix (with $O(\log n)$ unknown quantities) and the logarithmic operator converts the multiplicative Kronecker product structure into an additive one. Therefore, the logarithm of a Kronecker product 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 its consistency and provided a central limit theorem (CLT). However, their results require strong, albeit sufficient but not necessary, conditions; in particular they obtained Frobenius norm consistency of the estimator under a condition that at least $n/T \to 0$, which is very restrictive. On the contrary, other methodologies typically achieve *average* Frobenius norm consistency provided

[2]Peter Phillips has made some fundamental contributions to the large dimensional analysis. Phillips and Moon (1999) provided three asymptotic frameworks for analysing double-index (n, T) processes: sequential limit framework (e.g., $n \to \infty$ followed by $T \to \infty$), diagonal path limit framework (i.e., both n and T pass to infinity along some specific diagonal in the two dimensional array), and joint limit framework (i.e., $n, T \to \infty$ simultaneously). In particular, they provided a central limit theorem in joint limit framework for double-index processes (Theorem 2 of Phillips and Moon (1999)). However, the Lindeberg condition of the theorem is perhaps difficult to verify in practice. In Section B, we provide a variant (Theorem B.1), which relies on a Lyapounov's condition. Moreover, the variant allows the central limit theorem to kick in from either the cross-sectional or time dimension.

$s \log n/T \to 0$, where s is some sparsity index (e.g., see Bickel and Levina (2008) Theorem 2 with $q = 0$).[3]

In this article, we relax the rate restriction on n imposed by Hafner et al. (2019) and allow n to be possibly larger than T. We propose a new covariance matrix estimator called the *quadratic form* estimator based on the Kronecker product model. Our estimator averages elements of the sample covariance matrix, so we obtain a rate improvement by averaging. In particular, under a cross-sectional weak dependence condition, the quadratic form estimator achieves *relative* Frobenius norm consistency provided $\log^3 n/T \to 0$. Moreover, this method automatically produces a symmetric and positive definite covariance matrix estimator, unlike some of the sparsifying methods considered by Fan et al. (2016).

We apply our methodology to a concrete testing problem; we consider the null hypothesis $H_0 : \mu = \mu_0$, where μ is the mean of the large dimensional data y_t and μ_0 is some known vector. One practical example would be that y_t corresponds to differences between treated and controlled groups and we want to test whether these cross-sectional differences are different from zero. We define Lagrange multiplier (LM) and Wald test statistics based on our estimated precision matrix and establish their asymptotic distributions under both null and local alternatives of the form $H_1 : \mu = \mu_0 + \theta/\sqrt{T}$ for some vector θ. We also provide two results regarding testing linear restrictions of μ.

We compare our estimation and testing methods with Ledoit and Wolf (2004)'s linear shrinkage estimator and Ledoit and Wolf (2017)'s direct nonlinear shrinkage estimator in Monte Carlo simulations. Our methods perform very well in moderate-sized samples. In fact, they work well even in situations where a Kronecker product model is *misspecified* for a covariance matrix.

The rest of the article is structured as follows. In Section 2 we discuss the model and identification while in Section 3 we propose the quadratic form estimator. Section 4 gives the rate of convergence for the quadratic form estimator. In Section 5 we define LM and Wald test statistics and establish their asymptotic distributions under both null and local alternatives. We also consider testing linear restrictions of μ. Section 6 conducts Monte Carlo simulations comparing our approach with Ledoit and Wolf estimators. Section 7 concludes. All 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 $\operatorname{vec} A$ 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 $\operatorname{vec} A$ to $\operatorname{vec}(A^\intercal)$, i.e., $\operatorname{vec}(A^\intercal) = K_{m,n} \operatorname{vec}(A)$. If A is a symmetric $n \times n$ matrix, its $n(n-1)/2$ supradiagonal elements are redundant in the sense that they can be deduced from symmetry. If we eliminate these redundant elements from $\operatorname{vec} A$, we obtain a new $n(n+1)/2 \times 1$ vector, denoted $\operatorname{vech} A$. They are related by the full-column-rank, $n^2 \times n(n+1)/2$ *duplication matrix* D_n: $\operatorname{vec} A = D_n \operatorname{vech} A$. Conversely, $\operatorname{vech} A = D_n^+ \operatorname{vec} A$, where D_n^+ is $n(n+1)/2 \times n^2$ and the Moore-Penrose generalized inverse of D_n. In particular, $D_n^+ = (D_n^\intercal D_n)^{-1} D_n^\intercal$ 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 $\|x\|_\infty := \max_{1 \le i \le n} |x_i|$ denote the Euclidean (ℓ_2) norm and the element-wise maximum (ℓ_∞) 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 $A = (a_{i,j})_{1 \le i \le m, 1 \le j \le n}$, let $\|A\|_F := [\operatorname{tr}(A^\intercal A)]^{1/2} \equiv [\operatorname{tr}(AA^\intercal)]^{1/2} \equiv \|\operatorname{vec} A\|_2$, $\|A\|_1 := \sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|$, $\|A\|_{\ell_2} := \max_{\|x\|_2 = 1} \|Ax\|_2 \equiv \sqrt{\lambda_{\max}(A^\intercal A)}$, $\|A\|_{\ell_1} := \max_{1 \le j \le n} \sum_{i=1}^m |a_{i,j}|$, and $\|A\|_{\ell_\infty} := \max_{1 \le i \le m} \sum_{j=1}^n |a_{i,j}|$ denote the Frobenius (ℓ_2) norm, ℓ_1 norm, and spectral norm (ℓ_2-operator norm), maximum column sum matrix norm (ℓ_1-operator

[3] *Average* Frobenius norm means dividing a Frobenius norm by $\sqrt{n}$, while *relative* Frobenius norm means dividing a Frobenius norm by the Frobenius norm of a target matrix, say, the unknown covariance matrix. These two concepts are similar, but not exactly the same.

norm), and maximum row sum matrix norm (ℓ_∞-operator norm) of A, respectively. Note that $\|\cdot\|_\infty$ can also be applied to matrix A, i.e., $\|A\|_\infty = \max_{1\le i\le m, 1\le j\le n}|a_{i,j}|$; 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. An *absolute* positive constant refers to a constant independent of anything which is a function of n and/or T. We write $a \asymp b$ if there exist absolute constants $0 < c_1 \le c_2$ such that $c_1 b \le a \le c_2 b$. For real numbers a, b let $a \vee b$ denote $\max(a, b)$.

2 The Model and Identification

We now directly work with the high-level n-dimensional random vector y_t with $\mu := \mathbb{E} y_t$ and $\Sigma := \mathbb{E}[(y_t - \mu)(y_t - \mu)^\intercal]$ for every t. In particular, Σ takes the form of (1.1). For each j, Σ_j contains $n_j(n_j+1)/2-1$ (unrestricted) parameters. In total, model (1.1) contains $\sum_{j=1}^{v} n_j(n_j+1)/2 - (v-1)$ unknown parameters. This model is the same as considered in Hafner et al. (2019) except that we make a different identifying restriction. The implied form for Σ^{-1} is also Kronecker, i.e., $\Sigma^{-1} = \sigma^{-2} \times \Sigma_1^{-1} \otimes \cdots \otimes \Sigma_v^{-1}$.

We show that model (1.1) is indeed identified. First, the parameter σ is identified because

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

whence we have $\sigma^2 = \mathrm{tr}(\Sigma)/n$. We next consider identification of the remaining parameters based on the *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,

$$A = \begin{pmatrix} A_{-1;1,1} & \cdots & A_{-1;1,n_1} \\ & \ddots & \vdots \\ & & A_{-1;n_1,n_1} \end{pmatrix}. \tag{2.1}$$

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

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

Consider model (1.1), and let $\Sigma_{-1} := \Sigma_2 \otimes \cdots \otimes \Sigma_v$. Define the $n_1 \times n_1$ matrix $d^{(1)} := \mathrm{PTR}_{n_1}(\Sigma) = \sigma^2 \mathrm{tr}(\Sigma_{-1}) \times \Sigma_1$. Then $\Sigma_1 = d^{(1)}/(\mathrm{tr}(d^{(1)})/n_1)$. According to Definition 1.1(ii) of Filipiak et al. (2018), $\mathrm{PTR}_{n_1}(\Sigma) = \sum_{\ell=1}^{n_{-1}} (I_{n_1} \otimes e_{\ell,n_{-1}}^\intercal)\Sigma(I_{n_1} \otimes e_{\ell,n_{-1}})$, where $e_{\ell,n_{-1}}$ is the $n_{-1} \times 1$ elementary vector with one in position ℓ and zero elsewhere. In this sense, $d^{(1)}$ is a quadratic form of Σ.

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

$$\Sigma_{-h} := \begin{cases} \Sigma_{h+1} \otimes \cdots \otimes \Sigma_v \otimes \Sigma_1 \otimes \cdots \otimes \Sigma_{h-1} & \text{for } h = 2, \ldots, v-1 \\ \Sigma_1 \otimes \cdots \otimes \Sigma_{v-1} & \text{for } h = v \end{cases}$$

Note that Σ_{-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

$$\begin{aligned} \Sigma^{(h)} &:= K_{n_h\times\cdots\times n_v, n_1\times\cdots\times n_{h-1}} \Sigma K_{n_1\times\cdots\times n_{h-1}, n_h\times\cdots\times n_v} \\ &= K_{n_h\times\cdots\times n_v, n_1\times\cdots\times n_{h-1}} (\sigma^2 \times \Sigma_1 \otimes \cdots \otimes \Sigma_v) K_{n_1\times\cdots\times n_{h-1}, n_h\times\cdots\times n_v} \\ &= \sigma^2 \times \Sigma_h \otimes \Sigma_{h+1} \otimes \cdots \otimes \Sigma_v \otimes \Sigma_1 \otimes \cdots \otimes \Sigma_{h-1} = \sigma^2 \times \Sigma_h \otimes \Sigma_{-h}. \end{aligned} \tag{2.2}$$

Define the $n_h \times n_h$ matrix $d^{(h)} := \mathrm{PTR}_{n_h}(\Sigma^{(h)}) = \sigma^2 \mathrm{tr}(\Sigma_{-h}) \times \Sigma_h$. Then

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

3 Estimation

We observe an n-dimensional weakly stationary time series vector $\{y_t\}_{t=1}^T$ with mean μ and covariance matrix Σ. Define the sample covariance matrix

$$M_T := \frac{1}{T}\sum_{t=1}^T (y_t - \bar{y})(y_t - \bar{y})^\intercal,$$

where $\bar{y} := \frac{1}{T}\sum_{t=1}^T y_t$. Define $\hat{d}^{(1)} := \mathrm{PTR}_{n_1}(M_T)$. Then let $\tilde{\Sigma}_1 := \hat{d}^{(1)}/(\mathrm{tr}(\hat{d}^{(1)})/n_1)$. Likewise define the "permuted" sample covariance matrix

$$M_T^{(h)} := K_{n_h\times\cdots\times n_v, n_1\times\cdots\times n_{h-1}} M_T K_{n_1\times\cdots\times n_{h-1}, n_h\times\cdots\times n_v}, \tag{3.1}$$

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

$$\tilde{\Sigma}_h := \frac{\hat{d}^{(h)}}{\mathrm{tr}(\hat{d}^{(h)})/n_h}, \tag{3.2}$$

for $h = 1, \ldots, v$.

The *quadratic form* estimator $\tilde{\Sigma}$ for Σ is

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

$$\hat{\sigma}^2 := \frac{\mathrm{tr}(M_T)}{n}. \tag{3.3}$$

By Lemma 2.4 of Filipiak et al. (2018), if M_T is symmetric and positive semidefinite, then so are $\{\tilde{\Sigma}_j\}_{j=1}^v$ and hence $\tilde{\Sigma}$. Moreover, simulations show that even for positive semidefinite M_T, $\{\tilde{\Sigma}_j\}_{j=1}^v$ and hence $\tilde{\Sigma}$ will be positive definite. As a result, the quadratic form estimator $\tilde{\Sigma}^{-1}$ for Σ^{-1} is $\tilde{\Sigma}^{-1} = \hat{\sigma}^{-2} \times \tilde{\Sigma}_1^{-1} \otimes \cdots \otimes \tilde{\Sigma}_v^{-1}$. We stress that $\tilde{\Sigma}^{-1}$ exists even if $n > T$. The quadratic form estimator is closely related to the quasi-maximum likelihood estimation (QMLE), but has the particular advantage in large dimensions in the sense that it is in closed form.[4]

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 4.3, we should have a rate improvement for the quadratic form estimator. We formally establish this in Section 4.

4 The Rate of Convergence

In this section, we shall derive the rate of convergence for the quadratic form estimator. We make the following assumptions:

Assumption 4.1.

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

[4] In the previous version of this article, we introduced a variant of the quadratic form estimator, which was derived by replacing the partial trace operator with a partial sum operator. Because of inferiority of that variant, we no longer include it in the current version.

(ii)

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

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

$$\text{var}\left((y_t-\mu)\otimes(y_t-\mu)\right)\leq C\,\text{var}\left((z_t-\mu)\otimes(z_t-\mu)\right),$$

for some absolute positive constant C for every 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 a concentration inequality such as the Bernstein's inequality. Normal random vectors or random vectors that exhibit some exponential-type tail probability (e.g., subgaussianity, subexponentiality, semiexponentiality etc) satisfy this condition. Assumption 4.1(iii) supposes that the kurtosis matrix of y_t is of the same order of magnitude as if it were a normal random vector. We impose this restriction on the kurtosis matrix of y_t because not much research has touched on unrestricted kurtosis matrices in the large dimensional case.

Assumption 4.2.

(i) $\max_{1\leq j\leq v} n_j$ is an absolute positive constant.

(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) requires that the dimensions of the sub-matrices be 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 $v\to\infty$.

Assumption 4.3. *Let $0\leq\beta_1\leq 2$.*

$$\lim_{n\to\infty}\frac{1}{n^{\beta_1}}\|\Sigma\|_F^2=\lim_{n\to\infty}\frac{\sigma^4}{n^{\beta_1}}\left(\prod_{j=1}^{v}\|\Sigma_j\|_F^2\right)=\omega<\infty.$$

Assumption 4.3 characterises the cross-sectional dependence of $\{y_t\}_{t=1}^T$. According to Proposition 1 of Chudik and Pesaran (2013), $\{y_t\}_{t=1}^T$ is said to be *cross-sectionally weakly dependent.* The smaller β_1 is, the less cross-sectional dependence of $\{y_t\}_{t=1}^T$ is allowed and the stronger Assumption 4.3 is. When $\beta_1=2$, Assumption 4.3 is slack as we are not restricting cross-sectional dependence of $\{y_t\}_{t=1}^T$ at all ($\|\Sigma\|_F^2=O(n^2)$ in general). On the one hand, we would like to assume β_1 as close to 2 as possible to make Assumption 4.3 as weak as possible. On the other hand, the smaller β_1 is, the weaker cross-sectional dependence $\{y_t\}_{t=1}^T$ exhibits, and the faster rate of convergence the quadratic form estimator will be able to achieve. There is a trade off.

One important case is $\beta_1 = 1$. In this case one *sufficient* condition for Assumption 4.3 is that Σ has bounded maximum column sum matrix norm (i.e., $\|\Sigma\|_{\ell_1} = O(1)$) or bounded maximum row sum matrix norm (i.e., $\|\Sigma\|_{\ell_\infty} = O(1)$). To see this

$$\frac{1}{n}\|\Sigma\|_F^2 \leq \frac{1}{n}n\|\Sigma\|_{\ell_1}^2 = \frac{1}{n}n\|\Sigma\|_{\ell_\infty}^2 = O(1).$$

Note that for symmetric Σ, bounded maximum column sum matrix norm or bounded maximum row sum matrix norm implies that the maximum eigenvalue of Σ is bounded from the above by an absolute positive constant and the minimum eigenvalue of Σ^{-1} is bounded away from zero by an absolute positive constant: $1/(\lambda_{\min}(\Sigma^{-1})) = \lambda_{\max}(\Sigma) = \|\Sigma\|_{\ell_2} \leq \|\Sigma\|_{\ell_1} = \|\Sigma\|_{\ell_\infty} = O(1)$. 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).

Theorem 4.1. *Suppose Assumptions 4.1, 4.2 and 4.3 hold. If* $\log^3 n/T \to 0$ *as* $n, T \to \infty$, *then we have*

(i)

$$\frac{\|\tilde{\Sigma} - \Sigma\|_F}{\|\Sigma\|_F} = O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right).$$

(ii)

$$\frac{\|\tilde{\Sigma}^{-1} - \Sigma^{-1}\|_F}{\|\Sigma^{-1}\|_F} = O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right).$$

(iii)

$$\frac{\|\tilde{\Sigma} - \Sigma\|_1}{\|\Sigma\|_1} = O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right).$$

(iv)

$$\frac{\|\tilde{\Sigma}^{-1} - \Sigma^{-1}\|_1}{\|\Sigma^{-1}\|_1} = O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right).$$

(v)

$$\frac{\|\tilde{\Sigma} - \Sigma\|_{\ell_2}}{\|\Sigma\|_{\ell_2}} = O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right).$$

(vi)

$$\frac{\|\tilde{\Sigma}^{-1} - \Sigma^{-1}\|_{\ell_2}}{\|\Sigma^{-1}\|_{\ell_2}} = O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right).$$

The reason that we divide the Frobenius norm of estimation error, say, $\|\tilde{\Sigma} - \Sigma\|_F$, by the Frobenius norm of the target, i.e., $\|\Sigma\|_F$, is to define a proper notion of "consistency". This is necessary because the cross-sectional dimension n is growing to infinity. In this case, 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 estimation error is standard in the large dimensional case, but in the literature scholars tend to divide the Frobenius norm of estimation 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). The same reasoning applies to the ℓ_1 and the spectral norm of the estimation error.

