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

Natalia Bailey
Queen Mary, University of London

M. Hashem Pesaran
Department of Economics & USC Dornsife INET,
University of Southern California, USA, and Trinity College, Cambridge

L. Vanessa Smith
University of York

September 14, 2016

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 maximum degree of dependence of the underlying observations when their pair-wise correlation is zero, 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.

JEL Classifications: C13, C58

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

*The authors would like to thank Jianqing Fan, Alex Chudik, George Kapetanios, Yuan Liao, Ron Smith, Michael Wolf, the associate editor and two 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. The authors also wish to acknowledge financial support under ESRC Grant ES/I031626/1.

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 macro-econometric 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 bio-informatics, 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, $\mathbf{\Sigma} = (\sigma_{ij})$, 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 $\mathbf{\Sigma}$ is to appropriately restrict the off-diagonal elements of its sample estimate denoted by $\hat{\mathbf{\Sigma}}$. 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 of $\mathbf{\Sigma}$ such as the Cholesky decomposition (see Pourahmadi (1999), Pourahmadi (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), Bickel and Levina (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 $\hat{\mathbf{\Sigma}}$ 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 sparseness is loosely defined as the presence of a sufficient number of zeros on each row of $\mathbf{\Sigma}$ 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) and Fan et al. (2013) show that such regularisation techniques can be applied even if the underlying population covariance matrix is not sparse, so long as the non-sparseness 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), Daniels and Kass (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, possibly combined with cross validation, 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, $\widetilde{\boldsymbol{R}}_{MT}=(\tilde{\rho}_{ij})$, 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 $\boldsymbol{R}$ in spectral norm at the rate of $O_p\left(\frac{m_N}{\sqrt{T}}\right)$, where m_N is the maximum number of non-zero elements in the off-diagonal rows of $\boldsymbol{R}$. This compares favourably with the corresponding $O_p\left(m_N\sqrt{\frac{\log(N)}{T}}\right)$ rate established in the literature. Furthermore, it is shown that the spectral norm result holds even if N rises faster than T, with the expansion rate of N in terms of T given by $N=O\left(T^{3/2-\epsilon}\right)$, for some small positive constant, ϵ. Similarly, it is established that the MT estimator converges in Frobenius norm at the rate of $O_p\left(\sqrt{\frac{m_N N}{T}}\right)$. 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 rate of $O_p\left(\sqrt{\frac{m_N N\log(N)}{T}}\right)$, assuming the observations are Gaussian. We also establish conditions under which our proposed estimator consistently recovers the support of the population covariance matrix under non-Gaussian observations, and show that the true positive rate tends to zero 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 Supplementary Appendices.

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 exists $N_0 \geq 1$ 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 $\|\boldsymbol{A}\|_{spec} = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$, its Frobenius norm by $\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{A}'\boldsymbol{A})}$. Note that $\|\boldsymbol{A}\|_{spec} = \varrho(\boldsymbol{A})$. $\to_p$ denotes convergence in probability, and $\to_d$ convergence in distribution. $K, K_0, K_1, C, \varkappa, c_\delta, c_d, \varepsilon_0$ and ϵ are finite positive constants, independent of N and 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 $\mathbf{\Sigma}$, denoted by $\hat{\mathbf{\Sigma}}$, with elements

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

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

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

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

is bounded in N, where $I(A)$ is an indicator function that takes the value of 1 if A holds and zero otherwise. The remaining $N(N - m_N - 1)$ non-diagonal elements of $\mathbf{\Sigma}$ are zero.

A comprehensive discussion of the concept of sparsity applied to $\mathbf{\Sigma}$ and alternative ways of defining it are provided in El Karoui (2008) and Bickel and Levina (2008a). Definition 1 is a natural choice when considering concurrently the problems of regularisation of $\hat{\mathbf{\Sigma}}$ and support recovery of $\mathbf{\Sigma}$. We also make the following assumption about the bivariate moments of (x_{it}, x_{jt}).

Assumption 2 *The T observations $\{(x_{i1}, x_{j1}), (x_{i2}, x_{j2}), \ldots, (x_{iT}, x_{jT})\}$ are independent draws from a common bivariate distribution 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 $|\rho_{ij}| < 1$. Further, $E|y_{it}|^{2s} < K < \infty$, for some positive integer $s \geq 3$, where $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$. Specifically, the following moments exist*

$$\begin{aligned} \mu_{ij}(2,2) &= E\left(y_{it}^2 y_{jt}^2\right), \ \mu_{ij}(3,1) = E(y_{it}^3 y_{jt}), \ \mu_{ij}(1,3) = E(y_{it} y_{jt}^3), \\ \mu_{ij}(4,0) &= E(y_{it}^4), \ \textit{and } \mu_{ij}(0,4) = E(y_{jt}^4). \end{aligned}$$

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 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 $t = 1, 2, \ldots, T$, 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 $\Phi^{-1}(\cdot)$ 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 $H_{01}, H_{02}, \ldots, H_{0n}$, and suppose we are provided with the corresponding test statistics, $Z_{1T}, Z_{2T}, \ldots, Z_{nT}$, with separate rejection rules given by (using a two-sided alternative)

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

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 generalized 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

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

Hence to achieve $FWER_T \leq p$, it is sufficient to set $p_{iT} \leq p/n$. 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 whether the pairs (y_{it}, y_{jt}), for all $i \neq j$ display dependence when $\rho_{ij} = 0$, and on the relative rate at which N and T rise.

As will be shown in Section 2.1, when $\rho_{ij}=0$ for all i and j, $i\neq j$, the degree of dependence is defined by the parameter $\varphi_{\max}=\sup_{ij}\left(\varphi_{ij}\right)$ where $\varphi_{ij}=E\left(y_{it}^{2}y_{jt}^{2}\left|\rho_{ij}=0\right.\right)>0$. In the case when y_{it} and y_{jt} are independent, then $\varphi_{\max}=1$. In general, where the value of $\varphi_{\max}$ is not known we propose to set δ by cross validation.

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

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

where

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

Finally, the MT estimator of $\boldsymbol{\Sigma}$ is now given by

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

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

2.1 Theoretical properties of the MT estimator

Next we investigate the asymptotic properties of the MT estimator defined by (5). We begin with the following proposition.

Proposition 1 *Let $y_{it}=(x_{it}-\mu_{i})/\sqrt{\sigma_{ii}}$, where $\mu_{i}=E(x_{it})$, $|\mu_{i}|<K$, and $\sigma_{ii}=Var(x_{it})$, $0<\sigma_{ii}<K$, for all i and t, and suppose that Assumption 2 holds. Consider the sample correlation coefficient defined by (4) which can also be expressed in terms of y_{it} as*

$$\hat{\rho}_{ij,T}=\frac{\sum_{t=1}^{T}\left(y_{it}-\bar{y}_{i}\right)\left(y_{jt}-\bar{y}_{j}\right)}{\left[\sum_{t=1}^{T}\left(y_{it}-\bar{y}_{i}\right)^{2}\right]^{1/2}\left[\sum_{t=1}^{T}\left(y_{jt}-\bar{y}_{j}\right)^{2}\right]^{1/2}}. \tag{7}$$

Then

$$\rho_{ij,T}=E\left(\hat{\rho}_{ij,T}\right)=\rho_{ij}+\frac{K_{m}(\boldsymbol{\theta}_{ij})}{T}+O\left(T^{-2}\right), \tag{8}$$

$$\omega_{ij,T}^{2}=Var\left(\hat{\rho}_{ij,T}\right)=\frac{K_{v}(\boldsymbol{\theta}_{ij})}{T}+O\left(T^{-2}\right), \tag{9}$$

where

$$K_{m}(\boldsymbol{\theta}_{ij})=-\frac{1}{2}\rho_{ij}(1-\rho_{ij}^{2})+\frac{1}{8}\left\{3\rho_{ij}\left[\kappa_{ij}(4,0)+\kappa_{ij}(0,4)\right]-4\left[\kappa_{ij}(3,1)+\kappa_{ij}(1,3)\right]+2\rho_{ij}\kappa_{ij}(2,2)\right\}, \tag{10}$$

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

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

and $\boldsymbol{\theta}_{ij} = (\rho_{ij}, \mu_{ij}(0,4) + \mu_{ij}(4,0), \mu_{ij}(3,1) + \mu_{ij}(1,3), \mu_{ij}(2,2))'$. *Furthermore,* $|K_m(\boldsymbol{\theta}_{ij})| < K$, $K_v(\boldsymbol{\theta}_{ij}) = \lim_{T\to\infty}\left[TVar\left(\hat{\rho}_{ij,T}\right)\right]$, *and* $K_v(\boldsymbol{\theta}_{ij}) < K$.

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

Remark 1 *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 when* $\rho_{ij} = 0$, *by (11),*

$$
\varphi_{ij} := K_v(\boldsymbol{\theta}_{ij}\left|\rho_{ij} = 0\right.) = E\left(y_{it}^2 y_{jt}^2\left|\rho_{ij} = 0\right.\right) > 0, \tag{12}
$$

and by (10),

$$
\psi_{ij} := K_m(\boldsymbol{\theta}_{ij}\left|\rho_{ij} = 0\right.) = -0.5\left[E\left(y_{it}^3 y_{jt}\left|\rho_{ij} = 0\right.\right) + E\left(y_{it} y_{jt}^3\left|\rho_{ij} = 0\right.\right)\right]. \tag{13}
$$

Note also 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* $\psi_{ij} = 0$.

To establish probability bounds on $\hat{\rho}_{ij,T}$, following Bhattacharya and Ghosh (1978), we set out conditions under which a formal Edgeworth expansion can be established for the standardised correlation coefficient, $z_{ij,T}$. But to simplify the exposition we introduce the following general assumption first.

Assumption 3 *The standardised correlation coefficients,* $z_{ij,T} = \left[\hat{\rho}_{ij,T} - E\left(\hat{\rho}_{ij,T}\right)\right]/\sqrt{Var\left(\hat{\rho}_{ij,T}\right)}$, *for all* i *and* j $(i \neq j)$ *pairs admit the Edgeworth expansion*

$$
\begin{aligned}
\Pr\left(z_{ij,T} \le a_{ij,T}\left|\mathcal{P}_{ij}\right.\right) &= F_{ij,T}(a_{ij,T}\left|\mathcal{P}_{ij}\right.) \qquad (14) \\
&= \Phi(a_{ij,T}) + \phi\left(a_{ij,T}\right)\left[\sum_{r=1}^{s-2} T^{-r/2} G_r\left(a_{ij,T}\left|\mathcal{P}_{ij}\right.\right) + O\left(T^{-(s-1)/2}\right)\right],
\end{aligned}
$$

for some positive integer $s \ge 3$, *where* $E\left(\hat{\rho}_{ij,T}\right)$ *and* $Var\left(\hat{\rho}_{ij,T}\right)$ *are respectively defined by (8) and (9) of Proposition 1,* $\Phi(a_{ij,T})$ *and* $\phi\left(a_{ij,T}\right)$ *are the cumulative distribution and density functions of the standard Normal* $(0,1)$, *respectively, and* $G_r\left(a_{ij,T}\left|\mathcal{P}_{ij}\right.\right)$, $r = 1, 2, \ldots, s-2$, *are polynomials in* $a_{ij,T}$, *whose coefficients depend on the parameters of the underlying bivariate distribution of* x_{it} *and* x_{jt}, *for* $t = 1, 2, \ldots, T$, *which are denoted by* $\mathcal{P}_{ij}$.

