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

Natalia Bailey
Queen Mary, University of London

M. Hashem Pesaran
USC and Trinity College, Cambridge

L. Vanessa Smith
University of York

January 22, 2015

Abstract

This paper proposes a novel regularisation method for the estimation of large covariance matrices, using insights from the multiple testing (MT) literature. The method tests the statistical significance of individual pair-wise correlations and sets to zero those elements that are not statistically significant, taking account of the multiple testing nature of the problem. The procedure is straightforward to implement and is readily adapted to deal with non-Gaussian observations. By using the inverse of the normal distribution at a predetermined significance level, it circumvents the challenge of evaluating the theoretical constant arising in the rate of convergence of existing thresholding estimators, and hence does not require cross-validation. We compare the small sample performance of the proposed MT estimator to a number of other regularisation techniques in the literature using Monte Carlo experiments. We find that the MT estimator performs well and tends to outperform the other estimators, particularly when the cross-sectional dimension, N, is larger than the time series dimension, T. If the inverse covariance matrix is also of interest, then we propose a shrinkage version of the MT estimator that ensures positive definiteness.

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 large 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); Pesaran and Yamagata (2012)) when the number of assets is large. In global macro-econometric modelling with many domestic and foreign channels of interactions, large 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, high-dimensional covariance matrices are required when inferring large-scale 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.

Assuming that the $N \times N$ dimensional population covariance matrix, $\mathbf{\Sigma}$, is invertible, one way of obtaining a suitable estimator is to appropriately restrict the off-diagonal elements of its sample estimate denoted by $\hat{\mathbf{\Sigma}}$. Numerous methods have been developed to address this challenge, predominantly in the statistics literature. See Pourahmadi (2011) for an extensive review and references therein. Some approaches are regression-based and make use of suitable decompositions of $\mathbf{\Sigma}$ such as the Cholesky decomposition (see Pourahmadi (1999, 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 approaches 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 $\hat{\mathbf{\Sigma}}$ can either be set to their sample estimates or can be adjusted downward. This relates to the concepts of 'hard' and 'soft' thresholding, respectively. The thresholding approach traditionally assumes that the underlying (true) covariance matrix is *sparse*, where sparseness is loosely defined as the presence of a sufficient number of zeros on each row of $\mathbf{\Sigma}$ such that it is absolute summable row (column)-wise. 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 thresholding method retains symmetry of the sample covariance matrix but does not necessarily deliver a positive definite estimator of $\mathbf{\Sigma}$ if N is large relative to T. The main difficulty in applying this approach lies in the estimation of the thresholding parameter. The method of cross-validation is primarily used for this purpose which is computationally intensive 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 $\mathbf{\Sigma}$ is defined as a weighted average of the sample covariance matrix and an invertible covariance matrix estimator known as the shrinkage target. 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 and produce estimators that are positive definite by construction, 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).

In this paper, we propose an alternative thresholding procedure using a multiple testing (MT) estimator which is simple to apply, does not require cross-validation, and is readily adapted to deal with non-Gaussian observations. It makes use of insights from the multiple testing literature and sets to zero the off-diagonal elements of $\mathbf{\Sigma} = (\sigma_{ij})$ that are statistically insignificant. The nominal significance level of the test, p, is suitably adjusted for the cross section dimension, N, depending on whether $\sigma_{ij} = 0$ implies independence. The procedure is shown to be equivalent to the application of the multiple testing procedure due to Bonferroni (1935) to the individual rows of $\mathbf{\Sigma}$, separately, when $\sigma_{ij} = 0$ implies independence, and to all distinct non-diagonal elements of $\mathbf{\Sigma}$, if $\sigma_{ij} = 0$ does not imply independence. It is shown that the MT estimator of $\mathbf{R}$, the correlation matrix corresponding to $\mathbf{\Sigma}$, converges to its true value at the rate of $O_p\left(\sqrt{m_N N/T}\right)$ under the Frobenius norm, where T is the number of observations, and m_N is the maximum number of non-zero elements in the off-diagonal rows of $\boldsymbol{R}$. The MT estimator also consistently recovers the support of the true covariance matrix irrespective of whether $\sigma_{ij} = 0$ implies independence or not.

Since traditional thresholding, including the multiple testing approach, does not necessarily lead to a positive definite matrix, a shrinkage version of the MT estimator (denoted by S-MT) is also proposed and is shown to be positive definite.

The small sample performance of the MT estimator is investigated using a Monte Carlo simulation study, and its properties are compared to a number of extant regularisation 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 when compared with the widely used regularisation methods considered in the literature, especially when N is large relative to T. The MT procedure also dominates other regularisation techniques 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, derives its asymptotic properties, and discusses the S-MT estimator. The small sample properties of the MT estimator and its shrinkage version 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 $N \times N$ matrix $\boldsymbol{A} = (a_{ij})$ by $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$, respectively, its trace by $tr(\boldsymbol{A}) = \sum_{i=1}^{N} a_{ii}$, its maximum absolute column sum norm by $\|\boldsymbol{A}\|_1 = \max_{1\leq j\leq N}\left(\sum_{i=1}^{N}|a_{ij}|\right)$, its maximum absolute row sum norm by $\|\boldsymbol{A}\|_\infty = \max_{1\leq i\leq N}\left(\sum_{j=1}^{N}|a_{ij}|\right)$, its Frobrenius norm by $\|\boldsymbol{A}\|_F = \sqrt{tr(\boldsymbol{A}'\boldsymbol{A})}$, its spectral (or operator) norm by $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$. $a_N = O(b_N)$ states the deterministic sequence $\{a_N\}$ is at most of order b_N, $\boldsymbol{x}_N = O_p(\mathbf{y}_N)$ states the vector of random variables, $\boldsymbol{x}_N$, is at most of order $\mathbf{y}_N$ in probability, and $\boldsymbol{x}_N = o_p(\mathbf{y}_N)$ is of smaller order in probability than $\mathbf{y}_N$, $\rightarrow_p$ denotes convergence in probability, and $\rightarrow_d$ convergence in distribution. All asymptotics are carried out under $N \rightarrow \infty$ *jointly* with $T \rightarrow \infty$.

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 (Ω, F, P), and denote the covariance matrix of $\boldsymbol{x}_t = (x_{1t}, x_{2t}, ..., x_{Nt})'$ by

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

where $E(\boldsymbol{x}_t) = \boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_N)'$, and $\boldsymbol{\Sigma}$ is an $N \times N$ symmetric, positive definite real matrix with (i,j) element, σ_{ij}.

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

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

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

Assumption 1 *The population covariance matrix, $\boldsymbol{\Sigma} = (\sigma_{ij})$, is sparse in the sense that m_N defined by*

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

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

We also make the following assumption about the bivariate moments of (x_{it}, x_{jt}).