Note that there are two terms on the right side. The term $O_p\left(\log^2 n/T\right)$ exists because we need to estimate the unknown μ. If we knew μ, this term would not be present.[5] The rate of convergence, $\left(\log^3 n/(n^{2-\beta_1}T)\right)^{1/2}$, contains an additional, non-standard item $\sqrt{n^{2-\beta_1}}$ in the denominator. This non-standard item exists because of the cross-sectional weak dependence condition (Assumption 4.3). If $\beta_1 = 2$ (i.e., we are not restricting cross-sectional dependence of $\{y_t\}_{t=1}^T$ at all), this term vanishes. The rate of convergence of the quadratic form estimator then becomes $(\log^3 n/T)^{1/2}$, which is comparable to the convergence rates of other existent estimators in the large dimensional case.

Take part (i) of the theorem as an illustration. If $\beta_1 = 2$ and we knew μ, we have $\left\|\tilde{\Sigma} - \Sigma\right\|_F = O_p\left(\|\Sigma\|_F(\log^3 n/T)^{1/2}\right)$. A typical threshold estimator $\hat{\Sigma}_{\text{thres}}$ has $\|\hat{\Sigma}_{\text{thres}} - \Sigma^\dagger\|_F = O_p\left((sn\log n/T)^{1/2}\right)$, where $\Sigma^\dagger$ is some sparse truth and s is its 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 $\|\Sigma^\dagger\|_F = O(\sqrt{sn})$ under the sparsity model. If one assumes $\|\Sigma^\dagger\|_F \asymp \sqrt{sn}$, one can write $\|\hat{\Sigma}_{\text{thres}} - \Sigma^\dagger\|_F = O_p\left(\|\Sigma^\dagger\|_F(\log n/T)^{1/2}\right)$. Then the two rates of convergence only differ by a logarithmic factor.

Because of the cross-sectional weak dependence condition (Assumption 4.3), the quadratic form estimator is able to achieve a faster rate of convergence than a typical estimator does.

5 Test Statistics

We apply our methodology to the testing issue. We consider the problem of testing the null hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$.

The classical Wald test statistic (based on the sample covariance matrix M_T) is not defined when $n \geq T$; there is a large literature that proposes alternative test statistics. Bai and Saranadasa (1996) proposed a statistic based on $\|\bar{y}\|_2^2$, 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 the limiting normal distribution of the centred and scaled version of this under cross-sectional weak dependence conditions. Fan et al. (2015) considered a Wald test 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 a 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 now define the Lagrange multiplier (LM) test statistic

$$LM_{n,T} = T(\bar{y} - \mu_0)^\intercal \tilde{\Sigma}_{\mu_0}^{-1}(\bar{y} - \mu_0), \tag{5.1}$$

where $\tilde{\Sigma}_{\mu_0}$ is the quadratic form estimator assuming that we know $\mu = \mu_0$. The Wald test statistic is

$$W_{n,T} = T(\bar{y} - \mu_0)^\intercal \tilde{\Sigma}^{-1}(\bar{y} - \mu_0), \tag{5.2}$$

which is the Hotelling T^2-statistic based on the quadratic form estimator. We next present the large sample properties of the binity $LM_{n,T}$ and $W_{n,T}$. We make one more cross-sectional dependence assumption.

[5] If we knew μ, the estimation procedure in Section 3 applies to $M_T^0 := T^{-1}\sum_{t=1}^T (y_t - \mu)(y_t - \mu)^\intercal$ instead of M_T.

Assumption 5.1. *Let* $0 \leq \beta_2 \leq 2$.

$$\lim_{n\to\infty} \frac{1}{n^{\beta_2}} \|\Sigma^{-1}\|_1 = \lim_{n\to\infty} \frac{1}{n^{\beta_2}\sigma^2} \left(\prod_{j=1}^{v} \|\Sigma_j^{-1}\|_1 \right) = \omega' < \infty.$$

The bigger β_2 is, the weaker Assumption 5.1 is. This is because it is putting less restriction on the cross-sectional dependence of Σ^{-1}. When $\beta_2 = 2$, Assumption 5.1 is slack as in essence we are not restricting anything. On the one hand, we wish to assume β_2 as close to 2 as possible to make Assumption 5.1 as weak as possible. On the other hand, we wish to assume that β_2 is as small as possible so that our methodology could accommodate an n as large as possible.

One important case is $\beta_2 = 1$. In this case, a *sufficient* condition for Assumption 5.1 is that Σ^{-1} has bounded maximum column sum matrix norm (i.e., $\|\Sigma^{-1}\|_{\ell_1} = O(1)$) or bounded maximum row sum matrix norm (i.e., $\|\Sigma^{-1}\|_{\ell_\infty} = O(1)$). To see this

$$\frac{1}{n}\|\Sigma^{-1}\|_1 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|(\Sigma^{-1})_{i,j}\right| \leq \max_{1\leq i\leq n}\sum_{j=1}^{n}\left|(\Sigma^{-1})_{i,j}\right| = \|\Sigma^{-1}\|_{\ell_\infty} = \|\Sigma^{-1}\|_{\ell_1} = O(1).$$

Note that for symmetric Σ^{-1}, bounded maximum column sum matrix norm or bounded maximum row sum matrix norm implies that the maximum eigenvalue of Σ^{-1} is bounded from the above by an absolute positive constant and the minimum eigenvalue of Σ is bounded away from zero by an absolute positive constant: $1/(\lambda_{\min}(\Sigma)) = \lambda_{\max}(\Sigma^{-1}) = \|\Sigma^{-1}\|_{\ell_2} \leq \|\Sigma^{-1}\|_{\ell_1} = \|\Sigma^{-1}\|_{\ell_\infty} = O(1)$. 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).

Theorem 5.1. *Suppose Assumptions 4.1, 4.2, 4.3, and 5.1 hold. We make the following assumptions:*

(a)

$$\frac{n^{2\beta_2+\beta_1-3}\log^5 n}{T} = o(1).$$

(b) Consider the Cholesky decomposition of Σ*, i.e.,* $\Sigma = LL^{\intercal}$*, where* L *is a nonsingular lower triangular matrix* L *with positive diagonal elements. Assume that* $x_t := L^{-1}(y_t - \mu)$ *is cross-sectionally independent for any* t*, and for some* $\delta > 0$

$$\limsup_{n,T\to\infty} \max_{1\leq i\leq n} \max_{1\leq t\leq T} \mathbb{E}\left|x_{t,i}\right|^{4+2\delta} < \infty.$$

Then under $H_0 : \mu = \mu_0$*, as* $n, T \to \infty$*,*

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

If one additionally assumes

$$\frac{n^{\beta_2-\frac{1}{2}} \cdot \log^3 n}{T} = o(1), \tag{5.3}$$

then under $H_0 : \mu = \mu_0$*, as* $n, T \to \infty$*,*

$$\frac{W_{n,T} - n}{\sqrt{2n}} \xrightarrow{d} N(0,1). \tag{5.4}$$

For the LM test, if we want to allow the interesting case of $n/T \to \infty$, then assumption (a) necessarily implies that $2\beta_2 + \beta_1 < 4$, which restricts both β_2 and β_1. In the special case of $\beta_1 = \beta_2 = 1$, assumption (a) is reduced to $\log^5 n/T = o(1)$, which is a weak condition.

Assumption (b) is standard in the literature. Fan et al. (2015) maintained normality (their Assumption 4.1(i)), which is a special case of assumption (b). Pesaran and Yamagata (2012) also maintained assumption (b) (their Assumption 2a). Assumption (b) implicitly assumes that $\lambda_{\min}(\Sigma)$ is bounded away from zero by an absolute positive constant, which strengthens Assumption 4.2(ii). Also note that $\text{var}(x_t) = I_n$, so strengthening from cross-sectional uncorrelatedness to cross-sectional independence in assumption (b) is rather innocuous. In addition, we assume that the $(4+2\delta)$th moment of $x_{t,i}$ is (uniformly in i and t) finite for n, T sufficiently large, which is also a weak assumption. Under the more restricted sequential limit ($T \to \infty$ and then $n \to \infty$), $\sqrt{T}(\bar{y} - \mu_0)$ is approximately normal so the limiting properties could be calculated for the non-normal case as if normality held. However, in our framework of joint limit, such procedure breaks down, so we make assumption (b).

In the low-dimensional case (n fixed, $T \to \infty$), LM test statistic $LM_{n,T}$ and Wald test statistic $W_{n,T}$ are asymptotically equivalent in the sense that they all converge in distribution to χ^2_n.[6] In the large dimensional case ($n, T \to \infty$), Theorem 5.1 shows that $LM_{n,T}$ and $W_{n,T}$ are, again, asymptotically equivalent. Wald test requires an additional rate restriction (5.3), which is the price we pay for estimating Σ^{-1} under the alternative $H_1 : \mu \neq \mu_0$.

Recall that a typical threshold estimator $\hat{\Sigma}_{\text{thres}}$ has $\|\hat{\Sigma}^{-1}_{\text{thres}} - (\Sigma^\dagger)^{-1}\|_{\ell_2} = O_p\left(s(\log n/T)^{1/2}\right)$, where $\Sigma^\dagger$ is some sparse truth and s is its sparsity index (see Bickel and Levina (2008) Theorem 1 with $q = 0$). For this rate of convergence, a result like (5.4) 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 (5.4) for their Wald test statistics in the CAPM context.

In our case of Wald test, if we also want to allow the interesting large dimension case of $n/T \to \infty$, then assumption (a) and (5.3) necessarily imply $2\beta_2 + \beta_1 < 4$ and $\beta_2 < 3/2$, respectively. For example, we can choose the special case $\beta_1 = \beta_2 = 1$, so assumption (a) and (5.3) reduce to

$$\frac{\log^5 n}{T} = o(1), \qquad \frac{n^{\frac{1}{2}} \cdot \log^3 n}{T} = o(1),$$

the latter of which is the binding rate condition and the same as the rate condition in Assumption 4.2 of Fan et al. (2015).

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

5.1 Power Investigation

In this section, we analyse the asymptotic distributions of the proposed test statistics under the alternative hypothesis $H_1 : \mu \neq \mu_0$. In particular, we shall focus on a sequence of local alternatives $H_1 : \mu = \mu_T := \mu_0 + \theta/\sqrt{T}$, where $\max_{1 \leq i \leq n} |\theta_i| = O(\sqrt{\log n})$. We focus on the Wald test without loss of generality.

Theorem 5.2. *Suppose Assumptions 4.1, 4.2, 4.3, and 5.1 hold. We make the following additional assumptions:*

[6] The finite sample performance of these statistics is known to vary. Park and Phillips (1988) established higher order approximations for Wald test of nonlinear restrictions in the finite dimensional case, and showed how to improve performance of the test statistic. It may be possible to apply their methodology to the large dimensional case.

(a) (i)
$$\frac{n^{2\beta_2+\beta_1-3}\cdot\log^5 n}{T}=o(1),$$

(ii)
$$\frac{n^{\beta_2-\frac{1}{2}}\cdot\log^3 n}{T}=o(1).$$

(b) *Consider the Cholesky decomposition of* Σ, *i.e.,* $\Sigma=LL^\intercal$, *where* L *is an* $n\times n$ *nonsingular lower triangular matrix with positive diagonal elements. Assume that* $x_t:=L^{-1}(y_t-\mu)$ *is cross-sectionally independent for any* t, *and for some* $\delta>0$

$$\limsup_{n,T\to\infty}\max_{1\le i\le n}\max_{1\le t\le T}\mathbb{E}\left|x_{t,i}\right|^{4+2\delta}<\infty$$
$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left|(L^{-1}\theta)_i\right|^{2+\delta}<\infty.$$

Then under $H_1:\mu=\mu_0+\theta/\sqrt{T}$,

$$\frac{W_{n,T}-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}-\frac{\theta^\intercal\Sigma^{-1}\theta}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}\xrightarrow{d}N(0,1).$$

The preceding theorem shows that the asymptotic distribution of $(W_{n,T}-n)/\sqrt{2n+4\theta^\intercal\Sigma^{-1}\theta}$ under H_1 has a center $\theta^\intercal\Sigma^{-1}\theta/\sqrt{2n+4\theta^\intercal\Sigma^{-1}\theta}$. Note that

$$\frac{\theta^\intercal\Sigma^{-1}\theta}{\sqrt{2n+4\theta^\intercal\Sigma^{-1}\theta}}\ge\frac{\theta^\intercal\theta\lambda_{\min}(\Sigma^{-1})}{\sqrt{2n+4\theta^\intercal\theta\lambda_{\max}(\Sigma^{-1})}}=\frac{\theta^\intercal\theta/\lambda_{\max}(\Sigma)}{\sqrt{2n+4\theta^\intercal\theta/\lambda_{\min}(\Sigma)}}.$$

In the special case of $0<\lambda_{\min}(\Sigma)<\lambda_{\max}(\Sigma)<\infty$, we see that the test has power against local alternatives that satisfy $\max_{1\le i\le n}|\theta_i|=O(\sqrt{\log n})$ and $\theta^\intercal\theta=O(n^{\delta_a})$, where $\delta_a\ge 1/2$, and power tending to one in the case where $\delta_a>1/2$. This specification requires that θ has a sufficiently large number of non-zero elements. It does not require that all the elements of θ are non-zero.

5.2 Testing Linear Restrictions of μ

In this section, we consider testing linear restrictions of μ using two approaches. We first consider $H_0:R\mu=r$, where R is a $q\times n$ matrix of rank q. We assume that q is a fixed number; this case covers applications where a finite number of linear restrictions are coming from economic theory.

Theorem 5.3. *Suppose Assumptions 4.1, 4.2 and 4.3 hold. We also make the following assumptions:*

(a) $\lambda_{\min}(\Sigma)$ *is bounded away from zero by an absolute positive constant.*

(b) *Consider* $H_0:R\mu=r$, *where* R *is a* $q\times n$ *matrix of rank* q *for any fixed* n *and* $n\to\infty$ *(*q *is a fixed number). Moreover,* R *and* r *are rescaled in such a way that* $\lambda_{\min}(RR^\intercal)$ *is bounded away from zero by an absolute constant, and*

$$\lambda_{\max}(RR^\intercal)\|\Sigma\|_{\ell_2}\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}+\frac{\log^2 n}{T}\right)=o(1).\tag{5.5}$$

Then under $H_0 : R\mu = r$, *if* $\log^3 n/T \to 0$ *as* $n, T \to \infty$,

$$W^*_{n,T} := T(R\bar{y} - r)^\intercal (R\tilde{\Sigma}R^\intercal)^{-1}(R\bar{y} - r) \xrightarrow{d} \chi^2_q.$$

Assumption (a) strengthens Assumption 4.2(ii) slightly, which is a mild condition. A sufficient condition for (5.5) in assumption (b) is $\lambda_{\max}(RR^\intercal)$ is bounded from the above by an absolute positive constant and $\|\Sigma\|_{\ell_2} < \infty$. The requirement of $\lambda_{\min}(RR^\intercal)$ and $\lambda_{\max}(RR^\intercal)$ being bounded away from zero and from the above by absolute positive constants, respectively, could be achieved by normalising each row of R to have ℓ_2 norm of 1.

We next take another approach to derive simultaneous confidence intervals for *all* linear combinations of μ.

Lemma 5.1. *Suppose Assumptions 4.1, 4.2, 4.3, and 5.1 hold. Simultaneously for all* $\phi \in \mathbb{R}^n$, *the unknown* μ *satisfies the following inequalities with confidence* $1 - \alpha$*:*

$$\frac{T\left[\phi^\intercal(\bar{y} - \mu)\right]^2 / \phi^\intercal\tilde{\Sigma}\phi - n}{\sqrt{2n}} < z_\alpha,$$

as $n, T \to \infty$, *where* z_α *is the upper* α *percentile of a standard normal.*

One disadvantage of this approach is that the confidence region for μ could be conservative.

6 Simulation Study

In this section, we provide some Monte Carlo simulations that evaluate performance of our procedures.

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

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

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

$$\frac{1}{n^{\beta_1}}\|\Sigma\|_F^2 = \frac{1}{n^{\beta_1}}\prod_{j=1}^{v} 2(1 + \rho_j^2) = n^{1-\beta_1}\prod_{j=1}^{v}(1 + \rho_j^2).$$

Since $\prod_{j=1}^v(1 + \rho_j^2) \geq 1$, Assumption 4.3 necessarily implies $\beta_1 \geq 1$. When $\beta_1 = 1$, $\frac{1}{n^{\beta_1}}\|\Sigma\|_F^2 = \prod_{j=1}^v(1 + \rho_j^2)$, 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). When $\beta_1 > 1$, Assumption 4.3 is satisfied if $\prod_{j=1}^v(1 + \rho_j^2) = O(n^{\beta_1 - 1})$.

