School of Economics and Finance

A Multiple Testing Approach to the Regularisation of Large Sample Correlation Matrices

Natalia Bailey, M. Hashem Pesaran and L. Vanessa Smith

Working Paper No. 764 December 2015 ISSN 1473-0278

Queen Mary
University of London

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

Natalia Bailey
Queen Mary, University of London

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

L. Vanessa Smith
University of York

December 14, 2015

Abstract

This paper proposes a regularisation method for the estimation of large covariance matrices that uses insights from the multiple testing (MT) literature. The approach tests the statistical significance of individual pair-wise correlations and sets to zero those elements that are not statistically significant, taking account of the multiple testing nature of the problem. By using the inverse of the normal distribution at a predetermined significance level, it circumvents the challenge of estimating the theoretical constant arising in the rate of convergence of existing thresholding estimators, and hence it is easy to implement and does not require cross-validation. The MT estimator of the sample correlation matrix is shown to be consistent in the spectral and Frobenius norms, and in terms of support recovery, so long as the true covariance matrix is sparse. The performance of the proposed MT estimator is compared to a number of other estimators in the literature using Monte Carlo experiments. It is shown that the MT estimator performs well and tends to outperform the other estimators, particularly when the cross section dimension, N, is larger than the time series dimension, T.

JEL Classifications: C13, C58
Keywords: Sparse correlation matrices, High-dimensional data, Multiple testing, Thresholding, Shrinkage

*The authors would like to thank Alex Chudik, Jianqing Fan, George Kapetanios, Yuan Liao, Ron Smith and Michael Wolf for valuable comments and suggestions, as well as Elizaveta Levina and Martina Mincheva for helpful email correspondence with regard to implementation of their approaches. The authors also wish to acknowledge financial support under ESRC Grant ES/I031626/1.

1 Introduction

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

Importantly, the issue of consistently estimating the population covariance matrix, $\mathbf{\Sigma} = (\sigma_{ij})$, becomes particularly challenging when the number of variables, N, is larger than the number of observations, T. In this case, one way of obtaining a suitable estimator for $\mathbf{\Sigma}$ is to appropriately restrict the off-diagonal elements of its sample estimate denoted by $\hat{\mathbf{\Sigma}}$. Numerous methods have been developed to address this challenge, predominantly in the statistics literature. See Pourahmadi (2011) for an extensive review and references therein. Some approaches are regression-based and make use of suitable decompositions of $\mathbf{\Sigma}$ such as the Cholesky decomposition (see Pourahmadi (1999, 2000), Rothman et al. (2010), Abadir et al. (2014), among others). Others include banding or tapering methods as proposed, for example, by Bickel and Levina (2004, 2008a) and Wu and Pourahmadi (2009), which assume that the variables under consideration follow a natural ordering. Two popular regularisation techniques in the literature that do not make use of any ordering assumptions are those of thresholding and shrinkage.

Thresholding involves setting off-diagonal elements of the sample covariance matrix that are in absolute terms below certain threshold values to zero. This approach includes 'universal' thresholding put forward by El Karoui (2008) and Bickel and Levina (2008b), and 'adaptive' thresholding proposed by Cai and Liu (2011). Universal thresholding applies the same thresholding parameter to all off-diagonal elements of the unconstrained sample covariance matrix, while adaptive thresholding allows the threshold value to vary across the different off-diagonal elements of the matrix. Furthermore, the selected non-zero elements of $\hat{\mathbf{\Sigma}}$ can either be set to their sample estimates or can be adjusted downward. This relates to the concepts of 'hard' and 'soft' thresholding, respectively. The thresholding approach traditionally assumes that the underlying (population) covariance matrix is *sparse*, where sparseness is loosely defined as the presence of a sufficient number of zeros on each row of $\mathbf{\Sigma}$ such that it is absolute summable row (column)-wise, or more generally in the sense defined by El Karoui (2008). However, Fan et al. (2011, 2013) show that such regularisation techniques can be applied even if the underlying population covariance matrix is not sparse, so long as the non-sparseness is characterised by an approximate factor structure. The main challenge in applying this approach lies in the estimation of the thresholding parameter. The method of cross-validation is primarily used for this purpose which has its own limitations 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 - see Friedman (1989). A number of shrinkage targets have been considered in the literature that take advantage of *a priori* knowledge of the data characteristics under investigation. Examples of covariance matrix targets can be found in Ledoit and Wolf (2003), Daniels and Kass (1999, 2001), Fan et al. (2008), and Hoff (2009), among others. Ledoit and Wolf (2004) suggest a modified shrinkage estimator that involves a linear combination of the unrestricted sample covariance matrix with the identity matrix. This is recommended by the authors for more general situations where no natural shrinking target exists. On the whole, shrinkage estimators tend to be stable, but yield inconsistent estimates if the purpose of the analysis is the estimation of the true and false positive rates of the underlying true sparse covariance matrix (the so called 'support recovery' problem).

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

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

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

Notation: We denote the largest and the smallest eigenvalues of the $N \times N$ real symmetric matrix $\boldsymbol{A} = (a_{ij})$ 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 spectral radius by $\varrho(\boldsymbol{A}) = |\lambda_{\max}(\boldsymbol{A})|$, its spectral (or operator) norm by $\|\boldsymbol{A}\|_{spec} = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$, its Frobenius norm by $\|\boldsymbol{A}\|_F = \sqrt{tr(\boldsymbol{A}'\boldsymbol{A})}$. Note that $\|\boldsymbol{A}\|_{spec} = \varrho(\boldsymbol{A})$. $a_N = O(b_N)$ 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

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

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

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

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

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

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

Assumption 2 *The* T *observations* $\{(x_{i1}, x_{j1}), (x_{i2}, x_{j2}), ..., (x_{iT}, x_{jT})\}$ *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 $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$*, and* $E\left(y_{it}^r y_{jt}^s\right) = \mu_{ij}(r,s)$*, for all* $r, s \geq 0$*.*

We follow the hard thresholding literature but, as noted above, we employ multiple testing rather than cross-validation to decide on the threshold value. More specifically, we set to zero those elements of $\boldsymbol{R} = (\rho_{ij})$ that are statistically insignificant and therefore determine the threshold value as a part of a multiple testing strategy rather than by cross-validation. We apply the thresholding procedure explicitly to the correlations rather than the covariances. This has the added advantage that one can use a so-called 'universal' threshold rather than making entry-dependent adjustments, which in turn need to be estimated when thresholding is applied to covariances. This feature is in line with the method of Bickel and Levina (2008b) or El Karoui (2008) but shares the properties of the adaptive thresholding estimator developed by Cai and Lui (2011).

Specifically, denote the sample correlation of x_{it} and x_{jt}, computed over $t = 1, 2, ..., T$, 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. Using the union bound, we have

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

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

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

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

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

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

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

2.1 Theoretical properties of the MT estimator

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

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

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

Then

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

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

where

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

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

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

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 $\lim_{T\to\infty}\left[TVar\left(\hat{\rho}_{ij,T}\right)\right] = K_v(\boldsymbol{\theta}_{ij})$. Finally, the boundedness of $|K_{\rm m}(\boldsymbol{\theta}_{ij})|$ and $K_v(\boldsymbol{\theta}_{ij})$ 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–Schwarz inequalities as given below:

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

and

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

■

Remark 1 *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*

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

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* $G_s(a_{ij,T}|\mathcal{P}_{ij})$, $s = 1, 2, ...$ *are polynomials in* $a_{ij,T}$*, whose coefficients depend on the underlying bivariate distribution of the observations* (x_{it},x_{jt} *for* $t = 1, 2, ..., T$) *which is denoted by* $\mathcal{P}_{ij}$.

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

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

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

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

and

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

Then so long as $N/\sqrt{T} \to 0$ *we have*

$$E\left\|\widetilde{\boldsymbol{R}}_{MT} - \mathbf{R}\right\|_{spec} = O\left(\frac{m_N}{\sqrt{T}}\right), \tag{13}$$

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

Proof. See Appendix. ■

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

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

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

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

Theorem 2 *(Convergence under the Frobenius norm) 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* $\boldsymbol{R} = (\rho_{ij})$*. 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

$$\kappa_{\max} \leq lim_{N\to\infty} \frac{\ln[f(N)]}{\ln(N)}, \tag{14}$$

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

$$E\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F = O\left(\sqrt{\frac{m_N N}{T}}\right), \tag{15}$$

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

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

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

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

Remark 4 *While in practice we find it reasonable to set* $\kappa_{\max}$ *no greater than two, further research is required in determining this data dependent measure, which is beyond the scope of the present paper. Convergence under the spectral norm is less demanding than under the Frobenius norm. Unlike Theorem 2, the statement of Theorem 1 does not require a condition on* $\kappa_{\max}$*. This implies that under the spectral norm, when controlling the errors in estimation of* $\mathbf{R}$*, it is sufficient for* $f(N)$ *to rise linearly with* N *irrespective of the value of* $\kappa_{\max}$*.*

Remark 5 *The orders of convergence in (13) and (15) are in line with the results in the thresholding literature. See, for example, Theorem 1 of Cai and Liu (2011, CL), and Bickel and Levina (2008b, BL), with $q = 0$, that state the convergence rate using the spectral norm in terms of probability,* $\left\| \widetilde{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\|_{spec} = O_p\left(m_N \sqrt{\frac{\log(N)}{T}} \right)$, *where $\widetilde{\mathbf{\Sigma}}$ is the thresholded estimator of $\hat{\mathbf{\Sigma}}$ using either the CL or BL approaches. Similarly, Theorem 2 of Bickel and Levina (2008b), with $q = 0$, using the Frobenius norm under the Gaussianity assumption, obtains a convergence rate of* $\left\| \widetilde{\mathbf{\Sigma}} - \mathbf{\Sigma} \right\|_{F} = O_p\left(\sqrt{\frac{m_N N \log(N)}{T}} \right)$. *In fact (13) and (15) are improvements on the existing rates since the $\log(N)$ factor is absent in both cases. The rate of $O_p\left(\sqrt{\frac{m_N}{T}}\right)$ is achieved in the shrinkage literature as well if the assumption of sparseness is imposed. Here m_N also can be assumed to rise with N in which case the rate of convergence becomes slower. This compares with a rate of $O_p\left(\sqrt{N/T}\right)$ for the sample covariance (correlation) matrix - see Theorem 3.1 in Ledoit and Wolf (2004 - LW). Note that LW use an unconventional definition for the Frobenius norm (see their Definition 1 p. 376).*

Remark 6 *Results (13) and (15) also hold if a concept of 'approximate' sparseness is used in place of Assumption 1, such that m_N is defined more generally as $m_N = \max_{i \leq N} \sum_{j=1}^{N} |\sigma_{ij}|^q$, for some $q \in [0, 1)$. See Bickel and Levina (2008b) or Fan et al. (2013).*

Remark 7 *It is interesting to note that application of the Bonferroni procedure to the problem of testing $\rho_{ij} = 0$ for all $i \neq j$, is equivalent to setting $f(N) = N(N-1)/2$. Our theoretical results suggest that this is too conservative if $\rho_{ij} = 0$ implies x_{it} and x_{jt} are independent, but could be appropriate otherwise. In our Monte Carlo study we consider observations with linear and non-linear dependence, and experiment with $f(N) = N - 1$ and $f(N) = N(N-1)/2$. We find that the simulation results conform closely to our theoretical findings.*

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

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

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

given by

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

$$FPR = \frac{\sum\sum_{i\neq j} I(\tilde{\rho}_{ij,T} \neq 0, \ \text{and } \rho_{ij} = 0)}{\sum\sum_{i\neq j} I(\rho_{ij} = 0)}, \tag{17}$$

where $\hat{\rho}_{ij,T}$ *is the pair-wise correlation coefficient defined by (7),* $c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) > 0$*,* $0 < p < 1$*,* $f(N)$ *is an increasing function such that* $c_p(N) \to \infty$*, as* $N \to \infty$*,* $\ln f(N)/T \to 0$ *and* $c_p(N)/\sqrt{T} \to 0$*, as* N *and* $T \to \infty$*. Suppose also that Assumptions 1-3 hold. Then with probability tending to 1,* $TPR = 1$*, and with probability tending to 1,* $FPR = 0$*, if there exist* N_0 *and* T_0 *such that for* $N > N_0$ *and* $T > T_0$*,* $\sqrt{T}\rho_{\min} - c_p(N) > 0$*, where* $\rho_{\min} = \min_{ij} \left|\rho_{ij}\right| > 0$*.*
Proof. *See Appendix.* ■

Remark 8 *The proof of support recovery does not depend on* $\mu_{ij}(2,2)$*. Also it only requires that* $f(N)$ *rises with* N *linearly. For example, setting* $f(N) = N - 1$ *it is easily seen that* $\ln f(N)/T = \ln(N-1)/T \to 0$*, as* N *and* $T \to \infty$*, and* $c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) \to \infty$*, as* $N \to \infty$*, and conditions of Theorem 3 are met. Interestingly, this suggests that consistent support recovery is ensured if Bonferroni's MT procedure is applied to* $\mathbf{R}$ *(or* $\mathbf{\Sigma}$*) 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.*

3 Monte Carlo simulations

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

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

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

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

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

where $\varepsilon_{it} \sim IIDN(0,1)$, and $\chi^2_{\mathrm{v},t}$ is a chi-squared random variate with $\mathrm{v} > 4$ degrees of freedom, distributed independently of ε_{it} for all i and t. As fourth-order moments are required by Assumption 2 we set $\mathrm{v} = 8$ to ensure that $E\left(y_{it}^4\right)$ exists and $\kappa_{\max} \leq 2$. Note that under $\rho_{ij} = 0$, $\kappa_{ij} = \mu_{ij}(2, 2 \left| \rho_{ij} = 0 \right.) = (\mathrm{v}{-}2)/(\mathrm{v}{-}4)$, and with $\mathrm{v} = 8$ 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 2 are met we must set $f(N) = N\left(N-1\right)/2$. (See also Remark 3 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 2 to be met.

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

$$\boldsymbol{x}_t = \mathbf{a} + \boldsymbol{\gamma} f_t + \mathbf{D}^{1/2}\mathbf{y}_t, \tag{18}$$

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

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

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

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

where $\boldsymbol{A}_1 = \boldsymbol{A} + \epsilon \boldsymbol{I}_{N/2}$, $\boldsymbol{A} = (a_{ij})_{1 \leq i,j \leq N/2}$, $a_{ij} = (1 - \frac{|i-j|}{10})_+$ with $\epsilon = \max(-\lambda_{\min}(A), 0) + 0.01$ 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, $\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}^{-1/2} = \mathbf{P}\mathbf{P}'$, 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 $i = N/2 + 1, N/2 + 1, ..., N$.