The forms of the polynomial functions $G_r(.|\mathcal{P}_{ij})$, $r = 1, 2....$, is the same for all i and j pairs, and only differ in terms of the parameters of the underlying bivariate distribution of (x_{it}, x_{jt}). The following proposition provides a set of sufficient conditions under which Assumption 3 holds. Bhattacharya and Ghosh (1988) and Lahiri (2010) provide further developments that allow some relaxations of these conditions.

Proposition 2 *For a given i and j, let $\boldsymbol{\xi}_{ij,t} = (y_{it}, y_{jt}, y_{it}^2, y_{jt}^2, y_{it}y_{jt})'$, and suppose that $\boldsymbol{\xi}_{ij,t}$, for $t = 1, 2, \ldots, T$, are random draws from a common distribution, $G_{ij}(\boldsymbol{\xi})$, which is absolutely continuous with non-zero density on subsets of $\mathbb{R}^5$. Suppose further that $E|y_{it}|^{2s} < \infty$, for some positive integer $s \geq 3$. Then the Edgeworth expansion in (14) is formally valid.*

Given Assumptions 1-3, first we establish the rate of convergence of the MT estimator under the spectral (or operator) norm which implies convergence in eigenvalues and eigenvectors (see El Karoui (2008), and Bickel and Levina (2008b)).

Theorem 1 *(Convergence under the 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 $\boldsymbol{R} = (\rho_{ij})$. Suppose that Assumptions 1, 2 and 3 hold. Let $f(N) = c_\delta N^\delta$ and $T = c_d N^d$, where c_δ, c_d, and δ are finite positive constants, and $d > 2/3$. Further, suppose that there exist finite T_0 and N_0 such that for all $T > T_0$ and $N > N_0$,*

$$1 - \frac{p}{2f(N)} > 0, \tag{15}$$

where $0 < p < 1$. Consider values of δ that satisfy condition

$$\delta > (1 - 0.5d)\,\varphi_{\max}, \tag{16}$$

where $\varphi_{\max} = \sup_{ij} E\left(y_{it}^2 y_{jt}^2 \,|\rho_{ij} = 0\right) > 0$, and $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ (see Assumption 2). Then

$$E\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_{spec} = O\left(\frac{m_N}{\sqrt{T}}\right), \tag{17}$$

where m_N is defined by (3), and $\widetilde{\boldsymbol{R}}_{MT} = (\tilde{\rho}_{ij,T}) = \hat{\rho}_{ij,T} I\left[\left|\hat{\rho}_{ij,T}\right| > T^{-1/2} c_p(N)\right]$, with $c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) > 0$.

Under the conditions of Theorem 1, and since by Assumptions 1 and 2, $\lambda_{\min}(\boldsymbol{R}) \geq \varepsilon_0 > 0$, then the eigenvalues of $\widetilde{\boldsymbol{R}}_{MT}$ are bounded away from zero with probability approaching 1, and we have

$$\begin{aligned}
\left\|\left(\widetilde{\boldsymbol{R}}_{MT}\right)^{-1} - \boldsymbol{R}^{-1}\right\|_{spec} &= \left\|\left(\widetilde{\boldsymbol{R}}_{MT}\right)^{-1}\left(\boldsymbol{R} - \widetilde{\boldsymbol{R}}_{MT}\right)\boldsymbol{R}^{-1}\right\|_{spec} \\
&\leq \lambda_{\min}\left(\widetilde{\boldsymbol{R}}_{MT}\right)^{-1}\left\|\boldsymbol{R} - \widetilde{\boldsymbol{R}}_{MT}\right\|_{spec}\lambda_{\min}(\boldsymbol{R})^{-1} \\
&= O_p\left(\frac{m_N}{\sqrt{T}}\right).
\end{aligned}$$

Also see Appendix A of Fan et al. (2013) and proof of Lemma A.1 in Fan et al. (2011).

Similarly, we establish the rate of convergence of the MT estimator under the Frobenius norm.

Theorem 2 *(Convergence under the Frobenius 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 $\mathbf{R} = (\rho_{ij})$. Suppose that Assumptions 1, 2 and 3 hold. Let $f(N) = c_\delta N^\delta$ and $T = c_d N^d$, where c_δ, c_d, and δ are finite positive constants, and $d > 2/3$. Further, suppose that there exist finite T_0 and N_0 such that for all $T > T_0$ and $N > N_0$*

$$1 - \frac{p}{2f(N)} > 0, \tag{18}$$

where $0 < p < 1$. Consider values of δ that satisfy the condition

$$\delta > (2 - d)\, \varphi_{\max}, \tag{19}$$

where $\varphi_{\max} = \sup_{ij} E\left(y_{it}^2 y_{jt}^2 \left| \rho_{ij} = 0\right.\right) > 0$, $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ (see Assumption 2). Further,

$$\sqrt{T}\rho_{\min} - c_p(N) \to \infty, \tag{20}$$

where $\rho_{\min} = \min_{ij}(\left|\rho_{ij}\right|, \rho_{ij} \neq 0)$ and $c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) > 0$. Then we have

$$E\left\|\widetilde{\mathbf{R}}_{MT} - \mathbf{R}\right\|_F = O\left(\sqrt{\frac{m_N N}{T}}\right), \tag{21}$$

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

Remark 2 *In view of the conditions of Theorems 1 and 2, we note that (15) or (18) is met for any $\delta > 0$. Further, conditions (16) and (19) imply 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 $\rho_{ij} = 0$). Importantly, for a given d and $\varphi_{\max}$, convergence under the spectral norm, (17), requires a lower δ than under the Frobenius norm, (21). Both norm results hold even if N rises faster than T, so long as $d > 2/3$. In the case where N and T are of the same orders 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 spectral norm result, (17), requires $\delta > 1/2$, and the Frobenius norm result, (21), requires $\delta > 1$. Finally, note that by allowing for $\varphi_{\max}$ to differ from unity our analysis applies to non-Gaussian processes.*

Remark 3 *Condition (20) can be written as*

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

Since $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$ (see Lemma 3), this final condition is 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 $\ln(N)/N^d$, for some $d > 0$. Note that for values of $d < 1$ we allow N to rise much faster than T.

Remark 4 *The orders of convergence in (17) and (21) are in line with the results in the thresholding literature. See, for example, Theorem 1 of Cai and Liu (2011), - CL, and Bickel and Levina (2008a) - BL, with* $q = 0$*, that state the convergence rate using the spectral norm in terms of probability,* $\left\|\tilde{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{spec} = O_p\left(m_N\sqrt{\frac{\log(N)}{T}}\right)$*, where* $\tilde{\mathbf{\Sigma}}$ *is the thresholded estimator of* $\hat{\mathbf{\Sigma}}$ *using either the CL or BL approaches. Similarly, Theorem 2 of Bickel and Levina (2008a), with* $q = 0$*, using the Frobenius norm under the Gaussianity assumption, obtains a convergence rate of* $\left\|\tilde{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_F = O_p\left(\sqrt{\frac{m_N N \log(N)}{T}}\right)$*. In fact (17) and (21) are improvements on the existing rates since the* $\log(N)$ *factor is absent in both cases. The rate of* $O_p\left(\sqrt{\frac{m_N}{T}}\right)$ *is achieved in the shrinkage literature as well, if the assumption of sparseness is imposed. Here* m_N *also can be assumed to rise with* N *in which case the rate of convergence becomes slower. This compares with a rate of* $O_p\left(\sqrt{N/T}\right)$ *for the sample covariance (correlation) matrix - see Theorem 3.1 in Ledoit and Wolf (2004) - LW. Note that LW use an unconventional definition for the Frobenius norm (see their Definition 1 p. 376).*

Remark 5 *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 rate at which* N *and* T *rise. In our Monte Carlo study we consider observations complying with* $\varphi_{\max} = 1$ *and* $\varphi_{\max} = 1.5$*, and experiment with* $\delta = \{1, 2\}$*. We also present results where* δ *is estimated by cross validation over the range* $\{1 - 2.5\}$*. We find that the simulation results conform closely to our theoretical findings.*

Consider now the issue of consistent support recovery of $\boldsymbol{R}$ (or $\mathbf{\Sigma}$), which is defined in terms of true positive rate (TPR), false positive rate (FPR), and false discovery rate (FDR) statistics. Consistent support recovery requires $TPR \to 1$, $FPR \to 0$ and $FDR \to 0$, with probability 1 as N and $T \to \infty$, and does not follow immediately from the results obtained above on the convergence rates of different estimators of $\boldsymbol{R}$. The problem is addressed in the following theorem.

