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

Natalia Bailey
Queen Mary, University of London

M. Hashem Pesaran
University of Southern California, USA, and Trinity College, Cambridge

L. Vanessa Smith
University of York

November 5, 2015

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. By using the inverse of the normal distribution at a predetermined significance level, it circumvents the challenge of estimating the theoretical constant arising in the rate of convergence of existing thresholding estimators, and hence it is easy to implement and does not require cross-validation. 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 the cross section dimension, $N$, is larger than the time series dimension, $T$.

JEL Classifications: C13, C58

Keywords: Sparse correlation matrices, High-dimensional data, Multiple testing, Thresholding, Shrinkage

*The authors would like to thank Alex Chudik, Jianqing Fan, George Kapetanios, Yuan Liao, Ron Smith and Michael Wolf 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, $\Sigma = (\sigma_{ij})$, 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 $\Sigma$ is to appropriately restrict the off-diagonal elements of its sample estimate denoted by $\hat{\Sigma}$. 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 $\Sigma$ such as the Cholesky decomposition (see Pourahmadi (1999, 2000), Rothman et al. (2010), Abadir et al. (2014), among others). Others include banding or tapering methods as proposed, for example, by Bickel and Levina (2004, 2008a) 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 (2008b), 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 $\hat{\Sigma}$ 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 $\Sigma$ such that it is absolute summable row (column)-wise, or more generally in the sense defined by El Karoui (2008). However, Fan et al. (2011, 2013) show that such regularisation techniques can be applied even if the underlying population covariance matrix is not sparse, so long as the non-sparseness is characterised by an approximate factor structure. The main challenge in applying this approach lies in the estimation of the thresholding parameter. The method of cross-validation is primarily used for this purpose which has its own limita-
tions and may not be appropriate in applications where the underlying model generating the observations is unstable over time.

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

This paper considers an alternative to cross-validation by making use of a multiple testing (MT) approach to set the thresholding parameter. The idea has been suggested by El Karoui (2008, p. 2748) but has not been theoretically developed in the literature. As noted by El Karoui, hard thresholding can also be implemented by testing the $N(N-1)/2$ null hypotheses that $\sigma_{ij} = 0$, for all $i \neq j$. However, such tests will not be standard and their critical value must be determined from the knowledge of the inferential problem and the fact that $N$ and $T$ both tend to infinity. The MT approach can readily accommodate both Gaussian and non-Gaussian observations and does not require cross-validation which is often quite time consuming to apply. The MT procedure is shown to be equivalent to the application of the multiple testing procedure due to Bonferroni (1935) to the individual rows of $\Sigma$, separately, when $\sigma_{ij} = 0$ implies independence, and to all distinct non-diagonal elements of $\Sigma$, if $\sigma_{ij} = 0$ does not imply independence. We show that the MT estimator of $\mathbf{R}$, the correlation matrix associated with $\Sigma$, converges 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 $\mathbf{R}$. This compares favourably with the corresponding $O_p \left( m_N \sqrt{\frac{\log(N)}{T}} \right)$ rate established in the literature. Similarly, we show that the MT estimator converges in Frobenius norm at the rate of $O_p \left( \sqrt{\frac{m_N N}{T}} \right)$, 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 (2008b), who establish the rate of $O_p \left( \sqrt{\frac{m_N N \log(N)}{T}} \right)$, assuming the observations are Gaussian. The MT estimator also consistently recovers the support of the population covariance matrix under non-Gaussian observations.

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 simulation results are provided in a Supplementary Appendix.

Notation: We denote the largest and the smallest eigenvalues of the \( N \times N \) real symmetric matrix \( A = (a_{ij}) \) by \( \lambda_{\text{max}}(A) \) and \( \lambda_{\text{min}}(A) \), respectively, its trace by \( tr(A) = \sum_{i=1}^{N} a_{ii} \), its maximum absolute column sum norm by \( \|A\|_1 = \max_{1 \leq i \leq N} \left( \sum_{j=1}^{N} |a_{ij}| \right) \), its maximum absolute row sum norm by \( \|A\|_\infty = \max_{1 \leq i \leq N} \left( \sum_{j=1}^{N} |a_{ij}| \right) \), its spectral (or operator) norm by \( \|A\|_{\text{spec}} = \lambda_{\text{max}}^{1/2}(A'A) \), its Frobenius norm by \( \|A\|_{F} = \sqrt{tr(A'A)} \). Note that \( \|A\|_{\text{spec}} = \rho(A) \). \( a_{N} = O(b_{N}) \) states the deterministic sequence \( \{a_{N}\} \) is at most of order \( b_{N} \), \( x_{N} = O_{p}(y_{N}) \) states the vector of random variables, \( x_{N} \), is at most of order \( y_{N} \) in probability, and \( x_{N} = o_{p}(y_{N}) \) is of smaller order in probability than \( y_{N} \), \( \to_{p} \) denotes convergence in probability, and \( \to_{d} \) convergence in distribution. All asymptotics are carried out under \( N \to \infty \) jointly with \( T \to \infty \).

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 \( (\Omega, F, P) \), and denote the covariance matrix of \( x_{t} = (x_{1t}, x_{2t}, ..., x_{Nt})' \) by

\[
\text{Var} (x_{t}) = \Sigma = E \left[ (x_{t} - \mu) (x_{t} - \mu)' \right],
\]

where \( E(x_{t}) = \mu = (\mu_{1}, \mu_{2}, ..., \mu_{N})' \), and \( \Sigma \) is an \( N \times N \) symmetric, positive definite real matrix with \( (i, j) \) element, \( \sigma_{ij} \).

We consider the regularisation of the sample covariance matrix estimator of \( \Sigma \), denoted by \( \hat{\Sigma} \), with elements

\[
\hat{\sigma}_{ij,T} = T^{-1} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i}) (x_{jt} - \bar{x}_{j}), \text{ for } i, j = 1, 2, ..., N,
\]

where \( \bar{x}_{i} = T^{-1} \sum_{t=1}^{T} x_{it} \). To this end we assume that \( \Sigma \) is (exactly) sparse defined as follows.
Assumption 1 The population covariance matrix, $\Sigma = (\sigma_{ij})$, where $\lambda_{\min}(\Sigma) \geq \varepsilon_0 > 0$, is sparse in the sense that $m_N$ defined by

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

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 $\Sigma$ are zero.

A comprehensive discussion of the concept of sparsity applied to $\Sigma$ and alternative ways of defining it are provided in El Karoui (2008) and Bickel and Levina (2008b). Definition 1 is a natural choice when considering concurrently the problems of regularisation of $\hat{\Sigma}$ and support recovery of $\Sigma$. 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 drawn from a general bivariate distribution with mean $\mu_i = E(x_{it})$, $|\mu_i| < K$, variance $\sigma_{ii} = \text{Var}(x_{it})$, $0 < \sigma_{ii} < K$, and correlation coefficient $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ satisfying $0 < \rho_{\text{min}} < |\rho_{ij}| < \rho_{\text{max}} < 1$. Also, it is assumed that the following finite higher-order moments exist

$$
\begin{align*}
\mu_{ij}(2,2) &= E(y_{it}^2 y_{jt}^2), \quad \mu_{ij}(3,1) = E(y_{it}^3 y_{jt}), \quad \text{and} \quad \mu_{ij}(1,3) = E(y_{it} y_{jt}^3), \\
\mu_{ij}(4,0) &= E(y_{it}^4) < K, \quad \text{and} \quad \mu_{ij}(0,4) = E(y_{jt}^4) < K,
\end{align*}
$$

where $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$, and $E(y_{it} y_{jt}^s) = \mu_{ij}(r,s)$, for all $r, s \geq 0$.

We follow the hard thresholding literature but, as noted above, we employ multiple testing rather than cross-validation to decide on the threshold value. More specifically, we set to zero those elements of $R = (\rho_{ij})$ that are statistically insignificant and therefore determine the threshold value as a part of a multiple testing strategy rather than by cross-validation. 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 (2008b) or El Karoui (2008) but shares the properties of the adaptive thresholding estimator developed by Cai and Lui (2011).

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

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

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 $T^{-1/2} \Phi^{-1} \left( \frac{p}{2} \right)$ as the threshold for $|\hat{\rho}_{ij,T}|$, 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 \( \Sigma \) and fails 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 \( H_{01}, H_{02}, ..., H_{0n} \), and suppose we are provided with the corresponding test statistics, \( Z_{1T}, Z_{2T}, ..., Z_{nT} \), 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

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

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 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

\[
\Pr \left[ \bigcup_{i=1}^{n} \left( |Z_{iT}| > CV_{iT} | H_{0i} \right) \right] \leq \sum_{i=1}^{n} \Pr \left( |Z_{iT}| > CV_{iT} | H_{0i} \right)
\]

\[
\leq \sum_{i=1}^{n} p_{iT}.
\]

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 \( f(N) \) and then derive conditions on \( f(N) \) which ensure consistent support recovery and a suitable convergence rate of the error in estimation of \( R = (\rho_{ij}) \). In particular, we show that the spectral norm of \( R \) and its support recovery can be consistently estimated so long as \( f(N) \) rises linearly in \( N \), and does not depend on whether \( x_{it} \) and \( x_{jt} \) are dependently distributed when \( \rho_{ij} = 0 \). However, we show that under the Frobenius norm the form of \( f(N) \) depends on whether the pairs \((x_{it}, x_{jt})\), for all \( i \neq j \) display non-linear dependence, in the sense that they are dependent even if \( \rho_{ij} = 0 \). As will be shown in Section 2.1, under the null hypothesis, \( H_{0,ij} : \rho_{ij} = 0 \) for all \( i \) and \( j \), \( i \neq j \), the degree of non-linear dependence is defined by the
parameter $\kappa_{\text{max}} = \sup_{ij} (\kappa_{ij})$ where $\kappa_{ij} = [\mu_{ij}(2, 2)|\rho_{ij} = 0]$. Under independence, $\kappa_{\text{max}} = 1$ and $f(N) = N$, while under non-linear dependence we have $\kappa_{\text{max}} > 1$ and $f(N) = O(N^{\kappa_{\text{max}}})$.

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

$$\tilde{\rho}_{ij} = \hat{\rho}_{ij} I \{ |\hat{\rho}_{ij}| > T^{-1/2} c_p(N) \}, \ i = 1, 2, ..., N - 1, \ j = i + 1, ..., N,$$

where

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

It is evident that since $c_p(N)$ is selected a priori and does not need to be estimated, the multiple testing procedure in (5) is also computationally simple to implement. This contrasts with traditional thresholding approaches which face the challenge of evaluating the theoretical constant, $C$, arising in the rate of convergence of their estimators. A separate cross-validation procedure is typically employed for the estimation of $C$ that has its own limitations.