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 $N_b = \lceil N^{\delta} \rceil$, and set the remaining middle elements of $\mathbf{b}$ to zero. The resulting population correlation matrix $\boldsymbol{R}$ is defined by

$$\boldsymbol{R} = \boldsymbol{I}_N + \boldsymbol{bb}' - diag\,(\boldsymbol{bb}'), \tag{19}$$

for which $\sqrt{T}\rho_{\min} - c_p(N) > 0$ and $\rho_{\min} = \min_{ij} |\rho_{ij}| > 0$, in line with Theorem 3. The degree of sparseness of $\boldsymbol{R}$ 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 (19). 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 $0 < q < 1$), namely the correlation matrices corresponding to an AR(1) and spatial AR(1), SAR(1), process respectively, along with their associated simulation results can be found in the Supplementary Appendix D.

3.1 Finite sample positive definiteness

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

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

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

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT}\,(\xi) = \xi\boldsymbol{I}_N + (1-\xi)\widetilde{\boldsymbol{R}}_{MT}, \tag{20}$$

with shrinkage parameter $\xi \in (\xi_0, 1]$, and ξ_0 being the minimum value of ξ that produces a non-singular $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0)$ matrix. Alternative ways of computing the optimal weights on the two matrices can be entertained. We choose to calibrate, ξ, since opting to use ξ_0 in (20), as suggested in El Karoui (2008), does not necessarily provide a well-conditioned estimate of $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$. Accordingly, we set ξ by solving the following optimisation problem

$$\xi^* = \arg\min_{\xi_0+\epsilon\leq\xi\leq 1} \left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \right\|_F^2, \tag{21}$$

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

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

Further details on the $S\text{-}MT$ procedure, the optimisation of (21) and choice of reference matrix $\mathbf{R}_0$ are available in the Supplementary Appendix C.

3.2 Alternative estimators and evaluation metrics

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

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

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

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

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

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

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

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

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

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

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

In accordance with the theoretical results in Theorem 2 and in view of Remark 3, we consider two versions of the MT estimator depending on the choice of $f(N) = \{N-1, N(N-1)/2\}$. The $BL_{\hat{C}}$, 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 3.1. $BL_{\hat{C}^*}$ and $CL_{\hat{C}^*}$ estimators are obtained by applying the additional FLM adjustments. The shrinkage estimator, $LW_{\hat{\Sigma}}$, is invertible by construction. In the case of the MT estimators where regularisation is performed on the correlation matrix the associated covariance matrix is estimated as $\hat{\mathbf{D}}^{1/2}\widetilde{\boldsymbol{R}}_{MT}\hat{\mathbf{D}}^{1/2}$.

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

$$\left\|\boldsymbol{\Sigma}-\mathring{\boldsymbol{\Sigma}}\right\|_{spec} \text{ and } \left\|\boldsymbol{\Sigma}-\mathring{\boldsymbol{\Sigma}}\right\|_{F}, \tag{22}$$

where $\mathring{\boldsymbol{\Sigma}}$ is set to one of the following estimators $\{\widetilde{\boldsymbol{\Sigma}}_{MT_{N-1}}, \widetilde{\boldsymbol{\Sigma}}_{MT_{N(N-1)/2}}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{CL,2}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT_{N-1}}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT_{N(N-1)/2}}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}^*}, \hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. The threshold values, $\hat{C}$ and $\hat{C}^*$, are obtained by cross-validation (see Supplementary Appendix B.3 for details). Both norms are also computed for the difference between $\boldsymbol{\Sigma}^{-1}$, the population inverse of $\boldsymbol{\Sigma}$, and the estimators $\{\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT_{N-1}}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT_{N(N-1)/2}}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\boldsymbol{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. Further, we investigate the ability of the thresholding estimators to recover the support of the true covariance matrix via the true positive rate (TPR) and false positive rate (FPR), as defined by (16) and (17), respectively. The statistics TPR and FPR are not relevant to the shrinkage estimator $LW_{\hat{\Sigma}}$ and will not be reported for this estimator.

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

We begin by investigating the sensitivity of the MT estimator to the choice of the p-value, p, and the scaling factor $f(N)$ used in the formulation of $c_p(N)$ defined by (6). For this purpose we consider the typical significance levels used in the literature, namely $p = \{0.01, 0.05, 0.10\}$, and $f(N) = \{N-1, N(N-1)/2\}$. Table 1 summarises the spectral and Frobenius norm losses (averaged over 2000 replications) for both Monte Carlo designs A and B, and for both distributional error assumptions (Gaussian and multivariate t). First, we note that neither of the norms is much affected by the choice of the p values when the scaling factor is $N(N-1)/2$, irrespective of whether the observations are drawn from a Gaussian or a multivariate t distribution. Perhaps this is to be expected since for N sufficiently large the effective p-value which is given by $2p/N(N-1)$ is very small and the test outcomes are more likely to be robust to the changes in the values of p as compared to the case when the scaling factor used is $N-1$. The results in Table 1 also confirm our theoretical finding of Theorem 2 that in the case of Gaussian observations, where $\kappa_{\max} = 1$, the scaling factor $N-1$ is likely to perform better as compared to $N(N-1)/2$, but the reverse is true if the observations are multivariate t distributed and the scaling factor $N(N-1)/2$ is to be preferred (see also Remark 3). We also note that all the norm losses rise with N given that T is kept at 100 in all the experiments. We obtain similar results when we consider other Monte Carlo designs with approximately sparse covariance matrices. To save space the results for these designs are provided in the Supplementary Appendix D. Overall, we find that the results are more robust when the scaling factor $N(N-1)/2$ is used.

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

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

Tables 2 and 3 summarise the results for the Monte Carlo designs A and B, respectively. Based on the results of Section 3.3, we provide norm comparisons for the MT estimator using the scaling factor $N(N-1)/2$, and the conventional significance level of $p = 0.05$. Initially, we consider the threshold estimators, MT, BL and the two versions of the CL estimators (CL_2 and $CL_{\hat{C}}$) 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 $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}$.

3.5 Norm comparisons of inverse estimators

Although the theoretical focus of this paper has been on estimation of $\mathbf{\Sigma}$ rather than its inverse, it is still of interest to see how well $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT}$, $\widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}$, $\widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}$, and $\hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}$ estimate $\mathbf{\Sigma}^{-1}$, assuming that $\mathbf{\Sigma}^{-1}$ is well defined. Table 4 provides average norm losses for Monte Carlo design B whose $\mathbf{\Sigma}$ is positive definite. $\mathbf{\Sigma}$ for design A is ill-conditioned and will not be considered any further here. As can be seen from the results in Table 4, $\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT}$ 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 appears to be less sensitive to the 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.

3.6 Support recovery statistics

Table 5 reports the true positive and false positive rates (TPR and FPR) for the support recovery of $\mathbf{\Sigma}$ using the multiple testing and thresholding estimators. In the comparison set we include 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 $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}$. Again we use 100 replications due to the use of cross-validation in the implementation of BL and CL thresholding. We include the MT estimators for both choices of the scaling factor, $f(N) = N-1$ and $f(N) = N(N-1)/2$, computed at $p = 0.05$, to see if our theoretical result, namely that for consistent support recovery only the linear scaling factor, $N-1$, is needed, is borne out by the simulations. For consistent support recovery we would like to see

FPR values near zero and TPR values near unity. As can be seen from Table 5, the FPR values of all estimators are very close to zero, so any comparisons of different estimators must be based on the TPR values. Comparing the results for $\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}$ and $\widetilde{\mathbf{\Sigma}}_{MT_{N(N-1)/2}}$ we find that as predicted by the theory (Theorem 3 and Remark 8), TPR values of $\widetilde{\mathbf{\Sigma}}_{MT_{N-1}}$ are closer to unity as compared to the TPR values of $\widetilde{\mathbf{\Sigma}}_{MT_{N(N-1)/2}}$. Similar results are obtained for the MT estimators for different choices of the p values. Table 6 provides results for $p = \{0.01, 0.05, 0.10\}$, and for $f(N) = \{N-1, N(N-1)/2\}$ using 2,000 replications. In this table it is further evident that, in line with the conclusions of Section 3.3, both the TPR and the FPR statistics are relatively robust to the choice of the p values irrespective of the scaling factor, $f(N)$, or whether the observations are drawn from a Gaussian or a multivariate t distribution. This is especially true under design B, since for this specification we explicitly control for the number of non-zero elements in $\mathbf{\Sigma}$, that ensures the conditions of Theorem 3 are met.

Turning to a comparison with other estimators in Table 5, we find that the MT and CL estimators perform substantially better than the BL estimator. Further, allowing for 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, again note that TPR values are higher for design B. 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.

3.7 Computational demands of the different thresholding methods

Table 7 reports the relative execution times of the different thresholding methods studied. All times are relative to the time it takes to carry out the computations for the $MT_{N(N-1)/2}$ estimator. It took 0.010, 0.013, and 0.016 seconds to apply the MT method in Matlab to a sample of $N = \{30, 100, 200\}$, respectively, and $T = 100$ observations using a desktop pc. The slight difference in execution time between MT_{N-1} and $MT_{N(N-1)/2}$ amounts to the stricter condition imposed by the p-value on the $MT_{N(N-1)/2}$ procedure, which produces a slightly sparser version of $\widetilde{\mathbf{\Sigma}}$. In contrast, the $BL_{\hat{C}}$ and $CL_{\hat{C}}$ thresholding approaches are computationally much more demanding. Their computations took between about 18 and 412293 times longer than the MT approach, for the same sample sizes and computer hardware. The $BL_{\hat{C}}$ method was less demanding than the $CL_{\hat{C}}$ method - it took between about 18 and 500 times longer than the MT approach. Even CL_2, which does not require estimation of the threshold parameter, took up to 17 times longer than the MT approach. Thus, compared with other thresholding methods MT has a clear computational advantage.

4 Concluding Remarks

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

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

The numerical properties of the proposed estimator are investigated using Monte Carlo simulations. It is shown that the MT estimator performs well, and generally better than the other estimators proposed in the literature. The simulations also show that in terms of spectral and Frobenius norm losses, the MT estimator is reasonably robust to the choice of p in the threshold criterion, $\left|\hat{\rho}_{ij}\right| > T^{-1/2}\Phi^{-1}\left(1 - \frac{p}{2f(N)}\right)$, 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 $T = 100$

Monte Carlo design A						
	$f(N) = N-1$			$f(N) = N(N-1)/2$		
N	$MT_{N-1}(.01)$	$MT_{N-1}(.05)$	$MT_{N-1}(.10)$	$MT_{\frac{N(N-1)}{2}}(.01)$	$MT_{\frac{N(N-1)}{2}}(.05)$	$MT_{\frac{N(N-1)}{2}}(.10)$
$\mathbf{u}_{it}\sim$ 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)

Note: Norm losses are averages over 2,000 replications. Simulation standard deviations are given in parentheses. The MT estimators are defined in Section 3.2.

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

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
			$\mathbf{u}_{it}\sim$ Gaussian			
	Error matrices $(\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}})$					
$MT_{N(N-1)/2}$	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 $(\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}})$					
$MT_{N(N-1)/2}$	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)

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

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

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
			$\mathbf{u}_{it}\sim$ Gaussian			
	Error matrices $(\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}})$					
$MT_{N(N-1)/2}$	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 $(\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}})$					
$MT_{N(N-1)/2}$	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 note to Table 2.

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

	$N=30$		$N=100$		$N=200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices $(\mathbf{\Sigma}^{-1} - \mathring{\mathbf{\Sigma}}^{-1})$					
			$\mathbf{u}_{it}\sim$ **Gaussian**			
$S\text{-}MT_{N(N-1)/2}$	4.42(1.22)	2.66(0.31)	15.62(2.68)	5.87(0.46)	13.89(2.29)	5.45(0.36)
$BL_{\hat{C}^*}$	$3.8\times10^3(2.4\times10^4)$	19.56(58.88)	$1.2\times10^3(1.1\times10^4)$	12.16(33.25)	41.07(143.74)	7.66(3.17)
$CL_{\hat{C}^*}$	$1.9\times10^3(1.7\times10^4)$	10.92(42.39)	51.99(241.39)	8.16(4.23)	28.45(24.37)	7.35(1.11)
$LW_{\hat{\Sigma}}$	11.03(0.58)	4.26(0.09)	31.04(0.64)	8.62(0.06)	31.81(0.21)	9.40(0.05)
			$\mathbf{u}_{it}\sim$ **multivariate** $t-$**distributed with 8 degrees of freedom**			
$S\text{-}MT_{N(N-1)/2}$	3.42(1.55)	2.44(0.38)	12.39(3.01)	5.49(0.54)	11.23(4.12)	5.54(0.66)
$BL_{\hat{C}^*}$	$157.26(1.0\times10^3)$	6.11(11.28)	$349.35(3.1\times10^3)$	9.80(17.03)	28.58(22.06)	7.77(1.04)
$CL_{\hat{C}^*}$	85.82(546.85)	5.53(7.84)	$517.27(4.8\times10^3)$	10.07(21.25)	25.61(3.55)	7.54(0.50)
$LW_{\hat{\Sigma}}$	12.08(1.19)	4.48(0.20)	31.78(1.32)	8.74(0.23)	32.06(1.00)	9.50(0.33)

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

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

		Monte Carlo design A							Monte Carlo design B				
N		MT_{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

Note: TPR is the true positive and FPR is the false positive rates defined by (16) and (17), respectively. MT estimators are computed with $p = 0.05$. For a description of other estimators see the note to Table 2. The TPR and FPR numbers are averages over 100 replications.

