The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers
www.cesifo.org/wp

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

Natalia Bailey
M. Hashem Pesaran
L. Vanessa Smith

CESIFO WORKING PAPER NO. 4834
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
JUNE 2014

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 4834*

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

Abstract

This paper proposes a novel regularisation method for the estimation of large covariance matrices, which makes use of insights from the multiple testing literature. The method tests the statistical significance of individual pair-wise correlations and sets to zero those elements that are not statistically significant, taking account of the multiple testing nature of the problem. The procedure is straightforward to implement, and does not require cross validation. By using the inverse of the normal distribution at a predetermined significance level, it circumvents the challenge of evaluating the theoretical constant arising in the rate of convergence of existing thresholding estimators. We compare the performance of our multiple testing (*MT*) estimator to a number of thresholding and shrinkage estimators in the literature in a detailed Monte Carlo simulation study. Results show that our *MT* estimator performs well in a number of different settings and tends to outperform other estimators, particularly when the cross-sectional dimension, *N*, is larger than the time series dimension, *T*: If the inverse covariance matrix is of interest then we recommend a shrinkage version of the *MT* estimator that ensures positive definiteness.

JEL-Code: C130, C580.

Keywords: sparse correlation matrices, high-dimensional data, multiple testing, thresholding, shrinkage.

Natalia Bailey
Queen Mary, University of London
n.bailey@qmul.ac.uk

M. Hashem Pesaran
University of Southern California
pesaran@usc.edu

L. Vanessa Smith
University of York
vanessa.smith@york.ac.uk

May 2014
The authors would like to thank Alex Chudik for valuable comments and suggestions, as well as Jianqing Fan, 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

Robust estimation of large covariance matrices is a problem that features prominently in a number of areas of multivariate statistical analysis (Anderson (2003)). In finance it arises in portfolio selection and optimisation (Ledoit and Wolf (2003)), risk management (Fan et al. (2008)) and testing of capital asset pricing models (Sentana (2009); Pesaran and Yamagata (2012)) when the number of assets is large. In global macro-econometric modelling with many domestic and foreign channels of interaction, large error covariance matrices must be estimated for impulse response analysis and bootstrapping (Pesaran et al. (2004); Dees et al. (2007)). In the area of bio-informatics, high-dimensional covariance matrices are required when inferring large-scale gene association networks (Carroll (2003); Schäfer and Strimmer (2005)). Large covariance matrices are further encountered in fields including meteorology, climate research, spectroscopy, signal processing and pattern recognition.

Assuming that the $N \times N$ dimensional population covariance matrix, $\mathbf{\Sigma}$, is invertible, one way of obtaining a suitable estimator is to appropriately restrict the off-diagonal elements of its sample equivalence denoted by $\hat{\mathbf{\Sigma}}$. Numerous methods have been developed to address this challenge, predominantly in the statistics literature. 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 rely on a natural ordering among variables and are thus better suited to the analysis of certain types of data. Two popular approaches in the literature that do not make use of any ordering assumptions include those of shrinkage and thresholding. See also Pourahmadi (2011) for an extensive review of general linear models (GLS) and regularisation based methods for estimation of the covariance matrix.

The idea of shrinkage dates back to the seminal work of Stein (1956) who proposed the shrinkage approach in the context of regression models so as to minimize the mean square error of the regression coefficients. The method intentionally introduces a bias in the estimates with the aim of reducing the variance. In the context of covariance matrix estimation the estimated covariances are shrunk towards zero element-wise. More formally, the shrinkage estimator is defined as a weighted average of the sample covariance matrix and an invertible covariance matrix estimator known as the shrinkage target. A number of shrinkage targets have been considered in the literature that take advantage of *a priori* knowledge of the data characteristics under investigation. For example, Ledoit and Wolf (2003) in a study of stock market returns consider Sharpe (1963) and Fama and French (1997) market based covariance matrix specifications as targets.[1] Ledoit and Wolf (2004, LW) suggest a modified shrinkage estimator that involves a convex 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. Numerous other estimators based on the same concept but using different shrinkage targets are proposed in the literature such as by Haff (1980, 1991), Lin and Perlam (1985), Dey and Srinivasan (1985), and Donoho et al. (1995). On the whole, shrinkage estimators are considered to be stable, robust and produce positive definite covariance matrices by construction.

Thresholding is an alternative regularisation technique that involves setting off-diagonal elements of the sample covariance matrix that are in absolute terms below a certain 'threshold' value(s), 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

[1] Other shrinkage targets include the 'diagonal common variance', the 'common covariance', the 'diagonal unequal variance', the 'perfect positive correlation' and the 'constant correlation' target. Examples of structured covariance matrix targets can be found in Daniels and Kass (1999, 2001), Fan et al. (2008) and Hoff (2009), among others.

vary across the different off-diagonal elements of the matrix. Furthermore, the selected non-zero elements of $\hat{\mathbf{\Sigma}}$ can either be set at their sample estimates or can be somewhat adjusted downward. This relates to the concepts of 'hard' and 'soft' thresholding, respectively. The thresholding approach traditionally assumes that the underlying (true) covariance matrix is *sparse*, where sparseness is loosely defined as the presence of a sufficient number of zeros on each row of $\mathbf{\Sigma}$ such that it is absolute summable row (column)-wise. However**,** Fan, Liao and Mincheva (2011, 2013) show that such regularization techniques can be applied to $\hat{\mathbf{\Sigma}}$ even if the underlying population covariance matrix is not sparse, so long as the non-sparseness is characterised by an approximate factor structure.[2] The thresholding method retains symmetry of the sample covariance matrix but does not necessarily deliver a positive definite estimate of $\mathbf{\Sigma}$ if N is large relative to T. The main difficulty in applying this approach lies in the estimation of the thresholding parameter. The method of cross-validation is primarily used for this purpose which is rather convoluted, computationally intensive and not appropriate for all applications. Indeed, cross-validation assumes stability of the underlying covariance matrix over time which may not be the case in many applications in economics and finance.[3]

In this paper, we propose an alternative thresholding procedure using a multiple testing (MT) estimator which is simple and practical to implement. As suggested by its name, it makes use of insights from the multiple testing literature to test the statistical significance of all pair-wise covariances or correlations, and is invariant to the ordering of the underlying variables. It sets the elements associated with the statistically insignificant correlations to zero, and retains the significant ones. We apply the multiple testing procedure to the sample correlation matrix denoted by $\hat{\boldsymbol{R}}$**,** rather than $\hat{\mathbf{\Sigma}}$**,** so as to preserve the variance components of $\hat{\mathbf{\Sigma}}$. Further, we counteract the problem of size distortions due to the nature of multiple testing by use of Bonferroni (1935, 1936) and Holm (1979) corrections. We compare the absolute values of the non-diagonal entries of $\hat{\boldsymbol{R}}$ with a parameter determined by the inverse of the normal distribution at a prespecified significance level, p. The MT estimator is shown to be reasonably robust to the typical choices of p used in the literature (10% or 5%), and converges to the population correlation matrix $\boldsymbol{R}$ at a rate of $O_p\left(\sqrt{\frac{m_N N}{T}}\right)$ under the Frobenius norm, where m_N is bounded in N, and could represent the number of non-zero off-diagonal elements in each row of $\boldsymbol{R}$.

In many applications, an estimate of the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ is required. Since traditional thresholding, including our multiple testing approach, does not necessarily lead to a positive definite matrix, we recommend supplementary shrinkage applied to our regularised MT correlation matrix when required. To this end, we propose a LW type shrinkage approach where the associated shrinkage parameter is derived from the minimisation of the squared Frobenius norm of the difference between two inverse matrices: an estimate of the inverse matrix of interest (our MT estimator), and the inverse of a suitable reference matrix. We denote this shrinkage version of the MT estimator by S-MT. We also consider a LW type shrinkage estimator applied directly to the sample correlation matrix $\hat{\boldsymbol{R}}$, when the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ is of interest. This shrinkage estimator is denoted by $\hat{\boldsymbol{R}}_{LW}$.

We compare the small sample performance of the MT, S-MT and $\hat{\boldsymbol{R}}_{LW}$ estimators with a number of extant regularised estimators in the literature for large-dimensional covariance matrices in an extended Monte Carlo simulation study. We consider two *approximately* sparse and two *exactly* sparse covariance structures. The simulation results show that the proposed multiple testing and shrinkage based estimators are robust to the different covariance matrix specifications employed, and perform favourably when compared with the widely used regularisation methods considered in

[2] Earlier work by Fan, Fan and Lv (2008) use a strict factor model to impose sparseness on the covariance matrix. Friedman, Hastie and Tibshirani (2008) apply the lasso penalty to loadings in principal component analysis to achieve a sparse representation.

[3] Other contributions to the thresholding literature include the work of Huang et al. (2006), Rothman et al. (2009), Cai and Zou (2011, 2012), and Wang and Zou (2010), among others.

our study, especially when N is large relative to T.

The rest of the paper is organised as follows: Section 2 outlines some preliminaries and definitions. Section 3 introduces our multiple testing (MT) procedure and presents its theoretical properties. Section 4 discusses issues of invertibility of the MT estimator in finite samples and advances our recommended S-MT and $\hat{\boldsymbol{R}}_{LW}$ estimators. Section 5 provides an overview of a number of existing key regularisation techniques. The small sample properties of the MT estimator, its adjusted shrinkage version (S-MT) and $\hat{\boldsymbol{R}}_{LW}$ are investigated in Section 6. Finally Section 7 concludes.

The largest and the smallest eigenvalues of the $N \times N$ matrix $\boldsymbol{A} = (a_{ij})$ are denoted by $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ respectively, $tr(\boldsymbol{A}) = \sum_{i=1}^{N} a_{ii}$ is its trace, $\|\boldsymbol{A}\|_1 = \max_{1\leq j\leq N}\left\{\sum_{i=1}^{N}|a_{ij}|\right\}$ is its maximum absolute column sum norm, $\|\boldsymbol{A}\|_\infty = \max_{1\leq i\leq N}\left\{\sum_{j=1}^{N}|a_{ij}|\right\}$ is its maximum absolute row sum norm, $\|\boldsymbol{A}\|_F = \sqrt{tr(\boldsymbol{A}'\boldsymbol{A})}$ is its Frobenius norm, and $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$ is its spectral (or operator) norm. When $\boldsymbol{A}$ is a vector, both $\|\boldsymbol{A}\|_F$ and $\|\boldsymbol{A}\|$ are equal to the Euclidean norm.

2 Large covariance matrix estimation: Some preliminaries

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). i can rise indefinitely ($i \to \infty$) and denotes units of an unordered population. Conversely, the time dimension t explicitly refers to an ordered set, and can too tend to infinity ($t \to \infty$). We assume that for each $t \in T$, x_{it} is cross-sectionally weakly dependent (CWD), as defined in Chudik et al. (2011). The covariance matrix of $\boldsymbol{x}_t = (x_{1t}, ..., x_{Nt})'$ is given by

$$Var(\boldsymbol{x}_t) = E\left(\boldsymbol{x}_t\boldsymbol{x}_t'\right) = (\sigma_{ij,t}) = \boldsymbol{\Sigma}_t, \tag{1}$$

where, for simplicity of exposition and without loss of generality it is assumed that $E(\boldsymbol{x}_t) = \boldsymbol{0}$, $\boldsymbol{\Sigma}_t$ is an $N \times N$ symmetric, positive definite real matrix with its $(i,j)^{th}$ element, $\sigma_{ij,t}$, given by

$$\begin{aligned} \sigma_{ii,t} &= E\left[x_{it} - E(x_{it})\right]^2 < K, \\ \sigma_{ij,t} &= E\left[\left(x_{it} - E(x_{it})\right)\left(x_{jt} - E(x_{jt})\right)\right], \end{aligned} \tag{2}$$

for $i, j = 1, ..., N$, $t = 1, ..., T$, $\sigma_{ii,t} > 0$ and K is a finite generic constant independent of N. The diagonal elements of $\boldsymbol{\Sigma}_t$ are represented by the $N \times N$ diagonal matrix $\mathbf{D}_t$, such that

$$\boldsymbol{D}_t = diag(\sigma_{11,t}, \sigma_{22,t}, ..., \sigma_{NN,t}). \tag{3}$$

Following the literature we now introduce the concepts of *approximate* and *exact* sparseness of a matrix.

Definition 1 *The $N \times N$ matrix $\boldsymbol{A} = (a_{ij})$ is approximately sparse if, for some $q \in [0, 1)$,*

$$m_N = \max_{i\leq N}\sum_{j\leq N}|a_{ij}|^q \leq c_0(N),\ N \to \infty.$$

Exact sparseness is established when setting $q = 0$. Then, $m_N = \max_{i\leq N}\sum_{j\leq N} I(a_{ij} \neq 0)$ is the maximum number of non-zero elements in each row and is bounded in N, where $I(.)$ denotes the indicator function.

Given the above definition and following Remark 2.2 and Proposition 2.1(a) of Chudik et al. (2011), it follows that under the assumption that x_{it} is CWD, then each row/column of $\boldsymbol{\Sigma}_t$ can only have a finite number of non-zero elements, namely $\|\boldsymbol{\Sigma}_t\|_1 = O(1)$. See also Bailey et al. (2013) and Pesaran (2013).

The estimation of $\mathbf{\Sigma}_t$ gives rise to three main challenges: the sample covariance matrix $\hat{\mathbf{\Sigma}}_t$ becomes firstly ill-conditioned and secondly non-invertible as N increases relative to T, and thirdly $\mathbf{\Sigma}_t$ is likely to become unstable for T sufficiently large. The statistics literature thus far has predominantly focused on tackling the first two problems while largely neglecting the third. On the other hand, in the finance literature time variations in $\mathbf{\Sigma}_t$ are allowed when using conditionally heteroskedastic models such as the Dynamic Conditional Correlation (DCC) model of Engle (2002) or its generalization in Pesaran and Pesaran (2010). However, the DCC approach still requires $T > N$ and it is not applicable when N is large relative to T. This is because the sample correlation matrix is used as the estimator of the unconditional correlation matrix which is assumed to be time invariant.

One can adopt a non-parametric approach to time variations in variances (volatilities) and covariances and base the sample estimate of the covariance matrix on high frequency observations. As measures of volatility (often referred to as realized volatility) intra-day log price changes are used in the finance literature. See, for example, Andersen et al. (2003), and Barndorff-Nielsen and Shephard (2002, 2004). The idea of realized volatility can be adapted easily for use in macro-econometric models by summing squares of daily returns within a given quarter to construct a quarterly measure of market volatility. Also, a similar approach can be used to compute realized measures of correlations, thus yielding a realized correlation matrix. However, such measures are based on a relatively small number of time periods. For example, under the best case scenario where intra-daily observations are available, weekly estimates of realized variance and covariances are based typically on 48 intra-daily price changes and 5 trading days, namely $T = 240$, which is less than the number of securities often considered in practice in portfolio optimisation problems. T can be increased by using rolling windows of observations over a number of weeks or months, but there is a trade off between maintaining stability of the covariance matrix and the size of the time series observations. As T is increased, by considering longer time spans, the probability of the covariance matrix remaining stable over that time span is accordingly reduced.

In this paper we assume that T is sufficiently small so that $\mathbf{\Sigma}_t$ remains constant over the selected time horizon and we concentrate on addressing the remaining two challenges in the estimation of $\mathbf{\Sigma}_t$. We suppress subscript t in $\mathbf{\Sigma}_t$ and $\boldsymbol{D}_t$ and evaluate the sample covariance matrix estimator of $\mathbf{\Sigma}$, denoted by $\hat{\mathbf{\Sigma}}$, with elements

$$\hat{\sigma}_{ij} = T^{-1} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j), \text{ for } i, j = 1, ..., N \tag{4}$$

where $\bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$. The diagonal elements of $\hat{\mathbf{\Sigma}}$ are collected in $\hat{\boldsymbol{D}} = diag(\hat{\sigma}_{ii}, i = 1, 2, ..., N)$.

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

We propose a regularisation method that follows the thresholding literature, where typically, as mentioned in the introduction, non-diagonal elements of the sample covariance matrix that fall below a certain level or 'threshold' in absolute terms are set to zero. Our method tests the statistical significance of all distinct pair-wise covariances or correlations of the sample covariance matrix $\hat{\mathbf{\Sigma}}$, $N(N-1)/2$ in total. As such, this family of tests is prone to size distortions arising from possible dependence across the individual pair-wise tests. We take into account these 'multiple testing' problems in estimation, in an effort to improve support recovery of the true covariance matrix. Our multiple testing (MT) approach is applied to the sample correlation matrix which is arguably more appropriate than the sample covariance matrix, is invariant to the ordering of the variables under consideration, and it is computationally simple to implement.

Suppose that x_{it}, $i = 1, ..., N$, $t = 1, ..., T$, are cross-sectionally weakly correlated with a sparse covariance matrix $\mathbf{\Sigma}$ defined in (1), and with diagonal elements collected in (3), where subscript t

has been suppressed. Consider the $N \times N$ correlation matrix corresponding to $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}^{-1/2} = (\rho_{ij}), \text{ where } \boldsymbol{D} = diag\left(\boldsymbol{\Sigma}\right),$$

with

$$\rho_{ij} = \rho_{ji} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \ i,j = 1,...,N$$

where σ_{ij} is given in (2). We base our thresholding procedure on the correlation matrix. The reasons for opting to work with the correlation matrix rather than its covariance counterpart are twofold. First, the transformation from $\boldsymbol{R}$ to $\boldsymbol{\Sigma}$ leaves the diagonal elements of $\boldsymbol{\Sigma}$ unaffected which is a desirable property in many financial applications. Second, given that all entries in $\boldsymbol{R}$ are bounded from above and below ($-1 \leq \rho_{ij} \leq 1$, $i,j = 1,...,N$), potentially one can use a so called 'universal' parameter to identify the non-zero elements in $\boldsymbol{R}$ rather than making entry-dependent adjustments which in turn need to be estimated**.** This feature is in line with the method of Bickel and Levina (2008b) but shares the properties of the adaptive thresholding estimator developed by Cai and Lui (2011). Both of these approaches are outlined below in Section 5.

The sample correlation matrix, $\hat{\boldsymbol{R}} = (\hat{\rho}_{ij})$, is given by

$$\hat{\boldsymbol{R}} = \hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{D}}^{-1/2},$$

with elements

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} = \frac{\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\left(x_{jt} - \bar{x}_j\right)}{\left(\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)^2\right)^{1/2}\left(\sum_{t=1}^{T}\left(x_{jt} - \bar{x}_j\right)^2\right)^{1/2}}, \ i = 1,2,...,N, \ t = 1,2,...,T.$$

Now for a sufficiently large T, the correlation coefficients $\hat{\rho}_{ij}$ are approximately normally distributed as[4]

$$\hat{\rho}_{ij} \sim N\left(\mu_{ij}, \omega_{ij}^2\right), \tag{5}$$

where (using Fisher's (1915) bias correction - see also Soper (1913)) we have

$$\mu_{ij} = \rho_{ij} - \frac{\rho_{ij}(1-\rho_{ij}^2)}{2T} \text{ and } \omega_{ij}^2 = \frac{(1-\rho_{ij}^2)^2}{T}.$$

Joint tests of $\rho_{ij} = 0$ for $i = 1,2,...,N-1$, $j = i+1,...,N$ can now be carried out, allowing for the cross dependence of the individual tests using a suitable multiple testing (MT) procedure. This yields the following MT estimator of $\mathbf{R}$,

$$\widetilde{\boldsymbol{R}}_{MT} = \left(\tilde{\rho}_{ij}\right) = \left[\hat{\rho}_{ij} I(\sqrt{T}\left|\hat{\rho}_{ij}\right| > b_N)\right], \ i = 1,2,...,N-1, \ j = i+1,...,N. \tag{6}$$

where

$$b_N = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right). \tag{7}$$

The indicator function $I(.)$ used in (6), is in line with the concept of 'hard' thresholding whereby all elements of $\hat{\boldsymbol{\Sigma}}$ or $\hat{\boldsymbol{R}}$ that drop below a certain level in absolute terms are set to zero. The remaining ones are equated to their original sample covariance or correlation coefficients. Multiple testing (MT) does not consider functions used in the 'soft' thresholding literature (see for example Antoniadis and Fan (2001), Rothman et al. (2009), and Cai and Liu (2011)).

Parameter b_N is of special importance. It is determined by the inverse of the cumulative distribution function of the standard normal variate, $\Phi^{-1}(.)$, using a prespecified overall test size,

[4]Other functions of $\hat{\rho}_{ij}$, such as the Fisher's transformation can also be used. But our simulation exercises suggested that there is little to choose between $\hat{\rho}_{ij}$ or its Fisher transform.

p, selected for the joint testing problem. The size of the test is normalised by $f(N)$, which controls for the multiple testing nature of the testing problem in (6). As mentioned above, testing the null hypothesis that $\rho_{ij} = 0$ for $i = 1, 2, ..., N-1$, $j = i+1, ..., N$ can result in spurious outcomes, especially when N is larger than T, due to the multiple tests being conducted across the $N(N-1)/2$ distinct elements of $\hat{\boldsymbol{R}}$.