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

$$\tilde{\Sigma}_{MT} = \tilde{D}^{1/2} \tilde{R}_{MT} \tilde{D}^{1/2},$$

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

### 2.1 Theoretical properties of the MT estimator

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 $\sigma_{ii} = \text{Var}(x_{it})$, $0 < \sigma_{ii} < K$, 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

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

Then

$$\rho_{ij,T} = E(\hat{\rho}_{ij,T}) = \rho_{ij} + \frac{K_m(\theta_{ij})}{T} + O(T^{-2}),$$

$$\omega_{ij,T}^2 = \text{Var}(\hat{\rho}_{ij,T}) = \frac{K_v(\theta_{ij})}{T} + O(T^{-2}),$$

where

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

$$K_v(\theta_{ij}) = \text{Var}(\rho_{ij}) = \frac{1}{2} \rho_{ij}(1 - \rho_{ij}^2) + \frac{1}{4} \left\{ 3\rho_{ij} [\kappa_{ij}(4, 0) + \kappa_{ij}(0, 4)] - 4 [\kappa_{ij}(3, 1) + \kappa_{ij}(1, 3)] + 2\rho_{ij} \kappa_{ij}(2, 2) \right\},$$

$$\omega_{ij,T}^2 = \text{Var}(\hat{\rho}_{ij,T}) = \frac{K_v(\theta_{ij})}{T} + O(T^{-2}).$$
\[ K_v(\theta_{ij}) = (1 - \rho_{ij}^2)^2 + \frac{1}{4} \left\{ \rho_{ij}^2 [\kappa_{ij}(4, 0) + \kappa_{ij}(0, 4)] - 4\rho_{ij} [\kappa_{ij}(3, 1) + \kappa_{ij}(1, 3)] + 2(2 + \rho_{ij}^2)\kappa_{ij}(2, 2) \right\}, \]

\( K_v(\theta_{ij}) = (1 - \rho_{ij}^2)^2 + \frac{1}{4} \left\{ \rho_{ij}^2 [\kappa_{ij}(4, 0) + \kappa_{ij}(0, 4)] - 4\rho_{ij} [\kappa_{ij}(3, 1) + \kappa_{ij}(1, 3)] + 2(2 + \rho_{ij}^2)\kappa_{ij}(2, 2) \right\}, \]

Remark 1

Finally, the boundedness of results (8) and (9) is established in Gayen (1951) 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 \( |K_m(\theta_{ij})| < K \), \( K_v(\theta_{ij}) = \lim_{T \to \infty} [TVar(\hat{\rho}_{ij,T})] \), and \( K_v(\theta_{ij}) < K \).

Proof of Proposition 1. The results for \( E(\hat{\rho}_{ij,T}) \) and \( Var(\hat{\rho}_{ij,T}) \) 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 \( |K_m(\theta_{ij})| < K \), \( K_v(\theta_{ij}) = \lim_{T \to \infty} [TVar(\hat{\rho}_{ij,T})] \), and \( K_v(\theta_{ij}) < K \).

Remark 1

From Gayen (1951) p.232 (eq (54)bis) it follows that \( K_m(\theta_{ij}) > 0 \) for each correlation coefficient \( \rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} \) satisfying \( 0 < \rho_{\min} < |\rho_{ij}| < \rho_{\max} < 1 \). Further, under the null \( H_{0,ij} : \rho_{ij} = 0 \), (11) becomes \( K_m(\theta_{ij}) = 1 + \kappa_{ij}(2, 2) = \mu_{ij}(2, 2) > 0 \).

We introduce the following assumption which is inspired from the above proposition.

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

\[ Pr(z_{ij,T} \leq a_{ij,T} | P_{ij}) = F_{ij,T}(a_{ij,T} | P_{ij}) = \Phi(a_{ij,T} + T^{-1/2} \phi(a_{ij,T}) G_1(a_{ij,T} | P_{ij}) + T^{-1} \phi(a_{ij,T}) G_2(a_{ij,T} | P_{ij}) + \ldots, \]

where \( E(\hat{\rho}_{ij,T}) \) and \( Var(\hat{\rho}_{ij,T}) \) are defined by (8) and (9) of Proposition 1, \( \Phi(a_{ij,T}) \) and \( \phi(a_{ij,T}) \) are the cumulative distribution and density functions of the standard Normal \((0, 1), \)
respectively, and \( G_s(a_{ij,T} | \mathcal{P}_{ij}) \), \( s = 1, 2, ... \) are polynomials in \( a_{ij,T} \), whose coefficients depend on the underlying bivariate distribution of the observations \((x_{it}, x_{jt} \text{ for } t = 1, 2, ..., T)\) which is denoted by \( \mathcal{P}_{ij} \).

**Remark 2** While Assumption C1 of Cai and Liu (2011) characterising the tail-property of \( y_{it} \) can be used, we opt to focus on the standardised correlation coefficient, \( z_{ij,T} \). This is a self-normalised process where \( E(\hat{\rho}_{ij,T}) \) and \( \text{Var}(\hat{\rho}_{ij,T}) \) are given by (8) and (9) respectively. Then, for a finite \( T \), all moments of \( z_{ij,T} \) exist and as \( T \to \infty, z_{ij,T} \to_d z \sim N(0,1) \). Hence, following the theorem of Sargan (1976) on p.423 the Edgeworth expansion is 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 (2008a)).

**Theorem 1** *(Convergence under the spectral norm)* Denote the sample correlation coefficient of \( x_{it} \) and \( x_{jt} \) over \( t = 1, 2, ..., T \) by \( \hat{\rho}_{ij,T} \) (as defined in (7) of Proposition 1) and the population correlation matrix by \( \mathbf{R} = (\rho_{ij}) \). Suppose that Assumptions 1-3 hold. Let \( f(N) \) be an increasing function of \( N, p \) a finite constant \((0 < p < 1)\), and suppose there exist finite \( T_0 \) and \( N_0 \) such that for all \( T > T_0 \) and \( N > N_0 \),

\[
1 - \frac{p}{2f(N)} > 0
\]

and

\[
\ln f(N)/T \to 0, \text{ as } N \text{ and } T \to \infty.
\]

Then so long as \( N/\sqrt{T} \to 0 \) we have

\[
E \left\| \mathbf{R}_{MT} - \mathbf{R} \right\|_{\text{spec}} = O \left( \frac{m_N}{\sqrt{T}} \right),
\]

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

**Proof.** See Appendix.

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

\[
\left\| \left( \mathbf{R}_{MT} \right)^{-1} - \mathbf{R}^{-1} \right\|_{\text{spec}} = \left\| \left( \mathbf{R}_{MT} \right)^{-1} \left( \mathbf{R} - \mathbf{R}_{MT} \right) \mathbf{R}^{-1} \right\|_{\text{spec}} \\
\leq \lambda_{\min} \left( \mathbf{R}_{MT} \right)^{-1} \left\| \mathbf{R} - \mathbf{R}_{MT} \right\|_{\text{spec}} \lambda_{\min} (\mathbf{R})^{-1} \\
= O_p \left( \frac{m_N}{\sqrt{T}} \right).
\]

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) Denote the sample correlation coefficient of $x_{it}$ and $x_{jt}$ over $t = 1, 2, ..., T$ by $\hat{\rho}_{ij,T}$ (as defined in (7) of Proposition 1) and the population correlation matrix by $R = (\rho_{ij})$. Suppose that Assumptions 1-3 hold. Let $f(N)$ be an increasing function of $N$, $p$ a finite constant ($0 < p < 1$), and suppose there exist finite $T_0$ and $N_0$ such that for all $T > T_0$ and $N > N_0$,

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

$$\ln f(N)/T \to 0, \text{ as } N \text{ and } T \to \infty,$$

and

$$\kappa_{\max} \leq \lim_{N \to \infty} \frac{\ln[f(N)]}{\ln(N)},$$

where $\kappa_{\max} = \sup_{ij} [\kappa_{ij}], \kappa_{ij} = [\mu_{ij}(2,2) \big| \rho_{ij} = 0], \text{ with } \mu_{ij}(2,2) \text{ defined in Assumption 2.}$ Then as long as $N/\sqrt{T} \to 0$ we have

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

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

Proof. See Appendix. ■

Remark 3 For the convergence of the Frobenius norm, $E \left\| \tilde{R}_{MT} - R \right\|_F$, at the rate of $O_p \left( \sqrt{m_N N/T} \right)$, the rate at which $f(N)$ rises with $N$ is dictated by the magnitude of $\kappa_{\max}$. For example if $\kappa_{\max} = 1$, setting $f(N) = N - 1$ meets all the conditions of Theorem 2. But for values of $\kappa_{\max} > 1$, we need $f(N)$ to rise with $N$ at a faster rate. For $\kappa_{\max} \in (1, 2]$, it is sufficient to set $f(N) = N(N - 1)/2$. It is easily seen that in this case

$$\lim_{N \to \infty} \left[ \frac{\ln f(N)}{\ln(N)} - \kappa_{\max} \right] = \lim_{N \to \infty} \left[ \frac{\ln(N) + \ln(N - 1) - \ln(2)}{\ln(N)} - \kappa_{\max} \right] = 2 - \kappa_{\max},$$

and the conditions are met if $\kappa_{\max} \leq 2$. Similarly, for $\kappa_{\max} \leq 3$ we need to specify $f(N) = O(N^3)$.

Remark 4 While in practice we find it reasonable to set $\kappa_{\max}$ no greater than two, further research is required in determining this data dependent measure, which is beyond the scope of the present paper. Convergence under the spectral norm is less demanding than under the Frobenius norm. Unlike Theorem 2, the statement of Theorem 1 does not require a condition on $\kappa_{\max}$. This implies that under the spectral norm, when controlling the errors in estimation of $R$, it is sufficient for $f(N)$ to rise linearly with $N$ irrespective of the value of $\kappa_{\max}$. 
Remark 5 The orders of convergence in (13) and (15) 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 (2008b, BL), with $q = 0$, that state the convergence rate using the spectral norm in terms of probability, $\| \hat{\Sigma} - \Sigma \|_{\text{spec}} = O_p \left( m_N \sqrt{\log(N)/T} \right)$, where $\hat{\Sigma}$ is the thresholded estimator of $\Sigma$ using either the CL or BL approaches. Similarly, Theorem 2 of Bickel and Levina (2008b), with $q = 0$, using the Frobenius norm under the Gaussianity assumption, obtains a convergence rate of $\| \hat{\Sigma} - \Sigma \|_F = O_p \left( \sqrt{m_N N \log(N)/T} \right)$. In fact (13) and (15) are improvements on the existing rates since the $\log(N)$ factor is absent in both cases. The rate of $O_p \left( \sqrt{m_N / T} \right)$ 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 6 Results (13) and (15) also hold if a concept of ‘approximate’ sparseness is used in place of Assumption 1, such that $m_N$ is defined more generally as $m_N = \max_{i \leq N} \sum_{j=1}^{N} |\sigma_{ij}|^q$, for some $q \in [0, 1)$. See Bickel and Levina (2008b) or Fan et al. (2013).

Remark 7 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 is too conservative if $\rho_{ij} = 0$ implies $x_{it}$ and $x_{jt}$ are independent, but could be appropriate otherwise. In our Monte Carlo study we consider observations with linear and non-linear dependence, and experiment with $f(N) = N - 1$ and $f(N) = N(N - 1)/2$. We find that the simulation results conform closely to our theoretical findings.

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

**Theorem 3 (Support Recovery)** Consider the true positive rate (TPR) and the false positive rate (FPR) statistics computed using the multiple testing estimator

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


given by

\[
TPR = \frac{\sum \sum I(\hat{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} \neq 0)}{\sum \sum I(\rho_{ij} \neq 0)} \tag{16}
\]

\[
FPR = \frac{\sum \sum I(\hat{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum \sum I(\rho_{ij} = 0)}, \tag{17}
\]

where \( \hat{\rho}_{ij,T} \) is the pair-wise correlation coefficient defined by (7), \( c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) > 0, \) \( 0 < p < 1, \) \( f(N) \) is an increasing function such that \( c_p(N) \to \infty, \) as \( N \to \infty, \) \( \ln f(N)/T \to 0 \) and \( c_p(N)/\sqrt{T} \to 0, \) as \( N \) and \( T \to \infty. \) Suppose also that Assumptions 1-3 hold. Then with probability tending to 1, \( TPR = 1, \) and with probability tending to 1, \( FPR = 0, \) if 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 \( \rho_{\min} = \min_{ij} |\rho_{ij}| > 0. \)

**Proof.** See Appendix. ■

**Remark 8** The proof of support recovery does not depend on \( \mu_{ij}(2,2). \) Also it only requires that \( f(N) \) rises with \( N \) linearly. For example, setting \( f(N) = N - 1 \) it is easily seen that \( \ln f(N)/T = \ln(N - 1)/T \to 0, \) as \( N \) and \( T \to \infty, \) and \( c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) \to \infty, \) as \( N \to \infty, \) and conditions of Theorem 3 are met. Interestingly, this suggests that consistent support recovery is ensured if Bonferroni’s MT procedure is applied to \( \mathbf{R} \) (or \( \Sigma \)) row-wise. One is likely to encounter loss of power if Bonferroni’s procedure is applied to all the distinct off-diagonal elements of \( \mathbf{R}. \) A similar argument can be made for Holm’s MT procedure, although the application of Holm’s procedure row-wise can result in contradictions due to the symmetry of the correlation matrix.