Table 6: Support recovery statistics for the multiple testing estimator computed with $p = 0.01, 0.05, 0.10$ - $T = 100$

		Monte Carlo design A								Monte Carlo design B					
		$p = 0.01$		$p = 0.05$		$p = 0.10$				$p = 0.01$		$p = 0.05$		$p = 0.10$	
N		MT_{N-1}	$MT_{N(N-1)/2}$	MT_{N-1}	$MT_{N(N-1)/2}$	MT_{N-1}	$MT_{N(N-1)/2}$	N		MT_{N-1}	$MT_{N(N-1)/2}$	MT_{N-1}	$MT_{N(N-1)/2}$	MT_{N-1}	$MT_{N(N-1)/2}$
							$\mathbf{u}_{it}\sim$ Gaussian								
30	TPR	0.75	0.69	0.80	0.73	0.81	0.75	30	TPR	1.00	0.98	1.00	0.99	1.00	1.00
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
100	TPR	0.65	0.55	0.69	0.59	0.71	0.61	100	TPR	1.00	0.98	1.00	0.99	1.00	0.99
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
200	TPR	0.62	0.51	0.66	0.54	0.68	0.56	200	TPR	0.99	0.94	1.00	0.97	1.00	0.98
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
							$\mathbf{u}_{it}\sim$ multivariate $t-$distributed with 8 degrees of freedom								
30	TPR	0.76	0.69	0.80	0.73	0.82	0.75	30	TPR	0.99	0.97	1.00	0.99	1.00	0.99
	FPR	0.00	0.00	0.01	0.00	0.01	0.00		FPR	0.00	0.00	0.01	0.00	0.01	0.00
100	TPR	0.65	0.56	0.69	0.59	0.71	0.61	100	TPR	0.99	0.96	1.00	0.98	1.00	0.99
	FPR	0.00	0.00	0.00	0.00	0.01	0.00		FPR	0.00	0.00	0.00	0.00	0.01	0.00
200	TPR	0.62	0.51	0.66	0.55	0.68	0.56	200	TPR	0.99	0.92	0.99	0.96	1.00	0.97
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00

Note: TPR is the true positive and FPR is the false positive rates defined by (16) and (17), respectively. The TPR and FPR numbers are averages over 2,000 replications.

Table 7: Relative execution time for the different thresholding methods

	$T=100$		
	$N=30$	$N=100$	$N=200$
$MT_{N(N-1)/2}$	1.000	1.000	1.000
MT_{N-1}	1.176	1.239	1.110
$BL_{\hat{C}}$	18.25	102.2	500.9
CL_2	1.304	5.746	17.39
$CL_{\hat{C}}$	1278	59907	412293

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

Appendix: Mathematical proofs of theorems for the MT estimator

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

Proof of Theorem 1. Consider the spectral norm,

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

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

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

Also

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

Hence,

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

and

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

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

$$\begin{aligned}E\left(\sum_j\left|\tilde{\rho}_{ij,T}-\rho_{ij}\right|\right)\leq\ & E\left[\sum_j\left|(\hat{\rho}_{ij,T}-\rho_{ij})I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\right)\right|\right]+\\ & E\left[\sum_j\left|\rho_{ij}I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|\leq c_p(N)\right)\right|\right],\end{aligned}\qquad(23)$$

or

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

where

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

Consider now the orders of these three terms $\mathcal{A}_i, \mathcal{B}_i$, and $\mathcal{C}_i$ in turn, starting with $\mathcal{A}_i$. Since under Assumption 2, $0<\rho_{\min}<\left|\rho_{ij}\right|<\rho_{\max}<1$, then uniformly over all i,

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

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

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

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

Consider now $\mathcal{B}_i$ and note that since $\hat{\rho}_{ij,T}=\omega_{ij,T}z_{ij,T}+\rho_{ij,T}$ (to simplify the notation we use $\omega_{ij,T}^2$ 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 inequality, $\mathcal{B}_i\leq\mathcal{B}_{i1}+\mathcal{B}_{i2}$, where

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

Using (8) and (9)

$$\omega_{ij,T} = \frac{K_v^{1/2}(\boldsymbol{\theta}_{ij})}{T^{1/2}}+O\left(T^{-3/2}\right), \tag{24}$$

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

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

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

By the Cauchy-Schawrz inequality, we have

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

since

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

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

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

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

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

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

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

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

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

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

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

where $\kappa_{ij} = \left[\mu_{ij}(2,2)\,|\rho_{ij}=0\right]$ 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 $\mathcal{C}_{i2} \to 0$ as N and $T \to \infty$, uniformly for all i, since ψ_{ij}^2 and κ_{ij} are bounded and $c_p(N) \to \infty$.

Finally, considering $\mathcal{C}_{i1}$ we note that (since $\omega_{ij,T} > 0$),

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

and by the Cauchy-Schawrz inequality,

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

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

Collecting the results for the orders of convergence of $\mathcal{A}_i, \mathcal{B}_i$, and $\mathcal{C}_i$ given above, overall we obtain a convergence rate of order $O(\frac{m_N}{T^{1/2}})$ uniformly for all i, where $i = 1,2,...,N$. Therefore, (13) follows as required. ■

Proof of Theorem 2. Consider the squared Frobenius norm,

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

and recall that

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

Hence

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

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

$$\sum\sum_{i\neq j}E\left(\tilde{\rho}_{ij,T}-\rho_{ij}\right)^2=A+B+C, \tag{26}$$

where

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

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

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

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

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

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

Consider now B. Recalling that $\hat{\rho}_{ij,T} = \omega_{ij,T} z_{ij,T} + \rho_{ij,T}$ we have the following decomposition of B, $B = B_1 + B_2 + 2B_3$, where

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

Again, using (8) and (9)

$$\omega_{ij,T}^2 = \frac{K_v(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right), \tag{27}$$

$$\left(\rho_{ij,T} - \rho_{ij}\right)^2 = \frac{K_{\mathrm{m}}^2(\boldsymbol{\theta}_{ij})}{T^2} + O\left(T^{-3}\right), \tag{28}$$

$$\left(\rho_{ij,T} - \rho_{ij}\right)\omega_{ij,T} = \frac{K_v^{1/2}(\boldsymbol{\theta}_{ij}) K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{T^{3/2}} + O\left(T^{-5/2}\right). \tag{29}$$

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

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

Since $\sup_{ij} K_v(\boldsymbol{\theta}_{ij})$ 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 $O\left(\frac{N m_N}{T}\right)$. 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

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

and

$$\begin{aligned} B_3 &= \sum\sum_{i\neq j, \rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right) \omega_{ij,T} E\left\{ z_{ij,T} \left[I \left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| > c_p(N) \left| \rho_{ij} \neq 0 \right.\right)\right]\right\} \\ &= \sum\sum_{i\neq j, \rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right) \omega_{ij,T} E\left\{ z_{ij,T} - z_{ij,T} \left[I \left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) \left| \rho_{ij} \neq 0 \right.\right)\right]\right\} \\ &= -\sum\sum_{i\neq j, \rho_{ij}\neq 0} \left(\rho_{ij,T} - \rho_{ij}\right) \omega_{ij,T} E\left[z_{ij,T} I \left(\left| \sqrt{T}\hat{\rho}_{ij,T} \right| \leq c_p(N) \left| \rho_{ij} \neq 0 \right.\right)\right]. \end{aligned} \tag{30}$$

Also, from Lemma 4

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

and from Lemma 2

$$\begin{aligned} E\left[zI\left(L_{ij,T} \leq z \leq U_{ij,T}\left|\rho_{ij} \neq 0\right.\right)\right] &= \phi\left(\frac{-c_p(N) - \sqrt{T}\rho_{ij} + O\left(\frac{1}{\sqrt{T}}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(\frac{1}{T}\right)}}\right) \\ &\quad -\phi\left(\frac{c_p(N) - \sqrt{T}\rho_{ij} + O\left(\frac{1}{\sqrt{T}}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(\frac{1}{T}\right)}}\right), \end{aligned} \tag{31}$$

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 (29) and noting that $K_v^{-1/2}(\boldsymbol{\theta}_{ij})K_{\mathrm{m}}(\theta_{ij})$ 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 (26):

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

Starting with the simpler terms, we first note that

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

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 $\psi_{ij} = \left[\mu_{ij}(3,1)+\mu_{ij}(1,3)\right]/2$. 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

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

But using Lemma 4, Lemma 2 and (31) 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

$$\begin{aligned}
C_1 &= \sum\sum_{i\neq j,\rho_{ij}=0} \omega_{ij,T}^2 E\left[z_{ij,T}^2 I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \\
&= \sum\sum_{i\neq j,\rho_{ij}=0} \omega_{ij,T}^2 E\left\{z_{ij,T}^2\left[1 - I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\} \\
&\leq \frac{N(N-m_N-1)}{T} \sup_{ij}\left[\kappa_{ij} + O\left(\frac{1}{T}\right)\right] \\
&\quad \times \sup_{ij} E\left\{z_{ij,T}^2\left[1 - I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\}. \qquad (32)
\end{aligned}$$

But using Lemma 4

$$\begin{aligned}
&\lim_{T\to\infty} E\left\{z_{ij,T}^2\left[1 - I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\} \qquad (33) \\
= &\lim_{T\to\infty} E\left\{z^2\left[1 - I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\left|\rho_{ij}=0\right.\right)\right]\right\},
\end{aligned}$$

and then Lemma 2

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

where $U_{ij,T}(0)$ and $L_{ij,T}(0)$ are given by (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

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

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

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

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

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

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

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

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

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

or

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

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

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

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 (15) follows as desired. ■

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

$$
FPR = |FPR| = \frac{\sum\sum_{i\neq j} I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)|\rho_{ij}=0\right)}{N(N-m_N-1)}. \tag{34}
$$

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

Taking the expectation of (34) 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 (16) 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

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

Since $|X| = X$, then

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

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

$$\lim_{N,T\to\infty} P(|TPR - 1| > \delta) \leq \frac{1}{\delta} \lim_{N,T\to\infty} \sup_{ij} \Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| < c_p(N)|\rho_{ij} \neq 0\right). \tag{35}$$

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

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

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

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

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

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

Hence, more generally $\lim_{N,T\to\infty}\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| < c_p(N)|\rho_{ij}\neq 0\right) = 0$, if $c_p(N) - \sqrt{T}\left|\rho_{ij}\right| \to -\infty$, for all $\rho_{ij}\neq 0$, or equivalently if $\sqrt{T}\rho_{\min} - c_p(N) \to \infty$, where $\rho_{\min} = \min_{ij}\left|\rho_{ij}\right|$ 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 (35) 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., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of American Statistical Association, 96, 1348–1360.

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.

Friedman, J. H. (1989). Regularized discriminant analysis. Journal of the American Statistical Association, 84, 165–175.

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

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

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

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

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

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. Econometric Reviews, 34, 1089–1117.

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.

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.

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

Supplementary appendix to: 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

December 14, 2015

Appendix A Technical Lemmas

A.1 Statement of technical lemmas

We begin by stating a few technical lemmas that are needed for the proof of the main results.

Lemma 1 *Consider the sample correlation coefficient, $\hat{\rho}_{ij,T}$, defined by (7) and suppose that Assumptions 2 and 3 hold. Then*

$$\lim_{a_{ij,T}\rightarrow\pm\infty}\left\{e^{\frac{1-\epsilon}{2}a_{ij,T}^{2}}\left[F_{ij,T}(a_{ij,T}\left|\mathcal{P}_{ij}\right.)-\Phi(a_{ij,T})\right]\right\}=0, \tag{A.1}$$

for some small positive ϵ.

Lemma 2 *Suppose that $z\sim N(0,1)$, then*

$$E\left[zI(L\leq z\leq U)\right]=\phi\left(L\right)-\phi(U), \tag{A.2}$$

and

$$E\left[z^{2}I(L\leq z\leq U)\right]=\left[\Phi\left(U\right)-\Phi\left(L\right)\right]+L\phi(L)-U\phi(U). \tag{A.3}$$

Lemma 3 *Let $c_p(N)=\Phi^{-1}\left(1-\frac{p}{2f(N)}\right)$, where $0<p<1$, $f(N)$ is an increasing function of N, and suppose there exist finite T_0 and N_0 such that for all $N>N_0$*

$$1-\frac{p}{2f(N)}>0, \tag{A.4}$$

and as N and $T\rightarrow\infty$

$$\frac{\ln f(N)}{T}\rightarrow 0. \tag{A.5}$$

Then

$$c_p(N)\leq\sqrt{2\left[\ln f(N)-\ln(p)\right]}, \tag{A.6}$$

and for all $N>N_0$ and $T>T_0$, $c_p(N)/\sqrt{T}$ is bounded and

$$\frac{c_p(N)}{\sqrt{T}}\rightarrow 0, \tag{A.7}$$

as N and $T\rightarrow\infty$.

Lemma 4 *Consider the standardised sample correlation coefficient $z_{ij,T}=\frac{\left[\hat{\rho}_{ij,T}-E\left(\hat{\rho}_{ij,T}\right)\right]}{\sqrt{Var\left(\hat{\rho}_{ij,T}\right)}}$, where $\hat{\rho}_{ij,T}$ is defined by (7) and $E\left(\hat{\rho}_{ij,T}\right)$ and $Var\left(\hat{\rho}_{ij,T}\right)>0$ are given by (8) and (9), respectively. Suppose that $c_p(N)=\Phi^{-1}\left(1-\frac{p}{2f(N)}\right)$, and conditions (A.4) and (A.5) hold. Then for all i and j, there exist N_0 and T_0 such that for $N>N_0$ and $T>T_0$*

$$\begin{aligned}\lim_{T\rightarrow\infty}E\left\{z_{ij,T}^{s}\left[I\left(\left|\hat{\rho}_{ij,T}\right|\leq\frac{c_p(N)}{\sqrt{T}}\right)\right]\right\} &= \lim_{T\rightarrow\infty}E\left[z_{ij,T}^{s}I\left(L_{ij,T}\leq z_{ij,T}\leq U_{ij,T}\right)\right]\\ &= \lim_{T\rightarrow\infty}E\left[z^{s}I\left(L_{ij,T}\leq z\leq U_{ij,T}\right)\right],\end{aligned} \tag{A.8}$$

for $s=0,1,2,\ldots$, where

$$U_{ij,T}=\frac{c_p(N)-\sqrt{T}E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}},\ L_{ij,T}=\frac{-c_p(N)-\sqrt{T}E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}} \tag{A.9}$$

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

Lemma 5 *Consider the cumulative distribution function of a standard normal variate, defined by*

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du.$$

Then for $x > 0$

$$\Phi(-x) = 1 - \Phi(x) \leq \frac{1}{2} \exp(-\frac{x^2}{2}). \tag{A.10}$$

Lemma 6 *Consider the sample correlation coefficient,* $\hat{\rho}_{ij,T}$*, defined by (7) and suppose that Assumptions 2 and 3 hold, then there exists* N_0 *and* T_0 *such that for all* $N > N_0$ *and* $T > T_0$ [1]

$$\Pr\left(\left| \sqrt{T} \hat{\rho}_{ij,T} \right| > c_p(N) \left| \rho_{ij} = 0 \right. \right) \leq K e^{-\frac{1-\epsilon}{2} \frac{c_p^2(N)}{\kappa_{ij}}} [1 + o(1)] \tag{A.11}$$

where $\kappa_{ij} = \left[\mu_{ij}(2,2) \left| \rho_{ij} = 0 \right. \right]$*,* $\mu_{ij}(2,2)$ *is defined under Assumption 2, and* ϵ *is a small positive constant. Further, if* $\left| \rho_{ij} \right| > c_p(N)/\sqrt{T}$ *we have*

$$\Pr\left(\left| \sqrt{T} \hat{\rho}_{ij,T} \right| < c_p(N) | \rho_{ij} \neq 0 \right) \leq K e^{\frac{-1}{2} \frac{T\left(\left| \rho_{ij} \right| - \frac{c_p(N)}{\sqrt{T}} \right)^2}{K_v(\boldsymbol{\theta}_{ij})}} [1 + o(1)], \tag{A.12}$$

where $K_v(\boldsymbol{\theta}_{ij})$ *is given by (11),*

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)} \right) > 0, \tag{A.13}$$

