Journal of Econometrics 208 (2019) 507–534

Contents lists available at ScienceDirect

Journal of Econometrics

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

A multiple testing approach to the regularisation of large sample correlation matrices

Natalia Bailey[a], M. Hashem Pesaran[b,c], L. Vanessa Smith[d,*]

[a] Department of Econometrics and Business Statistics, Monash University, Australia
[b] Department of Economics, University of Southern California, United States
[c] Trinity College, Cambridge, United States
[d] Department of Economics and Related Studies, University of York, UK

ARTICLE INFO

Article history:
Received 13 November 2015
Received in revised form 7 September 2018
Accepted 19 October 2018
Available online 5 November 2018

JEL classification:
C13
C58

Keywords:
High-dimensional data
Multiple testing
Non-Gaussian observations
Sparsity
Thresholding
Shrinkage

ABSTRACT

This paper proposes a regularisation method for the estimation of large covariance matrices that uses insights from the multiple testing (*MT*) literature. The approach tests the statistical significance of individual pair-wise correlations and sets to zero those elements that are not statistically significant, taking account of the multiple testing nature of the problem. The effective p-values of the tests are set as a decreasing function of N (the cross section dimension), the rate of which is governed by the nature of dependence of the underlying observations, and the relative expansion rates of N and T (the time dimension). In this respect, the method specifies the appropriate thresholding parameter to be used under Gaussian and non-Gaussian settings. The *MT* estimator of the sample correlation matrix is shown to be consistent in the spectral and Frobenius norms, and in terms of support recovery, so long as the true covariance matrix is sparse. The performance of the proposed *MT* estimator is compared to a number of other estimators in the literature using Monte Carlo experiments. It is shown that the *MT* estimator performs well and tends to outperform the other estimators, particularly when N is larger than T.

1. Introduction

Improved estimation of covariance matrices is a problem that features prominently in a number of areas of multivariate statistical analysis. In finance it arises in portfolio selection and optimisation (Ledoit and Wolf, 2003), risk management (Fan et al., 2008) and testing of capital asset pricing models (Sentana, 2009). In global macroeconometric modelling with many domestic and foreign channels of interactions, error covariance matrices must be estimated for impulse response analysis and bootstrapping (Pesaran et al., 2004; Dees et al., 2007). In the area of bioinformatics, covariance matrices are required when inferring gene association networks (Carroll, 2003; Schäfer and Strimmer, 2005). Such matrices are further encountered in fields including meteorology, climate research, spectroscopy, signal processing and pattern recognition.

Importantly, the issue of consistently estimating the population covariance matrix, $\boldsymbol{\Sigma} = (\sigma_{ij})$, becomes particularly challenging when the number of variables, N, is larger than the number of observations, T. In this case, one way of obtaining a suitable estimator for $\boldsymbol{\Sigma}$ is to appropriately restrict the off-diagonal elements of its sample estimate denoted by $\hat{\boldsymbol{\Sigma}}$. Numerous methods have been developed to address this challenge, predominantly in the statistics literature. See Pourahmadi (2011) for an extensive review and references therein. Some approaches are regression-based and make use of suitable decompositions

* Correspondence to: Department of Economics, University of York, Heslington, York, YO105 DD, UK.
E-mail address: vanessa.smith@york.ac.uk (L.V. Smith).

https://doi.org/10.1016/j.jeconom.2018.10.006
0304-4076/

of $\boldsymbol{\Sigma}$ such as the Cholesky decomposition (see Pourahmadi (1999, 2000), Rothman et al. (2010), Abadir et al. (2014), among others). Others include banding or tapering methods as proposed, for example, by Bickel and Levina (2004, 2008b) and Wu and Pourahmadi (2009), which assume that the variables under consideration follow a natural ordering. Two popular regularisation techniques in the literature that do not make use of any ordering assumptions are those of thresholding and shrinkage.

Thresholding involves setting off-diagonal elements of the sample covariance matrix that are in absolute terms below certain threshold values to zero. This approach includes 'universal' thresholding put forward by El Karoui (2008) and Bickel and Levina (2008a), and 'adaptive' thresholding proposed by Cai and Liu (2011). Universal thresholding applies the same thresholding parameter to all off-diagonal elements of the unconstrained sample covariance matrix, while adaptive thresholding allows the threshold value to vary across the different off-diagonal elements of the matrix. Furthermore, the selected non-zero elements of $\hat{\boldsymbol{\Sigma}}$ can either be set to their sample estimates or can be adjusted downward. This relates to the concepts of 'hard' and 'soft' thresholding, respectively. The thresholding approach traditionally assumes that the underlying (population) covariance matrix is *sparse*, where sparsity is loosely defined as the presence of a sufficient number of zeros on each row of $\boldsymbol{\Sigma}$ such that it is absolute summable row (column)-wise, or more generally in the sense defined by El Karoui (2008). However, Fan et al. (2011, 2013) show that such regularisation techniques can be applied even if the underlying population covariance matrix is not sparse, so long as the non-sparsity is characterised by an approximate factor structure. The main challenge in applying this approach lies in the estimation of the thresholding parameter, which is primarily calibrated by cross-validation.

In contrast to thresholding, the shrinkage approach reduces all sample estimates of the covariance matrix towards zero element-wise. More formally, the shrinkage estimator of $\boldsymbol{\Sigma}$ is defined as a weighted average of the sample covariance matrix and an invertible covariance matrix estimator known as the shrinkage target - see Friedman (1989). A number of shrinkage targets have been considered in the literature that take advantage of *a priori* knowledge of the data characteristics under investigation. Examples of covariance matrix targets can be found in Ledoit and Wolf (2003), Daniels and Kass (1999, 2001), Fan et al. (2008), and Hoff (2009) , among others. Ledoit and Wolf (2004) suggest a modified shrinkage estimator that involves a linear combination of the unrestricted sample covariance matrix with the identity matrix. This is recommended by the authors for more general situations where no natural shrinking target exists. On the whole, shrinkage estimators tend to be stable, but yield inconsistent estimates if the purpose of the analysis is the estimation of the true and false positive rates of the underlying true sparse covariance matrix (the so called 'support recovery' problem).

This paper considers an alternative approach using a multiple testing (MT) procedure to set the thresholding parameter. A similar idea has been suggested by El Karoui (2008) — p. 2748, who considers testing the $N(N-1)/2$ null hypotheses that $\sigma_{ij}=0$, for all $i \neq j$, jointly. But no formal theory has been developed in the literature for this purpose. In our application of this idea we focus on testing the significance of the correlation coefficients, $\rho_{ij}=\sigma_{ij}/\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}$ for all $i \neq j$, which avoids the scaling problem associated with the use of σ_{ij}, and allows us to obtain a universal threshold for all i and j pairs. We use ideas from the multiple testing literature to control the rate at which the spectral and Frobenius norms of the difference between the true correlation matrix $\boldsymbol{R}=(\rho_{ij})$, and our proposed estimator of it, $\widetilde{\boldsymbol{R}}_{MT}=(\tilde{\rho}_{ij,T})$, tends to zero, and will not be particularly concerned with controlling the overall size of the joint $N(N-1)/2$ tests of $\rho_{ij}=0$, for all $i \neq j$.

We establish that $\widetilde{\boldsymbol{R}}_{MT}$ converges to $\mathbf{R}$ in spectral norm at the rate of $O_p\left(\frac{m_N\, c_p(N)}{\sqrt{T}}\right)$, where m_N is the maximum number of non-zero elements in the off-diagonal rows of $\mathbf{R}$, $c_p(N)=\Phi^{-1}\left(1-\frac{p}{2N^{\delta}}\right)$, $\Phi^{-1}(.)$ is the inverse of the cumulative distribution of a standard normal variate, p is the nominal size of the test, and the choice of $\delta > 0$ is related to the degree of non-Gaussianity of the underlying observations. This is equivalent to the corresponding $O_p\left(m_N\sqrt{\frac{\ln(N)}{T}}\right)$ rate established for the threshold estimator of $\boldsymbol{\Sigma}$ in the literature, considering that $c_p^2(N)/\ln(N) \to 2\delta$ as $N \to \infty$. The main difference between the two approaches is that we use a multiple testing critical value to set the threshold, whilst the literature uses cross validation. It is perhaps also worth noting that our results are established under weaker moment conditions than sub-Gaussianity typically assumed in the literature while comparable to the polynomial-type tail conditions considered in Bickel and Levina (2008a) or Cai and Liu (2011).

In terms of the Frobenius norm, we show that the MT estimator converges at the rate of $O_p\left(\sqrt{\frac{m_N N}{T}}\right)$, for suitable choices of the critical value function in our MT procedure. This result holds even if the underlying observations are non-Gaussian. To the best of our knowledge, the only work that addresses the theoretical properties of the thresholding estimator for the Frobenius norm is Bickel and Levina (2008a), who establish the slower rate of $O_p\left(\sqrt{\frac{m_N N\ln(N)}{T}}\right)$. We also establish conditions under which our proposed estimator consistently recovers the support of the population covariance matrix under Gaussian and non-Gaussian observations, and show that the true positive rate tends to one with probability 1, and the false positive rate and the false discovery rate tend to zero with probability 1, even if N tends to infinity faster than T. We provide conditions under which these results hold.

The performance of the MT estimator is investigated using a Monte Carlo simulation study, and its properties are compared to a number of extant regularised estimators in the literature. The simulation results show that the proposed multiple testing estimator is robust to the typical choices of p used in the literature (10%, 5% and 1%), and performs favourably compared to the other estimators, especially when N is large relative to T. The MT procedure also dominates other regularised estimators when the focus of the analysis is on support recovery.

The rest of the paper is organised as follows: Section 2 outlines some preliminaries, introduces the *MT* procedure and derives its asymptotic properties. The small sample properties of the *MT* estimator are investigated in Section 3. Concluding remarks are provided in Section 4. Some of the technical proofs and additional material are provided in an online supplement.

Notations

$O(.)$ and $o(.)$ denote the Big O and Little o notations, respectively. If $\{f_N\}_{N=1}^{\infty}$ is any real sequence and $\{g_N\}_{N=1}^{\infty}$ is a sequence of positive real numbers, then $f_N = O(g_N)$ if there exists a positive finite constant K such that $|f_N|/g_N \leq K$ for all N. $f_N = o(g_N)$ if $f_N/g_N \to 0$ as $N \to \infty$. $O_p(.)$ and $o_p(.)$ are the equivalent orders in probability. If $\{f_N\}_{N=1}^{\infty}$ and $\{g_N\}_{N=1}^{\infty}$ are both positive sequences of real numbers, then $f_N = \ominus(g_N)$ if there exist $N_0 \geq 1$ and positive finite constants K_0 and K_1, such that $\inf_{N\geq N_0}(f_N/g_N) \geq K_0$, and $\sup_{N\geq N_0}(f_N/g_N) \leq K_1$. The largest and the smallest eigenvalues of the $N \times N$ real symmetric matrix $\boldsymbol{A} = (a_{ij})$ are denoted by $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$, respectively, its trace by $\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{N} a_{ii}$, its maximum absolute column sum norm by $\|\boldsymbol{A}\|_1 = \max_{1\leq j\leq N}\left(\sum_{i=1}^{N}|a_{ij}|\right)$, its maximum absolute row sum norm by $\|\boldsymbol{A}\|_\infty = \max_{1\leq i\leq N}\left(\sum_{j=1}^{N}|a_{ij}|\right)$, its spectral radius by $\varrho(\boldsymbol{A}) = |\lambda_{\max}(\boldsymbol{A})|$, its spectral (or operator) norm by $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$, its Frobenius norm by $\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{A}'\boldsymbol{A})}$. $\xrightarrow{a.s.}$ denotes almost sure convergence, and $\xrightarrow{p}$ convergence in probability. $K, K_0, K_1, C, \varkappa, c_\delta, c_d, \varepsilon_0, \epsilon, \gamma$ and η are finite positive constants, independent of N and T. $\sup_{i,t}$ will be used to denote $\sup_{1\leq i\leq N, 1\leq t\leq T}$. All asymptotics are carried out under N and $T \to \infty$, *jointly*.

2. Regularising the sample correlation matrix: A multiple testing (MT) approach

Let $\{x_{it},\ i \in N,\ t \in T\}$, $N \subseteq \mathbb{N}$, $T \subseteq \mathbb{Z}$, be a double index process where x_{it} is defined on a suitable probability space (Ω, F, P), and denote the covariance matrix of $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ by

$$Var(\boldsymbol{x}_t) = \boldsymbol{\Sigma} = E\left[(\boldsymbol{x}_t - \boldsymbol{\mu})(\boldsymbol{x}_t - \boldsymbol{\mu})'\right], \tag{1}$$

where $E(\boldsymbol{x}_t) = \boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)'$, and $\boldsymbol{\Sigma}$ is an $N \times N$ symmetric, positive definite real matrix with (i,j) element, σ_{ij}. We assume that x_{it} is independent over time, t. We consider the regularisation of the sample covariance matrix estimator of $\boldsymbol{\Sigma}$, denoted by $\hat{\boldsymbol{\Sigma}}$, with elements

$$\hat{\sigma}_{ij,T} = T^{-1}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)\left(x_{jt} - \bar{x}_j\right),\ \text{for } i,j = 1, 2, \ldots, N, \tag{2}$$

where $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$. To this end we assume that $\boldsymbol{\Sigma}$ is (exactly) sparse defined as follows:

Assumption 1. The population covariance matrix, $\boldsymbol{\Sigma} = (\sigma_{ij})$, where $\lambda_{\min}(\boldsymbol{\Sigma}) \geq \varepsilon_0 > 0$, is sparse in the sense that m_N defined by

$$m_N = \max_{1\leq i\leq N}\sum_{j=1}^{N} I\left(\sigma_{ij} \neq 0\right), \tag{3}$$

is $O\left(N^{\vartheta}\right)$ for some $0 \leq \vartheta < 1/2$, where $I(A)$ is an indicator function that takes the value of 1 if A holds and zero otherwise.

A comprehensive discussion of the concept of sparsity applied to $\boldsymbol{\Sigma}$ and alternative ways of defining it are provided in El Karoui (2008) and Bickel and Levina (2008a).

Remark 1. The concept of sparsity defined by (3) is particularly suited to economic applications where the focus of the analysis is often on connections in a given network, or support recovery of $\boldsymbol{\Sigma}$.[1] But our analysis can be readily extended to allow for approximate sparsity entertained in the literature, with (3) replaced by

$$m_{q,N} = \max_{1\leq i\leq N}\sum_{j=1}^{N}\left|\sigma_{ij}\right|^{q},$$

for $0 \leq q < 1$, and $m_{q,N} = O\left(N^{\vartheta}\right)$, with $0 \leq \vartheta < 1/2$. To simplify the exposition we focus on the concept of exact sparsity as defined by Assumption 1.

We follow the hard thresholding literature but, as noted above, we employ multiple testing to decide on the threshold value. More specifically, we set to zero those elements of $\boldsymbol{R} = (\rho_{ij})$ that are statistically insignificant and therefore determine the threshold value as part of a multiple testing strategy. We apply the thresholding procedure explicitly to the correlations

[1] A similar argument is also made in Fan et al. (2011).

rather than the covariances. This has the added advantage that one can use a so-called 'universal' threshold rather than making entry-dependent adjustments, which in turn need to be estimated when thresholding is applied to covariances. This feature is in line with the method of Bickel and Levina (2008a) or El Karoui (2008) but shares the properties of the adaptive thresholding estimator developed by Cai and Liu (2011).

Specifically, denote the sample correlation of x_{it} and x_{jt}, computed over $t = 1, 2, \dots, T$, by

$$\hat{\rho}_{ij,T} = \hat{\rho}_{ji,T} = \frac{\hat{\sigma}_{ij,T}}{\sqrt{\hat{\sigma}_{ii,T}\hat{\sigma}_{jj,T}}}, \tag{4}$$

where $\hat{\sigma}_{ij,T}$ is defined by (2). For a given i and j, it is well known that under $H_{0,ij}: \sigma_{ij} = 0$, $\sqrt{T}\hat{\rho}_{ij,T}$ is asymptotically distributed as $N(0, 1)$ for T sufficiently large. This suggests using $T^{-1/2}\Phi^{-1}\left(1 - \frac{p}{2}\right)$ as the threshold for $\left|\hat{\rho}_{ij,T}\right|$, where $\Phi^{-1}(.)$ is the inverse of the cumulative distribution of a standard normal variate, and p is the chosen nominal size of the test, typically taken to be 1% or 5%. However, since there are in fact $N(N-1)/2$ such tests and N is large, then using the threshold $T^{-1/2}\Phi^{-1}\left(1 - \frac{p}{2}\right)$ for all $N(N-1)/2$ pairs of correlation coefficients will yield inconsistent estimates of $\mathbf{\Sigma}$ and fail to recover its support.

A popular approach to the multiple testing problem is to control the overall size of the $n = N(N-1)/2$ tests jointly (known as family-wise error rate) rather than the size of the individual tests. Let the family of null hypotheses of interest be $H_{01}, H_{02}, \dots, H_{0n}$, and suppose we are provided with the corresponding test statistics, $Z_{1T}, Z_{2T}, \dots, Z_{nT}$, with separate rejection rules given by (using a two-sided alternative)

$$\Pr\left(|Z_{iT}| > CV_{iT} \,|H_{0i}\right) \le p_{iT},$$

where CV_{iT} is some suitably chosen critical value of the test, and p_{iT} is the observed p-value for H_{0i}. Consider now the family-wise error rate (FWER) defined by

$$FWER_T = \Pr\left[\cup_{i=1}^{n}\left(|Z_{iT}| > CV_{iT} \,|H_{0i}\right)\right],$$

and suppose that we wish to control $FWER_T$ to lie below a pre-determined value, p. One could also consider other generalised error rates (see for example Abramovich et al. (2006) or Romano et al. (2008)). Bonferroni (1935) provides a general solution, which holds for all possible degrees of dependence across the separate tests. Using the union bound, we have

$$\begin{aligned}\Pr\left[\cup_{i=1}^{n}\left(|Z_{iT}| > CV_{iT} \,|H_{0i}\right)\right] &\le \sum_{i=1}^{n}\Pr\left(|Z_{iT}| > CV_{iT} \,|H_{0i}\right)\\ &\le \sum_{i=1}^{n} p_{iT}.\end{aligned}$$

Hence to achieve $FWER_T \le p$, it is sufficient to set $p_{iT} \le p/n$. Alternative multiple testing procedures advanced in the literature that are less conservative than the Bonferroni procedure can also be employed. One prominent example is the step-down procedure proposed by Holm (1979) that, similar to the Bonferroni approach, does not impose any further restrictions on the degree to which the underlying tests depend on each other. More recently, Romano and Wolf (2005) proposed step-down methods that reduce the multiple testing procedure to the problem of sequentially constructing critical values for single tests. Such extensions can be readily considered but will not be pursued here.