Theorem 3 *(Support Recovery) Consider the true positive rate (TPR), the false positive rate (FPR), and the false discovery rate (FDR) statistics defined by*

$$TPR = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} \neq 0)}{\sum\sum_{i \neq j} I(\rho_{ij} \neq 0)} \tag{22}$$

$$FPR = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum\sum_{i \neq j} I(\rho_{ij} = 0)}, \tag{23}$$

$$FDR = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum\sum_{i \neq j} I(\rho_{ij} \neq 0)}, \tag{24}$$

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),* $c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) > 0$*,* $0 < p < 1$*,* $c_p(N) \to \infty$ *as* $N \to \infty$*, and* $c_p(N)/\sqrt{T} \to 0$ *as* N *and* $T \to \infty$*. Let* $f(N) = c_\delta N^\delta$ *and* $T = c_d N^d$*, where* c_δ, c_d*,* δ *and* d *are finite positive constants. Suppose also that Assumptions 1, 2 and 3 hold, and that there exist* N_0 *and* T_0 *such that for* $N > N_0$ *and* $T > T_0$*,*

$$\sqrt{T}\rho_{\min} - c_p(N) > 0,$$

where $\rho_{\min} = \min_{ij}(\left|\rho_{ij}\right|, \rho_{ij} \neq 0) > 0$*. Then with probability tending to 1,* $TPR = 1$*, and* $FPR = 0$*, for any* $\delta > 0$*. If it is further assumed that for some small positive constant* ϵ*,* $\delta > (1-\epsilon)^{-1}\varphi_{\max}$*, where* $\varphi_{\max} = \sup_{ij} E\left(y_{it}^2 y_{jt}^2 \left| \rho_{ij} = 0\right.\right) > 0$*,* $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ *(see Assumption 2), then with probability tending to 1,* $FDR = 0$*.*

Remark 6 *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 $c_p(N)/\sqrt{T} \to 0$ *is met for any finite* $d > 0$*. For a discussion of the remaining condition* $\sqrt{T}\rho_{\min} - c_p(N) > 0$*, see Remark 3. Hence, the support recovery results in the above theorem hold even if* N *is much larger than* T*. It is also worth emphasizing that condition* $\delta > (1-\epsilon)^{-1}\varphi_{\max}$ *is required for FDR to tend to zero, but not for the results on TPR and FPR.*

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. As mentioned earlier 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

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

where $\varepsilon_{it} \sim IIDN(0,1)$, and $\chi^2_{\mathrm{v},t}$ is a chi-squared random variate with $\mathrm{v} > 4$ degrees of freedom, distributed independently of ε_{it} for all i and t. As sixth-order moments are required by Assumption 2 we set $\mathrm{v} = 8$ to ensure that $E\left(y_{it}^6\right)$ exists and $\varphi_{\max} \leq 2$. Note that under $\rho_{ij} = 0$, $\varphi_{ij} = E\left(y_{it}^2 y_{jt}^2 \left|\rho_{ij} = 0\right.\right) = (\mathrm{v}{-}2)/(\mathrm{v}{-}4)$, and with $\mathrm{v} = 8$ we have $\varphi_{ij} = \varphi_{\max} = 1.5$. Therefore, in the case where the standardised errors are multivariate t-distributed to ensure that conditions of both Theorems 1 and 2 are met we set $\delta = 2$. (See also Remark 2 and Lemma 7 in the Supplementary Appendix A). 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 8 of the 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 then sufficient for conditions of both Theorems 1 and 2 to be met, given the (N, T) combinations considered. In order to verify and calibrate the values of δ corresponding to the alternative processes generating y_{it}, we also consider an estimated version δ. For this purpose we use an analogous cross-validation procedure to CL over a specified range with end points $\delta_{\min} = 1$ and $\delta_{\max} = 2.5$, and increments of 0.1 (see the Supplementary Appendix B for further details).

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

$$\boldsymbol{x}_t = \mathbf{a} + \boldsymbol{\gamma} f_t + \mathbf{D}^{1/2}\mathbf{y}_t, \tag{25}$$

where $\boldsymbol{D} = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{NN})$, $\mathbf{a} = (a_1, a_2, \ldots, a_N)'$ 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 (25). 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 $IIDN\left(1,1\right)$ 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).

In accordance with our theoretical assumptions we consider two *exactly* sparse covariance (correlation) matrices:

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

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

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

$$\boldsymbol{R} = \boldsymbol{I}_N + \boldsymbol{b}\boldsymbol{b}' - \operatorname{diag}\left(\boldsymbol{b}\boldsymbol{b}'\right), \tag{26}$$

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 (26). 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 $\sigma_{ii} \sim IID\,(1/2 + \chi^2(2)/4)$, $i = 1, 2, \ldots, N$.

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), Rothman et al. (2009), Yuan and Lin (2007), and 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. (2013), 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 smoothly clipped absolute deviation by Fan and Li (2001) and 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),

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT}\,(\xi) = \xi\boldsymbol{I}_N + (1 - \xi)\widetilde{\boldsymbol{R}}_{MT}, \tag{27}$$

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 (27), 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

$$\xi^* = \arg \min_{\xi_0+\epsilon\leq\xi\leq 1} \left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \right\|_F^2, \tag{28}$$

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

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

Further details on the S-MT procedure, the optimisation of (28) 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 $\mathbf{\Sigma}$ where required, we obtain the following estimates of $\mathbf{\Sigma}$:

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

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

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

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

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

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

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

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

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

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

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

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

In accordance with the theoretical results in Theorems 1 and 2 and in view of Remark 5, 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-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 $\mathbf{\Sigma}$:

$$\left\|\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}}\right\|_{spec} \text{ and } \left\|\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}}\right\|_{F}, \tag{29}$$

where $\mathring{\mathbf{\Sigma}}$ is set to one of the following estimators $\{\widetilde{\mathbf{\Sigma}}_{MT,1}$, $\widetilde{\mathbf{\Sigma}}_{MT,2}$, $\widetilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}$, $\widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}$, $\widetilde{\mathbf{\Sigma}}_{CL,2}$, $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}$, $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT,1}$, $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT,2}$, $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT,\hat{\delta}}$, $\widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}$, $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}$, $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. The threshold values, $\hat{\delta}$, $\hat{C}$ and $\hat{C}^*$, are obtained by cross-validation (see Supplementary Appendix B.3 for details). Both norms are also computed for the difference between $\mathbf{\Sigma}^{-1}$, the population inverse of $\mathbf{\Sigma}$, and the estimators $\{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,1}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,2}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. 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 (22) and (23), 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 confirm our theoretical findings of Theorems 1 and 2 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 and the scaling factor using $\delta = 2$ is to be preferred (see also Remark 2).

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.

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 $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}$, $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT,2}$ and $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT,\hat{\delta}}$) with $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$. In such comparisons $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$ performs relatively better, nevertheless, $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$ is still dominated by $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT,2}$ and $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT,\hat{\delta}}$, with a few exceptions in the case of design A and primarily when $N = 30$. However, no clear ordering emerges when we compare $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$ with $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}$.

3.5 Norm comparisons of inverse estimators

Although the theoretical focus of this paper has been on estimation of $\mathbf{\Sigma}$ rather than its inverse, it is still of interest to see how well $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,2}$, $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}$, $\widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}$, $\widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}$, and $\hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}$ estimate $\mathbf{\Sigma}^{-1}$, assuming that $\mathbf{\Sigma}^{-1}$ is well defined. Table 4 provides average norm losses for Monte Carlo design B whose $\mathbf{\Sigma}$ is positive definite. $\mathbf{\Sigma}$ 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 $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,2}$ and $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}$ perform much better than $\widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}$ and $\widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}$ for Gaussian and multivariate t-distributed observations. In fact, the average spectral norms for $\widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}$ and $\widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}$ 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 $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,2}$ and $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}$ 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 $\mathbf{\Sigma}$ using the multiple testing and thresholding estimators. In the comparison set we include three versions of the MT estimator ($\widetilde{\mathbf{\Sigma}}_{MT,1}$, $\widetilde{\mathbf{\Sigma}}_{MT,2}$ and $\widetilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}$), $\widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}$, $\widetilde{\mathbf{\Sigma}}_{CL,2}$, and $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}$. 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 $\widetilde{\mathbf{\Sigma}}_{MT,1}$ and $\widetilde{\mathbf{\Sigma}}_{MT,2}$ we find that as predicted by the theory (Theorem 3 and Remark 6), TPR values of $\widetilde{\mathbf{\Sigma}}_{MT,1}$ are closer to unity as compared to the TPR values of $\widetilde{\mathbf{\Sigma}}_{MT,2}$. This is supported by the TPR values of $\widetilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}$ 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 2,000 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 $\mathbf{\Sigma}$, that ensures the conditions of Theorem 3 are met.

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 $\widetilde{\mathbf{\Sigma}}_{MT,1}$ or $\widetilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}$ do best in recovering the support of $\mathbf{\Sigma}$ 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 Computational demands of the different thresholding methods

Table 7 reports the relative execution times 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. 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 validations. 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 $CL_{\hat{C}}$. Turning to the relative execution times of $MT_{\hat{\delta}}$ and $BL_{\hat{C}}$, we find that $BL_{\hat{C}}$ is somewhere between 24 and 2634 faster to compute than $MT_{\hat{\delta}}$. However, the computational advantage of $BL_{\hat{C}}$ over $MT_{\hat{\delta}}$ procedure is to be weighed against the much more favourable performance of $MT_{\hat{\delta}}$ over $CL_{\hat{C}}$.

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, $\hat{\mathbf{\Sigma}}$, becomes ill-conditioned and is not a satisfactory estimator of the population covariance.

A regularisation estimator is proposed which uses multiple testing to calibrate the threshold value. It is shown that the resultant estimator has a convergence rate of $\left(m_N T^{-1/2}\right)$ under the spectral norm and $(m_N N/T)^{1/2}$ under the Frobenius norm, where T is the number of observations, and m_N is bounded in N (the dimension of $\mathbf{\Sigma}$), which provide slightly better rates than the convergence rates established in the literature for other regularised covariance matrix estimators. Our results derived under the spectral and Frobenius norms explicitly relate the scaling function in the multiple testing problem to the possible dependence of the underlying data when $\rho_{ij} = 0$, for all i and j, $i \neq j$, and the relative rate at which N and T rise. These results are valid under both Gaussian and non-Gaussian assumptions. This compliments the existing theoretical results in the literature for the Frobenius norm of the thresholding estimator derived only under the assumption of Gaussianity. As compared to the threshold estimators that use cross-validation, the MT estimator is also computationally simple and fast to implement.

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 $\delta = 1$.

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
					$\mathbf{u}_{it}\sim$ 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)
				$\mathbf{u}_{it}\sim$ 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
		$\mathbf{u}_{it}\sim$ 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)
	$\mathbf{u}_{it}\sim$ 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.

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
				$\mathbf{u}_{it} \sim$ 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)
				$\mathbf{u}_{it} \sim$ 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
	$\mathbf{u}_{it} \sim$ 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)
$\mathbf{u}_{it} \sim$ 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)

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.

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 $(\mathbf{\Sigma} - \mathring{\mathbf{\Sigma}})$					
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 $(\mathbf{\Sigma} - \mathring{\mathbf{\Sigma}})$					
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. $\mathring{\mathbf{\Sigma}} = \{\tilde{\mathbf{\Sigma}}_{MT,2}, \tilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}, \tilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \tilde{\mathbf{\Sigma}}_{CL,2}, \tilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \tilde{\mathbf{\Sigma}}_{S\text{-}MT,2}, \tilde{\mathbf{\Sigma}}_{S\text{-}MT,\hat{\delta}}, \tilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \tilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. $\tilde{\mathbf{\Sigma}}_{MT,2}$, $\tilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}$, $\tilde{\mathbf{\Sigma}}_{S\text{-}MT,2}$ and $\tilde{\mathbf{\Sigma}}_{S\text{-}MT,\hat{\delta}}$ 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 $\hat{\delta}$ being estimated by cross-validation. BL is Bickel and Levina universal thresholding, CL is Cai and Liu adaptive thresholding, $\tilde{\mathbf{\Sigma}}_{MT,2}$ and $\tilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}$ are based on MT_2 and $MT_{\hat{\delta}}$. $\tilde{\mathbf{\Sigma}}_{S\text{-}MT,2}$ and $\tilde{\mathbf{\Sigma}}_{S\text{-}MT,\hat{\delta}}$ apply supplementary shrinkage to $\tilde{\mathbf{\Sigma}}_{MT,2}$ and $\tilde{\mathbf{\Sigma}}_{MT,\hat{\delta}}$, $\tilde{\mathbf{\Sigma}}_{BL,\hat{C}}$ and $\tilde{\mathbf{\Sigma}}_{CL,\hat{C}}$ are based on $\hat{C}$ which is obtained by cross-validation, $\tilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}$ and $\tilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}$ employ the further adjustment to the cross-validation coefficient, $\hat{C}^*$, proposed by Fan et al. (2013), $\tilde{\mathbf{\Sigma}}_{CL,2}$ is CL's estimator with $C = 2$ (the theoretical value of C). $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$ is Ledoit and Wolf's shrinkage estimator applied to the sample covariance matrix.

