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

Natalia Bailey
Queen Mary, University of London

M. Hashem Pesaran
USC and Trinity College, Cambridge

L. Vanessa Smith
University of York

March 15, 2014

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 approaches in the literature in a detailed Monte Carlo simulation study. Results show that the estimates of the covariance matrix based on MT procedure perform well in a number of different settings and tend to outperform other estimators proposed in the literature, particularly when the cross-sectional dimension, N, is larger than the time series dimension, T. Finally, we investigate the relative performance of the proposed estimators in the context of two important applications in empirical finance when $N \gg T$, namely testing the CAPM hypothesis and optimising the asset allocation of a risky portfolio. For this purpose the inverse covariance matrix is of interest and we recommend a shrinkage version of the MT estimator that ensures positive definiteness.

JEL Classifications: C13,C58

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

*Financial support from the ESRC under grant ES/I031626/1 is gratefully acknowledged. We would like to thank Jianqing Fan, Elizaveta Levina and Martina Mincheva for helpful email correspondence with regard to implementation of their approaches.

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); Pesaran and Zaffaroni (2009)), risk management (Fan, Fan and Lv (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, Schuermann and Weiner (2004); Dees, di Mauro, Pesaran, and Smith (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, Bickel, Levina and Zhu (2010), Abadir, Distaso and Zikes (2012), among others). Others include banding or tapering methods as proposed by Bickel and Levina (2004, 2008a) and Wu and Pourahmadi (2009), which are better suited to the analysis of longitudinal data as they take advantage of the natural *ordering* of the underlying observations. On the other hand, popular methods of regularisation of $\hat{\mathbf{\Sigma}}$ exist in the literature that do not make use of such ordering assumptions. These include the two broad approaches of shrinkage and thresholding.[1]

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 its variance. In the context of variance-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-French (1997) market based covariance matrix specifications as targets.[2] Ledoit and Wolf (2004) 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, Johstone, Kerkyacharian and Pickard (1995). On the whole shrinkage estimators are considered to be stable, robust and produce positive definite covariance matrices by construction. However, they focus on shrinking the over-dispersed sample covariance eigenvalues but not the corresponding eigenvectors which are also inconsistent under shrinkage (Johnstone and Lu (2004)). Further, their implemetation involves weights that themselves depend on unknown parameters.

Thresholding is an alternative regularisation technique that involves setting off-diagonal ele-

[1] See Pourahmadi (2011) for an extensive review of general linear models (GLS) and regularisation based methods for estimation of the covariance matrix.

[2] 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), Hoff (2009) and Fan, Fan and Lv (2008), among others.

ments of the sample covariance matrix that are in absolute value below certain 'threshold' value(s), to zero. This approach includes 'universal' thresholding put forward by Bickel and Levina (2008b), and 'adaptive' thresholding proposed by Cai and Liu (2011). Universal thresholding applies the same thresholding parameter to all off-diagonal elements of the unconstrained sample covariance matrix, while adaptive thresholding allows the threshold value to vary across the different off-diagonal elements of the matrix. Also, the selected non-zero elements of $\hat{\mathbf{\Sigma}}$ can either be set at their sample estimates or somewhat adjusted downward. These relate 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 the 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.[3] The thresholding method retains symmetry of the sample variance-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 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.[4]

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 multiple testing problem 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 the number of non-diagonal elements in each row of $\boldsymbol{R}$ that are non-zero, which is assumed to be bounded in N.

In many applications, including those to be considered in this paper, an estimate of the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ is required. Since traditional thresholding does not necessarily lead to a positive definite matrix, 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 Σ^{-1}. See, for example, Efron (1975), D'Aspremont, Banerjee and Elghaoui (2008), Rothman, Bickel, Levina, and Zhu (2008), Yuan and Lin (2007), and Peng, Wang, Zhou and Zhu (2009). The existing approaches are rather complex and computationally extensive. In addition, though convergence to Σ^{-1} is achieved in this manner, the same methods can not be used to estimate a reliable $\mathbf{\Sigma}$ estimate. If both the covariance matrix and its inverse are of interest, the shrinkage approach is to be recommended. We propose a shrinkage estimator, $\hat{\boldsymbol{R}}_{LW}$, that is applied to the sample correlation matrix $\hat{\boldsymbol{R}}$ rather than $\hat{\mathbf{\Sigma}}$ in order to

[3] 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.