$0 < p < 1$*, and* $f(N)$ *is an increasing function of* N *such that*

$$\ln f(N)/T \to 0, \text{ as } N \text{ and } T \to \infty. \tag{A.14}$$

Lemma 7 *Consider the data generating process*

$$\mathbf{y}_t = \mathbf{P}\mathbf{u}_t,$$

where $\mathbf{y}_t$ *and* $\mathbf{u}_t$ *are* $N \times 1$ *vectors of random variables, and* $\mathbf{P}$ *is an* $N \times N$ *matrix of fixed constants, such that* $\mathbf{P}\mathbf{P}' = \mathbf{R}$*, where* $\mathbf{R}$ *is a correlation matrix. Suppose that* $\mathbf{u}_t$ *follows a multivariate t-distribution with v degrees of freedom generated as*

$$\mathbf{u}_t = \left(\frac{\mathrm{v} - 2}{\chi^2_{v,t}} \right)^{1/2} \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})' \sim IIDN(\mathbf{0}, \mathbf{I}_N)$*, and* $\chi^2_{\mathrm{v},t}$ *is a chi-squared random variate with* $\mathrm{v} > 4$ *degrees of freedom distributed independently of* $\boldsymbol{\varepsilon}_t$*. Then we have that*

$$\mu_{ij}(2,2) = E(y_{it}^2 y_{jt}^2) = \frac{(\mathrm{v} - 2)\left[(\mathbf{p}_i'\mathbf{p}_i)^2 + (\mathbf{p}_i'\mathbf{p}_j)^2 \right]}{(\mathrm{v} - 4)},$$

where $\mathbf{p}_i'$ *is the* i^{th} *row of* $\mathbf{P}$*. In the case where* $\mathbf{P} = \mathbf{I}_N$*,* $\mu_{ij}(2,2) = (\mathrm{v} - 2)/(\mathrm{v} - 4)$ *and*

$$E(y_{it}^2 y_{jt}) = E(y_{jt}^2 y_{it}) = 0.$$

Lemma 8 *Fat-tailed shocks do not necessarily generate* $\mu_{ij}(2,2) > 1$*.*

[1] To simplify the notation we have dropped explicit reference to $\mathcal{P}_{ij}$, the underlying bivariate distribution of the observations.

A.2 Proofs of lemmas for the MT estimator

Proof of Lemma 1. Under (12), and noting that

$$e^{\frac{1-\epsilon}{2}a_{ij,T}^2}\phi\left(a_{ij,T}\right)=e^{\frac{1-\epsilon}{2}a_{ij,T}^2}(2\pi)^{-1/2}\exp\left(-\frac{1}{2}a_{ij,T}^2\right)=(2\pi)^{-1/2}\exp\left(-\frac{\epsilon}{2}a_{ij,T}^2\right),$$

we have

$$\begin{aligned} e^{\frac{1-\epsilon}{2}a_{ij,T}^2}\left[F_{ij,T}(a_{ij,T}\,|\mathcal{P}_{ij})-\Phi(a_{ij,T})\right] &= (2\pi)^{-1/2}\exp\left(-\frac{\epsilon}{2}a_{ij,T}^2\right) \\ &\times\left[T^{-1/2}G_1\left(a_{ij,T}\,|\mathcal{P}_{ij}\right)+T^{-1}G_2\left(a_{ij,T}\,|\mathcal{P}_{ij}\right)+....,\right]. \end{aligned}$$

and the desired result follows noting that $a_{ij,T}^s\exp\left(-\frac{\epsilon}{2}a_{ij,T}^2\right)\to 0$ as $a_{ij,T}\to\pm\infty$, for all $s\geq 0$. This result holds for a fixed T, and as $T\to\infty$. ■

Proof of Lemma 2. Denote the density of the standard normal distribution by $\phi(z)=(2\pi)^{-1/2}e^{-(1/2)z^2}$, then

$$E\left[zI(L\leq z\leq U)\right]=\int_L^U z(2\pi)^{-1/2}e^{-(1/2)z^2}dz=\left[-\phi(z)\right]_L^U=\phi\left(L\right)-\phi(U).$$

Similarly, to prove (A.3) note that $E\left[z^2I(L\leq z\leq U)\right]=\int_L^U z^2\phi(z)dz$. Hence, integrating by parts, we have

$$\int_L^U z^2\phi(z)dz=\left[-z\phi(z)\right]_L^U+\int_L^U\phi(z)dz=\left[\Phi\left(U\right)-\Phi\left(L\right)\right]+L\phi(L)-U\phi(U),$$

as required. ■

Proof of Lemma 3. First note that

$$\Phi^{-1}\left(z\right)=\sqrt{2}\,\mathrm{erf}^{-1}(2z-1),\ \ z\in(0,1),$$

where $\Phi(x)$ is cumulative distribution function of a standard normal variate, and $\mathrm{erf}(x)$ is the error function defined by

$$\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\textstyle\int_0^x e^{-u^2}du. \tag{A.15}$$

Consider now the inverse complementary error function $\mathrm{erfc}^{-1}(x)$ given by

$$\mathrm{erf\,c}^{-1}(1-x)=\mathrm{erf}^{-1}(x).$$

Using results in Chiani et al. (2003, p.842) we have

$$\mathrm{erf\,c}^{-1}(x)\leq\sqrt{-\ln(x)}.$$

Applying the above results to $c_p(N)$ we have

$$\begin{aligned} c_p(N) &= \Phi^{-1}\left(1-\frac{p}{2f(N)}\right) \\ &= \sqrt{2}\,\mathrm{erf}^{-1}\left[2\left(1-\frac{p}{2f(N)}\right)-1\right] \\ &= \sqrt{2}\,\mathrm{erf}^{-1}\left(1-\frac{p}{f(N)}\right)=\sqrt{2}\,\mathrm{erf\,c}^{-1}\left(\frac{p}{f(N)}\right) \\ &\leq \sqrt{2}\sqrt{-\ln\left(\frac{p}{f(N)}\right)}=\sqrt{2\left[\ln f(N)-\ln(p)\right]}. \end{aligned}$$

Hence, in view of condition (A.5), and noting that p is fixed, then $c_p(N)\sqrt{T}$ is bounded in N and T, and result (A.7) follows noting that $c_p(N)/\sqrt{T} \leq \sqrt{2\left[\ln f(N) - \ln(p)\right]/T} \to 0$, as N and $T \to \infty$.

■

Proof of Lemma 4. We first note that since $Var\left(\hat{\rho}_{ij,T}\right) > 0$

$$\begin{aligned} I\left(\left|\hat{\rho}_{ij,T}\right| \leq \frac{c_p(N)}{\sqrt{T}}\right) &= I\left(\frac{-c_p(N)}{\sqrt{T}} \leq \hat{\rho}_{ij,T} \leq \frac{c_p(N)}{\sqrt{T}}\right) \\ &= I\left(\frac{\frac{-c_p(N)}{\sqrt{T}} - E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\hat{\rho}_{ij,T}\right)}} \leq \frac{\hat{\rho}_{ij,T} - E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\hat{\rho}_{ij,T}\right)}} \leq \frac{\frac{c_p(N)}{\sqrt{T}} - E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\hat{\rho}_{ij,T}\right)}}\right) \\ &= I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\right). \qquad \text{(A.16)} \end{aligned}$$

Also, since $\hat{\rho}_{ij,T}$ is a correlation coefficient, $\left|\hat{\rho}_{ij,T}\right| < 1$, and for a finite $T > T_0$, $Var\left(\hat{\rho}_{ij,T}\right) > 0$, then

$$\left|z_{ij,T}\right| < \frac{\left|\hat{\rho}_{ij,T}\right| + \left|E\left(\hat{\rho}_{ij,T}\right)\right|}{\sqrt{Var\left(\hat{\rho}_{ij,T}\right)}} < 2\sup_{i,j}\left(\frac{1}{\sqrt{Var\left(\hat{\rho}_{ij,T}\right)}}\right) < K.$$

Hence all moments of $z_{ij,T}$ exist for T finite. Furthermore, it is well known that $z_{ij,T} \to_d N(0,1)$ as $T \to \infty$. Therefore, all moments of $z_{ij,T}$ exist for all values of $T > T_0$, and by the *second limit-theorem* (see, for example, Rao and Kendall (1950, p. 228))

$$E\left(z_{ij,T}^s\right) \to E\left(z^s\right), \text{ as } T \to \infty, \text{ for all } s = 1, 2, ...$$

Furthermore, since $I\left(L_{ij,T} \leq z_{ij,T} \leq U_{ij,T}\right) = I\left(\left|\hat{\rho}_{ij,T}\right| \leq \frac{c_p(N)}{\sqrt{T}}\right) \leq c_p(N)/\sqrt{T}$, and under conditions (A.4) and (A.5), $c_p(N)/\sqrt{T}$ is bounded (see Lemma 3). Then for all $N > N_0$ we must also have

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

as required. ■

Proof of Lemma 5. Using results in Chiani et al. (2003, eq. (5)) we have

$$\text{erf c}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-u^2} du \leq \exp(-x^2), \qquad \text{(A.17)}$$

where erf $c(x)$ is the complement of the erf(x) function defined by (A.15). But

$$1 - \Phi(x) = (2\pi)^{-1/2}\int_x^\infty e^{-\frac{u^2}{2}} du = \frac{1}{2}\text{erf c}\left(\frac{x}{\sqrt{2}}\right),$$

and using (A.17) we have

$$1 - \Phi(x) = \frac{1}{2}\text{erf c}\left(\frac{x}{\sqrt{2}}\right) \leq \frac{1}{2}\exp\left[-\left(\frac{x}{\sqrt{2}}\right)^2\right] = \frac{1}{2}\exp\left(-\frac{x^2}{2}\right).$$

■

Proof of Lemma 6. We first note that

$$\begin{aligned} \Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| \leq c_p(N)\right) &= \Pr\left(-c_p(N) \leq \sqrt{T}\hat{\rho}_{ij,T} \leq c_p(N)\right) \\ &= \Pr\left(L_{ij} \leq \frac{\sqrt{T}\left[\hat{\rho}_{ij,T} - E\left(\hat{\rho}_{ij,T}\right)\right]}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}} \leq U_{ij}\right), \end{aligned}$$

where

$$U_{ij}=\frac{c_p(N)-\sqrt{T}E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}},\ L_{ij}=\frac{-c_p(N)-\sqrt{T}E\left(\hat{\rho}_{ij,T}\right)}{\sqrt{Var\left(\sqrt{T}\hat{\rho}_{ij,T}\right)}}. \tag{A.18}$$

Using (8) and (9), we also note that under $\rho_{ij}=0$, and setting $\psi_{ij}=0.5\left[\mu_{ij}(3,1)+\mu_{ij}(1,3)\right]$

$$\begin{aligned} E\left(\hat{\rho}_{ij,T}\left|\rho_{ij}=0\right.\right) &= \frac{-\psi_{ij}}{T}+O\left(T^{-2}\right), \\ Var\left(\hat{\rho}_{ij,T}\left|\rho_{ij}=0\right.\right) &= \frac{\kappa_{ij}}{T}+O\left(T^{-2}\right), \end{aligned}$$

where $\kappa_{ij}=\left[\mu_{ij}(2,2)\left|\rho_{ij}=0\right.\right]$, and

$$\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|\leq c_p(N)\left|\rho_{ij}=0\right.\right)=F_{ij,T}\left[U_{ij,T}(0)\right]-F_{ij,T}\left[L_{ij,T}(0)\right]$$

where

$$U_{ij,T}(0)=\frac{c_p(N)+\frac{\psi_{ij}\left(\rho_{ij}=0\right)}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij}+O\left(T^{-1}\right)}},\ \ L_{ij,T}(0)=\frac{-c_p(N)+\frac{\psi_{ij}\left(\rho_{ij}=0\right)}{\sqrt{T}}+O\left(T^{-3/2}\right)}{\sqrt{\kappa_{ij}+O\left(T^{-1}\right)}}. \tag{A.19}$$

Hence,

$$\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\left|\rho_{ij}=0\right.\right)=1-F_{ij,T}\left[U_{ij,T}(0)\right]+F_{ij,T}\left[L_{ij,T}(0)\right]. \tag{A.20}$$

Setting $a_{ij,T}=U_{ij,T}(0)$ we have that (recall by assumption $\sup_{ij}\left|\psi_{ij}\right|<K$)

$$a_{ij,T}^2=\frac{c_p^2(N)}{\kappa_{ij}}+O\left[\frac{c_p(N)}{\sqrt{T}}\right]+O\left(T^{-1}\right).$$

By Lemma 3, $c_p(N)/\sqrt{T}=o(1)$, as N and $T\to\infty$ (see (A.7)), and hence

$$a_{ij,T}^2=\frac{c_p^2(N)}{\kappa_{ij}}+o(1). \tag{A.21}$$

Therefore, in view of (A.1) established in Lemma 1 and (A.21), we have (for some small positive ϵ)

$$\begin{aligned} F_{ij,T}\left[U_{ij,T}(0)\right] &= \Phi\left[U_{ij,T}(0)\right]+Ke^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1+o(1)\right], \\ F_{ij,T}\left[L_{ij,T}(0)\right] &= \Phi\left[L_{ij,T}(0)\right]+Ke^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1+o(1)\right]. \end{aligned}$$

Substituting the above results in (A.20) yields

$$\begin{aligned} \Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\left|\rho_{ij}=0\right.\right) &= 1-\Phi\left[U_{ij,T}(0)\right]+\Phi\left[L_{ij,T}(0)\right] \\ &\quad +Ke^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1+o(1)\right], \end{aligned}$$

or

$$\begin{aligned} \Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right|>c_p(N)\left|\rho_{ij}=0\right.\right) &= \Phi\left[-U_{ij,T}(0)\right]+\Phi\left[L_{ij,T}(0)\right] \\ &\quad +Ke^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1+o(1)\right]. \end{aligned} \tag{A.22}$$

Since by assumption $\left|\psi_{ij}\right| < K$, and $c_p(N)$ is an increasing function of N then there must exist N_0 and T_0 such that for values of $N > N_0$ and $T > T_0$

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

and

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

and by Lemma 5 we have

$$\begin{aligned}
\Phi\left[-U_{ij,T}(0)\right] &\leq \frac{1}{2}\exp\left\{-\frac{\left[c_p(N) + \frac{\psi_{ij}(\rho_{ij}=0)}{\sqrt{T}} + O\left(T^{-3/2}\right)\right]^2}{2\left[\kappa_{ij} + O\left(T^{-1}\right)\right]}\right\} \\
&= \frac{1}{2}e^{-\frac{1}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1 + O\left(\frac{c_p(N)}{\sqrt{T}}\right) + O\left(T^{-1}\right)\right] \qquad \text{(A.23)} \\
&= \frac{1}{2}e^{-\frac{1}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1 + o(1)\right].
\end{aligned}$$

Similarly,

$$\Phi\left[L_{ij,T}(0)\right] \leq \frac{1}{2}e^{-\frac{1}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1 + o(1)\right]. \qquad \text{(A.24)}$$

Substituting the above results in (A.22) now yields

$$\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right) \leq \left[e^{-\frac{1}{2}\frac{c_p^2(N)}{\mu_{ij}(2,2)}} + Ke^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\mu_{ij}(2,2)}}\right]\left[1 + o(1)\right],$$

or[2]

$$\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij} = 0\right.\right) \leq Ke^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\kappa_{ij}}}\left[1 + o(1)\right],$$

as required.