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 $(\boldsymbol{\Sigma} - \mathring{\hat{\boldsymbol{\Sigma}}})$					
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 $(\boldsymbol{\Sigma} - \mathring{\hat{\boldsymbol{\Sigma}}})$					
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 for Monte Carlo design B - $T = 100$

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices $(\mathbf{\Sigma}^{-1} - \mathring{\mathbf{\Sigma}}^{-1})$					
			$\mathbf{u}_{it} \sim$ **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)
$BL_{\hat{C}^*}$	$3.8\times10^3(2.4\times10^4)$	19.56(58.88)	$1.2\times10^3(1.1\times10^4)$	12.16(33.25)	41.07(143.74)	7.66(3.17)
$CL_{\hat{C}^*}$	$1.9\times10^3(1.7\times10^4)$	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)
$BL_{\hat{C}^*}$	$157.26(1.0\times10^3)$	6.11(11.28)	$349.35(3.1\times10^3)$	9.80(17.03)	28.58(22.06)	7.77(1.04)
$CL_{\hat{C}^*}$	85.82(546.85)	5.53(7.84)	$517.27(4.8\times10^3)$	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: $\mathring{\mathbf{\Sigma}}^{-1} = \{\tilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,2}, \tilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT,\hat{\delta}}, \tilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \tilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. See also the note to Table 2.

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	$CL_{\hat{C}}$
								$\mathbf{u}_{it} \sim$ **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
								$\mathbf{u}_{it} \sim$ **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 (22) and (23), 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	$MT_{\hat{\delta}}$
$\mathbf{u}_{it}\sim$ 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
$\mathbf{u}_{it}\sim$ 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 (22) and (23), 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 2000 replications

Table 7: Relative execution times 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.

Appendix: Mathematical proofs of theorems for the MT estimator

The lemmas referred to in this Appendix are described and proved in the Supplementary Appendix A, which will be 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 (8) and (9) hold for any law of dependence between x_{it} and x_{jt}. See, in particular, p. 228 and equations (53) and (54) in Gayen (1951). Using (9) and (11) we have $\lim_{T\to\infty}\left[TVar\left(\hat{\rho}_{ij,T}\right)\right] = K_v(\boldsymbol{\theta}_{ij})$. Finally, the boundedness of $|K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})|$ and $K_v(\boldsymbol{\theta}_{ij})$ follows directly from the assumption that the sixth-order moment of y_{it} exists for all i and t. The existence of the other moments, $E(y_{it}^3 y_{jt})$ and $E\left(y_{it}^2 y_{jt}^2\right)$, follows by application of Holder's and Cauchy–Schwarz inequalities as given below:

$$\left|E(y_{it}^2 y_{jt}^2)\right| \leq \left[E(|y_{it}|^4)\right]^{1/2}\left[E\left(|y_{jt}|^4\right)\right]^{1/2} < K$$

and

$$\begin{aligned}\left|E(y_{it}y_{jt}^3)\right| &\leq E(\left|y_{it}y_{jt}^3\right|) \leq \left[E(|y_{it}|^4)\right]^{1/4}\left[E\left(\left|y_{jt}^3\right|^{4/3}\right)\right]^{3/4} \\ &= \left[E(|y_{it}|^4)\right]^{1/4}\left[E\left(|y_{jt}|^4\right)\right]^{3/4} = E(|y_{it}|^4) < K.\end{aligned}$$

■

Proof of Proposition 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 notations we are dropping the subscripts ij. 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 $\hat{\rho}_{ij,T}$, 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. 424. 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 $\sqrt{T}\left[H\left(\bar{\boldsymbol{\xi}}_T\right) - H\left(\boldsymbol{\mu}_\xi\right)\right] = \sqrt{T}\ \left(\hat{\rho}_{ij,T} - \rho_{ij}\right)$ and Theorem 2 of Bhattacharya and Ghosh (1978) can be applied to $\hat{\rho}_{ij,T}$. Note that as required by this theorem, $\boldsymbol{\xi}_t$, 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 $\mathbf{y}_t = (y_{it}, y_{jt})'$, $H\left(\boldsymbol{\xi}\right)$ is continuous and differentiable in $\boldsymbol{\xi}$, and all derivatives of $H\left(\boldsymbol{\xi}\right)$ 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 $k = 1,2,\ldots,5$, and for some positive integer $s \geq 3$. ■

In what follows we suppress subscript MT from $\widetilde{\boldsymbol{R}}_{MT}$ for notational convenience.

Proof of Theorem 1. Consider the spectral norm,

$$\left\|\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right\|_{spec} = \lambda_{\max}^{1/2}\left[\left(\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right)'\left(\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right)\right] = \lambda_{\max}^{1/2}\left[\left(\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right)^2\right] = \left|\lambda_{\max}\left[\left(\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right)\right]\right|,$$

and note that (see Horn and Johnson (1985) on p.297)

$$\left|\lambda_{\max}\left[\left(\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right)\right]\right| \leq \left\|\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right\|_\infty = \max_{1\leq i\leq N}\sum_j \left|\tilde{\rho}_{ij,T} - \rho_{ij}\right|.$$

Also

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

Hence,

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

and

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

For any given i, where $i = 1,2,\ldots,N$, and taking expectations, we obtain

$$\begin{aligned} E\left(\sum_j \left|\tilde{\rho}_{ij,T} - \rho_{ij}\right|\right) \leq\ & E\left[\sum_j \left|(\hat{\rho}_{ij,T} - \rho_{ij}) I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\right)\right|\right] + \\ & E\left[\sum_j \left|\rho_{ij} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\right)\right|\right], \end{aligned} \tag{30}$$

or

$$\sum_j E\left(\left|\tilde{\rho}_{ij,T} - \rho_{ij}\right|\right) \leq \mathcal{A}_i + \mathcal{B}_i + \mathcal{C}_i,$$

where

$$\begin{aligned} \mathcal{A}_i &= \sum_{j,\rho_{ij}\neq 0} E\left[\left|\rho_{ij}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right], \\ \mathcal{B}_i &= \sum_{j,\rho_{ij}\neq 0} E\left[\left|(\hat{\rho}_{ij,T} - \rho_{ij}) I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right|\right], \\ \mathcal{C}_i &= \sum_{j,\rho_{ij}=0} E\left[\left|\hat{\rho}_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right)\right]. \end{aligned}$$

Consider now the orders of these three terms $\mathcal{A}_i, \mathcal{B}_i$, and $\mathcal{C}_i$ in turn, starting with $\mathcal{A}_i$. 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 uniformly over all i,

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

and using (A.11) of Lemma 6 when $\left|\rho_{ij}\right| > c_p(N)/\sqrt{T}$, we have

$$\begin{aligned}
\mathcal{A}_i &\leq m_N \rho_{\max} \sup_{ij} K e^{\frac{-1}{2} \frac{T\left[\left|\rho_{ij}\right| - \frac{c_p(N)}{\sqrt{T}}\right]^2}{K_v(\boldsymbol{\theta}_{ij})}} \left[1 + o(1)\right] \\
&\leq m_N \rho_{\max} \sup_{ij} K e^{\frac{-1}{2} \frac{T\left[\rho_{\min} - \frac{c_p(N)}{\sqrt{T}}\right]^2}{\sup_{ij} K_v(\boldsymbol{\theta}_{ij})}} \left[1 + o(1)\right].
\end{aligned}$$

Recalling that $\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$ and that m_N is bounded in N, it then readily follows that $\mathcal{A}_i$ is of order $O(e^{-T})$ or $O(e^{-N^d})$. Therefore, $\mathcal{A}_i$ is uniformly bounded for all i and N, and tends to zero as $N \to \infty$.

Consider now $\mathcal{B}_i$ and note that since $\hat{\rho}_{ij,T} = \omega_{ij,T} z_{ij,T} + \rho_{ij,T}$ (to simplify the notation we use $\omega^2_{ij,T}$ and $\rho_{ij,T}$ for $Var\left(\hat{\rho}_{ij,T}\right)$ and $E\left(\hat{\rho}_{ij,T}\right)$, respectively) we have the following inequality, $\mathcal{B}_i \leq \mathcal{B}_{i1} + \mathcal{B}_{i2}$, where

$$\begin{aligned}
\mathcal{B}_{i1} &= \sum_{j,\rho_{ij}\neq 0} E\left[\left|\omega_{ij,T} z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right], \\
\mathcal{B}_{i2} &= \sum_{j,\rho_{ij}\neq 0} E\left[\left|\rho_{ij,T} - \rho_{ij}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right].
\end{aligned}$$

Using (8) and (9),

$$\omega_{ij,T} = \frac{K_v^{1/2}(\boldsymbol{\theta}_{ij})}{T^{1/2}} + O\left(T^{-3/2}\right), \tag{31}$$

$$\rho_{ij,T} - \rho_{ij} = \frac{K_m(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right). \tag{32}$$

Hence (noting that m_N is bounded in N and T, and $\omega_{ij,T} > 0$), $\mathcal{B}_{i1}$ becomes

$$\begin{aligned}
\mathcal{B}_{i1} &\leq \sum_{j,\rho_{ij}\neq 0} \omega_{ij,T} E\left[\left|z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right] \\
&\leq \frac{m_N}{T^{1/2}} \left[\sup_{ij} K_v^{1/2}(\boldsymbol{\theta}_{ij})\right] \sup_{ij} \left\{E\left(\left|z_{ij,T}\right|\right) - E\left[\left|z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right]\right\} + O\left(\frac{m_N}{T^{3/2}}\right).
\end{aligned}$$

But, $\sup_{ij} E\left(\left|z_{ij,T}\right|\right) = 2\phi(0) = \sqrt{2/\pi}$ by Lemma 2, $\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$, and

$$\begin{aligned}
&\lim_{T\to\infty} E\left[\left|z_{ij,T}\right|^s I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right] \\
&= \lim_{T\to\infty} E\left[\left|z\right|^s I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right] \qquad (33)
\end{aligned}$$

for $s=0,1,2,\ldots$, by Lemma 4, where $z\sim N(0,1)$ and

$$U_{ij,T}=\frac{c_p(N)-\sqrt{T}E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}},\ L_{ij,T}=\frac{-c_p(N)-\sqrt{T}E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}}.$$

Finally, since $c_p(N)$ is an increasing function of N, then there exists N_0 and T_0 such that for all $N>N_0$ and $T>T_0$, $U_{ij,T}>0$ and $L_{ij,T}<0$, we have (by Lemma 2, for $s=1$),

$$E\left[|z|\,I\left(L_{ij,T}\leq z\leq U_{ij,T}\right)\right]=2\phi(0)-\phi\left(L_{ij,T}\right)-\phi(U_{ij,T}).$$

Hence, (33) is bounded, as $T\to\infty$. It readily follows that $\mathcal{B}_{i1}$ is of order $O(\frac{m_N}{T^{1/2}})$, uniformly for all i.