Assumption 2 *The T observations $\{(x_{i1}, x_{j1}), (x_{i2}, x_{j2}), ..., (x_{iT}, x_{jT})\}$ are drawn from a general bivariate distribution with mean $\mu_i = E(x_{it})$, $|\mu_i| < K$, variance $\sigma_{ii} = Var(x_{it})$, $0 < \sigma_{ii} < K$, and correlation coefficient $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ satisfying $0 < \rho_{\min} < |\rho_{ij}| < \rho_{\max} < 1$. Also, it is assumed that the following finite higher-order moments exist*

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

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

The regularisation method that we propose follows the hard thresholding literature, where the non-diagonal elements of the sample covariance matrix (in absolute value) that fall below a certain 'threshold' are set to zero. The distinguishing feature of our approach is its reliance on multiple testing rather than on cross-validation to obtain the thresholds. We also apply the thresholding procedure 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) 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 $t = 1, 2, ..., T$, by

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

where $\hat{\sigma}_{ij,T}$ is defined by (2). For a given i and j, it is well known that under $H_{0,ij} : \sigma_{ij} = 0$, $\sqrt{T}\hat{\rho}_{ij,T}$ is asymptotically distributed as $N(0,1)$ for T sufficiently large. This suggests using $T^{-1/2}\Phi^{-1}\left(1 - \frac{p}{2}\right)$ as the threshold for $|\hat{\rho}_{ij,T}|$, 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 $\mathbf{\Sigma}$ 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

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

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

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

Hence to achieve $FWER_T \leq p$, it is sufficient to set $p_{iT} \leq p/n$. Alternative multiple testing procedures advanced in the literature that are less conservative than the Bonferroni procedure can also be employed. One prominent example is the step-down procedure proposed by Holm (1979) that, similar to the Bonferroni approach, does not impose any further restrictions on the degree to which the underlying tests depend on each other. More recently, Romano and Wolf (2005) propose computer intensive stepdown 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 $\mathbf{R} = (\rho_{ij})$. In particular, we show that consistent support recovery can be achieved so long as $f(N)$ rises linearly in N, but for controlling the errors in estimation of $\mathbf{R}$, we might require $f(N)$ to be a quadratic or even a cubic function of N. The latter case arises if the underlying observations, x_{it}, display non-linear dependence in the sense that zero correlations do not imply independence.

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

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

where

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

It is evident that since $c_p(N)$ is given 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. The computationally intensive cross-validation procedure is typically employed for the estimation of C.

Remark 1 *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 $f(N) = N(N-1)/2$ which 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 small sample results conform closely to our theoretical findings.*

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

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

where $\hat{\boldsymbol{D}} = diag(\hat{\sigma}_{11,T}, \hat{\sigma}_{22,T}, ..., \hat{\sigma}_{NN,T})$. Note that the MT procedure can be applied to de-factored observations as well when adjusted for the loss of degrees of freedom associated with the de-factoring.

2.1 Theoretical properties of the MT estimator

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

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

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

Then

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

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

where

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

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

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

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

Proof of Proposition 1. The results for $E\left(\hat{\rho}_{ij,T}\right)$ and $Var\left(\hat{\rho}_{ij,T}\right)$ are established in Gayen (1951) using a bivariate Edgeworth expansion approach. This confirms earlier findings obtained by Tschuprow (1925, English Translation, 1939) who shows that results (8) and (9) hold for any law of dependence between x_{it} and x_{jt}. See, in particular, p. 228 and equations (53) and (54) in Gayen (1951). Using (9) and (11) we have $\lim_{T\to\infty}\left[TVar\left(\hat{\rho}_{ij,T}\right)\right] = K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})$. Finally, the boundedness of $|K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})|$ and $K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})$ follows directly from the assumption that the fourth-order moment of y_{it} exists for all i and t. The existence of the other moments, $E(y_{it}^3 y_{jt})$ and $E\left(y_{it}^2 y_{jt}^2\right)$, follows by application of Holder's and Cauchy–Schwartz inequalities as given below:

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

and

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

■

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

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

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

$$\begin{aligned}\Pr\left(z_{ij,T} \leq a_{ij,T}\left|\mathcal{P}_{ij}\right.\right) &= F_{ij,T}(a_{ij,T}\left|\mathcal{P}_{ij}\right.) \\ &= \Phi(a_{ij,T}) + T^{-1/2}\phi\left(a_{ij,T}\right)G_1\left(a_{ij,T}\left|\mathcal{P}_{ij}\right.\right) + T^{-1}\phi\left(a_{ij,T}\right)G_2\left(a_{ij,T}\left|\mathcal{P}_{ij}\right.\right) + ...,\end{aligned} \tag{12}$$

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

Remark 3 *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 $E\left(\hat{\rho}_{ij,T}\right)$ and $Var\left(\hat{\rho}_{ij,T}\right)$ 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, now we establish the rate of convergence of the MT estimator under the Frobenius norm.

Theorem 1 *(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 $\mathbf{R} = (\rho_{ij})$. 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 $N > N_0$,*

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

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

and

$$\kappa_{\max} \leq lim_{N\to\infty} \frac{\ln [f(N)]}{\ln(N)}, \tag{13}$$

where $\kappa_{\max} = \sup_{ij} [\kappa_{ij}]$, $\kappa_{ij} = \left[\mu_{ij}(2,2) \left| \rho_{ij} = 0\right.\right]$, with $\mu_{ij}(2,2)$ defined in Assumption 2. Then

$$E \left\| \widetilde{\mathbf{R}}_{MT} - \mathbf{R} \right\|_F^2 = \sum\sum_{i \neq j} E(\tilde{\rho}_{ij,T} - \rho_{ij})^2 = O\left(\frac{m_N N}{T}\right), \tag{14}$$

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

Proof. *See Appendix.* ■

Result (14) implies that $N^{-1} \left\| \widetilde{\mathbf{R}}_{MT} - \mathbf{R} \right\|_F^2 = O_p\left(\frac{m_N}{T}\right)$ which is in line with the existing results in the thresholding literature that use the Frobenius norm. See, for example, Theorem 2 with $q = 0$ in Bickel and Levina (2008b). The same rate of $O_p(m_N/T)$ is achieved in the shrinkage literature 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 $O_p(N/T)$ 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 4 *For the convergence of the Frobenius norm at the rate of $O_p(m_N N/T)$, 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 1. 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 $f(N) = O(N^3)$.

Finally, we evaluate the conditions for consistent support recovery of the true correlation matrix via the true positive rate (TPR) and the false positive rate (FPR) defined below.

Theorem 2 *(Support Recovery) Consider the true positive rate (TPR) and the false positive rate (FPR) statistics computed using the multiple testing estimator*

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

given by

$$TPR = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} \neq 0)}{\sum\sum_{i \neq j} I(\rho_{ij} \neq 0)} \tag{15}$$

$$FPR = \frac{\sum\sum_{i \neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum\sum_{i \neq j} I(\rho_{ij} = 0)}, \tag{16}$$

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 $\rho_{\min} = \min_{ij}\left|\rho_{ij}\right| > 0$.

Proof. *See Appendix.* ■

Remark 5 *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 $\ln f(N)/T = \log(N-1)/T \to 0$, 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 2 are met. Interestingly, this suggests that consistent support recovery is ensured if Bonferroni's MT procedure is applied to $\mathbf{R}$ (or $\mathbf{\Sigma}$) 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.*

2.2 Positive definiteness of the MT estimator

As with other thresholding approaches, multiple testing preserves the symmetry of $\hat{\boldsymbol{R}}$ and is invariant to the ordering of the variables but it does not ensure positive definiteness of the estimated covariance matrix. Bickel and Levina (2008b) provide an asymptotic condition that ensures positive definiteness, which is not met unless T is sufficiently large relative to N.