Likewise

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

so that via Lemma B.3 in Section B

$$\frac{1}{n^{\beta_2}}\|\Sigma^{-1}\|_1 = \frac{1}{n^{\beta_2}}\prod_{j=1}^{v}\|\Sigma_j^{-1}\|_1 = n^{1-\beta_2}\prod_{j=1}^{v}\frac{1}{1 - |\rho_j|} = n^{1-\beta_2}\frac{1}{\prod_{j=1}^v(1 - |\rho_j|)}.$$

Since $\prod_{j=1}^{v}\left(1-|\rho_j|\right) \leq 1$, Assumption 5.1 necessarily implies $\beta_2 \geq 1$. When $\beta_2 = 1$, $\frac{1}{n^{\beta_2}}\|\Sigma^{-1}\|_1 = 1/\prod_{j=1}^{v}\left(1-|\rho_j|\right)$, the denominator of which 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). When $\beta_2 > 1$, Assumption 5.1 is satisfied if $\left[\prod_{j=1}^{v}\left(1-|\rho_j|\right)\right]^{-1} = O(n^{\beta_2-1})$.[7]

We consider $\mu = 0$, $n = 2^v$, and $\rho_j = \rho^j$ for $j = 1, \ldots, v$. The number of Monte Carlo simulations is 1000. We compare the quadratic form estimator 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).[8]

The first evaluation criterion is the *relative mean square error* (MSE) in terms of Σ. Given a generic estimator $\hat{\Sigma}_G$ of the covariance matrix Σ, we compute

$$\frac{\mathbb{E}\|\hat{\Sigma}_G - \Sigma\|_F^2}{\|\Sigma\|_F^2}$$

where the expectation operator is taken with respect to all the simulations. Often the precision matrix Σ^{-1} is of more interest than Σ, so we also compute the MSE of the estimator of Σ^{-1}:

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

where the expectation operator is taken with respect to all the 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 next calculate

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

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 Σ 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

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

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

Finally, we consider testing $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. We compute sizes of LM and Wald tests (Theorem 5.1). The significance level is 5%. To investigate power, we generate μ_i as $\mu_i \overset{i.i.d.}{\sim} N(0,1)/\sqrt{T}$ for $i = 1, 2, \ldots, \lfloor n^{0.7} \rfloor$, where $\lfloor x \rfloor$ is the largest integer less than or equal to x; $\mu_i = 0$ for $i = \lfloor n^{0.7} \rfloor + 1, \lfloor n^{0.7} \rfloor + 2, \ldots, n$. These also require invertibility of $\hat{\Sigma}_G$.

The results are reported in Tables 1-3. In Table 1, we set $T = 252$ and $v = 10$ so that $n = 2^v = 1024$; we set $\rho = 0.5, 0.7, 0.85$. First, consider the top panel ($\rho = 0.5$). For the MSE in terms of Σ (i.e., MSE-1), all the estimators beat the sample covariance matrix M_T by a large margin. The quadratic form estimator $\tilde{\Sigma}$ also outperformed the LW04 and LW17 estimators considerably. For the MSE in terms of Σ^{-1} (i.e., MSE-2), a similar pattern exists. Note that the MSE-2 cannot be computed for M_T because M_T is not invertible when $n > T$. For the PRIAL in terms of Σ (i.e., PRIAL-1), again $\tilde{\Sigma}$ is better than the LW04 and LW17 estimators.

[7] Furthermore, the largest eigenvalue of Σ is $\prod_{j=1}^{v}(1+|\rho_j|)$, which converges as $v \to \infty$ if and only if $\sum_{j=1}^{v}|\rho_j|$ converges (Knopp (1947) Theorem 28.3).

[8] 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.

	M_T	$\tilde{\Sigma}$	LW04	LW17
		$\rho = 0.5$		
MSE-1	2.989	0.000	0.242	0.243
MSE-2	NA	0.000	0.311	0.308
PRIAL-1	0	1.000	0.919	0.919
size of LM	NA	0.051	1.000	1.000
size of Wald	NA	0.050	0.085	0.093
		$\rho = 0.7$		
MSE-1	1.760	0.000	0.429	0.430
MSE-2	NA	0.000	0.722	0.715
PRIAL-1	0	1.000	0.756	0.756
size of LM	NA	0.050	1.000	1.000
size of Wald	NA	0.051	0.158	0.164
		$\rho = 0.85$		
MSE-1	0.501	0.001	0.320	0.316
MSE-2	NA	0.002	0.980	0.980
PRIAL-1	0	0.998	0.360	0.370
size of LM	NA	0.051	1.000	1.000
size of Wald	NA	0.060	0.334	0.329

Table 1: M_T, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, quadratic form estimator, Ledoit and Wolf (2004)'s linear shrinkage estimator, and Ledoit and Wolf (2017)'s direct nonlinear shrinkage estimator, respectively. MSE-1 and MSE-2 are the MSE in terms of Σ and Σ^{-1}, respectively. PRIAL-1 is the PRIAL in terms of Σ. $T = 252$ and $n = 2^{10} = 1024$. 0.000 means less than 0.001.

The sample covariance matrix M_T has zero PRIAL-1 by definition. The superiority of $\tilde{\Sigma}$ in this experiment is expected because the true covariance matrix is indeed a Kronecker product.

Considering the size of Wald test, we realize that the quadratic form estimator $\tilde{\Sigma}$ has the correct size while the LW04 and LW17 estimators are over-sized. Note that Wald test is not defined for M_T because M_T is not invertible. Size of LM test is similar to that of Wald test for $\tilde{\Sigma}$, but LM test seems to perform poorly for both the LW04 and LW17 estimators. Undoubtedly, the quadratic form estimator $\tilde{\Sigma}$ is the best performing estimator.

As we increase the "mother" correlation parameter ρ from 0.5 to 0.85, performance of $\tilde{\Sigma}$ remains unchanged across all five criteria. In terms of MSE-1, performance of M_T improves while performances of LW04 and LW17 estimators initially worsen and then improve. In terms of MSE-2, PRIAL-1, the size of the LM test, and the size of the Wald test, the performances of both the LW04 and LW17 estimators worsen. Again the quadratic form estimator $\tilde{\Sigma}$ is the best performing estimator.

Next, we fix ρ at 0.7 and examine effects of n and T; the results are reported in Table 2. If we fix T at 252 and increase v (and hence n), in terms of MSE-1, all the estimators except the quadratic form estimator $\tilde{\Sigma}$ worsen. The same pattern is observed when we use the MSE-2 criterion instead (the sample covariance matrix M_T dropped out in this case). In terms of PRIAL-1, we see that all the candidate estimators are becoming increasingly superior to M_T. As n increases, size of Wald test worsens for all the estimators except $\tilde{\Sigma}$; a similar pattern is observed for LM test. If we increase T from 252 to 504, all the estimators improve in terms of both the MSE-1 and MSE-2 criteria. Also sizes of Wald and LM tests in general improve for all the estimators.

The results of power investigation are reported in Table 3. We see that power of the quadratic form estimator $\tilde{\Sigma}$ is very good for the specified local alternative. Powers of the LW04 and LW17

	M_T	$\tilde{\Sigma}$	LW04	LW17	M_T	$\tilde{\Sigma}$	LW04	LW17
	$n = 2^9$, $T = 252$				$n = 2^9$, $T = 504$			
MSE-1	0.882	0.001	0.346	0.345	0.442	0.000	0.249	0.246
MSE-2	NA	0.001	0.676	0.656	NA	0.000	0.601	0.531
PRIAL-1	0	0.999	0.608	0.609	0	0.999	0.438	0.443
size of LM	NA	0.039	1.000	1.000	NA	0.053	1.000	1.000
size of Wald	NA	0.041	0.153	0.149	NA	0.058	0.148	0.151
	$n = 2^{10}$, $T = 252$				$n = 2^{10}$, $T = 504$			
MSE-1	1.760	0.000	0.429	0.430	0.882	0.000	0.345	0.344
MSE-2	NA	0.000	0.722	0.715	NA	0.000	0.677	0.659
PRIAL-1	0	1.000	0.756	0.756	0	1.000	0.608	0.610
size of LM	NA	0.050	1.000	1.000	NA	0.059	1.000	1.000
size of Wald	NA	0.051	0.158	0.164	NA	0.062	0.168	0.168
	$n = 2^{11}$, $T = 252$				$n = 2^{11}$, $T = 504$			
MSE-1	3.514	0.000	0.489	0.490	1.760	0.000	0.429	0.429
MSE-2	NA	0.000	0.747	0.744	NA	0.000	0.723	0.717
PRIAL-1	0	1.000	0.861	0.861	0	1.000	0.756	0.757
size of LM	NA	0.057	1.000	1.000	NA	0.057	1.000	1.000
size of Wald	NA	0.067	0.202	0.221	NA	0.060	0.169	0.181

Table 2: M_T, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, quadratic form estimator, Ledoit and Wolf (2004)'s linear shrinkage estimator, and Ledoit and Wolf (2017)'s direct nonlinear shrinkage estimator, respectively. MSE-1 and MSE-2 are the MSE in terms of Σ and Σ^{-1}, respectively. PRIAL-1 is the PRIAL in terms of Σ. $\rho = 0.7$. 0.000 means less than 0.001.

	M_T	$\tilde{\Sigma}$	LW04	LW17	M_T	$\tilde{\Sigma}$	LW04	LW17
	$\rho = 0.5$, $n = 2^9$				$\rho = 0.5$, $n = 2^{10}$			
power of LM	NA	0.890	0.730	0.866	NA	0.925	0.999	1.000
power of Wald	NA	0.905	0.689	0.734	NA	0.942	0.750	0.780
	$\rho = 0.7$, $n = 2^9$				$\rho = 0.7$, $n = 2^{10}$			
power of LM	NA	1.000	1.000	1.000	NA	1.000	1.000	1.000
power of Wald	NA	1.000	0.742	0.833	NA	1.000	0.746	0.806
	$\rho = 0.85$, $n = 2^9$				$\rho = 0.85$, $n = 2^{10}$			
power of LM	NA	1.000	1.000	1.000	NA	1.000	1.000	1.000
power of Wald	NA	1.000	0.990	1.000	NA	1.000	0.981	0.977

Table 3: M_T, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, 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. $T = 252$. Powers are not size-adjusted.

estimators in terms of Wald test are less good. Powers of the LW04 and LW17 estimators in terms of the LM test come at a price of their sizes.

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 by Ledoit and Wolf (2004). We still assume that $y_t \sim N(\mu, \Sigma)$. The true covariance matrix Σ is *diagonal* without loss of generality. The diagonal entries Σ_{ii} (i.e., the eigenvalues of Σ) are log normally distributed: $\log \Sigma_{ii} \sim N(\mu_{\text{LW}}, \sigma_{\text{LW}}^2)$. Ledoit and Wolf (2004) defined the *grand mean* μ_{g} of and *cross-sectional dispersion* α^2 of the eigenvalues of Σ as, respectively,

$$\mu_{\text{g}} := \frac{1}{n}\sum_{i=1}^{n}\Sigma_{ii} \qquad \alpha^2 := \frac{1}{n}\sum_{i=1}^{n}(\Sigma_{ii}-\mu_{\text{g}})^2.$$

In the Monte Carlo simulations, we re-define μ_{g} and α^2 as the corresponding population counterparts:

$$\mu_{\text{g}} = \mathbb{E}\Sigma_{ii} = e^{\mu_{\text{LW}}+\sigma_{\text{LW}}^2/2} \qquad \alpha^2 = \operatorname{var}\Sigma_{ii} = e^{2(\mu_{\text{LW}}+\sigma_{\text{LW}}^2)} - e^{2\mu_{\text{LW}}+\sigma_{\text{LW}}^2}.$$

Ledoit and Wolf (2004) set $\mu_{\text{g}} = 1$, so we can solve $\mu_{\text{LW}} = -\log(1+\alpha^2)/2$ and $\sigma_{\text{LW}}^2 = \log(1+\alpha^2)$, 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 Σ_{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), $\mu = 0, n = 20$, $T = 40$, and $\alpha^2 = 0.5$.

There are a few Kronecker products which we can consider to approximate Σ (see Hafner et al. (2019) for more discussions of model selection). The possible Kronecker factorizations are $5\times 2\times 2$, 4×5, 2×10. Within each Kronecker factorization, we can further permute the Kronecker sub-matrices to obtain different Kronecker models. We experiment *all* the Kronecker products and compare with the LW04 and LW17 estimators. All the estimators do not know $\mu = 0$ and have to estimate it, except in the case of LM test.

The results are reported in Table 4. The first observation is that 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 MSE-1 and MSE-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 size of LM test, $\tilde{\Sigma}$ $(2\times 5\times 2)$ has the correct size while the LW04 and LW17 estimators are under-sized. In terms of size of Wald test, all candidate estimators are slightly over-sized.

We next vary α^2. We base the comparisons on the $2\times 5\times 2$ Kronecker product factorization. The results are reported in Table 5. As α^2 increases, performance of M_T actually improves in terms of MSE-1 and MSE-2. On the other hand, performances of $\tilde{\Sigma}$, the LW04 and LW17 estimators worsen in terms of MSE-1, MSE-2, PRIAL-1 and PRIAL-2. The worsening performance of $\tilde{\Sigma}$ is not surprising because α^2 can be interpreted as the distance of Σ from a Kronecker product model. The worsening performance of the LW04 estimator has also been documented by Ledoit and Wolf (2004). As α^2 increases, $\tilde{\Sigma}$ has roughly correct size for LM test while both the LW04 and LW17 estimators are under-sized. In terms of Wald test, all the candidate estimators are slightly over-sized.

	M_T	$\tilde{\Sigma}$ $(5 \times 2 \times 2)$	$\tilde{\Sigma}$ $(2 \times 5 \times 2)$	$\tilde{\Sigma}$ $(2 \times 2 \times 5)$	$\tilde{\Sigma}$ (4×5)
MSE-1	0.446	0.137	0.136	0.137	0.140
MSE-2	6.876	0.154	0.153	0.154	0.163
PRIAL-1	0	0.684	0.685	0.682	0.675
PRIAL-2	0	0.977	0.977	0.977	0.976
size of LM	0.004	0.043	0.050	0.038	0.038
size of Wald	0.690	0.092	0.087	0.081	0.094
	$\tilde{\Sigma}$ (5×4)	$\tilde{\Sigma}$ (2×10)	$\tilde{\Sigma}$ (10×2)	LW04	LW17
MSE-1	0.139	0.189	0.188	0.113	0.129
MSE-2	0.163	0.293	0.288	0.122	0.148
PRIAL-1	0.679	0.570	0.571	0.738	0.702
PRIAL-2	0.976	0.957	0.958	0.982	0.978
size of LM	0.041	0.035	0.028	0.022	0.015
size of Wald	0.100	0.163	0.167	0.074	0.083

Table 4: M_T, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, quadratic form estimator (factorisations given in parentheses), Ledoit and Wolf (2004)'s linear shrinkage estimator, and Ledoit and Wolf (2017)'s direct nonlinear shrinkage estimator, respectively. MSE-1 and MSE-2 are the MSE in terms of Σ and Σ^{-1}, respectively. PRIAL-1 and PRIAL-2 are the PRIAL in terms of Σ and Σ^{-1}, respectively. $n = 20, T = 40, \alpha^2 = 0.5$.

	M_T	$\tilde{\Sigma}$ $(2 \times 5 \times 2)$	LW04	LW17	M_T	$\tilde{\Sigma}$ $(2 \times 5 \times 2)$	LW04	LW17
	$\alpha^2 = 0.25$				$\alpha^2 = 0.50$			
MSE-1	0.492	0.077	0.050	0.070	0.446	0.136	0.113	0.129
MSE-2	7.405	0.089	0.048	0.086	6.876	0.153	0.122	0.148
PRIAL-1	0	0.843	0.898	0.856	0	0.685	0.738	0.702
PRIAL-2	0	0.988	0.993	0.988	0	0.977	0.982	0.978
size of LM	0.004	0.042	0.035	0.020	0.004	0.050	0.022	0.015
size of Wald	0.690	0.083	0.064	0.066	0.690	0.087	0.074	0.083
	M_T	$\tilde{\Sigma}$ $(2 \times 5 \times 2)$	LW04	LW17	M_T	$\tilde{\Sigma}$ $(2 \times 5 \times 2)$	LW04	LW17
	$\alpha^2 = 0.75$				$\alpha^2 = 1$			
MSE-1	0.396	0.195	0.154	0.167	0.353	0.243	0.173	0.184
MSE-2	6.311	0.241	0.194	0.204	5.807	0.335	0.259	0.246
PRIAL-1	0	0.469	0.589	0.557	0	0.218	0.469	0.440
PRIAL-2	0	0.959	0.966	0.966	0	0.934	0.948	0.953
size of LM	0.004	0.058	0.017	0.013	0.004	0.067	0.017	0.013
size of Wald	0.690	0.091	0.087	0.093	0.690	0.106	0.090	0.091

Table 5: M_T, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, quadratic form estimator (factorisations given in parentheses), Ledoit and Wolf (2004)'s linear shrinkage estimator, and Ledoit and Wolf (2017)'s direct nonlinear shrinkage estimator, respectively. MSE-1 and MSE-2 are the MSE in terms of Σ and Σ^{-1}, respectively. PRIAL-1 and PRIAL-2 are the PRIAL in terms of Σ and Σ^{-1}, respectively. $n = 20, T = 40$.

We finally 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: $2 \times 2 \times 2 \times 2$, 4×4, $4 \times 2 \times 2$ and 2×8. For the second variation, we use the Kronecker product factorizations: $5 \times 2 \times 2 \times 2$, $5 \times 2 \times 4$, 5×8 and $10 \times 2 \times 2$. We also considered permutations of sub-matrices for each factorization, but the performances remained relatively unchanged, so we do not report them in the interest of space. The results are reported in Table 6.

Consider the top panel of Table 6 first. All the candidate estimators beat the sample covariance matrix M_T. Performance of the quadratic form estimator $\tilde{\Sigma}$ is relatively robust to the Kronecker product factorizations ($2\times2\times2\times2$, 4×4 and $4\times2\times2$); the best performing one is $4 \times 2 \times 2$. In terms of MSE-1, MSE-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 size of LM test, $\tilde{\Sigma}$ ($4 \times 2 \times 2$) has the correct size while both the LW04 and LW17 estimators are under-sized. In terms of size of Wald test, all the candidate estimators are slightly over-sized.

Next consider the bottom panel of Table 6. 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 MSE-1, MSE-2 and PRIAL-1, $\tilde{\Sigma}$ ($5 \times 2 \times 2 \times 2$) is comparable to the LW04 and LW17 estimators. In terms of size of LM test, $\tilde{\Sigma}$ ($5 \times 2 \times 2 \times 2$) and the LW04 estimator have correct size while the LW17 estimator is slightly over-sized. In terms of size of Wald test, all the candidate estimators are slightly over-sized.