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

$$\begin{aligned}\mathcal{B}_{i2} &= \sum_{j,\rho_{ij}\neq 0}\left|\rho_{ij,T}-\rho_{ij}\right|E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\,|\rho_{ij}\neq 0\right)\right]\\ &\leq m_N\left[\frac{|K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})|}{T}+O\left(T^{-2}\right)\right]=O\left(\frac{m_N}{T}\right),\end{aligned}$$

uniformly for all i. Overall, therefore, $\mathcal{B}_i=O(\frac{m_N}{T^{1/2}})$.

Consider now $\mathcal{C}_i$ and note that $\mathcal{C}_i\leq\mathcal{C}_{i1}+\mathcal{C}_{i2}$, where

$$\begin{aligned}\mathcal{C}_{i1} &= \sum_{j,\rho_{ij}=0}E\left[\left|\omega_{ij,T}z_{ij,T}\right|I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\,|\rho_{ij}=0\right)\right],\\ \mathcal{C}_{i2} &= \sum_{j,\rho_{ij}=0}E\left[\left|\rho_{ij,T}\right|I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\,|\rho_{ij}=0\right)\right].\end{aligned}$$

Starting with $\mathcal{C}_{i2}$, first we note that

$$\begin{aligned}\mathcal{C}_{i2} &= \sum_{j,\rho_{ij}=0}E\left[\left|\rho_{ij,T}\right|I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\,|\rho_{ij}=0\right)\right]\\ &= \sum_{j,\rho_{ij}=0}\left|\rho_{ij,T}\right|E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\,|\rho_{ij}=0\right)\right]\\ &\leq \frac{(N-m_N-1)}{T}\sup_{ij}\left[K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})\,|\rho_{ij}=0\right]\sup_{ij}E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\,|\rho_{ij}=0\right)\right],\end{aligned}$$

and $E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\,|\rho_{ij}=0\right)\right]\leq 1$. Using (32) and (A.10) of Lemma 6 (and evaluating these expressions under $\rho_{ij}=0$) we have

$$\mathcal{C}_{i2}\leq K\frac{(N-m_N-1)\left[\psi_{\max}+O\left(T^{-1}\right)\right]}{T}e^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\left[1+o(1)\right],$$

where $\varphi_{\max}=\sup_{ij}\left(\varphi_{ij}\right)>0$ with φ_{ij} defined by (12), and $\psi_{\max}=\sup_{ij}\left(\left|\psi_{ij}\right|\right)$ with ψ_{ij} defined by (13). Strictly speaking, $\mu_{ij}(3,1)$ and $\mu_{ij}(1,3)$ in the above expression are also defined under $\rho_{ij} = 0$, but since ψ_{ij} do not enter the asymptotic results this is not made explicit to simplify the notation. Then, we have

$$\mathcal{C}_{i2} \leq K_1 e^{\ln\left(\frac{N^{1-d}}{c_d}\right) - \frac{1-\epsilon}{2}\frac{c_p^2(N)}{\varphi_{\max}}} \left[1 + o(1)\right],$$

or

$$\mathcal{C}_{i2} \leq K_1 e^{-\left(\frac{1-\epsilon}{2\varphi_{\max}}\right)\ln N\left[\frac{c_p^2(N)}{\ln N} - \frac{2(1-d)\varphi_{\max}}{1-\epsilon}\right]} \left[1 + o(1)\right],$$

where $K_1 > 0$. Therefore, so long as $\lim_{N\to\infty} c_p^2(N)/\ln(N) > \frac{2(1-d)\varphi_{\max}}{1-\epsilon}$, where ϵ is a small positive constant, then $\mathcal{C}_{i2} \to 0$ as $N \to \infty$, uniformly for all i. But, from Lemma 3 we have that $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$. Hence, $\mathcal{C}_{i2} \to 0$, as $N \to \infty$, uniformly for all i, if $\delta > \frac{(1-d)\varphi_{\max}}{1-\epsilon}$.

Finally, consider $\mathcal{C}_{i1}$ and note that using (31) we have (since $\omega_{ij,T} > 0$),

$$\begin{aligned}
\mathcal{C}_{i1} &= \sum_{j,\rho_{ij}=0} E\left[\left|\omega_{ij,T} z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right)\right] \\
&= \sum_{j,\rho_{ij}=0} \omega_{ij,T} E\left[\left|z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right)\right] \\
&\leq \frac{(N - m_N - 1)}{T}\left[\sup_{ij} K_v^{1/2}(\boldsymbol{\theta}_{ij})\right] \\
&\quad \times \sup_{ij} E\left[\sqrt{T}\left|z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right)\right] + O\left(\frac{N - m_N - 1}{T^{3/2}}\right).
\end{aligned}$$

Since $K_v^{1/2}(\boldsymbol{\theta}_{ij})$ is bounded, it is then sufficient to find conditions under which

$$\lim_{T,N\to\infty} \frac{N}{T} E\left[\sqrt{T}\left|z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right)\right] = 0, \text{ uniformly in all } i \text{ and } j.$$

To this end using Lemma 4, first we note that

$$\begin{aligned}
&\lim_{T,N\to\infty} E\left[\sqrt{T}\left|z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right)\right] \\
=\ &\lim_{T,N\to\infty} E\left[\sqrt{T}\left|z_{ij,T}\right| - \sqrt{T}\left|z_{ij,T}\right| I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij} = 0\right.\right)\right] \\
=\ &\lim_{T,N\to\infty} E\left[\sqrt{T}\left|z\right| - \sqrt{T}\left|z\right| I\left(L_{ij,T}(0) \leq z \leq U_{ij,T}(0)\left|\rho_{ij} = 0\right.\right)\right], \text{ for all } i \text{ and } j,
\end{aligned}$$

where $z \sim N(0,1)$ and

$$U_{ij,T}(0) = \frac{c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij} + O\left(T^{-1}\right)}}, \tag{34}$$

and

$$L_{ij,T}(0) = \frac{-c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij} + O\left(T^{-1}\right)}}. \tag{35}$$

We also note that, since $\varphi_{ij} > 0$ and $c_p(N)$ is an increasing function of N, then there exists N_0 and T_0 such that for all $N > N_0$ and $T > T_0$, $U_{ij,T}(0) > 0$ and $L_{ij,T}(0) < 0$. Hence, by (A.4) in Lemma 2 we have

$$\begin{aligned} & E\left[\sqrt{T}\,|z| - \sqrt{T}\,|z|\, I\left(L_{ij,T}(0) \leq z \leq U_{ij,T}(0)\,|\rho_{ij} = 0\right)\right] \\ = & \sqrt{T}\left\{2\phi(0) - 2\phi(0) + \phi\left[L_{ij,T}(0)\right] + \phi\left[U_{ij,T}(0)\right]\right\} \\ = & \sqrt{T}\left\{\phi\left[L_{ij,T}(0)\right] + \phi\left[U_{ij,T}(0)\right]\right\}. \end{aligned}$$

But

$$\sqrt{T}\phi\left[L_{ij,T}(0)\right] = (2\pi)^{-1/2}\, e^{0.5\ln(T) - 0.5L^2_{ij,T}(0)},$$

and noting that $c_p(N)/\sqrt{T} \to 0$ as N and $T \to \infty$, then it readily follows that

$$\begin{aligned} \lim_{N,T\to\infty} \frac{N}{T}(2\pi)^{1/2}\sqrt{T}\phi\left[L_{ij,T}(0)\right] &= \lim_{N\to\infty}\left[e^{\ln\left(\frac{c_d N^{1-d}}{c_d}\right) + 0.5\ln(c_d N^d) - 0.5\frac{c_p^2(N)}{\varphi_{ij}} + o(1)}\right] \\ &\leq K \lim_{N\to\infty}\left\{e^{\frac{-0.5\ln N}{\varphi_{\max}}\left[\frac{c_p^2(N)}{\ln N} - \frac{1-0.5d}{0.5}\varphi_{\max}\right] + o(1)}\right\}. \end{aligned}$$

Recall that $\varphi_{\max} > 0$, and from Lemma 3, $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$. Hence,

$$\lim_{N,T\to\infty} \frac{N}{T}(2\pi)^{1/2}\sqrt{T}\phi\left[L_{ij,T}(0)\right] = 0$$

so long as $\delta > (1 - 0.5d)\,\varphi_{\max}$. It is also easily seen that under the same condition, $\lim_{N,T\to\infty}(2\pi)^{1/2}\sqrt{T}\phi\left[U_{ij,T}(0)\right] = 0$. Finally, note that we consider the remainder term of $\mathcal{C}_{i1}$ which is $O\left(\frac{N-m_N-1}{T^{3/2}}\right) = O\left(N^{1-3d/2}\right)$ and tends to zero for $d > 2/3$. Hence, $\mathcal{C}_{i1} \to 0$, as N and $T \to \infty$, uniformly for all i, if $\delta > (1 - 0.5d)\,\varphi_{\max}$, and $d > 2/3$.