In our application we scale p by a general function of N, which we denote by $f(N) = c_\delta N^\delta$, where c_δ and δ are finite positive constants, and then derive conditions on δ which ensure consistent support recovery and a suitable convergence rate of the error in estimation of $\boldsymbol{R} = (\rho_{ij})$. In particular, we show that the choice of δ depends on the nature of dependence of the pairs (y_{it}, y_{jt}), for all $i \ne j$, and on the relative rate at which N and T rise. As will be shown in Section 2.1, the degree of dependence is defined by $K_v = \sup_{ij} K_v(\boldsymbol{\theta}_{ij})$ where $\boldsymbol{\theta}_{ij}$ is a vector of cumulants of (y_{it}, y_{jt}). When $\rho_{ij} = 0$ for all i and j, $i \ne j$, this parameter is given by $\varphi_{\max} = \sup_{ij}\left(\varphi_{ij}\right)$ where $\varphi_{ij} = E\left(y_{it}^2 y_{jt}^2 \,|\rho_{ij} = 0\right) > 0$. In the case where y_{it} and y_{jt} are independent under the null, then $\varphi_{\max} = 1$.

Specifically, the multiple testing (MT) estimator of $\boldsymbol{R}$, denoted by $\widetilde{\boldsymbol{R}}_{MT} = \left(\tilde{\rho}_{ij,T}\right)$, is given by

$$\tilde{\rho}_{ij,T} = \hat{\rho}_{ij,T} I\left[\left|\hat{\rho}_{ij,T}\right| > T^{-1/2}c_p(N)\right],\ i = 1, 2, \dots, N-1,\ j = i+1, \dots, N, \tag{5}$$

where

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right), \tag{6}$$

with $f(N) = c_\delta N^\delta$, c_δ, $\delta > 0$. The corresponding MT estimator of $\mathbf{\Sigma}$ is given by

$$\widetilde{\mathbf{\Sigma}}_{MT} = \hat{\boldsymbol{D}}^{1/2}\widetilde{\boldsymbol{R}}_{MT}\hat{\boldsymbol{D}}^{1/2}, \tag{7}$$

where $\hat{\boldsymbol{D}} = \mathrm{diag}(\hat{\sigma}_{11,T}, \hat{\sigma}_{22,T}, \dots, \hat{\sigma}_{NN,T})$. The MT procedure can also be applied to de-factored observations following the de-factoring approach of Fan et al. (2011, 2013).

2.1. Theoretical properties of the MT estimator

To investigate the asymptotic properties of the *MT* estimator defined by (5) we make the following assumption on the bivariate distribution of x_{it} and x_{jt}, for any $i \neq j$, and $t = 1, 2, \ldots, T$.

Assumption 2. Let $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ with mean $\mu_i = E(x_{it})$, $|\mu_i| < K$, variance $\sigma_{ii} = Var(x_{it})$, $0 < \sigma_{ii} < K$, and correlation coefficient $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$, where $\sigma_{ij} = E(y_{it}y_{jt})$, and $\left|\rho_{ij}\right| < 1$. Suppose that $\sup_{i,t} E\,|y_{it}|^{2s} < K$ for some positive integer $s \geq 3$, and let $\boldsymbol{\xi}_{ij,t} = (y_{it}, y_{jt}, y_{it}^2, y_{jt}^2, y_{it}y_{jt})'$ such that for any $i \neq j$ the time series observations $\boldsymbol{\xi}_{ij,t}$, $t = 1, 2, \ldots, T$, are random draws from a common distribution which is absolutely continuous with non-zero density on subsets of $\mathbb{R}^5$.[2]

We begin our theoretical derivations with the following proposition.

Proposition 1. *Let $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$, and suppose that Assumption 2 holds. Consider the sample correlation coefficient given by (4), and note that*

$$\hat{\rho}_{ij,T} = \frac{\sum_{t=1}^{T} (y_{it} - \bar{y}_i)\left(y_{jt} - \bar{y}_j\right)}{\left[\sum_{t=1}^{T} (y_{it} - \bar{y}_i)^2\right]^{1/2}\left[\sum_{t=1}^{T} \left(y_{jt} - \bar{y}_j\right)^2\right]^{1/2}}. \tag{8}$$

Then

$$\rho_{ij,T} = E\left(\hat{\rho}_{ij,T}\right) = \rho_{ij} + \frac{K_m\left(\boldsymbol{\theta}_{ij}\right)}{T} + O\left(T^{-2}\right), \tag{9}$$

$$\omega_{ij,T}^2 = Var\left(\hat{\rho}_{ij,T}\right) = \frac{K_v(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right), \tag{10}$$

uniformly in the i and j ($i \neq j$) pairs, where[3]

$$K_m(\boldsymbol{\theta}_{ij}) = -\frac{1}{2}\rho_{ij}(1-\rho_{ij}^2) + \frac{3}{8}\rho_{ij}\left[\kappa_{ij}(4,0)+\kappa_{ij}(0,4)\right] - \frac{1}{2}\left[\kappa_{ij}(3,1)+\kappa_{ij}(1,3)\right] + \frac{1}{4}\rho_{ij}\kappa_{ij}(2,2), \tag{11}$$

$$K_v(\boldsymbol{\theta}_{ij}) = (1-\rho_{ij}^2)^2 + \frac{1}{4}\rho_{ij}^2\left[\kappa_{ij}(4,0)+\kappa_{ij}(0,4)\right] - \rho_{ij}\left[\kappa_{ij}(3,1)+\kappa_{ij}(1,3)\right] + \frac{1}{2}(2+\rho_{ij}^2)\kappa_{ij}(2,2), \tag{12}$$

$$\kappa_{ij}(4,0) = E(y_{it}^4) - 3,\ \ \kappa_{ij}(0,4) = E(y_{jt}^4) - 3,$$
$$\kappa_{ij}(3,1) = E(y_{it}^3 y_{jt}) - 3\rho_{ij},\ \ \kappa_{ij}(1,3) = E(y_{it} y_{jt}^3) - 3\rho_{ij},$$
$$\kappa_{ij}(2,2) = E(y_{it}^2 y_{jt}^2) - 2\rho_{ij}^2 - 1,$$

$\boldsymbol{\theta}_{ij} = (\rho_{ij}, \kappa_{ij}(0,4)+\kappa_{ij}(4,0), \kappa_{ij}(3,1)+\kappa_{ij}(1,3), \kappa_{ij}(2,2))'$, $\sup_{ij}\left|K_m(\boldsymbol{\theta}_{ij})\right| < K$ *and* $\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$. *Under the additional assumption that y_{it} are Gaussian the above expressions simplify to $K_m(\boldsymbol{\theta}_{ij}) = -\frac{1}{2}\rho_{ij}(1-\rho_{ij}^2)$ and $K_v(\boldsymbol{\theta}_{ij}) = (1-\rho_{ij}^2)^2$, and it follows that $\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) = 1$.*

All proofs are given in Appendix A with supporting lemmas and technical details provided in an online supplement.

Remark 2. From Gayen (1951) p. 232 (Eq. (54)bis) it follows that $K_v(\boldsymbol{\theta}_{ij}) > 0$ for all correlation coefficients $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$, such that $\left|\rho_{ij}\right| < 1$. Further, in the case where $\rho_{ij} = 0$, by (12),

$$\varphi_{ij} := K_v(\boldsymbol{\theta}_{ij}\left|\rho_{ij}=0\right.) = E\left(y_{it}^2 y_{jt}^2\left|\rho_{ij}=0\right.\right) > 0, \tag{13}$$

and by (11),

$$\psi_{ij} := K_m(\boldsymbol{\theta}_{ij}\left|\rho_{ij}=0\right.) = -0.5\left[E\left(y_{it}^3 y_{jt}\left|\rho_{ij}=0\right.\right) + E\left(y_{it}y_{jt}^3\left|\rho_{ij}=0\right.\right)\right]. \tag{14}$$

Note in addition that when y_{it} and y_{jt} are independently distributed, then $\varphi_{ij} = E\left(y_{it}^2\right)E\left(y_{jt}^2\right) = 1$, and $\psi_{ij} = 0$. This is also the case when y_{it} are Gaussian.

The next proposition assists in establishing probability bounds on $\hat{\rho}_{ij,T}$.

Proposition 2. *Consider the standardised correlation coefficient*

$$z_{ij,T} = \left(\hat{\rho}_{ij,T} - \rho_{ij,T}\right)/\omega_{ij,T}, \tag{15}$$

---

2 The restrictions on the common distribution imply that Cramér's condition holds. See p. 45 of Hall (1992) for further details.

3 See also Eqs. (38) and (39) of Gayen (1951).

where $\hat{\rho}_{ij,T}$ is defined by (4), $\rho_{ij,T}$ and $\omega^2_{ij,T}$ are defined by (9) and (10), respectively, and suppose that Assumptions 1 and 2 hold, and for all i and j ($i \neq j$) $\sup_{ij} E\left(\left|z_{ij,T}\right|^s\right) < K$, for some finite integer $s \geq 3$. Then the cumulative distribution function of $z_{ij,T}$, denoted by $F_{ij,T}(x) = \Pr\left(z_{ij,T} \leq x\right)$, has the following Edgeworth expansion

$$\Pr\left(z_{ij,T} \leq x\right) = \Phi(x) + \sum_{r=1}^{s-2} T^{-r/2} g_r(x)\,\phi(x) + O\left[T^{-(s-1)/2}\right] \tag{16}$$

uniformly in $x \in \mathbb{R}$, where $\Phi(x)$ and $\phi(x)$ are the distribution and density functions of the standard Normal $(0, 1)$, respectively, and $g_r(x)$, for $r = 1, 2, \ldots, s-2$, are finite-order polynomials in x of degree $3r-1$ whose coefficients do not depend on x. Furthermore, for all finite $s \geq 3$, and $a_T > 0$, we have

$$\Pr\left(z_{ij,T} \leq -a_T\right) \leq Ke^{-\frac{1}{2}a_T^2} + O\left(T^{-\frac{(s-2)}{2}} a_T^{3(s-2)-1} e^{-\frac{1}{2}a_T^2}\right) + O\left[T^{-(s-1)/2}\right], \tag{17}$$

and

$$\Pr\left(z_{ij,T} > a_T\right) \leq Ke^{-\frac{1}{2}a_T^2} + O\left(T^{-\frac{(s-2)}{2}} a_T^{3(s-2)-1} e^{-\frac{1}{2}a_T^2}\right) + O\left[T^{-(s-1)/2}\right]. \tag{18}$$

Remark 3. This proposition establishes a bound on the probability of $\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > T^{-1/2}c_p(N)$ without requiring sub-Gaussianity, at the expense of the additional order term, $O\left[T^{-(s-1)/2}\right]$, which relates the bound to the order of the moments of $z_{ij,T}$.

Remark 4. It is also possible to use the Berry–Essen inequality and Cramer-type moderate deviation to obtain the probability bounds in the above proposition which could result in better bounds. On this see, in particular, Delaigle et al. (2011), and the recent contributions by Zhou et al. (2018) and Fan et al. (2018) who make use of a robust covariance estimator to tackle heavy-tailed data that yields exponential-type deviation bounds under mild moment conditions. While the focus of these contributions is primarily on large scale dependence-adjusted multiple testing of the mean, application of their approach to our problem could lead to weaker moment conditions.[4]

Using the probability bounds (17) and (18) we first establish the rate of convergence of the *MT* estimator under the spectral norm which implies convergence in eigenvalues and eigenvectors (see El Karoui (2008), and Bickel and Levina (2008b)).

Theorem 1 (*Convergence under spectral norm*)**.** *Consider the sample correlation coefficient of x_{it} and x_{jt}, defined by $\hat{\rho}_{ij,T}$ (see (4)), and denote the associated population correlation matrix by $\boldsymbol{R} = (\rho_{ij})$. Let $T = c_d N^d$, with $c_d > 0$, and suppose that Assumptions 1 and 2 hold. Further, let*

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right),$$

where $0 < p < 1$, $f(N) = c_\delta N^\delta$, with $c_\delta, \delta > 0$. Suppose also that there exists N_0 such that for all $N > N_0$,

$$1 - \frac{p}{2f(N)} > 0, \tag{19}$$

$$\rho_{\min} > c_p(N)/\sqrt{T}, \tag{20}$$

where $\rho_{\min} = \min_{ij}(\left|\rho_{ij}\right|, \rho_{ij} \neq 0)$, and

$$c_p(N)/\sqrt{T} = o(1). \tag{21}$$

Consider values of δ that satisfy condition

$$\delta > \frac{2K_v}{(1-\gamma)^2}, \tag{22}$$

for some small positive constant γ, where $K_v = \sup_{ij} K_v(\boldsymbol{\theta}_{ij})$ and $K_v(\boldsymbol{\theta}_{ij})$ is defined by (12). Then for all values of $d > 4/(s-1)$, where s is defined by Assumption 2, we have

$$\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\| = O_p\left(\frac{m_N c_p(N)}{\sqrt{T}}\right), \tag{23}$$

where m_N is defined by (3), and the multiple testing threshold estimator, $\widetilde{\boldsymbol{R}}_{MT}$, is defined by $\widetilde{\boldsymbol{R}}_{MT} = \left(\tilde{\rho}_{ij,T}\right)$, with $\tilde{\rho}_{ij,T} = \hat{\rho}_{ij,T} I\left[\left|\hat{\rho}_{ij,T}\right| > T^{-1/2}c_p(N)\right]$.

[4] We are grateful to Jianqing Fan (Co-Editor) for drawing our attention to this alternative approach which could form the basis of future investigations.

Remark 5. The term $c_p(N)$ in (23) directly corresponds to the term $\sqrt{\ln(N)}$ obtained in the literature for the probability order, $O_p\left(\frac{m_N\sqrt{\ln(N)}}{\sqrt{T}}\right)$, of the spectral norm of threshold estimators of $\boldsymbol{R}$. This result follows since $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$, for $\delta > 0$.[5]

Remark 6. The parameter d, which controls the rate at which T rises with N, is required to be sufficiently large such that $d > 4/(s-1)$, and $T^{-1/2}c_p(N) = o(1)$ hold. But from result (a) of Lemma 2 in the online supplement, we have $N^{-d/2}c_p(N) = O\left(\sqrt{N^{-d}\ln(N)}\right)$, and condition $T^{-1/2}c_p(N) = o(1)$ will be met if $N^{-d}\ln(N) = o(1)$. Further, recall that the validity of the Edgeworth expansion that underlies our analysis requires s to be finite, and hence condition $d > 0$ will follow from the moment condition $d > 4/(s-1)$, for $s \geq 3$ required by Assumption 2.

Remark 7. Condition (19) is met for $\delta > 0$ and N sufficiently large. Condition (20) can be written as

$$\rho_{\min}^2 > \frac{c_p^2(N)}{T} = \frac{c_p^2(N)}{c_d N^d} = c_d^{-1}\left[\frac{c_p^2(N)}{\ln(N)}\right]\left[\frac{\ln(N)}{N^d}\right].$$

Once again since $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$, then condition (20) will be satisfied for any $\delta > 0$, even if $\rho_{\min}$ tends to zero with N, so long as the rate at which $\rho_{\min}$ tends to zero is slower than $\sqrt{\ln(N)/N^d}$.

Remark 8. Note that under Gaussianity where $K_v = \sup_{ij} K_v(\theta_{ij}) = 1$, condition (22) becomes $\delta > 2$. In general, the spectral norm result requires δ to be set above $2\sup_{ij} K_v(\theta_{ij})$, which turns out to be larger than the value of δ required for the Frobenius norm obtained in the theorem below.

Theorem 2 (*Convergence under Frobenius norm*)**.** *Suppose that conditions of Theorem 1 hold, but (22) is replaced by the weaker condition on δ*

$$\delta > (2-d)\,\varphi_{\max}, \tag{24}$$

where $\varphi_{\max} = \sup_{ij} E\left(y_{it}^2 y_{jt}^2 \,\middle|\, \rho_{ij} = 0\right) > 0$, $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ (see Assumption 2), δ and d are the exponents in $f(N) = c_\delta N^\delta$, and $T = c_d N^d$, with $c_\delta, c_d > 0$. Then for all values of

$$d > \max\left(\frac{2+\vartheta}{s-1}, \frac{4}{s+1}\right), \tag{25}$$

where s is defined by Assumption 2, and ϑ $(0 \leq \vartheta < 1/2)$ is the degree of sparsity of the correlation matrix, $\mathbf{R}$, defined by condition (3) in Assumption 1, we have

$$E\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F = O\left(\sqrt{\frac{m_N N}{T}}\right), \tag{26}$$

and

$$\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F = O_p\left(\sqrt{\frac{m_N N}{T}}\right), \tag{27}$$

where $m_N = O(N^\vartheta)$, and the multiple testing threshold estimator, $\widetilde{\boldsymbol{R}}_{MT}$, is defined by $\widetilde{\boldsymbol{R}}_{MT} = \left(\tilde{\rho}_{ij,T}\right)$, with $\tilde{\rho}_{ij,T} = \hat{\rho}_{ij,T} I\left[\left|\hat{\rho}_{ij,T}\right| > T^{-1/2}c_p(N)\right]$.

Remark 9. For the Frobenius norm result to hold condition (24) implies that δ should be set at a sufficiently high level, determined by d (the relative expansion rates of N and T), and $\varphi_{\max}$ (the maximum degree of dependence between y_{it} and y_{jt} when $\rho_{ij} = 0$). The Frobenius norm result holds even if N rises faster than T, so long as $c_p(N)/\sqrt{T} = o(1)$ and a sufficient number of moments exists such that condition (25) is met. In the case where N and T are of the same order of magnitude (namely, $d = 1$), and where y_{it} and y_{jt} are independently distributed when $\rho_{ij} = 0$ (namely, $\varphi_{\max} = 1$), then the Frobenius norm results, (26) and (27), require $\delta > 1$.

Remark 10. The number of moments, s, of y_{it} required for the convergence results (23), (26) or (27) to hold is related to the relative rate of expansion of N and T, d. For $d = 1$, $s = 5$ moments of $\hat{\rho}_{ij}$ (which requires x_{it} to have 10 moments) are sufficient to achieve the spectral or Frobenius norm results. Additional moments are required if N is to rise faster than T.

5 See part (b) of Lemma 2 in the online supplement.

Remark 11. The convergence rate of $O_p\left(\sqrt{\frac{m_N N}{T}}\right)$ obtained for the MT estimator under the Frobenius norm compares favourably to the corresponding rate of $O_p\left(\sqrt{\frac{m_N N \ln(N)}{T}}\right)$ obtained for the threshold estimator. See, for example, Theorem 2 of Bickel and Levina (2008a), BL. The slower rate of convergence achieved by BL under the Frobenius norm arises from the fact that their result is derived by explicitly using their spectral norm convergence rate. On the other hand, we consider the derivation of the Frobenius norm convergence rate directly and independently of our spectral norm results. Furthermore, the sparsity condition of Assumption 1 sets an upper bound, m_N, on the number of non-zero units in the rows (columns) of the population covariance matrix $\boldsymbol{\Sigma}$, but it is silent as to the number of rows (columns) of $\boldsymbol{\Sigma}$ with m_N non-zero elements. Whilst this ambiguity does not impact the convergence rate obtained for the spectral norm, it does affect the Frobenius norm. In many economic applications it might be known that only a finite number of rows of $\boldsymbol{\Sigma}$, say k, have at most m_N non-zero elements with the rest of the rows only containing a fixed number of non-zero elements, say m_0, which is bounded in N. Under this notion of sparsity the convergence rate of the Frobenius norm will be given by

$$\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F = O_p\left(\sqrt{\frac{km_N}{T}}\right) + O_p\left(\sqrt{\frac{(N-k)m_0}{T}}\right),$$

which has a more favourable convergence rate as compared to (27).