By looking at Tables 4 and 6 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 MSE-1 and MSE-2, performances of the best performing quadratic form estimator as well as the LW04 and LW17 estimators worsen as n/T increases. In terms of size of LM test, the best performing quadratic form estimator always has the correct size, while sizes of Wald tests increase monotonocally with n/T.

7 Concluding Remarks

We have proposed a new estimator of the Kronecker product model for covariance matrices - the quadratic form estimator. We establish the rate of convergence and use the estimated precision matrix to form LM and Wald test statistics. The asymptotic distributions of these test statistics are established under both null and local alternative hypotheses. Testing linear restrictions of the unknown mean vector is also investigated. In Monte Carlo simulations, the quadratic form estimator performs well both when the Kronecker product model is correctly specified and when it is misspecified.

We remark on a number of possible extensions. One can generalize to allow weakly time series dependent data (see Hafner et al. (2019) for some work in this direction), and perhaps to where the spectral density matrix is Kronecker product factored. We may also consider the two-sample case where $\Sigma_1 := \mathbb{E}[(y_{1,t} - \mu_1)(y_{1,t} - \mu_1)^\intercal]$ ($n \times n$), $\Sigma_2 := \mathbb{E}[(y_{2,t} - \mu_2)(y_{2,t} - \mu_2)^\intercal]$ ($n \times n$), $\mu_1 := \mathbb{E}(y_{1,t})$, and $\mu_2 := \mathbb{E}(y_{2,t})$. Cho and Phillips (2018) showed that the hypothesis of $\Sigma_1 = \Sigma_2$ can be tested based on $\text{tr}(\Sigma_1\Sigma_2^{-1}) = n$; if both the covariance matrices have a conformable Kronecker product structure, this simplifies to $\text{tr}(\Sigma_{1,1}\Sigma_{2,1}^{-1}) \times \cdots \times \text{tr}(\Sigma_{1,v}\Sigma_{2,v}^{-1}) = n$.

$n/T = 0.32$	M_T	$\tilde{\Sigma}$ $(2\times2\times2\times2)$	$\tilde{\Sigma}$ (4×4)	$\tilde{\Sigma}$ $(4\times2\times2)$	$\tilde{\Sigma}$ (2×8)	LW04	LW17
MSE-1	0.292	0.118	0.122	0.120	0.145	0.098	0.109
MSE-2	1.491	0.134	0.142	0.137	0.190	0.110	0.118
PRIAL-1	0	0.580	0.571	0.576	0.492	0.655	0.618
PRIAL-2	0	0.907	0.902	0.905	0.870	0.924	0.919
size of LM	0.013	0.057	0.050	0.050	0.041	0.023	0.019
size of Wald	0.373	0.081	0.090	0.080	0.133	0.072	0.074
$n/T = 2$	M_T	$\tilde{\Sigma}$ $(5\times2\times2\times2)$	$\tilde{\Sigma}$ $(5\times2\times4)$	$\tilde{\Sigma}$ (5×8)	$\tilde{\Sigma}$ $(10\times2\times2)$	LW04	LW17
MSE-1	1.684	0.168	0.175	0.216	0.234	0.159	0.196
MSE-2	NA	0.182	0.194	0.286	0.337	0.151	0.164
PRIAL-1	0	0.898	0.894	0.870	0.860	0.904	0.882
PRIAL-2	NA	NA	NA	NA	NA	NA	NA
size of LM	NA	0.051	0.054	0.050	0.049	0.051	0.070
size of Wald	NA	0.155	0.159	0.224	0.260	0.129	0.140

Table 6: M_T, $\tilde{\Sigma}$, LW04 and LW17 stand for the sample covariance matrix, quadratic form estimator (factorisations given in parentheses), Ledoit and Wolf (2004)'s linear shrinkage estimator, and Ledoit and Wolf (2017)'s direct nonlinear shrinkage estimator, respectively. MSE-1 and MSE-2 are the MSE in terms of Σ and Σ^{-1}, respectively. PRIAL-1 and PRIAL-2 are the PRIAL in terms of Σ and Σ^{-1}, respectively. $\alpha^2 = 0.5$.

A Appendix

A.1 Proof of Lemma 4.1

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

$$v \le \log n / \log\left(\min_{1\le j\le v} n_j\right) = O(\log n).$$

For part (ii):

$$\max_{1\le j\le v} \lambda_{\max}(\Sigma_j) \le \max_{1\le j\le v} \mathrm{tr}(\Sigma_j) = \max_{1\le j\le v} n_j < \infty.$$

□

A.2 Proof of Theorem 4.1

We first give an auxiliary lemma and an auxiliary theorem 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) Both $\max_{1\le j\le v}\|\Sigma_j\|_F$ *and* $\max_{1\le j\le v}\|\Sigma_j^{-1}\|_F$ *are bounded from the above by absolute positive constants. Moreover both* $\min_{1\le j\le v}\|\Sigma_j\|_F$ *and* $\min_{1\le j\le v}\|\Sigma_j^{-1}\|_F$ *are bounded away from zero by absolute positive constants.*

(ii) Both $\max_{1\le j\le v}\|\Sigma_j\|_1$ *and* $\max_{1\le j\le v}\|\Sigma_j^{-1}\|_1$ *are bounded from the above by absolute positive constants. Moreover both* $\min_{1\le j\le v}\|\Sigma_j\|_1$ *and* $\min_{1\le j\le v}\|\Sigma_j^{-1}\|_1$ *are bounded away from zero by absolute positive constants.*

(iii) *Both* $\max_{1\le j\le v}\|\Sigma_j\|_{\ell_2}$ *and* $\max_{1\le j\le v}\|\Sigma_j^{-1}\|_{\ell_2}$ *are bounded from the above by absolute positive constants. Moreover both* $\min_{1\le j\le v}\|\Sigma_j\|_{\ell_2}$ *and* $\min_{1\le j\le v}\|\Sigma_j^{-1}\|_{\ell_2}$ *are bounded away from zero by absolute positive constants.*

Proof of Lemma A.1. For part (i), note that

$$\lambda_{\min}(\Sigma_j)\le\lambda_{\max}(\Sigma_j)\le\|\Sigma_j\|_F\le\sqrt{n_j}\lambda_{\max}(\Sigma_j)$$

whence we deduce that $\max_{1\le j\le v}\|\Sigma_j\|_F$ is bounded from the above by an absolute positive constant and $\min_{1\le j\le v}\|\Sigma_j\|_F$ is bounded away from zero by an absolute positive constant via Assumption 4.2 and Lemma 4.1. Similarly, we have

$$\frac{1}{\lambda_{\max}(\Sigma_j)}=\lambda_{\min}(\Sigma_j^{-1})\le\lambda_{\max}(\Sigma_j^{-1})\le\|\Sigma_j^{-1}\|_F\le\sqrt{n_j}\lambda_{\max}(\Sigma_j^{-1})=\sqrt{n_j}\frac{1}{\lambda_{\min}(\Sigma_j)},$$

whence we deduce that $\max_{1\le j\le v}\|\Sigma_j^{-1}\|_F$ is bounded from the above by an absolute positive constant and $\min_{1\le j\le v}\|\Sigma_j^{-1}\|_F$ is bounded away from zero by an absolute positive constant via Assumption 4.2 and Lemma 4.1.

For part (ii), note that

$$\|\Sigma_j\|_F\le\|\Sigma_j\|_1\le n_j\|\Sigma_j\|_F$$
$$\|\Sigma_j^{-1}\|_F\le\|\Sigma_j^{-1}\|_1\le n_j\|\Sigma_j^{-1}\|_F$$

whence we deduce that part (ii) holds via part (i).

For part (iii), we have

$$\max_{1\le j\le v}\|\Sigma_j\|_{\ell_2}\le\max_{1\le j\le v}\|\Sigma_j\|_F$$
$$\max_{1\le j\le v}\|\Sigma_j^{-1}\|_{\ell_2}\le\max_{1\le j\le v}\|\Sigma_j^{-1}\|_F$$

whence we could deduce the first half of the statement via part (i). Next,

$$\min_{1\le j\le v}\|\Sigma_j\|_{\ell_2}=\min_{1\le j\le v}\lambda_{\max}(\Sigma_j)\ge\min_{1\le j\le v}\lambda_{\min}(\Sigma_j)$$

which is bounded away from zero by an absolute positive constant via Assumption 4.2(ii). Finally,

$$\min_{1\le j\le v}\|\Sigma_j^{-1}\|_{\ell_2}=\min_{1\le j\le v}\lambda_{\max}(\Sigma_j^{-1})\ge\min_{1\le j\le v}\lambda_{\min}(\Sigma_j^{-1})=\min_{1\le j\le v}\frac{1}{\lambda_{\max}(\Sigma_j)}=\frac{1}{\max_{1\le j\le v}\lambda_{\max}(\Sigma_j)}$$

which is bounded away from zero by an absolute positive constant via Lemma 4.1(ii). □

A.2.2 Theorem A.1

Theorem A.1. *Suppose Assumptions 4.1, 4.2 and 4.3 hold. Then*

(i)

$$\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}}\left|\hat{d}_{i,j}^{(h)}-d_{i,j}^{(h)}\right|=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

(ii) *We have* $\mathrm{tr}(d^{(h)})/(n_hn_{-h})=\sigma^2>0$ *for* $h=1,\dots,v$. *Also,*

$$\max_{1\le h\le v}\frac{1}{n_hn_{-h}}\left|\mathrm{tr}(\hat{d}^{(h)})-\mathrm{tr}(d^{(h)})\right|=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

As a result, $\min_{1\le h\le v}\mathrm{tr}(\hat{d}^{(h)})/(n_hn_{-h})$ *is bounded away from zero by an absolute positive constant in probability.*

(iii)

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h-\Sigma_h\|_\infty=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\left|[\tilde{\Sigma}_h]_{i,j}-[\Sigma_h]_{i,j}\right|=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right),$$

where $[\tilde{\Sigma}_h]_{i,j}$ and $[\Sigma_h]_{i,j}$ are the (i,j)th entry of $\tilde{\Sigma}_h$ and Σ_h, respectively.

(iv)

$$\left|\hat{\sigma}^2-\sigma^2\right|=O_p\left(\sqrt{\frac{1}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

(v)

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h-\Sigma_h\|_F=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

(vi)

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h^{-1}-\Sigma_h^{-1}\|_F=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

(vii)

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h-\Sigma_h\|_1=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

(viii)

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h^{-1}-\Sigma_h^{-1}\|_1=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

(ix)

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h-\Sigma_h\|_{\ell_2}=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

(x)

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h^{-1}-\Sigma_h^{-1}\|_{\ell_2}=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

Proof. For part (i), note that $d_{i,j}^{(h)}=\operatorname{tr}\Sigma_{\{[i,j]\}}^{(h)}$, where $\Sigma_{\{[i,j]\}}^{(h)}$ is the $[i,j]$th block of $\Sigma^{(h)}$ (each block is $n_{-h}\times n_{-h}$ dimensional) for $i,j=1,\dots,n_h$. Similarly, $\hat{d}_{i,j}^{(h)}=\operatorname{tr}M_{T,\{[i,j]\}}^{(h)}$, where $M_{T,\{[i,j]\}}^{(h)}$ is the $[i,j]$th block of $M_T^{(h)}$ (each block is $n_{-h}\times n_{-h}$ dimensional). Write

$$\hat{d}_{i,j}^{(h)}=\operatorname{tr}M_{T,\{[i,j]\}}^{(h)}=\operatorname{tr}M_{T,\{[i,j]\}}^{0,(h)}-\operatorname{tr}\left(\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right)_{\{[i,j]\}}=:\hat{d}_{i,j}^{0,(h)}-\operatorname{tr}\left(\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right)_{\{[i,j]\}}$$

where

$$\begin{aligned}
M_T^0&:=\frac{1}{T}\sum_{t=1}^T(y_t-\mu)(y_t-\mu)^\intercal\\
M_T^{0,(h)}&:=K_{n_h\times\cdots\times n_v,n_1\times\cdots\times n_{h-1}}M_T^0K_{n_1\times\cdots\times n_{h-1},n_h\times\cdots\times n_v}\\
\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}&:=K_{n_h\times\cdots\times n_v,n_1\times\cdots\times n_{h-1}}(\bar{y}-\mu)(\bar{y}-\mu)^\intercal K_{n_1\times\cdots\times n_{h-1},n_h\times\cdots\times n_v},
\end{aligned}$$

$M^{0,(h)}_{T,\{[i,j]\}}$ is the $[i,j]$th block of $M^{0,(h)}_T$ (each block is $n_{-h}\times n_{-h}$ dimensional), and $\left(\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right)_{\{[i,j]\}}$ is the $[i,j]$th block of $\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}$ (each block is $n_{-h}\times n_{-h}$ dimensional). Thus we have

$$\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}}\left|\hat{d}^{(h)}_{i,j}-d^{(h)}_{i,j}\right|$$
$$\le \max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}}\left|\hat{d}^{0,(h)}_{i,j}-d^{(h)}_{i,j}\right| + \max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}}\left|\operatorname{tr}\left(\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right)_{\{[i,j]\}}\right| \quad (A.1)$$

We consider the first term of (A.1) first. Note that $\mathbb{E}[\hat{d}^{0,(h)}_{i,j}]=d^{(h)}_{i,j}$. Write for some $M>0$

$$\mathbb{P}\left(\max_{1\le h\le v}\max_{1\le i,j\le n_h}\sqrt{\frac{n^{2-\beta_1}T}{\log n}}\frac{1}{n_{-h}}|\hat{d}^{0,(h)}_{i,j}-d^{(h)}_{i,j}|>M\right)=\mathbb{P}\left(\bigcup_{1\le h\le v}\bigcup_{1\le i,j\le n_h}\left\{\sqrt{\frac{n^{2-\beta_1}T}{\log n}}\frac{1}{n_{-h}}|\hat{d}^{0,(h)}_{i,j}-d^{(h)}_{i,j}|>M\right\}\right)$$
$$\le\sum_{h=1}^{v}\sum_{i=1}^{n_h}\sum_{j=1}^{n_h}\mathbb{P}\left(\sqrt{\frac{n^{2-\beta_1}T}{\log n}}\frac{1}{n_{-h}}|\hat{d}^{0,(h)}_{i,j}-d^{(h)}_{i,j}|>M\right)\le\frac{n^{2-\beta_1}T\sum_{h=1}^{v}\sum_{i=1}^{n_h}\sum_{j=1}^{n_h}\operatorname{var}(\hat{d}^{0,(h)}_{i,j}/n_{-h})}{\log n\cdot M^2}$$
$$\le\frac{v\max_{1\le h\le v}n_h^2 n^{2-\beta_1}T\max_{1\le h\le v}\max_{1\le i,j\le n_h}\operatorname{var}(\hat{d}^{0,(h)}_{i,j}/n_{-h})}{\log n\cdot M^2}$$

where the second inequality is due to Chebyshev's inequality. We now show that

$$\max_{1\le h\le v}\max_{1\le i,j\le n_h}\operatorname{var}(\hat{d}^{0,(h)}_{i,j}/n_{-h})=O\left(\frac{1}{n^{2-\beta_1}T}\right).$$

For arbitrary $i,j=1,\ldots,n_h$,

$$\operatorname{var}(\hat{d}^{0,(h)}_{i,j}/n_{-h})=\frac{1}{n_{-h}^2}\operatorname{var}\left(\sum_{k=1}^{n_{-h}}\left[M^{0,(h)}_{T,\{[i,j]\}}\right]_{kk}\right)=\frac{1}{n_{-h}^2}\operatorname{var}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{n_{-h}}\dot{y}^{(h)}_{t,(i-1)n_{-h}+k}\dot{y}^{(h)}_{t,(j-1)n_{-h}+k}\right)$$
$$=\frac{1}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}\operatorname{cov}\left(\dot{y}^{(h)}_{t,(i-1)n_{-h}+k}\dot{y}^{(h)}_{t,(j-1)n_{-h}+k},\dot{y}^{(h)}_{t,(i-1)n_{-h}+\ell}\dot{y}^{(h)}_{t,(j-1)n_{-h}+\ell}\right)$$
$$\le\frac{C}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}\operatorname{cov}\left(\dot{z}^{(h)}_{t,(i-1)n_{-h}+k}\dot{z}^{(h)}_{t,(j-1)n_{-h}+k},\dot{z}^{(h)}_{t,(i-1)n_{-h}+\ell}\dot{z}^{(h)}_{t,(j-1)n_{-h}+\ell}\right), \quad (A.2)$$

where $\dot{y}^{(h)}_t:=K_{n_h\times\cdots\times n_v,n_1\times\cdots\times n_{h-1}}(y_t-\mu)$ such that $\mathbb{E}[\dot{y}^{(h)}_t\dot{y}^{(h)\intercal}_t]=\Sigma^{(h)}$ and $\dot{z}^{(h)}_t$ 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

$$\operatorname{var}\left(\operatorname{vec}(\dot{z}^{(h)}_t\dot{z}^{(h)\intercal}_t)\right)=\operatorname{var}\left(\dot{z}^{(h)}_t\otimes\dot{z}^{(h)}_t\right)=2D_nD_n^+(\Sigma^{(h)}\otimes\Sigma^{(h)})=\left(I_{n^2}+K_{n,n}\right)(\Sigma^{(h)}\otimes\Sigma^{(h)}),$$