[4] Other contributions to the thresholding literature include the work of Huang, Liu, Pourahmadi, and Liu (2006), Rothman, Levina and Zhu (2009), Cai and Zou (2009, 2010), and Wang and Zou (2010), among others.

avoid distortions to its variance components. It is motivated by the work of Schäfer and Strimmer (2005), which in turn is based on the theoretical results of Ledoit and Wolf (2003). Their procedure, however, ignores the bias of the empirical correlation coefficients which we take into account in the case of our proposed estimator, $\hat{\boldsymbol{R}}_{LW}$. This estimator can also be used in conjunction with our multiple testing approach. In light of this, we consider a supplementary shrinkage estimator applied to our regularised MT correlation matrix. In this case, the shrinkage parameter is derived from the minimisation of the squared Frobenius norm of the difference between two inverse matrices: a recursive estimate of the inverse matrix of interest (which we take as the MT estimator), and the inverse of a suitable reference matrix (which we take to be $\hat{\boldsymbol{R}}_{LW}$). This supplementary shrinkage estimator will be denoted by S-MT.

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

We further evaluate and compare the multiple testing approach to existing thresholding and shrinkage techniques, in the context of testing the Fama-French (2004) capital asset pricing model (CAPM) and implementing portfolio optimization using a similar factor setting. For this purpose we make use of the S-MT estimator as the inverse of the estimated covariance matrix is required. Key challenges in tackling these problems are explored and discussed.

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 $\hat{\boldsymbol{R}}_{LW}$ and S-MT 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. Applications to testing the Fama- French CAPM and portfolio optimization are found in Section 7. Finally Section 8 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 Frobrenius 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 i, x_{it} is covariance stationary over t, and for each t, x_{it} is cross-sectionally weakly dependent (CWD), as defined in Chudik, Pesaran and Tosetti (2011). For each $t \in T$ the variance-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 $E(\boldsymbol{x}_t) = \mathbf{0}$, $\boldsymbol{\Sigma}_t$ is an $N \times N$ symmetric, positive definite real matrix with its $(i,j)^{th}$ element, $\sigma_{ij,t}$, given by

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

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 $\mathbf{A}$ is approximately sparse if, for some $q \in [0,1)$,*

$$m_N = \max_{i \leq N} \sum_{j \leq N} |a_{ij,t}|^q$$

does not grow too fast as $N \to \infty$. Exact sparseness is established when setting $q = 0$. Then, $m_N = \max_{i \leq N} \sum_{j \leq N} I\left(a_{ij,t} \neq 0\right)$ is the maximum number of non-zero elements in each row and is bounded in N, where $I\left(.\right)$ denotes the indicator function.

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

The estimation of $\boldsymbol{\Sigma}_t$ gives rise to three main challenges: the sample $\boldsymbol{\Sigma}_t$ becomes firstly ill-conditioned and secondly non-invertible as N increases relative to T, and thirdly $\boldsymbol{\Sigma}_t$ 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 $\boldsymbol{\Sigma}_t$ 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, Bollerslev, Diebold and Labys (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 Σ_t. We suppress subscript t in $\mathbf{\Sigma}_t$ and $\boldsymbol{D}_t$ and evaluate the sample variance-covariance matrix estimator of $\mathbf{\Sigma}$, denoted by $\hat{\mathbf{\Sigma}}$, with elements

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

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

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 directly to the sample correlation matrix. This ensures the preservation of the variance components of $\hat{\mathbf{\Sigma}}$ upon transformation, which is imperative when considering portfolio optimisation and risk management. Our method is invariant to the ordering of the variables under consideration, it is computationally efficient and suitable for application in the case of high frequency observations making it considerably robust to changes in $\mathbf{\Sigma}$ over time.

Recall the cross-sectionally weakly correlated units x_{it}, $i = 1, ..., N$, $t = 1, ..., T$, with sparse variance-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 $\mathbf{\Sigma}$ given by

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

with

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

where σ_{ij} is given in (2). The reasons for opting to work with the correlation matrix rather than its covariance counterpart are primarily twofold. First, the main diagonal of $\boldsymbol{R}$ is set to unity element-wise by construction. This implies that the transformation of $\boldsymbol{R}$ back to $\mathbf{\Sigma}$ leaves the diagonal elements of $\mathbf{\Sigma}$ unaffected, 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 threholding estimator developed by Cai and Lui (2011).[5]

We proceed to the sample correlation matrix, $\hat{\boldsymbol{R}} = (\hat{\rho}_{ij})$,

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

[5] Both approaches are outlined in Section 5.

with elements

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} = \frac{\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(x_{jt}-\bar{x}_j)}{\left(\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2\right)^{1/2}\left(\sum_{t=1}^{T}(x_{jt}-\bar{x}_j)^2\right)^{1/2}},\ i=1,2,...,N,\ t=1,2,...,T, \tag{5}$$

where $\hat{\sigma}_{ij}$ and $\hat{\sigma}_{ii}$ are given in (4).