Remark 12. It is interesting to note that application of the Bonferroni procedure to the problem of testing $\rho_{ij} = 0$ for all $i \neq j$, is equivalent to setting $f(N) = N(N-1)/2$. Our theoretical results suggest that this can be too conservative if $\rho_{ij} = 0$ implies y_{it} and y_{jt} are independent, but could be appropriate otherwise depending on the relative rates at which N and T rise. In our Monte Carlo study we consider $\delta = \{1, 2\}$, that corresponds to $\varphi_{\max} = \{1, 1.5\}$.

Consider now the issue of consistent support recovery of $\boldsymbol{R}$ (or $\boldsymbol{\Sigma}$) for $T = T(N) = c_d N^d$ and $N \to \infty$, which is defined in terms of the true positive rate (TPR_N), false positive rate (FPR_N), and false discovery rate (FDR_N) statistics. Consistent support recovery requires $TPR_N \to 1$, $FPR_N \to 0$ and $FDR_N \to 0$, with probability 1 (almost surely) as $N \to \infty$, and does not follow immediately from the results obtained above on the convergence rates of different estimators of $\boldsymbol{R}$. This is addressed in the following theorem.

Theorem 3 (*Support Recovery*)**.** *Suppose that Assumption 1 and 2 hold, and let*

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right),$$

where $0 < p < 1$, $f(N) = c_\delta N^\delta$, with $c_\delta, \delta > 0$, and $T = c_d N^d$, with $c_d > 0$. Further, suppose that there exists N_0 such that for all $N > N_0$,

$$1 - \frac{p}{2f(N)} > 0,$$

$$\rho_{\min} > c_p(N)/\sqrt{T}, \tag{28}$$

where $\rho_{\min} = \min_{ij}(\left|\rho_{ij}\right|, \rho_{ij} \neq 0)$, and

$$c_p(N)/\sqrt{T} = o(1). \tag{29}$$

Consider the true positive rate (TPR_N), the false positive rate (FPR_N), and the false discovery rate (FDR_N) statistics defined by

$$TPR_N = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} \neq 0)}{\sum\sum_{i \neq j} I(\rho_{ij} \neq 0)} \tag{30}$$

$$FPR_N = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum\sum_{i \neq j} I(\rho_{ij} = 0)}, \tag{31}$$

$$FDR_N = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum\sum_{i \neq j} I(\rho_{ij} \neq 0)}, \tag{32}$$

computed using the multiple testing estimator

$$\tilde{\rho}_{ij,T} = \hat{\rho}_{ij,T} I\left[\left|\hat{\rho}_{ij,T}\right| > T^{-1/2} c_p(N)\right],$$

where $\hat{\rho}_{ij,T}$ is the pair-wise correlation coefficient defined by (4). Then as $N \to \infty$ we have:

$$TPR_N \xrightarrow{a.s.} 1, \text{ for } \delta > 0, \text{ and } d > 2/(s-1)$$
$$FPR_N \xrightarrow{a.s.} 0, \text{ for } \delta > \varphi_{\max}, \text{ and } d > 2/(s-1)$$
$$FDR_N \xrightarrow{a.s.} 0, \text{ for } \delta > (2-\vartheta)\varphi_{\max}, \text{ and } d > 2\,(2-\vartheta)\,/(s-1)$$

where ϑ $(0 \leq \vartheta < 1/2)$ is the degree of sparsity of the correlation matrix, $\mathbf{R}$, defined by condition (3) in Assumption 1, $\varphi_{\max} = \sup_{ij} E\left(y_{it}^2 y_{jt}^2 \left| \rho_{ij} = 0\right.\right) > 0$, with $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ (see Assumption 2). Further, as $N \to \infty$, $TPR_N \to 1$ and $FPR_N \to 0$ in probability for any $\delta > 0$ and $d > 2/(s-1)$, and $FDR_N \to 0$ in probability if $\delta > (1-\vartheta)\varphi_{\max}$, and $d > 2\,(1-\vartheta)\,/(s-1)$.

Remark 13. We note that

$$\frac{c_p^2(N)}{T} \leq \frac{2\left[\ln(N) - \ln(p)\right]}{c_d N^d},$$

and hence condition $T^{-1/2}c_p(N) = o(1)$ will be met if $N^{-d}\ln(N) = o(1)$. Also, since under Assumption 2 $s \geq 3$, it follows from the moment conditions on d that $d > 0$. For a discussion of the remaining conditions on δ, d, and $\rho_{\min} > T^{-1/2}c_p(N) > 0$, see the above Remarks. In general, the conditions needed for the support recovery results to hold when N is much larger than T are less restrictive as compared to the conditions needed for the validity of the results on the spectral and Frobenius norms.

3. Monte Carlo simulations

We investigate the numerical properties of the proposed multiple testing (MT) estimator using Monte Carlo simulations. We compare our estimator with a number of thresholding and shrinkage estimators proposed in the literature, namely the thresholding estimators of Bickel and Levina (2008a) – BL – and Cai and Liu (2011) — CL, and the shrinkage estimator of LW. The thresholding methods of BL and CL require the computation of a theoretical constant, C, that arises in the rate of their convergence. For this purpose, cross-validation is typically employed which we use when implementing these estimators. For the CL approach we also consider the theoretical value of $C = 2$ derived by the authors in the case of Gaussianity. A review of these estimators along with details of the associated cross-validation procedure can be found in the Supplementary Appendix B.

We begin by generating the standardised variates, y_{it}, as

$$\mathbf{y}_t = \boldsymbol{P}\mathbf{u}_t, \; t = 1, 2, \ldots, T,$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$, and $\boldsymbol{P}$ is the Cholesky factor associated with the choice of the correlation matrix $\boldsymbol{R} = \boldsymbol{P}\boldsymbol{P}'$. We consider two alternatives for the errors, u_{it}: (i) the benchmark Gaussian case where $u_{it} \sim IIDN(0, 1)$ for all i and t, and (ii) the case where u_{it} follows a multivariate t-distribution with v degrees of freedom generated as

$$u_{it} = \left(\frac{\text{v}-2}{\chi^2_{\text{v},t}}\right)^{1/2} \varepsilon_{it}, \text{ for } i = 1, 2, \ldots, N,$$

where $\varepsilon_{it} \sim IIDN(0, 1)$, and $\chi^2_{\text{v},t}$ is a chi-squared random variate with $\text{v} > 4$ degrees of freedom, distributed independently of ε_{it} for all i and t. In order to investigate the robustness of our results to the moment conditions, we experiment with a relatively low degrees of freedom for the t-distribution and set $\text{v} = 8$, which ensures that $E\left(y_{it}^6\right)$ exists and $\varphi_{\max} \leq 2$. Note that under $\rho_{ij} = 0$, $\varphi_{ij} = E\left(y_{it}^2 y_{jt}^2 \left| \rho_{ij} = 0\right.\right)$ =(v−2)/(v−4), and with $\text{v} = 8$ we have $\varphi_{ij} = \varphi_{\max} = 1.5$. Given our theoretical findings, it is most likely that we obtain better results if we experiment with higher degrees of freedom. One could further allow for fat-tailed ε_{it} shocks, say, though fat-tail shocks alone (e.g. generating u_{it} as such) do not necessarily result in $\varphi_{ij} > 1$ as shown in Lemma 6 in the online supplementary Appendix A. The same is true for normal shocks under case (i) where $E\left(y_{it}^2 y_{jt}^2\right) = 1$ whether $\boldsymbol{P} = \boldsymbol{I}_N$ or not. In such cases setting $\delta = 1$ is likely to be sufficient for the Frobenius norms given the (N, T) combinations considered. But for the spectral norm a larger value of δ might be necessary. In order to verify and calibrate the values of δ corresponding to the alternative processes generating y_{it}, we also consider an estimated version of δ. For this purpose we use a cross-validation procedure that corresponds to those used for the BL and CL methods respectively. Details can be found in Section 3.7.

Next, the non-standardised variates $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ are generated as

$$\boldsymbol{x}_t = \mathbf{a} + \boldsymbol{\gamma} f_t + \mathbf{D}^{1/2}\mathbf{y}_t, \tag{33}$$

where $\boldsymbol{D} = \text{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{NN})$, $\mathbf{a} = (a_1, a_2, \ldots, a_N)'$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N)'$.

We report results for $N = \{30, 100, 200\}$ and $T = 100$, for the baseline case where $\gamma = 0$ and $a = 0$ in (33). The properties of the MT procedure when factors are included in the data generating process are also investigated by drawing γ_i and a_i as $IIDN\,(1, 1)$ for $i = 1, 2, \ldots, N$, and generating f_t, the common factor, as a stationary AR(1) process, but to save space these results are made available upon request. Under both settings we focus on the residuals from an OLS regression of $\boldsymbol{x}_t$ on an intercept and a factor (if needed).

Given our interest in both the problems of regularisation of $\hat{\boldsymbol{\Sigma}}$ and support recovery of $\boldsymbol{\Sigma}$, we consider two exactly sparse covariance (correlation) matrices:

Monte Carlo design A: Following Cai and Liu (2011) we consider the banded matrix

$$\boldsymbol{\Sigma} = \left(\sigma_{ij}\right) = \text{diag}(\boldsymbol{A}_1, \boldsymbol{A}_2),$$

where $\boldsymbol{A}_1 = \boldsymbol{A} + \epsilon \boldsymbol{I}_{N/2}$, $\boldsymbol{A} = (a_{ij})_{1\leq i,j\leq N/2}$, $a_{ij} = (1 - \frac{|i-j|}{10})_+$ with $\epsilon = \max(-\lambda_{\min}(\boldsymbol{A}), 0) + 0.01$ to ensure that $\boldsymbol{A}$ is positive definite, and $\boldsymbol{A}_2 = 4\boldsymbol{I}_{N/2}$. $\boldsymbol{\Sigma}$ is a two-block diagonal matrix, $\boldsymbol{A}_1$ is a banded and sparse covariance matrix, and $\boldsymbol{A}_2$ is a diagonal matrix with 4 along the diagonal. Matrix $\boldsymbol{P}$ is obtained numerically by applying the Cholesky decomposition to the correlation matrix, $\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}^{-1/2} = \mathbf{PP}'$, where the diagonal elements of $\boldsymbol{D}$ are given by $\sigma_{ii} = 1 + \epsilon$, for $i = 1, 2, \ldots, N/2$ and $\sigma_{ii} = 4$, for $i = N/2 + 1, N/2 + 1, \ldots, N$.

Monte Carlo design B: We consider a covariance structure that explicitly controls for the number of non-zero elements of the population correlation matrix. First we draw the $N \times 1$ vector $\mathbf{b} = (b_1, b_2, \ldots, b_N)'$ with elements generated as *Uniform* (0.7, 0.9) for the first and last N_b ($< N$) elements of $\mathbf{b}$, where $N_b = \left[N^{\beta}\right]$, and set the remaining middle elements of $\mathbf{b}$ to zero. The resulting population correlation matrix $\boldsymbol{R}$ is defined by

$$\boldsymbol{R} = \boldsymbol{I}_N + \boldsymbol{b}\boldsymbol{b}' - \text{diag}\left(\boldsymbol{b}\boldsymbol{b}'\right), \tag{34}$$

for which $\sqrt{T}\rho_{\min} - c_p(N) > 0$ and $\rho_{\min} = \min_{ij}\left(\left|\rho_{ij}\right|, \rho_{ij} \neq 0\right) > 0$, in line with Theorem 3. The degree of sparseness of $\boldsymbol{R}$ is determined by the value of the parameter β. We are interested in weak cross-sectional dependence, so we focus on the case where $\beta < 1/2$ following Pesaran (2015), and set $\beta = 0.25$. Matrix $\boldsymbol{P}$ is then obtained by applying the Cholesky decomposition to $\boldsymbol{R}$ defined by (34). Further, we set $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{R}\boldsymbol{D}^{1/2}$, where the diagonal elements of $\boldsymbol{D}$ are given by $\sigma_{ii} \sim IID\left(1/2 + \chi^2(2)/4\right)$, $i = 1, 2, \ldots, N$.

3.1. Finite sample positive definiteness

As with other thresholding approaches, multiple testing preserves the symmetry of $\hat{\boldsymbol{R}}$ and is invariant to the ordering of the variables but it does not ensure positive definiteness of the estimated covariance matrix when $N > T$.

A number of methods have been developed in the literature that produce sparse inverse covariance matrix estimates which make use of a penalised likelihood (D'Aspremont et al., 2008; Rothman et al., 2008, 2009; Yuan and Lin, 2007; Peng et al., 2009) or convex optimisation techniques that apply suitable penalties such as a logarithmic barrier term (Rothman, 2012), a positive definiteness constraint (Xue et al., 2012), an eigenvalue condition (Liu et al. (2014), Fryzlewicz (2013), Fan et al. (2013) — FLM). Most of these approaches are rather complex and computationally extensive.

A simpler alternative, which conceptually relates to soft thresholding (such as the smoothly clipped absolute deviation by Fan and Li (2001) and the adaptive lasso by Zou (2006)), is to consider a convex linear combination of $\widetilde{\boldsymbol{R}}_{MT}$ and a well-defined target matrix which is known to result in a positive definite matrix. In what follows, we opt to set as benchmark target the $N \times N$ identity matrix, $\mathbf{I}_N$, in line with one of the methods suggested by El Karoui (2008). The advantage of doing so lies in the fact that the same support recovery achieved by $\widetilde{\boldsymbol{R}}_{MT}$ is maintained and the diagonal elements of the resulting correlation matrix do not deviate from unity. Given the similarity of this adjustment to the shrinking method, we dub this step shrinkage on our multiple testing estimator (S-MT),

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT}\left(\xi\right) = \xi \boldsymbol{I}_N + (1 - \xi)\widetilde{\boldsymbol{R}}_{MT}, \tag{35}$$

with shrinkage parameter $\xi \in (\xi_0, 1]$, and ξ_0 being the minimum value of ξ that produces a non-singular $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0)$ matrix. Alternative ways of computing the optimal weights on the two matrices can be entertained. We choose to calibrate, ξ, since opting to use ξ_0 in (35), as suggested in El Karoui (2008), does not necessarily provide a well-conditioned estimate of $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$. Accordingly, we set ξ by solving the following optimisation problem

$$\xi^* = \arg \min_{\xi_0+\epsilon\leq\xi\leq1} \left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \right\|_F^2, \tag{36}$$

where ϵ is a small positive constant, and $\boldsymbol{R}_0$ is a reference invertible correlation matrix. Finally, we construct the corresponding covariance matrix as

$$\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT}\left(\xi^*\right) = \hat{\boldsymbol{D}}^{1/2}\widetilde{\boldsymbol{R}}_{S\text{-}MT}\left(\xi^*\right)\hat{\boldsymbol{D}}^{1/2}.$$

Further details on the S-MT procedure, the optimisation of (36) and choice of reference matrix $\mathbf{R}_0$ are available in the Supplementary Appendix C.

3.2. Alternative estimators and evaluation metrics

Using the earlier set up and the relevant adjustments to achieve positive definiteness of the estimators of $\boldsymbol{\Sigma}$ where required, we obtain the following estimates of $\boldsymbol{\Sigma}$:

MT_1: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\boldsymbol{\Sigma}}_{MT}$) using $\delta = 1$ ($\widetilde{\boldsymbol{\Sigma}}_{MT,1}$)

MT_2: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\boldsymbol{\Sigma}}_{MT}$) using $\delta = 2$ ($\widetilde{\boldsymbol{\Sigma}}_{MT,2}$)

$MT_{\hat{\delta}}$: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\boldsymbol{\Sigma}}_{MT}$) using cross-validated δ ($\widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}$)

$BL_{\hat{C}}$: BL thresholding on the sample covariance matrix using cross-validated C ($\widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}$)

CL_2: CL thresholding on the sample covariance matrix using the theoretical value of $C = 2$ ($\widetilde{\boldsymbol{\Sigma}}_{CL,2}$)

$CL_{\hat{C}}$: CL thresholding on the sample covariance matrix using cross-validated C ($\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}$)

$S\text{-}MT_1$: supplementary shrinkage applied to MT_1 ($\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,1}$)

$S\text{-}MT_2$: supplementary shrinkage applied to MT_2 ($\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,2}$)

$S\text{-}MT_{\hat{\delta}}$: supplementary shrinkage applied to $MT_{\hat{\delta}}$ ($\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,\hat{\delta}}$)

$BL_{\hat{C}*}$: BL thresholding using the Fan et al. (2013) – FLM – cross-validation adjustment procedure for estimating C to ensure positive definiteness ($\widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}*}$)

$CL_{\hat{C}*}$: CL thresholding using the FLM cross-validation adjustment procedure for estimating C to ensure positive definiteness ($\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}*}$)

$LW_{\hat{\Sigma}}$: LW shrinkage on the sample covariance matrix ($\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}$).

In accordance with the theoretical results and in view of Remark 12, we consider three versions of the MT estimator depending on the choice of $\delta = \left\{1, 2, \hat{\delta}\right\}$. The $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$ estimators apply the thresholding procedure without ensuring that the resultant covariance estimators are invertible. The next six estimators yield invertible covariance estimators. The $S\text{-}MT$ estimators are obtained using the supplementary shrinkage approach described in Section 3.1. $BL_{\hat{C}*}$ and $CL_{\hat{C}*}$ estimators are obtained by applying the additional FLM adjustments. The shrinkage estimator, $LW_{\hat{\Sigma}}$, is invertible by construction. In the case of the MT estimators where regularisation is performed on the correlation matrix, the associated covariance matrix is estimated as $\hat{\mathbf{D}}^{1/2}\widetilde{\boldsymbol{R}}_{MT}\hat{\mathbf{D}}^{1/2}$.

For both Monte Carlo designs A and B, we compute the spectral and Frobenius norms of the deviations of each of the regularised covariance matrices from their respective population $\boldsymbol{\Sigma}$:

$$\left\|\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}}\right\| \text{ and } \left\|\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}}\right\|_F, \tag{37}$$

where $\mathring{\boldsymbol{\Sigma}}$ is set to one of the following estimators $\{\widetilde{\boldsymbol{\Sigma}}_{MT,1}, \widetilde{\boldsymbol{\Sigma}}_{MT,2}, \widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{CL,2}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,1}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,2}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,\hat{\delta}}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}*}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}*}, \hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. The threshold values, $\hat{\delta}$, $\hat{C}$ and $\hat{C}^*$, are obtained by cross-validation (see Section 3.7 and supplementary Appendix B.3 for details). Both norms are also computed for the difference between $\boldsymbol{\Sigma}^{-1}$, the population inverse of $\boldsymbol{\Sigma}$, and the estimators $\{\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,1}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,2}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}*}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}*}, \hat{\boldsymbol{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. Further, we investigate the ability of the thresholding estimators to recover the support of the true covariance matrix via the true positive rate (TPR) and false positive rate (FPR), as defined by (30) and (31), respectively. The statistics TPR and FPR are not relevant to the shrinkage estimator $LW_{\hat{\Sigma}}$ and will not be reported for this estimator.