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 would be to apply shrinkage to $\widetilde{\boldsymbol{R}}_{MT}$ which is known to result in a positive definite matrix. Specifically, following Ledoit and Wolf (2004) we set as benchmark target the $N \times N$ identity matrix $\boldsymbol{I}_N$, and introduce the shrinkage estimator (S-MT)

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi) = \xi \boldsymbol{I}_N + (1-\xi)\widetilde{\boldsymbol{R}}_{MT}, \tag{17}$$

where the shrinkage parameter $\xi \in (\xi_0, 1]$, and ξ_0 is the minimum value of ξ that produces a non-singular $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0)$ matrix. First note that shrinkage is again deliberately implemented on the correlation matrix $\widetilde{\boldsymbol{R}}_{MT}$ rather than on $\widetilde{\boldsymbol{\Sigma}}_{MT}$. In this way we ensure that no shrinkage is applied to the variances. Second, shrinkage is applied to the non-zero elements of $\widetilde{\boldsymbol{R}}_{MT}$, and as a result the shrinkage estimator, $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$, also consistently recovers the support of $\mathbf{R}$, since it has the same support recovery property as $\widetilde{\boldsymbol{R}}_{MT}$.

The shrinkage parameter, ξ, is calibrated solving the following optimisation problem

$$\xi^* = \arg \min_{\xi_0+\epsilon \leq \xi \leq 1} \left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \right\|_F^2, \tag{18}$$

where ϵ is a small positive constant, and $\mathbf{R}_0$ is a reference invertible correlation matrix. Let $\boldsymbol{A} = \boldsymbol{R}_0^{-1}$ and $\boldsymbol{B}(\xi) = \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi)$. Note that since $\boldsymbol{R}_0$ and $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$ are symmetric

$$\left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \right\|_F^2 = tr\left(\boldsymbol{A}^2\right) - 2tr[\boldsymbol{A}\boldsymbol{B}(\xi)] + tr[\boldsymbol{B}^2(\xi)].$$

The first order condition for the above optimisation problem is given by

$$\frac{\partial \left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \right\|_F^2}{\partial \xi} = -2tr\left(\boldsymbol{A}\frac{\partial \boldsymbol{B}(\xi)}{\partial \xi}\right) + 2tr\left(\boldsymbol{B}(\xi)\frac{\partial \boldsymbol{B}(\xi)}{\partial \xi}\right),$$

where

$$\begin{aligned} \frac{\partial \boldsymbol{B}(\xi)}{\partial \xi} &= -\widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi)\left(\boldsymbol{I}_N - \tilde{\boldsymbol{R}}_{MT}\right)\widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \\ &= -\boldsymbol{B}(\xi)\left(\boldsymbol{I}_N - \tilde{\boldsymbol{R}}_{MT}\right)\boldsymbol{B}(\xi). \end{aligned}$$

Hence, ξ^* is obtained as the solution of

$$f(\xi) = -tr\left[(\mathbf{A} - \boldsymbol{B}(\xi))\boldsymbol{B}(\xi)\left(\mathbf{I}_N - \widetilde{\boldsymbol{R}}_{MT}\right)\boldsymbol{B}(\xi)\right] = 0,$$

where $f(\xi)$ is an analytic differentiable function of ξ for values of ξ close to unity, such that $\boldsymbol{B}(\xi)$ exists. The resulting $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi^*)$ is guaranteed to be positive definite since

$$\lambda_{\min}\left[\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi)\right] = \xi\lambda_{\min}(\boldsymbol{I}_N) + (1-\xi)\lambda_{\min}\left(\widetilde{\boldsymbol{R}}_{MT}\right) > 0,$$

for any $\xi \in [\xi_0, 1]$, where $\xi_0 = \max\left(\frac{\epsilon-\lambda_{\min}\left(\widetilde{\boldsymbol{R}}_{MT}\right)}{1-\lambda_{\min}\left(\widetilde{\boldsymbol{R}}_{MT}\right)}, 0\right)$.

Having obtained the shrinkage estimator $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$, using ξ^* in (17), we construct the corresponding covariance matrix as

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

Implementation of the above procedure requires the use of a suitable reference matrix $\boldsymbol{R}_0$. Our experimentations suggest that the shrinkage estimator of Ledoit and Wolf (2004, LW) applied to the correlation matrix is likely to work well in practice, and is to be recommended. Schäfer and Strimmer (2005) consider LW shrinkage on the correlation matrix. In our application we also take account of the small sample bias of the correlation coefficients. Further details are available in the Supplementary Appendix B.

3 Small sample properties

We investigate the small sample properties of the proposed multiple testing (MT) estimator using Monte Carlo simulations. We compare our estimator with a number of thresholding and shrinkage type 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 C.

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

$$\mathbf{y}_t = \boldsymbol{P}\mathbf{u}_t,\ t = 1, ..., T,$$

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

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

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 ε_{it} for all i and t. As fourth-order moments are required by Assumption 2 we set $v = 8$ to ensure that $E\left(y_{it}^4\right)$ exists. We also note that under $\rho_{ij} = 0$, $\kappa_{ij} = \mu_{ij}(2, 2\left|\rho_{ij} = 0\right.) = (v-2)/(v-4)$, and with $v = 8$ we have $\kappa_{ij} = \kappa_* = 1.5$. Therefore, in the case where the standardised errors are multivariate t-distributed to ensure that conditions of Theorem 1 are met we must set $f(N) = N\left(N-1\right)/2$. (See also Remark

4 and Lemma 7 in the Supplementary Appendix A). One could further allow for fat-tailed ε_{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 $\boldsymbol{P} = \boldsymbol{I}_N$ or not. In such cases setting $f(N) = N - 1$ is then sufficient for conditions of Theorem 1 to be met.

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

$$\boldsymbol{x}_t = \mathbf{a} + \boldsymbol{\gamma} f_t + \mathbf{D}^{1/2}\mathbf{y}_t, \tag{19}$$

where $\boldsymbol{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})$, $\mathbf{a} = (a_1, a_2, ..., a_N)'$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_N)'$. We focus on the baseline case where $\boldsymbol{\gamma} = \mathbf{0}$ and $\mathbf{a} = \mathbf{0}$. The properties of the MT procedure when factors are included in the DGP are also investigated by drawing γ_i and a_i as $IIDN\,(1,1)$ for $i = 1, 2, ..., N$, and generating f_t, the common factor, as a stationary AR(1) process. 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

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

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

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 $Uniform\,(0.7, 0.9)$ for the first and last N_b $(< N)$ elements of $\mathbf{b}$, where $N_b = \left[N^\delta\right]$, and set the remaining middle elements of $\mathbf{b}$ to zero. The resulting population correlation matrix $\boldsymbol{R}$ is defined by

$$\boldsymbol{R} = \boldsymbol{I}_N + \boldsymbol{bb}' - diag\,(\boldsymbol{bb}')\,. \tag{20}$$

The degree of sparseness of $\boldsymbol{R}$ is determined by the value of the parameter δ. 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 $\boldsymbol{P}$ is then obtained by applying the Cholesky decomposition to $\boldsymbol{R}$ defined by (20). Further, we set $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{R}\boldsymbol{D}^{1/2}$, where the diagonal elements of $\boldsymbol{D}$ are given by $\sigma_{ii} \sim IID\,(1/2 + \chi^2(2)/4)$, $i = 1, 2, ..., N$.