Consider now the case where $\rho_{ij} \neq 0$ and note that

$$\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| < c_p(N)|\rho_{ij} \neq 0\right) = F_{ij,T}\left[U_{ij,T}(\rho_{ij})\right] - F_{ij,T}\left[L_{ij,T}(\rho_{ij})\right], \qquad \text{(A.25)}$$

where

$$\begin{aligned}
U_{ij,T}(\rho_{ij}) &= \frac{c_p(N) - \sqrt{T}\rho_{ij} - \frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(T^{-1}\right)}}, \qquad \text{(A.26)} \\
L_{ij,T}(\rho_{ij}) &= \frac{-c_p(N) - \sqrt{T}\rho_{ij} - \frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(T^{-1}\right)}}, \qquad \text{(A.27)}
\end{aligned}$$

[2] Note that

$$\frac{e^{-\frac{1}{2}\frac{c_p^2(N)}{\mu_{ij}(2,2)}}}{e^{-\frac{1-\epsilon}{2}\frac{c_p^2(N)}{\mu_{ij}(2,2)}}} = e^{-\frac{\epsilon}{2}\frac{c_p^2(N)}{\mu_{ij}(2,2)}} \to 0, \text{ as } c_p^2(N) \to \infty.$$

$|K_{\rm m}(\boldsymbol{\theta}_{ij})| < K$, and $0 < K_v(\boldsymbol{\theta}_{ij}) < K$. Suppose that $\rho_{ij} > 0$. Then $\sqrt{T}\rho_{ij} + c_p(N) \to \infty$ and $\sqrt{T}\rho_{ij} - c_p(N) \to \infty$, as N and $T \to \infty$ (recall that $c_p(N)/\sqrt{T} \to 0$ with N and $T \to \infty$). Again using (A.26) and (A.27) for $a_{ij,T}$ in (A.1) we have

$$\begin{aligned} F_{ij,T}\left[U_{ij,T}(\rho_{ij})\right] &= \Phi\left[U_{ij,T}(\rho_{ij})\right] + Ke^{\frac{-1}{2}\frac{\left[c_p(N)-\sqrt{T}\rho_{ij}\right]^2}{K_{\rm v}(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right], \\ F_{ij,T}\left[L_{ij,T}(\rho_{ij})\right] &= \Phi\left[L_{ij,T}(\rho_{ij})\right] + Ke^{\frac{-1}{2}\frac{\left[c_p(N)+\sqrt{T}\rho_{ij}\right]^2}{K_{\rm v}(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right]. \end{aligned}$$

Hence

$$\begin{aligned} \Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| < c_p(N)|\rho_{ij} \neq 0\right) &= \Phi\left[U_{ij,T}(\rho_{ij})\right] - \Phi\left[L_{ij,T}(\rho_{ij})\right] \\ &\quad + Ke^{\frac{-1}{2}\frac{\left[c_p(N)-\sqrt{T}\rho_{ij}\right]^2}{K_v(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right] \\ &\quad + Ke^{\frac{-1}{2}\frac{\left[c_p(N)+\sqrt{T}\rho_{ij}\right]^2}{K_v(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right]. \end{aligned}$$

Further, since $\Phi\left[L_{ij,T}(\rho_{ij})\right] \geq 0$, then

$$\Phi\left(\left[U_{ij,T}(\rho_{ij})\right]\right) - \Phi\left(\left[L_{ij,T}(\rho_{ij})\right]\right) \leq \Phi\left(\frac{c_p(N) - \sqrt{T}\rho_{ij} - \frac{K_{\rm m}(\boldsymbol{\theta}_{ij})}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(T^{-1}\right)}}\right).$$

Also, there exists N_0 and T_0 such that for $\rho_{ij} > 0$, and all $N > N_0$ and $T > T_0$, we have (using Lemma 5)

$$\Phi\left(\frac{c_p(N) - \sqrt{T}\rho_{ij} - \frac{K_{\rm m}(\boldsymbol{\theta}_{ij})}{\sqrt{T}} + O\left(T^{-3/2}\right)}{\sqrt{K_v(\boldsymbol{\theta}_{ij}) + O\left(T^{-1}\right)}}\right) \leq \frac{1}{2}e^{\frac{-1}{2}\frac{\left[c_p(N)-\sqrt{T}\rho_{ij}\right]^2}{K_v(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right],$$

and hence

$$\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| < c_p(N)|\rho_{ij} > 0\right) \leq Ke^{\frac{-1}{2}\frac{\left[c_p(N)-\sqrt{T}\rho_{ij}\right]^2}{K_v(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right].$$

A similar result can also be obtained for $\rho_{ij} < 0$, yielding the overall result

$$\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| < c_p(N)|\rho_{ij} \neq 0\right) \leq Ke^{\frac{-1}{2}\frac{T\left[\left|\rho_{ij}\right| - \frac{c_p(N)}{\sqrt{T}}\right]^2}{K_v(\boldsymbol{\theta}_{ij})}}\left[1+o(1)\right].$$

■

Proof of Lemma 7. We first note that

$$\begin{aligned} E\left(\frac{1}{\chi^2_{\rm v,t}}\right) &= \frac{1}{{\rm v}-2}, Var\left(\frac{1}{\chi^2_{\rm v,t}}\right) = \frac{2}{({\rm v}-2)^2({\rm v}-4)} \\ E\left(\frac{1}{\chi^2_{\rm v,t}}\right)^2 &= \frac{2}{({\rm v}-2)^2({\rm v}-4)} + \left(\frac{1}{{\rm v}-2}\right)^2 = \frac{{\rm v}-2}{({\rm v}-2)^2({\rm v}-4)}. \end{aligned} \tag{A.28}$$

Then

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = E\left[\left(\frac{{\rm v}-2}{\chi^2_{\rm v}}\right)\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right] = E\left(\frac{{\rm v}-2}{\chi^2_{\rm v,t}}\right)E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \mathbf{I}_N,$$

and

$$E(\mathbf{y}_t) = 0,\ E\left(\mathbf{y}_t\mathbf{y}_t'\right) = \mathbf{P}\mathbf{P}' = \mathbf{R}.$$

It is clear that y_{it} has mean zero and a unit variance. Denote the i^{th} row of $\mathbf{P}$ by $\mathbf{p}_i'$ and note that $y_{it} = \mathbf{p}_i'\mathbf{u}_t = \left(\frac{\text{v}-2}{\chi^2_{\text{v},t}}\right)^{1/2}\mathbf{p}_i'\boldsymbol{\varepsilon}_t$, and hence

$$\mu_{ij}(2,2) = E(y_{it}^2 y_{jt}^2) = E\left[\left(\frac{\text{v}-2}{\chi^2_{\text{v},t}}\right)^2 \left(\mathbf{p}_i'\boldsymbol{\varepsilon}_t\right)^2 \left(\mathbf{p}_j'\boldsymbol{\varepsilon}_t\right)^2\right],$$

and since $\boldsymbol{\varepsilon}_t$ and $\chi^2_{\text{v},t}$ are distributed independently using (A.28) we have

$$\mu_{ij}(2,2) = \frac{(\text{v}-2)^3}{(\text{v}-2)^2(\text{v}-4)} E\left[\left(\boldsymbol{\varepsilon}_t'\mathbf{A}_i\boldsymbol{\varepsilon}_t\right)\left(\boldsymbol{\varepsilon}_t'\mathbf{A}_j\boldsymbol{\varepsilon}_t\right)\right],$$

where $\mathbf{A}_i = \mathbf{p}_i\mathbf{p}_i'$. But since $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{I}_N)$, using results in Magnus (1978) we have

$$\begin{aligned} E\left[\left(\boldsymbol{\varepsilon}_t'\mathbf{A}_i\boldsymbol{\varepsilon}_t\right)\left(\boldsymbol{\varepsilon}_t'\mathbf{A}_j\boldsymbol{\varepsilon}_t\right)\right] &= tr\left(\mathbf{p}_i\mathbf{p}_i'\right) tr\left(\mathbf{p}_j\mathbf{p}_j'\right) + tr\left(\mathbf{p}_i\mathbf{p}_i'\mathbf{p}_j\mathbf{p}_j'\right) \\ &= \left(\mathbf{p}_i'\mathbf{p}_i\right)^2 + \left(\mathbf{p}_i'\mathbf{p}_j\right)^2. \end{aligned}$$

Hence

$$\mu_{ij}(2,2) = \frac{(\text{v}-2)\left[\left(\mathbf{p}_i'\mathbf{p}_i\right)^2 + \left(\mathbf{p}_i'\mathbf{p}_j\right)^2\right]}{(\text{v}-4)}.$$

When $\mathbf{P}$ is an identity matrix then $\mathbf{p}_i'\mathbf{p}_i = 1$ and $\mathbf{p}_i'\mathbf{p}_j = 0$, and hence $\mu_{ij}(2,2) = (\text{v}-2)/(\text{v}-4)$. Also

$$E(y_{it}^2 y_{jt}) = E\left[\left(\frac{\text{v}-2}{\chi^2_{\text{v},t}}\right)^{3/2}\right] E\left[\left(\boldsymbol{\varepsilon}_t'\mathbf{A}_i\boldsymbol{\varepsilon}_t\right)\mathbf{p}_j'\boldsymbol{\varepsilon}_t\right] = 0.$$

■

Proof of Lemma 8. Consider the data generating process $\mathbf{y}_t = \mathbf{P}\mathbf{u}_t$ where the elements of $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$, u_{it}, are generated as a standardized independent chi-squared distribution with v_i degrees of freedom, namely

$$u_{it} = \frac{\chi^2_{it}(\text{v}_i) - \text{v}_i}{\sqrt{2\text{v}_i}}, \text{ for all } i \text{ and } t.$$

Then it is clear that $E(u_{it}) = 0$, $E(u_{it}^2) = 1$, and also $E(u_{it}^2 u_{jt}^2) = E(u_{it}^2)E(u_{jt}^2) = 1$, and $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{I}_N$. Let $\mathbf{p}_i'$ be the i^{th} row of $\mathbf{P}$ and note that

$$\begin{aligned} E\left(y_{it}y_{jt}\right) &= \mathbf{p}_i'E\left(\mathbf{u}_t\mathbf{u}_t'\right)\mathbf{p}_j = \mathbf{p}_i'\mathbf{p}_j = \rho_{ij} \\ \mathbf{p}_i'\mathbf{p}_i &= \sum_{r=1}^{N} p_{ir}^2 = 1. \end{aligned}$$

Also

$$\begin{aligned} E\left(y_{it}^2 y_{jt}^2\right) &= E\left[\left(\mathbf{p}_i'\mathbf{u}_t\mathbf{u}_t'\mathbf{p}_i\right)\left(\mathbf{p}_j'\mathbf{u}_t\mathbf{u}_t'\mathbf{p}_j\right)\right] \\ &= \sum_r\sum_{r'}\sum_s\sum_{s'} p_{ir}p_{ir'}p_{js}p_{js'}E(u_{rt}u_{r't}u_{st}u_{s't}). \end{aligned}$$

But

$$\begin{aligned} E(u_{rt}u_{r't}u_{st}u_{s't}) &= 0 \text{ if } r \neq r' \text{ or } s \neq s' \\ &= E(u_{rt}^2 u_{st}^2) = 1 \text{ if } r = r' \text{ and } s = s', \end{aligned}$$

and hence

$$E\left(y_{it}^2 y_{jt}^2\right) = \sum_r \sum_s p_{ir}^2 p_{js}^2 = \left(\sum_{r=1}^{N} p_{ir}^2\right)^2 = 1.$$

Therefore, fat-tailed shocks do not necessarily generate $\mu_{ij}(2,2) > 1$. ■

Appendix B An overview of key regularisation techniques

Here we provide an overview of three main covariance estimators proposed in the literature which we use in our Monte Carlo experiments for comparative analysis, namely the thresholding methods of Bickel and Levina (2008b), and Cai and Liu (2011), and the shrinkage approach of Ledoit and Wolf (2004).

B.1 Bickel-Levina (BL) thresholding

The method developed by Bickel and Levina (2008b, BL) employs 'universal' thresholding of the sample covariance matrix $\hat{\mathbf{\Sigma}} = (\hat{\sigma}_{ij})$, $i, j = 1, 2, ..., N$. Under this approach $\mathbf{\Sigma}$ is required to be sparse as they define on p. 2580. The BL thresholding estimator is given by

$$\widetilde{\mathbf{\Sigma}}_{BL,C} = \left(\hat{\sigma}_{ij} I\left[|\hat{\sigma}_{ij}| \geq C\sqrt{\frac{\log(N)}{T}}\right]\right), \ i = 1, 2, ..., N-1, \ j = i+1, i+2, ..., N \quad \text{(B.29)}$$

where $I(.)$ is an indicator function and C is a positive constant which is unknown. The choice of thresholding function - $I(.)$ - implies that (B.29) implements 'hard' thresholding. The consistency rate of the BL estimator is $m_N\sqrt{\frac{\log(N)}{T}}$ under the spectral norm of the error matrix $\left(\widetilde{\mathbf{\Sigma}}_{BL,C} - \mathbf{\Sigma}\right)$. The potential computational burden in the implementation of this approach is the estimation of the thresholding parameter, C. This is usually calibrated by a separate cross-validation (CV) procedure. The quality of the performance of the BL estimator is therefore rooted in the specification chosen for the implementation of CV.[3] Further, cross-validation performs well only when $\mathbf{\Sigma}$ is assumed to be stable over time. Details of the BL cross-validation procedure are given in Section B.3.