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.2) is some element of $\left(I_{n^2}+K_{n,n}\right)(\Sigma^{(h)}\otimes\Sigma^{(h)})$. We need to determine the exact position of the summand on the right side of (A.2) in $\left(I_{n^2}+K_{n,n}\right)(\Sigma^{(h)}\otimes\Sigma^{(h)})$. 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 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

$$\operatorname{cov}\left(\dot{z}^{(h)}_{t,(i-1)n_{-h}+k}\dot{z}^{(h)}_{t,(j-1)n_{-h}+k},\dot{z}^{(h)}_{t,(i-1)n_{-h}+\ell}\dot{z}^{(h)}_{t,(j-1)n_{-h}+\ell}\right)$$

corresponds to

$$(\Sigma^{(h)} \otimes \Sigma^{(h)})_{\left\{ \left[(i-1)n_{-h}+k,(i-1)n_{-h}+\ell\right], \left[(j-1)n_{-h}+k,(j-1)n_{-h}+\ell\right] \right\}}. \tag{A.3}$$

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, 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

$$\operatorname{cov}\left(\dot{z}^{(h)}_{t,(i-1)n_{-h}+k}\dot{z}^{(h)}_{t,(j-1)n_{-h}+k}, \dot{z}^{(h)}_{t,(i-1)n_{-h}+\ell}\dot{z}^{(h)}_{t,(j-1)n_{-h}+\ell}\right)$$

corresponds to

$$\begin{aligned}
&[K_{n,n}(\Sigma^{(h)} \otimes \Sigma^{(h)})]_{\left\{ \left[(i-1)n_{-h}+k,(i-1)n_{-h}+\ell\right], \left[(j-1)n_{-h}+k,(j-1)n_{-h}+\ell\right] \right\}} \\
&= (\Sigma^{(h)} \otimes \Sigma^{(h)})_{\left\{ \left[(j-1)n_{-h}+k,(i-1)n_{-h}+\ell\right], \left[(i-1)n_{-h}+k,(j-1)n_{-h}+\ell\right] \right\}}.
\end{aligned} \tag{A.4}$$

Using (A.3) and (A.4), we have

$$
\begin{aligned}
&\max_{1\le h\le v}\max_{1\le i,j\le n_h}\operatorname{var}(\hat{d}_{i,j}^{0,(h)}/n_{-h})\\
&=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}\operatorname{cov}\left(\dot{y}_{t,(i-1)n_{-h}+k}^{(h)}\dot{y}_{t,(j-1)n_{-h}+k}^{(h)},\dot{y}_{t,(i-1)n_{-h}+\ell}^{(h)}\dot{y}_{t,(j-1)n_{-h}+\ell}^{(h)}\right)\\
&\le\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{C}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}\operatorname{cov}\left(\dot{z}_{t,(i-1)n_{-h}+k}^{(h)}\dot{z}_{t,(j-1)n_{-h}+k}^{(h)},\dot{z}_{t,(i-1)n_{-h}+\ell}^{(h)}\dot{z}_{t,(j-1)n_{-h}+\ell}^{(h)}\right)\\
&=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{C}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}(\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]\}}\\
&\quad+\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{C}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}(\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]\}}\\
&=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{C}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}\left(\Sigma_{(i-1)n_{-h}+k,(i-1)n_{-h}+\ell}^{(h)}\cdot\Sigma_{(j-1)n_{-h}+k,(j-1)n_{-h}+\ell}^{(h)}\right)\\
&\quad+\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{C}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}\left(\Sigma_{(j-1)n_{-h}+k,(i-1)n_{-h}+\ell}^{(h)}\cdot\Sigma_{(i-1)n_{-h}+k,(j-1)n_{-h}+\ell}^{(h)}\right)\\
&=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{C\sigma^4}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}\Big[[\Sigma_h]_{i,i}\cdot[\Sigma_{-h}]_{k,\ell}\cdot[\Sigma_h]_{j,j}\cdot[\Sigma_{-h}]_{k,\ell}+[\Sigma_h]_{j,i}\cdot[\Sigma_{-h}]_{k,\ell}\cdot[\Sigma_h]_{i,j}\cdot[\Sigma_{-h}]_{k,\ell}\Big]\\
&=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\left([\Sigma_h]_{i,i}\cdot[\Sigma_h]_{j,j}+[\Sigma_h]_{i,j}\cdot[\Sigma_h]_{j,i}\right)\frac{C\sigma^4}{n_{-h}^2T}\sum_{k=1}^{n_{-h}}\sum_{\ell=1}^{n_{-h}}[\Sigma_{-h}]_{k,\ell}^2\\
&=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\left([\Sigma_h]_{i,i}\cdot[\Sigma_h]_{j,j}+[\Sigma_h]_{i,j}\cdot[\Sigma_h]_{j,i}\right)\frac{C\sigma^4}{n_{-h}^2T}\|\Sigma_{-h}\|_F^2\le\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{2C[\Sigma_h]_{i,i}[\Sigma_h]_{j,j}\sigma^4}{n_{-h}^2T}\|\Sigma_{-h}\|_F^2\\
&=\max_{1\le h\le v}\frac{O(1)\sigma^4}{n_{-h}^2T}\|\Sigma_{-h}\|_F^2=\max_{1\le h\le v}\frac{O(1)n_h^2\sigma^4}{n^2T\|\Sigma_h\|_F^2}\left(\|\Sigma_{-h}\|_F^2\|\Sigma_h\|_F^2\right)=O\left(\frac{1}{n^{2-\beta_1}T}\right)\max_{1\le h\le v}\left(\frac{\sigma^4}{n^{\beta_1}}\|\Sigma_{-h}\|_F^2\|\Sigma_h\|_F^2\right)\\
&=O\left(\frac{1}{n^{2-\beta_1}T}\right),
\end{aligned}
$$

where the second inequality is due to Cauchy-Schwarz inequality $[\Sigma_h]_{i,j}\le\sqrt{[\Sigma_h]_{i,i}}\sqrt{[\Sigma_h]_{j,j}}$ using the fact that Σ_h is a covariance matrix, the fourth last equality uses the fact that $\max_{1\le h\le v}\max_{1\le i\le n_h}[\Sigma_h]_{i,i}\le\max_{1\le h\le v}\lambda_{\max}(\Sigma_h)<\infty$, the second last equality is due to Lemma A.1, and the last equality is due to Assumption 4.3. We hence have

$$
\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}}\left|\hat{d}_{i,j}^{0,(h)}-d_{i,j}^{(h)}\right|=O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right). \tag{A.5}
$$

We now consider the second term of (A.1).

$$
\begin{aligned}
&\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}}\left|\operatorname{tr}\left(\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right)_{\{[i,j]\}}\right|=\max_{1\le h\le v}\max_{1\le i,j\le n_h}\frac{1}{n_{-h}}\left|\sum_{k=1}^{n_{-h}}\left[\left(\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right)_{\{[i,j]\}}\right]_{kk}\right|\\
&\le\max_{1\le h\le v}\max_{1\le i,j\le n_h}\max_{1\le k\le n_{-h}}\left|\left[\left(\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right)_{\{[i,j]\}}\right]_{kk}\right|\le\max_{1\le h\le v}\left\|\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]^{(h)}\right\|_\infty\\
&=\left\|(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right\|_\infty=\left[\max_{1\le i\le n}\left|\frac{1}{T}\sum_{t=1}^{T}(y_{t,i}-\mathbb{E}y_{t,i})\right|\right]^2=O_p\left(\frac{\log n}{T}\right)
\end{aligned}\tag{A.6}
$$

where the last equality is due to Lemma B.1 in Section B. Inserting (A.5) and (A.6) into (A.1) delivers part (i).

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

$$\operatorname{tr}(d^{(h)})/(n_h n_{-h}) = \frac{1}{n_{-h}}\sigma^2 \operatorname{tr}(\Sigma_{-h}) = \sigma^2 \frac{1}{n_{-h}} \operatorname{tr}(\Sigma_{h+1}) \times \cdots \times \operatorname{tr}(\Sigma_v) \times \operatorname{tr}(\Sigma_1) \times \cdots \times \operatorname{tr}(\Sigma_{h-1})$$
$$= \sigma^2 > 0.$$

Now write

$$\max_{1\le h\le v} \frac{1}{n_h n_{-h}} \left|\operatorname{tr}(\hat{d}^{(h)}) - \operatorname{tr}(d^{(h)})\right| = \max_{1\le h\le v} \frac{1}{n_h n_{-h}} \left|\sum_{i=1}^{n_h} (\hat{d}^{(h)}_{i,i} - d^{(h)}_{i,i})\right| \le \max_{1\le h\le v} \frac{1}{n_h n_{-h}} \sum_{i=1}^{n_h} \left|\hat{d}^{(h)}_{i,i} - d^{(h)}_{i,i}\right|$$
$$\le \max_{1\le h\le v} \max_{1\le i\le n_h} \frac{1}{n_{-h}} \left|\hat{d}^{(h)}_{i,i} - d^{(h)}_{i,i}\right| = O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log n}{T}\right),$$

where the last equality is due to part (i). The last part of part (ii) also follows.

For part (iii), write

$$\max_{1\le h\le v} \max_{1\le i,j\le n_h} \left|[\tilde{\Sigma}_h]_{i,j} - [\Sigma_h]_{i,j}\right| \le \max_{1\le h\le v} \max_{1\le i,j\le n_h} \left|\frac{\hat{d}^{(h)}_{i,j}}{\operatorname{tr}(\hat{d}^{(h)})/n_h} - \frac{d^{(h)}_{i,j}}{\operatorname{tr}(\hat{d}^{(h)})/n_h}\right|$$
$$+ \max_{1\le h\le v} \max_{1\le i,j\le n_h} \left|\frac{d^{(h)}_{i,j}}{\operatorname{tr}(\hat{d}^{(h)})/n_h} - \frac{d^{(h)}_{i,j}}{\operatorname{tr}(d^{(h)})/n_h}\right| \tag{A.7}$$

Consider the first term on the right side of (A.7).

$$\max_{1\le h\le v} \max_{1\le i,j\le n_h} \left|\frac{\hat{d}^{(h)}_{i,j}}{\operatorname{tr}(\hat{d}^{(h)})/n_h} - \frac{d^{(h)}_{i,j}}{\operatorname{tr}(\hat{d}^{(h)})/n_h}\right| = \max_{1\le h\le v} \max_{1\le i,j\le n_h} \frac{1}{\operatorname{tr}(\hat{d}^{(h)})/(n_h n_{-h})} \frac{1}{n_{-h}} \left|\hat{d}^{(h)}_{i,j} - d^{(h)}_{i,j}\right|$$
$$= O_p(1) \max_{1\le h\le v} \max_{1\le i,j\le n_h} \frac{1}{n_{-h}} \left|\hat{d}^{(h)}_{i,j} - d^{(h)}_{i,j}\right| = O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log n}{T}\right),$$

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.7).

$$\max_{1\le h\le v} \max_{1\le i,j\le n_h} \left|\frac{d^{(h)}_{i,j}}{\operatorname{tr}(\hat{d}^{(h)})/n_h} - \frac{d^{(h)}_{i,j}}{\operatorname{tr}(d^{(h)})/n_h}\right| = \max_{1\le h\le v} \max_{1\le i,j\le n_h} \frac{\left|\frac{\operatorname{tr}(d^{(h)})}{n_h n_{-h}} - \frac{\operatorname{tr}(\hat{d}^{(h)})}{n_h n_{-h}}\right|}{\left|\frac{\operatorname{tr}(\hat{d}^{(h)})}{n_h n_{-h}} \cdot \frac{\operatorname{tr}(d^{(h)})}{n_h n_{-h}}\right|} \frac{1}{n_{-h}} \left|d^{(h)}_{i,j}\right|$$
$$= \left[O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log n}{T}\right)\right] \max_{1\le h\le v} \max_{1\le i,j\le n_h} \frac{|d^{(h)}_{i,j}|}{n_{-h}}$$
$$= \left[O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log n}{T}\right)\right] \max_{1\le h\le v} \max_{1\le i,j\le n_h} \frac{\sigma^2 |[\Sigma_h]_{i,j}| \operatorname{tr}(\Sigma_{-h})}{n_{-h}}$$
$$= \left[O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log n}{T}\right)\right] \max_{1\le h\le v} \max_{1\le i,j\le n_h} \sigma^2 |[\Sigma_h]_{i,j}|$$
$$= \left[O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log n}{T}\right)\right] \sigma^2 \max_{1\le h\le v} \lambda_{\max}(\Sigma_h)$$
$$= O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log n}{T}\right),$$

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), write

$$
\begin{aligned}
\left|\hat{\sigma}^2-\sigma^2\right| &= \left|\frac{1}{n}\operatorname{tr}(M_T-\Sigma)\right| = \left|\frac{1}{n}\operatorname{tr}\left(M_T^0-(\bar{y}-\mu)(\bar{y}-\mu)^\intercal-\Sigma\right)\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n}\left(M_{T,i,i}^0-\left[(\bar{y}-\mu)(\bar{y}-\mu)^\intercal\right]_{i,i}-\Sigma_{i,i}\right)\right| \\
&\leq \left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[(y_{t,i}-\mu_i)^2-\mathbb{E}(y_{t,i}-\mu_i)^2\right]\right| + \left[\max_{1\leq i\leq n}\left|\frac{1}{T}\sum_{t=1}^{T}(y_{t,i}-\mu_i)\right|\right]^2 \\
&= \left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\dot{y}_{t,i}^2-\mathbb{E}\dot{y}_{t,i}^2\right]\right| + O_p\left(\frac{\log n}{T}\right) \qquad \text{(A.8)}
\end{aligned}
$$

where $\dot{y}_{t,i} := y_{t,i}-\mu_i$, and the last equality is due to Lemma B.1 in Section B. We now establish a rate for the first term in (A.8). For some $M>0$,

$$
\mathbb{P}\left(\left|\frac{\sqrt{n^{2-\beta_1}T}}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\dot{y}_{t,i}^2-\mathbb{E}\dot{y}_{t,i}^2)\right|>M\right) \leq \frac{n^{2-\beta_1}T\operatorname{var}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{y}_{t,i}^2\right)}{M^2}.
$$

We now show $\operatorname{var}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}y_{t,i}^2\right)=O(1/(n^{2-\beta_1}T))$.

$$
\begin{aligned}
&\operatorname{var}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{y}_{t,i}^2\right) = \frac{1}{T}\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n}\dot{y}_{t,i}^2\right) = \frac{1}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}\left(\dot{y}_{t,i}\dot{y}_{t,i},\dot{y}_{t,j}\dot{y}_{t,j}\right) \\
&\leq \frac{C}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}\left(\dot{z}_{t,i}\dot{z}_{t,i},\dot{z}_{t,j}\dot{z}_{t,j}\right) = \frac{C}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left((\Sigma\otimes\Sigma)_{\{[i,j],[i,j]\}}+\left(K_{n,n}(\Sigma\otimes\Sigma)\right)_{\{[i,j],[i,j]\}}\right) \\
&= \frac{2C}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\Sigma\otimes\Sigma)_{\{[i,j],[i,j]\}} = \frac{2C}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Sigma_{i,j}\cdot\Sigma_{i,j} = \frac{2C}{Tn^{2-\beta_1}}\frac{1}{n^{\beta_1}}\|\Sigma\|_F^2 = O\left(\frac{1}{Tn^{2-\beta_1}}\right)
\end{aligned}
$$

where the first equality is due to independence over t of Assumption 4.1(i), the first inequality is due to Assumption 4.1(iii), the third and fourth equalities are due to the similar arguments which we used to prove part (i), and the last equality is due to Assumption 4.3. Thus we have

$$
\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\dot{y}_{t,i}^2-\mathbb{E}\dot{y}_{t,i}^2)\right| = O_p\left(\sqrt{\frac{1}{n^{2-\beta_1}T}}\right).
$$

Substituting this into (A.8) delivers part (iv).

For part (v), we have

$$
\max_{1\leq h\leq v}\|\tilde{\Sigma}_h-\Sigma_h\|_F \leq \max_{1\leq h\leq v} n_h\|\tilde{\Sigma}_h-\Sigma_h\|_\infty = O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right)
$$

where the last equality is due to part (iii). For part (vi), invoke Lemma B.4 and use that $\max_{1\leq h\leq v}\|\Sigma_h^{-1}\|_F=O(1)$ in Lemma A.1.