An additional two covariance specifications based on *approximately* sparse matrices as defined in Bickel and Levina (2008b, p. 2580 for $0 < q < 1$), along with their associated simulation results can be found in the Supplementary Appendix D.

3.1 Alternative estimators and evaluation metrics

Using the above set up we obtain the following estimates of $\mathbf{\Sigma}$:

MT_{N-1}: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\mathbf{\Sigma}}_{MT}$) using $f(N) = N-1$ ($\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}$)

$MT_{N(N-1)/2}$: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\mathbf{\Sigma}}_{MT}$) using $f(N) = N(N-1)/2$ ($\widetilde{\mathbf{\Sigma}}_{MT_{N(N-1)/2}}$)

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

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

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

$S\text{-}MT_{N-1}$: supplementary shrinkage applied to MT_{N-1} ($\widetilde{\mathbf{\Sigma}}_{S\text{-}MT_{N-1}}$)

$S\text{-}MT_{N(N-1)/2}$: supplementary shrinkage applied to $MT_{N(N-1)/2}$ ($\widetilde{\mathbf{\Sigma}}_{S\text{-}MT_{N(N-1)/2}}$)

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

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

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

In accordance with the theoretical results in Theorem 1 and in view of Remark 4, we consider two versions of the MT estimator depending on the choice of $f(N) = \{N-1, N(N-1)/2\}$. The $BL_{\hat{C}}$, and CL_2 and $CL_{\hat{C}}$ estimators apply the thresholding procedure without ensuring that the resultant covariance estimators are invertible. The next five estimators yield invertible covariance estimators. The $S\text{-}MT$ estimators are obtained using the supplementary shrinkage approach described in Section 2.2. $BL_{\hat{C}^*}$ and $CL_{\hat{C}^*}$ estimators are obtained by applying the additional FLM adjustments. The shrinkage estimator, $LW_{\hat{\Sigma}}$, is invertible by construction. In the case of the MT estimators where regularisation is performed on the correlation matrix the associated covariance matrix is estimated as $\hat{\mathbf{D}}^{1/2}\widetilde{\boldsymbol{R}}_{MT}\hat{\mathbf{D}}^{1/2}$.

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

$$\|\boldsymbol{A}_{\mathring{\mathbf{\Sigma}}}\| = \left\|\mathbf{\Sigma} - \mathring{\mathbf{\Sigma}}\right\|_F, \tag{21}$$

where $\mathring{\mathbf{\Sigma}}$ is set to one of the following estimators $\{\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}, \widetilde{\mathbf{\Sigma}}_{MT_{N(N-1)/2}}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_{N-1}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_{N(N-1)/2}}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. The threshold values, $\hat{C}$ and $\hat{C}^*$, are obtained by cross-validation (see Supplementary Appendix C.3 for details). We also evaluate the spectral norm of the difference given in (21), denoted by $\|.\|$. Both norms are also computed for the difference between $\mathbf{\Sigma}^{-1}$, the true inverse of $\mathbf{\Sigma}$, and the estimators $\{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_{N-1}}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_{N(N-1)/2}}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. Further, we investigate the ability of the thresholding estimators to recover the support of the true covariance matrix via the true positive rate (TPR) and false positive rate (FPR), as defined by (15) and (16), respectively. The statistics TPR and FPR are not relevant to the shrinkage estimator $LW_{\hat{\Sigma}}$ and will not be reported for this estimator.

We report results for $N = \{30, 100, 200\}$ and $T = 100$, in the case where $\boldsymbol{\gamma} = \mathbf{0}$ and $\mathbf{a} = \mathbf{0}$ in (19). Similar results are obtained for $\boldsymbol{\gamma} \neq \mathbf{0}$ and $\mathbf{a} \neq \mathbf{0}$, but to save space they are made available upon request.

3.2 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 that in the case of Gaussian observations, where $\kappa_{\max} = 1$, 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 Theorem 1 and Remark 4). 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 Appendix D.

Overall, we find that the results are more robust when the scaling factor $N(N-1)/2$ is used, particularly considering that the value of $\kappa_{\max}$ is typically unknown and must be estimated.

3.3 Norm comparisons of MT, BL, CL, and LW estimators

In comparing our proposed estimators with those proposed in the literature we were forced to consider a fewer number of Monte Carlo replications and report the results with norm losses averaged over only 100 replications, solely due to the computationally intensive nature of the cross-validation procedure used in the implementation of BL and CL thresholding, especially for large values of N. However, the results reported here are in line with the simulation set up and the results reported by BL and CL.

Tables 2 and 3 summarise the results for the Monte Carlo designs A and B, respectively. We provide norm comparisons for the MT estimator with $p = 0.05$ and the scaling factor $N(N-1)/2$. Initially, we consider the threshold estimators, MT, BL and the two versions of the CL estimators (CL_2 and $CL_{\hat{C}}$) 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 $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}$ and $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT}$) with $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$. In such comparisons $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$ performs relatively better, nevertheless, $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$ is still dominated by $\widetilde{\mathbf{\Sigma}}_{S\text{-}MT}$, with a few exceptions in the case of design A and primarily when $N = 30$. However, no clear ordering emerges when we compare $\hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}$ with $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}$.

3.4 Norm comparisons of inverse estimators

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

3.5 Support recovery statistics

Table 5 reports the true positive and false positive rates (TPR and FPR) for the support recovery of $\mathbf{\Sigma}$ using the multiple testing and thresholding estimators. In the comparison set we include two versions of the MT estimator ($\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}$ and $\widetilde{\mathbf{\Sigma}}_{MT_{N(N-1)/2}}$), $\widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}$, $\widetilde{\mathbf{\Sigma}}_{CL,2}$, and $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}$. 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 born 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 $\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}$ and $\widetilde{\mathbf{\Sigma}}_{MT_{N(N-1)/2}}$ we find that as predicted by the theory (Theorem 2 and Remark 5), TPR values of $\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}$ are closer to unity as compared to the TPR values obtained for $\widetilde{\mathbf{\Sigma}}_{MT_{N(N-1)/2}}$.

Turning to a comparison with other estimators, 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 $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$ to deteriorate noticeably while MT_{N-1} and $MT_{N(N-1)/2}$ remain remarkably stable. Finally, note that TPR values are higher for design B, since for this specification we explicitly control for the number of non-zero elements in $\mathbf{\Sigma}$, and ensure that conditions of Theorem 2 are met. Overall, the estimator $\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}$ does best in recovering the support of $\mathbf{\Sigma}$ 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.