As argued by BL, thresholding maintains the symmetry of $\hat{\mathbf{\Sigma}}$ but does not ensure positive definiteness of $\widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}$ in finite samples. BL show that their threshold estimator is positive definite if

$$\left\|\widetilde{\mathbf{\Sigma}}_{BL,C} - \widetilde{\mathbf{\Sigma}}_{BL,0}\right\|_{spec} \leq \epsilon \text{ and } \lambda_{\min}(\mathbf{\Sigma}) > \epsilon, \quad \text{(B.30)}$$

where $\|.\|_{spec}$ is the spectral or operator norm and ϵ is a small positive constant. This condition is not met unless T is sufficiently large relative to N. 'Universal' thresholding on $\hat{\mathbf{\Sigma}}$ performs best when the units x_{it}, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$ are assumed homoscedastic (i.e. $\sigma_{11} = \sigma_{22} = ... = \sigma_{NN}$).

B.2 Cai and Liu (CL) thresholding

Cai and Liu (2011, CL) proposed an improved version of the BL approach by incorporating the unit specific variances in their 'adaptive' thresholding procedure. In this way, unlike 'universal' thresholding on $\hat{\mathbf{\Sigma}}$, their estimator is robust to heteroscedasticity. Specifically, the thresholding estimator $\widetilde{\mathbf{\Sigma}}_{CL,C}$ is defined as

$$\widetilde{\mathbf{\Sigma}}_{CL,C} = \left(\hat{\sigma}_{ij} s_{\tau_{ij}}\left[|\hat{\sigma}_{ij}| \geq \tau_{ij}\right]\right), \ i = 1, 2, ..., N-1, \ j = i+1, i+2, ..., N \quad \text{(B.31)}$$

[3] Fang, Wang and Feng (2013) provide useful guidelines regarding the specification of various parameters used in cross-validation through an extensive simulation study.

where $\tau_{ij} > 0$ is an entry-dependent adaptive threshold such that $\tau_{ij} = \sqrt{\hat{\theta}_{ij}}\omega_T$,with $\hat{\theta}_{ij} = T^{-1}\sum_{i=1}^{T}(x_{it}x_{jt} - \hat{\sigma}_{ij})^2$ and $\omega_T = C\sqrt{\log(N)/T}$, for some constant $C > 0$. CL implement their approach using the general thresholding function $s_\tau(.)$ rather than $I(.)$, but point out that all their theoretical results continue to hold for the hard thresholding estimator. The consistency rate of the CL estimator is $C_0 m_N \sqrt{\log(N)/T}$ under the spectral norm of the error matrix $\left(\tilde{\mathbf{\Sigma}}_{CL,C} - \mathbf{\Sigma}\right)$. The parameter C can be fixed to a constant implied by theory ($C = 2$ in CL) or chosen via cross-validation. Details of the CL cross-validation procedure are provided in Section B.3.

As with the BL estimator, thresholding in itself does not ensure positive definiteness of $\tilde{\mathbf{\Sigma}}_{CL,\hat{C}}$. In light of condition (B.30), Fan, Liao and Mincheva (FLM) (2013) extend the CL approach and propose setting a lower bound on the cross-validation grid when searching for C such that the minimum eigenvalue of their threshold estimator is positive, $\lambda_{\min}\left(\tilde{\mathbf{\Sigma}}_{FLM,\hat{C}}\right) > 0$. This idea originated from Fryzlewicz (2013). Further details of this procedure can be found in Section B.3. We apply this extension to both BL and CL procedures (see Section B.3 for the relevant expressions).

B.3 Cross-validation for BL and CL

We perform a grid search for the choice of C over a specified range: $C = \{c : C_{\min} \leq c \leq C_{\max}\}$. In the BL procedure, we set $C_{\min} = \left|\min_{ij}\hat{\sigma}_{ij}\right|\sqrt{\frac{T}{\log N}}$ and $C_{\max} = \left|\max_{ij}\hat{\sigma}_{ij}\right|\sqrt{\frac{T}{\log N}}$ and impose increments of $\frac{(C_{\max}-C_{\min})}{N}$. In CL cross-validation, we set $C_{\min} = 0$ and $C_{\max} = 4$, and impose increments of c/N. In each point of this range, c, we use x_{it}, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$ and select the $N \times 1$ column vectors $\boldsymbol{x}_t = (x_{1t}, x_{2t}, ..., x_{Nt})'$, $t = 1, 2, ..., T$ which we randomly reshuffle over the t-dimension. This gives rise to a new set of $N \times 1$ column vectors $\boldsymbol{x}_t^{(s)} = \left(x_{1t}^{(s)}, x_{2t}^{(s)}, ..., x_{Nt}^{(s)}\right)'$ for the first shuffle $s = 1$. We repeat this reshuffling S times in total where we set $S = 50$. We consider this to be sufficiently large (FLM suggested $S = 20$ while BL recommended $S = 100$ - see also Fang, Wang and Feng (2013)). In each shuffle $s = 1, 2, ..., S$, we divide $\boldsymbol{x}^{(s)} = \left(\boldsymbol{x}_1^{(s)}, \boldsymbol{x}_2^{(s)}, ..., \boldsymbol{x}_T^{(s)}\right)$ into two subsamples of size $N \times T_1$ and $N \times T_2$, where $T_2 = T - T_1$. A theoretically 'justified' split suggested in BL is given by $T_1 = T\left(1 - \frac{1}{\log(T)}\right)$ and $T_2 = \frac{T}{\log(T)}$. In our simulation study we set $T_1 = \frac{2T}{3}$ and $T_2 = \frac{T}{3}$. Let $\hat{\mathbf{\Sigma}}_1^{(s)} = \left(\hat{\sigma}_{1,ij}^{(s)}\right)$, with elements $\hat{\sigma}_{1,ij}^{(s)} = T_1^{-1}\sum_{t=1}^{T_1} x_{it}^{(s)} x_{jt}^{(s)}$, and $\hat{\mathbf{\Sigma}}_2^{(s)} = \left(\hat{\sigma}_{2,ij}^{(s)}\right)$ with elements $\hat{\sigma}_{2,ij}^{(s)} = T_2^{-1}\sum_{t=T_1+1}^{T} x_{it}^{(s)} x_{jt}^{(s)}$, $i, j = 1, 2, ..., N$, denote the sample covariance matrices generated using T_1 and T_2 respectively, for each split s. We threshold $\hat{\mathbf{\Sigma}}_1^{(s)}$ as in (B.29) or (B.31) using $I(.)$ as the thresholding function, where both $\hat{\theta}_{ij}$ and ω_T are adjusted to

$$\hat{\theta}_{1,ij}^{(s)} = \frac{1}{T_1}\sum_{t=1}^{T_1}(x_{it}^{(s)} x_{jt}^{(s)} - \hat{\sigma}_{1,ij}^{(s)})^2,$$

and

$$\omega_{T_1}(c) = c\sqrt{\frac{\log(N)}{T_1}}.$$

Then (B.31) becomes

$$\tilde{\mathbf{\Sigma}}_1^{(s)}(c) = \left(\hat{\sigma}_{1,ij}^{(s)} I\left[\left|\hat{\sigma}_{1,ij}^{(s)}\right| \geq \tau_{1,ij}^{(s)}(c)\right]\right),$$

for each c, where

$$\tau_{1,ij}^{(s)}(c) = \sqrt{\hat{\theta}_{1,ij}^{(s)}}\omega_{T_1}(c) > 0,$$

and $\hat{\theta}_{1,ij}^{(s)}$ and $\omega_{T_1}(c)$ are defined above.

The following expression is computed for BL or CL,

$$\hat{G}(c) = \frac{1}{S}\sum_{s=1}^{S}\left\|\widetilde{\mathbf{\Sigma}}_1^{(s)}(c) - \widetilde{\mathbf{\Sigma}}_2^{(s)}\right\|_F^2, \tag{B.32}$$

for each c and

$$\hat{C} = \arg\min_{C_{\min}\leq c\leq C_{\max}} \hat{G}(c). \tag{B.33}$$

If several values of c attain the minimum of (B.33), then $\hat{C}$ is chosen to be the smallest one. The final estimator of the covariance matrix is then given by $\widetilde{\mathbf{\Sigma}}_{\hat{C}}$. The thresholding approach does not necessarily ensure that the resultant estimate, $\widetilde{\mathbf{\Sigma}}_{\hat{C}}$, is positive definite. To ensure that the threshold estimator is positive definite FLM (2013) propose setting a lower bound on the cross-validation grid for the search of C such that $\lambda_{\min}\left(\widetilde{\mathbf{\Sigma}}_{\hat{C}}\right) > 0$ - see Fryzlewicz (2013). Therefore, we modify (B.33) so that

$$\hat{C}^* = \arg\min_{C_{pd}+\epsilon\leq c\leq C_{\max}} \hat{G}(c), \tag{B.34}$$

where C_{pd} is the lowest c such that $\lambda_{\min}\left(\widetilde{\mathbf{\Sigma}}_{C_{pd}}\right) > 0$ and ϵ is a small positive constant. We do not conduct thresholding on the diagonal elements of the covariance matrices which remain in tact.

B.4 Ledoit and Wolf (LW) shrinkage

Ledoit and Wolf (2004, LW) considered a shrinkage estimator for regularisation which is based on a linear combination of the sample covariance matrix, $\hat{\mathbf{\Sigma}}$, and an identity matrix $\boldsymbol{I}_N$, and provide formulae for the appropriate weights. The LW shrinkage is expressed as

$$\hat{\mathbf{\Sigma}}_{LW} = \hat{\rho}_1\boldsymbol{I}_N + \hat{\rho}_2\hat{\mathbf{\Sigma}}, \tag{B.35}$$

with the estimated weights given by

$$\hat{\rho}_1 = m_T b_T^2/d_T^2,\ \ \hat{\rho}_2 = a_T^2/d_T^2$$

where

$$\begin{aligned} m_T &= N^{-1}tr\left(\hat{\mathbf{\Sigma}}\right),\ d_T^2 = N^{-1}tr\left(\hat{\mathbf{\Sigma}}^2\right) - m_T^2, \\ a_T^2 &= d_T^2 - b_T^2,\ b_T^2 = \min(\bar{b}_T^2, d_T^2), \end{aligned}$$

and

$$\bar{b}_T^2 = \frac{1}{NT^2}\sum_{t=1}^{T}\left\|\dot{\boldsymbol{x}}_t\dot{\boldsymbol{x}}_t' - \hat{\mathbf{\Sigma}}\right\|_F^2 = \frac{1}{NT^2}\sum_{t=1}^{T} tr\left[\left(\dot{\boldsymbol{x}}_t\dot{\boldsymbol{x}}_t'\right)\left(\dot{\boldsymbol{x}}_t\dot{\boldsymbol{x}}_t'\right)\right] - \frac{2}{NT^2}\sum_{t=1}^{T} tr\left(\dot{\boldsymbol{x}}_t'\hat{\mathbf{\Sigma}}\dot{\boldsymbol{x}}_t\right) + \frac{1}{NT}tr\left(\hat{\mathbf{\Sigma}}^2\right),$$

and noting that $\sum_{t=1}^{T} tr\left(\dot{\boldsymbol{x}}_t'\hat{\mathbf{\Sigma}}\dot{\boldsymbol{x}}_t\right) = \sum_{t=1}^{T} tr\left(\hat{\mathbf{\Sigma}}\sum_{t=1}^{T}\dot{\boldsymbol{x}}_t\dot{\boldsymbol{x}}_t'\right) = T\sum_{t=1}^{T} tr\left(\hat{\mathbf{\Sigma}}^2\right)$, we have

$$\bar{b}_T^2 = \frac{1}{NT^2}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\dot{x}_{it}^2\right)^2 - \frac{1}{NT}tr\left(\hat{\mathbf{\Sigma}}^2\right),$$

with $\dot{\boldsymbol{x}}_t = (\dot{x}_{1t}, \dot{x}_{2t}, ..., \dot{x}_{Nt})'$ and $\dot{x}_{it} = (x_{it} - \bar{x}_i)$.[4]

$\hat{\mathbf{\Sigma}}_{LW}$ is positive definite by construction. Thus, the inverse $\hat{\mathbf{\Sigma}}_{LW}^{-1}$ exists and is well conditioned.

[4]Note that LW scale the Frobenius norm by $1/N$, and use $\|\boldsymbol{A}\|_F^2 = tr(\boldsymbol{A}'\boldsymbol{A})/N$. See Definition 1 of Ledoit and Wolf (2004, p. 376). Here we use the standard notation for this norm.

Appendix C Shrinkage on MT estimator (S-MT)

Recall the shrinkage on the multiple testing estimator ($S\text{-}MT$) expression displayed in Section 3.1,

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi) = \xi \boldsymbol{I}_N + (1-\xi)\widetilde{\boldsymbol{R}}_{MT}, \tag{C.36}$$

where the $N \times N$ identity matrix $\boldsymbol{I}_N$ is set as benchmark target, the shrinkage parameter is denoted by $\xi \in (\xi_0, 1]$, and ξ_0 is the minimum value of ξ that produces a non-singular $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0)$ matrix. Note that shrinkage is 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. Further, 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 $\widetilde{\boldsymbol{R}}_{MT}$. With regard to the calibration of the shrinkage parameter, ξ, we solve the following optimisation problem

$$\xi^* = \arg \min_{\xi_0+\epsilon \leq \xi \leq 1} \left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\xi) \right\|_F^2, \tag{C.37}$$

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 $\xi_0 = \max\left(\frac{\epsilon - \lambda_{\min}(\widetilde{\boldsymbol{R}}_{MT})}{1-\lambda_{\min}(\widetilde{\boldsymbol{R}}_{MT})}, 0\right)$.

C.1 Derivation of S-MT shrinkage parameter

We need to solve $f(\xi) = 0$ for ξ^* such that $f(\xi^*) = 0$ for a given choice of $\mathbf{R}_0$.[5]

Abstracting from the subscripts, note that

$$f(1) = -tr\left[\left(\boldsymbol{R}^{-1} - \boldsymbol{I}_N\right)\left(\boldsymbol{I}_N - \widetilde{\boldsymbol{R}}\right)\right],$$

[5] The code for computing $\mathbf{R}_0$ of our choice is available upon request (see Section C.2).

or

$$\begin{aligned} f(1) &= -tr\left[-\boldsymbol{R}^{-1}\widetilde{\boldsymbol{R}}+\boldsymbol{R}^{-\mathbf{1}}-\boldsymbol{I}_N+\widetilde{\boldsymbol{R}}\right] \\ &= tr\left(\boldsymbol{R}^{-1}\widetilde{\boldsymbol{R}}\right)-tr\left(\boldsymbol{R}^{-\mathbf{1}}\right), \end{aligned}$$

which is generally non-zero. Also, $\xi = 0$ is ruled out, since $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(0) = \widetilde{\boldsymbol{R}}$ need not be non-singular.