Collecting the results for the orders of convergence of $\mathcal{A}_i$, $\mathcal{B}_i$, and $\mathcal{C}_i$ given above, overall we obtain a convergence rate of order $O(\frac{m_N}{T^{1/2}})$ uniformly for all i, where $i = 1, 2, \ldots, N$. Finally, (17) follows as required. ■

Proof of Theorem 2. Consider the squared Frobenius norm,

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

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

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

which can be decomposed as

$$\sum\sum_{i\neq j}E\left(\tilde{\rho}_{ij,T} - \rho_{ij}\right)^2 = \mathcal{D} + \mathcal{E} + \mathcal{F}, \tag{36}$$

where

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

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

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

and using (A.11) of Lemma 6 when $\left|\rho_{ij}\right| > c_p(N)/\sqrt{T}$, we have

$$\begin{aligned}
\mathcal{D} &\leq \rho_{\max}^2 N m_N \sup_{ij} K e^{\frac{-1}{2}\frac{T\left[\left|\rho_{ij}\right| - \frac{c_p(N)}{\sqrt{T}}\right]^2}{K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}}\left[1 + o(1)\right] \\
&\leq \rho_{\max}^2 N m_N \sup_{ij} K e^{\frac{-1}{2}\frac{T\left[\rho_{\min} - \frac{c_p(N)}{\sqrt{T}}\right]^2}{\sup_{ij} K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}}\left[1 + o(1)\right].
\end{aligned}$$

Recalling that $\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$ (including the case when $\rho_{ij} = 0$), m_N is bounded in N and $\frac{c_p(N)}{\sqrt{T}} \to 0$, as $T \to \infty$, it readily follows that $\mathcal{D}$ is of order $O(Ne^{-N^d})$. Therefore, $\mathcal{D}$ tends to zero as $N \to \infty$.

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

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

Again, using (8) and (9),

$$\omega_{ij,T}^2 = \frac{K_v(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right), \tag{37}$$

$$\left(\rho_{ij,T} - \rho_{ij}\right)^2 = \frac{K_{\mathrm{m}}^2(\boldsymbol{\theta}_{ij})}{T^2} + O\left(T^{-3}\right), \tag{38}$$

$$\left(\rho_{ij,T} - \rho_{ij}\right) \omega_{ij,T} = \frac{K_v^{1/2}(\boldsymbol{\theta}_{ij}) K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{T^{3/2}} + O\left(T^{-5/2}\right). \tag{39}$$

Hence (noting that m_N is bounded in N and T)

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

Since $\sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$ and $E\left(z_{ij,T}^2\right) = 1$, then it suffices to show that $E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right]$ is bounded. We have

$$\begin{aligned}
&\lim_{T\to\infty} E\left[z_{ij,T}^s I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right] \\
&= \lim_{T\to\infty} E\left[z^s I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N) \left|\rho_{ij} \neq 0\right.\right)\right]
\end{aligned}$$

for $s = 0, 1, 2, \ldots$, by Lemma 4, where $z \sim N(0, 1)$ and

$$U_{ij,T} = \frac{c_p(N) - \sqrt{T} E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}},\ L_{ij,T} = \frac{-c_p(N) - \sqrt{T} E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}}.$$

Then, since $c_p(N)$ is an increasing function of N, there exists N_0 and T_0 such that for all $N > N_0$ and $T > T_0$, $U_{ij,T} > 0$ and $L_{ij,T} < 0$, we have (by Lemma 2, for $s = 2$),

$$E\left[z^2 I\left(L_{ij,T} \leq z \leq U_{ij,T}\right)\right] = \Phi(U_{ij,T}) - \Phi(L_{ij,T}) + L_{ij,T}\phi\left(L_{ij,T}\right) - U_{ij,T}\phi(U_{ij,T}),$$

or

$$1 - E\left[z^2 I\left(L_{ij,T} \leq z \leq U_{ij,T}\right)\right] = \Phi(-U_{ij,T}) + \Phi(L_{ij,T}) - L_{ij,T}\phi\left(L_{ij,T}\right) + U_{ij,T}\phi(U_{ij,T}),$$

By symmetry $\Phi(U_{ij,T}) = \Phi(-U_{ij,T})$ and by Lemma 6,

$$\begin{aligned}
\Phi(U_{ij,T}) &\leq \frac{1}{2}e^{\frac{-1}{2}\frac{\left[c_p(N)-\sqrt{T}\rho_{ij}\right]^2}{K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right] \\
&\leq \frac{1}{2}e^{\frac{-1}{2}\frac{T\left[\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right]^2}{\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right],
\end{aligned}$$

so that

$$\frac{N}{T}\Phi(U_{ij,T}) \leq Ke^{(1-d)\ln(N)-\frac{c_d N^d\left[\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right]^2}{2\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right]. \tag{40}$$

Recall that $d > 2/3$, $\rho_{\min} > 0$, $0 < \sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$, and $c_p(N)/\sqrt{T} \to 0$ as N and $T \to \infty$. Distinguish between cases where $d \geq 1$, and $2/3 < d < 1$. Under the former both terms in the exponent of the exponential function on the right hand side of (40) are negative, and it readily follows that $\frac{N}{T}\Phi(U_{ij,T}) \to 0$, as N and $T \to \infty$. In the case where $2/3 < d < 1$, we write (40) as

$$\frac{N}{T}\Phi(U_{ij,T}) \leq Ke^{\frac{-[\ln(N)]\left(\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right)^2}{2\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}\left[\frac{c_d N^d}{\ln(N)}-\frac{2(1-d)\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}{\left(\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right)^2}\right]}\left[1+o(1)\right]. \tag{41}$$

Again, since $c_p(N)/\sqrt{T} \to 0$ as $T \to \infty$, $\rho_{\min} > 0$, $0 < \sup_{ij} K_v(\boldsymbol{\theta}_{ij}) < K$, and $2/3 < d < 1$, it follows that $\left[\frac{c_p(N)}{\sqrt{T}} - \rho_{\min}\right]^{-2}\left[2\left(1-d\right)\sup_{ij} K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})\right]$ is bounded in N, and there exists N_0 such that for all $N > N_0$, and $2/3 < d < 1$, $c_d N^d/\ln(N) > \left[\frac{c_p(N)}{\sqrt{T}} - \rho_{\min}\right]^{-2}\left[2\left(1-d\right)\sup_{ij} K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})\right]$. Therefore, the exponent of the exponential function on the right hand side of (41) is negative for all $N > N_0$, and hence $\frac{N}{T}\Phi(U_{ij,T}) \to 0$, as N and $T \to \infty$, even if $2/3 < d < 1$ (namely the case where N rises faster than T). Similar results follow for $\frac{N}{T}\Phi(L_{ij,T})$. Further,

$$\begin{aligned}
L_{ij,T}\phi\left(L_{ij,T}\right) &= \left[\frac{-c_p(N)-\sqrt{T}E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}}\right]\left(2\pi\right)^{-1/2}e^{-0.5L_{ij,T}^2} \\
&\leq \left\{\frac{\sqrt{T}\left[-\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}-\frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{T}+O\left(T^{-2}\right)\right]}{\sqrt{\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})+O\left(T^{-1}\right)}}\right\}\left(2\pi\right)^{-1/2}e^{-0.5\frac{T\left[\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right]^2}{\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}+o(1)} \\
&\leq K\sqrt{T}e^{-0.5\frac{T\left[\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right]^2}{\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}+o(1)}
\end{aligned}$$

where $\sup_{ij} K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) < K$ and $\frac{c_p(N)}{\sqrt{T}}$ is also bounded since $\frac{c_p(N)}{\sqrt{T}} \to 0$ as $T \to \infty$. Then,

$$\begin{aligned}
\frac{N}{T}L_{ij,T}\phi\left(L_{ij,T}\right) &\leq KN^{1-0.5d}e^{-0.5\frac{T\left[\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right]^2}{\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}+o(1)} \\
&= Ke^{-\left\{\frac{\left[\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right]^2}{2\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}\right\}\ln N\left\{\frac{N^d}{\ln N}-\frac{2(1-0.5d)\sup_{ij}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}{\left[\frac{c_p(N)}{\sqrt{T}}-\rho_{\min}\right]^2}\right\}+o(1)},
\end{aligned}$$

As before, for $d > 2/3$, we have $\frac{N}{T} L_{ij,T} \phi\left(L_{ij,T}\right) \to 0$, as N and $T \to \infty$. Similarly for $U_{ij,T}\phi(U_{ij,T})$.

Overall, $\lim_{N,T\to\infty} \frac{N}{T}\left\{1 - E\left[z^2 I\left(L_{ij,T} \leq z \leq U_{ij,T}\right)\right]\right\} = 0$. Hence, it readily follows that $\mathcal{E}_1$ is at most $O\left(\frac{m_N N}{T}\right)$, as $N, T \to \infty$.

Similarly, since $E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right] \leq 1$, we have

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

Finally,

$$\begin{aligned}
\mathcal{E}_3 &= \sum\sum_{i\neq j, \rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right) \omega_{ij,T} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right] \\
&= \sum\sum_{i\neq j, \rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right) \omega_{ij,T} E\left[z_{ij,T} - z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right] \\
&= -\sum\sum_{i\neq j, \rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right) \omega_{ij,T} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right]. \qquad (42)
\end{aligned}$$

Also, from Lemma 4

$$\lim_{T\to\infty} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right] = \lim_{T\to\infty} E\left[z I\left(L_{ij,T} \leq z \leq U_{ij,T}\left|\rho_{ij} \neq 0\right.\right)\right],$$

and from Lemma 2

$$\begin{aligned}
E\left[z I\left(L_{ij,T} \leq z \leq U_{ij,T}\left|\rho_{ij} \neq 0\right.\right)\right] &= \phi\left(\frac{-c_p(N) - \sqrt{T}\rho_{ij} + O\left(\frac{1}{\sqrt{T}}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(\frac{1}{T}\right)}}\right) \\
&\quad -\phi\left(\frac{c_p(N) - \sqrt{T}\rho_{ij} + O\left(\frac{1}{\sqrt{T}}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(\frac{1}{T}\right)}}\right), \qquad (43)
\end{aligned}$$

which is bounded in N and T. Since $\sqrt{T}\rho_{\min} - c_p(N) \to \infty$ as N and $T \to \infty$, then it is easily seen that $\lim_{T,N\to\infty} E\left[z I\left(L_{ij,T} \leq z \leq U_{ij,T}\left|\rho_{ij} \neq 0\right.\right)\right] = 0$. Hence, using (39) and noting that $K_v^{-1/2}(\boldsymbol{\theta}_{ij})K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})$ is bounded in T and $d > 2/3$, we have

$$\mathcal{E}_3 \leq K \sum\sum_{i\neq j, \rho_{ij}\neq 0} \left|\left(\rho_{ij,T} - \rho_{ij}\right)\omega_{ij,T}\right| = O\left(\frac{N m_N}{T^{3/2}}\right).$$

Overall, therefore, $\mathcal{E} = O\left(\frac{N m_N}{T}\right)$.