4 Concluding Remarks

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

A novel regularisation estimator is proposed that uses insights from the multiple testing (MT) literature to enhance the support of the true covariance matrix. It is applied to the sample correlation matrix. It is shown that the resultant estimator has a convergence rate of $(m_N N/T)^{1/2}$ under the Frobenius norm, where T is the number of observations, and m_N is bounded in N (the dimension of $\mathbf{\Sigma}$), which is comparable with the convergence rates established in the literature for other regularised covariance matrix estimators. But, unlike the threshold estimators which are calibrated using cross-validation, the proposed MT estimator is computationally very simple and fast to implement. To ensure that the resultant estimator of $\mathbf{\Sigma}$ is invertible we also propose a shrinkage on multiple testing (S-MT) estimator which is guaranteed to be positive definite.

The small sample properties of the proposed estimators are investigated using Monte Carlo simulations. It is shown that both MT and S-MT estimators perform 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 estimators are reasonably robust to the choice of p in the threshold criterion, $\left|\hat{\rho}_{ij}\right| > T^{-1/2}\Phi^{-1}\left(1 - \frac{p}{2f(N)}\right)$, 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 $T = 100$

	$f(N) = N-1$			$f(N) = N(N-1)/2$		
N	$MT_{N-1}(.05)$	$MT_{N-1}(.05)$	$MT_{N-1}(.10)$	$MT_{\frac{N(N-1)}{2}}(.01)$	$MT_{\frac{N(N-1)}{2}}(.05)$	$MT_{\frac{N(N-1)}{2}}(.10)$
Monte Carlo design A						
$\mathbf{u}_{it}\sim$ **Gaussian**						
	Spectral norm					
30	1.70(0.49)	1.68(0.49)	1.72(0.49)	1.84(0.50)	1.75(0.50)	1.71(0.50)
100	2.61(0.50)	2.51(0.50)	2.50(0.50)	3.02(0.50)	2.84(0.50)	2.76(0.50)
200	3.04(0.48)	2.92(0.49)	2.89(0.49)	3.58(0.47)	3.37(0.47)	3.29(0.47)
	Frobenius norm					
30	3.17(0.45)	3.14(0.50)	3.20(0.54)	3.41(0.42)	3.25(0.44)	3.19(0.44)
100	6.66(0.45)	6.51(0.51)	6.60(0.55)	7.57(0.41)	7.17(0.42)	7.00(0.42)
200	9.87(0.46)	9.60(0.53)	9.73(0.58)	11.49(0.41)	10.89(0.42)	10.63(0.42)
$\mathbf{u}_{it}\sim$ **multivariate $t-$distributed with 8 degrees of freedom**						
	Spectral norm					
30	2.26(1.08)	2.43(1.20)	2.55(1.27)	2.26(0.95)	2.24(1.03)	2.25(1.06)
100	3.85(4.84)	4.21(5.29)	4.47(5.48)	3.74(3.94)	3.71(4.28)	3.72(4.44)
200	4.49(3.46)	5.04(4.34)	5.45(4.77)	4.23(1.97)	4.19(2.37)	4.20(2.58)
	Frobenius norm					
30	4.07(1.14)	4.36(1.32)	4.62(1.40)	4.08(0.95)	4.03(1.06)	4.04(1.11)
100	8.88(5.17)	9.76(5.67)	10.51(5.88)	8.92(4.19)	8.74(4.57)	8.70(4.74)
200	12.96(4.23)	14.51(5.41)	15.82(5.95)	13.06(2.26)	12.71(2.77)	12.63(3.05)
Monte Carlo design B						
$\mathbf{u}_{it}\sim$ **Gaussian**						
	Spectral norm					
30	0.48(0.16)	0.50(0.16)	0.53(0.16)	0.49(0.18)	0.48(0.17)	0.48(0.17)
100	0.75(0.34)	0.76(0.32)	0.78(0.31)	0.85(0.41)	0.79(0.37)	0.77(0.37)
200	0.71(0.22)	0.74(0.20)	0.77(0.20)	0.81(0.31)	0.75(0.26)	0.73(0.26)
	Frobenius norm					
30	0.87(0.17)	0.92(0.18)	0.98(0.19)	0.87(0.18)	0.86(0.17)	0.86(0.17)
100	1.56(0.24)	1.66(0.24)	1.77(0.24)	1.64(0.31)	1.58(0.27)	1.57(0.27)
200	2.16(0.18)	2.32(0.20)	2.50(0.21)	2.22(0.22)	2.16(0.20)	2.15(0.20)
$\mathbf{u}_{it}\sim$ **multivariate $t-$distributed with 8 degrees of freedom**						
	Spectral norm					
30	0.70(0.39)	0.78(0.43)	0.84(0.45)	0.67(0.34)	0.67(0.37)	0.69(0.38)
100	1.16(0.97)	1.32(1.10)	1.42(1.18)	1.12(0.77)	1.10(0.83)	1.10(0.87)
200	1.36(1.73)	1.65(2.05)	1.83(2.20)	1.13(1.11)	1.14(1.28)	1.16(1.37)
	Frobenius norm					
30	1.23(0.43)	1.41(0.48)	1.54(0.51)	1.15(0.36)	1.18(0.39)	1.20(0.41)
100	2.40(1.12)	2.90(1.31)	3.26(1.40)	2.16(0.81)	2.16(0.90)	2.20(0.96)
200	3.57(2.14)	4.52(2.54)	5.18(2.72)	2.97(1.30)	3.01(1.53)	3.07(1.65)

Notes: Norm losses are averages over 2,000 replications. Simulation standard deviations are given in parentheses. The MT estimators are defined in Section 3.1.

Table 2: Spectral and Frobenius norm losses for different regularised covariance matrix estimators ($T = 100$) - Monte Carlo Design A

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
			$\mathbf{u}_{it}\sim$ Gaussian			
	Error matrices $(\mathbf{\Sigma} - \mathring{\mathbf{\Sigma}})$					
$MT_{N(N-1)/2}$	1.81(0.54)	3.31(0.42)	2.75(0.50)	7.11(0.42)	3.37(0.43)	10.91(0.39)
$BL_{\hat{C}}$	5.30(2.16)	7.61(1.23)	8.74(0.06)	16.90(0.10)	8.94(0.04)	24.26(0.13)
CL_2	1.87(0.55)	3.39(0.44)	2.99(0.49)	7.57(0.44)	3.79(0.47)	11.88(0.42)
$CL_{\hat{C}}$	1.82(0.58)	3.33(0.56)	2.54(0.50)	6.82(0.51)	3.02(0.46)	10.22(0.59)
$S\text{-}MT_{N(N-1)/2}$	3.20(0.79)	4.29(0.64)	5.73(0.34)	10.77(0.46)	6.40(0.21)	16.44(0.35)
$BL_{\hat{C}^*}$	7.09(0.10)	8.62(0.09)	8.74(0.06)	16.90(0.10)	8.94(0.04)	24.25(0.10)
$CL_{\hat{C}^*}$	7.05(0.16)	8.58(0.12)	8.71(0.07)	16.85(0.11)	8.94(0.04)	24.23(0.09)
$LW_{\hat{\Sigma}}$	2.99(0.47)	6.49(0.29)	5.20(0.34)	16.70(0.19)	6.28(0.20)	26.84(0.14)
		$\mathbf{u}_{it}\sim$ multivariate $t-$ distributed with 8 degrees of freedom				
	Error matrices $(\mathbf{\Sigma} - \mathring{\mathbf{\Sigma}})$					
$MT_{N(N-1)/2}$	2.16(0.76)	4.03(0.99)	3.43(1.09)	8.43(1.26)	3.97(0.92)	12.66(1.83)
$BL_{\hat{C}}$	6.90(0.82)	8.75(0.55)	8.74(0.10)	17.26(0.30)	9.00(0.42)	24.93(1.02)
CL_2	2.55(0.93)	4.53(1.00)	4.63(1.11)	10.35(1.48)	5.92(0.81)	16.43(1.74)
$CL_{\hat{C}}$	2.27(0.76)	4.24(0.94)	3.85(1.51)	9.44(2.33)	5.04(2.04)	15.65(4.71)
$S\text{-}MT_{N(N-1)/2}$	3.18(0.82)	4.68(0.82)	5.75(0.45)	11.33(0.62)	6.41(0.32)	17.10(0.74)
$BL_{\hat{C}^*}$	7.06(0.13)	8.84(0.30)	8.74(0.10)	17.25(0.31)	8.95(0.08)	24.84(0.55)
$CL_{\hat{C}^*}$	7.01(0.16)	8.77(0.30)	8.73(0.11)	17.23(0.29)	8.94(0.08)	24.77(0.53)
$LW_{\hat{\Sigma}}$	3.35(0.51)	7.35(0.50)	5.67(0.46)	18.04(0.45)	6.60(0.43)	28.18(0.53)