## 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 (2008b, 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 \( C = 2 \) proposed by the authors. 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

\[
y_t = Pu_t, \quad t = 1, 2, ..., T,
\]
where $y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $u_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$, and $P$ is the Cholesky factor associated with the choice of the correlation matrix $R = PP'$. 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

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

for $i = 1, 2, \ldots, N$.

where $\varepsilon_{it} \sim IIDN(0,1)$, and $\chi^2_{v,t}$ is a chi-squared random variate with $v > 4$ degrees of freedom, distributed independently of $\varepsilon_{it}$ for all $i$ and $t$. As fourth-order moments are required by Assumption 2 we set $v = 8$ to ensure that $E(y^4_{it})$ exists and $\kappa_{\text{max}} \leq 2$. Note that under $\rho_{ij} = 0$, $\kappa_{ij} = \mu_{ij}(2,2|\rho_{ij} = 0) = (v-2)/(v-4)$, and with $v = 8$ we have $\kappa_{ij} = \kappa_\ast = 1.5$. Therefore, in the case where the standardised errors are multivariate t-distributed to ensure that conditions of Theorem 2 are met we must set $f(N) = N(N-1)/2$. (See also Remark 3 and Lemma 7 in the Supplementary Appendix A). One could further allow for fat-tailed $\varepsilon_{it}$ shocks, though fat-tail shocks alone (e.g. generating $u_{it}$ as such) do not necessarily result in $\kappa_{ij} > 1$ as shown in Lemma 8 of the Supplementary Appendix A. The same is true for normal shocks under case (i), where $\mu_{ij}(2,2) = 1$ whether $P = I_N$ or not. In such cases setting $f(N) = N - 1$ is then sufficient for conditions of Theorem 2 to be met.

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

$$x_t = a + \gamma f_t + D^{1/2} y_t,$$

(18)

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

We report results for $N = \{30, 100, 200\}$ and $T = 100$, for the baseline case where $\gamma = 0$ and $a = 0$ in (18). The properties of the MT procedure when factors are included in the data generating process are also investigated by drawing $\gamma_i$ and $a_i$ as $IIDN(1,1)$ for $i = 1, 2, \ldots, N$, and generating $f_t$, the common factor, as a stationary AR(1) process, but to save space these results are made available upon request. Under both settings we focus on the residuals from an OLS regression of $x_t$ 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

$$\Sigma = (\sigma_{ij}) = \text{diag}(A_1, A_2),$$

where $A_1 = A + \epsilon I_{N/2}$, $A = (a_{ij})_{1 \leq i,j \leq N/2}$, $a_{ij} = 1 - \frac{|i-j|}{10}$, with $\epsilon = \max(-\lambda_{\text{min}}(A),0)+0.01$ to ensure that $A$ is positive definite, and $A_2 = 4I_{N/2}$. $\Sigma$ is a two-block diagonal matrix, $A_1$ is a banded and sparse covariance matrix, and $A_2$ is a diagonal matrix with 4 along the diagonal. Matrix $P$ is obtained numerically by applying the Cholesky decomposition to the correlation matrix, $R = D^{-1/2}\Sigma D^{-1/2} = PP'$, where the diagonal elements of $D$ 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 $\mathbf{b} = (b_1, b_2, ..., b_N)'$ with elements generated as $\text{Uniform}(0.7, 0.9)$ for the first and last $N_b (< N)$ elements of $\mathbf{b}$, where $N_b = \lfloor N^\delta \rfloor$, and set the remaining middle elements of $\mathbf{b}$ to zero. The resulting population correlation matrix $\mathbf{R}$ is defined by

$$R = I_N + b'b - \text{diag}(bb'),$$

(19)

for which $\sqrt{T}\rho_{\min} - c_p(N) > 0$ and $\rho_{\min} = \min_{ij} |\rho_{ij}| > 0$, in line with Theorem 3. The degree of sparseness of $\mathbf{R}$ is determined by the value of the parameter $\delta$. We are interested in weak cross-sectional dependence, so we focus on the case where $\delta < 1/2$ following Pesaran (2015), and set $\delta = 0.25$. Matrix $\mathbf{P}$ is then obtained by applying the Cholesky decomposition to $\mathbf{R}$ defined by (19). Further, we set $\Sigma = D^{1/2}RD^{1/2}$, where the diagonal elements of $\mathbf{D}$ are given by $\sigma_{ii} \sim \text{IID} (1/2 + \chi^2(2)/4)$, $i = 1, 2, ..., N$.

An additional two covariance specifications based on approximately sparse matrices as defined in Bickel and Levina (2008b, p. 2580 for $0 < q < 1$), namely the correlation matrices corresponding to an AR(1) and spatial AR(1), SAR(1), process respectively, along with their associated simulation results can be found in the Supplementary Appendix D.

### 3.1 Finite sample positive definiteness

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

A number of methods have been developed in the literature that produce sparse inverse covariance matrix estimates which make use of a penalised likelihood (D’Aspremont et al. (2008), Rothman et al. (2008, 2009), Yuan and Lin (2007), 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 $\hat{\mathbf{R}}_{MT}$ 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, $I_N$, 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 $\hat{\mathbf{R}}_{MT}$ 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 ($\mathbf{S-MT}$),

$$\tilde{\mathbf{R}}_{S-MT}(\xi) = \xi I_N + (1 - \xi)\hat{\mathbf{R}}_{MT},$$

(20)

with shrinkage parameter $\xi \in (\xi_0, 1]$, and $\xi_0$ being the minimum value of $\xi$ that produces a non-singular $\tilde{\mathbf{R}}_{S-MT}(\xi_0)$ matrix. Alternative ways of computing the optimal weights on the
two matrices can be entertained. We choose to calibrate, $\xi$, since opting to use $\xi_0$ in \(20\), as suggested in El Karoui (2008), does not necessarily provide a well-conditioned estimate of $\tilde{R}_{S-MT}$. Accordingly, we set $\xi$ by solving the following optimisation problem

$$
\xi^* = \arg \min_{\xi_0 + \epsilon \leq \xi \leq 1} \left\| R_0^{-1} - \tilde{R}^{-1}_{S-MT}(\xi) \right\|_F^2,
$$

where $\epsilon$ is a small positive constant, and $R_0$ is a reference invertible correlation matrix. Finally, we construct the corresponding covariance matrix as

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

Further details on the $S-MT$ procedure, the optimisation of \(21\) and choice of reference matrix $R_0$ 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 $\Sigma$ where required, we obtain the following estimates of $\Sigma$:

- $MT_{N-1}$: thresholding based on the $MT$ approach applied to the sample correlation matrix ($\tilde{\Sigma}_{MT}$) using $f(N) = N - 1$ ($\tilde{\Sigma}_{MT_{N-1}}$)
- $MT_{N(N-1)/2}$: thresholding based on the $MT$ approach applied to the sample correlation matrix ($\tilde{\Sigma}_{MT}$) using $f(N) = N(N - 1)/2$ ($\tilde{\Sigma}_{MT_{N(N-1)/2}}$)
- $BL_C$: BL thresholding on the sample covariance matrix using cross-validated $C$ ($\tilde{\Sigma}_{BL,C}$)
- $CL_2$: CL thresholding on the sample covariance matrix using the theoretical value of $C = 2$ ($\tilde{\Sigma}_{CL,2}$)
- $CL_\tilde{C}$: CL thresholding on the sample covariance matrix using cross-validated $C$ ($\tilde{\Sigma}_{CL,\tilde{C}}$)
- $S-MT_{N-1}$: supplementary shrinkage applied to $MT_{N-1}$ ($\tilde{\Sigma}_{S-MT_{N-1}}$)
- $S-MT_{N(N-1)/2}$: supplementary shrinkage applied to $MT_{N(N-1)/2}$ ($\tilde{\Sigma}_{S-MT_{N(N-1)/2}}$)
- $BL_{\tilde{C}}$: BL thresholding using the Fan, Liao and Mincheva (2013, FLM) cross-validation adjustment procedure for estimating $C$ to ensure positive definiteness ($\tilde{\Sigma}_{BL,\tilde{C}}$)
- $CL_{\tilde{C}}$: CL thresholding using the FLM cross-validation adjustment procedure for estimating $C$ to ensure positive definiteness ($\tilde{\Sigma}_{CL,\tilde{C}}$)
- $LW_{\tilde{C}}$: LW shrinkage on the sample covariance matrix ($\tilde{\Sigma}_{LW_{\tilde{C}}}$).

In accordance with the theoretical results in Theorem 2 and in view of Remark 3, we consider two versions of the $MT$ estimator depending on the choice of $f(N) = \{N - 1, N(N - 1)/2\}$. The $BL_{\tilde{C}}$ and $CL_2$ and $CL_{\tilde{C}}$ estimators apply the thresholding procedure without ensuring that the resultant covariance estimators are invertible. The next five estimators yield invertible covariance estimators. The $S-MT$ estimators are obtained using the supplementary shrinkage approach described in Section 3.1. $BL_{\tilde{C}}$ and $CL_{\tilde{C}}$ estimators are obtained by applying the additional FLM adjustments. The shrinkage estimator, $LW_{\tilde{C}}$, 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 $\hat{D}^{1/2} \hat{R}_{MT} \hat{D}^{1/2}$. 

14
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 $\Sigma$:

$$\|\Sigma - \hat{\Sigma}\|_{\text{spec}} \quad \text{and} \quad \|\Sigma - \hat{\Sigma}\|_{F},$$

(22)

where $\hat{\Sigma}$ is set to one of the following estimators $\{\hat{\Sigma}_{MT_{N-1}}, \hat{\Sigma}_{MT_{N(N-1)/2}}, \hat{\Sigma}_{BL,\hat{C}}, \hat{\Sigma}_{CL,2}, \hat{\Sigma}_{CL,\hat{C}}, \hat{\Sigma}_{s-MT_{N-1}}, \hat{\Sigma}_{s-MT_{N(N-1)/2}}, \hat{\Sigma}_{BL,\hat{C}^*}, \hat{\Sigma}_{CL,\hat{C}^*}, \hat{\Sigma}_{LW_{\hat{C}}}\}$. The threshold values, $\hat{C}$ and $\hat{C}^*$, are obtained by cross-validation (see Supplementary Appendix B.3 for details). Both norms are also computed for the difference between $\Sigma^{-1}$, the population inverse of $\Sigma$, and the estimators $\{\hat{\Sigma}_{s-MT_{N-1}}, \hat{\Sigma}_{s-MT_{N(N-1)/2}}, \hat{\Sigma}_{BL,\hat{C}^*}, \hat{\Sigma}_{CL,\hat{C}^*}, \hat{\Sigma}_{LW_{\hat{C}}}\}$. 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 (16) and (17), respectively. The statistics TPR and FPR are not relevant to the shrinkage estimator $LW_{\hat{C}}$ and will not be reported for this estimator.