Suppose that we are interested in a family of null hypotheses, $H_{01}, H_{02}, ..., H_{0r}$, and we are provided with corresponding test statistics, Z_{1T},Z_{2T}, ..., Z_{rT}, 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

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

and suppose that we wish to control $FWER_T$ to lie below a pre-determined value, p. Bonferroni (1935, 1936) provides a general solution, which holds for all possible degrees of dependence across the separate tests. By Boole's inequality we have

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

Hence to achieve $FWER_T \leq p$, it is sufficient to set $p_{iT} \leq p/r$.

However, as is known Bonferroni's procedure can be quite conservative and a number of alternative multiple testing procedures have been proposed in the literature. One prominent example is the step-down procedure proposed by Holm (1979) which is less conservative than the Bonferroni procedure, and does not impose any further restrictions on the degree to which the underlying tests depend on each other. If we abstract from the T subscript and order the p-values of the tests so that

$$p_{(1)} \leq p_{(2)} \leq .... \leq p_{(r)}$$

are associated with the null hypotheses, $H_{(01)}, H_{(02)}, ..., H_{(0r)}$, respectively, Holm's procedure rejects $H_{(01)}$ if $p_{(1)} \leq p/r$, rejects $H_{(01)}$ and $H_{(02)}$ if $p_{(2)} \leq p/(r-1)$, rejects $H_{(01)}, H_{(02)}$ and $H_{(03)}$ if $p_{(3)} \leq p/(r-2)$, and so on. Returning to (6) we observe that under the null i and j are unconnected, and $\hat{\rho}_{ij}$ is approximately distributed as $N\left(0, T^{-1}\right)$. Therefore, the p-values of the individual tests are (approximately) given by $p_{ij} = 2\left[1 - \Phi\left(\sqrt{T}\left|\hat{\rho}_{ij}\right|\right)\right]$ for $i = 1, 2, ..., N-1$, $j = i+1, ..., N$, with the total number of tests being carried out given by $r = N(N-1)/2$. To apply the Holm procedure we need to order these p-values in an ascending manner, which is equivalent to ordering $\left|\hat{\rho}_{ij}\right|$ in a descending manner. Denote the largest value of $\left|\hat{\rho}_{ij}\right|$ over all $i \neq j$, by $\left|\hat{\rho}_{(1)}\right|$, the second largest value by $\left|\hat{\rho}_{(2)}\right|$, and so on, to obtain the ordered sequence $\left|\hat{\rho}_{(s)}\right|$, for $s = 1, 2, ..., r$. Then the (i, j) pair associated with $\left|\hat{\rho}_{(s)}\right|$ are connected if $\left|\hat{\rho}_{(s)}\right| > T^{-1/2}\Phi^{-1}\left(1 - \frac{p/2}{N(N-1)/2-s+1}\right)$, otherwise disconnected, for $s = 1, 2, ..., N(N-1)/2$, where p is the pre-specified overall size of the test. Note that if the Bonferroni approach is implemented no such ordering is required and to see if the (i, j) pair is connected it suffices to assess whether $\left|\hat{\rho}_{ij}\right| > T^{-1/2}\Phi^{-1}\left(1 - \frac{p/2}{N(N-1)/2}\right)$.

There is also the issue of whether to apply the multiple testing procedure to all distinct $N(N-1)/2$ non-diagonal elements of $\hat{\boldsymbol{R}} = (\hat{\rho}_{ij})$ simultaneously, or to apply the procedure row-wise, by considering N separate families of $N-1$ tests defined by $\rho_{i^0 j} = 0$, for a given i^0, and $j = 1, 2, .., N$,

$j \neq i^0$. The theoretical results derived in (3.1) show that using $f(N) = N(N-1)/2$ in (7) rather than $f(N) = (N-1)$ provides a faster rate of convergence towards $\boldsymbol{R}$ under the Frobenius norm. However, simulation results of Section 6 indicate that in finite samples $f(N) = N-1$ can provide $\widetilde{\boldsymbol{R}}_{MT}$ estimates that perform equally well and even better than when $f(N) = N(N-1)/2$ is considered, depending on the setting. Note that multiple testing using the Holm approach can lead to contradictions if applied row-wise. To see this consider the simple case where $N = 3$ and p values for the three rows of $\hat{\boldsymbol{R}}$ are given by

$$\begin{pmatrix} - & p_1 & p_2 \\ p_1 & - & p_3 \\ p_2 & p_3 & - \end{pmatrix}.$$

Suppose that $p_1 < p_2 < p_3$. Then $\rho_{13} = 0$ is rejected if $p_2 < p$ when Holm's procedure is applied to the first row, and rejects $\rho_{13} = 0$ if $p_2 < p/2$ when the procedure is applied to the third row. To circumvent this problem in practice, if one of the ρ_{13} hypotheses is rejected but the other is accepted then we set both relevant elements in $\widetilde{\boldsymbol{R}}_{MT}$ to $\hat{\rho}_{13}$ using this example. The row-wise application of Bonferroni's procedure is not subject to this problem since it applies the same p-value of $p/(N-1)$ to all elements of $\hat{\boldsymbol{R}}$.[5]

After applying multiple testing to the unconditional sample correlation matrix, we recover the corresponding covariance matrix $\widetilde{\boldsymbol{\Sigma}}_{MT}$ by pre- and post-multiplying $\widetilde{\boldsymbol{R}}_{MT}$ by the square root of the diagonal elements of $\hat{\boldsymbol{\Sigma}}$, so that

$$\widetilde{\boldsymbol{\Sigma}}_{MT} = \hat{\boldsymbol{D}}^{1/2} \widetilde{\boldsymbol{R}}_{MT} \hat{\boldsymbol{D}}^{1/2}. \tag{8}$$

It is evident that since b_N is given and does not need to be estimated, the multiple testing procedure in (6) is also computationally simple to implement. This contrasts with traditional thresholding approaches which face the challenge of evaluating the theoretical constant, C, arising in the rate of convergence of their estimators. The computationally intensive cross validation procedure is typically employed for the estimation of C, which is further discussed in Section 5.

Finally, in the presence of factors in the data set $\boldsymbol{x}_t$ (as in the setting used in Fan, Liao and Mincheva (2011, 2013 - FLM)), we proceed as shown in FLM by estimating the covariance matrix of the residuals $\hat{\boldsymbol{u}}_t = (\hat{u}_{1t}, ..., \hat{u}_{Nt})'$ obtained from defactoring the data, $\hat{\boldsymbol{\Sigma}}_{\hat{u}}$, and applying the multiple testing approach to $\hat{\boldsymbol{\Sigma}}_{\hat{u}}$.[6] In this case, (6) is modified to correct for the degrees of freedom, m, associated with the defactoring regression:

$$\tilde{\rho}_{\hat{u},ij} = \hat{\rho}_{\hat{u},ij} I(\sqrt{T-m}\,|\hat{\rho}_{\hat{u},ij}| > b_N),\ i = 1, 2, ..., N-1,\ j = i+1, ..., N \tag{9}$$

where

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

An example of multiple testing applied to regression residuals is considered in our simulation study of Section 6. See also Bailey et al. (2013).

[5] Other multiple testing procedures can also be considered (see Efron (2010) for a recent review). But most of these methods tend to place undue prior restrictions on the dependence of the underlying test statistics while the Bonferroni and Holm methods are not subject to this problem.

[6] Assume a factor model:

$$y_{it} = \gamma_i' \hat{\mathbf{f}}_t + u_{it},\ i = 1, 2, ..., N;\ t = 2, ..., T,$$

where, $\hat{\mathbf{f}}_t$ is an $m \times 1$ vector of factors estimated through principal components (Bai (2003)) or cross-sectional averages (Pesaran (2006)), and $\gamma_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{i\ell})'$ is the associated vector of factor loadings. Then, the defactoring analysis entails running the above regressions and extracting the residuals:

$$\hat{u}_{it} = y_{it} - \hat{\gamma}_i' \hat{\mathbf{f}}_t,\ i = 1, 2, ..., N;\ t = 2, ..., T.$$

3.1 Theoretical properties of the MT estimator

In this subsection we investigate the asymptotic properties of the MT estimator defined in (6). We establish its rate of convergence under the Frobenius norm as well as the conditions for consistent support recovery via the true positive rate (TPR) and the false positive rate (FPR), to be defined below. We begin by stating a couple of assumptions that will be used in our proofs.

Assumption 1 *Let $\hat{\boldsymbol{R}}=(\hat{\rho}_{ij})$ be the sample correlation matrix, and suppose that (for sufficiently large T)*

$$\hat{\rho}_{ij} \sim N\left(\mu_{ij}, \omega_{ij}^2\right), \tag{10}$$

where

$$\mu_{ij} = E(\hat{\rho}_{ij}) = \rho_{ij} - \frac{\rho_{ij}(1-\rho_{ij}^2)}{2T} + \frac{G(\rho_{ij})}{T^2}, \tag{11}$$

$$\omega_{ij}^2 = Var(\hat{\rho}_{ij}) = \frac{(1-\rho_{ij}^2)^2}{T} + \frac{K(\rho_{ij})}{T^2}, \tag{12}$$

and $G(\rho_{ij})$ and $K(\rho_{ij})$ are bounded in ρ_{ij} and T, for all i and $j=1,2,...,N$.

The analytical expressions for the mean and variance of $\hat{\rho}_{ij}$ in (11) and (12) of Assumption 1 can be found in Soper et al. (1917).

Assumption 2 *The population correlation matrix, $\boldsymbol{R}=(\rho_{ij})$, is sparse according to Definition 1 such that only m_N of its non-diagonal elements in each row are non-zero satisfying the condition*

$$0 < \rho_{\min} < \left|\rho_{ij}\right| < \rho_{\max} < 1,$$

with m_N being bounded in N. The remaining $N(N-m_N-1)$ non-diagonal elements of $\boldsymbol{R}$ are zero.

Assumption 2 implies exact sparseness under Definition 1.

Theorem 1 ***(Rate of convergence)*** *Denote the sample correlation coefficient of x_{it} and x_{jt} over $t=1,2,...,T$ by $\hat{\rho}_{ij}$ and the population correlation matrix by $\boldsymbol{R}=(\rho_{ij})$, which obey Assumptions 1 and 2 respectively. Also let $f(N)$ be an increasing function of N, such that*

$$\frac{\ln\left[f(N)\right]}{T} = o(1), \text{ as } N \text{ and } T \to \infty.$$

Then

$$E\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F^2 = \sum_{i\neq j}\sum E(\tilde{\rho}_{ij} - \rho_{ij})^2 = O\left(\frac{m_N N}{T}\right), \tag{13}$$

where $\widetilde{\boldsymbol{R}}_{MT} = (\tilde{\rho}_{ij})$

$$\tilde{\rho}_{ij} = \hat{\rho}_{ij} I\left(\left|\hat{\rho}_{ij}\right| > \frac{b_N}{\sqrt{T}}\right), \text{ with } b_N = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right) > 0,$$

and p is a given overall Type I error.

Proof. *See Appendix A.* ■

Result (13) implies that $N^{-1}\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F^2 = O_p\left(\frac{m_N}{T}\right)$ which is in line with the existing results in the thresholding literature that use the Frobenius norm. See, for example, Theorem 2 with $q=0$ in Bickel and Levina (2008b). The same rate of $O_p\left(m_N/T\right)$ is achieved in the shrinkage literature if the assumption of sparseness is imposed. Here m_N can also be assumed to rise with N in which case the rate of convergence becomes slower. This compares with a rate of $O_p\left(N/T\right)$ for the sample covariance (correlation) matrix - see Theorem 3.1 in Ledoit and Wolf (2004). Note that LW use an unconventional definition for the Frobenius norm (see their Definition 1). Similar results can also be obtained for the spectral norm.

Theorem 2 ***(Support Recovery)*** *Consider the true positive rate (TPR) and the false positive rate (FPR) statistics computed using the multiple testing estimator* $\tilde{\rho}_{ij} = \hat{\rho}_{ij} I\left(\left|\hat{\rho}_{ij}\right| > \frac{b_N}{\sqrt{T}}\right)$*, given by*

$$TPR = \frac{\sum\sum_{i\neq j} I(\tilde{\rho}_{ij} \neq 0, \text{ and } \rho_{ij} \neq 0)}{\sum\sum_{i\neq j} I(\rho_{ij} \neq 0)} \tag{14}$$

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

respectively, where b_N *is defined as in Theorem 1, and* $\hat{\rho}_{ij}$ *and* ρ_{ij} *obey Assumptions 1 and 2, respectively. Then with probability tending to 1,* $FRP = 0$ *and* $TPR = 1$*, if* $\rho_{\min} = \min_{i\neq j}(\rho_{ij}) > \frac{b_N}{\sqrt{T}}$ *as* $N, T \to \infty$ *in any order.*

Proof. *See Appendix A.* ■

4 Positive definiteness of the covariance matrix estimator

As in the case of thresholding approaches, multiple testing preserves the symmetry of $\hat{\boldsymbol{R}}$ and is invariant to the ordering of the variables. However, it does not ensure positive definiteness of the estimated covariance matrix. Bickel and Levina (2008b) provide an asymptotic condition that ensures positive definiteness, which is not met unless T is sufficiently large relative to N**.** See Section 5 for the exact specification of this condition. Guillot and Rajaratnam (2012) demonstrate theoretically that retaining positive definiteness upon thresholding is governed by complex algebraic conditions. In particular, they show that the pattern of elements to be set to zero has to correspond to a graph which is a union of complete components.

A number of methods have been developed in the literature that produce sparse inverse covariance matrix estimates. A popular approach applies the penalised likelihood with a LASSO penalty to the off-diagonal terms of $\boldsymbol{\Sigma}^{-1}$. See, for example, D'Aspremont et al. (2008), Rothman et al. (2008), Yuan and Lin (2007), and Peng et al. (2009). More recent contributions propose a sparse positive definite covariance estimator obtained via convex optimisation, where sparseness is achieved by use of a suitable penalty. For example, Rothman (2012) uses a logarithmic barrier term, Xue et al. (2012) impose a positive definiteness constraint, while Liu et al. (2013) and Fan et al. (2013) enforce an eigenvalue condition.[7] Most of these approaches are rather complex and computationally extensive. Instead, if inversion of $\hat{\boldsymbol{R}}$ or $\hat{\boldsymbol{\Sigma}}$ is of interest we recommend the use of a Ledoit-Wolf (LW) type shrinkage estimator, either applied to the MT estimated correlation matrix, $\widetilde{\boldsymbol{R}}_{MT}$, or to the sample correlation, $\hat{\boldsymbol{R}}$, itself. The latter estimator is motivated by the work of Schäfer and Strimmer (2005) who draw on the theoretical results of LW. However, they do not account for the bias of the empirical correlation coefficients, which we do in our specification of $\hat{\boldsymbol{R}}_{LW}$.

Following Ledoit and Wolf (2004) (see Section 5 for a summary of their approach), we set as benchmark target the $N \times N$ identity matrix $\boldsymbol{I}_N$. Our shrinkage on multiple testing (S-MT) estimator is then defined by

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT} = \lambda \boldsymbol{I}_N + (1-\lambda)\widetilde{\boldsymbol{R}}_{MT}, \tag{16}$$

where the shrinkage parameter $\lambda \in (\lambda_0, 1]$, and λ_0 is the minimum value of λ that produces a non-singular $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\lambda_0)$ matrix. First note that shrinkage is again deliberately implemented on the

[7] Other related work includes that of Lam and Fan (2009), Rothman et al. (2009), Bien and Tibshirani (2011), Cai et al. (2011), and Yuan and Wang (2013).

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 volatility measures. Second, shrinkage is applied to the non-zero elements of $\widetilde{\boldsymbol{R}}_{MT}$, and as a result the shrinkage estimator, $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$, has the same optimal non-zero/zero patterns obtained for $\widetilde{\boldsymbol{R}}_{MT}$. This is in contrast to thresholding approaches that impose eigenvalue restrictions to achieve positive definiteness.

The shrinkage parameter used in (16) is derived from a grid search optimisation procedure described below that involves the inverse of two matrices. Specifically, we consider a reference correlation matrix, $\boldsymbol{R}_0$, which is selected to be well-conditioned, robust and positive definite. Next, over a grid of λ bounded from below and above by λ_0 and 1 respectively, $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\lambda)$ is evaluated. Since both $\boldsymbol{R}_0$ and $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\lambda)$ are positive definite, the difference of their inverses is compared over $\lambda \in (\lambda_0, 1]$ using the Frobenius norm. The shrinkage parameter, λ^*, is given by

$$\lambda^* = \arg \min_{\lambda_0+\epsilon \leq \lambda \leq 1} \left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\lambda) \right\|_F^2, \tag{17}$$

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

$$\left\| \boldsymbol{R}_0^{-1} - \widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\lambda) \right\|_F^2 = tr\left(\boldsymbol{A}^2\right) - 2tr[\boldsymbol{A}\boldsymbol{B}(\lambda)] + tr[\boldsymbol{B}^2(\lambda)]. \tag{18}$$

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

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

where

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

Hence, λ^* is obtained as the solution of

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

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

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

for any $\lambda \in [\lambda_0, 1]$, where $\lambda_0 = \max\left(\frac{\epsilon - \lambda_{\min}(\tilde{\boldsymbol{R}}_{MT})}{1 - \lambda_{\min}(\tilde{\boldsymbol{R}}_{MT})}, 0 \right)$. For more details of the above derivations and the grid search optimisation procedure see Appendix A.

Having obtained the shrinkage estimator $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$, using λ^* in (16), we construct the corresponding covariance matrix as

$$\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT} = \hat{\boldsymbol{D}}^{1/2} \widetilde{\boldsymbol{R}}_{S\text{-}MT} \hat{\boldsymbol{D}}^{1/2}. \tag{19}$$

Implementation of the above procedure requires the use of a suitable reference matrix $\boldsymbol{R}_0$. To this end, we propose using a LW type shrinkage estimator, $\hat{\boldsymbol{R}}_{LW}$, applied to the sample correlation matrix itself. This appears to work better in practice over the more natural choice of the identity matrix or even the generalised inverse of the sample correlation, which we experimented with. The same is true when compared to the correlation matrix derived from shrinking $\hat{\boldsymbol{\Sigma}}$ using the Ledoit

and Wolf (2004) method. In the simulations that follow, we use $\hat{\boldsymbol{R}}_{LW}$ as the reference matrix for $\widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT}$. However, $\hat{\boldsymbol{R}}_{LW}$, can also be used independently of the $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$ estimator. Thus, we also evaluate its performance when implemented on its own in obtaining the inverse of the covariance matrix, and make relevant recommendations.

Consider the following shrinkage estimator of $\boldsymbol{R}$,

$$\hat{\boldsymbol{R}}_{LW} = \xi \boldsymbol{I}_N + (1-\xi)\hat{\boldsymbol{R}},$$

with shrinkage parameter $\xi \in [0,1]$, where $\hat{\boldsymbol{R}} = (\hat{\rho}_{ij})$. The squared Frobenius norm of the error of estimating $\boldsymbol{R}$ by $\hat{\boldsymbol{R}}_{LW}(\xi)$ is given by

$$\begin{aligned} \left\| \hat{\boldsymbol{R}}_{LW}(\xi) - \boldsymbol{R} \right\|_F^2 &= \sum_{i \neq j} \sum \left[(1-\xi)\hat{\rho}_{ij} - \rho_{ij} \right]^2 \\ &= \sum_{i \neq j} \sum \left[\hat{\rho}_{ij} - \rho_{ij} - \xi \hat{\rho}_{ij} \right]^2. \end{aligned}$$

The main theoretical results for the shrinkage estimator based on the sample correlation matrix are summarised in the theorem below.

Theorem 3 *(**Rate of convergence and optimal shrinkage parameter**) Denote the sample correlation coefficient of x_{it} and x_{jt} over $t = 1, 2, ..., T$ by $\hat{\rho}_{ij}$ and the population correlation matrix by $\boldsymbol{R} = (\rho_{ij})$. Suppose also that Assumptions 1 and 2 are satisfied. Then*

$$E \left\| \hat{\boldsymbol{R}}_{LW}(\xi^*) - \boldsymbol{R} \right\|_F^2 = \sum_{i \neq j} \sum E \left[\hat{\rho}_{ij} - \rho_{ij} - \xi^* \hat{\rho}_{ij} \right]^2 = O\left(\frac{m_N N}{T} \right), \tag{20}$$

where ξ^* is the optimal value of the shrinkage parameter ξ, which is given by

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

Proof. See Appendix A. ■

Corollary 1 *Denote the sample correlation coefficient of x_{it} and x_{jt} over $t = 1, 2, ..., T$ by $\hat{\rho}_{ij}$ and the population correlation matrix by $\mathbf{R} = (\rho_{ij})$. Then*

$$\begin{aligned} E \left\| \hat{\boldsymbol{R}}_{LW}(\xi^*) - \boldsymbol{R} \right\|_F^2 &= \sum_{i \neq j} \sum E \left(\hat{\rho}_{ij} - \rho_{ij} \right)^2 - \frac{\left[\sum_{i \neq j} \sum E \left[\hat{\rho}_{ij} \left(\hat{\rho}_{ij} - \rho_{ij} \right) \right] \right]^2}{\sum_{i \neq j} \sum E \left(\hat{\rho}_{ij}^2 \right)} \\ &< \sum_{i \neq j} \sum E \left(\hat{\rho}_{ij} - \rho_{ij} \right)^2. \end{aligned}$$

Proof. See Appendix A. ■

From Corollary 1, assuming that T is sufficiently large so that ρ_{ij} can be reasonably accurately estimated by $\hat{\rho}_{ij}$, we would expect the shrinkage estimator to have smaller mean squared error than $\hat{\boldsymbol{R}}$. Recovery of the corresponding covariance matrix $\hat{\boldsymbol{\Sigma}}_{LW}(\xi^*)$ is performed as in (19).