Consider now the following decomposition of $\mathcal{F}$, in (36):

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

Starting with the simpler terms, first we note that

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

and $\sup_{ij} E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij}\right| > c_p(N)\,|\rho_{ij}=0\right)\right] \leq 1$. Using (8) and (A.10) of Lemma 6 (and evaluating these expressions under $\rho_{ij}=0$) we have

$$\mathcal{F}_2 \leq K\frac{N(N-m_N-1)\left[\psi^2_{\max} + O\left(T^{-1}\right)\right]}{T^2} e^{-\frac{1-\epsilon}{2}\frac{c^2_p(N)}{\varphi_{\max}}}\left[1+o(1)\right],$$

or

$$\begin{aligned}
\mathcal{F}_2 &\leq Ke^{2(1-d)\ln N - \frac{1-\epsilon}{2}\frac{c^2_p(N)}{\varphi_{\max}}}\left[1+o(1)\right] \\
&= Ke^{-\left(\frac{1-\epsilon}{2\varphi_{\max}}\right)\ln N\left[\frac{c^2_p(N)}{\ln N} - \frac{4(1-d)\varphi_{\max}}{1-\epsilon}\right]}\left[1+o(1)\right],
\end{aligned}$$

where $K>0$. But from Lemma 3 we have that $\lim_{N\to\infty} c^2_p(N)/\ln(N) = 2\delta$. Hence, $\mathcal{F}_2 \to 0$ as $N\to\infty$, so long as $\delta > \frac{2(1-d)\varphi_{\max}}{1-\epsilon}$, where ϵ is a small positive constant.

Similarly,

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

But using Lemma 4, Lemma 2 and (43) and evaluating the relevant expressions under $\rho_{ij}=0$, we have

$$\begin{aligned}
& \lim_{T,N\to\infty} \sqrt{T} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \le c_p(N) \left| \rho_{ij}=0 \right.\right)\right] \\
= & \lim_{T,N\to\infty} \sqrt{T} E\left[z I\left(L_{ij,T} \le z \le U_{ij,T} \left| \rho_{ij}=0 \right.\right)\right] \\
= & \lim_{N,T\to\infty} \sqrt{T}\phi\left[L_{ij,T}(0)\right] - \lim_{N,T\to\infty} \sqrt{T}\phi\left[U_{ij,T}(0)\right],
\end{aligned}$$

where $U_{ij,T}(0)$ and $L_{ij,T}(0)$ are given by (34) and (35), respectively. As in the proof of Theorem 1, we have $\lim_{N,T\to\infty}\sqrt{T}\phi\left[L_{ij,T}(0)\right]=0=\lim_{N,T\to\infty}\sqrt{T}\phi\left[U_{ij,T}(0)\right]$, such that

$$\sqrt{T}\phi\left[L_{ij,T}(0)\right] = (2\pi)^{-1/2} e^{0.5\ln(T)-0.5L^2_{ij,T}(0)}.$$

Noting that $c_p(N)/\sqrt{T}\to 0$ as N and $T\to\infty$, it readily follows that

$$\begin{aligned}
\lim_{N,T\to\infty}\left(\frac{N}{T}\right)^2 (2\pi)^{1/2}\sqrt{T}\phi\left[L_{ij,T}(0)\right] &= K\lim_{N\to\infty}\left[e^{2(1-d)\ln N+0.5d\ln N-0.5\frac{c_p^2(N)}{\varphi_{ij}}+o(1)}\right] \\
&\le K\lim_{N\to\infty}\left\{e^{\frac{-\ln N}{2\varphi_{\max}}\left[\frac{c_p^2(N)}{\ln N}-2(2-1.5d)\varphi_{\max}\right]+o(1)}\right\}.
\end{aligned}$$

Recall that $\varphi_{\max}=\sup_{ij}(\varphi_{ij})$ and $\varphi_{ij}=E\left(y_{it}^2y_{jt}^2\left|\rho_{ij}=0\right.\right)>0$ by (12). Hence

$$\lim_{N,T\to\infty}\left(\frac{N}{T}\right)^2(2\pi)^{1/2}\sqrt{T}\phi\left[L_{ij,T}(0)\right]=0,$$

so long as $\delta > (2-1.5d)\,\varphi_{\max}$. It is also easily seen that under the same condition, $\lim_{N,T\to\infty}\left(\frac{N}{T}\right)^2(2\pi)^{1/2}\sqrt{T}\phi\left[U_{ij,T}(0)\right]=0$. Therefore, $\mathcal{F}_3\to 0$, as N and $T\to\infty$.

Finally, considering $\mathcal{F}_1$ we note that

$$\begin{aligned}
\mathcal{F}_1 &= \sum\sum_{i\neq j,\rho_{ij}=0}\omega_{ij,T}^2 E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \\
&= \sum\sum_{i\neq j,\rho_{ij}=0}\omega_{ij,T}^2 E\left\{z_{ij,T}^2\left[1-I\left(L_{ij,T}\le z_{ij,T}\le U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\} \\
&\le \frac{N(N-m_N-1)}{T}\sup_{ij}\left[\varphi_{ij}+O\left(\frac{1}{T}\right)\right] \\
&\quad\times\sup_{ij}E\left\{z_{ij,T}^2\left[1-I\left(L_{ij,T}\le z_{ij,T}\le U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\}. \qquad (44)
\end{aligned}$$

But using Lemma 4

$$\begin{aligned}
& \lim_{T\to\infty}E\left\{z_{ij,T}^2\left[1-I\left(L_{ij,T}\le z_{ij,T}\le U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\} \qquad (45) \\
= & \lim_{T\to\infty}E\left\{z^2\left[1-I\left(L_{ij,T}\le z_{ij,T}\le U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\},
\end{aligned}$$

and then by Lemma 2

$$\begin{aligned}
& E\left\{z^2\left[1-I\left(L_{ij,T}\le z_{ij,T}\le U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\} = 1-E\left[z^2 I\left(L_{ij,T}\le z\le U_{ij,T}\left|\rho_{ij}=0\right.\right)\right] \\
= & 1-\left\{\Phi\left[U_{ij,T}(0)\right]-\Phi\left[L_{ij,T}(0)\right]+L_{ij,T}(0)\phi(L_{ij,T}(0))-U_{ij,T}(0)\phi\left[U_{ij,T}(0)\right]\right\} \\
= & \Phi\left[-U_{ij,T}(0)\right]+\Phi\left[L_{ij,T}(0)\right]+U_{ij,T}(0)\phi\left[U_{ij,T}(0)\right]-L_{ij,T}(0)\phi\left[L_{ij,T}(0)\right],
\end{aligned}$$

where $U_{ij,T}(0)$ and $L_{ij,T}(0)$ are given by (34) and (35), respectively. Since $\left|\psi_{ij}\right| < K$, then there exist N_0 and T_0 such that for $N > N_0$ and $T > T_0$, $c_p(N) - \frac{\left|\psi_{ij}\right|}{\sqrt{T}} > 0$, and using Lemma 5 (also see (A.25) and (A.26) of Lemma 6), we have

$$E\left\{z^2\left[1 - I\left(L_{ij,T} \leq z \leq U_{ij,T}\left|\rho_{ij} = 0\right.\right)\right]\right\} \leq \mathcal{G}_{1,ij} + \mathcal{G}_{2,ij},$$

where

$$\mathcal{G}_{1,ij} = \frac{1}{2}e^{\frac{-1}{2}\left[\frac{c_p(N)+\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij}+O\left(T^{-1}\right)}}\right]^2} + \frac{1}{2}e^{\frac{-1}{2}\left[\frac{c_p(N)-\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij}+O\left(T^{-1}\right)}}\right]^2},$$

and

$$\begin{aligned}\mathcal{G}_{2,ij} &= \left[\frac{c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij} + O\left(T^{-1}\right)}}\right] e^{\frac{-1}{2}\left[\frac{c_p(N)+\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij}+O\left(T^{-1}\right)}}\right]^2} \\ &\quad - \left[\frac{-c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij} + O\left(T^{-1}\right)}}\right] e^{\frac{-1}{2}\left[\frac{c_p(N)-\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij}+O\left(T^{-1}\right)}}\right]^2}.\end{aligned}$$

Then, for $\frac{N^2}{T}\,\mathcal{G}_{1,ij}$ we have

$$\begin{aligned}\lim_{N,T\to\infty} \frac{N^2}{T}\,\mathcal{G}_{1,ij} &= K \lim_{N\to\infty}\left[e^{\frac{-1}{2}\frac{c_p^2(N)}{\varphi_{ij}}+(2-d)\ln N+o(1)}\right] \\ &\leq K \lim_{N\to\infty}\left\{e^{\frac{-\ln(N)}{\varphi_{\max}}\left[\frac{c_p^2(N)}{\ln(N)}-(2-d)\varphi_{\max}\right]+o(1)}\right\}\end{aligned}$$

Since $\varphi_{\max} = \sup_{ij}(\varphi_{ij}) > 0$, then $\frac{N^2}{T}\,\mathcal{G}_{1,ij}$ tends to zero if $\delta > (2-d)\,\varphi_{\max}$. Similarly, for $\frac{N^2}{T}\,\mathcal{G}_{2,ij}$ we have

$$\begin{aligned}\frac{N^2}{T}\,\mathcal{G}_{2,ij} &= \left[\frac{\frac{N^2 c_p(N)}{T} + \frac{N^2\psi_{ij}}{T\sqrt{T}} + O\left(\frac{N^2 T^{-3/2}}{T}\right)}{\sqrt{\varphi_{ij} + O\left(T^{-1}\right)}}\right] e^{\frac{-1}{2}\left[\frac{c_p(N)+\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij}+O\left(T^{-1}\right)}}\right]^2} \\ &\quad - \left[\frac{-\frac{N^2 c_p(N)}{T} + \frac{N^2\psi_{ij}}{T\sqrt{T}} + O\left(\frac{N^2 T^{-3/2}}{T}\right)}{\sqrt{\varphi_{ij} + O\left(T^{-1}\right)}}\right] e^{\frac{-1}{2}\left[\frac{c_p(N)-\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\varphi_{ij}+O\left(T^{-1}\right)}}\right]^2},\end{aligned}$$

or

$$\begin{aligned}\frac{N^2}{T}\,\mathcal{G}_{2,ij} &\leq Ke^{\left\{\frac{-\ln(N)}{\varphi_{\max}}\left[\frac{c_p^2(N)}{2\ln(N)}-\varphi_{\max}\left(2-d+\frac{\ln(c_p(N))}{\ln(N)}\right)\right]+o(1)\right\}} \\ &\quad + Ke^{\left\{\frac{-\ln(N)}{\varphi_{\max}}\left[\frac{c_p^2(N)}{2\ln(N)}-\varphi_{\max}\left(2-d+\frac{\ln(c_p(N))}{\ln(N)}\right)\right]+o(1)\right\}}.\end{aligned}$$

But $\ln[c_p(N)] = \ln[O(\ln(N)]$ and $\ln(c_p(N))/\ln(N) \to 0$, as $N \to \infty$, and $\frac{N^2}{T}\,\mathcal{G}_{2,ij} \to 0$, so long as $\delta > (2-d)\,\varphi_{\max}$. Hence, using the above results in (44) we have $\mathcal{F}_1 \to 0$ as N and $T \to \infty$.