For part (vii), we have

$$
\max_{1\leq h\leq v}\|\tilde{\Sigma}_h-\Sigma_h\|_1 \leq \max_{1\leq h\leq v} n_h\|\tilde{\Sigma}_h-\Sigma_h\|_F = O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).
$$

For part (viii), we have

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h^{-1}-\Sigma_h^{-1}\|_1 \le \max_{1\le h\le v} n_h\|\tilde{\Sigma}_h^{-1}-\Sigma_h^{-1}\|_F = O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

For part (ix), we have

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h-\Sigma_h\|_{\ell_2} \le \max_{1\le h\le v}\|\tilde{\Sigma}_h-\Sigma_h\|_F = O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

For part (x), we have

$$\max_{1\le h\le v}\|\tilde{\Sigma}_h^{-1}-\Sigma_h^{-1}\|_{\ell_2} \le \max_{1\le h\le v}\|\tilde{\Sigma}_h^{-1}-\Sigma_h^{-1}\|_F = O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right).$$

□

A.2.3 Proof of Theorem 4.1

Proof of Theorem 4.1. For part (i),

$$\begin{aligned}&\big\|\tilde{\Sigma}-\Sigma\big\|_F/\|\Sigma\|_F=\big\|\hat{\sigma}^2\times\tilde{\Sigma}_1\otimes\cdots\otimes\tilde{\Sigma}_v-\sigma^2\times\Sigma_1\otimes\cdots\otimes\Sigma_v\big\|_F/\|\Sigma\|_F=\\&\big\|\hat{\sigma}^2\times\tilde{\Sigma}_1\otimes\cdots\otimes\tilde{\Sigma}_v-\hat{\sigma}^2\times\Sigma_1\otimes\cdots\otimes\Sigma_v+\hat{\sigma}^2\times\Sigma_1\otimes\cdots\otimes\Sigma_v-\sigma^2\times\Sigma_1\otimes\cdots\otimes\Sigma_v\big\|_F/\|\Sigma\|_F\\&\le\hat{\sigma}^2\big\|\tilde{\Sigma}_1\otimes\cdots\otimes\tilde{\Sigma}_v-\Sigma_1\otimes\cdots\otimes\Sigma_v\big\|_F/\|\Sigma\|_F+|\hat{\sigma}^2-\sigma^2|\big\|\Sigma_1\otimes\cdots\otimes\Sigma_v\big\|_F/\|\Sigma\|_F \qquad (A.9)\end{aligned}$$

We consider the first term in (A.9). By inserting terms like $\Sigma_1\otimes\tilde{\Sigma}_2\otimes\cdots\otimes\tilde{\Sigma}_v$ and the triangular inequality, we have

$$\begin{aligned}&\big\|\tilde{\Sigma}_1\otimes\cdots\otimes\tilde{\Sigma}_v-\Sigma_1\otimes\cdots\otimes\Sigma_v\big\|_F\le\\&\big\|\tilde{\Sigma}_1-\Sigma_1\big\|_F\prod_{\ell=2}^{v}\|\tilde{\Sigma}_\ell\|_F+\sum_{j=2}^{v-1}\left(\left[\prod_{k=1}^{j-1}\|\Sigma_k\|_F\right]\big\|\tilde{\Sigma}_j-\Sigma_j\big\|_F\left[\prod_{\ell=j+1}^{v}\|\tilde{\Sigma}_\ell\|_F\right]\right)+\left[\prod_{k=1}^{v-1}\|\Sigma_k\|_F\right]\big\|\tilde{\Sigma}_v-\Sigma_v\big\|_F.\end{aligned} \tag{A.10}$$

We first divide the first term of (A.10) by $\prod_{\ell=1}^{v}\|\Sigma_\ell\|_F$. We have

$$\begin{aligned}&\frac{\big\|\tilde{\Sigma}_1-\Sigma_1\big\|_F\prod_{\ell=2}^{v}\|\tilde{\Sigma}_\ell\|_F}{\prod_{\ell=1}^{v}\|\Sigma_\ell\|_F}=\frac{\big\|\tilde{\Sigma}_1-\Sigma_1\big\|_F}{\|\Sigma_1\|_F}\prod_{\ell=2}^{v}\frac{\|\tilde{\Sigma}_\ell\|_F}{\|\Sigma_\ell\|_F}\le\frac{\big\|\tilde{\Sigma}_1-\Sigma_1\big\|_F}{\|\Sigma_1\|_F}\prod_{\ell=2}^{v}\left[1+\frac{\|\tilde{\Sigma}_\ell-\Sigma_\ell\|_F}{\|\Sigma_\ell\|_F}\right]\\&\le\frac{\big\|\tilde{\Sigma}_1-\Sigma_1\big\|_F}{\|\Sigma_1\|_F}\left[1+\frac{\max_{1\le k\le v}\|\tilde{\Sigma}_k-\Sigma_k\|_F}{\min_{1\le k\le v}\|\Sigma_k\|_F}\right]^{v-1}.\end{aligned} \tag{A.11}$$

We next divide the summand of the second term of (A.10) by $\prod_{\ell=1}^{v}\|\Sigma_\ell\|_F$. We have for $j=2,\ldots,v-1$

$$\begin{aligned}&\frac{\left[\prod_{k=1}^{j-1}\|\Sigma_k\|_F\right]\big\|\tilde{\Sigma}_j-\Sigma_j\big\|_F\left[\prod_{\ell=j+1}^{v}\|\tilde{\Sigma}_\ell\|_F\right]}{\prod_{\ell=1}^{v}\|\Sigma_\ell\|_F}=\frac{\big\|\tilde{\Sigma}_j-\Sigma_j\big\|_F}{\|\Sigma_j\|_F}\prod_{\ell=j+1}^{v}\frac{\|\tilde{\Sigma}_\ell\|_F}{\|\Sigma_\ell\|_F}\\&\le\frac{\big\|\tilde{\Sigma}_j-\Sigma_j\big\|_F}{\|\Sigma_j\|_F}\prod_{\ell=j+1}^{v}\left[1+\frac{\|\tilde{\Sigma}_\ell-\Sigma_\ell\|_F}{\|\Sigma_\ell\|_F}\right]\le\frac{\big\|\tilde{\Sigma}_j-\Sigma_j\big\|_F}{\|\Sigma_j\|_F}\left[1+\frac{\max_{1\le k\le v}\|\tilde{\Sigma}_k-\Sigma_k\|_F}{\min_{1\le k\le v}\|\Sigma_k\|_F}\right]^{v-j}.\end{aligned} \tag{A.12}$$

We finally divide the third term of (A.10) by $\prod_{\ell=1}^{v}\|\Sigma_\ell\|_F$. We have

$$\frac{\left[\prod_{k=1}^{v-1}\|\Sigma_k\|_F\right]\left\|\tilde{\Sigma}_v-\Sigma_v\right\|_F}{\prod_{\ell=1}^{v}\|\Sigma_\ell\|_F}=\frac{\left\|\tilde{\Sigma}_v-\Sigma_v\right\|_F}{\|\Sigma_v\|_F}. \tag{A.13}$$

Thus we have

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

$$\leq\frac{\hat{\sigma}^2}{\sigma^2}\left(1+\frac{\max_{1\leq k\leq v}\|\tilde{\Sigma}_k-\Sigma_k\|_F}{\min_{1\leq k\leq v}\|\Sigma_k\|_F}\right)^{v-1}\sum_{j=1}^{v}\frac{\left\|\tilde{\Sigma}_j-\Sigma_j\right\|_F}{\|\Sigma_j\|_F}=\frac{\hat{\sigma}^2}{\sigma^2}O_p(1)\sum_{j=1}^{v}\frac{\left\|\tilde{\Sigma}_j-\Sigma_j\right\|_F}{\|\Sigma_j\|_F}$$

$$=O_p(1)\sum_{j=1}^{v}\left\|\tilde{\Sigma}_j-\Sigma_j\right\|_F=vO_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\vee\frac{\log n}{T}\right)=O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log^2 n}{T}\right)$$

where the first inequality is due to that $\|\Sigma\|_F=\sigma^2\prod_{j=1}^{v}\|\Sigma_j\|_F$ via Lemma B.3, (A.11), (A.12) and (A.13), the first equality is due to Lemma A.1 and Theorem A.1(v)[9], the second equality is due to Lemma A.1 and Theorem A.1(iv), and the third equality is due to Theorem A.1(v).

We now consider the second term in (A.9).

$$|\hat{\sigma}^2-\sigma^2|\left\|\Sigma_1\otimes\cdots\otimes\Sigma_v\right\|_F/\|\Sigma\|_F=\frac{|\hat{\sigma}^2-\sigma^2|}{\sigma^2}=O_p\left(\sqrt{\frac{1}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log n}{T}\right)$$

where the last equality is due to Theorem A.1(iv). Part (ii)-(vi) of the theorem could be established in a similar manner, so we omit the details. □

A.3 Proof of Theorem 5.1

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

A.3.1 Theorem A.2

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

Theorem A.2. *Consider $\{\varepsilon_{T,i}:1\leq i\leq n,n\geq 1,T\geq 1\}$, an array of real numbers $\{b_{T,i}:1\leq i\leq n,n\geq 1,T\geq 1\}$ and $Q_{n,T}:=\sum_{i=1}^{n}\varepsilon_{T,i}^2+\sum_{i=1}^{n}b_{T,i}\varepsilon_{T,i}$. Suppose that*

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

(ii)

$$\limsup_{T\to\infty}\sup_{n\geq 1}\sup_{1\leq i\leq n}\mathbb{E}\left|\varepsilon_{T,i}\right|^{4+2\delta}<\infty$$

$$\limsup_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}|b_{T,i}|^{2+\delta}<\infty$$

for some $\delta>0$.

[9]To see this:

$$\left(1+\frac{\max_{1\leq k\leq v}\|\tilde{\Sigma}_k-\Sigma_k\|_F}{\min_{1\leq k\leq v}\|\Sigma_k\|_F}\right)^{v-1}=\left(1+O_p\left(\sqrt{\frac{\log n}{n^{2-\beta_1}T}}\vee\frac{\log n}{T}\right)\right)^{O(\log n)}=O_p(1)$$

where the last equality could be deduced from the the facts those $\lim_{x\to\infty}(1+1/x)^x=e$ and $\log^3 n/T\to 0$.

(iii)

$$\liminf_{n,T\to\infty} \frac{1}{n}\operatorname{var}(Q_{n,T}) \geq C > 0$$

for some absolute positive constant C.

Then as $n, T \to \infty$,

$$\frac{Q_{n,T} - \mathbb{E}[Q_{n,T}]}{\sqrt{\operatorname{var}(Q_{n,T})}} \xrightarrow{d} N(0,1).$$

Proof. We can calculate that

$$\mathbb{E}[Q_{n,T}] = \mathbb{E}\left[\sum_{i=1}^{n} \varepsilon_{T,i}^2 + \sum_{i=1}^{n} b_{T,i}\varepsilon_{T,i}\right] = \sum_{i=1}^{n} \mathbb{E}\left[\varepsilon_{T,i}^2\right] =: \sum_{i=1}^{n} \sigma_{T,i}^2$$

$$Q_{n,T} - \mathbb{E}[Q_{n,T}] = \sum_{i=1}^{n}\left(\varepsilon_{T,i}^2 - \sigma_{T,i}^2 + b_{T,i}\varepsilon_{T,i}\right) =: \sum_{i=1}^{n} Y_{T,i}$$

$$\mathbb{E}[Y_{T,i}^2] = \mathbb{E}[\varepsilon_{T,i}^4] - \sigma_{T,i}^4 + b_{T,i}^2\sigma_{T,i}^2 + 2b_{T,i}\mathbb{E}\varepsilon_{T,i}^3$$

$$\operatorname{var}(Q_{n,T}) = \mathbb{E}\left[\sum_{i=1}^{n} Y_{T,i}\right]^2 = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{T,i}Y_{T,j}\right] = \sum_{i=1}^{n} \mathbb{E}[Y_{T,i}^2],$$

where the last equality is due to independence of $\varepsilon_{T,i}^2$ across i. We now show that

$$\frac{Q_{n,T} - \mathbb{E}[Q_{n,T}]}{\sqrt{\operatorname{var}(Q_{n,T})}} = \sum_{i=1}^{n} \frac{Y_{T,i}}{\sqrt{\operatorname{var}(Q_{n,T})}} \xrightarrow{d} N(0,1)$$

as $n, T \to \infty$. This boils down to verifying the Lyapounov's condition in Theorem B.1 part (b); that is, for some $\delta > 0$,

$$\lim_{n,T\to\infty} \sum_{i=1}^{n} \frac{1}{[\operatorname{var}(Q_{n,T})]^{1+\delta/2}} \mathbb{E}\left|Y_{T,i}\right|^{2+\delta} = 0.$$

Let's first find an upper bound for $\mathbb{E}\left|Y_{T,i}\right|^{2+\delta}$.

$$\begin{aligned}&\mathbb{E}\left|Y_{T,i}\right|^{2+\delta} = \mathbb{E}\left|\varepsilon_{T,i}^2 - \sigma_{T,i}^2 + b_{T,i}\varepsilon_{T,i}\right|^{2+\delta} \leq 3^{1+\delta}\left(\mathbb{E}|\varepsilon_{T,i}^2|^{2+\delta} + \mathbb{E}|\sigma_{T,i}^2|^{2+\delta} + |b_{T,i}|^{2+\delta}\mathbb{E}|\varepsilon_{T,i}|^{2+\delta}\right)\\ &= 3^{1+\delta}\left(\mathbb{E}|\varepsilon_{T,i}|^{4+2\delta} + \sigma_{T,i}^{4+2\delta} + |b_{T,i}|^{2+\delta}\mathbb{E}|\varepsilon_{T,i}|^{2+\delta}\right) \leq K_1 + K_2|b_{T,i}|^{2+\delta},\end{aligned}$$

for absolute positive constants K_1 and K_2 for sufficiently large T, where the first inequality is due to Loeve's c_r inequality, and the last inequality is due to the assumption (ii) of the theorem. Then we have

$$\sum_{i=1}^{n} \frac{\mathbb{E}\left|Y_{T,i}\right|^{2+\delta}}{[\operatorname{var}(Q_{n,T})]^{1+\delta/2}} \leq \frac{nK_1 + nK_2\left(\frac{1}{n}\sum_{i=1}^{n}|b_{T,i}|^{2+\delta}\right)}{[n^{-1}\operatorname{var}(Q_{n,T})]^{1+\delta/2}n^{1+\delta/2}} = \frac{K_1 + K_2\left(\frac{1}{n}\sum_{i=1}^{n}|b_{T,i}|^{2+\delta}\right)}{[n^{-1}\operatorname{var}(Q_{n,T})]^{1+\delta/2}n^{\delta/2}} \to 0$$

as $n, T \to \infty$, where the convergence to 0 relies on the assumption (ii) and (iii) of the theorem. □

A.3.2 Proof of Theorem 5.1

Proof of Theorem 5.1. Write

$$\begin{aligned}\frac{LM_{n,T} - n}{\sqrt{2n}} &= \frac{T(\bar{y} - \mu_0)^\intercal \tilde{\Sigma}_{\mu_0}^{-1}(\bar{y} - \mu_0) - n}{\sqrt{2n}}\\ &= \frac{T(\bar{y} - \mu_0)^\intercal \Sigma^{-1}(\bar{y} - \mu_0) - n}{\sqrt{2n}} + \frac{T(\bar{y} - \mu_0)^\intercal(\tilde{\Sigma}_{\mu_0}^{-1} - \Sigma^{-1})(\bar{y} - \mu_0)}{\sqrt{2n}}.\end{aligned}$$

We first show that under H_0 as $n, T \to \infty$,

$$\frac{T(\bar{y}-\mu_0)^\intercal \Sigma^{-1}(\bar{y}-\mu_0) - n}{\sqrt{2n}} \xrightarrow{d} N(0,1).$$

Write

$$\frac{T(\bar{y}-\mu_0)^\intercal \Sigma^{-1}(\bar{y}-\mu_0) - n}{\sqrt{2n}} = \frac{\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T (y_t-\mu_0)\right]^\intercal (L^{-1})^\intercal L^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T (y_t-\mu_0)\right] - n}{\sqrt{2n}}$$
$$=: \frac{\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t\right)^\intercal \left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t\right) - n}{\sqrt{2n}} =: \frac{z_T^\intercal z_T - n}{\sqrt{2n}} = \frac{\sum_{i=1}^n z_{T,i}^2 - n}{\sqrt{2n}} =: \frac{Q_{n,T}-n}{\sqrt{2n}}.$$

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

$$\begin{aligned}
\mathbb{E}[z_{T,i}] &= \mathbb{E}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T x_{t,i}\right] = 0\\
\operatorname{var}(z_T) &= \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t\right) = I_n\\
\mathbb{E}[Q_{n,T}] &= \mathbb{E}\left[\sum_{i=1}^n z_{T,i}^2\right] = \sum_{i=1}^n \mathbb{E}\left[z_{T,i}^2\right] = n\\
\operatorname{var}(Q_{n,T}) &= \operatorname{var}\left(\sum_{i=1}^n z_{T,i}^2\right) = \sum_{i=1}^n \operatorname{var}\left(z_{T,i}^2\right) = \sum_{i=1}^n \left[\mathbb{E}[z_{T,i}^4] - \left(\mathbb{E}[z_{T,i}^2]\right)^2\right]\\
&= \sum_{i=1}^n \left(\mathbb{E}[z_{T,i}^4] - 1\right) =: \sum_{i=1}^n (\gamma_{z,i}+2)
\end{aligned}$$

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

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

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

$$\mathbb{E}[z_{T,i}^4] = \mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_{t,i}\right)^4\right] = \frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\sum_{k=1}^T\sum_{\ell=1}^T \mathbb{E}\left[x_{t,i}x_{s,i}x_{k,i}x_{\ell,i}\right] \tag{A.14}$$

Note that the summand in (A.14) is non-zero only if $t = s = k = \ell$, $t = s \neq k = \ell$, $t = k \neq s = \ell$, $t = \ell \neq k = s$. First, consider the case $t = s = k = \ell$. Collecting all the summands in (A.14) satisfying this, we have

$$\frac{1}{T^2}\sum_{t=1}^T \mathbb{E}\left[x_{t,i}^4\right] = \frac{1}{T^2}\sum_{t=1}^T(\gamma_{x,t,i}+3) = \frac{1}{T^2}\sum_{t=1}^T \gamma_{x,t,i} + \frac{3}{T} \tag{A.15}$$

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

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

Second, consider the case $t = s \neq k = \ell$. Collecting all the summands in (A.14) satisfying this, we have

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{k=1\\ \neq t}}^{T}\mathbb{E}\left[x_{t,i}^2 x_{k,i}^2\right] = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{\substack{k=1\\ \neq t}}^{T}\mathbb{E}\left[x_{t,i}^2\right]\mathbb{E}\left[x_{k,i}^2\right] = \frac{T(T-1)}{T^2} = 1 - \frac{1}{T} \tag{A.16}$$

Likewise for cases $t = k \neq s = \ell$ and $t = \ell \neq k = s$, both sums are $1 - 1/T$. Substituting (A.15) and (A.16) into (A.14), we have

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

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

$$\operatorname{var}(Q_{n,T}) = \sum_{i=1}^{n}\left(\gamma_{z,i} + 2\right) = \sum_{i=1}^{n}\left(\frac{1}{T^2}\sum_{t=1}^{T}\gamma_{x,t,i} + 2\right) = 2n\left(1 + \frac{1}{2T}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{x,t,i}\right). \tag{A.17}$$

It remains to verify condition (ii)-(iii) of Theorem A.2. We have

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

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.2 is satisfied. Condition (ii) of Theorem A.2 is also satisfied: for some $\delta > 0$

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

by Theorem B.3 in Section B under assumption (b) of the theorem. Thus we have

$$\frac{T(\bar{y}-\mu_0)^\intercal\Sigma^{-1}(\bar{y}-\mu_0) - n}{\sqrt{2n}} = \frac{Q_{n,T} - n}{\sqrt{2n}} = \frac{Q_{n,T} - n}{\sqrt{2n\left(1 + \frac{1}{2T}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{x,t,i}\right)}}\big(1 + o(1)\big) \xrightarrow{d} N(0,1),$$

under H_0 as $n, T \to \infty$, where the second equality is due to

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

under the assumption (b) of the theorem, and the weak convergence is due to Theorem A.2.