5 An overview of key regularisation techniques

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

5.1 Bickel-Levina (BL) thresholding

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

$$\widetilde{\mathbf{\Sigma}}_{BL,C} = \left(\hat{\sigma}_{ij} I \left[|\hat{\sigma}_{ij}| \geq C \sqrt{\frac{\log N}{T}} \right] \right), \ i = 1, 2, ..., N-1, \ j = i+1, ..., N \tag{21}$$

where $I(.)$ is an indicator function and C is a positive constant which is unknown. The choice of thresholding function - $I(.)$ - implies that (21) implements 'hard' thresholding. The consistency rate of the BL estimator is $\sqrt{\frac{\log N}{T}}$ under the spectral norm of the error matrix $\left(\widetilde{\mathbf{\Sigma}}_{BL,C} - \mathbf{\Sigma}\right)$. The main challenge in the implementation of this approach is the estimation of the thresholding parameter, C, which is usually calibrated by cross validation.[8] Details of the BL cross validation procedure can be found in Appendix B.

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

$$\left\| \widetilde{\mathbf{\Sigma}}_{BL,C} - \widetilde{\mathbf{\Sigma}}_{BL,0} \right\| \leq \epsilon \text{ and } \lambda_{\min}(\mathbf{\Sigma}) > \epsilon, \tag{22}$$

where $\|.\|$ 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. Furthermore, it is generally acknowledged that the cross validation technique used for estimating C is computationally expensive. More importantly, cross validation performs well only when $\mathbf{\Sigma}$ is assumed to be stable over time. If a structural break occurs on either side of the cross validation split chosen over the T dimension then the estimate of C could be biased. Finally, 'universal' thresholding on $\hat{\mathbf{\Sigma}}$ performs best when the units x_{it}, $i = 1, ..., N$, $t = 1, ..., T$ are assumed homoscedastic (i.e. $\sigma_{11} = \sigma_{22} = ... = \sigma_{NN}$). Departure from such a setting can have a negative impact on the properties of the thresholding parameter.

5.2 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. More specifically, the thresholding estimator $\widetilde{\mathbf{\Sigma}}_{CL,C}$ is defined as

$$\widetilde{\mathbf{\Sigma}}_{CL,C} = \left(\hat{\sigma}_{ij} s_{\tau_{ij}} \left[|\hat{\sigma}_{ij}| \geq \tau_{ij} \right] \right), \ i = 1, 2, ..., N-1, \ j = i+1, ..., N \tag{23}$$

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

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

of the CL estimator is $\sqrt{\log N/T}$ under the spectral norm of the error matrix $\left(\widetilde{\mathbf{\Sigma}}_{CL,C} - \mathbf{\Sigma}\right)$. 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 Appendix B.

As with the BL estimator, thresholding in itself does not ensure positive definiteness of $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}$. In light of condition (22), Fan, Liao and Mincheva (FLM) (2011, 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 thresholded estimator is positive, $\lambda_{\min}\left(\widetilde{\mathbf{\Sigma}}_{FLM,\hat{C}}\right) > 0$. Further details of this procedure can be found in Appendix B. We apply this extension to both BL and CL procedures. The problem of $\widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}$ and $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}$ not being invertible in finite samples is then resolved. However, depending on the application, the selected C might not be necessarily optimal (see Appendix B for the relevant expressions). In other words, the properties of the constrained $\widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}$ and $\widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}$ can deviate noticeably from their respective unconditional versions.

5.3 Ledoit and Wolf (LW) shrinkage

Ledoit and Wolf (2004, LW) considered a shrinkage estimator for regularisation which is based on a convex 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

$$\hat{\mathbf{\Sigma}}_{LW} = \hat{\rho}_1 \boldsymbol{I}_N + \hat{\rho}_2 \hat{\mathbf{\Sigma}}, \tag{24}$$

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 $\dot{\boldsymbol{x}}_t = (\dot{x}_{1t}, ..., \dot{x}_{Nt})'$ and $\dot{x}_{it} = (x_{it} - \bar{x}_i)$.[9]

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

As explained in LW and in subsequent contributions to this literature, shrinkage can be seen as a trade-off between bias and variance in estimation of $\mathbf{\Sigma}$**,** as captured by the choices of ρ_1 and ρ_2. Note however that LW do not require these parameters to add up to unity, and it is possible for the shrinkage method to place little weight on the data (i.e. the correlation matrix). Of particular importance is the effect that LW shrinkage has on the diagonal elements of $\hat{\mathbf{\Sigma}}$ which renders it inappropriate for use in impulse response analysis where the size of the shock is calibrated to the standard deviation of the variables. Unlike the thresholding approaches considered in this paper, the LW methodology does not require $\mathbf{\Sigma}$ to be sparse.

[9] 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.

6 Small sample properties

Using Monte Carlo simulations we investigate the small sample properties of our proposed multiple testing (MT) estimators as compared to the other thresholding and shrinkage type estimators proposed in the literature and reviewed in Section 5. In what follows we present the MT results using the Bonferroni procedure. We obtain very similar results when we use the Holm approach, and to save space the MT results based on Holm procedure are provided in a supplementary appendix which is available on request.

Given the importance of the type of covariance matrix being estimated, we consider four experiments with four different types of covariance matrices.

(A) a first order autoregressive specification (AR);

(B) a first order spatial autoregressive model (SAR);

(C) a banded matrix with ordering used in CL (Model 1);

(D) a covariance structure that is based on a pre-specified number of non-zero off-diagonal elements.

The first two experiments produce standard covariance matrices used in the literature and comply to the *approximately* sparse covariance settings. The covariances in experiments C and D are examples of *exactly* sparse covariance matrices. Results are reported for $N = \{30, 100, 200, 400\}$ and $T = \{60, 100\}$.

As explained in Section 2, we are interested in our MT and shrinkage estimators producing covariance matrix estimates that are not only well-conditioned (and, when needed, invertible) but also relatively stable over time. For this purpose we conduct our simulation exercises using values of T that are relatively small but still sufficient to produce reliable covariance/correlation coefficient estimates. A robustness analysis is also conducted for these setups.

Experiment A In this experiment we set $\mathbf{\Sigma}$ to the covariance matrix of a first-order autoregressive process with coefficient, ϕ,

$$\mathbf{\Sigma} = (\sigma_{ij}) = \frac{1}{1-\phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{N-1} \\ \phi & 1 & & & \vdots \\ \phi^2 & \phi & \ddots & & \vdots \\ \vdots & \cdots & \cdots & \ddots & \phi \\ \phi^{N-1} & \cdots & \cdots & \phi & 1 \end{pmatrix}_{N \times N}.$$

For $|\phi| < 1$, this matrix has a well-defined inverse given by

$$\mathbf{\Sigma}^{-1} = \left(\sigma^{ij}\right) = \begin{pmatrix} 1 & -\phi & 0 & \cdots & 0 \\ -\phi & 1+\phi^2 & & & \vdots \\ 0 & -\phi & \ddots & & \vdots \\ \vdots & \cdots & -\phi & 1+\phi^2 & -\phi \\ 0 & \cdots & \cdots & -\phi & 1 \end{pmatrix}_{N \times N}.$$

The corresponding correlation matrix is given by $\boldsymbol{R} = \left(1-\phi^2\right)\mathbf{\Sigma}$, and it is easily seen that $\mathbf{\Sigma}^{-1} = \boldsymbol{Q}'\boldsymbol{Q}$, where

$$\boldsymbol{Q} = (q_{ij}) = \begin{pmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 \\ -\phi & 1 & & & \vdots \\ 0 & -\phi & \ddots & & \vdots \\ \vdots & \cdots & -\phi & 1 & 0 \\ 0 & \cdots & \cdots & -\phi & 1 \end{pmatrix}_{N \times N}.$$

The data generating process is then given by

$$\boldsymbol{Q}\boldsymbol{x}_t^{(r)} = \boldsymbol{\varepsilon}_t^{(r)},\ t = 1, ..., T. \tag{25}$$

Here $\boldsymbol{x}_t^{(r)}$= $(x_{1t}^{(r)}, x_{2t}^{(r)}, ..., x_{Nt}^{(r)})'$, $\boldsymbol{\varepsilon}_t^{(r)}$= $(\varepsilon_{1t}^{(r)}, \varepsilon_{2t}^{(r)}, ..., \varepsilon_{Nt}^{(r)})'$ and $\varepsilon_{it}^{(r)} \sim IIDN(0,1)$ are generated for each replication $r = 1, ..., R$.

Equivalently, (25) can be written as

$$\begin{aligned} x_{1t}^{(r)} &= \frac{1}{\sqrt{1-\phi^2}}\varepsilon_{1t}^{(r)}, \\ x_{it}^{(r)} &= \phi x_{i-1,t}^{(r)} + \varepsilon_{it}^{(r)},\ \text{for } i = 2, ..., N. \end{aligned}$$

We set $\phi = 0.7$. The sample covariance matrix of $\boldsymbol{x}_t^{(r)}$ is computed as

$$\hat{\boldsymbol{\Sigma}}^{(r)} = T^{-1}\sum_{t=1}^{T} \dot{\boldsymbol{x}}_t^{(r)} \dot{\boldsymbol{x}}_t^{(r)\prime}, \tag{26}$$

for each replication r, where $\dot{\boldsymbol{x}}_t^{(r)} = \left(\dot{x}_{1t}^{(r)}, ..., \dot{x}_{Nt}^{(r)}\right)'$, $\dot{x}_{it}^{(r)} = \left(x_{it}^{(r)} - \bar{x}_i^{(r)}\right)$ and $\bar{x}_i^{(r)} = T^{-1}\sum_{t=1}^{T} x_{it}^{(r)}$, for $i = 1, ..., N$. The corresponding sample correlation matrix, $\hat{\boldsymbol{R}}^{(r)}$ is expressed as

$$\hat{\boldsymbol{R}}^{(r)} = \hat{\boldsymbol{D}}^{-1/2(r)}\hat{\boldsymbol{\Sigma}}^{(r)}\hat{\boldsymbol{D}}^{-1/2(r)}, \tag{27}$$

where $\hat{\boldsymbol{D}}^{(r)}$=$diag(\ \hat{\sigma}_{ii}^{(r)},\ i = 1, 2, ..., N)$.

Experiment B Here we examine a standard first-order spatial autoregressive model (SAR). The data generating process for replication r is now given by

$$\begin{aligned} \boldsymbol{x}_t^{(r)} &= \vartheta \boldsymbol{W}\boldsymbol{x}_t^{(r)} + \boldsymbol{\varepsilon}_t^{(r)} \\ &= (\boldsymbol{I}_N - \vartheta \boldsymbol{W})^{-1}\boldsymbol{\varepsilon}_t^{(r)},\ t = 1, ..., T, \end{aligned} \tag{28}$$

where $\boldsymbol{x}_t^{(r)} = (x_{1t}^{(r)}, x_{2t}^{(r)}, ..., x_{Nt}^{(r)})'$, ϑ is the spatial autoregressive parameter, $\varepsilon_{it}^{(r)} \sim IIDN(0, \sigma_{ii})$, and $\sigma_{ii} \sim IID\left(\frac{1}{2} + \frac{\chi^2(2)}{4}\right)$. Therefore, $E(\sigma_{ii}) = 1$ and σ_{ii} is bounded away from zero, for $i = 1, ..., N$. The weights matrix $\boldsymbol{W}$ is row-standardized 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 intensity of cross-sectional dependence of $\boldsymbol{x}_t^{(r)}$ is measured by ϑ. We set $\vartheta = 0.4$. The population covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = (\boldsymbol{I}_N - \vartheta \boldsymbol{W})^{-1}\boldsymbol{D}(\boldsymbol{I}_N - \vartheta \boldsymbol{W}')^{-1},$$

where $\boldsymbol{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})$, its inverse by

$$\boldsymbol{\Sigma}^{-1} = (\boldsymbol{I}_N - \vartheta \boldsymbol{W}')\boldsymbol{D}^{-1}(\boldsymbol{I}_N - \vartheta \boldsymbol{W}),$$

and $\mathbf{R} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$.

We generate the sample covariance and correlation matrices $\hat{\boldsymbol{\Sigma}}$ and $\hat{\mathbf{R}}$ as in experiment A using (26) and (27).

Experiment C Following Model 1 of Cai and Liu (2011), we consider the banded matrix given by,

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

where $\boldsymbol{A}_1 = (\sigma_{ij})_{1\leq i,j\leq N/2}$, $\sigma_{ij} = (1 - \frac{|i-j|}{10})_+$ and $\boldsymbol{A}_2 = 4\boldsymbol{I}_{N/2}$. $\boldsymbol{\Sigma}$ is a two block diagonal (non-invertible) matrix, $\boldsymbol{A}_1$ is a banded and sparse covariance matrix, and $\boldsymbol{A}_2$ is a diagonal matrix with 4 along the diagonal. Here $\boldsymbol{x}_t^{(r)} = (x_{1t}^{(r)}, x_{2t}^{(r)}, ..., x_{Nt}^{(r)})'$ are generated as $IIDN$-variate random vectors from the normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

Experiment D Under this experiment we consider a covariance structure that explicitly controls for the number of non-zero elements of the population correlation matrix. First we draw $N \times 1$ vectors $\boldsymbol{b} = (b_1, b_2, ..., b_N)'$ as $Uniform\,(0.7, 0.9)$ for the first and last $N_b\,(< N)$ elements of $\boldsymbol{b}$, where $N_b = \left[N^{\delta}\right]$, and set the remaining middle elements of $\boldsymbol{b}$ to zero. The resulting population correlation matrix $\boldsymbol{R}$ is defined by

$$\boldsymbol{R} = \boldsymbol{I}_N + \boldsymbol{b}\boldsymbol{b}' - \check{\boldsymbol{B}}^2,$$

where $\check{\boldsymbol{B}}$=$diag\,(\boldsymbol{b})$ is of $N \times N$ dimension. 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 (2013), and set $\delta = 0.25$.

Further, we impose heteroskedasticity on the main diagonal of $\boldsymbol{\Sigma}$ by generating $\boldsymbol{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})$ such that $\sigma_{ii} \sim IID\ \left(1/2 + \chi^2(2)/4\right),\ i = 1, 2, ..., N$ as in Experiment B. Then, $\boldsymbol{\Sigma}$ becomes $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{R}\boldsymbol{D}^{1/2}$. We obtain the Cholesky factor of $\boldsymbol{R}$, $\boldsymbol{P}$, and generate $\boldsymbol{Q} = \boldsymbol{D}^{1/2}\boldsymbol{P}$ which is then used in the data generating process

$$\boldsymbol{x}_t^{(r)} = \boldsymbol{Q}\boldsymbol{\varepsilon}_t^{(r)},\ t = 1, ..., T. \tag{29}$$

6.1 Alternative estimators and evaluation metrics

We obtain estimates of $\boldsymbol{\Sigma}$ for all four experiments, using the alternative regularisation techniques described above. More specifically, we compute the following estimates:

MT_R: thresholding based on the MT approach applied row-wise to the sample correlation matrix $(\widetilde{\boldsymbol{\Sigma}}_{MT_R})$

MT_F: thresholding based on the MT approach applied to all distinct non-diagonal elements of the sample correlation matrix $(\widetilde{\boldsymbol{\Sigma}}_{MT_F})$

$BL_{\hat{C}}$: 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 $(\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}})$

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

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

$BL_{\hat{C}^*}$: BL thresholding using the FML 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 $(\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}})$

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

The first five estimates relate to the original thresholding techniques. With regard to the next four estimates, for the MT method we apply additional shrinkage and for both BL and CL thresholding procedures we further impose the FLM extension which ensures positive definiteness of the estimated matrices. The adjusted thresholding methods and shrinkage approaches are evaluated predominantly for comparison with the inverse covariance matrices.

Where regularisation is performed on the correlation matrix we reconstruct the corresponding covariance matrix in line with (8). Across all experiments we compute the spectral norm of the deviations of each of the regularised covariance matrices from their respective true $\mathbf{\Sigma}$:

$$\left\|\boldsymbol{A}_{\mathring{\mathbf{\Sigma}}}\right\| = \left\|\mathbf{\Sigma} - \mathring{\mathbf{\Sigma}}\right\|, \tag{30}$$

for $\mathring{\mathbf{\Sigma}} = \{\widetilde{\mathbf{\Sigma}}_{MT_R}, \widetilde{\mathbf{\Sigma}}_{MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$, where $\hat{C}$ and $\hat{C}^*$ are constants evaluated through cross-validation - over the full grid and a reduced grid suggested by Fan, Liao and Mincheva (2013) (see Appendix B for details). We also evaluate the Frobenius norm of the difference given in (30), denoted by $\|.\|_F$. With regard to the performance of the inverse covariance matrices we evaluate

$$\left\|\boldsymbol{B}_{\mathring{\mathbf{\Sigma}}^{-1}}\right\| = \left\|\mathbf{\Sigma}^{-1} - \mathring{\mathbf{\Sigma}}^{-1}\right\|, \tag{31}$$

for $\mathring{\mathbf{\Sigma}}^{-1} = \{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{R}}}\}$, where all estimates are positive definite. Again, we also calculate the Frobenius norm of the difference given in (31).

Note that as long as $\mathbf{\Sigma}$ is well defined (implying that $\left\|\mathbf{\Sigma}^{-1}\right\| = O(1)$) then for the inverses it holds that:

$$\begin{aligned} \left\|\mathbf{\Sigma}^{-1} - \mathring{\mathbf{\Sigma}}^{-1}\right\| &= \left\|\mathbf{\Sigma}^{-1}\left(\mathring{\mathbf{\Sigma}} - \mathbf{\Sigma}\right)\mathring{\mathbf{\Sigma}}^{-1}\right\| \\ &\leq \left\|\mathbf{\Sigma}^{-1}\right\| \left\|\mathring{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\| \left\|\mathring{\mathbf{\Sigma}}^{-1}\right\|. \end{aligned}$$

The condition $\left\|\mathbf{\Sigma}^{-1}\right\| = O(1)$ is satisfied for all experiments with the exception of experiment C, for which the population covariance matrix is not invertible.

We report the averages of the above norms over $R = 500$ replications, except for the BL and CL cross-validation procedures. For these procedures computations take a very long time to complete and we were forced to use a lower number of replications and a cursor grid structure. Specifically for the experiments with $N = 400$, calculations can take up to 12 weeks so we set the grid increments to 4 and reduced the number of replications to $R = 100$ in this case. The latter is in line with the BL and CL simulation specifications.[10]

Finally, we assess 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 in (14) and (15), respectively. These are only implemented for experiments C and D. Experiments A and B refer to approximately sparse matrix settings, implying the absence of zero elements in the true covariance matrix. Also, TPR and FPR are not applicable to shrinkage techniques.

6.2 Robustness analysis

In order to assess the robustness of our multiple testing (MT) and shrinkage methodologies we also conduct the following experiments:

1. We allow for departures from normality for the errors $\varepsilon_{it}^{(r)}$ in experiments A-D. Therefore, in each case we also generate $\varepsilon_{it}^{(r)} \sim IID((\chi^2(2) - 2)/4)$, for $i = 1, 2, ..., N$ and $r = 1, ..., R$ and repeat the steps in (26) and (27). We evaluate our results using the sample covariance matrix.

2. We consider a more complex setting where $\boldsymbol{x}_t^{(r)}$ represents a vector of error terms in a regression equation. We set $u_{it}^{(r)} = x_{it}^{(r)}$, for $i = 1, 2, ..., N$, $t = 1, 2, ..., T$ for notational convenience, where $u_{it}^{(r)}$ are constructed as in experiments A-D. Then for each replication r, we generate

$$y_{it}^{(r)} = \delta_i + \gamma_i z_{it}^{(r)} + u_{it}^{(r)}, \text{ for } i = 1, 2, ..., N, \ t = 1, 2, ..., T, \tag{32}$$

[10] Standard deviations for all estimates are available upon request.

where $\delta_i \sim IIDN(1,1)$, and

$$z_{it}^{(r)} = \zeta_i z_{i,t-1}^{(r)} + \sqrt{1-\zeta_i^2}\nu_{it}^{(r)}, \text{ for } i = 1,2,....,N, \ t = -49,...,0,1,...,T,$$

with $z_{i,-50} = 0$, and $\nu_{it} \sim IIDN(0,1)$. We discard the first 50 observations. The observed regressors, $z_{it}^{(r)}$, are therefore strictly exogenous and serially correlated, and could possibly also be cross-sectionally dependent. We set $\zeta_i = 0.9$. Further we allow for slope heterogeneity by generating $\gamma_i \sim IIDN(1,1)$ for $i = 1,2,....,N$. In this case, the multiple testing approach is corrected for the degrees of freedom. Hence, as in (9) $\sqrt{T}$ is replaced by $\sqrt{T-m}$, where m is equal to the number of regressors in (32) including the intercept.