### 3.3 Robustness of MT to the choice of the p-value and $f(N)$

We begin by investigating the sensitivity of the $MT$ estimator to the choice of the p-value, $p$, and the scaling factor $f(N)$ 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 $p = \{0.01, 0.05, 0.10\}$, and $f(N) = \{N-1, N(N-1)/2\}$. Table 1 summarises the spectral and Frobenius norm losses (averaged over 2000 replications) for both Monte Carlo designs A and B, 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 the scaling factor is $N(N-1)/2$, irrespective of whether the observations are drawn from a Gaussian or a multivariate $t$ distribution. Perhaps this is to be expected since for $N$ sufficiently large the effective $p$-value which is given by $2p/N(N-1)$ is very small and the test outcomes are more likely to be robust to the changes in the values of $p$ as compared to the case when the scaling factor used is $N - 1$. The results in Table 1 also confirm our theoretical finding of Theorem 2 that in the case of Gaussian observations, where $\kappa_{\text{max}} = 1$, the scaling factor $N - 1$ is likely to perform better as compared to $N(N-1)/2$, but the reverse is true if the observations are multivariate $t$ distributed and the scaling factor $N(N-1)/2$ is to be preferred (see also Remark 3). We also note that all the norm losses rise with $N$ given that $T$ is kept at 100 in all the experiments. We obtain similar results when we consider other Monte Carlo designs with approximately sparse covariance matrices. To save space the results for these designs are provided in the Supplementary covariance Appendix D. Overall, we find that the results are more robust when the scaling factor $N(N-1)/2$ is used.

### 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 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 $N(N - 1)/2$, and the conventional significance level of $p = 0.05$. Initially, we consider the threshold estimators, $MT$, $BL$ and the two versions of the $CL$ estimators ($CL_2$ and $CL_{C}$) without further adjustments to ensure invertibility. First, we note that the $MT$ and $CL$ estimators (both versions) dominate the $BL$ estimator in every case, without any exceptions and for both designs. 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 $eCL;^C$ and $eS-MT$) with $^LW$. In such comparisons $^LW$ performs relatively better, nevertheless, $^LW$ is still dominated by $S-MT$, with a few exceptions in the case of design A and primarily when $N = 30$. However, no clear ordering emerges when we compare $^LW$ with $eCL;^C$.

### 3.5 Norm comparisons of inverse estimators

Although the theoretical focus of this paper has been on estimation of $\Sigma$ rather than its inverse, it is still of interest to see how well $\Sigma^{-1}$, $\Sigma_{BL;^C}$, $\Sigma_{CL;^C}$, and $\Sigma_{LW}$ estimate $\Sigma^{-1}$, assuming that $\Sigma^{-1}$ is well defined. Table 4 provides average norm losses for Monte Carlo design B whose $\Sigma$ is positive definite. $\Sigma$ for design A is ill-conditioned and will not be considered any further here. As can be seen from the results in Table 4, $S-MT$ performs much better than $BL;^C$ and $CL;^C$ for Gaussian and multivariate $t$-distributed observations. In fact, the average spectral norms for $BL;^C$ and $CL;^C$ 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 $S-MT$ performs better than $LW$, for all sample sizes and irrespective of whether the observations are drawn as Gaussian or multivariate $t$.

### 3.6 Support recovery statistics

Table 5 reports the true positive and false positive rates (TPR and FPR) for the support recovery of $\Sigma$ using the multiple testing and thresholding estimators. In the comparison set we include two versions of the $MT$ estimator ($MT_{N-1}$ and $MT_{N(N-1)/2}$), $BL;^C$, $CL;^C$, and $CL;^C$. Again we use 100 replications due to the use of cross-validation in the implementation of BL and CL thresholding. We include the $MT$ estimators for both choices of the scaling factor, $f(N) = N-1$ and $f(N) = N(N-1)/2$, computed at $p = 0.05$, to see if our theoretical result, namely that for consistent support recovery only the linear scaling factor, $N - 1$, is needed, is borne out by the simulations. 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 $\tilde{\Sigma}_{MT_{N-1}}$ and $\tilde{\Sigma}_{MT_{N(N-1)/2}}$ we find that as predicted by the theory (Theorem 3 and Remark 8), TPR values of $\tilde{\Sigma}_{MT_{N-1}}$ are closer to unity as compared to the TPR values of $\tilde{\Sigma}_{MT_{N(N-1)/2}}$. 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 $f(N) = \{N - 1, N(N - 1)/2\}$ 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, $f(N)$, 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 $\Sigma$, 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 non-linear dependence in the errors causes the support recovery performance of $BL_{CL}$, $CL_2$ and $CL_{CL}$ to deteriorate noticeably while $MT_{N-1}$ and $MT_{N(N-1)/2}$ remain remarkably stable. Finally, again note that TPR values are higher for design B. Overall, the estimator $\tilde{\Sigma}_{MT_{N-1}}$ does best in recovering the support of $\Sigma$ as compared to other estimators, although the results of $CL$ and $MT$ for support recovery are 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 $MT_{N(N-1)/2}$ estimator. It took 0.010, 0.013, and 0.016 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 slight difference in execution time between $MT_{N-1}$ and $MT_{N(N-1)/2}$ amounts to the stricter condition imposed by the p-value on the $MT_{N(N-1)/2}$ procedure, which produces a slightly sparser version of $\tilde{\Sigma}$. In contrast, the $BL_{CL}$ and $CL_{CL}$ thresholding approaches are computationally much more demanding. Their computations took between about 18 and 412293 times longer than the MT approach, for the same sample sizes and computer hardware. The $BL_{CL}$ method was less demanding than the $CL_{CL}$ method - it took between about 18 and 500 times longer than the MT approach. Even $CL_2$, which does not require estimation of the threshold parameter, took up to 17 times longer than the MT approach. Thus, compared with other thresholding methods MT has a clear computational advantage.

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

A regularisation estimator is proposed which uses multiple testing rather than cross-validation to calibrate the threshold value. It is shown that the resultant estimator has a convergence rate of $(m_N T^{-1/2})$ 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 $\Sigma$), which provide slightly better rates than the convergence rates established in the literature for other regularised covariance matrix estimators. Our results derived under the Frobenius norm explicitly relate the scaling function in the multiple testing problem to the possible non-linear dependence of the underlying data, and together with the spectral norm 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, $|\hat{\rho}_{ij}| > T^{-1/2}\Phi^{-1}\left(1 - \frac{p}{2f(N)}\right)$, particularly when $f(N)$ is set to $N(N - 1)/2$. For support recovery, better results are obtained if $f(N)$ is set to $N - 1$. 
Table 1: Spectral and Frobenius norm losses for the $MT(p)$ estimator using significance levels $p = \{0.01, 0.05, 0.10\}$ and the scaling factors $f(N) = \{N - 1, N(N - 1)/2\}$, for $T = 100$

<table>
<thead>
<tr>
<th></th>
<th>$f(N) = N - 1$</th>
<th>$f(N) = N(N - 1)/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>$MT_{N-1}(0.01)$</td>
<td>$MT_{N-1}(0.05)$</td>
</tr>
<tr>
<td>Spectral norm</td>
<td>$u_{it} \sim$ Gaussian</td>
<td>$u_{it} \sim$ multivariate $t$--distributed with 8 degrees of freedom</td>
</tr>
<tr>
<td>30</td>
<td>1.70(0.49)</td>
<td>1.68(0.49)</td>
</tr>
<tr>
<td>100</td>
<td>2.61(0.50)</td>
<td>2.51(0.50)</td>
</tr>
<tr>
<td>200</td>
<td>3.04(0.48)</td>
<td>2.92(0.49)</td>
</tr>
<tr>
<td>Frobenius norm</td>
<td>$u_{it} \sim$ Gaussian</td>
<td>$u_{it} \sim$ multivariate $t$--distributed with 8 degrees of freedom</td>
</tr>
<tr>
<td>30</td>
<td>3.17(0.45)</td>
<td>3.14(0.50)</td>
</tr>
<tr>
<td>100</td>
<td>6.66(0.45)</td>
<td>6.51(0.51)</td>
</tr>
<tr>
<td>200</td>
<td>9.87(0.46)</td>
<td>9.60(0.53)</td>
</tr>
<tr>
<td>$u_{it} \sim$ Gaussian</td>
<td>$u_{it} \sim$ multivariate $t$--distributed with 8 degrees of freedom</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>2.26(1.08)</td>
<td>2.43(1.20)</td>
</tr>
<tr>
<td>100</td>
<td>3.85(4.84)</td>
<td>4.21(5.29)</td>
</tr>
<tr>
<td>200</td>
<td>4.49(3.46)</td>
<td>5.04(4.34)</td>
</tr>
<tr>
<td>Spectral norm</td>
<td>$u_{it} \sim$ Gaussian</td>
<td>$u_{it} \sim$ multivariate $t$--distributed with 8 degrees of freedom</td>
</tr>
<tr>
<td>30</td>
<td>0.48(0.16)</td>
<td>0.50(0.16)</td>
</tr>
<tr>
<td>100</td>
<td>0.75(0.34)</td>
<td>0.76(0.32)</td>
</tr>
<tr>
<td>200</td>
<td>0.71(0.22)</td>
<td>0.74(0.20)</td>
</tr>
<tr>
<td>Frobenius norm</td>
<td>$u_{it} \sim$ multivariate $t$--distributed with 8 degrees of freedom</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.87(0.17)</td>
<td>0.92(0.18)</td>
</tr>
<tr>
<td>100</td>
<td>1.56(0.24)</td>
<td>1.66(0.24)</td>
</tr>
<tr>
<td>200</td>
<td>2.16(0.18)</td>
<td>2.32(0.20)</td>
</tr>
</tbody>
</table>

Note: Norm losses are averages over 2,000 replications. Simulation standard deviations are given in parentheses. The MT estimators are defined in Section 3.2.
Table 2: Spectral and Frobenius norm losses for different regularised covariance matrix estimators (T = 100) - Monte Carlo design A

<table>
<thead>
<tr>
<th></th>
<th>N = 30</th>
<th>N = 100</th>
<th>N = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spectral</td>
<td>Frobenius</td>
<td>Spectral</td>
</tr>
<tr>
<td><strong>u_i ~ Gaussian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$MT_{N(N−1)/2}$</td>
<td>1.81(0.54)</td>
<td>3.31(0.42)</td>
<td>2.75(0.50)</td>
</tr>
<tr>
<td>$BL \hat{C}$</td>
<td>5.30(2.16)</td>
<td>7.61(1.23)</td>
<td>8.74(0.06)</td>
</tr>
<tr>
<td>$CL_2$</td>
<td>1.87(0.55)</td>
<td>3.39(0.44)</td>
<td>2.99(0.49)</td>
</tr>
<tr>
<td>$CL \hat{C}$</td>
<td>1.82(0.58)</td>
<td>3.33(0.56)</td>
<td>2.54(0.50)</td>
</tr>
<tr>
<td>$S-MT_{N(N−1)/2}$</td>
<td>3.20(0.79)</td>
<td>4.29(0.64)</td>
<td>5.73(0.34)</td>
</tr>
<tr>
<td>$BL \hat{C}^*$</td>
<td>7.09(1.10)</td>
<td>8.62(0.09)</td>
<td>8.74(0.06)</td>
</tr>
<tr>
<td>$CL \hat{C}^*$</td>
<td>7.05(0.16)</td>
<td>8.58(0.12)</td>
<td>8.71(0.07)</td>
</tr>
<tr>
<td>$LW_\Sigma$</td>
<td>2.99(0.47)</td>
<td>6.49(0.29)</td>
<td>5.20(0.34)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>u_i ~ multivariate t-distributed with 8 degrees of freedom</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$MT_{N(N−1)/2}$</td>
<td>2.16(0.76)</td>
<td>4.03(0.99)</td>
<td>3.43(1.09)</td>
<td>8.43(1.26)</td>
<td>3.97(0.92)</td>
<td>12.66(1.83)</td>
</tr>
<tr>
<td>$BL \hat{C}$</td>
<td>6.90(0.82)</td>
<td>8.75(0.55)</td>
<td>8.74(0.10)</td>
<td>17.26(0.30)</td>
<td>9.00(0.42)</td>
<td>24.93(1.02)</td>
</tr>
<tr>
<td>$CL_2$</td>
<td>2.55(0.93)</td>
<td>4.53(1.00)</td>
<td>4.63(1.11)</td>
<td>10.35(1.48)</td>
<td>5.92(0.81)</td>
<td>16.43(1.74)</td>
</tr>
<tr>
<td>$CL \hat{C}$</td>
<td>2.27(0.76)</td>
<td>4.24(0.94)</td>
<td>3.85(1.51)</td>
<td>9.44(2.33)</td>
<td>5.04(2.04)</td>
<td>15.65(4.71)</td>
</tr>
<tr>
<td>$S-MT_{N(N−1)/2}$</td>
<td>3.18(0.82)</td>
<td>4.68(0.82)</td>
<td>5.75(0.45)</td>
<td>11.33(0.62)</td>
<td>6.41(0.32)</td>
<td>17.10(0.74)</td>
</tr>
<tr>
<td>$BL \hat{C}^*$</td>
<td>7.06(0.13)</td>
<td>8.84(0.30)</td>
<td>8.74(0.10)</td>
<td>17.25(0.31)</td>
<td>8.95(0.08)</td>
<td>24.84(0.55)</td>
</tr>
<tr>
<td>$CL \hat{C}^*$</td>
<td>7.01(0.16)</td>
<td>8.77(0.30)</td>
<td>8.73(0.11)</td>
<td>17.23(0.29)</td>
<td>8.94(0.08)</td>
<td>24.77(0.53)</td>
</tr>
<tr>
<td>$LW_\Sigma$</td>
<td>3.35(0.51)</td>
<td>7.35(0.50)</td>
<td>5.67(0.46)</td>
<td>18.04(0.45)</td>
<td>6.60(0.43)</td>
<td>28.18(0.53)</td>
</tr>
</tbody>
</table>