The theorem would follow if we show that

$$\frac{T(\bar{y}-\mu_0)^\intercal(\tilde{\Sigma}_{\mu_0}^{-1} - \Sigma^{-1})(\bar{y}-\mu_0)}{\sqrt{2n}} = o_p(1).$$

We now show this.

$$\frac{T|(\bar{y}-\mu_0)^\intercal(\tilde{\Sigma}_{\mu_0}^{-1}-\Sigma^{-1})(y_t-\mu_0)|}{\sqrt{2n}} = \frac{\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-\mu_0)\right)^\intercal\left(\tilde{\Sigma}_{\mu_0}^{-1}-\Sigma^{-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-\mu_0)\right)\right|}{\sqrt{2n}}$$

$$= \frac{1}{\sqrt{2n}}\left|\sum_{i=1}^n\sum_{j=1}^n\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,i}-\mu_{0,i})\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,j}-\mu_{0,j})\right)(\tilde{\Sigma}_{\mu_0,i,j}^{-1}-\Sigma_{i,j}^{-1})\right|$$

$$\leq \frac{1}{\sqrt{2n}}\left(\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,i}-\mu_{0,i})\right|\right)^2\sum_{i=1}^n\sum_{j=1}^n|\tilde{\Sigma}_{\mu_0,i,j}^{-1}-\Sigma_{i,j}^{-1}|$$

$$= \frac{1}{\sqrt{2n}}\left(\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,i}-\mu_{0,i})\right|\right)^2\|\tilde{\Sigma}_{\mu_0}^{-1}-\Sigma^{-1}\|_1$$

$$= O_p\left(\frac{\log n}{\sqrt{n}}\right)\|\Sigma^{-1}\|_1 O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) = \frac{1}{n^{\beta_2}}\|\Sigma^{-1}\|_1 O_p\left(\sqrt{\frac{n^{2\beta_2-1}\log^5 n}{n^{2-\beta_1}T}}\right)$$

$$= O_p\left(\sqrt{\frac{n^{2\beta_2+\beta_1-3}\log^5 n}{T}}\right) = o_p(1)$$

where the fourth equality is due to Lemma B.1, the sixth equality is due to Assumption 5.1, and the last equality is due to the assumption (a) of the theorem.

For the Wald statistic, write

$$\frac{W_{n,T}-n}{\sqrt{2n}} = \frac{T(\bar{y}-\mu_0)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu_0)-n}{\sqrt{2n}}$$
$$= \frac{T(\bar{y}-\mu_0)^\intercal\Sigma^{-1}(\bar{y}-\mu_0)-n}{\sqrt{2n}} + \frac{T(\bar{y}-\mu_0)^\intercal(\tilde{\Sigma}^{-1}-\Sigma^{-1})(\bar{y}-\mu_0)}{\sqrt{2n}}.$$

We have already shown in the proof of the LM test that under the assumptions (a)-(b) of the theorem and under H_0 as $n, T\to\infty$,

$$\frac{T(\bar{y}-\mu_0)^\intercal\Sigma^{-1}(\bar{y}-\mu_0)-n}{\sqrt{2n}} \xrightarrow{d} N(0,1).$$

Display (5.4) would follow if we show that

$$\frac{T(\bar{y}-\mu_0)^\intercal(\tilde{\Sigma}^{-1}-\Sigma^{-1})(\bar{y}-\mu_0)}{\sqrt{2n}} = o_p(1).$$

We now show this.

$$\frac{T|(\bar{y}-\mu_0)^\intercal(\tilde{\Sigma}^{-1}-\Sigma^{-1})(y_t-\mu_0)|}{\sqrt{2n}} = \frac{\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-\mu_0)\right)^\intercal(\tilde{\Sigma}^{-1}-\Sigma^{-1})\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-\mu_0)\right)\right|}{\sqrt{2n}}$$

$$= \frac{1}{\sqrt{2n}}\left|\sum_{i=1}^n\sum_{j=1}^n\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,i}-\mu_{0,i})\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,j}-\mu_{0,j})\right)(\tilde{\Sigma}_{i,j}^{-1}-\Sigma_{i,j}^{-1})\right|$$

$$\leq \frac{1}{\sqrt{2n}}\left(\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,i}-\mu_{0,i})\right|\right)^2\sum_{i=1}^n\sum_{j=1}^n|\tilde{\Sigma}_{i,j}^{-1}-\Sigma_{i,j}^{-1}|$$

$$= \frac{1}{\sqrt{2n}}\left(\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_{t,i}-\mu_{0,i})\right|\right)^2\|\tilde{\Sigma}^{-1}-\Sigma^{-1}\|_1$$

$$= O_p\left(\frac{\log n}{\sqrt{n}}\right)\|\Sigma^{-1}\|_1\left[O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{\log^2 n}{T}\right)\right]$$

$$= \frac{1}{n^{\beta_2}}\|\Sigma^{-1}\|_1\left[O_p\left(\sqrt{\frac{n^{2\beta_2-1}\log^5 n}{n^{2-\beta_1}T}}\right)+O_p\left(\frac{n^{\beta_2-\frac{1}{2}}\log^3 n}{T}\right)\right]$$

$$= O_p\left(\sqrt{\frac{n^{2\beta_2+\beta_1-3}\log^5 n}{T}}\right)+O_p\left(\frac{n^{\beta_2-\frac{1}{2}}\log^3 n}{T}\right) = o_p(1)$$

where the fourth equality is due to Lemma B.1 and Theorem 4.1(iv), and the sixth equality is due to Assumption 5.1. □

A.4 Proof of Theorem 5.2

Proof. Write

$$\begin{aligned} W_{n,T} &= T(\bar{y}-\mu_0)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu_0) = T\left[\bar{y}-\mu_T+\mu_T-\mu_0\right]^\intercal\tilde{\Sigma}^{-1}\left[\bar{y}-\mu_T+\mu_T-\mu_0\right] \\ &= T(\bar{y}-\mu_T)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu_T)+2T(\mu_T-\mu_0)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu_T)+T(\mu_T-\mu_0)^\intercal\tilde{\Sigma}^{-1}(\mu_T-\mu_0) \\ &=: W_{n,T,1}+\theta^\intercal\tilde{\Sigma}^{-1}\theta \end{aligned}$$

whence we have

$$\frac{W_{n,T}-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}-\frac{\theta^\intercal\Sigma^{-1}\theta}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}} = \frac{W_{n,T,1}-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}+\frac{\theta^\intercal(\tilde{\Sigma}^{-1}-\Sigma_{-1})\theta}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}} \tag{A.18}$$

We first consider the first term on the right side of (A.18).

$$\begin{aligned} \frac{W_{n,T,1}-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}} &= \frac{T(\bar{y}-\mu_T)^\intercal\Sigma^{-1}(\bar{y}-\mu_T)+2T(\mu_T-\mu_0)^\intercal\Sigma^{-1}(\bar{y}-\mu_T)-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}} \\ &+\frac{T(\bar{y}-\mu_T)^\intercal(\tilde{\Sigma}^{-1}-\Sigma^{-1})(\bar{y}-\mu_T)+2T(\mu_T-\mu_0)^\intercal(\tilde{\Sigma}^{-1}-\Sigma^{-1})(\bar{y}-\mu_T)}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}} \end{aligned} \tag{A.19}$$

We show that the first term on the right side of (A.19) converges in distribution under the local alternatives.

$$\frac{T(\bar{y}-\mu_T)^\intercal\Sigma^{-1}(\bar{y}-\mu_T)+2T(\mu_T-\mu_0)^\intercal\Sigma^{-1}(\bar{y}-\mu_T)-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}$$
$$=\frac{\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-\mu_T)\right]^\intercal(L^{-1})^\intercal L^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-\mu_T)\right]+2\theta^\intercal(L^{-1})^\intercal L^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-\mu_T)\right]-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}$$
$$=:\frac{\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t\right)^\intercal\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t\right)+2(L^{-1}\theta)^\intercal\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t\right)-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}=:\frac{z_T^\intercal z_T+b_T^\intercal z_T-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}$$
$$=\frac{\sum_{i=1}^n z_{T,i}^2+\sum_{i=1}^n b_{T,i}z_{T,i}-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}=:\frac{Q_{n,T}-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^\intercal\Sigma^{-1}\theta\right)}}.$$

Note that for each $n\geq 1, T\geq 1$, $z_{T,1},\ldots,z_{T,n}$ are (mutually) independent under assumption (b) of the theorem and Assumption 4.1(i). Under H_1,

$$\mathbb{E}[z_{T,i}]=\mathbb{E}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T x_{t,i}\right]=0$$
$$\text{var}(z_T)=\text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t\right)=I_n$$
$$\mathbb{E}[Q_{n,T}]=\mathbb{E}\left[\sum_{i=1}^n z_{T,i}^2+\sum_{i=1}^n b_{T,i}z_{T,i}\right]=\sum_{i=1}^n\mathbb{E}\left[z_{T,i}^2\right]=n.$$

We next calculate $\text{var}(Q_{n,T})$.

$$\text{var}(Q_{n,T})=\text{var}\left(\sum_{i=1}^n z_{T,i}^2+\sum_{i=1}^n b_{T,i}z_{T,i}\right)=\sum_{i=1}^n\text{var}\left(z_{T,i}^2+b_{T,i}z_{T,i}\right)$$
$$=\sum_{i=1}^n\mathbb{E}\left[z_{T,i}^2+b_{T,i}z_{T,i}-\mathbb{E}z_{T,i}^2\right]^2=\sum_{i=1}^n\mathbb{E}\left[z_{T,i}^2+b_{T,i}z_{T,i}-1\right]^2$$
$$=\sum_{i=1}^n\left[\left(\mathbb{E}[z_{T,i}^4]-1\right)+2b_{T,i}\mathbb{E}[z_{T,i}^3]+b_{T,i}^2\right]=\sum_{i=1}^n\left(\gamma_{z,i}+2\right)+2\sum_{i=1}^n b_{T,i}\mathbb{E}[z_{T,i}^3]+\sum_{i=1}^n b_{T,i}^2$$

In (A.17), we have already calculated that

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

We now calculate $\mathbb{E}[z_{T,i}^3]$.

$$\mathbb{E}[z_{T,i}^3]=\mathbb{E}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T x_{t,i}\right]^3=\mathbb{E}\left[\frac{1}{T^{3/2}}\sum_{t=1}^T\sum_{s=1}^T\sum_{k=1}^T x_{t,i}x_{s,i}x_{k,i}\right]=\frac{1}{T^{3/2}}\sum_{t=1}^T\sum_{s=1}^T\sum_{k=1}^T\mathbb{E}\left[x_{t,i}x_{s,i}x_{k,i}\right]$$
$$=\frac{1}{T^{3/2}}\sum_{t=1}^T\mathbb{E}\left[x_{t,i}^3\right]$$

Backing up, we have

$$\operatorname{var}(Q_{n,T}) = 2n\left(1 + \frac{1}{2T}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{x,t,i}\right) + \frac{2}{T^{3/2}}\sum_{i=1}^{n} b_{T,i}\sum_{t=1}^{T}\mathbb{E}\left[x_{t,i}^{3}\right] + \sum_{i=1}^{n} b_{T,i}^{2}.$$

We now verify the conditions (ii) and (iii) of Theorem A.2. For the condition (ii), we have already verified in the proof of Theorem 5.1 that $\limsup_{T\to\infty}\sup_{n\geq 1}\sup_{1\leq i\leq n}\mathbb{E}\left|z_{T,i}\right|^{4+2\delta} < \infty$. Next,

$$\limsup_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}|b_{T,i}|^{2+\delta} = \limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left|2(L^{-1}\theta)_i\right|^{2+\delta} = 2^{2+\delta}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left|(L^{-1}\theta)_i\right|^{2+\delta} < \infty$$

via the assumption (b) of the theorem. Thus the condition (ii) of Theorem A.2 is met. Finally,

$$\begin{aligned}\frac{1}{n}\operatorname{var}(Q_{n,T}) &= 2\left(1 + \frac{1}{2T}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{x,t,i}\right) + \frac{2}{nT^{3/2}}\sum_{i=1}^{n} b_{T,i}\sum_{t=1}^{T}\mathbb{E}\left[x_{t,i}^{3}\right] + \frac{1}{n}\sum_{i=1}^{n} b_{T,i}^{2}\\ &= 2\left(1 + \frac{1}{2T}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{x,t,i}\right) + \frac{4}{\sqrt{T}}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(L^{-1}\theta)_i\mathbb{E}\left[x_{t,i}^{3}\right] + \frac{4}{n}\theta^{\intercal}\Sigma^{-1}\theta > 0\end{aligned}$$

for large enough n and T because $\gamma_{x,t,i} > -3$ for all t and i by definition of the excess kurtosis. Thus the condition (iii) of Theorem A.2 is met.

Thus we have

$$\begin{aligned}&\frac{T(\bar{y}-\mu_T)^{\intercal}\Sigma^{-1}(\bar{y}-\mu_T) + 2T(\mu_T-\mu_0)^{\intercal}\Sigma^{-1}(\bar{y}-\mu_T) - n}{\sqrt{2n\left(1+\frac{2}{n}\theta^{\intercal}\Sigma^{-1}\theta\right)}} = \frac{Q_{n,T}-n}{\sqrt{2n\left(1+\frac{2}{n}\theta^{\intercal}\Sigma^{-1}\theta\right)}}\\ &= \frac{Q_{n,T}-n}{\sqrt{\operatorname{var}(Q_{n,T})}}\frac{\sqrt{\operatorname{var}(Q_{n,T})}}{\sqrt{2n\left(1+\frac{2}{n}\theta^{\intercal}\Sigma^{-1}\theta\right)}} = \frac{Q_{n,T}-n}{\sqrt{\operatorname{var}(Q_{n,T})}}(1+o(1)) \xrightarrow{d} N(0,1)\end{aligned}$$

as $n, T\to\infty$.

We next show that the second term of on the right side of (A.19) is $o_p(1)$ under H_1

$$\begin{aligned}&\frac{T\left|(\bar{y}-\mu_T)^{\intercal}(\tilde{\Sigma}^{-1}-\Sigma^{-1})(\bar{y}-\mu_T)\right|}{\sqrt{2n\left(1+\frac{2}{n}\theta^{\intercal}\Sigma^{-1}\theta\right)}} + \frac{2T\left|(\mu_T-\mu_0)^{\intercal}(\tilde{\Sigma}^{-1}-\Sigma^{-1})(\bar{y}-\mu_T)\right|}{\sqrt{2n\left(1+\frac{2}{n}\theta^{\intercal}\Sigma^{-1}\theta\right)}}\\ &= \frac{\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t-\mu_T)\right)^{\intercal}\left(\tilde{\Sigma}^{-1}-\Sigma^{-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t-\mu_T)\right)\right|}{\sqrt{2n\left(1+\frac{2}{n}\theta^{\intercal}\Sigma^{-1}\theta\right)}} + \frac{2\left|\theta^{\intercal}(\tilde{\Sigma}^{-1}-\Sigma^{-1})\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t-\mu_T)\right)\right|}{\sqrt{2n\left(1+\frac{2}{n}\theta^{\intercal}\Sigma^{-1}\theta\right)}}\\ &\leq \frac{\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t-\mu_T)\right)^{\intercal}\left(\tilde{\Sigma}^{-1}-\Sigma^{-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t-\mu_T)\right)\right|}{\sqrt{2n}} + \frac{2\left|\theta^{\intercal}(\tilde{\Sigma}^{-1}-\Sigma^{-1})\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_t-\mu_T)\right)\right|}{\sqrt{2n}}\\ &\leq \frac{1}{\sqrt{2n}}\left(\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_{t,i}-\mu_{T,i})\right|\right)^{2}\|\tilde{\Sigma}^{-1}-\Sigma^{-1}\|_1\\ &\quad + \sqrt{\frac{2}{n}}\max_{1\leq i\leq n}|\theta_i|\left(\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_{t,i}-\mu_{T,i})\right|\right)\|\tilde{\Sigma}^{-1}-\Sigma^{-1}\|_1\\ &= O_p\left(\frac{\log n}{\sqrt{n}}\right)\|\Sigma^{-1}\|_1\left[O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right)\right]\\ &= \frac{1}{n^{\beta_2}}\|\Sigma^{-1}\|_1\left[O_p\left(\sqrt{\frac{n^{2\beta_2-1}\log^5 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{n^{\beta_2-\frac{1}{2}}\log^3 n}{T}\right)\right]\\ &= O_p\left(\sqrt{\frac{n^{2\beta_2+\beta_1-3}\log^5 n}{T}}\right) + O_p\left(\frac{n^{\beta_2-\frac{1}{2}}\log^3 n}{T}\right) = o_p(1)\end{aligned} \tag{A.20}$$

where the second equality is due to Lemma B.1 and Theorem 4.1(iv), and the fourth equality is due to Assumption 5.1.