6.3 Robustness of the MT procedures to the choice of the p-value

First, we investigate the robustness of our MT estimators to different levels of significance, p, used in the derivation of the theoretical threshold value, b_N, defined by (7). We experimented with $p = 5\%$ and 10%. The spectral and Frobenius norms (averaged over 500 replications) for all the four experiments are summarised in Table 1, and clearly show that the choice of p is of secondary importance for the performance of the MT type estimators of sample covariance or correlation matrices. There are minor differences in the average spectral and Frobenius norms for $MT_R(0.05)$ (or $MT_F(0.05)$) and $MT_R(0.10)$ (or $MT_F(0.10)$) for all N and T combinations and for all covariance matrix setups considered. As we noted earlier (see also Section 3.1), the multiple testing procedure applied to all distinct non-diagonal elements of $\hat{\boldsymbol{R}}$ (namely the MT_F estimator) is expected to have a faster rate of convergence to $\boldsymbol{R}$, at least when Bonferroni critical values are used. However, in small samples multiple testing by row (namely the MT_R estimator) appears to perform marginally better in most cases, although the differences between the MT_R and MT_F estimators diminish in most cases as T and N increase, with the MT_F version outperforming the row-wise version in some cases. See for example experiment D for $T = 100$. Also, as to be expected, the performance of the estimators (as measured by the norms) deteriorates with N for a given T, and improves with T for a given N. These results hold across all four experiments (A-D).

6.4 Comparative results

The average norm results for experiments A-D across the different regularisation estimators are summarised in Tables 2-5. In view of the discussion of the previous section, we provide results for the MT estimators using the Bonferroni critical values only at the 5% significance level. In all cases the top panel shows comparative results for the different regularisation estimators. The middle panel presents results for the estimated inverse matrices (when such inverses exist). Finally, the bottom panel gives the results for the shrinkage coefficients used in the shrinkage approaches that we consider. Note that in Table 4 the middle panel has been excluded because the population covariance matrix, $\boldsymbol{\Sigma}$, is non-invertible.

Starting with experiment A and focusing initially on the top panel of Table 2, the results show that multiple testing and thresholding in general ourperform the shrinkage technique under both norm specifications and especially for larger values of N. While not surprisingly the performance of all estimators improves as T increases from 60 to 100, the MT and other thresholding procedures continue to outperform the shrinkage estimators. For small N, MT_R, MT_F, $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$ behave similarly, however as N increases MT_R and in most cases MT_F outperform $BL_{\hat{C}}$ and CL_2. In general, $CL_{\hat{C}}$ performs better than MT_F though the difference between the two diminishes for larger values of N. When the positive definite condition is imposed the performance of all estimators deteriorates. However, $S\text{-}MT_R$ and $S\text{-}MT_F$ perform favourably relative to $BL_{\hat{C}^*}$ and $CL_{\hat{C}^*}$ across all (N,T) combinations. Finally, adaptive thresholding ($CL_{\hat{C}}$ and $CL_{\hat{C}^*}$) outperforms universal thresholding ($BL_{\hat{C}}$ and $BL_{\hat{C}^*}$), which is to be expected given the heteroskedasticity present in the data. Also, in line with results reported in Cai and Liu (2011), the CL procedure that uses the

theoretical thresholding parameter of 2 (CL_2) performs poorer than its cross-validated equivalent ($CL_{\hat{C}}$).

Moving on to the middle panel of Table 2, we find that the inverse covariance estimators S-MT_R and S-MT_F perform much better than $BL_{\hat{C}^*}$ and $CL_{\hat{C}^*}$. In fact, the average spectral norm for $CL_{\hat{C}^*}$ includes some sizeable outliers, especially for small N. Still, their more reliable Frobenius norm estimates are higher than those of shrinkage applied to the multiple testing estimators. Furthermore, while $LW_{\hat{\Sigma}}$ outperforms both S-MT_R and S-MT_F for $N = \{30, 100\}$, as N rises to 200 and 400 shrinkage applied to the MT estimators appears to perform better. Finally, of all estimators considered, shrinkage on the sample correlation matrix $LW_{\hat{R}}$ produces the lowest norm values across the N, T combinations. Interestingly, the shrinkage parameters of the bottom panel of Table 2 show that $LW_{\hat{\Sigma}}$ imposes a progressively lower weight on $\hat{\mathbf{\Sigma}}$ as N increases, even more so for smaller T. On the other hand, S-MT_R, S-MT_F and $LW_{\hat{R}}$ place comparatively more balanced weights on $\boldsymbol{I}_N$ and $\hat{\boldsymbol{R}}$ in this case across the range of (N, T) combinations.

Results for experiments B, C and D are summarized in Tables 3-5. On the whole, the results are qualitatively similar to those of experiment A, although the average value of the norms are lower, particularly for experiments B and D. Also, MT_R and MT_F now outperform both of the threshold cross-validated estimators $BL_{\hat{C}}$ and $CL_{\hat{C}}$. With regard to the inverse covariance matrix estimators, again $BL_{\hat{C}^*}$ and $CL_{\hat{C}^*}$ suffer from outlier realisations especially for smaller values of N. Further, $LW_{\hat{R}}$ and $LW_{\hat{\Sigma}}$ perform similarly for small N but as the cross section dimension rises $LW_{\hat{\Sigma}}$ clearly outperforms, especially in the case of experiment D. Overall, S-MT outperforms the other inverse covariance estimators across experiments. The results also clearly show adaptive thresholding to be superior to universal thresholding. Finally, although $LW_{\hat{\Sigma}}$ is computationally attractive as compared to the cross-validation based thresholding approaches, its performance still falls short of the equally computationally appealing MT and S-MT procedures. The $LW_{\hat{\Sigma}}$ estimator also has the additional disadvantage that it tends to shrink the sample covariance matrix excessively towards the identity matrix.

Table 6 presents results for support recovery of $\mathbf{\Sigma}$ using the multiple testing and thresholding approaches with no adjustments. Superiority of MT_R and MT_F over $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$ is again established when comparing the true positive rates (TPR) of the estimators (FPR are uniformly close to zero in all cases). As T rises the TPRs improve while as N increases they decrease, as expected. The only exception is $BL_{\hat{C}}$ in experiment D, which shows improvement from $N = 30$ to $N = 100$ for both values of T (60 and 100). TPRs are higher for experiment D, since for this experiment we explicitly control for the number of non-zero elements in $\mathbf{\Sigma}$, and ensure that conditions of Theorem 2 are met.

We next turn to the results obtained from the robustness analysis outlined in Section 6.2 applied to experiments A-D. Evaluating the estimated covariance matrices based on non-normal errors is of particular interest. In this case, a deterioration in the values of the average spectral and Frobenius norms is observed across all estimators and experiments. This is not surprising as most of these methods are based on the assumption of normality of the underlying data. However, the MT and S-MT procedures still outperform the remaining estimators in most instances. To see this we measured the relative performance of the considered key regularisation estimators against our MT and S-MT estimators across all (N, T) combinations and experiments. Specifically, we compare MT_F, $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$ to our preferred MT_R estimator, where no adjustments are made to these estimators, and S-MT_F, $BL_{\hat{C}^*}$, $CL_{\hat{C}^*}$ and $LW_{\hat{\Sigma}}$ (as well as $LW_{\hat{R}}$ for the inverse covariance matrices) to our adjusted S-MT_R estimator, where all these estimators are positive definite. Tables 7-10 show the relative values of the average spectral and Frobenius norms for all these estimators. As can be seen these values are predominantly greater than the benchmark of unity attached to MT_R and S-MT_R respectively. Values greater than one indicate that the corresponding estimators are underperforming relative to the preferred estimators. Interestingly, MT_F outperforms MT_R at times depending on the experiment, which supports the theoretical results of Section 3.1. Similar conclusions can be drawn by re-evaluating the support recovery of $\mathbf{\Sigma}$ following the original Table

6. By comparison, the TPR values for all estimators under non-normal errors shown in Table 11 are lower than their normal counterparts. However, the loss in support recovery is significantly less pronounced for MT_R and MT_F than for $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$, especially for experiment D and for large N. As before, support recovery improves in all cases when T increases. Finally, results based on the residuals from regression (32) in general are similar to the main results of Tables 2-5. The tables for this case are therefore relegated to the supplementary appendix.

Overall, both our proposed multiple testing (MT) and shrinkage on multiple testing (S-MT) estimators prove to be robust to the specification of the underlying covariance matrix $\mathbf{\Sigma}$. If the inverse covariance matrix is of interest S-MT and $LW_{\hat{R}}$ are more appropriate, while MT gives better covariance matrix estimates when positive definiteness is not required. Also, MT is robust to the choice of the family wise significance level, p, used in the calculation of b_N. Moreover, S-MT yields covariance matrix estimates that are closer to the sample correlation matrix as compared to the widely used LW shrinkage approach. No clear ordering emerges when S-MT and $LW_{\hat{R}}$ are compared, rather the outcome depends on the true covariance matrix under consideration.

7 Concluding Remarks

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

A novel regularisation estimator (MT) is proposed that uses insights from the multiple testing literature to enhance the support of the true covariance matrix. It is applied to the sample correlation matrix thus keeping the variance components of $\hat{\mathbf{\Sigma}}$ intact. It is shown that the resultant estimator has a convergence rate of $(m_N N/T)^{1/2}$ under the Frobenius norm, where m_N is bounded in N, which is comparable with the convergence rates established in the literature. Further, it is robust to random permutations of the underlying observations and it is computationally simple to implement. Multiple testing is also suitable for application to high frequency observations, rendering it robust to changes in the covariance matrix over time.

Monte Carlo simulation results provide support of the theoretical properties of our MT estimator. They show favourable performance of the proposed MT procedure (applied either by row or to the full matrix) compared with a number of key regularisation techniques in the literature. They further highlight the robustness of the MT estimator to different covariance matrix settings and deviations from the main assumptions of the underlying theory.

If the inverse of the covariance matrix is of interest, since traditional thresholding approaches including multiple testing do not necessarily produce a positive definite matrix, we recommend additional shrinkage of our regularised multiple testing estimator or a shrinkage estimator applied to the sample correlation matrix itself.

The problems of invertibility and robustness of estimated large covariance matrices to time variations of the underlying variances and covariances are topics that continue to concern the research community and are interesting areas for future study.

Table 1: Performance of Multipe Testing (MT) estimator under the Spectral and Frobenius norms of error matrices ($\widetilde{\mathbf{\Sigma}}_{MT_R} - \mathbf{\Sigma}$) and ($\widetilde{\mathbf{\Sigma}}_{MT_F} - \mathbf{\Sigma}$) at 5% and 10% significance levels

Normally distributed errors. Averages over 500 replications

	Experiment A							
	$N = 30$		$N = 100$		$N = 200$		$N = 400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	$T = 60$							
$MT_R(0.05)$	4.461	7.972	5.699	16.281	6.116	23.994	6.483	35.138
$MT_F(0.05)$	5.217	9.164	6.621	19.076	7.165	28.491	7.737	42.609
$MT_R(0.10)$	4.278	7.706	5.505	15.761	5.941	23.301	6.329	34.190
$MT_F(0.10)$	5.035	8.874	6.472	18.630	6.996	27.832	7.540	41.463
	$T = 100$							
$MT_R(0.05)$	3.529	6.313	4.550	12.888	4.946	19.016	5.343	27.884
$MT_F(0.05)$	4.058	7.190	5.392	15.283	5.894	23.026	6.297	34.347
$MT_R(0.10)$	3.407	6.101	4.403	12.513	4.795	18.500	5.195	27.177
$MT_F(0.10)$	3.920	6.967	5.246	14.840	5.759	22.398	6.194	33.539
	Experiment B							
	$T = 60$							
$MT_R(0.05)$	1.421	3.339	1.634	6.477	2.012	10.094	2.170	14.753
$MT_F(0.05)$	1.558	3.929	1.755	7.407	2.098	11.187	2.243	15.997
$MT_R(0.10)$	1.379	3.183	1.613	6.243	1.988	9.796	2.153	14.401
$MT_F(0.10)$	1.526	3.795	1.731	7.313	2.095	11.131	2.242	15.964
	$T = 100$							
$MT_R(0.05)$	1.030	2.306	1.284	4.561	1.618	7.258	1.750	10.836
$MT_F(0.05)$	1.225	2.872	1.523	6.163	1.939	9.896	2.046	14.659
$MT_R(0.10)$	1.000	2.238	1.251	4.365	1.565	6.916	1.712	10.332
$MT_F(0.10)$	1.173	2.703	1.495	5.904	1.906	9.590	2.022	14.335
	Experiment C							
	$T = 60$							
$MT_R(0.05)$	2.211	4.097	3.311	8.506	3.873	12.589	4.302	18.414
$MT_F(0.05)$	2.404	4.352	3.955	9.664	4.730	14.720	5.335	22.171
$MT_R(0.10)$	2.220	4.149	3.263	8.550	3.799	12.618	4.218	18.428
$MT_F(0.10)$	2.339	4.258	3.818	9.407	4.580	14.331	5.184	21.601
	$T = 100$							
$MT_R(0.05)$	1.686	3.123	2.523	6.453	2.879	9.496	3.223	13.892
$MT_F(0.05)$	1.739	3.214	2.841	7.093	3.310	10.726	3.775	16.116
$MT_R(0.10)$	1.724	3.192	2.516	6.560	2.859	9.614	3.194	14.030
$MT_F(0.10)$	1.707	3.160	2.766	6.931	3.230	10.474	3.689	15.750
	Experiment D							
	$T = 60$							
$MT_R(0.05)$	0.656	1.196	1.062	2.186	0.998	2.980	1.401	4.344
$MT_F(0.05)$	0.729	1.245	1.514	2.513	1.487	3.195	2.416	4.818
$MT_R(0.10)$	0.677	1.258	1.056	2.294	1.015	3.164	1.357	4.568
$MT_F(0.10)$	0.687	1.201	1.395	2.398	1.366	3.106	2.248	4.705
	$T = 100$							
$MT_R(0.05)$	0.488	0.902	0.763	1.653	0.730	2.310	0.920	3.331
$MT_F(0.05)$	0.468	0.852	0.798	1.589	0.743	2.154	1.120	3.224
$MT_R(0.10)$	0.510	0.959	0.786	1.772	0.767	2.492	0.951	3.608
$MT_F(0.10)$	0.467	0.853	0.780	1.574	0.723	2.141	1.054	3.178

Notes: For an $N \times N$ matrix $\boldsymbol{A} = (a_{ij})$, the spectral norm is given by: $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$, where $\lambda_{\max}(\boldsymbol{A})$ is its largest eigenvalue. For an $N \times N$ matrix $\boldsymbol{A} = (a_{ij})$, the Frobenius norm is given by: $\|\boldsymbol{A}\|_F = \sqrt{tr(\boldsymbol{A}'\boldsymbol{A})}$. MT_R=Multiple Testing by row, MT_F= Multiple Testing on full $\hat{\mathbf{R}}$ matrix (applied to all off-diagonal elements). Both estimators use the Bonferroni method at the 0.05 and 0.10 significance level.

Table 2: Comparison of regularisation estimators applied to sparse covariance matrix $\hat{\mathbf{\Sigma}}$

Experiment A - normally distributed errors. Averages over 500 replications.

	$N = 30$		$N = 100$		$N = 200$		$N = 400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}$)							
	$T = 60$							
MT_R	4.461	7.972	5.699	16.281	6.116	23.994	6.483	35.138
MT_F	5.217	9.164	6.621	19.076	7.165	28.491	7.737	42.609
$BL_{\hat{C}}$	4.284	7.497	5.648	16.028	6.384	24.347	6.963	36.414
CL_2	5.566	9.705	7.537	21.611	8.263	33.149	8.729	49.729
$CL_{\hat{C}}$	4.088	7.339	5.228	15.610	5.785	23.612	6.274	35.382
$S\text{-}MT_R$	5.827	8.801	7.176	18.501	7.574	27.585	7.883	40.666
$S\text{-}MT_F$	6.473	9.939	7.778	20.880	8.174	31.156	8.506	46.083
$BL_{\hat{C}^*}$	8.543	14.503	9.142	27.137	9.223	38.570	9.267	54.679
$CL_{\hat{C}^*}$	8.512	14.446	9.130	27.098	9.220	38.555	9.265	54.668
$LW_{\hat{\Sigma}}$	4.221	7.039	7.002	18.704	8.206	30.743	8.890	48.020
	$T = 100$							
MT_R	3.529	6.313	4.550	12.888	4.946	19.016	5.343	27.884
MT_F	4.058	7.190	5.392	15.283	5.894	23.026	6.297	34.347
$BL_{\hat{C}}$	3.336	5.829	4.383	12.439	4.893	18.775	5.496	28.182
CL_2	4.140	7.336	5.695	16.169	6.323	24.760	6.931	37.571
$CL_{\hat{C}}$	3.247	5.757	4.144	12.227	4.585	18.407	5.000	27.459
$S\text{-}MT_R$	4.837	7.149	6.208	15.409	6.678	23.282	7.067	34.784
$S\text{-}MT_F$	5.497	8.161	6.890	17.668	7.369	26.781	7.737	40.095
$BL_{\hat{C}^*}$	8.527	14.450	9.114	27.043	9.187	38.438	9.228	54.503
$CL_{\hat{C}^*}$	8.434	14.299	9.095	26.980	9.181	38.409	9.228	54.491
$LW_{\hat{\Sigma}}$	3.393	5.683	6.039	16.076	7.503	27.550	8.489	44.737
	Error matrices ($\mathbf{\Sigma}^{-1} - \hat{\mathbf{\Sigma}}^{-1}$)							
	$T = 60$							
$S\text{-}MT_R$	4.090	5.255	4.756	10.265	4.995	15.033	5.174	21.863
$S\text{-}MT_F$	4.087	5.007	4.452	9.993	4.559	15.132	4.718	22.756
$BL_{\hat{C}^*}$	5.683	7.348	5.868	13.663	5.941	19.403	6.002	27.487
$CL_{\hat{C}^*}$	2.5E+02	8.723	1.2E+02	14.302	6.298	19.404	7.520	27.514
$LW_{\hat{\Sigma}}$	2.523	4.187	4.038	10.674	4.666	16.953	5.074	25.610
$LW_{\hat{R}}$	2.216	3.920	3.421	9.028	3.818	13.865	3.995	20.560
	$T = 100$							
$S\text{-}MT_R$	3.547	5.076	4.311	10.071	4.615	14.823	4.864	21.651
$S\text{-}MT_F$	4.053	5.190	4.734	9.985	4.969	14.635	5.135	21.425
$BL_{\hat{C}^*}$	29.820	7.590	5.822	13.731	5.879	19.496	5.925	27.623
$CL_{\hat{C}^*}$	7.1E+03	13.561	6.9E+03	18.230	32.454	19.744	4.1E+02	29.356
$LW_{\hat{\Sigma}}$	1.927	3.480	3.511	9.463	4.285	15.764	4.846	24.669
$LW_{\hat{R}}$	1.712	3.368	3.042	8.254	3.601	13.124	3.896	19.965
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T = 60$							
$S\text{-}MT_R$	0.392	0.608	0.474	0.526	0.513	0.487	0.545	0.455
$S\text{-}MT_F$	0.414	0.586	0.494	0.506	0.534	0.466	0.564	0.436
$LW_{\hat{\Sigma}}$	0.443	0.770	0.898	0.534	1.202	0.377	1.458	0.244
$LW_{\hat{R}}$	0.157	0.843	0.306	0.694	0.377	0.623	0.425	0.575
	$T = 100$							
$S\text{-}MT_R$	0.315	0.685	0.401	0.599	0.445	0.555	0.484	0.516
$S\text{-}MT_F$	0.355	0.645	0.435	0.565	0.480	0.520	0.522	0.478
$LW_{\hat{\Sigma}}$	0.298	0.846	0.678	0.650	0.988	0.491	1.296	0.333
$LW_{\hat{R}}$	0.109	0.891	0.248	0.752	0.331	0.669	0.396	0.604

Notes: For the BL and CL methods, for $N = 400$ and $T = 60, 100$ we set the number of replications to 100 and the grid increment to 4. Here, $\hat{\mathbf{\Sigma}} = \{\widetilde{\mathbf{\Sigma}}_{MT_R}, \widetilde{\mathbf{\Sigma}}_{MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$ and $\hat{\mathbf{\Sigma}}^{-1} = \{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{R}}}\}$. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. $S\text{-}MT_R$=Shrinkage on MT by row; $S\text{-}MT_F$=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 3: Comparison of regularisation estimators applied to sparse covariance matrix $\hat{\mathbf{\Sigma}}$

Experiment B - normally distributed errors. Averages over 500 replications.