3.3. Robustness of MT to the choice of p-values

We begin by investigating the sensitivity of the MT estimator to the choice of the p-value, p, and the scaling factor determined by δ used in the formulation of $c_p(N)$ defined by (6). For this purpose we consider the typical significance levels used in the literature, namely $p = \{0.01, 0.05, 0.10\}$, $\delta = \{1, 2\}$, and a cross-validated version of δ, denoted by $\hat{\delta}$. Tables 1a and 1b summarise the spectral and Frobenius norm losses (averaged over 2000 replications) for Monte Carlo designs A and B respectively, and for both distributional error assumptions (Gaussian and multivariate t). First, we note that neither of the norms is much affected by the choice of the p values when setting $\delta = 1$ or 2 in the scaling factor, irrespective of whether the observations are drawn from a Gaussian or a multivariate t distribution. Similar results are also obtained using the cross validated version of δ. Perhaps this is to be expected since for N sufficiently large the effective p-value which is given by $2p/N^{\delta}$ is very small and the test outcomes are more likely to be robust to the choice of p values as compared to the choice of δ. The results in Tables 1a and 1b also show that in the case of Gaussian observations, where $\varphi_{\max} = 1$, the scaling factor using $\delta = 1$ is likely to perform better as compared to $\delta = 2$, but the reverse is true if the observations are multivariate t distributed under which the scaling factor using $\delta = 2$ is to be preferred.

It is also interesting that the performance of the MT procedure when using $\hat{\delta}$ is in line with our theoretical findings. The estimates of δ are closer to unity in the case of experiments with $\varphi_{\max} = 1$, and are closer to $\delta = 2$ in the case of experiments with $\varphi_{\max} = 1.5$. The average estimates of $\hat{\delta}$ shown in Tables 1a and 1b are also indicative that a higher value of δ is required when observations are multivariate t distributed. Finally, we note that the norm losses rise with N given that T is kept at 100 almost across the board in all the experiments. Overall, the simulation results support using a sufficiently high value of δ (say around 2) or its estimate, $\hat{\delta}$, obtained by cross validation.

3.4. Norm comparisons of MT, BL, CL, and LW estimators

In comparing our proposed estimators with those in the literature we consider a fewer number of Monte Carlo replications and report the results with norm losses averaged over 100 replications, given the use of the cross-validation procedure in the implementation of MT, BL and CL thresholding. This Monte Carlo specification is in line with the simulation set up of BL and CL. Our reported results are also in agreement with their findings.

Table 1a
Spectral and Frobenius norm losses for the MT estimator using significance levels $p = \{0.01, 0.05, 0.10\}$ and scaling factors with $\delta = \left\{1, 2, \hat{\delta}\right\}$, for $T = 100$.

Monte Carlo design A									
	$\delta = 1$			$\delta = 2$			$\hat{\delta}$		
$N\backslash p$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
	$\mathbf{u}_{it}$ ∼ Gaussian								
	Spectral norm								
30	1.70(0.49)	1.68(0.49)	1.71(0.49)	1.89(0.51)	1.79(0.50)	1.75(0.50)	1.71(0.49)	1.68(0.49)	1.69(0.49)
100	2.61(0.50)	2.51(0.50)	2.50(0.50)	3.11(0.50)	2.91(0.50)	2.84(0.50)	2.62(0.50)	2.52(0.50)	2.51(0.50)
200	3.04(0.48)	2.92(0.49)	2.89(0.49)	3.67(0.47)	3.46(0.47)	3.37(0.47)	3.05(0.48)	2.93(0.49)	2.90(0.49)
	Frobenius norm								
30	3.17(0.45)	3.14(0.50)	3.20(0.53)	3.49(0.42)	3.32(0.43)	3.26(0.43)	3.19(0.44)	3.13(0.48)	3.16(0.52)
100	6.67(0.45)	6.51(0.51)	6.60(0.55)	7.75(0.40)	7.34(0.41)	7.17(0.42)	6.70(0.45)	6.52(0.50)	6.57(0.54)
200	9.87(0.46)	9.60(0.53)	9.73(0.58)	11.76(0.40)	11.15(0.41)	10.89(0.42)	9.91(0.46)	9.62(0.52)	9.69(0.57)
	$\mathbf{u}_{it}$ ∼ multivariate t-distributed with 8 degrees of freedom								
	Spectral norm								
30	2.26(1.08)	2.42(1.20)	2.55(1.26)	2.29(0.90)	2.24(0.99)	2.24(1.03)	2.23(0.95)	2.32(1.04)	2.39(1.08)
100	3.85(4.84)	4.20(5.28)	4.46(5.48)	3.78(3.78)	3.71(4.12)	3.71(4.27)	3.67(3.81)	3.83(4.11)	3.93(4.21)
200	4.49(3.46)	5.04(4.34)	5.44(4.77)	4.26(1.80)	4.20(2.21)	4.19(2.37)	4.20(2.43)	4.45(2.78)	4.57(2.94)
	Frobenius norm								
30	4.06(1.14)	4.35(1.32)	4.60(1.40)	4.12(0.90)	4.04(1.00)	4.03(1.06)	4.03(1.00)	4.19(0.13)	4.32(1.19)
100	8.88(5.17)	9.75(5.67)	10.49(5.87)	9.04(4.04)	8.80(4.40)	8.74(4.57)	8.65(4.16)	9.09(4.48)	9.41(4.59)
200	12.96(4.23)	14.50(5.41)	15.81(5.95)	13.25(2.10)	12.85(2.54)	12.71(2.76)	12.57(2.97)	13.25(3.48)	13.73(3.67)

	Cross validated values of δ		
$N\backslash p$	0.01	0.05	0.10
	$\mathbf{u}_{it}$ ∼ Gaussian		
30	1.08(0.11)	1.10(0.12)	1.12(0.13)
100	1.04(0.06)	1.05(0.07)	1.06(0.08)
200	1.03(0.05)	1.03(0.06)	1.04(0.06)
	$\mathbf{u}_{it}$ ∼ multivariate t-distr. with 8 dof		
30	1.13(0.18)	1.19(0.22)	1.25(0.25)
100	1.12(0.18)	1.18(0.22)	1.23(0.25)
200	1.15(0.20)	1.20(0.23)	1.24(0.25)

Note: The MT approach is implemented using $\delta = 1$, $\delta = 2$, and $\hat{\delta}$, computed using cross-validation. Norm losses and estimates of δ, $\hat{\delta}$, are averages over 2,000 replications. Simulation standard deviations are given in parentheses.

Tables 2 and 3 summarise the results for the Monte Carlo designs A and B, respectively. Based on the results of Section 3.3, we provide norm comparisons for the MT estimator using the scaling factor where $\delta = 2$ and $\hat{\delta}$, and the conventional significance level of $p = 0.05$. Initially, we consider the threshold estimators, the two versions of MT (MT_2 and $MT_{\hat{\delta}}$) and CL (CL_2 and $CL_{\hat{C}}$) estimators, and BL without further adjustments to ensure invertibility. First, we note that the MT and CL estimators (both versions for each case) dominate the BL estimator in every case, and for both designs. MT performs better than CL, when comparing the versions of the two estimators using their respective theoretical thresholding values and their estimated equivalents. The outperformance of MT is more evident as N increases and when non-Gaussian observations are considered. The same is also true if we compare MT and CL estimators to the LW shrinkage estimator, although it could be argued that it is more relevant to compare the invertible versions of the MT and CL estimators (namely $\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}*}$, $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,\hat{\delta}}$) with $\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}$. In such comparisons $\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}$ performs relatively better, nevertheless, $\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}$ is still dominated by $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,\hat{\delta}}$, with a few exceptions in the case of design A and primarily when $N = 30$. However, no clear ordering emerges when we compare $\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}$ with $\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}*}$.

3.5. Norm comparisons of inverse estimators

Although the theoretical focus of this paper has been on estimation of $\boldsymbol{\Sigma}$ rather than its inverse, it is still of interest to see how well $\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,2}$, $\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}$, $\widetilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}*}$, $\widetilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}*}$, and $\hat{\boldsymbol{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}$ estimate $\boldsymbol{\Sigma}^{-1}$, assuming that $\boldsymbol{\Sigma}^{-1}$ is well defined. Table 4 provides average norm losses for Monte Carlo design B for which $\boldsymbol{\Sigma}$ is positive definite. $\boldsymbol{\Sigma}$ for design A is ill-conditioned and will not be considered any further here. As can be seen from the results in Table 4, both $\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}$ perform much better than $\widetilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}*}$ and $\widetilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}*}$ for Gaussian and multivariate t-distributed observations. In fact, the average spectral norms for $\widetilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}*}$ and $\widetilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}*}$ include some sizeable outliers, especially for $N \leq 100$. However, the ranking of the different estimators remains the same if we use the Frobenius norm which appears to be less sensitive to the outliers. It is also worth noting that

Table 1b
Spectral and Frobenius norm losses for the *MT* estimator using significance levels $p = \{0.01, 0.05, 0.10\}$ and scaling factors with $\delta = \left\{1, 2, \hat{\delta}\right\}$, for $T = 100$.

Monte Carlo design B									
	$\delta = 1$			$\delta = 2$			$\hat{\delta}$		
$N \backslash p$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
	$\mathbf{u}_{it}$ ∼ Gaussian								
	Spectral norm								
30	0.48(0.16)	0.50(0.16)	0.53(0.16)	0.50(0.20)	0.49(0.18)	0.48(0.17)	0.48(0.17)	0.49(0.16)	0.49(0.16)
100	0.75(0.34)	0.76(0.32)	0.78(0.31)	0.89(0.43)	0.81(0.39)	0.79(0.37)	0.76(0.35)	0.76(0.34)	0.76(0.34)
200	0.71(0.22)	0.74(0.20)	0.77(0.20)	0.85(0.33)	0.78(0.28)	0.75(0.26)	0.72(0.24)	0.72(0.22)	0.72(0.22)
	Frobenius norm								
30	0.87(0.17)	0.91(0.18)	0.97(0.19)	0.89(0.20)	0.87(0.17)	0.86(0.17)	0.86(0.17)	0.88(0.17)	0.88(0.17)
100	1.56(0.24)	1.66(0.24)	1.77(0.24)	1.67(0.34)	1.60(0.29)	1.58(0.27)	1.56(0.25)	1.58(0.24)	1.58(0.25)
200	2.16(0.18)	2.32(0.20)	2.50(0.21)	2.25(0.24)	2.19(0.21)	2.16(0.20)	2.15(0.18)	2.18(0.19)	2.18(0.20)
	$\mathbf{u}_{it}$ ∼ multivariate t-distributed with 8 degrees of freedom								
	Spectral norm								
30	0.70(0.39)	0.78(0.43)	0.84(0.45)	0.67(0.33)	0.67(0.35)	0.67(0.37)	0.67(0.33)	0.68(0.35)	0.68(0.36)
100	1.16(0.97)	1.32(1.10)	1.42(1.18)	1.15(0.75)	1.11(0.80)	1.10(0.83)	1.10(0.72)	1.10(0.77)	1.11(0.80)
200	1.36(1.73)	1.65(2.05)	1.83(2.20)	1.14(1.03)	1.13(1.21)	1.14(1.28)	1.16(1.06)	1.19(1.20)	1.20(1.27)
	Frobenius norm								
30	1.23(0.42)	1.40(0.48)	1.53(0.51)	1.15(0.35)	1.16(0.38)	1.17(0.39)	1.17(0.36)	1.19(0.38)	1.20(0.39)
100	2.39(1.12)	2.90(1.31)	3.25(1.40)	2.17(0.77)	2.15(0.86)	2.16(0.90)	2.17(0.76)	2.22(0.85)	2.24(0.89)
200	3.57(2.13)	4.52(2.54)	5.18(2.72)	2.97(1.21)	2.98(1.43)	3.01(1.53)	3.06(1.27)	3.17(1.48)	3.21(1.57)

	Cross validated values of δ		
$N \backslash p$	0.01	0.05	0.10
	$\mathbf{u}_{it}$ ∼ Gaussian		
30	1.27(0.27)	1.46(0.35)	1.61(0.36)
100	1.25(0.24)	1.43(0.31)	1.56(0.32)
200	1.23(0.22)	1.36(0.26)	1.49(0.27)
	$\mathbf{u}_{it}$ ∼ multivariate t-distr. with 8 dof		
30	1.45(0.38)	1.72(0.39)	1.87(0.35)
100	1.59(0.41)	1.76(0.40)	1.85(0.37)
200	1.68(0.44)	1.78(0.41)	1.85(0.39)

Note: The *MT* approach is implemented using $\delta = 1$, $\delta = 2$, and $\hat{\delta}$, computed using cross-validation. Norm losses and estimates of δ, $\hat{\delta}$, are averages over 2,000 replications. Simulation standard deviations are given in parentheses.

$\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}$ perform better than $LW_{\hat{\Sigma}}$, for all sample sizes and irrespective of whether the observations are drawn as Gaussian or multivariate t. Finally, using $\hat{\delta}$ rather than $\delta = 2$ when implementing the *MT* method improves the precision of the estimated inverse covariance matrix across all experiments.

3.6. Support recovery statistics

Table 5 reports the true positive and false positive rates (TPR and FPR) for the support recovery of $\boldsymbol{\Sigma}$ using the multiple testing and thresholding estimators. In the comparison set we include three versions of the *MT* estimator ($\widetilde{\boldsymbol{\Sigma}}_{MT,1}$, $\widetilde{\boldsymbol{\Sigma}}_{MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}$), $\widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}$, $\widetilde{\boldsymbol{\Sigma}}_{CL,2}$, and $\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}$. Again we use 100 replications due to the use of cross-validation in the implementation of *MT*, BL and CL thresholding. We include the *MT* estimators for choices of the scaling factor where $\delta = 1$ and $\delta = 2$, computed at $p = 0.05$, to see if our theoretical result, namely that for consistent support recovery only the linear scaling factor, where $\delta = 1$, is needed, is borne out by the simulations. Further, we implement *MT* using $\hat{\delta}$ to verify that the support recovery results under $MT_{\hat{\delta}}$ correspond more closely to those under MT_1, in line with the findings of Theorem 3. For consistent support recovery we would like to see *FPR* values near zero and *TPR* values near unity. As can be seen from Table 5, the *FPR* values of all estimators are very close to zero, so any comparisons of different estimators must be based on the *TPR* values. Comparing the results for $\widetilde{\boldsymbol{\Sigma}}_{MT,1}$ and $\widetilde{\boldsymbol{\Sigma}}_{MT,2}$ we find that as predicted by the theory (Theorem 3 and Remark 13), *TPR* values of $\widetilde{\boldsymbol{\Sigma}}_{MT,1}$ are closer to unity as compared to the *TPR* values of $\widetilde{\boldsymbol{\Sigma}}_{MT,2}$. This is supported by the *TPR* values of $\widetilde{\boldsymbol{\Sigma}}_{MT}$, $\hat{\delta}$ as well. Similar results are obtained for the *MT* estimators for different choices of the p values. Table 6 provides results for $p = \{0.01, 0.05, 0.10\}$, and for $\delta = \left\{1, 2, \hat{\delta}\right\}$ using 2000 replications. In this table it is further evident that, in line with the conclusions of Section 3.3, both the *TPR* and the *FPR* statistics are relatively robust to the choice of the p values irrespective of the scaling factor, or whether the observations are drawn from a Gaussian or a multivariate t distribution. This is especially true under design B, since for this specification we explicitly control for the number of non-zero elements in $\boldsymbol{\Sigma}$, that ensures the conditions of Theorem 3 are met.

Table 2
Spectral and Frobenius norm losses for different regularised covariance matrix estimators ($T = 100$) — Monte Carlo design A.

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	$\mathbf{u}_{it} \sim$ **Gaussian**					
	Error matrices $(\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}})$					
MT_2	1.85(0.53)	3.38(0.40)	2.83(0.50)	7.29(0.42)	3.45(0.43)	11.17(0.38)
$MT_{\hat{\delta}}$	1.75(0.55)	3.21(0.49)	2.44(0.50)	6.48(0.50)	2.95(0.45)	9.65(0.48)
$BL_{\hat{C}}$	5.30(2.16)	7.61(1.23)	8.74(0.06)	16.90(0.10)	8.94(0.04)	24.26(0.13)
CL_2	1.87(0.55)	3.39(0.44)	2.99(0.49)	7.57(0.44)	3.79(0.47)	11.88(0.42)
$CL_{\hat{C}}$	1.82(0.58)	3.33(0.56)	2.54(0.50)	6.82(0.51)	3.02(0.46)	10.22(0.59)
$S\text{-}MT_2$	3.36(0.78)	4.45(0.63)	5.83(0.34)	10.95(0.47)	6.47(0.21)	16.64(0.35)
$S\text{-}MT_{\hat{\delta}}$	2.67(0.81)	3.85(0.65)	5.08(0.40)	9.70(0.51)	5.79(0.27)	14.91(0.46)
$BL_{\hat{C}*}$	7.09(0.10)	8.62(0.09)	8.74(0.06)	16.90(0.10)	8.94(0.04)	24.25(0.10)
$CL_{\hat{C}*}$	7.05(0.16)	8.58(0.12)	8.71(0.07)	16.85(0.11)	8.94(0.04)	24.23(0.09)
$LW_{\hat{\Sigma}}$	2.99(0.47)	6.49(0.29)	5.20(0.34)	16.70(0.19)	6.28(0.20)	26.84(0.14)
	$\mathbf{u}_{it} \sim$ **multivariate** ***t*****-distributed with 8 degrees of freedom**					
	Error matrices $(\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}})$					
MT_2	2.17(0.72)	4.02(0.88)	3.44(0.98)	8.52(1.17)	4.00(0.83)	12.79(1.66)
$MT_{\hat{\delta}}$	2.27(0.88)	4.20(1.11)	3.59(1.39)	8.76(1.65)	4.32(1.53)	13.28(2.83)
$BL_{\hat{C}}$	6.90(0.82)	8.75(0.55)	8.74(0.10)	17.26(0.30)	9.00(0.42)	24.93(1.02)
CL_2	2.55(0.93)	4.53(1.00)	4.63(1.11)	10.35(1.48)	5.92(0.81)	16.43(1.74)
$CL_{\hat{C}}$	2.27(0.76)	4.24(0.94)	3.85(1.51)	9.44(2.33)	5.04(2.04)	15.65(4.71)
$S\text{-}MT_2$	3.28(0.80)	4.76(0.77)	5.84(0.45)	11.47(0.62)	6.48(0.32)	17.27(0.71)
$S\text{-}MT_{\hat{\delta}}$	2.86(0.92)	4.51(0.97)	5.30(0.52)	10.76(0.77)	6.00(0.39)	16.36(1.04)
$BL_{\hat{C}*}$	7.06(0.13)	8.84(0.30)	8.74(0.10)	17.25(0.31)	8.95(0.08)	24.84(0.55)
$CL_{\hat{C}*}$	7.01(0.16)	8.77(0.30)	8.73(0.11)	17.23(0.29)	8.94(0.08)	24.77(0.53)
$LW_{\hat{\Sigma}}$	3.35(0.51)	7.35(0.50)	5.67(0.46)	18.04(0.45)	6.60(0.43)	28.18(0.53)

Note: Norm losses are averages over 100 replications. Simulation standard deviations are given in parentheses. $\mathring{\boldsymbol{\Sigma}} = \{\widetilde{\boldsymbol{\Sigma}}_{MT,2}, \widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{CL,2}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,2}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,\hat{\delta}}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}*}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}*}, \hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. $\widetilde{\boldsymbol{\Sigma}}_{MT,2}$, $\widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}$, $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,\hat{\delta}}$ are computed using $p = 0.05$. (MT_2, $S\text{-}MT_2$) and ($MT_{\hat{\delta}}$, $S\text{-}MT_{\hat{\delta}}$) are thresholding based on multiple testing with critical value $\Phi^{-1}\left(1 - \frac{p}{2f(N)}\right)$, where $f(N) = N^2$ and $f(N) = N^{\hat{\delta}}$, respectively, with $\hat{\delta}$ estimated by cross-validation. *BL* is Bickel and Levina universal thresholding, *CL* is Cai and Liu adaptive thresholding, $\widetilde{\boldsymbol{\Sigma}}_{MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}$ are based on MT_2 and $MT_{\hat{\delta}}$. $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT,\hat{\delta}}$ apply supplementary shrinkage to $\widetilde{\boldsymbol{\Sigma}}_{MT,2}$ and $\widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}$, $\widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}$ and $\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}$ are based on $\hat{C}$ which is obtained by cross-validation, $\widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}*}$ and $\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}*}$ employ the further adjustment to the cross-validation coefficient, $\hat{C}^*$, proposed by Fan et al. (2013), $\widetilde{\boldsymbol{\Sigma}}_{CL,2}$ is CL's estimator with $C = 2$ (the theoretical value of C). $\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}$ is Ledoit and Wolf's shrinkage estimator applied to the sample covariance matrix.