Thus we need to assess whether $f(\xi) = 0$ has a solution in the range $\xi_0 < \xi < 1$, where ξ_0 is the minimum value of ξ such that $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0)$ is non-singular. First, we can compute ξ_0 by implementing naive shrinkage as an initial estimate:

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0) = \xi_0\boldsymbol{I}_N + (1-\xi_0)\widetilde{\boldsymbol{R}}.$$

The shrinkage parameter $\xi_0 \in [0, 1]$ is given by

$$\xi_0 = \max\left(\frac{\epsilon - \lambda_{\min}\left(\widetilde{\boldsymbol{R}}\right)}{1-\lambda_{\min}\left(\widetilde{\boldsymbol{R}}\right)}, 0\right),$$

where in our simulation study we set $\epsilon = 0.01$. Here, $\lambda_{\min}(\boldsymbol{A})$ stands for the minimum eigenvalue of matrix $\boldsymbol{A}$. If $\widetilde{\boldsymbol{R}}$ is already positive definite and $\lambda_{\min}\left(\widetilde{\boldsymbol{R}}\right) > 0$, then ξ_0 is automatically set to zero. Conversely, if $\lambda_{\min}\left(\widetilde{\boldsymbol{R}}\right) \leq 0$, then ξ_0 is set to the smallest possible value that ensures positivity of $\lambda_{\min}\left(\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0)\right)$.

Second, we implement the optimisation procedure. In our simulation study we employ a grid search for $\xi^* = \{\xi : \xi_0 + \epsilon \leq \xi \leq 1\}$ with increments of 0.005. The final ξ^* is given by

$$\xi^* = \arg\min_{\xi} \left[f(\xi)\right]^2.$$

C.2 Specification of reference matrix $\mathbf{R}_0$

Implementation of the above procedure requires the use of a suitable reference matrix $\boldsymbol{R}_0$. Our experimentations suggested 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 in what follows. We set as reference matrix $\mathbf{R}_0$ the shrinkage estimator of LW applied to the sample correlation matrix:

$$\hat{\mathbf{R}}_0 = \theta\boldsymbol{I}_N + (1-\theta)\hat{\mathbf{R}},$$

with shrinkage parameter $\theta \in [0, 1]$, and $\hat{\mathbf{R}} = (\hat{\rho}_{ij})$. The optimal value of the shrinkage parameter that minimizes the expectation of the squared Frobenius norm of the error of estimating $\mathbf{R}$ by $\hat{\mathbf{R}}_0$:

$$E\left\|\hat{\mathbf{R}}_0 - \mathbf{R}\right\|_F^2 = \sum_{i\neq j}\sum E\left(\hat{\rho}_{ij}-\rho_{ij}\right)^2 + \theta^2\sum_{i\neq j}\sum E\left(\hat{\rho}_{ij}^2\right) - 2\theta\sum_{i\neq j}\sum E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right], \tag{C.38}$$

is given by

$$\theta^* = \frac{\sum\sum_{i\neq j} E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right]}{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}^2\right)} = 1 - \frac{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}\rho_{ij}\right)}{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}^2\right)}, \tag{C.39}$$

with

$$\hat{\theta}^* = 1 - \frac{\sum\sum_{i\neq j}\hat{\rho}_{ij}\left[\hat{\rho}_{ij} - \frac{\hat{\rho}_{ij}(1-\hat{\rho}_{ij}^2)}{2T}\right]}{\frac{1}{T}\sum\sum_{i\neq j}(1-\hat{\rho}_{ij}^2)^2 + \sum\sum_{i\neq j}\left[\hat{\rho}_{ij} - \frac{\hat{\rho}_{ij}(1-\hat{\rho}_{ij}^2)}{2T}\right]^2}.$$

Note that $\lim_{T\to\infty}(\hat{\theta}^*) = 0$ for any N. However, in small samples values of $\hat{\theta}^*$ can be obtained that fall outside the range $[0,1]$. To avoid such cases, if $\hat{\theta}^* < 0$ then $\hat{\theta}^*$ is set to 0, and if $\hat{\theta}^* > 1$ it is set to 1, or $\hat{\theta}^{**} = \max(0, \min(1, \hat{\theta}^*))$.

Appendix D Additional Monte Carlo simulation results

D.1 Approximately sparse covariance matrix specifications

We present here two additional covariance (correlation) specifications based on approximately sparse matrices. These are considered in the context of the Monte Carlo setup of Section 3.

Monte Carlo design C: We follow Bickel and Levina (2008b) and set $\boldsymbol{R}$ to coincide with the correlation matrix of a first-order autoregressive process with coefficient, ϕ, given by

$$\boldsymbol{R} = \begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{N-1} \\ \phi & 1 & & & \vdots \\ \phi^2 & \phi & \ddots & & \vdots \\ \vdots & \cdots & \cdots & \ddots & \phi \\ \phi^{N-1} & \cdots & \cdots & \phi & 1 \end{pmatrix}.$$

The Cholesky factor, $\boldsymbol{P}$, for this specification is given by

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \phi & \sqrt{1-\phi^2} & \cdots & & 0 \\ \phi^2 & \phi\sqrt{1-\phi^2} & \cdots & & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{N-2} & \phi^{N-3}\sqrt{1-\phi^2} & \cdots & \sqrt{1-\phi^2} & 0 \\ \phi^{N-1} & \phi^{N-2}\sqrt{1-\phi^2} & \cdots & \phi\sqrt{1-\phi^2} & \sqrt{1-\phi^2} \end{pmatrix}.$$

Also, $\sigma_{ii} = 1/\left(1-\phi^2\right)$, $i = 1, 2, ..., N$. In this experiment we set $\phi = 0.7$, and hence we generate $\boldsymbol{x}_t = \left(1-\phi^2\right)^{-1/2}\mathbf{P}\mathbf{u}_t$, with $\mathbf{P}$ given above.

Monte Carlo design D: Under this specification $\boldsymbol{\Sigma}$ $(= \mathbf{D}^{1/2}\boldsymbol{R}\mathbf{D}^{1/2})$ is set to the covariance matrix of a standard first-order spatial autoregressive model (SAR) with coefficient ϑ and weight matrix, $\boldsymbol{W}$,

$$\boldsymbol{\Sigma} = (\sigma_{ij}) = (\boldsymbol{I}_N - \vartheta\boldsymbol{W})^{-1}\boldsymbol{\Lambda}(\boldsymbol{I}_N - \vartheta\boldsymbol{W}')^{-1}, \tag{D.40}$$

where $\boldsymbol{\Lambda} = diag(\lambda_{11}, \lambda_{22}, ..., \lambda_{NN})$, and $\boldsymbol{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})$ with $\sigma_{ii} \sim IID\left(1/2 + \chi^2(2)/4\right)$, $i = 1, 2, ..., N$. The weight matrix $\boldsymbol{W}$ is row-standardised with all units having two neighbours except for the first and last units that have only one neighbour

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ 1/2 & 0 & 1/2 & \cdots & \cdots & 0 & 0 \\ 0 & 1/2 & 0 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}_{N\times N}.$$

This ensures that the largest eigenvalue of $\boldsymbol{W}$ is unity and the degree of cross-sectional dependence is measured by ϑ. The correlation matrix in this case is given by

$$\mathbf{R} = \boldsymbol{D}^{-1/2}(\boldsymbol{I}_N - \vartheta\boldsymbol{W})^{-1}\boldsymbol{\Lambda}(\boldsymbol{I}_N - \vartheta\boldsymbol{W}')^{-1}\boldsymbol{D}^{-1/2},$$

with the associated Cholesky factor, $\boldsymbol{P}$, given by

$$\boldsymbol{P} = \boldsymbol{D}^{-1/2}(\boldsymbol{I}_N - \vartheta \boldsymbol{W})^{-1}\boldsymbol{\Lambda}^{1/2}.$$

To ensure that $Var(x_{it}) = \sigma_{ii}$, we need to set λ_{ii} such that

$$diag\left[(\boldsymbol{I}_N - \vartheta \boldsymbol{W})^{-1}\boldsymbol{\Lambda}(\boldsymbol{I}_N - \vartheta \boldsymbol{W}')^{-1}\right] = \mathbf{D}.$$

Computation of λ_{ii} can be done numerically. Let $d_i(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{22}, ..., \lambda_{NN})'$ be the i^{th} diagonal element of $(\boldsymbol{I}_N - \vartheta \boldsymbol{W})^{-1}\boldsymbol{\Lambda}(\boldsymbol{I}_N - \vartheta \boldsymbol{W}')^{-1}$, then we compute $\boldsymbol{\lambda}$ by solving the following optimisation problem

$$\min_{\boldsymbol{\lambda}} \sum_{i=1}^{N} \left[d_i(\boldsymbol{\lambda}) - \sigma_{ii}\right]^2.$$

The initial vector of $\boldsymbol{\lambda}$ is set to $\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})'$ generated as above.

All results are reported for $N = \{30, 100, 200\}$ and $T = 100$, for the case where $\boldsymbol{\gamma} = \mathbf{0}$ and $\mathbf{a} = \mathbf{0}$ in (19). Results for $\boldsymbol{\gamma} \neq \mathbf{0}$ and $\mathbf{a} \neq \mathbf{0}$ are very similar and are available upon request.

D.2 Additional results

Overall, similar conclusions are drawn when considering approximately sparse matrices in our experiments to those obtained under the exactly sparse Monte Carlo designs of Section 3.

D.2.1 Robustness of MT to the choice of the p-value and $f(N)$

In line with Table 1, Table D1 shows the sensitivity of the MT estimator to different levels of significance, p, and scaling factors $f(N)$ inherent in the theoretical critical value, $c_p(N)$, by way of average spectral and Frobenius norm losses over 2,000 replications for Monte Carlo designs C and D when $p = \{0.01, 0.05, 0.10\}$ and $f(N) = \{N-1, N(N-1)/2\}$, and under both distributional assumptions for the errors (Gaussian and multivariate t). Neither of the norms is affected much by the choice of p under the error specifications considered for all N. With regard to the scaling factor $f(N)$, under normality of the errors, where $\kappa_{\max} = 1$, both norms of MT_{N-1} outperform $MT_{N(N-1)/2}$ for designs C and D, which is expected given Theorem 2. Under non-linear dependence of the errors for Monte Carlo design C, MT_{N-1} still outperforms $MT_{N(N-1)/2}$. However, the difference between the two norms reduces considerably. On the other hand, for Monte Carlo design D, $MT_{N(N-1)/2}$ produces lower norms than MT_{N-1} almost uniformly when the spectral norm is considered, which is in line with the theory of Section 2.1.

D.2.2 Norm comparisons of MT, BL, CL, and LW estimators

Results when comparing our proposed estimators with those suggested in the literature (average norms over 100 replications) from Monte Carlo designs C and D are shown in Tables D2 and D3, respectively. As in Section 3.4, the MT estimators are computed using scaling factor $f(N) = N(N-1)/2$ and $p = 0.05$. In general, for both designs thresholding outperforms shrinkage across N. Since design C considers a correlation matrix, $BL_{\hat{C}}$ performs comparatively well while CL_2 outperforms $CL_{\hat{C}}$ as N increases. Design D analyses heteroskedastic data, hence in this case $BL_{\hat{C}}$ is outperformed by $CL_{\hat{C}}$, especially when looking at the Frobenius norms, whilst $CL_{\hat{C}}$ outperforms CL_2 across N as suggested in Cai and Liu (2011). Overall, $CL_{\hat{C}}$ performs best but the MT method records lower norms at times especially when the errors are non-linearly dependent (t-distributed), as shown in the bottom panel of Tables D2 and D3. Looking at the adjusted thresholding methods, they suffer universally compared to their unadjusted counterparts which is expected. For both designs, $S\text{-}MT_{N(N-1)/2}$ clearly outperforms $BL_{\hat{C}^*}$ and $CL_{\hat{C}^*}$ across all N.

D.2.3 Norm comparisons of inverse estimators

Finally, Tables D4 and D5 present norm results for the inverses of the regularisation methods we consider for designs C and D respectively. In line with Monte Carlo design B, $S\text{-}MT_{N(N-1)/2}$ outperforms $BL_{\hat{C}^*}$ and $CL_{\hat{C}^*}$ irrespective of whether the errors are Gaussian or t-distributed. The adjusted BL and CL methods are both prone to sizeable outliers, especially for smaller N. For design C, $LW_{\hat{\Sigma}}$ performs more favourably than $S\text{-}MT_{N(N-1)/2}$ for $N = \{30, 100\}$ under both Gaussian and non-linearly dependent errors but suffers as N increases to 200. For design D, however, $LW_{\hat{\Sigma}}$ is outperformed by the shrinkage on MT estimator uniformly across N.