	$N = 30$		$N = 100$		$N = 200$		$N = 400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}$)							
	$T = 60$							
MT_R	1.421	3.339	1.634	6.477	2.012	10.094	2.170	14.753
MT_F	1.558	3.929	1.755	7.407	2.098	11.187	2.243	15.997
$BL_{\hat{C}}$	1.615	3.941	1.983	7.625	2.106	11.250	2.277	16.048
CL_2	1.571	3.974	1.894	7.505	2.093	11.182	2.242	15.986
$CL_{\hat{C}}$	1.436	3.361	1.900	7.214	2.089	11.129	2.239	15.929
$S\text{-}MT_R$	1.438	3.329	1.657	6.389	2.003	9.860	2.144	14.428
$S\text{-}MT_F$	1.560	3.878	1.758	7.360	2.093	11.136	2.238	15.964
$BL_{\hat{C}^*}$	1.599	4.095	1.978	7.639	2.103	11.251	2.267	16.060
$CL_{\hat{C}^*}$	1.461	3.568	1.977	7.476	2.093	11.191	2.252	16.010
$LW_{\hat{\Sigma}}$	1.621	3.576	2.643	7.559	2.543	11.829	3.308	17.824
	$T = 100$							
MT_R	1.030	2.306	1.284	4.561	1.618	7.258	1.750	10.836
MT_F	1.225	2.872	1.523	6.163	1.939	9.896	2.046	14.659
$BL_{\hat{C}}$	1.214	2.705	1.574	5.843	1.911	9.915	2.145	15.584
CL_2	1.249	2.961	1.553	6.401	1.970	10.214	2.086	15.020
$CL_{\hat{C}}$	1.034	2.334	1.295	4.587	1.628	7.423	1.860	11.911
$S\text{-}MT_R$	1.142	2.535	1.409	4.966	1.702	7.633	1.814	11.216
$S\text{-}MT_F$	1.293	2.971	1.575	6.199	1.953	9.813	2.053	14.531
$BL_{\hat{C}^*}$	1.193	2.718	1.543	6.145	1.919	10.161	2.148	15.649
$CL_{\hat{C}^*}$	1.035	2.344	1.331	4.836	1.756	8.282	2.040	14.228
$LW_{\hat{\Sigma}}$	1.405	3.071	2.402	7.012	2.429	11.291	3.205	17.301
	Error matrices ($\mathbf{\Sigma}^{-1} - \hat{\mathbf{\Sigma}}^{-1}$)							
	$T = 60$							
$S\text{-}MT_R$	1.966	3.377	2.652	6.892	3.259	10.149	3.691	14.938
$S\text{-}MT_F$	2.584	3.926	3.157	8.029	3.723	11.584	4.078	16.669
$BL_{\hat{C}^*}$	1.4E+04	19.315	58.881	9.377	3.9E+03	15.321	14.009	17.017
$CL_{\hat{C}^*}$	2.1E+04	33.982	2.4E+04	23.651	44.094	12.593	16.774	17.064
$LW_{\hat{\Sigma}}$	2.971	3.874	3.715	8.438	4.932	12.850	5.832	18.870
$LW_{\hat{R}}$	1.969	3.539	4.809	8.773	6.958	13.956	8.767	20.919
	$T = 100$							
$S\text{-}MT_R$	1.296	2.650	1.891	5.438	2.436	8.036	2.854	11.934
$S\text{-}MT_F$	1.777	3.095	2.636	6.871	3.274	10.352	3.764	15.437
$BL_{\hat{C}^*}$	5.0E+03	23.048	4.2E+03	24.145	2.7E+04	30.297	43.825	17.318
$CL_{\hat{C}^*}$	3.0E+05	65.501	1.9E+05	1.0E+02	2.2E+07	3.6E+02	2.2E+03	31.662
$LW_{\hat{\Sigma}}$	2.338	3.374	3.406	7.993	4.735	12.515	5.744	18.663
$LW_{\hat{R}}$	1.333	2.982	2.805	7.349	4.381	12.101	5.719	18.967
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T = 60$							
$S\text{-}MT_R$	0.383	0.617	0.402	0.598	0.387	0.613	0.378	0.622
$S\text{-}MT_F$	0.329	0.671	0.327	0.673	0.303	0.697	0.312	0.688
$LW_{\hat{\Sigma}}$	0.591	0.517	0.871	0.257	1.011	0.162	1.086	0.105
$LW_{\hat{R}}$	0.341	0.659	0.436	0.564	0.461	0.539	0.474	0.526
	$T = 100$							
$S\text{-}MT_R$	1.296	2.650	0.417	0.583	0.415	0.585	0.408	0.592
$S\text{-}MT_F$	1.777	3.095	0.380	0.620	0.355	0.645	0.331	0.669
$LW_{\hat{\Sigma}}$	0.449	0.635	0.770	0.348	0.946	0.221	1.055	0.137
$LW_{\hat{R}}$	0.288	0.712	0.412	0.588	0.450	0.550	0.470	0.530

Notes: For the BL and CL methods, for $N = 400$ and $T = 60, 100$ we set the number of replications to 100 and the grid increment to 4. Here, $\hat{\mathbf{\Sigma}} = \{\widetilde{\mathbf{\Sigma}}_{MT_R}, \widetilde{\mathbf{\Sigma}}_{MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$ and $\hat{\mathbf{\Sigma}}^{-1} = \{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{R}}}\}$. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. $S\text{-}MT_R$=Shrinkage on MT by row; $S\text{-}MT_F$=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 4: Comparison of regularisation estimators applied to sparse covariance matrix $\hat{\mathbf{\Sigma}}$

Experiment C - normally distributed errors. Averages over 500 replications.

	$N = 30$		$N = 100$		$N = 200$		$N = 400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\mathbf{\Sigma} - \mathring{\mathbf{\Sigma}}$)							
	$T = 60$							
MT_R	2.211	4.097	3.311	8.506	3.873	12.589	4.302	18.414
MT_F	2.404	4.352	3.955	9.664	4.730	14.720	5.335	22.171
$BL_{\hat{C}}$	7.040	8.795	8.755	17.234	8.961	24.701	9.031	35.161
CL_2	2.661	4.641	5.138	11.183	6.477	17.786	7.468	27.640
$CL_{\hat{C}}$	2.381	4.394	3.574	9.404	4.316	14.278	5.024	21.375
$S\text{-}MT_R$	3.515	4.953	6.017	11.657	6.646	17.495	7.046	25.823
$S\text{-}MT_F$	4.171	5.520	6.603	12.652	7.213	18.988	7.609	28.107
$BL_{\hat{C}^*}$	7.091	8.804	8.755	17.233	8.961	24.701	9.031	35.172
$CL_{\hat{C}^*}$	7.059	8.769	8.747	17.207	8.958	24.671	9.030	35.131
$LW_{\hat{\Sigma}}$	3.532	7.675	5.853	18.451	6.707	28.593	7.182	42.720
	$T = 100$							
MT_R	1.686	3.123	2.523	6.453	2.879	9.496	3.223	13.892
MT_F	1.739	3.214	2.841	7.093	3.310	10.726	3.775	16.116
$BL_{\hat{C}}$	5.118	7.511	8.747	16.895	8.946	24.243	9.014	34.528
CL_2	1.781	3.279	3.084	7.534	3.786	11.748	4.585	18.160
$CL_{\hat{C}}$	1.738	3.230	2.634	6.816	3.002	10.180	3.395	15.206
$S\text{-}MT_R$	2.525	3.685	5.107	9.643	5.744	14.774	6.224	22.272
$S\text{-}MT_F$	3.088	4.152	5.775	10.764	6.381	16.366	6.823	24.506
$BL_{\hat{C}^*}$	7.082	8.609	8.747	16.898	8.946	24.241	9.014	34.534
$CL_{\hat{C}^*}$	7.038	8.563	8.721	16.852	8.937	24.215	9.011	34.504
$LW_{\hat{\Sigma}}$	2.989	6.497	5.246	16.722	6.267	26.843	6.935	41.115
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T = 60$							
$S\text{-}MT_R$	0.400	0.600	0.564	0.436	0.604	0.396	0.634	0.366
$S\text{-}MT_F$	0.471	0.529	0.595	0.405	0.628	0.372	0.655	0.345
$LW_{\hat{\Sigma}}$	1.015	0.586	1.633	0.335	1.925	0.217	2.124	0.136
	$T = 100$							
$S\text{-}MT_R$	0.277	0.723	0.483	0.517	0.533	0.467	0.572	0.428
$S\text{-}MT_F$	0.351	0.649	0.543	0.457	0.585	0.415	0.619	0.381
$LW_{\hat{\Sigma}}$	0.744	0.700	1.373	0.445	1.741	0.297	2.024	0.183

Notes: For the BL and CL methods, for $N = 400$ and $T = 60, 100$ we set the number of replications to 100 and the grid increment to 4. Here, $\mathring{\mathbf{\Sigma}} = \{\widetilde{\mathbf{\Sigma}}_{MT_R}, \widetilde{\mathbf{\Sigma}}_{MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. The population covariance matrix $\mathbf{\Sigma}$ does not have an inverse in this experiment hence results relating to matrix inverses are not provided. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. $S\text{-}MT_R$= Shrinkage on MT by row; $S\text{-}MT_F$=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 5: Comparison of regularisation estimators applied to sparse covariance matrix $\hat{\mathbf{\Sigma}}$

Experiment D - normally distributed errors. Averages over 500 replications.

	$N = 30$		$N = 100$		$N = 200$		$N = 400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}$)							
	$T = 60$							
MT_R	0.656	1.196	1.062	2.186	0.998	2.980	1.401	4.344
MT_F	0.729	1.245	1.514	2.513	1.487	3.195	2.416	4.818
$BL_{\hat{C}}$	1.436	1.931	2.635	3.512	2.735	3.985	3.722	5.566
CL_2	0.847	1.389	2.055	3.054	1.976	3.550	3.088	5.218
$CL_{\hat{C}}$	0.925	1.478	1.854	2.939	2.328	3.761	3.362	5.372
$S\text{-}MT_R$	0.783	1.309	1.540	2.472	1.304	3.042	1.964	4.436
$S\text{-}MT_F$	0.890	1.391	2.027	2.878	1.946	3.399	3.050	5.041
$BL_{\hat{C}^*}$	1.512	2.016	3.336	4.072	2.744	3.987	3.730	5.557
$CL_{\hat{C}^*}$	1.314	1.854	3.356	4.085	2.738	3.977	3.733	5.547
$LW_{\hat{\Sigma}}$	1.188	2.304	3.166	4.703	2.522	6.172	3.623	9.534
	$T = 100$							
MT_R	0.488	0.902	0.763	1.653	0.730	2.310	0.920	3.331
MT_F	0.468	0.852	0.798	1.589	0.743	2.154	1.120	3.224
$BL_{\hat{C}}$	0.879	1.308	1.237	2.120	2.544	3.508	3.526	4.909
CL_2	0.485	0.875	0.948	1.738	0.923	2.309	1.595	3.589
$CL_{\hat{C}}$	0.496	0.917	0.812	1.718	1.141	2.533	2.445	4.258
$S\text{-}MT_R$	0.647	1.056	1.415	2.093	1.083	2.421	1.364	3.409
$S\text{-}MT_F$	0.646	1.040	1.457	2.105	1.193	2.423	1.877	3.612
$BL_{\hat{C}^*}$	1.133	1.573	3.328	3.915	2.727	3.617	3.696	4.989
$CL_{\hat{C}^*}$	1.052	1.499	3.333	3.922	2.720	3.613	3.731	5.001
$LW_{\hat{\Sigma}}$	1.032	2.052	2.935	4.463	2.450	6.007	3.575	9.318
	Error matrices ($\mathbf{\Sigma}^{-1} - \hat{\mathbf{\Sigma}}^{-1}$)							
	$T = 60$							
$S\text{-}MT_R$	4.758	2.905	15.439	6.138	13.381	6.046	14.052	7.855
$S\text{-}MT_F$	5.283	3.031	17.860	6.501	16.941	6.540	18.114	8.513
$BL_{\hat{C}^*}$	7.1E+02	7.034	46.674	8.388	26.348	7.707	24.963	9.503
$CL_{\hat{C}^*}$	9.3E+04	21.119	29.780	8.096	34.349	7.834	45.816	9.851
$LW_{\hat{\Sigma}}$	12.420	4.558	31.907	8.771	31.988	9.478	31.854	12.568
$LW_{\hat{R}}$	5.187	4.452	15.736	12.584	15.080	19.470	18.160	30.113
	$T = 100$							
$S\text{-}MT_R$	4.529	2.683	15.394	5.865	12.790	5.363	11.037	6.435
$S\text{-}MT_F$	4.526	2.665	15.673	5.882	13.853	5.444	14.398	6.900
$BL_{\hat{C}^*}$	1.7E+04	19.022	2.7E+02	8.880	48.354	7.690	26.695	8.897
$CL_{\hat{C}^*}$	4.5E+02	6.177	8.1E+02	9.214	1.9E+02	8.419	40.085	9.033
$LW_{\hat{\Sigma}}$	10.861	4.240	30.981	8.611	31.783	9.400	31.841	12.526
$LW_{\hat{R}}$	4.850	3.720	16.168	10.239	14.347	16.032	13.104	26.403
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T = 60$							
$S\text{-}MT_R$	0.381	0.619	0.406	0.594	0.399	0.601	0.455	0.545
$S\text{-}MT_F$	0.424	0.576	0.496	0.504	0.540	0.460	0.614	0.386
$LW_{\hat{\Sigma}}$	0.579	0.375	0.735	0.180	0.842	0.091	0.871	0.067
$LW_{\hat{R}}$	0.423	0.577	0.467	0.533	0.483	0.517	0.485	0.515
	$T = 100$							
$S\text{-}MT_R$	0.352	0.648	0.394	0.606	0.364	0.636	0.335	0.665
$S\text{-}MT_F$	0.353	0.647	0.402	0.598	0.401	0.599	0.459	0.541
$LW_{\hat{\Sigma}}$	0.473	0.492	0.682	0.244	0.826	0.115	0.868	0.077
$LW_{\hat{R}}$	0.392	0.608	0.460	0.540	0.485	0.515	0.489	0.511

Notes: For the BL and CL methods, for $N = 400$ and $T = 60, 100$ we set the number of replications to 100 and the grid increment to 4. Here, $\hat{\mathbf{\Sigma}} = \{\widetilde{\mathbf{\Sigma}}_{MT_R}, \widetilde{\mathbf{\Sigma}}_{MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$ and $\hat{\mathbf{\Sigma}}^{-1} = \{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{R}}}\}$. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. $S\text{-}MT_R$=Shrinkage on MT by row; $S\text{-}MT_F$=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 6: Comparison of $\mathbf{\Sigma}$ support recovery produced by different thresholding estimators

Support recovery is measured by the True Positive Rate (TPR) and False Positive Rate (FPR)

Normally distributed errors. Averages over 500 replications.

	$N = 30$		$N = 100$		$N = 200$		$N = 400$	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
	Experiment C							
	$T = 60$							
MT_R	0.714	0.001	0.587	0.000	0.551	0.000	0.521	0.000
MT_F	0.619	0.000	0.455	0.000	0.402	0.000	0.357	0.000
$BL_{\hat{C}}$	0.013	0.002	0.000	0.000	0.000	0.000	0.000	0.000
CL_2	0.584	0.000	0.370	0.000	0.286	0.000	0.215	0.000
$CL_{\hat{C}}$	0.710	0.005	0.576	0.002	0.528	0.001	0.478	0.000
	$T = 100$							
MT_R	0.801	0.002	0.696	0.000	0.667	0.000	0.640	0.000
MT_F	0.735	0.000	0.596	0.000	0.553	0.000	0.514	0.000
$BL_{\hat{C}}$	0.324	0.048	0.000	0.000	0.000	0.000	0.000	0.000
CL_2	0.729	0.000	0.566	0.000	0.506	0.000	0.453	0.000
$CL_{\hat{C}}$	0.781	0.002	0.686	0.001	0.655	0.001	0.623	0.000
	Experiment D							
	$T = 60$							
MT_R	0.974	0.001	0.972	0.000	0.941	0.000	0.895	0.000
MT_F	0.869	0.000	0.832	0.000	0.649	0.000	0.468	0.000
$BL_{\hat{C}}$	0.187	0.001	0.325	0.000	0.009	0.000	0.006	0.000
CL_2	0.753	0.000	0.607	0.000	0.375	0.000	0.214	0.000
$CL_{\hat{C}}$	0.723	0.003	0.666	0.001	0.225	0.000	0.135	0.000
	$T = 100$							
MT_R	1.000	0.001	0.999	0.000	0.997	0.000	0.994	0.000
MT_F	0.993	0.000	0.986	0.000	0.969	0.000	0.915	0.000
$BL_{\hat{C}}$	0.686	0.002	0.852	0.001	0.101	0.000	0.051	0.000
CL_2	0.981	0.000	0.950	0.000	0.886	0.000	0.749	0.000
$CL_{\hat{C}}$	0.994	0.002	0.986	0.001	0.790	0.000	0.469	0.000

Notes: For the BL and CL methods, for $N = 400$ and $T = 60, 100$ we set the number of replications to 100 and the grid increment to 4.

MT_R=Multiple testing by row, MT_F=Multiple testing on the full $\hat{\mathbf{R}}$ matrix.

Both MT estimators use the Bonferroni method at the 5% significance level.

$BL_{\hat{C}}$=Bickel and Levina universal thresholding using a cross-validated parameter $\hat{C}$.

CL_2= Cai and Liu adaptive thresholding using the theoretical parameter of 2.

$CL_{\hat{C}}$= Cai and Liu adaptive thresholding using a cross-validated parameter $\hat{C}$.

Table 7: Relative performance of key regularisation estimators to the MT and S-MT estimators MT_F, $BL_{\hat{C}}$, CL_2, $CL_{\hat{C}}$ are compared to MT_R and S-MT_F, $BL_{\hat{C}^*}$, $CL_{\hat{C}^*}$, $LW_{\hat{\Sigma}}$ are compared to S-MT_R

Experiment A - non-normally distributed errors. Averages over 500 replications.

	$N=30$		$N=100$		$N=200$		$N=400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}$)							
	$T=60$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	1.122	1.121	1.134	1.152	1.142	1.172	1.138	1.192
$BL_{\hat{C}}$	0.998	0.952	1.095	1.035	1.149	1.081	1.245	1.203
CL_2	1.372	1.331	1.388	1.405	1.368	1.421	1.314	1.417
$CL_{\hat{C}}$	0.942	0.927	0.966	0.992	0.992	1.031	1.031	1.076
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	1.102	1.114	1.082	1.121	1.077	1.125	1.072	1.129
$BL_{\hat{C}^*}$	1.446	1.591	1.266	1.441	1.212	1.381	1.169	1.330
$CL_{\hat{C}^*}$	1.436	1.580	1.264	1.438	1.211	1.380	1.169	1.330
$LW_{\hat{\Sigma}}$	0.926	0.863	1.176	1.089	1.282	1.214	1.292	1.293
	$T=100$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	1.126	1.122	1.149	1.159	1.155	1.178	1.148	1.195
$BL_{\hat{C}}$	0.973	0.941	1.032	0.998	1.070	1.025	1.113	1.051
CL_2	1.320	1.305	1.443	1.436	1.466	1.498	1.460	1.546
$CL_{\hat{C}}$	0.938	0.917	0.945	0.968	0.962	0.996	0.984	1.028
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	1.129	1.128	1.103	1.135	1.095	1.136	1.088	1.137
$BL_{\hat{C}^*}$	1.728	1.944	1.451	1.719	1.360	1.620	1.293	1.541
$CL_{\hat{C}^*}$	1.710	1.923	1.448	1.714	1.358	1.618	1.293	1.540
$LW_{\hat{\Sigma}}$	0.960	0.896	1.279	1.187	1.445	1.369	1.521	1.514
	Error matrices ($\mathbf{\Sigma}^{-1} - \hat{\mathbf{\Sigma}}^{-1}$)							
	$T=60$							
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	0.997	0.962	0.960	0.983	0.952	1.011	0.967	1.039
$BL_{\hat{C}^*}$	2.4E+02	1.687	1.239	1.329	1.196	1.290	1.170	1.257
$CL_{\hat{C}^*}$	40.960	1.723	12.447	1.379	2.549	1.299	1.188	1.257
$LW_{\hat{\Sigma}}$	0.671	0.845	0.849	1.058	0.922	1.138	0.961	1.179
$LW_{\hat{R}}$	0.706	0.807	0.756	0.898	0.799	0.936	0.818	0.954
	$T=100$							
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	1.119	1.016	1.074	0.989	1.051	0.986	1.030	0.988
$BL_{\hat{C}^*}$	10.472	1.497	1.335	1.364	1.261	1.318	1.213	1.280
$CL_{\hat{C}^*}$	9.4E+03	3.341	3.5E+03	2.037	25.906	1.370	5.511	1.291
$LW_{\hat{\Sigma}}$	0.587	0.726	0.810	0.958	0.906	1.076	0.969	1.149
$LW_{\hat{R}}$	0.604	0.719	0.718	0.833	0.786	0.895	0.812	0.931
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T=60$							
S-MT_R	0.389	0.611	0.474	0.526	0.512	0.488	0.545	0.455
S-MT_F	0.409	0.591	0.487	0.513	0.526	0.474	0.555	0.445
$LW_{\hat{\Sigma}}$	0.503	0.738	0.926	0.520	1.215	0.370	1.461	0.243
$LW_{\hat{R}}$	0.156	0.844	0.304	0.696	0.375	0.625	0.424	0.576
	$T=100$							
S-MT_R	0.312	0.688	0.401	0.599	0.446	0.554	0.485	0.515
S-MT_F	0.351	0.649	0.433	0.567	0.477	0.523	0.518	0.482
$LW_{\hat{\Sigma}}$	0.341	0.824	0.701	0.639	1.002	0.484	1.301	0.330
$LW_{\hat{R}}$	0.108	0.892	0.246	0.754	0.329	0.671	0.395	0.605

Notes: For the BL and CL methods, for $N=400$ and $T=60,100$ we set the number of replications to 100 and the grid increment to 4. Here, $\hat{\mathbf{\Sigma}} = \{\widetilde{\mathbf{\Sigma}}_{MT_R}, \widetilde{\mathbf{\Sigma}}_{MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$ and $\hat{\mathbf{\Sigma}}^{-1} = \{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}, \hat{\mathbf{\Sigma}}^{-1}_{LW_{\hat{R}}}\}$. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. S-MT_R=Shrinkage on MT by row; S-MT_F=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 8: Relative performance of key regularisation estimators to the MT and S-MT estimators MT_F, $BL_{\hat{C}}$, CL_2, $CL_{\hat{C}}$ are compared to MT_R and S-MT_F, $BL_{\hat{C}^*}$, $CL_{\hat{C}^*}$, $LW_{\hat{\Sigma}}$ are compared to S-MT_R

Experiment B - non-normally distributed errors. Averages over 500 replications.