Then, assuming that for sufficiently large T the correlation coefficients $\hat{\rho}_{ij}$ are approximately normally distributed as

$$\hat{\rho}_{ij} \sim N\left(\mu_{ij}, \omega_{ij}^2\right), \tag{6}$$

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 $\mathbf{R}$,

$$\tilde{\boldsymbol{R}}_{MT} = \left(\tilde{\rho}_{ij}\right) = \left[\hat{\rho}_{ij} I(\sqrt{T}\left|\hat{\rho}_{ij}\right| > b_N)\right],\ i=1,2,...,N-1,\ j=i+1,...,N. \tag{7}$$

where[6]

$$b_N = \Phi^{-1}\left(1-\frac{p}{2f(N)}\right). \tag{8}$$

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 size, p, selected for the joint testing problem. It is clear that for relatively large T, $T\hat{\rho}_{ij}^2 \sim \chi_1^2$.[7] The size of the test is normalised by $f(N)$. This controls the size correction that is imposed on the individual tests in (7). We explain the reasoning behind the choice of $f(N)$, in what follows.

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 multiple tests being conducted simultaneously across the distinct elements of $\hat{\boldsymbol{R}}$. The overall size of the test can then suffer from distortions and needs to be controlled.

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)

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

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

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

[6] The indicator function $I(.)$ used in (7), is in line with the concept of 'hard' thresholding. Hard thresholding implies that all elements of $\hat{\boldsymbol{\Sigma}}$ or $\hat{\boldsymbol{R}}$ that drop below a certain level in absolute terms are set to zero and 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 Antoniadis and Fan (2001), Rothman, Levina and Zhu (2009), and Cai and Liu (2011) among others, or the smoothly clipped absolute deviation (SCAD) approach - see Fan (1997), and Fan and Li (2001).

[7] Note that in place of $\hat{\rho}_{ij}$, $i,j=1,...,N$ one can also use the Fisher transformation of $\hat{\rho}_{ij}$ which could provide a closer approximation to the normal distribution. But our simulation results suggest that in our application there is little to choose between using $\hat{\rho}_{ij}$ or its Fisher's transform.

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

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

Hence to achieve $FWER_T \leq p$, it is sufficient to set $p_{iT} \leq p/r$. Bonferroni's procedure can be quite conservative and tends to lack power. An alternative step-down procedure is proposed by Holm (1979) which is more powerful than Bonferroni's procedure, without imposing 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 (7) 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 $\left|\hat{\rho}_{(s)}\right| > \Phi^{-1}\left(1 - \frac{p/2}{r-s+1}\right)$, otherwise disconnected, for $s = 1, 2, ..., r$, where p is the pre-specified overall size of the test.[8] 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 $\left|\hat{\rho}_{ij}\right| > \Phi^{-1}\left(1 - \frac{p/2}{N(N-1)/2}\right)$.

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 of subsection (3.1) show that using $f(N) = N(N-1)/2$ in (8) rather than $f(N) = (N-1)$ as $N \to \infty$, provides a faster rate of convergence towards R under the Frobenius norm. However, simulation results of Section 6 indicate that in finite samples $f(N) = (N-1)$ can provide $\tilde{\boldsymbol{R}}_{MT}$ 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

[8] In the Monte Carlo experiments we consider both $p = 0.05$ and 0.10, but find that the MT method is reasonably robust to the choice of p.