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

Table 3: Spectral and Frobenius norm losses for different regularised covariance matrix estimators ($T = 100$) - Monte Carlo Design B

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
			$\mathbf{u}_{it}\sim$ Gaussian			
	Error matrices $(\mathbf{\Sigma} - \mathring{\hat{\mathbf{\Sigma}}})$					
$MT_{N(N-1)/2}$	0.48(0.16)	0.88(0.17)	0.84(0.36)	1.61(0.26)	0.70(0.21)	2.13(0.18)
$BL_{\hat{C}}$	0.91(0.50)	1.35(0.43)	1.40(0.95)	2.25(0.78)	2.53(0.55)	3.49(0.32)
CL_2	0.49(0.17)	0.90(0.18)	1.00(0.48)	1.77(0.44)	0.90(0.37)	2.30(0.30)
$CL_{\hat{C}}$	0.49(0.15)	0.92(0.17)	0.83(0.31)	1.71(0.28)	1.14(0.83)	2.54(0.58)
$S\text{-}MT_{N(N-1)/2}$	0.68(0.25)	1.08(0.20)	1.50(0.50)	2.14(0.35)	1.18(0.38)	2.40(0.24)
$BL_{\hat{C}^*}$	1.19(0.46)	1.63(0.40)	3.32(0.20)	3.90(0.14)	2.73(0.11)	3.61(0.08)
$CL_{\hat{C}^*}$	1.08(0.46)	1.53(0.46)	3.34(0.15)	3.92(0.06)	2.73(0.10)	3.61(0.08)
$LW_{\hat{\Sigma}}$	1.05(0.13)	2.07(0.10)	2.95(0.26)	4.47(0.09)	2.46(0.06)	6.01(0.03)
	$\mathbf{u}_{it}\sim$ multivariate $t-$distributed with 8 degrees of freedom					
	Error matrices $(\mathbf{\Sigma} - \mathring{\hat{\mathbf{\Sigma}}})$					
$MT_{N(N-1)/2}$	0.65(0.25)	1.13(0.25)	1.02(0.45)	2.11(0.51)	1.33(2.46)	3.19(2.81)
$BL_{\hat{C}}$	1.36(0.40)	1.84(0.35)	2.70(0.94)	3.58(0.74)	2.70(0.29)	4.08(0.67)
CL_2	0.71(0.29)	1.21(0.30)	1.69(0.70)	2.73(0.70)	1.62(0.57)	3.31(0.65)
$CL_{\hat{C}}$	0.80(0.39)	1.33(0.39)	2.03(1.08)	3.07(0.90)	2.19(0.78)	3.72(0.62)
$S\text{-}MT_{N(N-1)/2}$	0.69(0.26)	1.18(0.23)	1.37(0.53)	2.32(0.44)	1.30(0.80)	3.02(0.89)
$BL_{\hat{C}^*}$	1.49(0.26)	1.98(0.21)	3.33(0.24)	4.07(0.18)	2.77(0.37)	4.04(0.56)
$CL_{\hat{C}^*}$	1.26(0.40)	1.79(0.40)	3.35(0.17)	4.08(0.14)	2.73(0.14)	4.01(0.42)
$LW_{\hat{\Sigma}}$	1.13(0.15)	2.25(0.11)	3.14(0.21)	4.68(0.11)	2.52(0.08)	6.18(0.13)

See the notes to Table 2.

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

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices $(\boldsymbol{\Sigma}^{-1} - \mathring{\boldsymbol{\Sigma}}^{-1})$					
			$\mathbf{u}_{it}\sim$ Gaussian			
$S\text{-}MT_{N(N-1)/2}$	4.42(1.22)	2.66(0.31)	15.62(2.68)	5.87(0.46)	13.89(2.29)	5.45(0.36)
$BL_{\hat{C}^*}$	$3.8{\times}10^3(2.4{\times}10^4)$	19.56(58.88)	$1.2{\times}10^3(1.1{\times}10^4)$	12.16(33.25)	41.07(143.74)	7.66(3.17)
$CL_{\hat{C}^*}$	$1.9{\times}10^3(1.7{\times}10^4)$	10.92(42.39)	51.99(241.39)	8.16(4.23)	28.45(24.37)	7.35(1.11)
$LW_{\hat{\Sigma}}$	11.03(0.58)	4.26(0.09)	31.04(0.64)	8.62(0.06)	31.81(0.21)	9.40(0.05)
			$\mathbf{u}_{it}\sim$ multivariate $t-$distributed with 8 degrees of freedom			
$S\text{-}MT_{N(N-1)/2}$	3.42(1.55)	2.44(0.38)	12.39(3.01)	5.49(0.54)	11.23(4.12)	5.54(0.66)
$BL_{\hat{C}^*}$	$157.26(1.0{\times}10^3)$	6.11(11.28)	$349.35(3.1{\times}10^3)$	9.80(17.03)	28.58(22.06)	7.77(1.04)
$CL_{\hat{C}^*}$	85.82(546.85)	5.53(7.84)	$517.27(4.8{\times}10^3)$	10.07(21.25)	25.61(3.55)	7.54(0.50)
$LW_{\hat{\Sigma}}$	12.08(1.19)	4.48(0.20)	31.78(1.32)	8.74(0.23)	32.06(1.00)	9.50(0.33)

Notes: $\mathring{\boldsymbol{\Sigma}}^{-1} = \{\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}N(N-1)/2}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\boldsymbol{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. See also the notes to Table 2.