	$N=30$		$N=100$		$N=200$		$N=400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}$)							
	$T=60$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	1.035	1.103	0.982	1.073	0.929	1.044	0.896	1.017
$BL_{\hat{C}}$	1.050	1.194	1.039	1.108	0.921	1.050	0.888	1.020
CL_2	1.015	1.133	1.012	1.091	0.905	1.043	0.868	1.015
$CL_{\hat{C}}$	0.973	0.979	0.994	1.036	0.905	1.018	0.860	1.002
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	1.060	1.107	1.041	1.098	1.018	1.083	1.007	1.065
$BL_{\hat{C}^*}$	1.111	1.226	1.147	1.148	1.033	1.099	1.018	1.074
$CL_{\hat{C}^*}$	1.037	1.040	1.109	1.096	1.019	1.081	0.994	1.063
$LW_{\hat{\Sigma}}$	0.886	0.924	1.088	0.980	0.863	0.996	0.901	1.019
	$T=100$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	1.087	1.155	1.049	1.218	1.036	1.226	0.989	1.207
$BL_{\hat{C}}$	1.154	1.185	1.190	1.381	1.049	1.336	0.961	1.279
CL_2	1.143	1.273	1.141	1.320	1.028	1.286	0.980	1.253
$CL_{\hat{C}}$	1.005	1.003	1.004	0.999	0.964	0.997	0.925	1.061
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	1.079	1.121	1.067	1.175	1.085	1.200	1.060	1.198
$BL_{\hat{C}^*}$	1.142	1.155	1.232	1.345	1.140	1.326	1.079	1.284
$CL_{\hat{C}^*}$	1.006	0.967	1.047	0.996	1.057	1.053	1.049	1.173
$LW_{\hat{\Sigma}}$	1.071	1.123	1.337	1.257	1.091	1.286	1.197	1.314
	Error matrices ($\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1}$)							
	$T=60$							
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	1.138	1.107	1.079	1.111	1.041	1.093	1.008	1.077
$BL_{\hat{C}^*}$	1.0E+02	1.785	2.338	1.172	1.944	1.122	1.380	1.090
$CL_{\hat{C}^*}$	2.5E+05	15.995	189.473	1.893	268.017	1.497	7.0E+05	12.629
$LW_{\hat{\Sigma}}$	1.292	1.079	1.123	1.139	1.184	1.181	1.230	1.187
$LW_{\hat{R}}$	1.620	1.144	2.978	1.404	3.745	1.539	4.516	1.599
	$T=100$							
S-MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S-MT_F	1.245	1.138	1.287	1.214	1.265	1.229	1.233	1.225
$BL_{\hat{C}^*}$	4.5E+05	22.845	3.0E+03	2.836	3.190	1.397	1.432	1.326
$CL_{\hat{C}^*}$	3.1E+03	8.657	4.5E+04	13.837	1.8E+04	6.414	2.9E+02	2.132
$LW_{\hat{\Sigma}}$	1.643	1.238	1.516	1.382	1.652	1.454	1.714	1.458
$LW_{\hat{R}}$	1.362	1.184	2.119	1.407	2.673	1.569	3.243	1.668
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$
	$T=60$							
S-MT_R	0.389	0.611	0.426	0.574	0.429	0.571	0.445	0.555
S-MT_F	0.331	0.669	0.348	0.652	0.330	0.670	0.339	0.661
$LW_{\hat{\Sigma}}$	0.649	0.468	0.887	0.245	1.015	0.160	1.087	0.105
$LW_{\hat{R}}$	0.338	0.662	0.435	0.565	0.460	0.540	0.473	0.527
	$T=100$							
S-MT_R	0.364	0.636	0.429	0.571	0.434	0.566	0.437	0.563
S-MT_F	0.349	0.651	0.385	0.615	0.364	0.636	0.350	0.650
$LW_{\hat{\Sigma}}$	0.509	0.586	0.786	0.335	0.952	0.217	1.057	0.136
$LW_{\hat{R}}$	0.285	0.715	0.411	0.589	0.449	0.551	0.469	0.531

Notes: For the BL and CL methods, for $N=400$ and $T=60,100$ we set the number of replications to 100 and the grid increment to 4. Here, $\hat{\boldsymbol{\Sigma}}^{\star} = \{\widetilde{\boldsymbol{\Sigma}}_{MT_R}, \widetilde{\boldsymbol{\Sigma}}_{MT_F}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{CL,2}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT_R}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT_F}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}^*}, \hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}\}$ and $\hat{\boldsymbol{\Sigma}}^{\star-1} = \{\widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT_R}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT_F}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\boldsymbol{\Sigma}}^{-1}_{LW_{\hat{\Sigma}}}, \hat{\boldsymbol{\Sigma}}^{-1}_{LW_{\hat{R}}}\}$. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. S-MT_R=Shrinkage on MT by row; S-MT_F=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 9: Relative performance of key regularisation estimators to the MT and $S\text{-}MT$ estimators MT_F, $BL_{\hat{C}}$, CL_2, $CL_{\hat{C}}$ are compared to MT_R and $S\text{-}MT_F$, $BL_{\hat{C}^*}$, $CL_{\hat{C}^*}$, $LW_{\hat{\Sigma}}$ are compared to $S\text{-}MT_R$

Experiment C - non-normally distributed errors. Averages over 500 replications.

	$N = 30$		$N = 100$		$N = 200$		$N = 400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}$)							
	$T = 60$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	1.006	1.015	0.996	1.028	0.963	1.033	0.895	1.034
$BL_{\hat{C}}$	2.005	1.541	1.720	1.491	1.531	1.465	1.229	1.419
CL_2	1.200	1.149	1.274	1.205	1.259	1.234	1.109	1.250
$CL_{\hat{C}}$	1.061	1.064	1.014	1.048	1.001	1.058	0.957	1.104
$S\text{-}MT_R$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$S\text{-}MT_F$	1.106	1.055	1.079	1.050	1.069	1.051	1.048	1.053
$BL_{\hat{C}^*}$	1.732	1.441	1.395	1.315	1.309	1.281	1.201	1.248
$CL_{\hat{C}^*}$	1.724	1.437	1.394	1.312	1.308	1.280	1.200	1.248
$LW_{\hat{\Sigma}}$	1.054	1.259	1.148	1.434	1.143	1.519	0.967	1.542
	$T = 100$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	0.992	1.004	0.984	1.020	0.960	1.020	0.904	1.019
$BL_{\hat{C}}$	2.290	1.751	2.330	1.836	2.052	1.805	1.685	1.760
CL_2	1.090	1.097	1.160	1.148	1.228	1.188	1.178	1.216
$CL_{\hat{C}}$	1.019	1.037	0.987	1.026	0.964	1.020	0.939	1.035
$S\text{-}MT_R$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$S\text{-}MT_F$	1.123	1.060	1.119	1.072	1.101	1.069	1.081	1.066
$BL_{\hat{C}^*}$	2.332	1.755	1.676	1.528	1.527	1.464	1.406	1.410
$CL_{\hat{C}^*}$	2.312	1.746	1.672	1.525	1.525	1.463	1.406	1.408
$LW_{\hat{\Sigma}}$	1.197	1.400	1.413	1.710	1.440	1.866	1.298	1.944
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\boldsymbol{\Sigma}}$
	$T = 60$							
$S\text{-}MT_R$	0.400	0.600	0.562	0.438	0.602	0.398	0.633	0.367
$S\text{-}MT_F$	0.470	0.530	0.595	0.405	0.628	0.372	0.655	0.345
$LW_{\hat{\Sigma}}$	1.146	0.532	1.674	0.320	1.942	0.211	2.127	0.135
$LW_{\hat{R}}$	0.132	0.868	0.263	0.737	0.340	0.660	0.401	0.599
	$T = 100$							
$S\text{-}MT_R$	0.281	0.719	0.482	0.518	0.532	0.468	0.571	0.429
$S\text{-}MT_F$	0.356	0.644	0.542	0.458	0.584	0.416	0.618	0.382
$LW_{\hat{\Sigma}}$	0.866	0.649	1.417	0.428	1.762	0.289	2.030	0.180
$LW_{\hat{R}}$	0.089	0.911	0.203	0.797	0.287	0.713	0.362	0.638

Notes: For the BL and CL methods, for $N = 400$ and $T = 60, 100$ we set the number of replications to 100 and the grid increment to 4. Here, $\hat{\boldsymbol{\Sigma}} = \{\widetilde{\boldsymbol{\Sigma}}_{MT_R}, \widetilde{\boldsymbol{\Sigma}}_{MT_F}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{CL,2}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT_R}, \widetilde{\boldsymbol{\Sigma}}_{S\text{-}MT_F}, \widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}^*}, \hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}\}$. The population covariance matrix $\boldsymbol{\Sigma}$ does not have an inverse in this experiment hence results relating to matrix inverses are not provided. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. $S\text{-}MT_R$= Shrinkage on MT by row; $S\text{-}MT_F$=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 10: Relative performance of key regularisation estimators to the MT and $S\text{-}MT$ estimators MT_F, $BL_{\hat{C}}$, CL_2, $CL_{\hat{C}}$ are compared to MT_R and $S\text{-}MT_F$, $BL_{\hat{C}^*}$, $CL_{\hat{C}^*}$, $LW_{\hat{\Sigma}}$ are compared to $S\text{-}MT_R$

Experiment D - non-normally distributed errors. Averages over 500 replications.

	$N=30$		$N=100$		$N=200$		$N=400$	
	Norms		Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Error matrices ($\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}$)							
	$T=60$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	1.004	0.993	1.426	1.150	0.985	0.920	1.015	0.893
$BL_{\hat{C}}$	1.321	1.195	3.095	2.282	1.311	0.986	1.316	0.933
CL_2	1.163	1.119	2.817	2.210	1.236	0.973	1.282	0.927
$CL_{\hat{C}}$	1.181	1.122	2.711	2.170	1.249	0.977	1.249	0.926
$S\text{-}MT_R$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$S\text{-}MT_F$	1.040	1.017	1.316	1.164	1.103	0.993	1.178	0.991
$BL_{\hat{C}^*}$	1.320	1.205	2.190	2.049	1.367	1.056	1.392	1.030
$CL_{\hat{C}^*}$	1.249	1.174	2.184	2.048	1.355	1.054	1.386	1.029
$LW_{\hat{\Sigma}}$	1.016	1.100	2.997	2.168	1.172	1.019	1.267	1.048
	$T=100$							
MT_R	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MT_F	0.980	0.966	0.997	0.945	0.946	0.907	0.942	0.882
$BL_{\hat{C}}$	1.517	1.283	1.904	1.268	1.734	1.082	1.815	1.015
CL_2	1.149	1.107	1.696	1.221	1.370	1.016	1.518	0.983
$CL_{\hat{C}}$	1.123	1.082	1.332	1.108	1.438	1.030	1.520	0.980
$S\text{-}MT_R$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$S\text{-}MT_F$	0.997	0.988	1.038	0.994	1.053	0.972	1.133	0.963
$BL_{\hat{C}^*}$	1.574	1.317	2.060	1.411	1.760	1.136	1.910	1.087
$CL_{\hat{C}^*}$	1.310	1.204	2.060	1.412	1.749	1.133	1.909	1.086
$LW_{\hat{\Sigma}}$	1.239	1.310	2.195	1.447	1.551	1.299	1.752	1.341
	Error matrices ($\mathbf{\Sigma}^{-1} - \hat{\mathbf{\Sigma}}^{-1}$)							
	$T=60$							
$S\text{-}MT_R$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$S\text{-}MT_F$	1.076	1.016	1.157	1.059	1.186	0.992	1.196	0.979
$BL_{\hat{C}^*}$	2.598	1.291	3.918	1.590	1.725	1.071	1.566	1.028
$CL_{\hat{C}^*}$	44.853	1.805	5.190	1.642	8.573	1.169	17.846	1.142
$LW_{\hat{\Sigma}}$	2.352	1.205	2.079	1.434	2.180	0.980	2.010	0.934
$LW_{\hat{R}}$	1.519	1.449	1.592	2.522	2.719	2.537	3.585	2.858
	$T=100$							
$S\text{-}MT_R$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$S\text{-}MT_F$	0.992	0.989	1.033	0.998	1.110	0.987	1.249	0.974
$BL_{\hat{C}^*}$	57.098	1.923	8.877	1.366	2.008	1.168	16.354	1.219
$CL_{\hat{C}^*}$	35.382	2.082	6.362	1.320	32.792	1.400	2.129	1.103
$LW_{\hat{\Sigma}}$	2.540	1.359	2.039	1.260	2.457	1.246	2.720	1.222
$LW_{\hat{R}}$	1.160	1.348	1.079	1.704	1.441	2.477	2.320	3.022
	Srinkage parameters							
	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$	on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T=60$							
$S\text{-}MT_R$	0.395	0.605	0.406	0.594	0.418	0.582	0.455	0.545
$S\text{-}MT_F$	0.432	0.568	0.496	0.504	0.530	0.470	0.593	0.407
$LW_{\hat{\Sigma}}$	0.625	0.322	0.746	0.169	0.845	0.090	0.872	0.067
$LW_{\hat{R}}$	0.422	0.578	0.468	0.532	0.483	0.517	0.485	0.515
	$T=100$							
$S\text{-}MT_R$	0.356	0.644	0.395	0.605	0.369	0.631	0.344	0.656
$S\text{-}MT_F$	0.356	0.644	0.413	0.587	0.424	0.576	0.460	0.540
$LW_{\hat{\Sigma}}$	0.527	0.433	0.695	0.230	0.828	0.114	0.869	0.077
$LW_{\hat{R}}$	0.391	0.609	0.460	0.540	0.485	0.515	0.489	0.511

Notes: For the BL and CL methods, for $N=400$ and $T=60,100$ we set the number of replications to 100 and the grid increment to 4. Here, $\hat{\mathbf{\Sigma}} = \{\widetilde{\mathbf{\Sigma}}_{MT_R}, \widetilde{\mathbf{\Sigma}}_{MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}_{LW_\Sigma}\}$ and $\hat{\mathbf{\Sigma}}^{-1} = \{\widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_R}, \widetilde{\mathbf{\Sigma}}^{-1}_{S\text{-}MT_F}, \widetilde{\mathbf{\Sigma}}^{-1}_{BL,\hat{C}^*}, \widetilde{\mathbf{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\mathbf{\Sigma}}^{-1}_{LW_\Sigma}, \hat{\mathbf{\Sigma}}^{-1}_{LW_R}\}$. MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use the Bonferroni method at the 5% significance level. $S\text{-}MT_R$=Shrinkage on MT by row; $S\text{-}MT_F$=Shrinkage on MT on full $\hat{\mathbf{R}}$ matrix. BL=Bickel and Levina universal thresholding; CL= Cai and Liu adaptive thresholding. $\hat{C}$ uses a cross-validation parameter; $\hat{C}^*$ uses Fan, Liao and Michela grid adjustment; 2 is the CL optimal theoretical parameter; LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on the sample covariance matrix; $\hat{R}$ on the sample correlation matrix.

Table 11: Comparison of $\mathbf{\Sigma}$ support recovery produced by different thresholding estimators
Support recovery is measured by the True Positive Rate (TPR) and False Positive Rate (FPR)

Non-normally distributed errors. Averages over 500 replications.

	$N=30$		$N=100$		$N=200$		$N=400$	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
	Experiment C							
	$T=60$							
MT_R	0.710	0.002	0.588	0.001	0.552	0.000	0.520	0.000
MT_F	0.614	0.000	0.456	0.000	0.403	0.000	0.357	0.000
$BL_{\hat{C}}$	0.009	0.001	0.000	0.000	0.000	0.000	0.000	0.000
CL_2	0.450	0.000	0.277	0.000	0.213	0.000	0.156	0.000
$CL_{\hat{C}}$	0.655	0.009	0.536	0.003	0.480	0.001	0.377	0.001
	$T=100$							
MT_R	0.797	0.002	0.693	0.001	0.663	0.000	0.639	0.000
MT_F	0.730	0.000	0.592	0.000	0.550	0.000	0.514	0.000
$BL_{\hat{C}}$	0.200	0.031	0.000	0.000	0.000	0.000	0.000	0.000
CL_2	0.637	0.000	0.483	0.000	0.418	0.000	0.365	0.000
$CL_{\hat{C}}$	0.778	0.008	0.674	0.002	0.637	0.001	0.595	0.001
	Experiment D							
	$T=60$							
MT_R	0.941	0.003	0.954	0.001	0.898	0.001	0.859	0.000
MT_F	0.809	0.000	0.802	0.000	0.611	0.000	0.481	0.000
$BL_{\hat{C}}$	0.237	0.002	0.091	0.000	0.004	0.000	0.001	0.000
CL_2	0.312	0.000	0.193	0.000	0.092	0.000	0.051	0.000
$CL_{\hat{C}}$	0.367	0.003	0.288	0.001	0.094	0.000	0.001	0.000
	$T=100$							
MT_R	0.998	0.003	0.997	0.001	0.992	0.001	0.987	0.000
MT_F	0.986	0.000	0.974	0.000	0.941	0.000	0.895	0.000
$BL_{\hat{C}}$	0.322	0.001	0.395	0.000	0.025	0.000	0.007	0.000
CL_2	0.668	0.000	0.496	0.000	0.339	0.000	0.238	0.000
$CL_{\hat{C}}$	0.787	0.005	0.745	0.002	0.326	0.000	0.278	0.000

Notes: For the BL and CL methods, for $N=400$ and $T=60,100$ we set
the number of replications to 100 and the grid increment to 4.
MT_R=Multiple testing by row, MT_F=Multiple testing on the full $\hat{\mathbf{R}}$ matrix.
Both MT estimators use the Bonferroni method at the 5% significance level.
$BL_{\hat{C}}$=Bickel and Levina universal thresholding using a cross-validated parameter $\hat{C}$.
CL_2= Cai and Liu adaptive thresholding using the theoretical parameter of 2.
$CL_{\hat{C}}$= Cai and Liu adaptive thresholding using a cross-validated parameter $\hat{C}$.