Turning to a comparison with other estimators in Table 5, we find that the *MT* and *CL* estimators perform substantially better than the *BL* estimator. Further, allowing for dependence in the errors causes the support recovery performance of $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$ to deteriorate noticeably while MT_1, MT_2 and $MT_{\hat{\delta}}$ remain remarkably stable. Finally, again note that *TPR* values are higher for design B. Overall, the estimators $\widetilde{\boldsymbol{\Sigma}}_{MT,1}$ or $\widetilde{\boldsymbol{\Sigma}}_{MT,\hat{\delta}}$ do best in recovering the support of $\boldsymbol{\Sigma}$ as compared to other estimators, although the results of *CL* and *MT* for support recovery can be very close, which is in line with the comparative analysis carried out in terms of the relative norm losses of these estimators.

3.7. Cross-validation of δ

We calibrate δ, the parameter of the critical value function, $c_p(N)$, in the *MT* approach, by following closely the cross-validation procedure implemented in BL and CL. Importantly, Bickel and Levina (2008a) show theoretically the validity of this approach for the 'sample splitting', '2-fold cross-validation' and more general 'V-fold cross-validation' procedures.

More precisely, we perform a grid search for the choice of δ over the range: $\delta = \{c : \delta_{\min} \le c \le \delta_{\max}\}$. We set $\delta_{\min} = 1.0$ and $\delta_{\max} = 2.5$ and impose either fixed increments of 0.1 or N-dependent increments of $1/N$.[6] At each point of the range, c, we generate $x_{it}, i = 1, 2, \ldots, N, t = 1, 2, \ldots, T$ and select the $N \times 1$ column vectors $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$, $t = 1, 2, \ldots, T$ which we randomly reshuffle over the t-dimension. This yields a new set of $N \times 1$ column vectors $\boldsymbol{x}_t^{(s)} = \left(x_{1t}^{(s)}, x_{2t}^{(s)}, \ldots, x_{Nt}^{(s)}\right)'$ for the first shuffle $s = 1$. We repeat this reshuffling S times in total where we set $S = 50$. We consider this to be sufficiently large (FLM suggested $S = 20$ while BL recommended $S = 100$ — see also Fang et al. (2016)). For each shuffle $s = 1, 2, \ldots, S$, we divide $\boldsymbol{x}^{(s)} = \left(\boldsymbol{x}_1^{(s)}, \boldsymbol{x}_2^{(s)}, \ldots, \boldsymbol{x}_T^{(s)}\right)$ into two subsamples of size $N \times T_1$ and $N \times T_2$, where $T_2 = T - T_1$. The theoretically 'justified' split suggested in BL is given by $T_1 = T\left(1 - \frac{1}{\ln(T)}\right)$ and $T_2 = \frac{T}{\ln(T)}$. In our simulation study we set $T_1 = \frac{2T}{3}$ and $T_2 = \frac{T}{3}$. Let $\hat{\boldsymbol{\Sigma}}_1^{(s)} = \left(\hat{\sigma}_{1,ij}^{(s)}\right)$ with elements $\hat{\sigma}_{1,ij}^{(s)} = T_1^{-1}\sum_{t=1}^{T_1} x_{it}^{(s)} x_{jt}^{(s)}$ and $\hat{\boldsymbol{\Sigma}}_2^{(s)} = \left(\hat{\sigma}_{2,ij}^{(s)}\right)$ with

6 The sample size dependent alternative provides slight improvement in estimation precision for δ, but is computationally more expensive as N rises.

Table 3
Spectral and Frobenius norm losses for different regularised covariance matrix estimators ($T = 100$) — Monte Carlo design B.

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	$\mathbf{u}_{it} \sim$ **Gaussian**					
	Error matrices $(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}})$					
MT_2	0.49(0.18)	0.89(0.19)	0.87(0.37)	1.63(0.28)	0.73(0.24)	2.15(0.19)
$MT_{\hat\delta}$	0.48(0.14)	0.89(0.16)	0.79(0.31)	1.57(0.23)	0.67(0.18)	2.15(0.17)
$BL_{\hat C}$	0.91(0.50)	1.35(0.43)	1.40(0.95)	2.25(0.78)	2.53(0.55)	3.49(0.32)
CL_2	0.49(0.17)	0.90(0.18)	1.00(0.48)	1.77(0.44)	0.90(0.37)	2.30(0.30)
$CL_{\hat C}$	0.49(0.15)	0.92(0.17)	0.83(0.31)	1.71(0.28)	1.14(0.83)	2.54(0.58)
$S\text{-}MT_2$	0.68(0.27)	1.08(0.21)	1.53(0.53)	2.16(0.38)	1.23(0.41)	2.44(0.26)
$S\text{-}MT_{\hat\delta}$	0.66(0.23)	1.07(0.18)	1.45(0.44)	2.08(0.29)	1.12(0.30)	2.38(0.19)
$BL_{\hat C *}$	1.19(0.46)	1.63(0.40)	3.32(0.20)	3.90(0.14)	2.73(0.11)	3.61(0.08)
$CL_{\hat C *}$	1.08(0.46)	1.53(0.46)	3.34(0.15)	3.92(0.06)	2.73(0.10)	3.61(0.08)
$LW_{\hat\Sigma}$	1.05(0.13)	2.07(0.10)	2.95(0.26)	4.47(0.09)	2.46(0.06)	6.01(0.03)
	$\mathbf{u}_{it} \sim$ **multivariate** t**-distributed with 8 degrees of freedom**					
	Error matrices $(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}})$					
MT_2	0.64(0.24)	1.12(0.24)	1.05(0.45)	2.13(0.49)	1.29(2.32)	3.15(2.66)
$MT_{\hat\delta}$	0.66(0.25)	1.15(0.26)	1.03(0.42)	2.17(0.53)	1.30(1.90)	3.29(2.22)
$BL_{\hat C}$	1.36(0.40)	1.84(0.35)	2.70(0.94)	3.58(0.74)	2.70(0.29)	4.08(0.67)
CL_2	0.71(0.29)	1.21(0.30)	1.69(0.70)	2.73(0.70)	1.62(0.57)	3.31(0.65)
$CL_{\hat C}$	0.80(0.39)	1.33(0.39)	2.03(1.08)	3.07(0.90)	2.19(0.78)	3.72(0.62)
$S\text{-}MT_2$	0.69(0.26)	1.18(0.23)	1.41(0.57)	2.36(0.47)	1.32(0.79)	3.02(0.87)
$S\text{-}MT_{\hat\delta}$	0.69(0.25)	1.19(0.22)	1.36(0.49)	2.34(0.42)	1.30(0.78)	3.10(0.87)
$BL_{\hat C *}$	1.49(0.26)	1.98(0.21)	3.33(0.24)	4.07(0.18)	2.77(0.37)	4.04(0.56)
$CL_{\hat C *}$	1.26(0.40)	1.79(0.40)	3.35(0.17)	4.08(0.14)	2.73(0.14)	4.01(0.42)
$LW_{\hat\Sigma}$	1.13(0.15)	2.25(0.11)	3.14(0.21)	4.68(0.11)	2.52(0.08)	6.18(0.13)

See the note to Table 2.

Table 4
Spectral and Frobenius norm losses for the inverses of different regularised covariance matrix estimators ($T = 100$) — Monte Carlo design B.

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices $(\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1})$					
	$\mathbf{u}_{it} \sim$ **Gaussian**					
$S\text{-}MT_2$	4.44(1.23)	2.66(0.32)	15.81(2.63)	5.90(0.45)	14.24(2.37)	5.50(0.38)
$S\text{-}MT_{\hat\delta}$	4.36(1.22)	2.64(0.31)	15.25(2.78)	5.80(0.48)	13.36(2.47)	5.39(0.37)
$BL_{\hat C *}$	3.8×10^3 (2.4×10^4)	19.56(58.88)	1.2×10^3 (1.1×10^4)	12.16(33.25)	41.07(143.74)	7.66(3.17)
$CL_{\hat C *}$	1.9×10^3 (1.7×10^4)	10.92(42.39)	51.99(241.39)	8.16(4.23)	28.45(24.37)	7.35(1.11)
$LW_{\hat\Sigma}$	11.03(0.58)	4.26(0.09)	31.04(0.64)	8.62(0.06)	31.81(0.21)	9.40(0.05)
	$\mathbf{u}_{it} \sim$ **multivariate** t**-distributed with 8 degrees of freedom**					
$S\text{-}MT_2$	3.45(1.61)	2.44(0.39)	12.78(3.13)	5.55(0.55)	11.57(4.17)	5.58(0.66)
$S\text{-}MT_{\hat\delta}$	3.43(1.63)	2.45(0.40)	12.37(3.27)	5.51(0.59)	11.28(3.97)	5.65(0.67)
$BL_{\hat C *}$	157.26(1.0×10^3)	6.11(11.28)	349.35(3.1×10^3)	9.80(17.03)	28.58(22.06)	7.77(1.04)
$CL_{\hat C *}$	85.82(546.85)	5.53(7.84)	517.27(4.8×10^3)	10.07(21.25)	25.61(3.55)	7.54(0.50)
$LW_{\hat\Sigma}$	12.08(1.19)	4.48(0.20)	31.78(1.32)	8.74(0.23)	32.06(1.00)	9.50(0.33)

Note: $\hat{\boldsymbol{\Sigma}}^{-1} = \{\tilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,2}, \tilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT,\hat\delta}, \tilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat C *}, \tilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat C *}, \hat{\boldsymbol{\Sigma}}^{-1}_{LW_{\hat\Sigma}}\}$. See also the note to Table 2.

elements $\hat\sigma^{(s)}_{2,ij} = T_2^{-1}\sum_{t=T_1+1}^{T} x^{(s)}_{it}x^{(s)}_{jt}$, $i, j = 1, 2, \ldots, N$ denote the sample covariance matrices generated using T_1 and T_2 respectively, for each shuffle s. The corresponding sample correlation matrices are given by $\hat{\boldsymbol{R}}^{(s)}_1 = \left[\hat{\boldsymbol{D}}^{(s)}_1\right]^{-1/2} \hat{\boldsymbol{\Sigma}}^{(s)}_1 \left[\hat{\boldsymbol{D}}^{(s)}_1\right]^{-1/2}$ and $\hat{\boldsymbol{R}}^{(s)}_2 = \left[\hat{\boldsymbol{D}}^{(s)}_2\right]^{-1/2} \hat{\boldsymbol{\Sigma}}^{(s)}_2 \left[\hat{\boldsymbol{D}}^{(s)}_2\right]^{-1/2}$ respectively, where $\hat{\boldsymbol{D}}^{(s)}_i = \mathrm{diag}(\hat\sigma^{(s)}_{i,11}, \hat\sigma^{(s)}_{i,22}, \ldots, \hat\sigma^{(s)}_{i,NN})$, $i = 1,2$. We regularise $\hat{\boldsymbol{R}}^{(s)}_1$ using the MT method in (5) and compute the following expression,

$$\hat{J}(c) = \frac{1}{S}\sum_{s=1}^{S}\left\|\tilde{\boldsymbol{R}}^{(s)}_1(c) - \hat{\boldsymbol{R}}^{(s)}_2\right\|_F^2, \tag{38}$$

for each c and

$$\hat\delta = \arg\inf_{\delta_{\min}\le c\le\delta_{\max}} \hat{J}(c). \tag{39}$$

Table 5
Support recovery statistics for different multiple testing and thresholding estimators — $T = 100$.

Monte Carlo design A								Monte Carlo design B							
N		MT_1	MT_2	$MT_{\hat{\delta}}$	$BL_{\hat{C}}$	CL_2	$CL_{\hat{C}}$	N		MT_1	MT_2	$MT_{\hat{\delta}}$	$BL_{\hat{C}}$	CL_2	$CL_{\hat{C}}$
		$\mathbf{u}_{it}$ ~ **Gaussian**													
30	TPR	0.80	0.71	0.79	0.29	0.72	0.78	30	TPR	1.00	0.98	1.00	0.64	0.98	1.00
	FPR	0.00	0.00	0.00	0.04	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
100	TPR	0.69	0.57	0.69	0.00	0.56	0.68	100	TPR	1.00	0.98	1.00	0.80	0.94	0.99
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
200	TPR	0.66	0.53	0.66	0.00	0.50	0.65	200	TPR	1.00	0.96	0.99	0.11	0.88	0.78
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
		$\mathbf{u}_{it}$ ~ **multivariate t-distributed with 8 degrees of freedom**													
30	TPR	0.80	0.72	0.79	0.03	0.62	0.74	30	TPR	1.00	0.98	0.99	0.26	0.89	0.82
	FPR	0.01	0.00	0.00	0.00	0.00	0.00		FPR	0.01	0.00	0.00	0.00	0.00	0.00
100	TPR	0.69	0.58	0.67	0.00	0.43	0.57	100	TPR	1.00	0.97	0.98	0.27	0.70	0.57
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
200	TPR	0.66	0.53	0.64	0.00	0.35	0.47	200	TPR	0.99	0.93	0.95	0.05	0.57	0.30
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00

Note: TPR is the true positive rate and FPR is the false positive rate defined by (30) and (31), respectively. MT estimators are computed with $p = 0.05$. For a description of other estimators see the note to Table 2. The TPR and FPR numbers are averages over 100 replications.

Table 6
Support recovery statistics for the multiple testing estimator computed with $p = \{0.01, 0.05, 0.10\}$ — $T = 100$.

Monte Carlo design A											Monte Carlo design B										
		$p = 0.01$			$p = 0.05$			$p = 0.10$					$p = 0.01$			$p = 0.05$			$p = 0.10$		
N		MT_1	MT_2	$MT_{\hat{\delta}}$	MT_1	MT_2	$MT_{\hat{\delta}}$	MT_1	MT_2	$MT_{\hat{\delta}}$	N		MT_1	MT_2	$MT_{\hat{\delta}}$	MT_1	MT_2	$MT_{\hat{\delta}}$	MT_1	MT_2	$MT_{\hat{\delta}}$
		$\mathbf{u}_{it}$ ~ **Gaussian**																			
30	TPR	0.75	0.67	0.75	0.80	0.71	0.79	0.81	0.73	0.80	30	TPR	1.00	0.97	1.00	1.00	0.99	1.00	1.00	0.99	1.00
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
100	TPR	0.65	0.54	0.64	0.69	0.57	0.69	0.71	0.59	0.70	100	TPR	1.00	0.97	0.99	1.00	0.98	1.00	1.00	0.99	1.00
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
200	TPR	0.62	0.49	0.61	0.66	0.53	0.66	0.68	0.54	0.67	200	TPR	0.99	0.92	0.99	1.00	0.96	0.99	1.00	0.97	0.99
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		$\mathbf{u}_{it}$ ~ **multivariate t-distributed with 8 degrees of freedom**																			
30	TPR	0.76	0.68	0.74	0.80	0.81	0.78	0.81	0.73	0.79	30	TPR	0.99	0.96	0.98	1.00	0.98	0.99	1.00	0.99	0.99
	FPR	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01		FPR	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00
100	TPR	0.65	0.54	0.64	0.71	0.71	0.67	0.71	0.59	0.68	100	TPR	0.99	0.96	0.97	1.00	0.97	0.98	1.00	0.98	0.98
	FPR	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
200	TPR	0.62	0.50	0.60	0.68	0.68	0.63	0.68	0.55	0.65	200	TPR	0.99	0.91	0.93	0.99	0.94	0.95	1.00	0.96	0.96
	FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: TPR is the true positive rate and FPR is the false positive rate defined by (30) and (31), respectively. MT estimators are computed with $p = 0.05$. For a description of other estimators see the note to Table 2. The TPR and FPR numbers are averages over 2,000 replications.

The final estimator of the correlation matrix is then given by $\widetilde{\boldsymbol{R}}_{\hat{\delta}}$ and the associated covariance matrix estimator, $\widetilde{\boldsymbol{\Sigma}}_{\hat{\delta}}$, is computed as in (7).