Backing up, in (A.19), we hence have under H_1 as $n, T \to \infty$

$$\frac{W_{n,T,1} - n}{\sqrt{2n\left(1 + \frac{2}{n}\theta^\intercal \Sigma^{-1}\theta\right)}} \xrightarrow{d} N(0,1)$$

We finally consider the second term on the right side of (A.18).

$$\frac{\theta^\intercal(\tilde{\Sigma}^{-1} - \Sigma^{-1})\theta}{\sqrt{2n\left(1 + \frac{2}{n}\theta^\intercal \Sigma^{-1}\theta\right)}} = o_p(1)$$

because of the similar trick used in (A.20). Backing up, in (A.18), we hence have under H_1 as $n, T \to \infty$

$$\frac{W_{n,T} - n}{\sqrt{2n\left(1 + \frac{2}{n}\theta^\intercal \Sigma^{-1}\theta\right)}} - \frac{\theta^\intercal \Sigma^{-1}\theta}{\sqrt{2n\left(1 + \frac{2}{n}\theta^\intercal \Sigma^{-1}\theta\right)}} \xrightarrow{d} N(0,1)$$

□

A.5 Proof of Theorem 5.3

Proof of Theorem 5.3.

$$\begin{aligned} W^*_{n,T} &:= T(R\bar{y} - r)^\intercal (R\tilde{\Sigma}R^\intercal)^{-1}(R\bar{y} - r) \\ &= T(R\bar{y} - r)^\intercal (R\Sigma R^\intercal)^{-1}(R\bar{y} - r) - T(R\bar{y} - r)^\intercal \left[(R\tilde{\Sigma}R^\intercal)^{-1} - (R\Sigma R^\intercal)^{-1}\right](R\bar{y} - r) \end{aligned} \tag{A.21}$$

We now show that the first term of (A.21) is asymptotically chi square distributed under H_0. Since R has full row rank q and $\lambda_{\min}(\Sigma)$ is bounded away from zero by an absolute positive constant, $R\Sigma R^\intercal$ has full rank q. Consider the Cholesky decomposition of $R\Sigma R^\intercal = L_R L_R^\intercal$, where L_R is a $q \times q$ nonsingular lower triangular matrix with positive diagonal elements. Write

$$\begin{aligned} &T(R\bar{y} - r)^\intercal (R\Sigma R^\intercal)^{-1}(R\bar{y} - r) = T(R\bar{y} - r)^\intercal (L_R^{-1})^\intercal L_R^{-1}(R\bar{y} - r) \\ &= \left[\frac{1}{\sqrt{T}}\sum_{t=1}^T L_R^{-1}R(y_t - \mu)\right]^\intercal \left[\frac{1}{\sqrt{T}}\sum_{t=1}^T L_R^{-1}R(y_t - \mu)\right]. \end{aligned}$$

Note that $L_R^{-1}R(y_1 - \mu), L_R^{-1}R(y_2 - \mu), \ldots, L_R^{-1}R(y_T - \mu)$ are independent random vectors in $\mathbb{R}^q$ with mean zero and variance matrix I_q. Then we can invoke a version of the multivariate central limit theorem to show $T^{-1/2}\sum_{t=1}^T L_R^{-1}R(y_t - \mu) \xrightarrow{d} N(0, I_q)$ as $n, T \to \infty$, whence we have $T(R\bar{y} - r)^\intercal (R\Sigma R^\intercal)^{-1}(R\bar{y} - r) \xrightarrow{d} \chi^2_q$ as $n, T \to \infty$.

We now show that the second term of (A.21) is $o_p(1)$ under H_0.

$$\begin{aligned} &\left|T(R\bar{y} - r)^\intercal \left[(R\tilde{\Sigma}R^\intercal)^{-1} - (R\Sigma R^\intercal)^{-1}\right](R\bar{y} - r)\right| \\ &= \left|\sum_{i=1}^q \sum_{j=1}^q \left(\frac{1}{\sqrt{T}}\sum_{t=1}^T [R(y_t - \mu)]_i\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T [R(y_t - \mu)]_j\right)\left[(R\tilde{\Sigma}R^\intercal)^{-1}_{i,j} - (R\Sigma R^\intercal)^{-1}_{i,j}\right]\right| \\ &\leq \left(\max_{1\leq i\leq q}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T [R(y_t - \mu)]_i\right|\right)^2 \left\|(R\tilde{\Sigma}R^\intercal)^{-1} - (R\Sigma R^\intercal)^{-1}\right\|_1 \\ &= O_p(1)\left\|(R\tilde{\Sigma}R^\intercal)^{-1} - (R\Sigma R^\intercal)^{-1}\right\|_1. \end{aligned}$$

We need to find a rate for $\left\|(R\tilde{\Sigma}R^\intercal)^{-1} - (R\Sigma R^\intercal)^{-1}\right\|_1$. First, note that

$$\begin{aligned}
&\left\|R\tilde{\Sigma}R^\intercal - R\Sigma R^\intercal\right\|_1 \leq q^{3/2}\left\|R(\tilde{\Sigma}-\Sigma)R^\intercal\right\|_{\ell_2} \leq q^{3/2}\|R\|_{\ell_2}\|\tilde{\Sigma}-\Sigma\|_{\ell_2}\|R^\intercal\|_{\ell_2} \\
&= q^{3/2}\|R^\intercal\|_{\ell_2}^2\|\tilde{\Sigma}-\Sigma\|_{\ell_2} = q^{3/2}\lambda_{\max}(RR^\intercal)\|\tilde{\Sigma}-\Sigma\|_{\ell_2} \\
&= q^{3/2}\lambda_{\max}(RR^\intercal)\|\Sigma\|_{\ell_2}\left[O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right)\right] = o_p(1),
\end{aligned}$$

where the second last equality is due to Theorem 4.1(v), and the last equality is due to (5.5). Second,

$$\left\|(R\Sigma R^\intercal)^{-1}\right\|_1 \leq q^{3/2}\left\|(R\Sigma R^\intercal)^{-1}\right\|_{\ell_2} = q^{3/2}\lambda_{\max}\left[(R\Sigma R^\intercal)^{-1}\right] = \frac{q^{3/2}}{\lambda_{\min}\left[R\Sigma R^\intercal\right]} \leq \frac{q^{3/2}}{\lambda_{\min}(RR^\intercal)\lambda_{\min}(\Sigma)}$$
$$= O(1)$$

where the last inequality is due to Lemma B.2 in Section B and the last equality is due to the assumption of the theorem. Then via Lemma B.4 in Section B we have

$$\left\|(R\tilde{\Sigma}R^\intercal)^{-1} - (R\Sigma R^\intercal)^{-1}\right\|_1 = q^{3/2}\lambda_{\max}(RR^\intercal)\|\Sigma\|_{\ell_2}\left[O_p\left(\sqrt{\frac{\log^3 n}{n^{2-\beta_1}T}}\right) + O_p\left(\frac{\log^2 n}{T}\right)\right] = o_p(1).$$

Backing up, we have proved that the second term of (A.21) is $o_p(1)$ under H_0. □

A.6 Proof of Lemma 5.1

Proof of Lemma 5.1. The assumptions of the lemma allows us to invoke Theorem 5.1. Thus under $H_0 : \mu = \mu_0$, as $n, T \to \infty$,

$$\frac{W_{n,T} - n}{\sqrt{2n}} = \frac{T(\bar{y}-\mu_0)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu_0) - n}{\sqrt{2n}} \xrightarrow{d} N(0,1).$$

This implies that for any unknown μ

$$\mathbb{P}_\mu\left(\frac{T(\bar{y}-\mu)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu) - n}{\sqrt{2n}} < z_\alpha\right) \to 1-\alpha$$

as $n, T \to \infty$, where z_α is the upper α percentile of $N(0,1)$.

Invoking Lemma B.5 in Section B with $x = \bar{y} - \mu$ and $S = \tilde{\Sigma}$ yields: For any $\phi \in \mathbb{R}^n$

$$\left[\phi^\intercal(\bar{y}-\mu)\right]^2 \leq \phi^\intercal\tilde{\Sigma}\phi \cdot (\bar{y}-\mu)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu)$$

whence we have

$$\frac{\left[\phi^\intercal(\bar{y}-\mu)\right]^2}{\phi^\intercal\tilde{\Sigma}\phi} \leq (\bar{y}-\mu)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu).$$

Multiply both sides by T, minus n, and divide by $\sqrt{2n}$:

$$\frac{T\left[\phi^\intercal(\bar{y}-\mu)\right]^2/\phi^\intercal\tilde{\Sigma}\phi - n}{\sqrt{2n}} \leq \frac{T(\bar{y}-\mu)^\intercal\tilde{\Sigma}^{-1}(\bar{y}-\mu) - n}{\sqrt{2n}}.$$

Thus we asset with confidence $1-\alpha$ that the unknown μ satisfies *simultaneously for all* ϕ the inequalities:

$$\frac{T\left[\phi^\intercal(\bar{y}-\mu)\right]^2/\phi^\intercal\tilde{\Sigma}\phi - n}{\sqrt{2n}} < z_\alpha,$$

as $n, T \to \infty$. □

B Auxiliary Lemmas

Lemma B.1. *Suppose Assumption 4.1(i)-(ii) hold. Then we have*

$$\max_{1\le i\le n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(y_{t,i}-\mathbb{E}y_{t,i})\right| = O_p(\sqrt{\log n}).$$

Proof. Under Assumption 4.1(ii), we have, for $i = 1,\ldots,n$, $m = 2,3,\ldots,$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left|y_{t,i}-\mathbb{E}y_{t,i}\right|^m \le \frac{1}{T}\sum_{t=1}^{T}2^{m-1}\left(\mathbb{E}|y_{t,i}|^m+\mathbb{E}|\mathbb{E}y_{t,i}|^m\right) \le \frac{1}{T}\sum_{t=1}^{T}2^{m-1}\left(\mathbb{E}|y_{t,i}|^m+\mathbb{E}|y_{t,i}|^m\right)$$
$$= 2^m\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}|y_{t,i}|^m \le 2^m A^m \le 2m!A^m = \frac{m!}{2}A^{m-2}A^2 4$$

for some absolute positive constant A. Now invoke the Bernstein's inequality in Section B with $\sigma_0^2 = 4A^2$: For all $\epsilon > 0$

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

Invoking Corollary B.1 in Section B, we have

$$\max_{1\le i\le n}\left|\frac{1}{T}\sum_{t=1}^{T}(y_{t,i}-\mathbb{E}y_{t,i})\right| = O_p\left(\frac{\log n}{T}\vee\sqrt{\frac{\log n}{T}}\right) = O_p\left(\sqrt{\frac{\log n}{T}}\right).$$

The lemma follows. □

We next give two central limit theorems for double-index (n, T) processes.

Theorem B.1.

(a) *Suppose $Y_{n,t}$ is a random variable independent across $1 \le t \le T$ for $n \ge 1$ and $T \ge 1$. Assume that*

$$\mathbb{E}[Y_{n,t}] = 0 \qquad \mathbb{E}[Y_{n,T,t}^2] = \sigma_{n,t}^2.$$

Define

$$s_{n,T}^2 := \sum_{t=1}^{T}\sigma_{n,t}^2 \qquad \xi_{n,T,t} := \frac{Y_{n,t}}{s_{n,T}}.$$

Assume that $s_{n,T}^2 > 0$ for large enough n and T. Suppose the following Lyapounov's condition hold: For some $\delta > 0$,

$$\lim_{n,T\to\infty}\sum_{t=1}^{T}\frac{1}{s_{n,T}^{2+\delta}}\mathbb{E}\left|Y_{n,t}\right|^{2+\delta} = 0.$$

Then as $n, T \to \infty$

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

(b) *Suppose $Y_{T,i}$ is a random variable independent across $1 \le i \le n$ for $n \ge 1$ and $T \ge 1$. Assume that*

$$\mathbb{E}[Y_{T,i}] = 0 \qquad \mathbb{E}[Y_{T,i}^2] = \sigma_{T,i}^2.$$

Define

$$s_{n,T}^2 := \sum_{i=1}^{n} \sigma_{T,i}^2 \qquad \xi_{n,T,i} := \frac{Y_{T,i}}{s_{n,T}}.$$

Assume that $s_{n,T}^2 > 0$ for large enough n and T. Suppose the following Lyapounov's condition hold: For some $\delta > 0$,

$$\lim_{n,T\to\infty} \sum_{i=1}^{n} \frac{1}{s_{n,T}^{2+\delta}} \mathbb{E}\left|Y_{T,i}\right|^{2+\delta} = 0.$$

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)) □

Theorem B.2 (Bernstein's inequality)**.** *We let $Z_1, \dots, Z_T$ be independent random variables, satisfying for absolute positive constants A and σ_0^2*

$$\mathbb{E}Z_t = 0 \quad \forall t, \qquad \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}|Z_t|^m \le \frac{m!}{2} A^{m-2}\sigma_0^2, \quad m = 2, 3, \dots.$$

Let $\epsilon > 0$ be arbitrary. Then

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} Z_t\right| \ge \sigma_0^2\left[A\epsilon + \sqrt{2\epsilon}\right]\right) \le 2e^{-T\sigma_0^2\epsilon}.$$

Proof. Slightly adapted from Bühlmann and van de Geer (2011) p487. □

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 $1 \le i \le n$,*

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

for some absolute positive constants K and σ_0^2. Then

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

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 $\varepsilon > 0$. These exist $N_\varepsilon := 2/\varepsilon$, T_ε and $M_\varepsilon := \max(4K, 4\sigma_0)$ such that for all $n > N_\varepsilon$ and $T > T_\varepsilon$ we have

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

□

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

$$\lambda_{\min}\left(A\right)\lambda_{\min}\left(BB^{\intercal}\right) \leq \lambda_{\min}\left(BAB^{\intercal}\right) \leq \lambda_{\max}\left(BAB^{\intercal}\right) \leq \lambda_{\max}\left(A\right)\lambda_{\max}\left(BB^{\intercal}\right).$$

Proof. First, note that $BAB^{\intercal}$ is Hermitian. By Rayleigh-Ritz theorem, we have

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

On the other hand,

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

□

Lemma B.3. *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),

$$\|A \otimes B\|_F^2 = \operatorname{tr}\left[(A^{\intercal} \otimes B^{\intercal})(A \otimes B)\right] = \operatorname{tr}\left[A^{\intercal}A \otimes B^{\intercal}B\right] = \operatorname{tr}(A^{\intercal}A)\operatorname{tr}(B^{\intercal}B) = \|A\|_F^2\|B\|_F^2.$$

For part (ii),

$$\|A \otimes B\|_{\ell_2} = \sqrt{\text{maxeval}[(A \otimes B)^{\intercal}(A \otimes B)]} = \sqrt{\text{maxeval}[(A^{\intercal} \otimes B^{\intercal})(A \otimes B)]}$$
$$= \sqrt{\text{maxeval}[A^{\intercal}A \otimes B^{\intercal}B]} = \sqrt{\text{maxeval}[A^{\intercal}A]\text{maxeval}[B^{\intercal}B]} = \|A\|_{\ell_2}\|B\|_{\ell_2},$$

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

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

□

Lemma B.4. *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 $\max_{1\leq j\leq m}\|\Omega_{n,j}^{-1}\| = O_p(1)$ 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.

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

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

$$w_{j,n,T} := \frac{z_{j,n,T}}{(v_{j,n,T} + z_{j,n,T})v_{j,n,T}} \le x_{j,n,T},$$

whence we have $\max_{1\le j\le m} w_{j,n,T} \le \max_{1\le j\le m} x_{j,n,T} = O_p(a_{m,n,T}) = o_p(1)$. 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

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

where the second equality is due to the fact that $0 \le v_{j,n,T}w_{j,n,T} \le 1$ for any j. □

Theorem B.3. *Let $\{x_{t,i}\}$ be a double-index process having zero mean and being independent across $1 \le t \le T$ for $n \ge 1$ and $T \ge 1$. If there exists k, $k \ge 2$, such that*

$$\max_{n\ge1}\max_{1\le i\le n}\max_{T\ge1}\max_{1\le t\le T} \mathbb{E}|x_{t,i}|^k < \infty,$$

then we have

$$\max_{n\ge1}\max_{1\le i\le n}\max_{T\ge1} \mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{t,i}\right|^k \le K$$

for some absolute positive constant K.

Proof. Slightly adapted from Brillinger (1962). □

Lemma B.5 (Generalised Cauchy-Schwarz Inequality)**.** *For a positive definite matrix S and any vectors ϕ and x*

$$(\phi^\intercal x)^2 \le \phi^\intercal S\phi \cdot x^\intercal S^{-1}x.$$

Proof. See Lemma 5.3.2 (p178) of Anderson (1984). □

References

Akdemir, D. and A. K. Gupta (2011). Array variate random variables with multiway kronecker delta covariance matrix structure. *Journal of Algebraic Statistics 2*(1), 98–113.

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

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

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.

Brillinger, D. R. (1962). A note on the rate of convergence of a mean. *Biometrika 49*, 574–576.

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 Behaviorial 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.

Estienne, F., N. Matthijs, D. Massart, P. Ricoux, and D. Leibovici (2001). Multi-way modelling of high-dimensionality electroencephalographic data. *Chemometrics and Intelligent Laboratory Systems 58*, 59–72.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

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), 603–680.

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 (2019). Estimation of a multiplicative correlation structure in the large dimensional case. *Journal of Econometrics forthcoming*.

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), 1169–1193.

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.

Kroonenberg, P. M. (2008). *Applied Multiway Data Analysis*. John Wiley & Sons.

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*.

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.

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.

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. 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), 231–240.