Appendix A Mathematical Proofs

A.1 Lemmas and proofs for MT estimator

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

Lemma 1 *Suppose that $x \sim N(\rho, \sigma^2)$, then*

$$E\left[xI(a \leq x \leq b)\right] = \rho\left[\Phi\left(\frac{b-\rho}{\sigma}\right) - \Phi\left(\frac{a-\rho}{\sigma}\right)\right] + \sigma\left[\phi\left(\frac{a-\rho}{\sigma}\right) - \phi(\frac{b-\rho}{\sigma})\right], \tag{A.1}$$

and

$$E\left[x^2 I(a \leq x \leq b)\right] = \left(\sigma^2+\rho^2\right)\left[\Phi\left(\frac{b-\rho}{\sigma}\right) - \Phi\left(\frac{a-\rho}{\sigma}\right)\right] + \sigma\left(a+\rho\right)\phi(\frac{a-\rho}{\sigma}) - \sigma\left(b+\rho\right)\phi(\frac{b-\rho}{\sigma}). \tag{A.2}$$

Proof. Note that

$$E\left[xI(a \leq x \leq b)\right] = \int_a^b x(2\pi\sigma^2)^{-1/2} e^{-(1/2)(x-\rho)^2/\sigma^2} dx.$$

Let $z = (x-\rho)/\sigma$, then

$$E\left[xI(a \leq x \leq b)\right] = \int_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} (\sigma z + \rho)\phi(z)dz,$$

where $\phi(z) = (2\pi)^{-1/2}\exp(-0.5z^2)$. But

$$\int_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} (\sigma z + \rho)\phi(z)dz = \sigma\left[-\phi(z)\right]_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} + \rho\int_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} \phi(z)dz,$$

and hence

$$E\left[xI(a \leq x \leq b)\right] = \rho\left[\Phi\left(\frac{b-\rho}{\sigma}\right) - \Phi\left(\frac{a-\rho}{\sigma}\right)\right] + \sigma\left[\phi\left(\frac{a-\rho}{\sigma}\right) - \phi(\frac{b-\rho}{\sigma})\right],$$

which establishes (A.1). To prove (A.2) note that using the transformation $z = (x-\rho)/\sigma$ we have

$$E\left[x^2 I(a \leq x \leq b)\right] = \int_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} \left(\sigma^2 z^2 + \rho^2 + 2\rho\sigma z\right)\phi(z)dz.$$

But

$$\begin{aligned}
\int_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} z^2\phi(z)dz &= \left[-z\phi(z)\right]_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} + \Phi\left(\frac{b-\rho}{\sigma}\right) - \Phi\left(\frac{a-\rho}{\sigma}\right) \\
&= \Phi\left(\frac{b-\rho}{\sigma}\right) - \Phi\left(\frac{a-\rho}{\sigma}\right) - \left(\frac{b-\rho}{\sigma}\right)\phi(\frac{b-\rho}{\sigma}) + \left(\frac{a-\rho}{\sigma}\right)\phi(\frac{a-\rho}{\sigma}),
\end{aligned}$$

and

$$\int_{(a-\rho)/\sigma}^{(b-\rho)/\sigma} z\phi(z)dz = \phi(\frac{a-\rho}{\sigma}) - \phi(\frac{b-\rho}{\sigma}).$$

Therefore

$$E\left[x^2 I(a \leq x \leq b)\right] = \left(\sigma^2 + \rho^2\right)\left[\Phi\left(\frac{b-\rho}{\sigma}\right) - \Phi\left(\frac{a-\rho}{\sigma}\right)\right] + \sigma\left(a+\rho\right)\phi(\frac{a-\rho}{\sigma}) - \sigma\left(b+\rho\right)\phi(\frac{b-\rho}{\sigma}).$$

which establishes (A.2). ■

Lemma 2 *Let* $b_N = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right)$, *where* $p/\left[2f(N)\right]$ *is sufficiently small such that* $1 - \frac{p}{2f(N)} > 0$, *then*

$$b_N \leq \sqrt{2\left[\ln f(N) - \ln(p)\right]}. \tag{A.3}$$

Proof. First note that

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

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

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du. \tag{A.4}$$

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 b_N we have

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

■

Lemma 3 *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 $x > 0$

$$\Phi(-x) = 1 - \Phi(x) \leq \frac{1}{2}\exp(-\frac{x^2}{4}). \tag{A.5}$$

Proof. Using results in Chiani et al. (2003, p.840) we have

$$\text{erf c}(x) = \frac{2}{\sqrt{\pi}}\int_{x}^{\infty} e^{-u^2}du \leq \exp(-\frac{x^2}{2}), \tag{A.6}$$

where erf $c(x)$ is the complement of the erf(x) function defined by (A.4). 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.6) we have

$$1 - \Phi(x) = \frac{1}{2}\text{erf c}\left(\frac{x}{\sqrt{2}}\right) \leq \frac{1}{2}\exp\left[-\frac{1}{2}\left(\frac{x}{\sqrt{2}}\right)^2\right] = \frac{1}{2}\exp\left(-\frac{x^2}{4}\right).$$

■

Lemma 4 *(i) Under Assumption 1,*

$$E[I\left(\left|\hat{\rho}_{ij}\right| \leq \frac{b_N}{\sqrt{T}}\right)] = P(L_{ij} \leq z_{ij} \leq\ U_{ij}) = \Phi(U_{ij}) - \Phi(L_{ij}),$$

where $z_{ij} = (\hat{\rho}_{ij} - \mu_{ij})/\omega_{ij}$, b_N *is defined as in Lemma 2, and*

$$U_{ij} = \begin{cases} O(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}), \text{ if } \rho_{ij} \neq 0 \\ b_N, \text{ otherwise} \end{cases}, \text{ and } L_{ij} = \begin{cases} O(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}), \text{ if } \rho_{ij} \neq 0 \\ -b_N, \text{ otherwise} \end{cases}. \tag{A.7}$$

(ii) Under Assumptions 1 and 2,

$$\sum_{i\neq j,\rho_{ij}\neq 0}\sum E[I\left(\left|\hat{\rho}_{ij}\right| \leq \frac{b_N}{\sqrt{T}}|\rho_{ij} \neq 0\right)] \leq 2m_N N\Phi(\frac{b_N - \sqrt{T}\rho_{\min}}{1-\rho_{\min}^2}).$$

Proof. *(i) Under (10) of Assumption 1*

$$z_{ij} = \frac{\hat{\rho}_{ij} - \mu_{ij}}{\omega_{ij}} \sim N(0,1).$$

The required result follows trivially,

$$\begin{aligned} E[I\left(\left|\hat{\rho}_{ij}\right| \leq \frac{b_N}{\sqrt{T}}\right)] &= E[I\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2} \leq \frac{\hat{\rho}_{ij} - \mu_{ij}}{\omega_{ij}} \leq \frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)] \\ &= P(L_{ij} \leq z_{ij} \leq\ U_{ij}) = \Phi(U_{ij}) - \Phi(L_{ij}). \end{aligned}$$

(ii) From part (i) it follows that

$$\sum_{i\neq j,\rho_{ij}\neq 0}\sum E[I\left(\left|\hat{\rho}_{ij}\right| \leq \frac{b_N}{\sqrt{T}}|\rho_{ij} \neq 0\right)] = \sum_{i\neq j,\rho_{ij}\neq 0}\sum\left\{\Phi\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) - \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right\}.$$

Distinguishing between cases where ρ_{ij} *are strictly positive and negative the last expression in the above can be written as*

$$\begin{aligned} &\sum_{i\neq j,\rho_{ij}>0}\sum\left\{\Phi\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) - \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right\} + \sum_{i\neq j,\rho_{ij}<0}\sum\left\{\Phi\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) - \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right\} \\ = &\sum_{i\neq j,\rho_{ij}>0}\sum\left\{\Phi\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) - \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right\} + \sum_{i\neq j,\rho_{ij}<0}\sum\left\{\Phi\left(\frac{b_N + \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) - \Phi\left(\frac{-b_N + \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right\} \\ = &\ 2\sum_{i\neq j,|\rho_{ij}|>0}\sum\left\{\Phi\left(\frac{b_N - \sqrt{T}\left|\rho_{ij}\right|}{1-\rho_{ij}^2}\right) - \Phi\left(\frac{-b_N - \sqrt{T}\left|\rho_{ij}\right|}{1-\rho_{ij}^2}\right)\right\}. \end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{i\neq j,\rho_{ij}\neq 0}\sum E\left[I\left(L_{ij}\leq z_{ij}\leq U_{ij}\left|\rho_{ij}\neq 0\right.\right)\right] \\
\leq\ & 2m_N N\left[\Phi\left(\frac{b_N-\sqrt{T}\rho_{\min}}{1-\rho_{\min}^2}\right)-\Phi\left(\frac{-b_N-\sqrt{T}\rho_{\max}}{1-\rho_{\max}^2}\right)\right] \\
\leq\ & 2m_N N\Phi\left(\frac{b_N-\sqrt{T}\rho_{\min}}{1-\rho_{\min}^2}\right).
\end{aligned}$$

■

A.2 Proofs of theorems for MT estimator

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

Proof of Theorem 1. Consider

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

and note that

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

Hence

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

However,

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

and

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

Therefore, we have

$$\begin{aligned}
\sum_{i\neq j}\sum\left(\tilde{\rho}_{ij}-\rho_{ij}\right)^2 &= \sum_{i\neq j}\sum\left(\hat{\rho}_{ij}-\rho_{ij}\right)^2 I\left(\left|\hat{\rho}_{ij}\right|>\frac{b_N}{\sqrt{T}}\right)+\sum_{i\neq j}\sum\rho_{ij}^2\left[1-I\left(\left|\hat{\rho}_{ij}\right|>\frac{b_N}{\sqrt{T}}\right)\right] \\
&= \sum_{i\neq j}\sum\left(\hat{\rho}_{ij}-\rho_{ij}\right)^2 I\left(\left|\hat{\rho}_{ij}\right|>\frac{b_N}{\sqrt{T}}\right)+\sum_{i\neq j}\sum\rho_{ij}^2 I\left(\left|\hat{\rho}_{ij}\right|\leq\frac{b_N}{\sqrt{T}}\right). \qquad \text{(A.8)}
\end{aligned}$$

To simplify the derivations we write all the indicator functions in terms of $z_{ij}=(\hat{\rho}_{ij}-\mu_{ij})/\omega_{ij}$, with μ_{ij} and ω_{ij} defined in (11) and (12) of Assumption 1, respectively. Hence, from part (i) of Lemma 4 it follows that

$$I\left(\left|\hat{\rho}_{ij}\right|>\frac{b_N}{\sqrt{T}}\right)=1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\right),$$

where U_{ij} and L_{ij} are given in (A.7) of the same lemma.

Consider now a typical element in the first term of (A.8) and note that it can be rewritten as

$$\begin{aligned}
\left(\hat{\rho}_{ij}-\rho_{ij}\right)^2 I\left(\left|\hat{\rho}_{ij}\right|>\frac{b_N}{\sqrt{T}}\right) &= \left(\hat{\rho}_{ij}-\mu_{ij}+\mu_{ij}-\rho_{ij}\right)^2\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\right)\right] \\
&= \left[\omega_{ij}^2 z_{ij}^2+2\omega_{ij}\left(\mu_{ij}-\rho_{ij}\right)z_{ij}+\left(\mu_{ij}-\rho_{ij}\right)^2\right]\times\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\right)\right].
\end{aligned}$$

From (11) and (12) of assumption 1, we note that

$$\begin{aligned}
\left(\mu_{ij}-\rho_{ij}\right)^2 &= 0, \text{ if } \rho_{ij}=0, \\
\left(\mu_{ij}-\rho_{ij}\right)^2 &= \frac{\rho_{ij}^2(1-\rho_{ij}^2)^2}{4T^2}+O\left(T^{-3}\right)=O(T^{-2}), \text{ if } \rho_{ij}\neq 0.
\end{aligned}$$

and

$$\begin{aligned}
\omega_{ij}\left(\mu_{ij}-\rho_{ij}\right) &= 0 \text{ if } \rho_{ij}=0 \\
\omega_{ij}\left(\mu_{ij}-\rho_{ij}\right) &= \frac{(1-\rho_{ij}^2)}{\sqrt{T}}\left[1+O(T^{-1})\right]^{1/2}\left[-\frac{\rho_{ij}(1-\rho_{ij}^2)}{2T}+\frac{G(\rho_{ij})}{T^2}\right]=O\left(T^{-3/2}\right), \text{ if } \rho_{ij}\neq 0.
\end{aligned}$$

Collecting the various terms, we can now write

$$\begin{aligned} E\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F^2 &= \sum_{i\neq j}\sum E\left\{\left[\omega_{ij}^2 z_{ij}^2+\left(\mu_{ij}-\rho_{ij}\right)^2+2\omega_{ij}\left(\mu_{ij}-\rho_{ij}\right)z_{ij}\right]\times\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\right)\right]\right\} \\ &\quad +\sum_{i\neq j}\sum\rho_{ij}^2 E\left[I\left(L_{ij}\leq z_{ij}\leq U_{ij}\right)\right]. \end{aligned}$$

We now decompose each of the above sums into those with $\rho_{ij}=0$ and those where $\rho_{ij}\neq 0$, and write

$$\begin{aligned} E\left\|\widetilde{\boldsymbol{R}}-\boldsymbol{R}\right\|_F^2 &= \sum_{i\neq j,\ \rho_{ij}\neq 0}\sum E\left\{\begin{array}{c}\left[\omega_{ij}^2 z_{ij}^2+\left(\mu_{ij}-\rho_{ij}\right)^2+2\omega_{ij}\left(\mu_{ij}-\rho_{ij}\right)z_{ij}\right] \\ \times\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}\neq 0\right)\right]\end{array}\right\} \\ &\quad +\sum_{i\neq j,\rho_{ij}\neq 0}\sum\rho_{ij}^2 E\left[I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}\neq 0\right)\right] \\ &\quad +\sum_{i\neq j,\ \rho_{ij}=0}\sum E\left\{\omega_{ij}^2 z_{ij}^2\times\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}=0\right)\right]\right\}. \end{aligned} \tag{A.9}$$

Consider the three terms in the above expression starting with the second term. We distinguish between cases where ρ_{ij} are strictly positive and negative as in part (ii) of Lemma 4 from which it follows that

$$\begin{aligned} &\sum_{i\neq j,\rho_{ij}\neq 0}\sum\rho_{ij}^2 E\left[I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}\neq 0\right)\right] \\ \leq\ & 2\rho_{\max}^2 m_N N\Phi\left(\frac{b_N-\sqrt{T}\rho_{\min}}{1-\rho_{\min}^2}\right) \\ =\ & 2\rho_{\max}^2 m_N N\Phi\left[\frac{-\sqrt{T}\rho_{\min}\left(1-\frac{b_N}{\sqrt{T}\rho_{\min}}\right)}{1-\rho_{\min}^2}\right]. \end{aligned}$$

Using (A.3) of Lemma 2 and under our assumptions, $\frac{b_N}{\sqrt{T}\rho_{\min}}=o(1)$, and

$$N\Phi\left[\frac{-\sqrt{T}\rho_{\min}\left(1-\frac{b_N}{\sqrt{T}\rho_{\min}}\right)}{1-\rho_{\min}^2}\right]=O\left[N\Phi\left(\frac{-\sqrt{T}\rho_{\min}}{1-\rho_{\min}^2}\right)\right].$$

But by (A.5) of Lemma 3

$$N\Phi\left(\frac{-\sqrt{T}\rho_{\min}}{1-\rho_{\min}^2}\right)\leq\frac{1}{2}N\exp\left[\frac{-1}{4}\frac{T\rho_{\min}^2}{\left(1-\rho_{\min}^2\right)^2}\right]=o(1).$$

Note that this result *does not* require $N/T\to 0$, and holds even if N/T tends to a fixed constant.

Consider now the third term of (A.9)

$$\begin{aligned} &\sum_{i\neq j,\ \rho_{ij}=0}\sum E\left\{\omega_{ij}^2 z_{ij}^2\times\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}=0\right)\right]\right\} \\ =\ & \left[\frac{1}{T}+O(T^{-2})\right]\sum_{i\neq j,\ \rho_{ij}=0}\sum E\left\{z_{ij}^2\times\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}=0\right)\right]\right\}. \end{aligned}$$

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

But since under $\rho_{ij}=0$, $U_{ij}=b_N$ and $L_{ij}=-b_N$, we then have

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

and

$$\sum_{i\neq j,\ \rho_{ij}=0}\sum E\left\{\omega_{ij}^2 z_{ij}^2\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}=0\right)\right]\right\}\approx\frac{N(N-m_N-1)}{T}\left[2\Phi\left(-b_N\right)+2b_N\phi(b_N)\right].$$

However,

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

and hence

$$\begin{aligned} &\sum_{i\neq j,\ \rho_{ij}=0}\sum E\left\{\omega_{ij}^2 z_{ij}^2\left[1-I\left(L_{ij}\leq z_{ij}\leq U_{ij}\,\middle|\,\rho_{ij}=0\right)\right]\right\} \\ \approx\ & \frac{N(N-m_N-1)}{T}\left[\frac{p}{f(N)}+2(2\pi)^{-1/2}b_N\exp\left(\frac{-1}{2}b_N^2\right)\right]. \end{aligned}$$

The first term in the above expression is $o(1)$ if $f(N) = O(N^2)$ for N and T large. But we need the additional restriction of $N/T \to 0$, if $f(N) = O(N)$. To ensure that the second term tends to zero, we need $N/T \to 0$, as well as $Nb_N \exp\left(\frac{-1}{2}b_N^2\right)$ being bounded in N. Finally, consider the first term of (A.9), and note that

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

and

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

$$\begin{aligned} &\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum E\left\{\left[\omega_{ij}^2 z_{ij}^2 + \left(\mu_{ij} - \rho_{ij}\right)^2 + 2\omega_{ij}\left(\mu_{ij} - \rho_{ij}\right) z_{ij}\right] \times \left[1 - I\left(L_{ij} \leq z_{ij} \leq U_{ij}\left|\rho_{ij} \neq 0\right.\right)\right]\right\} \\ = &\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \left\{ \begin{array}{c} \omega_{ij}^2\left[\Phi\left(-U_{ij}\right) + \Phi\left(L_{ij}\right) + U_{ij}\phi(U_{ij}) - L_{ij}\phi(L_{ij})\right] + \\ \left(\mu_{ij} - \rho_{ij}\right)^2\left[\Phi\left(-U_{ij}\right) + \Phi\left(L_{ij}\right)\right] + 2\omega_{ij}\left(\mu_{ij} - \rho_{ij}\right)\left[-\phi(L_{ij}) + \phi(U_{ij})\right] \end{array} \right\}. \end{aligned}$$

Hence, using the expressions for U_{ij} and L_{ij} under $\rho_{ij} \neq 0$,

$$\begin{aligned} &\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \left\{ \begin{array}{c} \omega_{ij}^2\left\{ \begin{array}{c} \Phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) + \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) + \left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) \\ + \left(\frac{b_N + \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) \end{array} \right\} + \\ \left(\mu_{ij} - \rho_{ij}\right)^2\left[\Phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) + \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right] \\ +2\omega_{ij}\left(\mu_{ij} - \rho_{ij}\right)\left[\phi\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) - \phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right] \end{array} \right\} \\ = &\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \left[\omega_{ij}^2 + \left(\mu_{ij} - \rho_{ij}\right)^2\right]\left[\Phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) + \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right] \\ &+ \sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \omega_{ij}^2\left[\begin{array}{c} \left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) \\ + \left(\frac{b_N + \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) \end{array} \right] \\ &+2\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \omega_{ij}\left(\mu_{ij} - \rho_{ij}\right)\left[\phi\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) - \phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\right]. \end{aligned}$$

Since $\omega_{ij}^2 = O(T^{-1})$, and $\left(\mu_{ij} - \rho_{ij}\right) = O(T^{-1})$, and also $\Phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) + \Phi\left(\frac{-\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) < 2$, then

$$\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \left[\omega_{ij}^2 + \left(\mu_{ij} - \rho_{ij}\right)^2\right]\left[\Phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) + \Phi\left(\frac{-\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right)\right] < 2\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \left[\omega_{ij}^2 + \left(\mu_{ij} - \rho_{ij}\right)^2\right],$$

and

$$2\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \left[\omega_{ij}^2 + \left(\mu_{ij} - \rho_{ij}\right)^2\right] = O\left(\frac{m_N N}{T}\right).$$

Also,

$$\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right) = (2\pi)^{-1/2}\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\exp\left(\left[\frac{-1}{2}\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)^2\right]\right),$$

and

$$\begin{aligned} &\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \omega_{ij}^2\left[\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\phi\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right)\right] \\ = &(2\pi)^{-1/2}\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \omega_{ij}^2\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\exp\left(\left[\frac{-1}{2}\left(\frac{\sqrt{T}\rho_{ij} - b_N}{1-\rho_{ij}^2}\right)^2\right]\right) \\ = &(2\pi)^{-1/2}\sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \omega_{ij}^2\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right)\exp\left(\frac{-T\rho_{ij}^2}{2}\left(\frac{b_N^2}{T\rho_{ij}^2} + 1 - 2\frac{b_N}{\rho_{ij}\sqrt{T}}\right)\right). \end{aligned}$$

But by (A.3) of Lemma 2, $\frac{b_N^2}{T} = o(1)$, and $T\exp\left(\frac{-T\rho_{\min}^2}{2}\right) \to 0$ as $T \to \infty$, and

$$\begin{aligned}
& (2\pi)^{-1/2} \sum_{i\neq j,\ \rho_{ij}\neq 0}\sum \omega_{ij}^2 \left(\frac{b_N - \sqrt{T}\rho_{ij}}{1-\rho_{ij}^2}\right) \exp\left[\frac{-T\rho_{ij}^2}{2}\left(\frac{b_N^2}{T\rho_{ij}^2} + 1 - 2\frac{b_N}{\rho_{ij}\sqrt{T}}\right)\right] \\
= \quad & O\left[\frac{m_N N}{T}\sqrt{T}\exp\left(\frac{-T\rho_{\min}^2}{2}\right)\right] = o(1).
\end{aligned}$$

Overall, the order of the final term is given by

$$\begin{aligned}
& \sum_{i\neq j,\ \rho_{ij}\neq 0}\sum E\left\{\left[\omega_{ij}^2 z_{ij}^2 + \left(\mu_{ij}-\rho_{ij}\right)^2 + 2\omega_{ij}\left(\mu_{ij}-\rho_{ij}\right)z_{ij}\right] \times \left[1 - I\left(L_{ij} \leq z_{ij} \leq U_{ij}\,|\rho_{ij}\neq 0\right)\right]\right\} \\
= \quad & O\left(\frac{m_N N}{T}\right).
\end{aligned}$$