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 $\tilde{\boldsymbol{R}}_{MT}$ 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 $\hat{\boldsymbol{R}}$.[9]

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

$$\tilde{\boldsymbol{\Sigma}}_{MT}=\hat{\boldsymbol{D}}^{1/2}\tilde{\boldsymbol{R}}_{MT}\hat{\boldsymbol{D}}^{1/2}. \tag{9}$$

It is evident that since b_N is given and does not need to be estimated, the multiple testing procedure of (7) is also computationally efficient. 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 variance 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 $\hat{\boldsymbol{\Sigma}}_{\hat{u}}$.[10] In this case, (7) is modified to correct for the degrees of freedom, m, associated with the defactoring regression:

$$\tilde{\rho}_{\hat{u},ij} = \hat{\rho}_{\hat{u},ij} I(\sqrt{T-m}\left|\hat{\rho}_{\hat{u},ij}\right| > b_N),\ i = 1, 2, ..., N-1,\ j = i+1, ..., N \tag{10}$$

where

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

For an empirical application of the multiple testing approach using defactoring and the Holm procedure, see also Bailey, Holly and Pesaran (2013).

3.1 Theoretical properties of the MT estimator

In this subsection we investigate the asymptotic properties of the MT estimator defined in (7). 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 $\hat{\mathbf{R}}= (\hat{\rho}_{ij})$ be the sample correlation matrix, and suppose that (for sufficiently large T)*

$$\hat{\rho}_{ij} \sim N\left(\mu_{ij}, \omega_{ij}^2\right), \tag{11}$$

where

$$\mu_{ij} = E(\hat{\rho}_{ij}) = \rho_{ij} - \frac{\rho_{ij}(1-\rho_{ij}^2)}{2T} + \frac{G(\rho_{ij})}{T^2}, \tag{12}$$

$$\omega_{ij}^2 = Var(\hat{\rho}_{ij}) = \frac{(1-\rho_{ij}^2)^2}{T} + \frac{K(\rho_{ij})}{T^2}, \tag{13}$$

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

[9] Other multiple testing procedures can also be considered and Efron (2010) provides 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.

[10] We consider an example of multiple testing on regression residuals in our simulation study of Section 6.

The analytical expressions for the mean and variance of $\hat{\rho}_{ij}$ in (12) and (13) of Assumption 1 can be found in Soper, Young, Cave, Lee and Pearson (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 (or the sum of their absolute values is bounded in N).

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 $\mathbf{R} = (\rho_{ij})$, 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

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

if $N/T \to 0$ as N and $T \to \infty$, in any order, where $\tilde{\boldsymbol{R}}_{MT} = (\tilde{\rho}_{ij})$

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

and p is a given overall Type I error.
Proof. *See Appendix A.* ■

Result (14) implies that $\left\|\tilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F = O_p(\sqrt{\frac{m_N N}{T}})$ which proves $(m_N N)^{-1/2} T^{1/2}$ -consistency of the multiple testing correlation matrix estimator $\tilde{\boldsymbol{R}}_{MT}$ under the Frobenius 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*

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

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

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 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 which is essential in a number of empirical applications including the ones considered in Section 7. Bickel and Levina (2008b) provide an asymptotic condition that ensures positive definiteness.[11] In their theoretical work Guillot and Rajaratnam (2012) demonstrate in more depth 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. Apart from the penalised likelihood approach to tackle this problem as mentioned in the introduction, more recent contributions propose a sparse positive definite covariance estimator obtained via convex optimisation, where sparseness is achieved by use of a suitable penalty. Rothman (2012) uses a logarithmic barrier term, Xue, Ma and Zou (2012) impose a positive definiteness constraint, while Liu, Wang and Zhao (2013) and Fan, Liao and Mincheva (2013) enforce an eigenvalue condition.[12],[13] These approaches are rather complex and computationally quite extensive. Instead, if inversion of $\hat{\boldsymbol{R}}$ or $\hat{\boldsymbol{\Sigma}}$ is of interest, we recommend the use of LW type shrinkage estimator, but applied to the sample correlation, $\hat{\boldsymbol{R}}$, or the MT estimated correlation matrix, $\tilde{\boldsymbol{R}}_{MT}$. This is motivated by the work of Schäfer and Strimmer (2005) and the theoretical results of Ledoit and Wolf (2003). However, in Schäfer and Strimmer (2005) the bias of the empirical correlation coefficients is ignored, which we will take into account in our specification of $\hat{\boldsymbol{R}}_{LW}$. Compared with the Ledoit and Wolf (2004) shrinkage covariance estimator, $\hat{\boldsymbol{R}}_{LW}$ has the advantage of retaining the diagonal of $\hat{\boldsymbol{\Sigma}}$ which is important in finance applications for instance, where the diagonal elements of $\hat{\boldsymbol{\Sigma}}$ correspond to volatilities of asset returns.[14]