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

		Monte Carlo Design A							Monte Carlo Design B				
N		MT_{N-1}	$MT_{N(N-1)/2}$	$BL_{\hat{C}}$	CL_2	$CL_{\hat{C}}$	N		MT_{N-1}	$MT_{N(N-1)/2}$	$BL_{\hat{C}}$	CL_2	$CL_{\hat{C}}$
							$\mathbf{u}_{it}\sim$ Gaussian						
30	TPR	0.80	0.73	0.29	0.72	0.78	30	TPR	1.00	0.98	0.64	0.98	1.00
	FPR	0.00	0.00	0.04	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00
100	TPR	0.69	0.59	0.00	0.56	0.68	100	TPR	1.00	0.98	0.80	0.94	0.99
	FPR	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00
200	TPR	0.66	0.55	0.00	0.50	0.65	200	TPR	1.00	0.97	0.11	0.88	0.78
	FPR	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00
				$\mathbf{u}_{it}\sim$ multivariate $t-$distributed with 8 degrees of freedom									
30	TPR	0.80	0.73	0.03	0.62	0.74	30	TPR	1.00	0.99	0.26	0.89	0.82
	FPR	0.01	0.00	0.00	0.00	0.00		FPR	0.01	0.00	0.00	0.00	0.00
100	TPR	0.69	0.59	0.00	0.43	0.57	100	TPR	1.00	0.98	0.27	0.70	0.57
	FPR	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00
200	TPR	0.66	0.55	0.00	0.35	0.47	200	TPR	0.99	0.94	0.05	0.57	0.30
	FPR	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00

TPR is the true positive and FPR is the false positive rates defined by (15) and (16), respectively. MT estimators are computed with $p = 0.05$. For a description of other estimators see the notes to Table 2.

Appendix: Mathematical Proofs of theorems for the MT estimator

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

Proof of Theorem 1. Consider

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

and note that

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

Hence

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

However,

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

and

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

Therefore, we have

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

which can be decomposed as

$$\sum\sum_{i\neq j}E\left(\tilde{\rho}_{ij,T}-\rho_{ij}\right)^2=A+B+C, \tag{22}$$

where

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

Consider now the orders of the above three terms in turn, starting with A. Since under Assumption 2, $0 < \rho_{\min} < \left|\rho_{ij}\right| < \rho_{\max} < 1$, then

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

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

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

Recalling that $\sup_{ij} K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) < K$ and by assumption $\rho_{\min} > 0$, it then readily follows that A is of order $O(Ne^{-T})$ so that $A \to 0$, as N and $T \to \infty$. Note that this result *does not* require $N/T \to 0$, and holds even if N/T tends to a fixed constant.

Consider now B and note that since $\hat{\rho}_{ij,T} = \omega_{ij,T} z_{ij,T} + \rho_{ij,T}$ (to simplify the notation we use $\omega_{ij,T}^2$ and $\rho_{ij,T}$ for $Var\left(\hat{\rho}_{ij,T}\right)$ and $E\left(\hat{\rho}_{ij,T}\right)$, respectively) we have the following decomposition of B, $B = B_1 + B_2 + 2B_3$, where

$$\begin{aligned} B_1 &= \sum_{i\neq j,\rho_{ij}\neq 0}\sum \omega_{ij,T}^2 E\left[z_{ij,T}^2\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right], \\ B_2 &= \sum_{i\neq j,\rho_{ij}\neq 0}\sum \left(\rho_{ij,T}-\rho_{ij}\right)^2 E\left[\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right], \\ B_3 &= \sum_{i\neq j,\rho_{ij}\neq 0}\sum \left(\rho_{ij,T}-\rho_{ij}\right)\omega_{ij,T} E\left[z_{ij,T}\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right]. \end{aligned}$$

Using (8) and (9)

$$\omega_{ij,T}^2 = \frac{K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right), \tag{23}$$

$$\left(\rho_{ij,T}-\rho_{ij}\right)^2 = \frac{K_{\mathrm{m}}^2(\boldsymbol{\theta}_{ij})}{T^2} + O\left(T^{-3}\right), \tag{24}$$

$$\left(\rho_{ij,T}-\rho_{ij}\right)\omega_{ij,T} = \frac{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{T^{3/2}} + O\left(T^{-2}\right). \tag{25}$$

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

$$\begin{aligned} B_1 &= \sum_{i\neq j,\rho_{ij}\neq 0}\sum \omega_{ij,T}^2 E\left[z_{ij,T}^2\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right] \\ &\leq \frac{Nm_N}{T}\left[\sup_{ij} K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})\right]\sup_{ij}\left\{1 - E\left[z_{ij,T}^2\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right]\right\} + O\left(\frac{m_N N}{T^2}\right). \end{aligned}$$

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

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

$$\begin{aligned} B_2 &= \sum\sum_{i\neq j,\rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right)^2 E\left[\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right] \\ &\leq Nm_N\left[\frac{K_{\text{m}}^2(\boldsymbol{\theta}_{ij})}{T^2} + O\left(T^{-3}\right)\right] = O\left(\frac{Nm_N}{T^2}\right), \end{aligned}$$

and

$$\begin{aligned} B_3 &= \sum\sum_{i\neq j,\rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right)\omega_{ij,T} E\left[z_{ij,T}\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right] \\ &= \sum\sum_{i\neq j,\rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right)\omega_{ij,T} E\left[z_{ij,T} - z_{ij,T}\left(I\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right] \\ &= -\sum\sum_{i\neq j,\rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right)\omega_{ij,T} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\left|\rho_{ij} \neq 0\right.\right)\right]. \qquad (26) \end{aligned}$$

Also, from Lemma 4

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

and from Lemma 2

$$\begin{aligned} E\left[zI\left(L_{ij,T} \leq z \leq U_{ij,T}\left|\rho_{ij} \neq 0\right.\right)\right] &= \phi\left(\frac{-c_p(N) - \sqrt{T}\rho_{ij} + O\left(\frac{1}{\sqrt{T}}\right)}{\sqrt{K_{\text{v}}(\boldsymbol{\theta}_{ij}) + O\left(\frac{1}{T}\right)}}\right) \\ &\quad -\phi\left(\frac{c_p(N) - \sqrt{T}\rho_{ij} + O\left(\frac{1}{\sqrt{T}}\right)}{\sqrt{K_{\text{v}}(\boldsymbol{\theta}_{ij}) + O\left(\frac{1}{T}\right)}}\right), \qquad (27) \end{aligned}$$

which is bounded in N and T. Since $\sqrt{T}\rho_{\min} - c_p(N) \to \infty$ as N and $T \to \infty$, it is easily seen that $\lim_{T,N\to\infty} E\left[zI\left(L_{ij,T} \leq z \leq U_{ij,T}\left|\rho_{ij} \neq 0\right.\right)\right] = 0$. Hence, using (25) and noting that $K_{\text{v}}^{1/2}(\boldsymbol{\theta}_{ij})K_{\text{m}}(\theta_{ij})$ is bounded in T we have

$$B_3 \leq K \sum\sum_{i\neq j,\rho_{ij}\neq 0} \left|\left(\rho_{ij,T} - \rho_{ij}\right)\omega_{ij,T}\right| = O\left(\frac{Nm_N}{T^{3/2}}\right).$$

Overall, therefore, $B = O\left(\frac{Nm_N}{T}\right)$.