Table D1: Average spectral and Frobenius norm losses for the $MT(p)$ estimator using significance levels $p = \{0.01, 0.05, 0.10\}$ and scaling factors $f(N) = \{N-1, N(N-1)/2\}$, for $T = 100$

	Monte Carlo design C					
	$f(N)=N-1$			$f(N)=N(N-1)/2$		
N	$MT_{N-1}(.01)$	$MT_{N-1}(.05)$	$MT_{N-1}(.10)$	$MT_{\frac{N(N-1)}{2}}(.01)$	$MT_{\frac{N(N-1)}{2}}(.05)$	$MT_{\frac{N(N-1)}{2}}(.10)$
	$\mathbf{u}_{it}\sim$ **Gaussian**					
	Spectral norm					
30	3.85(0.58)	3.53(0.56)	3.39(0.55)	4.41(0.59)	4.07(0.59)	3.93(0.58)
100	4.88(0.41)	4.53(0.42)	4.38(0.43)	5.70(0.35)	5.38(0.38)	5.23(0.39)
200	5.31(0.32)	4.97(0.34)	4.82(0.35)	6.18(0.23)	5.91(0.27)	5.78(0.28)
	Frobenius norm					
30	6.83(0.40)	6.30(0.42)	6.09(0.44)	7.73(0.41)	7.19(0.40)	6.96(0.40)
100	4.88(0.41)	4.53(0.42)	4.38(0.43)	5.70(0.35)	5.38(0.38)	5.23(0.39)
200	5.31(0.32)	4.97(0.34)	4.82(0.35)	6.18(0.23)	5.91(0.27)	5.78(0.28)
	$\mathbf{u}_{it}\sim$ **multivariate $t-$distributed with 8 degrees of freedom**					
	Spectral norm					
30	4.21(0.82)	4.01(0.91)	3.94(0.97)	4.64(0.71)	4.38(0.76)	4.27(0.79)
100	5.61(4.35)	5.55(4.61)	5.59(4.75)	6.06(3.83)	5.86(4.05)	5.77(4.14)
200	6.08(2.51)	6.15(3.21)	6.29(3.57)	6.45(1.30)	6.30(1.63)	6.23(1.80)
	Frobenius norm					
30	7.40(0.80)	7.02(0.93)	6.90(0.99)	8.15(0.66)	7.69(0.74)	7.50(0.78)
100	15.20(4.25)	14.74(4.54)	14.71(4.68)	17.04(3.72)	16.24(3.93)	15.90(4.02)
200	22.12(2.59)	21.65(3.40)	21.76(3.83)	25.09(1.26)	23.99(1.59)	23.52(1.78)
	Monte Carlo design D					
	$\mathbf{u}_{it}\sim$ **Gaussian**					
	Spectral norm					
30	0.86(0.15)	0.78(0.15)	0.76(0.14)	1.02(0.13)	0.93(0.14)	0.89(0.15)
100	1.06(0.13)	0.97(0.14)	0.95(0.14)	1.21(0.09)	1.16(0.10)	1.14(0.11)
200	1.35(0.14)	1.25(0.15)	1.21(0.15)	1.54(0.10)	1.50(0.11)	1.47(0.12)
	Frobenius norm					
30	1.95(0.20)	1.73(0.18)	1.69(0.18)	2.46(0.19)	2.15(0.20)	2.02(0.20)
100	3.95(0.19)	3.45(0.20)	3.31(0.20)	5.08(0.13)	4.68(0.16)	4.48(0.17)
200	6.30(0.20)	5.54(0.22)	5.28(0.22)	8.00(0.11)	7.57(0.14)	7.33(0.16)
	$\mathbf{u}_{it}\sim$ **multivariate $t-$distributed with 8 degrees of freedom**					
	Spectral norm					
30	1.05(0.37)	1.04(0.43)	1.06(0.46)	1.13(0.29)	1.08(0.34)	1.06(0.36)
100	1.37(1.00)	1.46(1.16)	1.54(1.24)	1.35(0.71)	1.35(0.82)	1.35(0.87)
200	1.81(1.67)	1.97(2.01)	2.10(2.17)	1.72(1.01)	1.73(1.20)	1.74(1.29)
	Frobenius norm					
30	2.26(0.40)	2.16(0.46)	2.18(0.49)	2.61(0.30)	2.39(0.35)	2.30(0.38)
100	4.50(1.02)	4.41(1.24)	4.51(1.35)	5.23(0.66)	4.94(0.79)	4.80(0.84)
200	7.15(1.78)	7.10(2.24)	7.30(2.46)	8.19(0.94)	7.86(1.17)	7.69(1.29)

Note: Norm losses are averages over 2,000 replications. Simulation standard deviations are given in the parentheses. MT estimators are defined in Section 3.2.

Table D2: Spectral and Frobenius norm losses for different regularised covariance matrix estimators ($T = 100$) - Monte Carlo design C

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
			$\mathbf{u}_{it}\sim$ Gaussian			
	Error matrices ($\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}}$)					
$MT_{N(N-1)/2}$	4.10(0.65)	7.25(0.42)	5.34(0.37)	15.23(0.42)	5.93(0.29)	23.05(0.40)
$BL_{\hat{C}}$	3.32(0.73)	5.83(0.63)	4.34(0.49)	12.46(0.57)	4.96(0.50)	18.71(0.55)
CL_2	4.14(0.65)	7.36(0.46)	5.66(0.37)	16.14(0.42)	4.59(0.31)	18.36(0.50)
$CL_{\hat{C}}$	3.23(0.73)	5.77(0.59)	4.12(0.44)	12.20(0.51)	6.34(0.40)	24.78(0.49)
$S\text{-}MT_{N(N-1)/2}$	5.54(0.50)	8.23(0.59)	6.86(0.24)	17.58(0.51)	7.39(0.18)	26.81(0.48)
$BL_{\hat{C}^*}$	8.53(0.10)	14.44(0.07)	9.11(0.06)	27.05(0.04)	9.19(0.05)	38.44(0.04)
$CL_{\hat{C}^*}$	8.43(0.16)	14.28(0.21)	9.10(0.07)	27.00(0.11)	9.18(0.05)	38.42(0.08)
$LW_{\hat{\Sigma}}$	3.37(0.57)	5.68(0.49)	6.00(0.36)	16.05(0.40)	7.54(0.22)	27.57(0.31)
			$\mathbf{u}_{it}\sim$ multivariate $t-$ distributed with 8 degrees of freedom			
	Error matrices ($\boldsymbol{\Sigma} - \mathring{\boldsymbol{\Sigma}}$)					
$MT_{N(N-1)/2}$	4.47(0.99)	7.75(0.95)	5.55(0.59)	15.94(0.71)	6.31(1.11)	24.07(1.37)
$BL_{\hat{C}}$	4.26(1.44)	7.11(1.52)	5.78(1.15)	15.76(2.54)	6.86(1.34)	25.46(5.29)
CL_2	5.11(0.71)	8.94(0.94)	6.98(0.43)	19.90(1.14)	7.64(0.33)	30.34(1.55)
$CL_{\hat{C}}$	3.80(1.19)	6.72(1.20)	4.83(0.69)	14.40(1.65)	5.51(0.80)	22.03(3.04)
$S\text{-}MT_{N(N-1)/2}$	5.59(0.55)	8.41(0.61)	6.85(0.38)	17.69(0.65)	7.38(0.31)	26.74(0.80)
$BL_{\hat{C}^*}$	8.53(0.18)	14.51(0.13)	9.12(0.15)	27.14(0.11)	9.20(0.15)	38.60(0.18)
$CL_{\hat{C}^*}$	8.46(0.22)	14.40(0.21)	9.11(0.16)	27.11(0.14)	9.19(0.15)	38.57(0.19)
$LW_{\hat{\Sigma}}$	4.03(0.84)	6.64(0.81)	6.72(0.63)	17.95(0.75)	8.25(1.13)	29.97(0.92)

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

Table D3: Spectral and Frobenius norm losses for different regularised covariance matrix estimators ($T = 100$) - Monte Carlo design D

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
			$\mathbf{u}_{it}\sim$ Gaussian			
	Error matrices $(\mathbf{\Sigma} - \mathring{\hat{\mathbf{\Sigma}}})$					
$MT_{N(N-1)/2}$	0.93(0.13)	2.16(0.18)	1.16(0.09)	4.68(0.16)	1.50(0.12)	7.55(0.14)
$BL_{\hat{C}}$	0.91(0.16)	2.05(0.22)	1.20(0.14)	4.54(0.42)	1.46(0.16)	7.53(0.70)
CL_2	0.95(0.13)	2.22(0.19)	1.17(0.09)	4.89(0.15)	1.53(0.10)	7.82(0.12)
$CL_{\hat{C}}$	0.77(0.12)	1.76(0.19)	0.98(0.13)	3.50(0.18)	1.26(0.15)	5.58(0.26)
$S\text{-}MT_{N(N-1)/2}$	0.98(0.12)	2.24(0.17)	1.20(0.09)	4.72(0.16)	1.51(0.12)	7.49(0.14)
$BL_{\hat{C}*}$	0.92(0.14)	2.12(0.27)	1.21(0.15)	4.93(0.57)	1.50(0.15)	7.87(0.65)
$CL_{\hat{C}*}$	0.78(0.15)	1.82(0.33)	1.01(0.14)	3.84(0.63)	1.36(0.17)	6.36(0.93)
$LW_{\hat{\Sigma}}$	1.09(0.11)	2.36(0.10)	1.72(0.12)	5.43(0.07)	1.90(0.05)	8.85(0.04)
	$\mathbf{u}_{it}\sim$ multivariate $t-$ distributed with 8 degrees of freedom					
	Error matrices $(\mathbf{\Sigma} - \mathring{\hat{\mathbf{\Sigma}}})$					
$MT_{N(N-1)/2}$	1.03(0.16)	2.34(0.20)	1.30(0.35)	4.88(0.36)	1.93(2.35)	8.03(2.26)
$BL_{\hat{C}}$	1.16(0.18)	2.78(0.48)	1.50(0.21)	5.88(0.23)	1.68(0.25)	8.67(0.29)
CL_2	1.13(0.12)	2.76(0.20)	1.31(0.15)	5.52(0.19)	1.63(0.14)	8.49(0.26)
$CL_{\hat{C}}$	1.00(0.20)	2.21(0.34)	1.32(0.25)	5.03(0.88)	1.58(0.19)	8.08(0.89)
$S\text{-}MT_{N(N-1)/2}$	1.03(0.13)	2.33(0.17)	1.26(0.19)	4.79(0.23)	1.64(0.59)	7.62(0.50)
$BL_{\hat{C}*}$	1.15(0.16)	2.87(0.50)	1.47(0.18)	5.84(0.29)	1.64(0.14)	8.69(0.25)
$CL_{\hat{C}*}$	1.00(0.18)	2.34(0.49)	1.36(0.22)	5.33(0.74)	1.63(0.15)	8.49(0.54)
$LW_{\hat{\Sigma}}$	1.23(0.14)	2.65(0.13)	1.86(0.14)	5.78(0.14)	2.01(0.19)	9.23(0.16)

See the note to Table D2.

Table D4: Spectral and Frobenius norm losses for the inverses of different regularised covariance matrix estimators for Monte Carlo design C - $T = 100$

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices $(\mathbf{\Sigma^{-1}} - \mathring{\mathbf{\Sigma}}^{-1})$					
			$\mathbf{u}_{it}\sim$ **Gaussian**			
$S\text{-}MT_{N(N-1)/2}$	4.03(0.31)	5.19(0.25)	4.75(0.19)	10.00(0.21)	4.97(0.18)	14.62(0.20)
$BL_{\hat{C}^*}$	5.65(0.15)	7.37(0.16)	5.83(0.10)	13.75(0.10)	5.89(0.09)	19.50(0.11)
$CL_{\hat{C}^*}$	3.4x10^4(1.7x10^5)	28.62(173.93)	31.47(255.19)	14.07(3.85)	5.89(0.09)	19.46(0.14)
$LW_{\hat{\Sigma}}$	1.91(0.18)	3.49(0.12)	3.51(0.10)	9.45(0.16)	4.28(0.07)	15.75(0.15)
			$\mathbf{u}_{it}\sim$ **multivariate** $t-$ **distributed with 8 degrees of freedom**			
$S\text{-}MT_{N(N-1)/2}$	3.95(0.48)	5.21(0.33)	4.62(0.30)	9.83(0.50)	4.88(0.29)	14.23(0.77)
$BL_{\hat{C}^*}$	5.67(0.23)	7.37(0.19)	5.84(0.20)	13.69(0.28)	5.95(0.20)	19.45(0.38)
$CL_{\hat{C}^*}$	53.32(262.27)	8.37(5.52)	7.31(10.30)	13.75(0.54)	7.53(5.1x10^7)	19.47(2.4x10^3)
$LW_{\hat{\Sigma}}$	2.42(0.49)	4.03(0.53)	3.90(0.33)	10.39(0.65)	4.58(0.28)	16.70(0.74)

Note: $\mathring{\mathbf{\Sigma}}^{-1} = \{\tilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_{N(N-1)/2}}, \tilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \tilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. See also the notes to Table D2.

Table D5: Spectral and Frobenius norm losses for the inverses of different regularised covariance matrix estimators for Monte Carlo design D - $T = 100$

	$N = 30$		$N = 100$		$N = 200$	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices $(\mathbf{\Sigma^{-1}} - \mathring{\mathbf{\Sigma}}^{-1})$					
			$\mathbf{u}_{it}\sim$ **Gaussian**			
$S\text{-}MT_{N(N-1)/2}$	3.49(0.70)	4.39(0.34)	4.78(0.46)	9.32(0.29)	5.82(0.45)	13.93(0.23)
$BL_{\hat{C}^*}$	6.2x10^3(4.3x10^4)	32.11(72.33)	2.9x10^4(1.0x10^4)	33.02(46.10)	9.3x10^3(8.8x10^4)	31.84(92.70)
$CL_{\hat{C}^*}$	1.3x10^6(1.3x10^7)	152.75(1.1x10^4)	1.3x10^5(3.4x10^6)	116.64(348.34)	5.8x10^5(4.1x10^6)	197.02(735.94)
$LW_{\hat{\Sigma}}$	4.56(0.43)	4.94(0.16)	6.20(0.19)	11.14(0.15)	8.65(0.13)	17.22(0.13)
			$\mathbf{u}_{it}\sim$ **multivariate** $t-$ **distributed with 8 degrees of freedom**			
$S\text{-}MT_{N(N-1)/2}$	3.59(0.94)	4.38(0.41)	4.62(0.64)	8.99(0.51)	5.85(0.84)	13.50(0.69)
$BL_{\hat{C}^*}$	3.3x10^3(1.7x10^4)	24.83(53.16)	2.4x10^3(2.3x10^4)	17.26(46.75)	13.65(63.27)	16.09(1.63)
$CL_{\hat{C}^*}$	979.79(3.3x10^3)	22.62(23.69)	3.4x10^3(2.9x10^4)	23.80(55.00)	412.43(2.2x10^3)	19.87(17.46)
$LW_{\hat{\Sigma}}$	3.66(0.86)	4.62(0.45)	9.26(0.62)	11.94(0.58)	8.99(0.60)	17.63(0.70)

Note: $\mathring{\mathbf{\Sigma}}^{-1} = \{\tilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_{N(N-1)/2}}, \tilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \tilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}\}$. See also the notes to Table D2.

References

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

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

Chiani, M., Dardari, D. and Simon, M. K. (2003). New exponential bounds and approximations for the computation of error probability in fading channels. IEEE Transactions on Wireless Communications, 2, no. 4, 840–845.

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.

Fang, Y., Wang, B. and Feng, Y. (2013). Tuning parameter selection in regularised estimations of large covariance matrices. Available at: http://arxiv.org/abs/1308.3416.

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

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

Magnus, J R (1978). The moments of products of quadratic forms in Normal variables. Statist. Nerlandica 32, 201-210.

Rao, K. S. and Kendall, D. G. (1950) On the generalised second limit theorem in the calculus of probabilities. Biometrika 37, 224-230.

Queen Mary
University of London

This working paper has been produced by
the School of Economics and Finance at
Queen Mary University of London

School of Economics and Finance
Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/