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

$$\hat{\boldsymbol{R}}_{LW} = \xi \boldsymbol{I}_N + (1-\xi)\hat{\boldsymbol{R}}, \tag{17}$$

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

$$\begin{aligned} \left\| \hat{\boldsymbol{R}}_{LW}(\xi) - \boldsymbol{R} \right\|_F^2 &= \sum_{i \neq j}\sum \left[(1-\xi)\hat{\rho}_{ij} - \rho_{ij} \right]^2 \\ &= \sum_{i \neq j}\sum \left[\hat{\rho}_{ij} - \rho_{ij} - \xi\hat{\rho}_{ij} \right]^2. \end{aligned} \tag{18}$$

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*

$$N^{-1}E\left\| \hat{\boldsymbol{R}}_{LW}(\xi^*) - \boldsymbol{R} \right\|_F^2 = N^{-1}\sum_{i \neq j}\sum E\left[\hat{\rho}_{ij} - \rho_{ij} - \xi^*\hat{\rho}_{ij} \right]^2 = O\left(\frac{N}{T}\right), \tag{19}$$

[11] See Section 5 for the exact specification of this condition.

[12] Other related work includes that of Lam and Fan (2009), Rothman, Levina and Zhu (2009), Bien and Tibshirani (2011), Cai, Liu and Luo (2011), and Yuan and Wang (2013).

[13] We implement the method of Fan, Liao and Mincheva (2013) in our simulation study of Section 6. More details regading this method can be found in Section 5 and Appendix B.

[14] We discuss the effect of distorting the size of asset return volatilities in the context of portfolio optimisation in Section 7.

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

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

Proof. See Appendix A. ■

In deriving the results of Theorem 3 we follow Ledoit and Wolf (2004, LW) and use the scaled Frobenius norm, where $\|\boldsymbol{A}\|^2 = tr(\boldsymbol{A}'\boldsymbol{A})/N$ for a N-dimensional matrix $\boldsymbol{A}$, (see Definition 1 pp. 376 of LW).

Corollary 1

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

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 variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{LW}(\xi^*)$ is performed as in (23).

The shrinkage estimator $\hat{\boldsymbol{R}}_{LW}$ can also be used as a supplementary tool to achieve invertibility of our multiple testing estimator. Using a shrinkage parameter derived through a grid search optimisation procedure described below, positive definiteness of $\tilde{\boldsymbol{R}}_{MT}$ is then guaranteed.

Following Ledoit and Wolf (2004)[15], we set as benchmark target the $N \times N$ identity matrix $\boldsymbol{I}_N$. Then, our shrinkage on multiple testing (S-MT) estimator is defined by

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

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

First note that shrinkage is again deliberately implemented on the correlation matrix $\tilde{\boldsymbol{R}}_{MT}$ rather than on $\tilde{\boldsymbol{\Sigma}}_{MT}$. In this way we ensure that no shrinkage is applied to the volatility measures. Second, the shrinkage is applied to non-zero elements of $\tilde{\boldsymbol{R}}_{MT}$, and as a result the shrinkage estimator, $\tilde{\boldsymbol{R}}_{S\text{-}MT}$, has the same optimal non-zero/zero patterns achieved for $\tilde{\boldsymbol{R}}_{MT}$. This is in contrast to approaches that impose eigenvalue restrictions to achieve positive definiteness.

The criterion used in the final selection of the shrinkage parameter in (20) 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, $\tilde{\boldsymbol{R}}_{MT}(\lambda)$ is evaluated. Since both $\boldsymbol{R}_0$ and $\tilde{\boldsymbol{R}}_{MT}(\lambda)$ 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

$$\lambda^* = \arg \min_{\lambda_0 \leq \lambda \leq 1} \left\| \boldsymbol{R}_0^{-1} - \tilde{\boldsymbol{R}}_{MT}^{-1}(\lambda) \right\|_F^2. \tag{21}$$

[15] This approach is summarised in Section 5.