Note: Norm losses are averages over 100 replications. Simulation standard deviations are given in parentheses. $\hat{C} = \{\hat{C}_{MT_{N(N−1)/2}}, \hat{C}_{BL,\hat{C}}, \hat{C}_{CL,\hat{C}}, \hat{C}_{S-MT_{N(N−1)/2}}, \hat{C}_{BL,\hat{C}^*}, \hat{C}_{CL,\hat{C}^*}, \hat{C}_{LW_\Sigma}\}$. $\hat{C}_{MT_{N(N−1)/2}}$ and $\hat{C}_{S-MT_{N(N−1)/2}}$ are computed using $p = 0.05$. $BL$ is Bickel and Levina universal thresholding, $CL$ is Cai and Liu adaptive thresholding, $\hat{C}_{BL,\hat{C}}$ is based on $\hat{C}$ which is obtained by cross-validation, $\hat{C}_{BL,\hat{C}^*}$ employs the further adjustment to the cross-validation coefficient, $\hat{C}^*$, proposed by Fan, Liao and Mincheva (2013), $\hat{C}_{CL,2}$ is CL’s estimator with $C = 2$ (the theoretical value of $C$), $\hat{C}_{LW_\Sigma}$ 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

<table>
<thead>
<tr>
<th>(N)</th>
<th>(N = 30)</th>
<th>(N = 100)</th>
<th>(N = 200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spectral</td>
<td>Frobenius</td>
<td>Spectral</td>
</tr>
<tr>
<td>(u_i\sim \text{Gaussian})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(MT_{N(N-1)/2})</td>
<td>0.48(0.16)</td>
<td>0.88(0.17)</td>
<td>0.84(0.36)</td>
</tr>
<tr>
<td>(BL_{C^*})</td>
<td>0.91(0.50)</td>
<td>1.35(0.43)</td>
<td>1.40(0.95)</td>
</tr>
<tr>
<td>(CL_2)</td>
<td>0.49(0.17)</td>
<td>0.90(0.18)</td>
<td>1.00(0.48)</td>
</tr>
<tr>
<td>(CL_{C^*})</td>
<td>0.49(0.15)</td>
<td>0.92(0.17)</td>
<td>0.83(0.31)</td>
</tr>
<tr>
<td>(S-MT_{N(N-1)/2})</td>
<td>0.68(0.25)</td>
<td>1.08(0.20)</td>
<td>1.50(0.50)</td>
</tr>
<tr>
<td>(BL_{C^*})</td>
<td>1.19(0.46)</td>
<td>1.63(0.40)</td>
<td>3.32(0.20)</td>
</tr>
<tr>
<td>(CL_{C^*})</td>
<td>1.08(0.46)</td>
<td>1.53(0.46)</td>
<td>3.34(0.15)</td>
</tr>
<tr>
<td>(LW_{C})</td>
<td>1.05(0.13)</td>
<td>2.07(0.10)</td>
<td>2.95(0.26)</td>
</tr>
</tbody>
</table>

\(\sigma_i\sim \text{multivariate } t\text{–distributed with 8 degrees of freedom}\)

<table>
<thead>
<tr>
<th>(N = 30)</th>
<th>(N = 100)</th>
<th>(N = 200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(MT_{N(N-1)/2})</td>
<td>0.65(0.25)</td>
<td>1.13(0.25)</td>
</tr>
<tr>
<td>(BL_{C})</td>
<td>1.36(0.40)</td>
<td>1.84(0.35)</td>
</tr>
<tr>
<td>(CL_2)</td>
<td>0.71(0.29)</td>
<td>1.21(0.30)</td>
</tr>
<tr>
<td>(CL_{C^*})</td>
<td>0.80(0.39)</td>
<td>1.33(0.39)</td>
</tr>
<tr>
<td>(S-MT_{N(N-1)/2})</td>
<td>0.69(0.26)</td>
<td>1.18(0.23)</td>
</tr>
<tr>
<td>(BL_{C^*})</td>
<td>1.49(0.26)</td>
<td>1.98(0.21)</td>
</tr>
<tr>
<td>(CL_{C^*})</td>
<td>1.26(0.40)</td>
<td>1.79(0.40)</td>
</tr>
<tr>
<td>(LW_{C})</td>
<td>1.13(0.15)</td>
<td>2.25(0.11)</td>
</tr>
</tbody>
</table>

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 - $T = 100$

<table>
<thead>
<tr>
<th>$N = 30$</th>
<th>$N = 100$</th>
<th>$N = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Norms</strong></td>
<td><strong>Norms</strong></td>
<td><strong>Norms</strong></td>
</tr>
<tr>
<td><strong>Spectral</strong></td>
<td><strong>Frobenius</strong></td>
<td><strong>Spectral</strong></td>
</tr>
<tr>
<td><strong>Error matrices</strong> $(\mathbf{\Sigma}^{-1} - \hat{\mathbf{\Sigma}}^{-1})$</td>
<td><strong>u$\sim$ Gaussian</strong></td>
<td></td>
</tr>
<tr>
<td>$S-MT_{N(N-1)/2}$</td>
<td>4.42(1.22)</td>
<td>15.62(2.68)</td>
</tr>
<tr>
<td>$BL_{\hat{\mathcal{C}}*}$</td>
<td>$3.8 \times 10^3 (2.4 \times 10^4)$</td>
<td>$19.56 (58.88)$</td>
</tr>
<tr>
<td>$CL_{\hat{\mathcal{C}}*}$</td>
<td>$1.9 \times 10^3 (1.7 \times 10^4)$</td>
<td>$19.56 (58.88)$</td>
</tr>
<tr>
<td>$LW_{\hat{\mathcal{S}}}$</td>
<td>11.03(0.58)</td>
<td>4.26(0.09)</td>
</tr>
<tr>
<td><strong>u$\sim$ multivariate $t$-distributed with 8 degrees of freedom</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S-MT_{N(N-1)/2}$</td>
<td>3.42(1.55)</td>
<td>12.39(3.01)</td>
</tr>
<tr>
<td>$BL_{\hat{\mathcal{C}}*}$</td>
<td>$157.26 (1.0 \times 10^3)$</td>
<td>$6.11 (11.28)$</td>
</tr>
<tr>
<td>$CL_{\hat{\mathcal{C}}*}$</td>
<td>$85.82 (546.85)$</td>
<td>$5.53 (7.84)$</td>
</tr>
<tr>
<td>$LW_{\hat{\mathcal{S}}}$</td>
<td>12.08(1.19)</td>
<td>4.48(0.20)</td>
</tr>
</tbody>
</table>

Note: $\hat{\mathbf{\Sigma}}^{-1} = \{\hat{\mathbf{\Sigma}}_{N(N-1)/2}^{-1}, \hat{\mathbf{\Sigma}}_{BL_{\hat{\mathcal{C}}*}}^{-1}, \hat{\mathbf{\Sigma}}_{CL_{\hat{\mathcal{C}}*}}^{-1}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\mathcal{S}}}}^{-1}\}$. See also the note to Table 2.

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

<table>
<thead>
<tr>
<th>Monte Carlo design A</th>
<th>Monte Carlo design B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>$MT_{N-1}$</td>
</tr>
<tr>
<td><strong>u$\sim$ Gaussian</strong></td>
<td></td>
</tr>
<tr>
<td>30 TPR</td>
<td>0.80</td>
</tr>
<tr>
<td>30 FPR</td>
<td>0.00</td>
</tr>
<tr>
<td>100 TPR</td>
<td>0.69</td>
</tr>
<tr>
<td>100 FPR</td>
<td>0.00</td>
</tr>
<tr>
<td>200 TPR</td>
<td>0.66</td>
</tr>
<tr>
<td>200 FPR</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>u$\sim$ multivariate $t$-distributed with 8 degrees of freedom</strong></td>
<td></td>
</tr>
<tr>
<td>30 TPR</td>
<td>0.80</td>
</tr>
<tr>
<td>30 FPR</td>
<td>0.01</td>
</tr>
<tr>
<td>100 TPR</td>
<td>0.69</td>
</tr>
<tr>
<td>100 FPR</td>
<td>0.00</td>
</tr>
<tr>
<td>200 TPR</td>
<td>0.66</td>
</tr>
<tr>
<td>200 FPR</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: TPR is the true positive and FPR is the false positive rates defined by (16) and (17), 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 $p = 0.01, 0.05, 0.10$ - $T = 100$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$p = 0.01$</th>
<th>$p = 0.05$</th>
<th>$p = 0.10$</th>
<th>$N$</th>
<th>$p = 0.01$</th>
<th>$p = 0.05$</th>
<th>$p = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$MT_{N-1}$</td>
<td>$MT_{N(N-1)/2}$</td>
<td>$MT_{N-1}$</td>
<td>$MT_{N(N-1)/2}$</td>
<td>$MT_{N-1}$</td>
<td>$MT_{N(N-1)/2}$</td>
<td>$MT_{N-1}$</td>
</tr>
<tr>
<td>30</td>
<td>TPR 0.75</td>
<td>0.69</td>
<td>0.80</td>
<td>0.73</td>
<td>0.81</td>
<td>0.75</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>FPR 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>100</td>
<td>TPR 0.65</td>
<td>0.55</td>
<td>0.69</td>
<td>0.59</td>
<td>0.71</td>
<td>0.61</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>FPR 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>100</td>
</tr>
<tr>
<td>200</td>
<td>TPR 0.62</td>
<td>0.51</td>
<td>0.66</td>
<td>0.54</td>
<td>0.68</td>
<td>0.56</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>FPR 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>100</td>
</tr>
</tbody>
</table>