Considering the results for all the three terms together we note that the order of $E\left\|\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right\|_F^2$ depends on the order of $Nb_N\exp\left(\frac{-1}{2}b_N^2\right)$. But from (A.3) of Lemma 2 setting $b_N = \sqrt{2\left[\ln f(N) - \ln(p)\right]}$ we have

$$\begin{aligned}
Nb_N\exp\left(\frac{-1}{2}b_N^2\right) &= \frac{Np\sqrt{2\left[\ln f(N) - \ln(p)\right]}}{f(N)} \\
&= \begin{cases} O(\sqrt{\ln N}), \text{ if } f(N) = O(N) \\ O(\frac{\sqrt{\ln N}}{N}), \text{ if } f(N) = O(N^2) \end{cases},
\end{aligned}$$

and therefore $Nb_N\exp\left(\frac{-1}{2}b_N^2\right)$ will be bounded in N only if $f(N) = O(N^2)$. Consequently

$$E\left\|\widetilde{\boldsymbol{R}} - \boldsymbol{R}\right\|_F^2 = O\left(\frac{m_N N}{T}\right), \text{ if } f(N) = O(N^2). \tag{A.10}$$

■

Proof of Theorem 2. Consider first the FPR statistic given by (15) which can be written equivalently as

$$FPR = \frac{\sum_{i\neq j}\sum I\left(|\hat{\rho}_{ij}| > \frac{b_N}{\sqrt{T}}|\rho_{ij} = 0\right)}{N(N - m_N - 1)}. \tag{A.11}$$

Taking the expectation of (A.11) we have

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

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

As earlier, to simplify the derivations we will write all the indicator functions in terms of $z_{ij} = (\hat{\rho}_{ij} - \mu_{ij})/\omega_{ij}$ with μ_{ij} and ω_{ij} defined in (11) and (12) of assumption 1, respectively. Using the property

$$I\left(|\hat{\rho}_{ij}| > \frac{b_N}{\sqrt{T}}|\rho_{ij} = 0\right) = 1 - I\left(|\hat{\rho}_{ij}| \leq \frac{b_N}{\sqrt{T}}|\rho_{ij} = 0\right),$$

and taking expectations it follows from part (i) of Lemma 4 that

$$\begin{aligned}
E\left[I\left(|\hat{\rho}_{ij}| > \frac{b_N}{\sqrt{T}}|\rho_{ij} = 0\right)\right] &= 1 - P(L_{ij} \leq z_{ij} \leq U_{ij}|\rho_{ij} = 0), \\
&= 1 - [\Phi(b_N) - \Phi(-b_N)] \\
&= 2[1 - \Phi(b_N)] \\
&= 2\left\{1 - \Phi\left[\Phi^{-1}\left(1 - \frac{p/2}{f(N)}\right)\right]\right\} \\
&= \frac{p}{f(N)},
\end{aligned}$$

with U_{ij} and L_{ij} given in (A.7) of the same lemma. Hence, $E\,|FPR| = \frac{N(N-1)p}{N(N-m_N-1)f(N)} = \frac{(N-1)p}{(N-m_N-1)f(N)} \to 0$ as $N \to \infty$, so long as $f(N) \to \infty$. But by the Markov inequality applied to $|FPR|$ we have that

$$P(|FPR| > \epsilon) \leq \frac{E(|FPR|)}{\epsilon} = \frac{p}{\epsilon f(N)},$$

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

Consider next the TPR statistic given by (14) and set

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

As before $|X| = X$ and $P(|X| > \epsilon) \leq \frac{E|X|}{\epsilon}$. But

$$E(X) = E|X| = \frac{\sum\sum_{i\neq j} P\left(|\hat{\rho}_{ij}| < \frac{b_N}{\sqrt{T}} | \rho_{ij} \neq 0\right)}{\sum\sum_{i\neq j} I(\rho_{ij} \neq 0)},$$

and from part (i) of Lemma 4 we have that

$$\begin{aligned} P\left(|\hat{\rho}_{ij}| < \frac{b_N}{\sqrt{T}} | \rho_{ij} \neq 0\right) &= P(L_{ij} \leq z_{ij} \leq U_{ij} | \rho_{ij} \neq 0) \\ &= \Phi\left(\frac{b_N - \sqrt{T}\rho_{ij}}{1 - \rho_{ij}^2}\right) - \Phi\left(\frac{-b_N - \sqrt{T}\rho_{ij}}{1 - \rho_{ij}^2}\right). \end{aligned}$$

We can further distinguish between cases where ρ_{ij} are strictly positive and negative as in part (ii) of Lemma 4 from which it follows that

$$E|X| \leq \frac{2m_N N}{m_N N} \Phi\left(\frac{b_N - \sqrt{T}\rho_{\min}}{1 - \rho_{\min}^2}\right).$$

Hence

$$P(|TPR - 1| > \epsilon) \leq 2\Phi\left(\frac{b_N - \sqrt{T}\rho_{\min}}{1 - \rho_{\min}^2}\right),$$

and the desired result is established if $b_N - \sqrt{T}\rho_{\min} \to -\infty$ which is equivalent to $\rho_{\min} > \frac{b_N}{\sqrt{T}}$, as $N, T \to \infty$ in any order. ■

A.3 Proof of theorem and corollary for shrinkage estimator $\boldsymbol{R}_{LW}$

Proof of Theorem 3 and Corollary 1. This proof has two parts. In the first part we obtain the optimal value of the shrinkage parameter that minimizes the squared Frobenius norm of the error of estimating $\boldsymbol{R}$ by $\hat{\boldsymbol{R}}_{LW}$. In the second part we obtain the convergence rate of the shrinkage correlation matrix estimator under the derived shrinkage parameter.

Taking the expectation of $\left\|\hat{\boldsymbol{R}}_{LW} - \boldsymbol{R}\right\|_F^2$, with $\hat{\boldsymbol{R}}_{LW} = \xi \mathbf{I}_N + (1-\xi)\hat{\boldsymbol{R}}$, we have

$$E\left\|\hat{\boldsymbol{R}}_{LW} - \boldsymbol{R}\right\|_F^2 = \sum\sum_{i\neq j} E\left(\hat{\rho}_{ij} - \rho_{ij}\right)^2 + \xi^2 \sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}^2\right) - 2\xi \sum\sum_{i\neq j} E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij} - \rho_{ij}\right)\right], \tag{A.12}$$

and following Ledoit and Wolf (2003, 2004) and Schäfer and Strimmer (2005) the value of ξ that minimizes (A.12) is given by

$$\xi^* = \frac{\sum\sum_{i\neq j} E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij} - \rho_{ij}\right)\right]}{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}^2\right)} = 1 - \frac{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}\rho_{ij}\right)}{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}^2\right)}. \tag{A.13}$$

Using (11) of Assumption 1 we have that

$$b_{ij} = E(\hat{\rho}_{ij}) - \rho_{ij} = -\frac{\rho_{ij}(1 - \rho_{ij}^2)}{2T} + \frac{G(\rho_{ij})}{T^2}. \tag{A.14}$$

Thus, in terms of b_{ij} and $Var(\hat{\rho}_{ij})$, it follows that

$$1 - \xi^* = \frac{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}\rho_{ij}\right)}{\sum\sum_{i\neq j} E\left(\hat{\rho}_{ij}^2\right)} = \frac{\sum\sum_{i\neq j} \rho_{ij}(b_{ij} + \rho_{ij})}{\sum\sum_{i\neq j} Var\left(\hat{\rho}_{ij}\right) + \sum\sum_{i\neq j}(b_{ij} + \rho_{ij})^2}. \tag{A.15}$$

Substituting for (12) of Assumption 1 and (A.14) in (A.15) yields

$$1-\xi^{*}=\frac{\sum\sum_{i\neq j}\rho_{ij}(\rho_{ij}-\frac{\rho_{ij}(1-\rho_{ij}^{2})}{2T}+\frac{G(\rho_{ij})}{T^{2}})}{\sum\sum_{i\neq j}\left[\frac{(1-\rho_{ij}^{2})^{2}}{T}+\frac{K(\rho_{ij})}{T^{2}}\right]+\sum\sum_{i\neq j}\left[\rho_{ij}-\frac{\rho_{ij}(1-\rho_{ij}^{2})}{2T}+\frac{G(\rho_{ij})}{T^{2}}\right]^{2}}.$$

Hence, an estimator of ξ^{*} can be obtained (ignoring terms of order T^{-2}) as

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

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

Using (A.13) in (A.12) we have that

$$\begin{aligned}E\left\|\hat{\boldsymbol{R}}_{LW}-\boldsymbol{R}\right\|_{F}^{2} &= \sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}-\rho_{ij}\right)^{2}-\frac{\left[\sum\sum_{i\neq j}E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right]\right]^{2}}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)}\\ &< \sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}-\rho_{ij}\right)^{2},\end{aligned}$$

which postulates that the expected quadratic loss of the shrinkage sample covariance estimator is smaller than that of the sample covariance matrix, suggesting an improvement using the former compared to the latter. Further we have

$$\begin{aligned}\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}-\rho_{ij}\right)^{2} &= \sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)-2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)+\sum\sum_{i\neq j}\rho_{ij}^{2},\\ \left\{\sum\sum_{i\neq j}E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right]\right\}^{2} &= \left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)-\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\right]^{2}\\ &= \left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)\right]^{2}+\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\right]^{2}-2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right),\end{aligned}$$

and

$$E\left\|\hat{\boldsymbol{R}}_{LW}-\boldsymbol{R}\right\|_{F}^{2}=\frac{\left\{\begin{array}{c}\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)-2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)+\sum\sum_{i\neq j}\rho_{ij}^{2}\right]\\ -\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)\right]^{2}-\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\right]^{2}+2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\end{array}\right\}}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^{2}\right)}.$$

Hence,

$$
\begin{aligned}
E\left\|\hat{\boldsymbol{R}}_{LW}-\boldsymbol{R}\right\|_F^2 &= \frac{\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)-\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\right]^2}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)} \\
&= \frac{\sum\sum_{i\neq j}\rho_{ij}^2[\sum\sum_{i\neq j}Var\left(\hat{\rho}_{ij}\right)+\sum\sum_{i\neq j}(b_{ij}+\rho_{ij})^2]-\left[\sum\sum_{i\neq j}\rho_{ij}(b_{ij}+\rho_{ij})\right]^2}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)} \\
&= \frac{\left\{\begin{array}{c}\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}Var\left(\hat{\rho}_{ij}\right)+\left[\sum\sum_{i\neq j}\rho_{ij}^2\right]^2+\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}b_{ij}^2+2\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}b_{ij}\rho_{ij} \\ -\left[\sum\sum_{i\neq j}b_{ij}\rho_{ij}\right]^2-\left[\sum\sum_{i\neq j}\rho_{ij}^2\right]^2-2\left[\sum\sum_{i\neq j}b_{ij}\rho_{ij}\right]\left[\sum\sum_{i\neq j}\rho_{ij}^2\right]\end{array}\right\}}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)} \\
&= \frac{\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}Var\left(\hat{\rho}_{ij}\right)+\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}b_{ij}^2-\left[\sum\sum_{i\neq j}b_{ij}\rho_{ij}\right]^2}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}.
\end{aligned}
$$

Finally, using Assumptions 1 and 2, it follows from the above results that

$$E\left\|\hat{\boldsymbol{R}}_{LW}^*(\xi^*)-\mathbf{R}\right\|_F^2=O\left(\frac{m_N N}{T}\right),$$

which is in line with the result obtained by LW. ■

A.4 Derivation of the shrinkage parameter for shrinkage on MT (S-MT) estimator

Recall the expression for the function $f(\lambda)$ from Section 4

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

with $\boldsymbol{A}=\boldsymbol{R}_0^{-1}$ and $\boldsymbol{B}(\lambda)=\widetilde{\boldsymbol{R}}_{S\text{-}MT}^{-1}(\lambda)$. We need to solve $f(\lambda)=0$, for λ^* such that $f(\lambda^*)=0$ for a given choice of $\mathbf{R}_0$.

Abstracting from the subscripts, note that

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

or

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

which is generally non-zero. Also, $\lambda=0$ 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 $f(\lambda)=0$ has a solution in the range $\lambda_0<\lambda<1$, where λ_0 is the minimum value of λ such that $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\lambda_0)$ is non-singular. First, we can compute λ_0 by implementing naive shrinkage as an initial estimate:

$$\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\lambda_0)=\lambda_0\boldsymbol{I}_N+(1-\lambda_0)\widetilde{\boldsymbol{R}}.$$

The shrinkage parameter $\lambda_0\in[0,1]$ is given by

$$\lambda_0=\max\left(\frac{\epsilon-\lambda_{\min}\left(\widetilde{\boldsymbol{R}}\right)}{1-\lambda_{\min}\left(\widetilde{\boldsymbol{R}}\right)},0\right),$$

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 λ_0 is automatically set to zero. Conversely, if $\lambda_{\min}\left(\widetilde{\boldsymbol{R}}\right)\leq 0$, then λ_0 is set to the smallest possible value that ensures positivity of $\lambda_{\min}\left(\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\lambda_0)\right)$.

Second, we implement the optimisation procedure. In our simulation study and empirical applications we employ a grid search for $\lambda^* = \{\lambda : \lambda_0 + \epsilon \leq \lambda \leq 1\}$ with increments of 0.005. The final λ^* is given by

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

When $\lambda_0 = 0$ we still implement shrinkage to find the optimal shrinkage parameter (which might not be $\lambda^* = 0$).

Appendix B Cross validation for BL and CL

BL and CL cross validation with FLM extension: We perform a grid search for the choice of C over a specified range: $C = \{c : C_{\min} \leq c \leq C_{\max}\}$. In 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}, $i = 1, ..., N$, $t = 1, ..., T$ and select the $N \times 1$ column vectors $\boldsymbol{x}_t = (x_{1t}, ..., x_{Nt})'$, $t = 1, ..., T$ which we randomly reshuffle over the t-dimension. This gives rise to a new set of $N \times 1$ column vectors $\boldsymbol{x}_t^{(s)} = \left(x_{1t}^{(s)}, ..., x_{Nt}^{(s)}\right)'$ 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 $s = 1, ..., S$, we divide $\boldsymbol{x}^{(s)} = \left(\boldsymbol{x}_1^{(s)}, ..., \boldsymbol{x}_T^{(s)}\right)$ 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 $\hat{\boldsymbol{\Sigma}}_1^{(s)} = \left(\hat{\sigma}_{1,ij}^{(s)}\right)$, with elements $\hat{\sigma}_{1,ij}^{(s)} = T_1^{-1} \sum_{t=1}^{T_1} x_{it}^{(s)} x_{jt}^{(s)}$, and $\hat{\boldsymbol{\Sigma}}_2^{(s)} = \left(\hat{\sigma}_{2,ij}^{(s)}\right)$ with elements $\hat{\sigma}_{2,ij}^{(s)} = T_2^{-1} \sum_{t=T_1+1}^{T} x_{it}^{(s)} x_{jt}^{(s)}$, $i, j = 1, ..., N$, denote the sample covariance matrices generated using T_1 and T_2 respectively, for each split s. We threshold $\hat{\boldsymbol{\Sigma}}_1^{(s)}$ as in (21) or (23) using $I(.)$ as 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 (23) becomes

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

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,

$$\hat{G}(c) = \frac{1}{S} \sum_{s=1}^{S} \left\| \widetilde{\boldsymbol{\Sigma}}_1^{(s)}(c) - \widetilde{\boldsymbol{\Sigma}}_2^{(s)} \right\|_F^2, \tag{B.16}$$

for each c and

$$\hat{C} = \arg \min_{C_{\min} \leq c \leq C_{\max}} \hat{G}(c). \tag{B.17}$$

If several values of c attain the minimum of (B.17), then $\hat{C}$ is chosen to be the smallest one. The final estimator of the covariance matrix is then given by $\widetilde{\boldsymbol{\Sigma}}_{\hat{C}}$. The thresholding approach does not necessarily ensure that the resultant estimate, $\widetilde{\boldsymbol{\Sigma}}_{\hat{C}}$, is positive definite. To ensure that the threshold estimator is positive definite FLM (2011, 2013) propose setting a lower bound on the cross validation grid for the search of C such that $\lambda_{\min}\left(\widetilde{\boldsymbol{\Sigma}}_{\hat{C}}\right) > 0$. Therefore, we modify (B.17) so that

$$\hat{C}^* = \arg \min_{C_{pd} + \epsilon \leq c \leq C_{\max}} \hat{G}(c), \tag{B.18}$$

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

References

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

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71, 529-626.

Anderson, T. W. (2003). An introduction to multivariate statistics. Wiley, 3rd edition, New York.

Antoniadis, A. and Fan, J. (2001). Regularised wavelet approximations. Journal of American Statistical Association, 96, 939-967.

Bai, J. (2003). Inferential theory for factor models of large dimensions. Econometrica, 71, 1, 135-171.

Bailey, N., Holly, S. and Pesaran, M. H. (2013). A two stage approach to spatiotemporal analysis with strong and weak cross-sectional dependence. Center for Applied Financial Economics (CAFE) research paper no. 14.01.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B, 64, 253-280.

Barndor-Nielsen, O. E. and Shephard N. (2004). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. Econometrica, 72, 885-925.

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 regularization by thresholding. The Annals of Statistics, 36, 2577-2604.

Bien, J. and Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. Biometrika, 98, 807-820.

Bonferroni, C. (1935). Il Calcolo delle Assicurazioni su Gruppi di Teste. Studi in Onore del Professore Salvatore Ortu Carboni, Rome, Italy, 13–60.

Bonferroni, C. (1936). Teoria Statistica delle Classi e Calcolo delle Probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8, 3-62.

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

Cai, T., Liu, W. and Luo, X. (2011). A constrained ℓ_1 minimisation approach to sparse precision matrix estimation. Journal of American Statistical Association, 106, 494, 594-607.

Cai, T., and Zhou, H. (2011). Minimax estimation of large covariance matrices under ℓ_1 norm (with discussion). Statistica Sinica, 22, 1319-1378.

Cai, T. , and Zhou, H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. The Annals of Statistics, 40, 2389-2420.

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

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.

Chudik, A., Pesaran, M.H. and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. The Econometrics Journal, 14, C45-C90.

Daniels, M. J. and Kass, R. (1999). Nonconjugate Bayesian estimation of covariance matrices in hierarchical models. Journal of 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.

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. The Annals of Statistics, 13, 1581-1591.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Pickard, D. (1995). Wavelet shrinkage: Asymptopia? (with discussion). Journal of Royal Statistical Society, Series B, 57, 301-369.

Efron, B. (2010). Large-scale inference: Empirical Bayes methods for estimation, testing and prediction. Cambridge. Cambridge University Press.

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

Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. Journal of Business Economics & Statistics, 20, 339-350.

Fama, E. F. and French, K.R. (1997). Industry costs of equity. Journal of Financial Economics, 43, 153-193.

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

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

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

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

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coe∕ cient in samples from an indefinitely large population. Biometrika, 10, 507-521.

Friedman J., Hastie T. and Tibshirani R. (2008). Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9, 432-441.

Guillot, D. and Rajaratnam, B. (2012). Retaining positive definiteness in thresholded matrices. Linear Algebra and its Applications, 436, 4143-4160.

Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. The Annals of Statistics, 8, 586-597.

Haff, L. R. (1991). The variational form of certain Bayes estimators. The Annals of Statistics, 19, 1163-1190.

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.

Huang, J., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. Biometrika, 93, 85–98.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrices estimation. The Annals of Statistics, 37, 4254-4278.

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.

Lin, S. P. and Perlman, M. D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In Multivariate Analysis, 6, Ed. P. R. Krishnaiah, Amsterdam: North-Holland, 411-429.

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, H. M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, 74, 4, 967-1012.

Pesaran, M. H. (2013). Testing weak cross-sectional dependence in large panels. Forthcoming in Econometric Reviews.

Pesaran, M. H. and Pesaran, B. (2010). Conditional volatility and correlations of weekly returns and the VaR analysis of 2008 stock market crash. Special issue of Economic Modelling in honor of PAVB Swamy. Edited by Stephen G. Hall, Lawrence R. Klein, George S. Tavlas and Arnold Zellner, 27, 1398-1416.

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

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

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

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

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

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-dimentional 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.

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.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. Management Science, 9, 277–293.

Soper, H. E. (1913). On the probable error of the correlation coefficient to a second approximation. Biometrika, 9, 91-115.

Soper, E. H., Young, A. W., Cave, B. M., Lee, A. and Pearson, K. (1917). On the distribution of the correlation coefficient in small samples. Appendix II to the Papers of 'Student'and R. A. Fisher, Biometrika, 11, 328-413.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I 197-206. Univ. California Press, Berkeley.

Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high-frequency financial data. The Annals of Statistics, 38, 943-978.

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.

Yuan, T. and Wang, J. (2013). A coordinate descent algorithm for sparse positive definite matrix estimation. Statistical Analysis and Data Mining, 6, 431-442.