Collecting the results for the orders of convergence of $\mathcal{F}_1, \mathcal{F}_2$, and $\mathcal{F}_3$ given above, and those of $\mathcal{D}$ and $\mathcal{E}$, overall we obtain a convergence rate of order $O(m_N N/T)$, and (21) follows as desired. ■

Proof of Theorem 3. Consider first the FPR statistic given by (23) which can be written equivalently as

$$FPR = |FPR| = \frac{\sum\sum_{i\neq j} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)|\rho_{ij}=0\right)}{N(N-m_N-1)}. \tag{46}$$

Note that the elements of FPR are either 0 or 1 and so $|FPR| = FPR$.

Taking the expectation of (46) we have

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

But using (A.10) Lemma 6 we have

$$\begin{aligned} E\,|FPR| &\leq \frac{K\sum\sum_{i\neq j} e^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\varphi_{ij}}}\,[1+o(1)]}{N(N-m_N-1)} \\ &\leq Ke^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\,[1+o(1)] \end{aligned} \tag{47}$$

where $\varphi_{\max} = \sup_{ij}\varphi_{ij} < K$, by Assumption 2. Hence, $E\,|FPR| \to 0$ as N and $T \to \infty$, noting that $c_p^2(N) \to \infty$, and $\varphi_{\max} > 0$. Further, by the Markov inequality applied to $|FPR|$ we have that

$$P(|FPR| > \eta) \leq \frac{E(|FPR|)}{\eta} \leq \frac{K}{\eta}e^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\,[1+o(1)]\,,$$

for some $\eta > 0$. Therefore, $\lim_{N,T\to\infty} P(|FPR| > \eta) = 0$, and the required result is established. This holds irrespective of the order by which N and $T \to \infty$.

Consider now the TPR statistic given by (22) and note that

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

Hence

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

Since $|X| = X$, then

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

and using the Markov inequality, $P(|X|>\eta) \leq \frac{E|X|}{\eta}$, for some $\eta > 0$, we have

$$P(|TPR-1|>\eta) \leq \frac{1}{\eta} \sup_{ij} \Pr\left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| < c_p(N) | \rho_{ij} \neq 0 \right),$$

and

$$\lim_{N,T\to\infty} P(|TPR-1|>\eta) \leq \frac{1}{\eta} \lim_{N,T\to\infty} \sup_{ij} \Pr\left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| < c_p(N) | \rho_{ij} \neq 0 \right). \tag{48}$$

However, using (A.27), (A.28) and (A.29) of Lemma 6 we have

$$\begin{aligned} & \Pr\left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| < c_p(N) | \rho_{ij} \neq 0 \right) \\ = \; & F_{ij,T}\left(\frac{c_p(N) - \sqrt{T}\rho_{ij} - \frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(T^{-1}\right)}} \right) \\ & - F_{ij,T}\left(\frac{-c_p(N) - \sqrt{T}\rho_{ij} - \frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(T^{-1}\right)}} \right). \end{aligned}$$

Suppose that $\rho_{ij} > 0$, then as N and $T \to \infty$, $c_p(N) - \sqrt{T}\rho_{ij} \to -\infty$ and $-c_p(N) - \sqrt{T}\rho_{ij} \to -\infty$, and since $F_{ij,T}(\cdot)$ is a cumulative distribution function we must have

$$\lim_{N,T\to\infty} \Pr\left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| < c_p(N) | \rho_{ij} \neq 0 \right) = F_{ij,T}(-\infty) - F_{ij,T}(-\infty) = 0 - 0 = 0.$$

Similarly if $\rho_{ij} < 0$, then $c_p(N) - \sqrt{T}\rho_{ij} \to +\infty$ and $-c_p(N) - \sqrt{T}\rho_{ij} \to +\infty$, and we have

$$\lim_{N,T\to\infty} \Pr\left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| < c_p(N) | \rho_{ij} \neq 0 \right) = F_{ij,T}(+\infty) - F_{ij,T}(+\infty) = 1 - 1 = 0.$$

Hence, more generally $\lim_{N,T\to\infty} \Pr\left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| < c_p(N) | \rho_{ij} \neq 0 \right) = 0$, if $c_p(N) - \sqrt{T}\left|\rho_{ij}\right| \to -\infty$, for all $\rho_{ij} \neq 0$, or equivalently if $\sqrt{T}\rho_{\min} - c_p(N) \to \infty$, where $\rho_{\min} = \min_{ij}\left(\left|\rho_{ij}\right|, \rho_{ij} \neq 0 \right)$. But

$$\sqrt{T}\rho_{\min} - c_p(N) = \sqrt{T}\left(\rho_{\min} - \frac{c_p(N)}{\sqrt{T}} \right),$$

and $\sqrt{T}\rho_{\min} - c_p(N) \to \infty$, as N and $T \to \infty$, since by assumption there exists N_0 and T_0 such that for all $N > N_0$ and $T > T_0$, $\rho_{\min} > c_p(N)/\sqrt{T}$, and $c_p(N)/\sqrt{T} \to 0$. The latter is ensured since by assumption $\ln f(N)/T \to 0$ (see Lemma 3). Using these results in (48) it now follows that $\lim_{N,T\to\infty} P(|TPR-1|>\eta) \to 0$, as required.

Finally, consider the FDR statistic defined by (24), and note that

$$FDR = \left(\frac{N - m_N - 1}{m_N} \right) FPR.$$

Now using (47) we have (recall that $m_N > 0$ is bounded in N)

$$\begin{aligned} E\left|FDR\right| & \leq K\ N e^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\left[1+o(1)\right] \\ & = K e^{\frac{-(1-\epsilon)\ln N}{2\varphi_{\max}}\left[\frac{c_p^2(N)}{\ln N} - \frac{2\varphi_{\max}}{1-\epsilon}\right]\left[1+o(1)\right]}. \end{aligned}$$

However, by Lemma 3, $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$, and $\lim_{N\to\infty}\left[\frac{c_p^2(N)}{\ln N} - \frac{2\varphi_{\max}}{1-\epsilon}\right] = 2\left(\delta - \frac{\varphi_{\max}}{1-\epsilon}\right) > 0$, since by assumption $\delta > \frac{\varphi_{\max}}{1-\epsilon}$. Hence, $\lim_{N\to\infty} E\left|FDR\right| = 0$, as required. ■

References

Abadir, K. M., W. Distaso, and F. Zikes (2014). Design-free estimation of large variance matrices. *Journal of Econometrics 181*, 165–180.

Abramovich, F., Y. Benjamini, D. Donoho, and I. Johnstone (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Annals of Statistics 34*, 584–653.

Bhattacharya, R. N. and J. K. Ghosh (1978). On the validity of the formal Edgeworth expansion. *The Annals of Statistics 6*, 434–451.

Bhattacharya, R. N. and J. K. Ghosh (1988). On moment conditions for valid formal Edgeworth expansions. *Journal of Multivariate Analysis 27*, 68–79.

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

Bickel, P. J. and E. Levina (2008b). Regularised estimation of large covariance matrices. *The Annals of Statistics 36*, 199–227.

Bonferroni, C. (1935). Il calcolo delle assicurazioni su gruppi di teste. *Studi in Onore del Professore Salvatore Ortu Carboni, Rome, Italy*.

Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of American Statistical Association 106*, 672–684.

Carroll, R. J. (2003). Variances are not always nuissance parameters. *Biometrics 59*, 211–220.

Daniels, M. J. and R. Kass (1999). Nonconjugate Bayesian estimation of covariance matrices in hierarchical models. *Journal of American Statistical Association 94*, 1254–1263.

Daniels, M. J. and R. Kass (2001). Shrinkage estimators for covariance matrices. *Biometrics 57*, 1173–1184.

D'Aspremont, A., O. Banerjee, and L. Elghaoui (2008). First-order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Applications 30*, 56–66.

Dees, S., F. di Mauro, M. H. Pesaran, and L. V. Smith (2007). Exploring the international linkages of the euro area: A Global VAR analysis. *Journal of Applied Econometrics 22*, 1–38.

El Karoui, N. (2008). Operator norm consistent estimation of large dimensional sparse covariance matrices. *The Annals of Statistics 36*, 2717–2756.

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

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association 96*, 1348–1360.

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

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of Royal Statistical Society, Series B 75*, 1–44.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association 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.

Hoff, P. D. (2009). A hierarchical eigenmodel for pooled covariance estimation. *Journal of Royal Statistical Society, Series B 71*, 971–992.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6*, 65–70.

Horn, R. A. and C. R. Johnson (1985). *Matrix Analysis.* Cambridge University Press: Cambridge.

Lahiri, S. N. (2010). Edgeworth expansions for studentized statistics under weak dependence. *The Annals of Statistics 38*, 388–434.

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

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

Liu, H., L. Wang, and T. Zhao (2013). Sparse covariance matrix estimation with eigenvalue constraints. *Journal of Computational and Graphical Statistics. DOI:10.1080/10618600.2013.782818*.

Peng, J., P. Wang, N. Zhou, and J. Zhu (2009). Partial correlation estimation by joint sparse regression models. *Journal of American Statistical Association 104*, 735–746.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews 34*, 1089–1117.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). Modeling regional interdependencies using a Global Error Correcting Macroeconometric model. *Journal of Business and Economic Statistics 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. *Statistical Science 26*, 369–387.

Romano, P. J., A. M. Shaikh, and M. Wolf (2008). Formalised data snooping based on generalised error rates. *Econometric Theory 24*, 404–447.

Romano, P. J. and M. Wolf (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

Rothman, A. J. (2012). Positive definite estimators of large covariance matrices. *Biometrika 99*, 733–740.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics 2*, 494–515.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2010). A new approach to Cholesky-based estimation of high-dimensional covariance matrices. *Biometrika 97*, 539–550.

Rothman, A. J., E. Levina, and J. Zhu (2009). Generalised thresholding of large covariance matrices. *Journal of American Statistical Association 104*, 177–186.

Schäfer, J. and K. Strimmer (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional Genomics. *Statistical Applications in Genetics and Molecular Biology 4, Article 32*.

Sentana, E. (2009). The econometrics of mean-variance efficiency tests: A survey. *The Econometrics Journal 12*, 65–101.

Tschuprow, A. (1925). *Grundbegriffe und Grundprobleme der Korrelationstheori.* Leipzig: B.G. Teubner.

Wu, W. B. and M. Pourahmadi (2009). Banding sample covariance matrices for stationary processes. *Statistica Sinica 19*, 1755–1768.

Xue, L., S. Ma, and H. Zou (2012). Positive-definite ℓ_1-penalized estimation of large covariance matrices. *Journal of the American Statistical Association 107*, 1480–1491.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika 94*, 19–35.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of American Statistical Association 101*, 1418–1429.