3.8. Computational demands of the different thresholding methods

Table 7 reports the relative execution times in seconds of the different thresholding methods studied. All times are relative to the time it takes to carry out the computations for the MT_2 estimator. The computational times shown for the methods that use a calibrated threshold parameter (i.e. $MT_{\hat{\delta}}$, $BL_{\hat{C}}$ and $CL_{\hat{C}}$) assume a sample-dependent grid in their respective CV procedures. It took 0.010, 0.013, and 0.014 seconds to apply the MT method in Matlab to a sample of $N = \{30, 100, 200\}$, respectively, and $T = 100$ observations using a desktop PC. The execution times of MT_1 and MT_2 are very similar and differ only slightly across the experiments with different p-values. In contrast, the $BL_{\hat{C}}$ and $CL_{\hat{C}}$ thresholding approaches are computationally much more demanding. Their computations took between about 12 and 485257 times (depending on N) longer than the MT_2 approach, for the same sample sizes and computer hardware. The $BL_{\hat{C}}$ method was less demanding than the $CL_{\hat{C}}$ method — it took between about 12 and 584 times longer than the MT_2 approach. Even CL_2, which does not require estimation of the threshold parameter, took up to 19 times longer than the MT_2 approach. Thus, compared with other thresholding methods, MT_1 and MT_2 procedures have a clear computational advantage over the CL and BL procedures. This is not a surprising outcome, considering that MT_1 and MT_2 do not involve cross validation. But we find similar computational advantages for the MT procedure when we compare its cross-validated version, $MT_{\hat{\delta}}$, with $CL_{\hat{C}}$. The execution times of $MT_{\hat{\delta}}$ were between 1278 and 482038 faster than $CL_{\hat{C}}$. But when compared to $BL_{\hat{C}}$, we find that $BL_{\hat{C}}$ is somewhere between 24 and 2634 faster to compute than $MT_{\hat{\delta}}$. However, when using a fixed point increment in the implementation of the $MT_{\hat{\delta}}$ procedure, the computational advantage of $BL_{\hat{C}}$ over $MT_{\hat{\delta}}$ disappears.

Table 7
Relative execution times in seconds of different thresholding methods.

	$T = 100$		
	$N = 30$	$N = 100$	$N = 200$
MT_2	1.000	1.000	1.000
MT_1	0.996	0.971	1.017
$MT_{\hat{\delta}}$	35.84	497.4	3219
$BL_{\hat{C}}$	11.53	106.3	584.8
CL_2	1.924	5.629	19.12
$CL_{\hat{C}}$	1314	63481	485257

Note: All times are relative to the MT_2 estimator. See Table 2 for a note on the thresholding methods.

4. Concluding remarks

This paper considers regularisation of large covariance matrices particularly when the cross section dimension N of the data under consideration exceeds the time dimension T. In this case the sample covariance matrix, $\hat{\boldsymbol{\Sigma}}$, becomes ill-conditioned and is not a satisfactory estimator of the population covariance.

A regularisation estimator is proposed which makes use of insights from the multiple testing literature to obtain threshold values for sample correlation coefficients. The proposed MT estimator of the correlation coefficient (ρ_{ij}) is set to zero when the sample correlation, in absolute value, is below the threshold, otherwise the MT estimator is set to the sample correlation coefficient. It is shown that the resultant estimator has a convergence rate of the order of $m_N c_p(N)/\sqrt{T}$ under the spectral norm, and $\sqrt{m_N N/T}$ under the Frobenius norm, where N is the number of units each observed T times, m_N measures the degree of sparsity of the population correlation matrix, and $c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2N^{\delta}}\right)$, where $\Phi^{-1}(.)$ is the inverse of the cumulative distribution of a standard normal variate, and p is the nominal size of the test. $c_p(N)$ directly corresponds to $\sqrt{\ln(N)}$. Our analysis allows for non-Gaussian observations and provides guidance as to the choice of critical value function for thresholding in terms of the degree to which underlying observations are dependent even if $\rho_{ij} = 0$. The choice of δ depends on the degree of non-Gaussianity of the underlying observations and yields spectral norm results that are similar to the rates obtained in the literature. But for the Frobenius norm we obtain better rates than those established in the literature.

The numerical properties of the proposed estimator are investigated using Monte Carlo simulations. It is shown that the *MT* estimator performs well, and generally better than the other estimators proposed in the literature. The simulations also show that in terms of spectral and Frobenius norm losses, the *MT* estimator is reasonably robust to the choice of p in the threshold criterion, $\left|\hat{\rho}_{ij}\right| > T^{-1/2}\Phi^{-1}\left(1 - \frac{p}{2f(N)}\right)$, where $f(N) = c_{\delta}N^{\delta}$, with c_{δ} and δ being finite positive constants, particularly when setting $\delta = 2$. For support recovery, better results are obtained if $\delta = 1$.

Acknowledgements

The authors are grateful to Jianqing Fan (Co-Editor), an Associate Editor and three anonymous reviewers for valuable comments and suggestions, as well as Elizaveta Levina and Martina Mincheva for helpful email correspondence with regard to implementation of their approaches. They would also like to thank Alex Chudik, George Kapetanios, Yuan Liao, Ron Smith and Michael Wolf for useful comments. Financial support under ESRC, United Kingdom Grant ES/I031626/1 is gratefully acknowledged by the authors.

Appendix A. Mathematical proofs of theorems for the MT estimator

The lemmas referred to in this Appendix are stated and proved in a supplement which is available online.

Proof of Proposition 1. The results for $E\left(\hat{\rho}_{ij,T}\right)$ and $Var\left(\hat{\rho}_{ij,T}\right)$ are established in Gayen (1951) using a bivariate Edgeworth expansion approach. This confirms earlier findings obtained by Tschuprow (1925) (English Translation, 1939) who shows that results (9) and (10) hold for any law of dependence between x_{it} and x_{jt}. See, in particular, p. 228 and Eqs. (53) and (54) in Gayen (1951). To see that these results hold uniformly in the i and j ($i \neq j$) pairs, first note we have that $\sup_{ij}\left|\rho_{ij}\right| < 1$. Using (10) and (12) $\lim_{T\to\infty}\left[TVar\left(\hat{\rho}_{ij,T}\right)\right] = K_v(\theta_{ij})$. The uniform boundedness of $\left|K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})\right|$ and $K_v(\boldsymbol{\theta}_{ij})$ follows directly from Assumption 2 that the sixth-order moment of y_{it} is uniformly bounded and application of Holder's and the Cauchy–Schwarz inequalities. Application of these inequalities establishes the uniform boundedness of the moments $E(y_{it}^3 y_{jt})$ and $E\left(y_{it}^2 y_{jt}^2\right)$ as given below:

$$\begin{aligned}
\sup_{ij,t}\left|E(y_{it}^2 y_{jt}^2)\right| &\leq \sup_{ij,t} E(\left|y_{it}^2 y_{jt}^2\right|) \leq \sup_{ij,t}\left\{\left[E(|y_{it}|^4)\right]^{1/2}\left[E\left(\left|y_{jt}\right|^4\right)\right]^{1/2}\right\} \\
&\leq \sup_{i,t}\left[E(|y_{it}|^4)\right]^{1/2}\sup_{j,t}\left[E\left(\left|y_{jt}\right|^4\right)\right]^{1/2} < K
\end{aligned}$$

and

$$\sup_{ij,t}\left|E(y_{it}y_{jt}^3)\right| \le \sup_{ij,t} E(\left|y_{it}y_{jt}^3\right|) \le \sup_{ij,t}\left\{\left[E(|y_{it}|^4)\right]^{1/4}\left[E\left(\left|y_{jt}^3\right|^{4/3}\right)\right]^{3/4}\right\}$$
$$= \sup_{ij,t}\left\{\left[E(|y_{it}|^4)\right]^{1/4}\left[E\left(\left|y_{jt}\right|^{4}\right)\right]^{3/4}\right\} = \sup_{i,t} E(|y_{it}|^4) < K.$$

The remaining terms included in $O\left(T^{-2}\right)$ in (9) and (10) as can be seen from Gayen (1951) are also a function of ρ_{ij} and $\kappa_{ij}(.,.)$ up to order $\kappa_{ij}(4,0)$. Hence, the results of Gayen (1951) hold uniformly across all i and j pair of correlations.

Consider now the case where y_{it} for all i are Gaussian. Then $E(y_{it}^4) = 3$, and for all $i \neq j$ we have

$$y_{it} = \rho_{ij}y_{jt} + \eta_{jt},$$

where $E(\eta_{jt}) = 0$, $Var(\eta_{jt}) = 1-\rho_{ij}^2$, and η_{jt} and y_{jt} are independently distributed. Hence,

$$E(y_{jt}^3y_{it}) = E[y_{jt}^3(\rho_{ij}y_{jt} + \eta_{jt})] = \rho_{ij}E(y_{jt}^4) = 3\rho_{ij}.$$

$$E(y_{it}^2y_{jt}^2) = E[y_{jt}^2(\rho_{ij}^2y_{jt}^2 + \eta_{jt}^2 + 2\rho_{ij}y_{jt}\eta_{jt})] = 3\rho_{ij}^2 + (1-\rho_{ij}^2) = 1+2\rho_{ij}^2.$$

Using the above results it now follows that

$$\kappa_{ij}(4,0) = \kappa_{ij}(0,4) = 0,\ \ \kappa_{ij}(3,1) = \kappa_{ij}(1,3) = 0,$$

which in turn establishes, when used in (11) and (12), that $K_{\mathrm{m}}(\boldsymbol{\theta}_{ij}) = -\frac{1}{2}\rho_{ij}(1-\rho_{ij}^2)$ and $K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) = (1-\rho_{ij}^2)^2$. ∎

Proof of Proposition 2. Under Assumption 2, for a given i and j, set $\boldsymbol{\xi}_t = (y_{it}, y_{jt}, y_{it}^2, y_{jt}^2, y_{it}y_{jt})' = (\xi_{1t}, \xi_{2t}, \ldots, \xi_{5t})'$, where $y_{it} = (x_{it}-\mu_i)/\sqrt{\sigma_{ii}}$. To simplify the notation we drop the subscripts i, j. Define

$$\bar{\boldsymbol{\xi}}_T = T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_t = \left(\bar{\xi}_{1T}, \bar{\xi}_{2T}, \ldots, \bar{\xi}_{5T}\right)',$$

and note that by Assumption 2, $\boldsymbol{\xi}_t$, for $t = 1, 2, \ldots, T$, are random draws from a common distribution with non-zero density, the elements of $\boldsymbol{\xi}_t$ are continuously differentiable functions of $\mathbf{y}_t = (y_{it}, y_{jt})'$. $\hat{\rho}_{ij,T}$, the sample correlation coefficient of x_{it} and x_{jt}, can be written as

$$\hat{\rho}_{ij,T} = H\left(\bar{\boldsymbol{\xi}}_T\right) = \frac{\bar{\xi}_{5T} - \bar{\xi}_{1T}\bar{\xi}_{2T}}{\left(\bar{\xi}_{3T} - \bar{\xi}_{1T}^2\right)^{1/2}\left(\bar{\xi}_{4T} - \bar{\xi}_{2T}^2\right)^{1/2}},$$

where $\bar{\xi}_{3T} > \bar{\xi}_{1T}^2$, and $\bar{\xi}_{4T} > \bar{\xi}_{2T}^2$. See also Bhattacharya and Ghosh (1978) — p. 434. It is also easily seen that $\boldsymbol{\mu}_{\xi} = E(\bar{\boldsymbol{\xi}}_T) = (0,0,1,1,\rho_{ij})'$, and $H\left(\boldsymbol{\mu}_{\xi}\right) = \rho_{ij}$, and hence $\sqrt{T}\left[H\left(\bar{\boldsymbol{\xi}}_T\right) - H\left(\boldsymbol{\mu}_{\xi}\right)\right] = \sqrt{T}\left(\hat{\rho}_{ij,T} - \rho_{ij}\right)$, where $H(\boldsymbol{\xi})$ is continuous and differentiable in $\boldsymbol{\xi}$, and all derivatives of $H(\boldsymbol{\xi})$ are continuous in a neighbourhood of $\boldsymbol{\mu}_{\xi}$; $1, \xi_{1t}, \xi_{2t}, \ldots, \xi_{5t}$ are linearly independent, and $E\left|\xi_{kt}\right|^s < \infty$, for $k = 1, 2, \ldots, 5$, for some positive integer $s \geq 3$. Hence, Theorem 2 of Bhattacharya and Ghosh (1978) can be applied to $\hat{\rho}_{ij,T}$, which establishes the validity of the Edgeworth expansion, (16). To prove (17) using (16) we first note that (for some $a_T > 0$)

$$\begin{aligned}
\Pr\left(z_{ij,T} > a_T\right) &= 1 - \Pr\left(z_{ij,T} \le a_T\right) \\
&= 1 - \Phi(a_T) - \sum_{r=1}^{s-2} T^{-r/2} g_r\left(a_T\right)\phi(a_T) + O\left[T^{-(s-1)/2}\right] \\
&= \Phi(-a_T) - (2\pi)^{-1/2}\exp\left(-\frac{a_T^2}{2}\right)\sum_{r=1}^{s-2} T^{-r/2} g_r\left(a_T\right) + O\left[T^{-(s-1)/2}\right],
\end{aligned}$$

and by the inequality (A.1) (in the online supplement), we have

$$\Pr\left(z_{ij,T} > a_T\right) \le \frac{1}{2}\exp\left(-\frac{a_T^2}{2}\right) + (2\pi)^{-1/2}\exp\left(-\frac{a_T^2}{2}\right)\sum_{r=1}^{s-2} T^{-r/2}\left|g_r\left(a_T\right)\right| + O\left[T^{-(s-1)/2}\right]. \tag{40}$$

But $g_r\left(a_T\right)$ is a polynomial of degree $3r-1$ in a_T, which is odd for even r, and even for odd r. For $r = 1$ and $r = 2$ we have

$$|g_1(x)| \le |g_{11}| + |g_{12}|\,|x|^2, \text{ and } |g_2(x)| \le |g_{21}|\,|x| + |g_{22}|\,|x|^3 + |g_{23}|\,|x|^5,$$

where g_{ij} are fixed coefficients that depend on the cumulants of $\boldsymbol{\xi}$. Result (17) now follows from (40) by separating the constant terms of $g_r(a_T)$ from the powers of a_T. Similarly, using (16) for $a_T > 0$ we have

$$\Pr\left(z_{ij,T} \le -a_T\right) = \Phi(-a_T) - \sum_{r=1}^{s-2} T^{-r/2} g_r\left(-a_T\right)\phi(a_T) + O\left[T^{-(s-1)/2}\right],$$

which upon using (A.1) yields

$$\Pr\left(z_{ij,T} \leq -a_T\right) \leq \frac{1}{2}\exp\left(-\frac{a_T^2}{2}\right) + (2\pi)^{-1/2}\exp\left(-\frac{a_T^2}{2}\right)\sum_{r=1}^{s-2} T^{-r/2}\left|g_r\left(-a_T\right)\right| + O\left[T^{-(s-1)/2}\right],$$

and result (18) follows. ■

Proof of Theorem 1. First we note that (see Horn and Johnson (1985) — p. 297)

$$\left\|\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right\| \leq \left\|\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right\|_\infty = \max_{1\leq i\leq N}\sum_j\left|\tilde{\rho}_{ij,T} - \rho_{ij}\right|,$$

where

$$\tilde{\rho}_{ij,T} = \hat{\rho}_{ij,T} I\left[\left|\hat{\rho}_{ij,T}\right| > \theta\left(N, T\right)\right],\ i = 1, 2, \ldots, N-1,\ j = i+1, \ldots, N,$$

$\theta\left(N, T\right) = T^{-1/2}c_p(N)$, and $c_p\left(N\right) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) > 0$. Note that $\theta\left(N, T\right) > 0$, and $\theta\left(N, T\right) = o(1)$ by assumption. Let $\rho_{ij}^* = \rho_{ij} I\left[\left|\rho_{ij}\right| > \theta\left(N, T\right)\right]$, and further note that

$$\max_i\sum_j\left|\tilde{\rho}_{ij,T} - \rho_{ij}\right| \leq \max_i\sum_j\left|\tilde{\rho}_{ij,T} - \rho_{ij}^*\right| + \max_i\sum_j\left|\rho_{ij}^* - \rho_{ij}\right|, \tag{41}$$

where to simplify the notation we will be using $\max_i$ for $\max_{1\leq i\leq N}$. We begin with the second term of (41) and write

$$\begin{aligned}\max_i\sum_j\left|\rho_{ij}^* - \rho_{ij}\right| &= \max_i\sum_j\left|\rho_{ij} I\left[\left|\rho_{ij}\right| > \theta\left(N, T\right)\right] - \rho_{ij}\right| = \\ &\leq \max_i\sum_j\left|\rho_{ij}\right| I\left[\left|\rho_{ij}\right| \leq \theta\left(N, T\right)\right].\end{aligned}$$

But $\left|\rho_{ij}\right| I\left[\left|\rho_{ij}\right| \leq \theta\left(N, T\right)\right] \leq \theta\left(N, T\right)$, and hence, in view of (3) we have

$$\max_i\sum_j\left|\rho_{ij}^* - \rho_{ij}\right| < K\theta\left(N, T\right)\max_i\left(\sum_{j, \rho_{ij}\neq 0} 1\right) = O\left[\theta\left(N, T\right) m_N\right]. \tag{42}$$

Consider now the first term of (41), and following Bickel and Levina (2008a) note that

$$\begin{aligned}\max_i\sum_j\left|\tilde{\rho}_{ij,T} - \rho_{ij}^*\right| &\leq \max_i\sum_j\left|\hat{\rho}_{ij,T}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta\left(N, T\right),\ \left|\rho_{ij}\right| < \theta\left(N, T\right)\right] \\ &\quad + \max_i\sum_j\left|\rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N, T\right),\ \left|\rho_{ij}\right| > \theta\left(N, T\right)\right] \\ &\quad + \max_i\sum_j\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta\left(N, T\right),\ \left|\rho_{ij}\right| > \theta\left(N, T\right)\right] \\ &= \mathcal{A} + \mathcal{B} + \mathcal{C}.\end{aligned} \tag{43}$$

Starting with $\mathcal{C}$ we have

$$\mathcal{C} = \max_i\sum_j\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta\left(N, T\right),\ \left|\rho_{ij}\right| > \theta\left(N, T\right)\right].$$

But $I\left[\left|\hat{\rho}_{ij,T}\right| > \theta\left(N, T\right),\ \left|\rho_{ij}\right| > \theta\left(N, T\right)\right] \leq I\left[\left|\rho_{ij}\right| > \theta\left(N, T\right)\right]$ and also $I\left(\left|\rho_{ij}\right| > \theta\left(N, T\right)\middle|\rho_{ij} = 0\right) = 0$. Hence

$$\begin{aligned}\mathcal{C} &\leq \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right|\max_i\sum_j I[\left|\rho_{ij}\right| > \theta\left(N, T\right)] \\ &\leq \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right|\max_i\sum_j I[\left|\rho_{ij}\right| > \theta\left(N, T\right)\left|\rho_{ij} \neq 0\right.] \\ &\quad + \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right|\max_i\sum_j I[\left|\rho_{ij}\right| > \theta\left(N, T\right)\left|\rho_{ij} = 0\right.] \\ &= \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right|\max_i\sum_j I[\left|\rho_{ij}\right| > \theta\left(N, T\right)\left|\rho_{ij} \neq 0\right.] \leq \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| m_N.\end{aligned} \tag{44}$$