$u_d \sim$ multivariate $t$-distributed with 8 degrees of freedom

<table>
<thead>
<tr>
<th>$N$</th>
<th>$p = 0.01$</th>
<th>$p = 0.05$</th>
<th>$p = 0.10$</th>
<th>$N$</th>
<th>$p = 0.01$</th>
<th>$p = 0.05$</th>
<th>$p = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$MT_{N-1}$</td>
<td>$MT_{N(N-1)/2}$</td>
<td>$MT_{N-1}$</td>
<td>$MT_{N(N-1)/2}$</td>
<td>$MT_{N-1}$</td>
<td>$MT_{N(N-1)/2}$</td>
<td>$MT_{N-1}$</td>
</tr>
<tr>
<td>30</td>
<td>TPR 0.76</td>
<td>0.69</td>
<td>0.80</td>
<td>0.73</td>
<td>0.82</td>
<td>0.75</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>FPR 0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>30</td>
</tr>
<tr>
<td>100</td>
<td>TPR 0.65</td>
<td>0.56</td>
<td>0.69</td>
<td>0.59</td>
<td>0.71</td>
<td>0.61</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>FPR 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>100</td>
</tr>
<tr>
<td>200</td>
<td>TPR 0.62</td>
<td>0.51</td>
<td>0.66</td>
<td>0.55</td>
<td>0.68</td>
<td>0.56</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>FPR 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>200</td>
</tr>
</tbody>
</table>

Note: TPR is the true positive and FPR is the false positive rates defined by (16) and (17), respectively. The TPR and FPR numbers are averages over 2,000 replications.
Table 7: Relative execution time for the different thresholding methods

<table>
<thead>
<tr>
<th></th>
<th>( T = 100 )</th>
<th>( N = 30 )</th>
<th>( N = 100 )</th>
<th>( N = 200 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( MT_{N(N-1)/2} )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>( MT_{N-1} )</td>
<td>1.176</td>
<td>1.239</td>
<td>1.110</td>
<td></td>
</tr>
<tr>
<td>( BL_C )</td>
<td>18.25</td>
<td>102.2</td>
<td>500.9</td>
<td></td>
</tr>
<tr>
<td>( CL_2 )</td>
<td>1.304</td>
<td>5.746</td>
<td>17.39</td>
<td></td>
</tr>
<tr>
<td>( CL_C )</td>
<td>1278</td>
<td>59907</td>
<td>412293</td>
<td></td>
</tr>
</tbody>
</table>

Note: All times are relative to the \( MT_{N(N-1)/2} \) estimator. See Table 2 for a note on the thresholding methods.

Appendix: Mathematical proofs of theorems for the MT estimator

In what follows we suppress subscript \( MT \) from \( \hat{R}_{MT} \) for notational convenience. All statements and proofs of technical lemmas are relegated to the Supplementary Appendix A.

Proof of Theorem 1. Consider the spectral norm,

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

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

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

Also

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

Hence,

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

and

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

For any given \( i \), where \( i = 1, 2, ..., N \), and taking expectations, we obtain

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

(23)
or

\[ \sum_j E \left( |\hat{\rho}_{ij,T} - \rho_{ij}| \right) \leq A_i + B_i + C_i, \]

where

\[ A_i = \sum_j E \left[ \rho_{ij} \left| I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) |\rho_{ij} \neq 0 \right) \right. \right], \]

\[ B_i = \sum_j E \left[ (\hat{\rho}_{ij,T} - \rho_{ij}) I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) |\rho_{ij} \neq 0 \right) \right], \]

\[ C_i = \sum_j E \left[ \hat{\rho}_{ij,T} \left| I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) |\rho_{ij} = 0 \right) \right. \right]. \]

Consider now the orders of these three terms \( A_i, B_i, \) and \( C_i \) in turn, starting with \( A_i \).

Since under Assumption 2, \( 0 < \rho_{\min} < |\rho_{ij}| < \rho_{\max} < 1 \), then uniformly over all \( i \),

\[ A_i \leq m_N \rho_{\max} \sup_{ij} E \left[ I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) |\rho_{ij} \neq 0 \right) \right], \]

\[ = m_N \rho_{\max} \sup_{ij} \Pr \left[ I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) |\rho_{ij} \neq 0 \right) \right], \]

and using equation (A.12) of Lemma 6 we have

\[ A_i \leq m_N \rho_{\max} \sup_{ij} K e^{\frac{1}{2} \left[ \frac{c_p(N) - \sqrt{T}|\rho_{ij}|^2}{K_v(\theta_{ij})} \right]} \left[ 1 + o(1) \right], \]

\[ \leq m_N \rho_{\max} \sup_{ij} K e^{\frac{1}{2} \left[ \frac{\rho_{\min} - c_p(N)}{K_v(\theta_{ij})} \right]} \left[ 1 + o(1) \right]. \]

Recalling that \( \sup_{ij} K_v(\theta_{ij}) < K \) and by assumption \( \rho_{\min} > 0 \), it then readily follows that \( A_i \) is of order \( O(e^{-T}) \) so that \( A_i \) is uniformly bounded for all \( i \) and \( N \) (recalling that \( m_N \) is bounded in \( N \)), and tends to zero as \( N \) and \( T \to \infty \), jointly.

Consider now \( B_i \) and note that since \( \hat{\rho}_{ij,T} = \omega_{ij,T} z_{ij,T} + \rho_{ij,T} \) (to simplify the notation we use \( \omega_{ij,T} \) and \( \rho_{ij,T} \) for \( Var (\hat{\rho}_{ij,T}) \) and \( E (\hat{\rho}_{ij,T}) \), respectively) we have the following inequality, \( B_i \leq B_{i1} + B_{i2} \), where

\[ B_{i1} = \sum_{j, \rho_{ij} \neq 0} E \left[ \omega_{ij,T} z_{ij,T} I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) |\rho_{ij} \neq 0 \right) \right], \]

\[ B_{i2} = \sum_{j, \rho_{ij} \neq 0} E \left[ \rho_{ij,T} - \rho_{ij} I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) |\rho_{ij} \neq 0 \right) \right]. \]

Using (8) and (9)

\[ \omega_{ij,T} = \frac{K_v^{1/2}(\theta_{ij})}{T^{1/2}} + O \left( T^{-3/2} \right), \]

\[ \rho_{ij,T} - \rho_{ij} = \frac{K_{m}(\theta_{ij})}{T} + O \left( T^{-2} \right). \]

25
Hence (noting that \( m_N \) is bounded in \( N \) and \( T \), and \( \omega_{ij,T} > 0 \), \( \mathcal{B}_{i1} \) becomes

\[
\mathcal{B}_{i1} \leq \sum_{j, \rho_{ij} \neq 0} \omega_{ij,T} E \left[ |z_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) |\rho_{ij} \neq 0 \right) \right]
\]

\[
\leq \frac{m_N}{T^{1/2}} \left[ \sup_{ij} K_{ij}^{1/2}(\theta_{ij}) \sup_{ij} \left\{ 1 - E \left[ |z_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| \leq c_p(N) |\rho_{ij} \neq 0 \right) \right] \right\} \right] + O \left( \frac{m_N}{T^{3/2}} \right).
\]

By the Cauchy-Schwarz inequality, we have

\[
E \left[ |z_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| \leq c_p(N) |\rho_{ij} \neq 0 \right) \right] \leq \left[ E \left( |z_{ij,T}|^2 \right) \right]^{1/2} \left\{ E \left[ I \left( |\sqrt{T} \hat{\rho}_{ij,T}| \leq c_p(N) |\rho_{ij} \neq 0 \right) \right] \right\}^{1/2} < K,
\]

since

\[
\left| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| \leq c_p(N) |\rho_{ij} \neq 0 \right) \right|^2 = I \left( |\sqrt{T} \hat{\rho}_{ij,T}| \leq c_p(N) |\rho_{ij} \neq 0 \right),
\]

the second moment of \( z_{ij,T} \) exists (see Proposition 1), and \( E I \left( |\sqrt{T} \hat{\rho}_{ij,T}| \leq c_p(N) |\rho_{ij} \neq 0 \right) \) is bounded. Further, \( \sup_{ij} K_i(\theta_{ij}) < K \), hence it readily follows that \( \mathcal{B}_{i1} \) is of order \( O \left( \frac{m_N}{T^{3/2}} \right) \), uniformly for all \( i \).

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

\[
\mathcal{B}_{i2} = \sum_{j, \rho_{ij} \neq 0} |\rho_{ij,T} - \rho_{ij}| E \left[ I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) |\rho_{ij} \neq 0 \right) \right]
\]

\[
\leq m_N \left[ \frac{|K_m(\theta_{ij})|}{T} + O(T^{-2}) \right] = O \left( \frac{m_N}{T} \right),
\]

uniformly for all \( i \). Overall, therefore, \( \mathcal{B}_i = O \left( \frac{m_N}{T^{3/2}} \right) \) uniformly for all \( i \).

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

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

\[
\mathcal{C}_{i2} = \sum_{j, \rho_{ij} = 0} E \left[ |\rho_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) |\rho_{ij} = 0 \right) \right].
\]

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

\[
\mathcal{C}_{i2} = \sum_{j, \rho_{ij} = 0} E \left[ |\rho_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) |\rho_{ij} = 0 \right) \right]
\]

\[
= \sum_{j, \rho_{ij} = 0} |\rho_{ij,T}| E \left[ I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) |\rho_{ij} = 0 \right) \right]
\]

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

and \( E I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) |\rho_{ij} = 0 \right) \leq 1 \). Using (25) and equation (A.11) of Lemma 6 (and evaluating these expressions under \( \rho_{ij} = 0 \)) we have

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

26
where \( \kappa_{ij} = \left[ \mu_{ij}(2, 2) \right| \rho_{ij} = 0 \] and \( \psi_{ij} = \left[ \mu_{ij}(3, 1) + \mu_{ij}(1, 3) \right]/2 \). Strictly speaking, \( \mu_{ij}(3, 1) \) and \( \mu_{ij}(1, 3) \) in the above expression are also defined under \( \rho_{ij} = 0 \), but since \( \psi_{ij} \) do not enter the asymptotic results we do not make this conditioning explicit to simplify the notation. Therefore, so long as \( N/T \) tends to a finite constant then \( C_{i2} \to 0 \) as \( N \) and \( T \to \infty \), uniformly for all \( i \), since \( \psi_{ij}^2 \) and \( \kappa_{ij} \) are bounded and \( c_p(N) \to \infty \).

Finally, considering \( C_{i1} \) we note that (since \( \omega_{ij,T} > 0 \)),

\[
C_{i1} = \sum_{j, \hat{\rho}_{ij}=0} E \left[ |\omega_{ij,T} z_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) | \rho_{ij} = 0 \right) \right] \\
= \sum_{j, \hat{\rho}_{ij}=0} \omega_{ij,T} E \left[ |z_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) | \rho_{ij} = 0 \right) \right] \\
\leq \frac{(N - m_N - 1)}{T^{1/2}} \left[ \sup_{ij} K_1^{1/2}(\theta_{ij}) \right] \\
\times \sup_{ij} E \left[ |z_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) | \rho_{ij} = 0 \right) \right] + O \left( \frac{(N - m_N - 1)}{T^{3/2}} \right).
\]

and by the Cauchy-Schwarz inequality,

\[
E \left[ |z_{ij,T}| I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) | \rho_{ij} = 0 \right) \right] \\
\leq \left[ E \left( |z_{ij,T}|^2 \right) \right]^{1/2} \left[ E \left[ I \left( |\sqrt{T} \hat{\rho}_{ij,T}| > c_p(N) | \rho_{ij} = 0 \right) \right] \right]^{1/2} < K,
\]

since \( E \left[ I \left( |\sqrt{T} \hat{\rho}_{ij,T}| \leq c_p(N) | \rho_{ij} = 0 \right) \right] \leq 1 \) and the second moment of \( z_{ij,T} \) exists. Hence, \( C_{i1} \) is bounded as \( N \) and \( T \to \infty \), uniformly for all \( i \), so long as \( N/\sqrt{T} \to 0 \).