Consider now the following decomposition of C, in (22):

$$\begin{aligned} C &= \sum_{i\neq j,\rho_{ij}=0}\sum E\left[\hat{\rho}_{ij,T}^{2} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \\ &= \sum_{i\neq j,\rho_{ij}=0}\sum \omega_{ij,T}^{2} E\left[z_{ij,T}^{2} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \\ &\quad + \sum_{i\neq j,\rho_{ij}=0}\sum \rho_{ij,T}^{2} E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \\ &\quad + 2\sum_{i\neq j,\rho_{ij}=0}\sum \rho_{ij,T}\omega_{ij,T} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \\ &= C_1 + C_2 + C_3. \end{aligned}$$

Starting with the simpler terms, we first note that

$$\begin{aligned} C_2 &= \sum_{i\neq j,\rho_{ij}=0}\sum \rho_{ij,T}^{2} E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \\ &\leq N(N-m_N-1)\sup_{ij}\left(\rho_{ij,T}^{2}\left|\rho_{ij}=0\right.\right)\sup_{ij} E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right], \end{aligned}$$

and $\sup_{ij} E\left[I\left(\left|\sqrt{T}\hat{\rho}_{ij}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \leq 1$. Using (8) and equation (A.11) of Lemma 6 (and evaluating these expressions under $\rho_{ij}=0$) we have

$$C_2 \leq K\frac{N(N-m_N-1)\sup_{ij}\left(\psi_{ij}^{2}+O\left(T^{-1}\right)\right)}{T^{2}} e^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\sup_{ij}\kappa_{ij}}}\left[1+o(1)\right],$$

where $\kappa_{ij} = \left[\mu_{ij}(2,2)\left|\rho_{ij}=0\right.\right]$, 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 ψ_{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 $C_2 \to 0$ as N and $T \to \infty$, since ψ_{ij}^{2} and κ_{ij} are bounded and $c_p(N) \to \infty$.

Similarly

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

But using Lemma 4, Lemma 2 and (27) and evaluating the relevant expressions under $\rho_{ij}=0$, we have

$$
\begin{aligned}
& \lim_{T,N\to\infty} E\left[z_{ij,T} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N) \left|\rho_{ij}=0\right.\right)\right] \\
= & \lim_{T,N\to\infty} E\left[z I\left(L_{ij,T} \leq z \leq U_{ij,T} \left|\rho_{ij}=0\right.\right)\right] \\
= & \lim_{N,T\to\infty} \phi\left(\frac{-c_p(N)+\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij}+O\left(\frac{1}{T}\right)}}\right) - \lim_{N,T\to\infty} \phi\left(\frac{c_p(N)+\frac{\psi_{ij}}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij}+O\left(\frac{1}{T}\right)}}\right) = 0.
\end{aligned}
$$

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

$$
\begin{aligned}
C_1 &= \sum\sum_{i\neq j,\rho_{ij}=0} \omega_{ij,T}^2 E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N) \left|\rho_{ij}=0\right.\right)\right] \\
&= \sum\sum_{i\neq j,\rho_{ij}=0} \omega_{ij,T}^2 E\left[z_{ij,T}^2 \left[1 - I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\right)\right] \left|\rho_{ij}=0\right.\right] \\
&\leq \frac{N(N-m_N-1)}{T} \sup_{ij}\left[\kappa_{ij}+O\left(\frac{1}{T}\right)\right] \\
&\quad \times \sup_{ij} E\left\{z_{ij,T}^2 \left[1 - I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\right)\right] \left|\rho_{ij}=0\right.\right\}. \qquad (28)
\end{aligned}
$$

But using Lemma 4

$$
\begin{aligned}
& \lim_{T\to\infty} E\left\{z_{ij,T}^2 \left[1 - I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\right)\right] \left|\rho_{ij}=0\right.\right\} \qquad (29) \\
= & \lim_{T\to\infty} E\left\{z^2 \left[1 - I\left(L_{ij,T} \leq z \leq U_{ij,T}\right)\right] \left|\rho_{ij}=0\right.\right\},
\end{aligned}
$$

and then Lemma 2

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

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

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

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

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

where

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

and

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

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

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

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

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

or

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

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 (28) we have

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

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 (14) follows as desired. ■

Proof of Theorem 2. Consider first the FPR statistic given by (16) which can be written equivalently as

$$FPR = |FPR| = \frac{\sum\sum_{i\neq j} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)|\rho_{ij}=0\right)}{N(N-m_N-1)}. \qquad (30)$$

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

Taking the expectation of (30) we have

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

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

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

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 $c_p^2(N) \to \infty$, and $\kappa_{\max} < K$. Further, by the Markov inequality applied to $|FPR|$ we have that

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

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 (15) and note that

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

Hence

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

Since $|X| = X$, then

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

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

$$\lim_{N,T\to\infty}P(|TPR-1|>\delta)\leq\frac{1}{\delta}\lim_{N,T\to\infty}\sup_{ij}\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|<c_p(N)|\rho_{ij}\neq 0\right). \quad (31)$$

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

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

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

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

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

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

Hence, more generally $\lim_{N,T\to\infty}\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|<c_p(N)|\rho_{ij}\neq 0\right)=0$, if $c_p(N)-\sqrt{T}\left|\rho_{ij}\right|\to-\infty$, for all $\rho_{ij}\neq 0$, or equivalently if $\sqrt{T}\rho_{\min}-c_p(N)\to\infty$, where $\rho_{\min}=\min_{ij}\left|\rho_{ij}\right|$ 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 (31) it now follows that $\lim_{N,T\to\infty}P(|TPR-1|>\delta)\to 0$, as required. ■

References

Abadir, K. M., Distaso, W. and Zikes, F. (2014). Design-free estimation of large variance matrices. Journal of Econometrics, forthcoming.

Bickel, P. J. and Levina, E. (2004). Some theory for Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. Bernoulli, 10, 989–1010.

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

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

Bonferroni, C. (1935). Il calcolo delle assicurazioni su gruppi di teste. Studi in Onore del Professore Salvatore Ortu Carboni, Rome, Italy, 13–60.

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

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

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

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

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

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

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

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

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

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

Fryzlewicz, P. (2013). High-dimensional volatility matrix estimation via wavelets and thresholding. Biometrika, 100, 921–938.

Gayen, A. K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. Biometrika, 38, 219-247.

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

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

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

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

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

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

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. Forthcoming in Econometric Reviews.

Pesaran M. H., Schuermann, T. and Weiner, S. M. (2004). Modeling regional interdependencies using a Global Error Correcting Macroeconometric model. Journal of Business & Economic Statistics, 22, 129–162.

Pesaran, M. H. and Yamagata, T. (2012). Testing CAPM with a large number of assets. IZA Discussion Papers 6469, Institute for the Study of Labor (IZA).

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. Biometrika, 86, 677–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. Biometrika, 87, 425–435.

Pourahmadi, M. (2011). Covariance estimation: GLM and regularisation perspectives. Statistical Science, 26, 369–387.

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

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

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

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

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

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

Sargan, J. D. (1976). Econometric estimators and the edgeworth approximation. Econometrica, 44(3), 421–448.

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

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

Tschuprow, A. (1925). Grundbegriffe und grundprobleme der korrelationstheorie. Leipzig: B.G. Teubner.

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

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

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