However, using (A.3) of Lemma 3 in the online supplement and noting that $c_p^2(N) = T\theta^2(N,T)$, we have

$$\sup_{ij} \Pr\left[\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > \theta(N,T)\right] \le Ke^{-\frac{1}{2}\frac{c_p^2(N)}{K_v}} + O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{K_v}\right]^{\frac{3(s-2)-1}{2}} e^{-\frac{1}{2}\frac{c_p^2(N)}{K_v}}\right) + O\left(T^{-(s-1)/2}\right),$$

where $K_v = \sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$, and $K_v(\boldsymbol{\theta}_{ij})$ is defined by (12), with $K_v(\boldsymbol{\theta}_{ij}) > 0$. By the first-order Bonferroni inequality we have

$$\Pr\left[\max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > \theta(N,T)\right] \le KN^2e^{-\frac{1}{2}\frac{c_p^2(N)}{K_v}} + O\left(N^2T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{K_v}\right]^{\frac{3(s-2)-1}{2}} e^{-\frac{1}{2}\frac{c_p^2(N)}{K_v}}\right) + O\left(N^2T^{-(s-1)/2}\right),$$

which can also be written as (noting that the middle term of the above is dominated by the first term)

$$\Pr\left[\max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > \theta(N,T)\right] = O\left(N^2e^{-\frac{1}{2}\frac{c_p^2(N)}{K_v}}\right) + O\left(N^2T^{-(s-1)/2}\right). \tag{45}$$

Also, using result (b) of Lemma 2 in the online supplement, $e^{-\frac{1}{2}\frac{c_p^2(N)}{K_v}} = O\left(N^{-\delta/K_v}\right)$, and therefore

$$\Pr\left[\max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > \theta(N,T)\right] = O\left(N^{2-\delta/K_v}\right) + O\left(N^2T^{-(s-1)/2}\right). \tag{46}$$

Since $T = O(N^d)$, then it follows that

$$\max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| = O_p\left[\theta(N,T)\right], \tag{47}$$

so long as $\delta > 2K_v$ and $d > 4/(s-1)$. Using this result in (44) now yields

$$\mathcal{C} = O_p\left[m_N\theta(N,T)\right]. \tag{48}$$

For $\mathcal{A}$ we have

$$\begin{aligned}\mathcal{A} &= \max_i \sum_j \left|\hat{\rho}_{ij,T}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta(N,T), \ \left|\rho_{ij}\right| < \theta(N,T)\right] \\ &\le \max_i \sum_j \left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta(N,T), \ \left|\rho_{ij}\right| < \theta(N,T)\right] + \max_i \sum_j \left|\rho_{ij}\right| I\left[\left|\rho_{ij}\right| < \theta(N,T)\right] \\ &\le \mathcal{A}_1 + \mathcal{A}_2,\end{aligned}$$

where

$$\mathcal{A}_2 = \max_i \sum_j \left|\rho_{ij}\right| I\left[\left|\rho_{ij}\right| < \theta(N,T)\right] = O\left[\theta(N,T)\, m_N\right]. \tag{49}$$

Also, for any $\gamma \in (0,1)$, we have

$$\begin{aligned}\mathcal{A}_1 &= \max_i \sum_j \left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta(N,T), \ \left|\rho_{ij}\right| < \theta(N,T)\right] \\ &\le \max_i \sum_j \left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta(N,T), \ \left|\rho_{ij}\right| \le \gamma\theta(N,T)\right] \\ &\quad + \max_i \sum_j \left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| > \theta(N,T), \ \gamma\theta(N,T) < \left|\rho_{ij}\right| < \theta(N,T)\right] \\ &\le \mathcal{A}_{11} + \mathcal{A}_{12},\end{aligned}$$

where

$$\mathcal{A}_{11} = \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| \max_i \sum_j I\left[\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > (1-\gamma)\theta(N,T)\right], \tag{50}$$

$$\mathcal{A}_{12} = \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| \max_i \sum_j I\left[\left|\hat{\rho}_{ij,T}\right| > \theta(N,T), \ \gamma\theta(N,T) < \left|\rho_{ij}\right| < \theta(N,T)\right]. \tag{51}$$

But

$$\Pr\left[\max_i \sum_j I\left[\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > (1-\gamma)\theta\left(N,T\right)\right] > 0\right] = \Pr\left[\max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > (1-\gamma)\theta\left(N,T\right)\right], \tag{52}$$

and by (45) we have,

$$\Pr\left[\max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > (1-\gamma)\theta\left(N,T\right)\right] = O\left(N^2 e^{-\frac{1}{2}\frac{(1-\gamma)^2 c_p^2(N)}{K_v}}\right) + O\left(N^2 T^{-(s-1)/2}\right).$$

Using a similar line of reasoning as above there exist $\delta > 2K_v/(1-\gamma)^2$ and $d > 4/(s-1)$ such that

$$\Pr\left[\max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > (1-\gamma)\theta\left(N,T\right)\right] \to 0.$$

Using this result in conjunction with (52) and (47) in (50) it follows that $\mathcal{A}_{11} = O_p\left[m_N\theta\left(N,T\right)\right]$. For $\mathcal{A}_{12}$, we first note that

$$\begin{aligned}
&\max_i \sum_j I\left[\left|\hat{\rho}_{ij,T}\right| > \theta\left(N,T\right),\ \gamma\theta\left(N,T\right) < \left|\rho_{ij}\right| < \theta\left(N,T\right)\right] \\
&\le \max_i \sum_j I\left[\gamma\theta\left(N,T\right) < \left|\rho_{ij}\right| < \theta\left(N,T\right)\right] \\
&= \max_i \sum_j I\{\left[\gamma\theta\left(N,T\right) < \left|\rho_{ij}\right| < \theta\left(N,T\right)\right]\left|\rho_{ij} \ne 0\right\} \le m_N,
\end{aligned}$$

which gives $\mathcal{A}_{12} = O_p\left[m_N\theta\left(N,T\right)\right]$, and together with the result for $\mathcal{A}_{11}$ we have $\mathcal{A}_1 = O_p\left[\theta\left(N,T\right)m_N\right]$. Overall using (49) we obtain

$$\mathcal{A} = O_p[m_N\theta\left(N,T\right)]. \tag{53}$$

Finally, for $\mathcal{B}$ we have

$$\begin{aligned}
\mathcal{B} &= \max_i \sum_j \left|\rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right),\ \left|\rho_{ij}\right| > \theta\left(N,T\right)\right] \\
&\le \max_i \sum_j \left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right),\ \left|\rho_{ij}\right| > \theta\left(N,T\right)\right] \\
&\quad + \max_i \sum_j \left|\hat{\rho}_{ij,T}\right| I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right),\ \left|\rho_{ij}\right| > \theta\left(N,T\right)\right] \\
&= \mathcal{B}_1 + \mathcal{B}_2.
\end{aligned}$$

But as before

$$\begin{aligned}
\mathcal{B}_1 &= \max_i \sum_j \left|\hat{\rho}_{ij,T} - \rho_{ij}\right| I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right),\ \left|\rho_{ij}\right| > \theta\left(N,T\right)\right] \\
&\le \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| \max_i \sum_j I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right),\ \left|\rho_{ij}\right| > \theta\left(N,T\right)\right] \\
&= \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| \max_i \sum_j I\left\{\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right),\ \left|\rho_{ij}\right| > \theta\left(N,T\right)\right]\left|\rho_{ij} \ne 0\right.\right\} \\
&\le m_N \max_{ij}\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| = O_p\left[m_N\theta\left(N,T\right)\right],
\end{aligned} \tag{54}$$

$$\begin{aligned}
\mathcal{B}_2 &= \max_i \sum_j \left|\hat{\rho}_{ij,T}\right| I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right),\ \left|\rho_{ij}\right| > \theta\left(N,T\right)\right] \\
&\le \max_i \sum_{j,\rho_{ij}\ne 0} \left|\hat{\rho}_{ij,T}\right| I\left[\left|\hat{\rho}_{ij,T}\right| < \theta\left(N,T\right)\right] \\
&\le \theta\left(N,T\right)m_N.
\end{aligned}$$

Hence $\mathcal{B}_2 = O_p\left[\theta\left(N,T\right)m_N\right]$, which in conjunction with (54) yields,

$$\mathcal{B} = O_p\left[\theta\left(N,T\right)m_N\right]. \tag{55}$$

Substituting results from (53), (55) and (48) in (43), and using the outcome with (42) in (41) we obtain

$$\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\| = O_p\left[\theta\left(N,T\right)m_N\right] = O_p\left(\frac{m_N c_p\left(N\right)}{\sqrt{T}}\right),$$

as required. ■

Proof of Theorem 2. Consider the squared Frobenius norm,

$$\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F^2 = \sum\sum_{i\neq j}(\tilde{\rho}_{ij,T}-\rho_{ij})^2,$$

and recall that

$$\tilde{\rho}_{ij,T}-\rho_{ij} = \left(\hat{\rho}_{ij,T}-\rho_{ij}\right) I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right) - \rho_{ij}\left[1-I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\right].$$

Hence

$$\begin{aligned}\left(\tilde{\rho}_{ij,T}-\rho_{ij}\right)^2 \;=\; &\left(\hat{\rho}_{ij,T}-\rho_{ij}\right)^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right) + \rho_{ij}^2\left[1-I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\right]^2\\ &-2\rho_{ij}\left(\hat{\rho}_{ij,T}-\rho_{ij}\right) I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\left[1-I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\right].\end{aligned}$$

However,

$$I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\left[1-I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\right] = 0,$$

and

$$\left[1-I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\right]^2 = 1-I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right).$$

Therefore, we have

$$\begin{aligned}\sum\sum_{i\neq j}\left(\tilde{\rho}_{ij,T}-\rho_{ij}\right)^2 \;=\; &\sum\sum_{i\neq j}\left(\hat{\rho}_{ij,T}-\rho_{ij}\right)^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\\ &+\sum\sum_{i\neq j}\rho_{ij}^2\left[1-I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\right]\\ =\; &\sum\sum_{i\neq j}\left(\hat{\rho}_{ij,T}-\rho_{ij}\right)^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\\ &+\sum\sum_{i\neq j}\rho_{ij}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\right).\end{aligned}$$

Taking expectations we have the following decomposition

$$E\left(\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F^2\right) = \sum\sum_{i\neq j}E\left(\tilde{\rho}_{ij,T}-\rho_{ij}\right)^2 = \mathcal{D}+\mathcal{E}+\mathcal{F}, \tag{56}$$

where

$$\begin{aligned}\mathcal{D} \;&=\; \sum\sum_{i\neq j,\rho_{ij}\neq 0}\rho_{ij}^2 E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right],\\ \mathcal{E} \;&=\; \sum\sum_{i\neq j,\rho_{ij}\neq 0}E\left[\left(\hat{\rho}_{ij,T}-\rho_{ij}\right)^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right],\\ \mathcal{F} \;&=\; \sum\sum_{i\neq j,\rho_{ij}=0}E\left[\hat{\rho}_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right].\end{aligned}$$

Consider now the orders of the above three terms in turn, starting with $\mathcal{D}$. We have $\rho_{\min} = \min_{ij}\left(\left|\rho_{ij}\right|, \rho_{ij}\neq 0\right)$ and $\rho_{\max} = \max_{ij}\left(\left|\rho_{ij}\right|, \rho_{ij}\neq 0\right)$ such that $\rho_{\max} < 1$. Then

$$\begin{aligned}\mathcal{D} \;&\leq\; \rho_{\max}^2 N m_N \sup_{ij} E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right]\\ &=\; \rho_{\max}^2 N m_N \sup_{ij}\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij}\neq 0\right.\right),\end{aligned}$$

and using (A.6) of Lemma 3 in the online supplement for $\left|\rho_{ij}\right| > c_p(N)/\sqrt{T}$, we have

$$\mathcal{D} \leq K\rho_{\max}^2 Nm_N K e^{-\frac{T\left(\rho_{\min}-T^{-1/2}c_p(N)\right)^2}{2K_v}}\left[1+O\left(T^{\frac{2(s-2)-1}{2}}\right)\right]+O\left(Nm_N T^{-(s-1)/2}\right),$$

where $K_v = \sup_{ij}\left|K_v(\boldsymbol{\theta}_{ij})\right| < K$. By assumption $T^{-1/2}c_p(N) = o(1)$, and since $\rho_{\min} > 0$, then the first term of the above will tend to zero with N and $T \to \infty$. Therefore, $\mathcal{D}$ is of order $O(Nm_N N^{-d(s-1)/2}) = O(N^{1+\vartheta-d(s-1)/2})$, and $\mathcal{D}$ tends to zero as $N \to \infty$, for values of $d > 2(1+\vartheta)/(s-1)$ and under (21), where by assumption $0 \leq \vartheta < 1/2$.

Consider now $\mathcal{E}$. Recalling that $\hat{\rho}_{ij,T} = \omega_{ij,T}z_{ij,T} + \rho_{ij,T}$ we have the following decomposition of $\mathcal{E}$, $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 + 2\mathcal{E}_3$, where

$$\mathcal{E}_1 = \sum\sum_{i\neq j,\rho_{ij}\neq 0}\omega_{ij,T}^2 E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right],$$

$$\mathcal{E}_2 = \sum\sum_{i\neq j,\rho_{ij}\neq 0}\left(\rho_{ij,T}-\rho_{ij}\right)^2 E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right],$$

$$\mathcal{E}_3 = \sum\sum_{i\neq j,\rho_{ij}\neq 0}\left(\rho_{ij,T}-\rho_{ij}\right)\omega_{ij,T} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right].$$

Again, using (9) and (10),

$$\omega_{ij,T}^2 = \frac{K_v(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right), \tag{57}$$

$$\left(\rho_{ij,T}-\rho_{ij}\right)^2 = \frac{K_m^2(\boldsymbol{\theta}_{ij})}{T^2} + O\left(T^{-3}\right), \tag{58}$$

$$\left(\rho_{ij,T}-\rho_{ij}\right)\omega_{ij,T} = \frac{K_v^{1/2}(\boldsymbol{\theta}_{ij})K_m(\boldsymbol{\theta}_{ij})}{T^{3/2}} + O\left(T^{-5/2}\right). \tag{59}$$

$$\begin{aligned}\mathcal{E}_1 &= \sum\sum_{i\neq j,\rho_{ij}\neq 0}\omega_{ij,T}^2 E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right]\\ &\leq \frac{Nm_N}{T}\left[\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) + O\left(T^{-1}\right)\right]\sup_{ij} E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right].\end{aligned}$$

Using (A.7) in Lemma 4 in the online supplement with $r = 2$, and noting that $\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$, then $\mathcal{E}_1 = O\left(m_N N/T\right)$. Similarly, since $E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right] \leq 1$, we have

$$\begin{aligned}\mathcal{E}_2 &= \sum\sum_{i\neq j,\rho_{ij}\neq 0}\left(\rho_{ij,T}-\rho_{ij}\right)^2 E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right]\\ &\leq Nm_N\left[\frac{K_m^2(\boldsymbol{\theta}_{ij})}{T^2} + O\left(T^{-3}\right)\right] = O\left(\frac{Nm_N}{T^2}\right).\end{aligned}$$

Also, using (A.7) in Lemma 4 in the online supplement with $r = 1$, and using (59) we have

$$\begin{aligned}\mathcal{E}_3 &= \sum\sum_{i\neq j,\rho_{ij}\neq 0}\left(\rho_{ij,T}-\rho_{ij}\right)\omega_{ij,T} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}\neq 0\right)\right]\\ &\leq KNm_N\left[\frac{\sup_{ij} K_v^{1/2}(\boldsymbol{\theta}_{ij})K_m(\boldsymbol{\theta}_{ij})}{T^{3/2}} + O\left(T^{-5/2}\right)\right] = O\left(\frac{Nm_N}{T^{3/2}}\right).\end{aligned}$$

Therefore, overall $\mathcal{E} = O\left(\frac{m_N N}{T}\right)$. Consider now the following decomposition of $\mathcal{F}$, in (56):

$$\begin{aligned}\mathcal{F} &= \sum\sum_{i\neq j,\rho_{ij}=0} E\left[\hat{\rho}_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right]\\ &= \sum\sum_{i\neq j,\rho_{ij}=0}\omega_{ij,T}^2 E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right]\\ &\quad+\sum\sum_{i\neq j,\rho_{ij}=0}\rho_{ij,T}^2 E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right]\end{aligned}$$

$$+2\sum_{i\neq j,\rho_{ij}=0}\sum \rho_{ij,T}\omega_{ij,T}E\left[z_{ij,T}I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right]$$
$$= \mathcal{F}_1+\mathcal{F}_2+\mathcal{F}_3.$$

Consider $\mathcal{F}_1$ and using (57) note that

$$\mathcal{F}_1 \leq \frac{N(N-m_N-1)}{T}\left[\sup_{ij} K_v(\boldsymbol{\theta}_{ij})+O\left(T^{-1}\right)\right]\sup_{ij} E\left[z^2_{ij,T}I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right].$$

Then using (A.8) in Lemma 4 of the online supplement with $r=2$, we have

$$\mathcal{F}_1 \leq K\frac{N(N-m_N-1)}{T}\left[\sup_{ij} K_v(\boldsymbol{\theta}_{ij})+O\left(T^{-1}\right)\right]\times$$
$$\left[Ke^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}+O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{\varphi_{\max}}\right]^{\frac{3(s-2)-1}{2}}e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right)+O\left(T^{-(s-1)/2}\right)\right],$$

where $\varphi_{\max}=\sup_{ij}\varphi_{ij}>0$. Now noting that $\sup_{ij}\left|K_v(\boldsymbol{\theta}_{ij})\right|=K_v<\infty$, and $m_N/N=o(1)$, then

$$\mathcal{F}_1 = O\left(N^2T^{-1}e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right)+O\left(N^2T^{-1}T^{-(s-1)/2}\right)$$
$$= O\left(e^{-\left(\frac{1}{2\varphi_{\max}}\right)\ln N\left[\frac{c_p^2(N)}{\ln N}-2(2-d)\varphi_{\max}\right]}\right)+O\left(N^{2-d(s-1)/2-d}\right).$$

Therefore, since $\lim_{N\to\infty}c_p^2(N)/\ln(N)=2\delta$ (see result (b) of Lemma 2 in the online supplement), then $\mathcal{F}_1\to 0$, as $N\to\infty$, if $\delta>(2-d)\,\varphi_{\max}$, and $d>4/(s+1)$. Similarly, using (59), we have

$$\mathcal{F}_2 = \sum_{i\neq j,\rho_{ij}=0}\sum \rho^2_{ij,T}E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right]$$
$$\leq N(N-m_N-1)\left\{\frac{\sup_{ij}\left[K_{\mathrm{m}}^2(\boldsymbol{\theta}_{ij})\,|\rho_{ij}=0\right]}{T^2}+O\left(T^{-3}\right)\right\}\sup_{ij} E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right].$$