Collecting the results for the orders of convergence of \( A_i, B_i, \) and \( C_i \) given above, overall we obtain a convergence rate of order \( O \left( \frac{m_N}{T^{3/2}} \right) \) uniformly for all \( i \), where \( i = 1, 2, ..., N \). Therefore, (13) follows as required.

**Proof of Theorem 2.** Consider the squared Frobenius norm,

\[
\| \tilde{R} - R \|_F^2 = \sum_{i \neq j} (\tilde{\rho}_{ij,T} - \rho_{ij})^2,
\]

and recall that

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

Hence

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

However,

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

27
\[
1 - I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right) \right] = 1 - I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right).
\]
Therefore, we have
\[
\sum_{i \neq j} \sum_{\rho_{ij} \neq 0} (\hat{\rho}_{ij,T} - \rho_{ij})^2 = \sum_{i \neq j} (\hat{\rho}_{ij,T} - \rho_{ij})^2 I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right) + \sum_{i \neq j} \rho_{ij}^2 \left[ 1 - I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right) \right] = \sum_{i \neq j} (\hat{\rho}_{ij,T} - \rho_{ij})^2 I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right) + \sum_{i \neq j} \rho_{ij}^2 I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) \right),
\]
which can be decomposed as
\[
\sum_{i \neq j} E (\hat{\rho}_{ij,T} - \rho_{ij})^2 = A + B + C,
\] (26)
where
\[
A = \sum_{i \neq j, \rho_{ij} \neq 0} \rho_{ij}^2 E \left[ I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) \right) \right], \quad B = \sum_{i \neq j, \rho_{ij} \neq 0} E \left[ (\hat{\rho}_{ij,T} - \rho_{ij})^2 I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right) \right], \quad C = \sum_{i \neq j, \rho_{ij} = 0} E \left[ \rho_{ij,T} I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right) \right].
\]

Consider now the orders of the above three terms in turn, starting with \( A \). Since under Assumption 2, \( 0 < \rho_{\min} < |\rho_{ij}| < \rho_{\max} < 1 \), then
\[
A \leq \rho_{\max}^2 N m_N \sup_{ij} E \left[ I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) \right) \right] = \rho_{\max}^2 N m_N \sup_{ij} \Pr \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) \right),
\]
and using Lemma 6, equation (A.12), we have
\[
A \leq \rho_{\max}^2 N m_N \sup_{ij} K e^{- \frac{\left( c_p(N) - \sqrt{T} \rho_{ij} \right)^2}{c_p(N)}} \left[ 1 + o(1) \right].
\]

Recalling that \( \sup_{ij} K_{\nu}(\theta_{ij}) < K \) and by assumption \( \rho_{\min} > 0 \), it then readily follows that \( A \) is of order \( O(N e^{-T}) \) so that \( A \rightarrow 0 \), as \( N \) and \( T \rightarrow \infty \). Note that this result does not require \( N/T \rightarrow 0 \), and holds even if \( N/T \) tends to a fixed constant.
Consider now $B$. Recalling that $\hat{\rho}_{ij,T} = \omega_{ij,T}z_{ij,T} + \rho_{ij,T}$ we have the following decomposition of $B$, $B = B_1 + B_2 + 2B_3$, where

$$B_1 = \sum_{i \neq j, \rho_{ij} \neq 0} \omega_{ij,T}^2 E \left[ z_{ij,T}^2 \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} \neq 0 \right) \right],$$

$$B_2 = \sum_{i \neq j, \rho_{ij} \neq 0} (\rho_{ij,T} - \rho_{ij})^2 E \left[ \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} \neq 0 \right) \right],$$

$$B_3 = \sum_{i \neq j, \rho_{ij} \neq 0} (\rho_{ij,T} - \rho_{ij}) \omega_{ij,T} E \left[ z_{ij,T} \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} \neq 0 \right) \right].$$

Again, using (8) and (9)

$$\omega_{ij,T}^2 = \frac{K_v(\theta_{ij})}{T} + O(T^{-2}), \quad (27)$$

$$(\rho_{ij,T} - \rho_{ij})^2 = \frac{K_m^2(\theta_{ij})}{T^2} + O(T^{-3}), \quad (28)$$

$$(\rho_{ij,T} - \rho_{ij}) \omega_{ij,T} = \frac{K_v^{1/2}(\theta_{ij})K_m(\theta_{ij})}{T^{3/2}} + O(T^{-5/2}). \quad (29)$$

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

$$B_1 = \sum_{i \neq j, \rho_{ij} \neq 0} \omega_{ij,T}^2 E \left[ z_{ij,T}^2 \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} \neq 0 \right) \right] \leq \frac{N m_N}{T} \left[ \sup_{ij} K_v(\theta_{ij}) \right] \sup_{ij} \left\{ 1 - E \left[ z_{ij,T}^2 \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) | \rho_{ij} \neq 0 \right) \right] \right\} + O \left( \frac{m_N N}{T^2} \right).$$

Since $\sup_{ij} K_v(\theta_{ij})$ and $E \left[ z_{ij,T}^2 \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) | \rho_{ij} \neq 0 \right) \right]$ are bounded, it then readily follows that $B_1$ is at most $O \left( \frac{N m_N}{T^2} \right)$. In fact $\lim_{T \to \infty} E \left[ z_{ij,T}^2 \left( I \left| \sqrt{T}\hat{\rho}_{ij} \right| \leq c_p(N) | \rho_{ij} \neq 0 \right) \right] = 0$ if $\sqrt{T}\rho_{\min} - c_p(N) \to \infty$, as $N$ and $T \to \infty$, which can be easily shown.

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

$$B_2 = \sum_{i \neq j, \rho_{ij} \neq 0} (\rho_{ij,T} - \rho_{ij})^2 E \left[ \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} \neq 0 \right) \right] \leq N m_N \left[ \frac{K_m^2(\theta_{ij})}{T^2} + O(T^{-3}) \right] = O \left( \frac{N m_N}{T^2} \right),$$

and

$$B_3 = \sum_{i \neq j, \rho_{ij} \neq 0} (\rho_{ij,T} - \rho_{ij}) \omega_{ij,T} E \left[ z_{ij,T} \left( I \left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} \neq 0 \right) \right]$$

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

$$= - \sum_{i \neq j, \rho_{ij} \neq 0} (\rho_{ij,T} - \rho_{ij}) \omega_{ij,T} E \left[ z_{ij,T} I \left( \left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) | \rho_{ij} \neq 0 \right) \right]. \quad (30)$$
Also, from Lemma 4
\[
\lim_{T \to \infty} E \left[ z_{ij,T} I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) \right| \rho_{ij} \neq 0 \right) \right] = \lim_{T \to \infty} E \left[ z I \left( L_{ij,T} \leq z \leq U_{ij,T} \right| \rho_{ij} \neq 0 \right) ,
\]
and from Lemma 2
\[
E \left[ z I \left( L_{ij,T} \leq z \leq U_{ij,T} \right| \rho_{ij} \neq 0 \right) = \phi \left( \frac{-c_p(N) - \sqrt{T} \rho_{ij} + O \left( \frac{1}{\sqrt{T}} \right)}{\sqrt{K_v(\theta_{ij})} + O \left( \frac{1}{T} \right)} \right) - \phi \left( \frac{c_p(N) - \sqrt{T} \rho_{ij} + O \left( \frac{1}{\sqrt{T}} \right)}{\sqrt{K_v(\theta_{ij})} + O \left( \frac{1}{T} \right)} \right),
\]
which is bounded in \( N \) and \( T \). Since \( \sqrt{T} \rho_{\min} - c_p(N) \to \infty \) as \( N \) and \( T \to \infty \), it is easily seen that \( \lim_{T,N \to \infty} E \left[ z I \left( L_{ij,T} \leq z \leq U_{ij,T} \right| \rho_{ij} \neq 0 \right) \right] = 0 \). Hence, using (29) and noting that \( K_{1/2}^{1/2}(\theta_{ij})K_m(\theta_{ij}) \) is bounded in \( T \) we have
\[
B_3 \leq K \sum_{i \neq j, \rho_{ij} \neq 0} \left| (\rho_{ij,T} - \rho_{ij}) \omega_{ij,T} \right| = O \left( \frac{N m_N}{T^{3/2}} \right).
\]
Overall, therefore, \( B = O \left( \frac{N m_N}{T} \right) \).

Consider now the following decomposition of \( C \), in (26):
\[
C = \sum_{i \neq j, \rho_{ij} = 0} \sum_{i \neq j} \sum_{\rho_{ij} = 0} E \left[ \rho_{ij,T}^2 I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right| \rho_{ij} = 0 \right] \]
\[
= \sum_{i \neq j, \rho_{ij} = 0} \omega_{ij,T}^2 E \left[ z_{ij,T}^2 I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right| \rho_{ij} = 0 \right] \]
\[
+ \sum_{i \neq j, \rho_{ij} = 0} \rho_{ij,T}^2 E \left[ I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right| \rho_{ij} = 0 \right] \]
\[
+ 2 \sum_{i \neq j, \rho_{ij} = 0} \rho_{ij,T} \omega_{ij,T} E \left[ z_{ij,T} I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right| \rho_{ij} = 0 \right] \]
\[
= C_1 + C_2 + C_3.
\]
Starting with the simpler terms, we first note that
\[
C_2 = \sum_{i \neq j, \rho_{ij} = 0} \rho_{ij,T}^2 E \left[ I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right| \rho_{ij} = 0 \right] \]
\[
\leq N(N - m_N - 1) \sup_{i,j} (\rho_{ij,T}^2 \left| \rho_{ij} = 0 \right) \sup_{i,j} E \left[ I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \right| \rho_{ij} = 0 \right] ,
\]
and \( \sup_{i,j} E \left[ I \left( \left| \sqrt{T} \hat{\rho}_{ij} \right| > c_p(N) \right| \rho_{ij} = 0 \right] \leq 1 \). Using (8) and equation (A.11) of Lemma 6 (and evaluating these expressions under \( \rho_{ij} = 0 \)) we have
\[
C_2 \leq K \frac{N(N - m_N - 1) \sup_{i,j} (\psi_{ij}^2 + O(T^{-1}))}{T^2} e^{-\frac{1}{2} \frac{c_p^2(N)}{\sup_{i,j} \psi_{ij}} \left[ 1 + o(1) \right] ,}
\]
30
where $\kappa_{ij} = [\mu_{ij}(2, 2) | \rho_{ij} = 0]$, and $\psi_{ij} = [\mu_{ij}(3, 1) + \mu_{ij}(1, 3)] / 2$. Therefore, so long as $N/T$ tends to a finite constant then $C_2 \to 0$ as $N$ and $T \to \infty$, since $\psi_{ij}^2$ and $\kappa_{ij}$ are bounded and $c_p(N) \to \infty$.

Similarly
\[
C_3 = \sum_{i \neq j, \rho_{ij} = 0} \rho_{ij,T} \omega_{ij,T} E \left[ z_{ij,T} I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} = 0 \right) \right] \\
= -\sum_{i \neq j, \rho_{ij} = 0} \rho_{ij,T} \omega_{ij,T} E \left[ z_{ij,T} I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) | \rho_{ij} = 0 \right) \right] \\
\leq \frac{N(N - m_N - 1)}{T^{3/2}} \sup_{ij} \left( |\psi_{ij}| + O(T^{-1}) \right) \sup_{ij} \left( \sqrt{\kappa_{ij} + O(T^{-1})} \right) \\
\times \sup_{ij} E \left[ z_{ij,T} I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) | \rho_{ij} = 0 \right) \right].
\]

But using Lemma 4, Lemma 2 and (31) and evaluating the relevant expressions under $\rho_{ij} = 0$, we have
\[
\lim_{T,N \to \infty} E \left[ z_{ij,T} I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| \leq c_p(N) | \rho_{ij} = 0 \right) \right] \\
= \lim_{T,N \to \infty} E \left[ z I \left( L_{ij,T} \leq z \leq U_{ij,T} | \rho_{ij} = 0 \right) \right] \\
= \lim_{N,T \to \infty} \phi \left( -c_p(N) + \frac{\psi_{ij} \sqrt{T}}{\sqrt{\kappa_{ij}}} + O(T^{-3/2}) \right) - \lim_{N,T \to \infty} \phi \left( c_p(N) + \frac{\psi_{ij} \sqrt{T}}{\sqrt{\kappa_{ij}}} + O(T^{-3/2}) \right) = 0.
\]

Hence, $C_3 \to 0$ as $N$ and $T \to \infty$, so long as $N/\sqrt{T} \to 0$, since $c_p(N) \to \infty$ with $N$.

Finally, considering $C_1$ we note that
\[
C_1 = \sum_{i \neq j, \rho_{ij} = 0} \omega_{ij,T}^2 E \left[ z_{ij,T}^2 I \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) | \rho_{ij} = 0 \right) \right] \\
= \sum_{i \neq j, \rho_{ij} = 0} \omega_{ij,T}^2 E \left[ z_{ij,T}^2 \left[ 1 - I \left( L_{ij,T} \leq z_{ij,T} \leq U_{ij,T} | \rho_{ij} = 0 \right) \right] \right] \\
\leq \frac{N(N - m_N - 1)}{T} \sup_{ij} \left[ \kappa_{ij} + O \left( \frac{1}{T} \right) \right] \\
\times \sup_{ij} E \left[ z_{ij,T}^2 \left[ 1 - I \left( L_{ij,T} \leq z_{ij,T} \leq U_{ij,T} | \rho_{ij} = 0 \right) \right] \right].
\]

But using Lemma 4
\[
\lim_{T \to \infty} E \left[ z_{ij,T}^2 \left[ 1 - I \left( L_{ij,T} \leq z_{ij,T} \leq U_{ij,T} | \rho_{ij} = 0 \right) \right] \right] = \lim_{T \to \infty} E \left[ z^2 \left[ 1 - I \left( L_{ij,T} \leq z \leq U_{ij,T} | \rho_{ij} = 0 \right) \right] \right] = 0
\]
and then Lemma 2
\[
E \left\{ z^2 \left[ 1 - I \left( L_{ij,T} \leq z \leq U_{ij,T} | \rho_{ij} = 0 \right) \right] \right\} = 1 - E \left[ z^2 I \left( L_{ij,T} \leq z \leq U_{ij,T} | \rho_{ij} = 0 \right) \right]
\]
\[
= 1 - \{ \Phi \left[ U_{ij,T}(0) \right] - \Phi \left[ L_{ij,T}(0) \right] + L_{ij,T}(0) \phi \left( L_{ij,T}(0) \right) - U_{ij,T}(0) \phi \left( U_{ij,T}(0) \right) \}
\]
\[
= \Phi \left[ -U_{ij,T}(0) \right] + \Phi \left[ L_{ij,T}(0) \right] + U_{ij,T}(0) \phi \left[ U_{ij,T}(0) \right] - L_{ij,T}(0) \phi \left[ L_{ij,T}(0) \right],
\]

31
where \( U_{ij,T}(0) \) and \( L_{ij,T}(0) \) are given by (A.19) which we reproduce here for convenience:

\[
U_{ij,T}(0) = \frac{c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}}, \quad L_{ij,T}(0) = \frac{-c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}}.
\]

Since \(|\psi_{ij}| < K\), then there exist \( N_0 \) and \( T_0 \) such that for \( N > N_0 \) and \( T > T_0 \), \( c_p(N) - \frac{|\psi_{ij}|}{\sqrt{T}} > 0 \), and using Lemma 5 (also see (A.23) and (A.24) of Lemma 6), we have

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

where

\[
D_{1,ij} = \frac{1}{2} e^{-\frac{1}{4} \left( \frac{c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right)^2} + \frac{1}{2} e^{-\frac{1}{4} \left( \frac{c_p(N) - \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right)^2},
\]

and

\[
D_{2,ij} = \left[ \frac{c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right]^{-\frac{1}{4} e^{-\frac{1}{2} \left( \frac{c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right)^2}} - \left[ \frac{-c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right]^{-\frac{1}{4} e^{-\frac{1}{2} \left( \frac{c_p(N) - \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right)^2}}.
\]

Then, for \( N D_{1,ij} \) we have

\[
\lim_{N,T \to \infty} N D_{1,ij} = \lim_{N \to \infty} \left[ e^{-\frac{1}{2} \left( \frac{c_p^2(N)}{\kappa_{ij} + \ln(N)} \right)} \right] = \lim_{N \to \infty} \left[ e^{-\frac{1}{2} \left( \frac{c_p^2(N)}{2 \ln(N)} - \kappa_{ij} \right)} \right].
\]

Since \( \kappa_{ij} > 0 \), then \( N D_{1,ij} \) tends to a finite constant or zero if \( \lim_{N \to \infty} \left( \frac{c_p^2(N)}{2 \ln(N)} \right) \geq \kappa_{ij} \). But using (A.6) of Lemma 3, we have

\[
\frac{\ln[f(N)] - \ln(p)}{\ln(N)} \geq \frac{c_p^2(N)}{2 \ln(N)} \geq \kappa_{\max},
\]

where \( \kappa_{\max} = \sup_{ij} (\kappa_{ij}) \). Next, for \( N D_{2,ij} \) we have

\[
N D_{2,ij} = \left[ \frac{N c_p(N) + \frac{N \psi_{ij}}{\sqrt{T}} + O\left(NT^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right]^{-\frac{1}{4} e^{-\frac{1}{2} \left( \frac{c_p(N) + \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right)^2}} - \left[ \frac{-N c_p(N) + \frac{N \psi_{ij}}{\sqrt{T}} + O\left(NT^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right]^{-\frac{1}{4} e^{-\frac{1}{2} \left( \frac{c_p(N) - \frac{\psi_{ij}}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij} + O(T^{-1})}} \right)^2}}.
\]
or

\[ N \, D_{2,ij} = \left[ 1 + \frac{\psi_{ij}}{c_p(N)\sqrt{T}} + O \left( N^{-1}T^{-3/2} \right) \right] \frac{1}{\sqrt{\kappa_{ij} + O(T^{-1})}} \left\{ \frac{-\ln(\kappa_{ij})}{\kappa_{ij}} \left[ \frac{\kappa_{ij} - \ln(\kappa_{ij})}{\kappa_{ij}} \right] \right\} \left\{ \frac{-\ln(\kappa_{ij})}{\kappa_{ij}} \left[ \frac{\kappa_{ij} - \ln(\kappa_{ij})}{\kappa_{ij}} \right] \right\} \left\{ \frac{-\ln(\kappa_{ij})}{\kappa_{ij}} \left[ \frac{\kappa_{ij} - \ln(\kappa_{ij})}{\kappa_{ij}} \right] \right\} \].

Then \( N \, D_{2,ij} \) tends to a finite constant for all \( i \) and \( j \) as long as \( \frac{\ln(c_p(N))}{\ln(N)} \to c \). Hence, for \( N/T \) tending to a constant and using the above results in (32) we have

\[ C_1 \leq \frac{(N - m_N - 1)}{T} \sup_{ij} \left[ \kappa_{ij} + O(T^{-1}) \right] \sup_{ij} (N D_{2,ij}). \]

Hence, \( C_1 \) must be at most \( O(N/T) \), since by assumption \( \lim_{N \to \infty} \frac{\ln(f(N))}{\ln(N)} \geq \kappa_{\max} \).

Collecting the results for the orders of convergence of \( C_1, C_2, \) and \( C_3 \) given above, and those of \( A \) and \( B \), overall we obtain a convergence rate of order \( O(m_N N/T) \), and (15) follows as desired. 

**Proof of Theorem 3.** Consider first the \( FPR \) statistic given by (17) which can be written equivalently as

\[ FPR = \left| FPR \right| = \sum_{i \neq j} \sum_{i \neq j} I \left( \left| \sqrt{T} \hat{p}_{ij,T} \right| > c_p(N) | \hat{p}_{ij} = 0 \right) \frac{1}{N(N - m_N - 1)}. \]  

(34)

Note that the elements of \( FPR \) are either 0 or 1 and so \( \left| FPR \right| = FPR \).

Taking the expectation of (34) we have

\[ E \left| FPR \right| = \sum_{i \neq j} \sum_{i \neq j} \text{Pr} \left( \left| \sqrt{T} \hat{p}_{ij,T} \right| > c_p(N) | \hat{p}_{ij} = 0 \right) \frac{1}{N(N - m_N - 1)}. \]

But using Lemma 6 (equation (A.11)) we have (recall that \( \kappa_{ij} = \left[ \mu_{ij}(2,2) | \hat{p}_{ij} = 0 \right] \))

\[ E \left| FPR \right| \leq K \sum_{i \neq j} e^{-\frac{1-\epsilon}{2} \frac{\epsilon_p^2(N)}{\kappa_{ij}}} \frac{1 + o(1)}{N(N - m_N - 1)} \leq Ke^{-\frac{1-\epsilon}{2} \frac{\epsilon_p^2(N)}{\kappa_{\max}}} \frac{1 + o(1)}{N(N - m_N - 1)} \]

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

\[ P(|FPR| > \delta) \leq \frac{E(|FPR|)}{\delta} \leq K \frac{e^{-\frac{1-\epsilon}{2} \frac{\epsilon_p^2(N)}{\kappa_{\max}}} [1 + o(1)]}{\delta}, \]
for some $\delta > 0$. Therefore, $\lim_{N,T \to \infty} P(|FPR| > \delta) = 0$, 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 (16) and note that

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

Hence

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

Since $|X| = X$, then

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

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

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

and

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

(35)

However, using (A.25), (A.26) and (A.27) of Lemma 6 we have

$$\Pr \left( \left| \sqrt{T} \hat{\rho}_{ij,T} \right| < c_p(N)|\rho_{ij} \neq 0 \right)$$

$$= F_{ij,T} \left( \frac{c_p(N) - \sqrt{T} \rho_{ij} - \frac{K_u(\theta_{ij})}{\sqrt{T}} + O(T^{-3/2})}{\sqrt{K_v(\theta_{ij}) + O(T^{-1})}} \right)$$

$$- F_{ij,T} \left( \frac{-c_p(N) - \sqrt{T} \rho_{ij} - \frac{K_u(\theta_{ij})}{\sqrt{T}} + O(T^{-3/2})}{\sqrt{K_v(\theta_{ij}) + O(T^{-1})}} \right).$$

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 \( c_p(N) - \sqrt{T} |\rho_{ij}| \to -\infty \), for all \( \rho_{ij} \neq 0 \), or equivalently if \( \sqrt{T} \rho_{\min} - c_p(N) \to \infty \), where \( \rho_{\min} = \min_{ij} |\rho_{ij}| \) for \( \rho_{ij} \neq 0 \). 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 \( T \), since by assumption there exists \( N_0 \) and \( T_0 \) such that for all \( N > N_0 \) and \( T > T_0 \), \( \rho_{\min} - c_p(N)/\sqrt{T} > 0 \), and \( c_p(N)/\sqrt{T} \to 0 \). The latter is ensured since by assumption \( \ln f(N)/T \to 0 \) (see also Lemma 3). Using these results in (35) it now follows that \( \lim_{N,T \to \infty} P(|TPR - 1| > \delta) \to 0 \), as required.

References