Now using (A.4) of Lemma 3 in the online supplement we have

$$\mathcal{F}_2 \leq K\frac{N(N-m_N-1)\left[\psi^2_{\max}+O\left(T^{-1}\right)\right]}{T^2}\times$$
$$\left[Ke^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}+O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{\varphi_{\max}}\right]^{\frac{3(s-2)-1}{2}}e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right)+O\left(T^{-(s-1)/2}\right)\right],$$

where $\psi_{\max}=\sup_{ij}\psi_{ij}<K$, and ψ_{ij} is defined by (14). Once again, since $m_N/N=o(1)$, then

$$\mathcal{F}_2 = O\left(e^{2(1-d)\ln N-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right)+O\left(N^2T^{-2}T^{-(s-1)/2}\right),$$

and following similar arguments as above, it follows that $\mathcal{F}_2\to 0$ as $N\to\infty$, if $\delta>2\,(1-d)\,\varphi_{\max}$, and $d>4/(s+3)$. Both of these conditions are met if $\delta>(2-d)\varphi_{\max}$ and $d>4/(s+1)$, since $(2-d)\varphi_{\max}>2\,(1-d)\,\varphi_{\max}$, and $s>0$. Consider now $\mathcal{F}_3$ and, using (9) and (10) evaluated at $\rho_{ij}=0$, note that

$$\mathcal{F}_3 = \sum_{i\neq j,\rho_{ij}=0}\sum \rho_{ij,T}\omega_{ij,T}E\left[z_{ij,T}I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right]$$
$$\leq \sum_{i\neq j,\rho_{ij}=0}\sum \left|\rho_{ij,T}\right|\left|\omega_{ij,T}\right|E\left[\left|z_{ij,T}\right|I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}=0\right)\right]$$
$$\leq N(N-m_N-1)\left[\frac{\sup_{ij}\left|\psi_{ij}\right|}{T}+O\left(T^{-2}\right)\right]\left[\frac{\sup_{ij}\sqrt{\varphi_{ij}}}{\sqrt{T}}+O\left(T^{-3/2}\right)\right]$$
$$\times\sup_{ij} E\left[\left|z_{ij,T}\right|I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\,|\rho_{ij}=0\right)\right].$$

Further using (A.8) in Lemma 4 of the online supplement with $r=1$, we obtain (recall that $\sup_{ij}\left|\psi_{ij}\right|<K$ and $\sup_{ij}\sqrt{\varphi_{ij}}<K$)

$$\mathcal{F}_3 \leq K\frac{N(1-m_N/N-1/N)}{T^{3/2}}\times$$
$$\left[Ke^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}+O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{\varphi_{\max}}\right]^{\frac{3(s-2)-1}{2}}e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right)+O\left(T^{-(s-1)/2}\right)\right],$$

which establishes that $\mathcal{F}_3\to 0$, as $N\to\infty$, if $\delta>(2-d)\,\varphi_{\max}$ and $d>4/(s+1)$ (using the same type of derivations as above). Therefore, overall $\mathcal{F}\to 0$, under the same conditions. Using this result, together with the results obtained for $\mathcal{D}$ and $\mathcal{E}$ above, in (56) we obtain $E\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F^2=O\left(\frac{m_N N}{T}\right)$, and (26) follows as required. Also by the Markov inequality

$$\Pr\left(\sqrt{\frac{T}{m_N N}}\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F\geq\varepsilon\right)=\Pr\left(\frac{T}{m_N N}\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F^2\geq\varepsilon^2\right)\leq\frac{\frac{T}{m_N N}E\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F^2}{\varepsilon^2}\leq\frac{K}{\varepsilon^2},$$

for some small $\varepsilon>0$. Hence,

$$\sqrt{\frac{T}{m_N N}}\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F=O_p(1),$$

and result (27) follows. ■

Proof of Theorem 3. Recall that $T=c_dN^d$, $c_d>0$, and consider first the FPR_N statistic given by (31) which can be written equivalently as

$$FPR_N=|FPR_N|=\frac{\sum\sum_{i\neq j}I\left(\left|\hat{\rho}_{ij,T}\right|>T^{-1/2}c_p(N)|\rho_{ij}=0\right)}{N(N-m_N-1)}. \tag{60}$$

Note that the elements of FPR_N are either 0 or 1 and so $|FPR_N|=FPR_N$. Taking the expectation of (60) we have

$$E\left|FPR_N\right|=\frac{\sum\sum_{i\neq j}\Pr\left(\left|\hat{\rho}_{ij,T}\right|>T^{-1/2}c_p(N)|\rho_{ij}=0\right)}{N(N-m_N-1)}\leq\sup_{ij}\Pr\left(\left|\hat{\rho}_{ij,T}\right|>T^{-1/2}c_p(N)|\rho_{ij}=0\right).$$

Hence, using (A.4) in Lemma 3 of the online supplement we have

$$E\left|FPR_N\right|=O\left(e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right)+O\left(T^{-(s-1)/2}\right),$$

where $\varphi_{\max}=\sup_{ij}\varphi_{ij}<K$ by Assumption 2 (see also Proposition 1). Hence, as $N\to\infty$ for any $d>0$ (recalling that $T=O(N^d)$), $E\left|FPR_N\right|\to 0$, noting that $c_p^2(N)\to\infty$, and $\varphi_{\max}>0$. Further, by the Markov inequality applied to $|FPR_N|$ we have, for some $\eta>0$,

$$\Pr(|FPR_N|>\eta)\leq\frac{E(|FPR_N|)}{\eta}=O\left(e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right)+O\left(T^{-(s-1)/2}\right). \tag{61}$$

It therefore follows that $\lim_{N\to\infty}\Pr(|FPR_N|>\eta)=0$, and $FPR_N\xrightarrow{p}0$ as $N\to\infty$, for any $d>0$. For almost sure convergence by the Borel Cantelli lemma it suffices to show that

$$\sum_{N=1}^{\infty}\Pr(|FPR_N|>\eta)<\infty. \tag{62}$$

From result (b) of Lemma 2 we have

$$e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}=O\left(N^{-\frac{\delta}{\varphi_{\max}}}\right), \tag{63}$$

and from (61) it follows that (for $\eta>0$)

$$\Pr(|FPR_N|>\eta)=O\left(N^{-\frac{\delta}{\varphi_{\max}}}\right)+O\left(N^{-d(s-1)/2}\right). \tag{64}$$

Hence (62) holds if $\sum_{N=1}^{\infty}N^{-\frac{\delta}{\varphi_{\max}}}$ and $\sum_{N=1}^{\infty}N^{-d(s-1)/2}$ converge, which is ensured if $\delta>\varphi_{\max}$, and $d>2/(s-1)$, which establishes that $FPR_N\xrightarrow{a.s.}0$, as $N\to\infty$.

Consider now the TPR_N statistic given by (30) and note that

$$TPR_N = \frac{\sum\sum_{i\neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} \neq 0)}{\sum\sum_{i\neq j} I(\rho_{ij} \neq 0)}.$$

Hence

$$X_N = 1 - TPR_N = \frac{\sum\sum_{i\neq j} I(\tilde{\rho}_{ij,T} = 0, \text{ and } \rho_{ij} \neq 0)}{Nm_N}.$$

Since $|X_N| = X_N$, then

$$E\,|X_N| = E(X_N) = \frac{\sum\sum_{i\neq j} \Pr\left(\left|\hat{\rho}_{ij,T}\right| < T^{-1/2}c_p(N)|\rho_{ij} \neq 0\right)}{Nm_N} \leq \sup_{ij} \Pr\left(\left|\hat{\rho}_{ij,T}\right| < T^{-1/2}c_p(N)|\rho_{ij} \neq 0\right).$$

From (A.6) of Lemma 3 of the online supplement we further have that

$$\begin{aligned} E\,|X_N| &\leq \sup_{ij} \Pr\left(\left|\hat{\rho}_{ij,T}\right| < T^{-1/2}c_p(N)|\rho_{ij} \neq 0\right) \\ &\leq K \mathrm{e}^{-\frac{T\left(\rho \min - T^{-1/2}c_p(N)\right)^2}{2K_v}} \left[1 + O\left(T^{\frac{2(s-2)-1}{2}}\right)\right] + O\left(T^{-(s-1)/2}\right), \end{aligned}$$

where $\rho_{\min} = \min_{ij}\left(\left|\rho_{ij}\right|, \rho_{ij} \neq 0\right) > 0$, and $K_v = \sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$. Hence, since by assumption $T^{-1/2}c_p(N) = o(1)$, and $T = c_d N^d$, with $c_d, d > 0$, it follows that $\lim_{N\to\infty} E\,|X_N| = 0$, as $N \to \infty$. Further, by the Markov inequality, $\Pr(|X_N| > \eta) \leq \frac{E|X_N|}{\eta}$ for some $\eta > 0$, and it follows that

$$\Pr(|TPR_N - 1| > \eta) \leq \frac{E(|TPR_N - 1|)}{\eta} = O\left(\mathrm{e}^{-\frac{T\left(\rho \min - T^{-1/2}c_p(N)\right)^2}{2K_v}}\right) + O\left(N^{-d(s-1)/2}\right). \tag{65}$$

Once again since by assumption $T^{-1/2}c_p(N) = o(1)$, $d > 0$, and $\rho_{\min} > 0$, then for any $\eta > 0$, $\lim_{N\to\infty} \Pr(|TPR_N - 1| > \eta) = 0$, and $TPR_N \xrightarrow{p} 1$, as $N \to \infty$. For almost sure convergence it is further required that

$$\sum_{N=1}^{\infty} \Pr(|TPR_N - 1| > \eta) < \infty. \tag{66}$$

From (65) we have that

$$\sum_{N=1}^{\infty} \Pr(|TPR_N - 1| > \eta) = O\left(\sum_{N=1}^{\infty} a_N\right) + O\left(\sum_{N=1}^{\infty} b_N\right),$$

where (setting $c_d = 1$ to simplify the notations)

$$a_N = e^{\frac{-N^d}{2K_v}\left(\rho_{\min} - N^{-d/2}c_p(N)\right)^2}, \text{ and } b_N = N^{-d(s-1)/2}.$$

Hence, for (66) to hold, the series $\sum_{N=1}^{\infty} a_N$ and $\sum_{N=1}^{\infty} b_N$ must converge. Using the direct comparison test for convergence of infinite series, this will be the case if

$$N^{-d(s-1)/2} \leq N^{-1-\epsilon}, \tag{67}$$

and

$$e^{\frac{-N^d}{2K_v}\left(\rho_{\min} - T^{-1/2}c_p(N)\right)^2} \leq N^{-1-\epsilon} \tag{68}$$

for all $N \geq N_0$, where N_0 is some finite positive integer, since for $\epsilon > 0$, we have $\sum_{N=1}^{\infty} N^{-1+\epsilon} < K$. Condition (68) can be written equivalently as

$$\frac{1}{2K_v}\left(\rho_{\min} - T^{-1/2}c_p(N)\right)^2 > (1+\epsilon)\,T^{-1}\ln(N),$$

which is satisfied since by assumption $N^{-d/2}c_p(N) = T^{-1//2}c_p(N) = o(1)$, $d > 0$, $\rho_{\min} > 0$, and K_v is a bounded positive constant. Hence, under the conditions of the theorem it follows that $TPR_N \xrightarrow{a.s.} 1$ as $N \to \infty$, if $d > 2/(s-1)$.

Finally, consider the FDR statistic defined by (32), and note that

$$FDR_N = \left(\frac{N - m_N - 1}{m_N}\right) FPR_N.$$

Now noting that $\frac{(N-m_N-1)}{m_N}=\ominus(N^{1-\vartheta})$, and using (64) we have

$$E\left|FDR_N\right| = O\left(N^{1-\vartheta}N^{-\frac{\delta}{\varphi\max}}\right) + O\left(N^{1-\vartheta}N^{-d(s-1)/2}\right).$$

Hence, $\lim_{N\to\infty} E\left|FDR_N\right| = 0$, as $N \to \infty$, if $\delta > (1-\vartheta)\varphi_{\max}$ and $d > 2\left(1-\vartheta\right)/(s-1)$. Also, applying Markov inequality to $|FDR_N|$, for some $\eta > 0$ we have

$$\Pr(|FDR_N| > \eta) \leq \frac{E(|FDR_N|)}{\eta} = O\left(N^{1-\vartheta}N^{-\frac{\delta}{\varphi\max}}\right) + O\left(N^{1-\vartheta}N^{-d(s-1)/2}\right). \tag{69}$$

Almost sure convergence requires

$$\sum_{N=1}^{\infty} \Pr(|FDR_N| > \eta) < \infty, \tag{70}$$

and using (69) this follows if $\delta > (2-\vartheta)\varphi_{\max}$ and $d > 2\left(2-\vartheta\right)/(s-1)$, then $FDR_N \overset{a.s.}{\to} 1$ as $N \to \infty$.. ■

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2018.10.006.

References

Abadir, K.M., Distaso, W., Zikes, F., 2014. Design-free estimation of large variance matrices. J. Econometrics 181, 165–180.
Abramovich, F., Benjamini, Y., Donoho, D., Johnstone, I., 2006. Adapting to unknown sparsity by controlling the false discovery rate. Ann. Statist. 34, 584–653.
Bhattacharya, R.N., Ghosh, J.K., 1978. On the validity of the formal edgeworth Expansion. Ann. Statist. 6, 434–451.
Bickel, P.J., Levina, E., 2004. Some theory for Fisher's linear discriminant function, 'Naive Bayes', and some alternatives when there are many more variables than observations. Bernoulli 10, 989–1010.
Bickel, P.J., Levina, E., 2008a. Covariance regularisation by thresholding. Ann. Statist. 36, 2577–2604.
Bickel, P.J., Levina, E., 2008b. Regularised estimation of large covariance matrices. Ann. Statist. 36, 199–227.
Bonferroni, C., 1935. Il calcolo delle assicurazioni su gruppi di teste. In: Studi in Onore del Professore Salvatore Ortu Carboni, Rome, Italy.
Cai, T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. J. Amer. Statist. Assoc. 106, 672–684.
Carroll, R.J., 2003. Variances are not always nuissance parameters. Biometrics 59, 211–220.
Daniels, M.J., Kass, R., 1999. Nonconjugate bayesian Estimation of Covariance Matrices in Hierarchical Models. J. Amer. Statist. Assoc. 94, 1254–1263.
Daniels, M.J., Kass, R., 2001. Shrinkage estimators for covariance matrices. Biometrics 57, 1173–1184.
D'Aspremont, A., Banerjee, O., Elghaoui, L., 2008. First-order methods for sparse covariance selection. SIAM J. Matrix Anal. Appl. 30, 56–66.
Dees, S., di Mauro, F., Pesaran, M.H., Smith, L.V., 2007. Exploring the international linkages of the euro area: A global VAR analysis. J. Appl. Econometrics 22, 1–38.
Delaigle, A., Hall, P., Jin, J., 2011. Robustness and accuracy of methods for high dimensional data analysis based on student's t-statistic. J. R. Stat. Soc. Ser. B 73, 283–301.
El Karoui, N., 2008. Operator norm consistent estimation of large dimensional sparse covariance matrices. Ann. Statist. 36, 2717–2756.
Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. J. Econometrics 147, 186–197.
Fan, J., Ke, Y., Sun, Q., Zhou, W. -X., 2018. Farmtest: factor-adjusted robust multiple testing with approximate false discovery control. J. Amer. Stat. Assoc. (forthcoming).
Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96, 1348–1360.
Fan, J., Liao, Y., Mincheva, M., 2011. High dimensional covariance matrix estimation in approximate factor models. Ann. Statist. 39, 3320–3356.
Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc. Ser. B 75, 1–44.
Fang, Y., Wang, B., Feng, Y., 2016. Tuning parameter selection in regularised estimations of large covariance matrices. J. Stat. Comput. Simul. 86, 494–509.
Friedman, J.H., 1989. Regularized discriminant analysis. J. Amer. Statist. Assoc. 84, 165–175.
Fryzlewicz, P., 2013. High-dimensional volatility matrix estimation via wavelets and thresholding. Biometrika 100, 921–938.
Gayen, A.K., 1951. The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. Biometrika 38, 219–247.
Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag New York, Inc.
Hoff, P.D., 2009. A hierarchical eigenmodel for pooled covariance estimation. J. R. Stat. Soc. Ser. B 71, 971–992.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70.
Horn, R.A., Johnson, C.R., 1985. Matrix Analysis. Cambridge University Press: Cambridge.
Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Finance 10, 603–621.
Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. J. Multivariate Anal. 88, 365–411.
Liu, H., Wang, L., Zhao, T., 2014. Sparse covariance matrix estimation with eigenvalue constraints. J. Comput. Graph. Statist. 23, 439–459.
Peng, J., Wang, P., Zhou, N., Zhu, J., 2009. Partial correlation estimation by joint sparse regression models. J. Amer. Statist. Assoc. 104, 735–746.
Pesaran, M.H., 2015. Testing weak cross-sectional dependence in large panels. Econometric Rev. 34, 1089–1117.
Pesaran, M.H., Schuermann, T., Weiner, S.M., 2004. Modeling regional interdependencies using a global Error Correcting Macroeconometric Model. J. Bus. Econom. Statist. 22, 129–162.
Pourahmadi, M., 1999. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. Biometrika 86, 677–690.
Pourahmadi, M., 2000. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. Biometrika 87, 425–435.
Pourahmadi, M., 2011. Covariance estimation: glm and Regularisation Perspectives. Statist. Sci. 26, 369–387.
Romano, P.J., Shaikh, A.M., Wolf, M., 2008. Formalised data snooping based on generalised error rates. Econom. Theory 24, 404–447.
Romano, P.J., Wolf, M., 2005. Exact and approximate stepdown methods for multiple hypothesis testing. J. Amer. Statist. Assoc. 100 (469), 94–108.
Rothman, A.J., 2012. Positive definite estimators of large covariance matrices. Biometrika 99, 733–740.
Rothman, A.J., Bickel, P.J., Levina, E., Zhu, J., 2008. Sparse permutation invariant covariance estimation. Electron. J. Stat. 2, 494–515.

Rothman, A.J., Bickel, P.J., Levina, E., Zhu, J., 2010. A new approach to cholesky-based Estimation of High-dimensional Covariance Matrices. Biometrika 97, 539–550.
Rothman, A.J., Levina, E., Zhu, J., 2009. Generalised thresholding of large covariance matrices. J. Amer. Statist. Assoc. 104, 177–186.
Schäfer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. 4, Article 32.
Sentana, E., 2009. The econometrics of mean-variance efficiency tests: a survey. Econom. J. 12, 65–101.
Tschuprow, A., 1925. Grundbegriffe und grundprobleme der Korrelationstheori. B.G. Teubner, Leipzig.
Wu, W.B., Pourahmadi, M., 2009. Banding sample covariance matrices for stationary processes. Statist. Sinica 19, 1755–1768.
Xue, L., Ma, S., Zou, H., 2012. Positive-definite ℓ_1-penalized Estimation of Large Covariance Matrices. J. Amer. Statist. Assoc. 107, 1480–1491.
Yuan, M., Lin, Y., 2007. Model selection and estimation in the Gaussian graphical model. Biometrika 94, 19–35.
Zhou, W. -X., Bose, K., Fan, J., Liu, H., 2018. A new perspective on robust m-Estimation: Finite Sample Theory and Applications to Dependence-Adjusted Multiple Testing. Ann. Statist. 46, 1904–1931.
Zou, H., 2006. The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc. 101, 1418–1429.