Let $\boldsymbol{A}=\boldsymbol{R}_0^{-1}$ and $\boldsymbol{B}(\lambda)=\tilde{\boldsymbol{R}}_{MT}^{-1}(\lambda)$. Note that since $\boldsymbol{R}_0$ and $\tilde{\boldsymbol{R}}_{MT}$ are symmetric

$$\left\|\boldsymbol{R}_0^{-1}-\tilde{\boldsymbol{R}}_{MT}^{-1}(\lambda)\right\|_F^2=tr\left(\boldsymbol{A}^2\right)-2tr[\boldsymbol{A}\boldsymbol{B}(\lambda)]+tr[\boldsymbol{B}^2(\lambda)]. \tag{22}$$

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

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

where

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

Hence, λ^* is given by the solution of

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

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 $\tilde{\boldsymbol{R}}_{S\text{-}MT}(\lambda^*)$ is guaranteed to be positive definite. This follows from Banerjee, Ghaoui and D'Asperton (2008) who show that if a recursive procedure is initialised with a positive definite matrix, then the subsequent iterates remain positive definite. For more details of the above derivations and the grid search optimisation procedure see Appendix A.

Having obtained the shrinkage estimator $\tilde{\boldsymbol{R}}_{S\text{-}MT}$ using λ^* in (20), we construct the corresponding covariance matrix as:

$$\tilde{\boldsymbol{\Sigma}}_{S\text{-}MT}=\hat{\boldsymbol{D}}^{1/2}\tilde{\boldsymbol{R}}_{S\text{-}MT}\hat{\boldsymbol{D}}^{1/2}. \tag{23}$$

An important aspect of the above method is the choice of the reference matrix $\boldsymbol{R}_0$. In our simulation study of Section 6 we considered various choices for the reference correlation matrix. These included the identity matrix, the generalised inverse of the sample correlation matrix, the correlation matrix derived from shrinking $\hat{\boldsymbol{\Sigma}}$ using the Ledoit and Wolf (2004) method and our proposed $\hat{\boldsymbol{R}}_{LW}$ shrinkage approach described above. Our results showed that $\hat{\boldsymbol{R}}_{LW}$ offers superior performance for $\tilde{\boldsymbol{R}}_{S\text{-}MT}$ in finite samples relative to the other reference matrices. For further details see Section 6.

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 the 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 $\hat{\boldsymbol{\Sigma}}=(\hat{\sigma}_{ij})$, $i,j=1,...,N$. Under this approach $\boldsymbol{\Sigma}$ is required to be sparse according to Definition 1. The BL thresholding estimator is given by

$$\tilde{\boldsymbol{\Sigma}}_{BL,C}=\left(\hat{\sigma}_{ij}I\left[|\hat{\sigma}_{ij}|\geq C\sqrt{\frac{\log N}{T}}\right]\right),\ i=1,2,...,N-1,\ j=i+1,...,N \tag{24}$$

where $I(.)$ is an indicator function and C is a positive constant which is unknown. The choice of thresholding function - $I(.)$ - implies that (24) implements 'hard' thresholding. The consistency rate of the BL estimator is $\sqrt{\frac{\log N}{T}}$ under the spectral norm of error matrix $\left(\tilde{\mathbf{\Sigma}}_{BL,C}-\mathbf{\Sigma}\right)$. The main challenge in the implementation of this approach is the estimation of the thresholding parameter, C, which is usually calibrated by cross validation. [16]

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

$$\left\|\tilde{\mathbf{\Sigma}}_{BL,C}-\tilde{\mathbf{\Sigma}}_{BL,0}\right\| \leq \varepsilon \text{ and } \lambda_{\min}(\mathbf{\Sigma}) > \varepsilon, \tag{25}$$

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 to 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 $\tilde{\mathbf{\Sigma}}_{CL,C}$ is defined as

$$\tilde{\mathbf{\Sigma}}_{CL,C} = \left(\hat{\sigma}_{ij} I\left[|\hat{\sigma}_{ij}| \geq \tau_{ij}\right]\right),\ i=1,2,...,N-1,\ j=i+1,...,N \tag{26}$$

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 $C>0$. The consistency rate of the CL estimator is $\sqrt{\frac{\log N}{T}}$ under the spectral norm of the error matrix $\left(\tilde{\mathbf{\Sigma}}_{CL,C}-\mathbf{\Sigma}\right)$. Parameter C can be fixed to a constant implied by theory ($C=2$ in CL) or chosen via cross validation.[17] Similar concerns to BL regarding cross validation also apply here.

As with the BL estimator, thresholding in itself does not ensure positive definiteness of $\tilde{\mathbf{\Sigma}}_{CL,\hat{C}}$. In light of condition (25), Fan, Liao and Mincheva (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, $\lambda_{\min}\left(\tilde{\mathbf{\Sigma}}_{FLM,\hat{C}}\right)>0$ - for more details see Appendix B. We apply this extension to both BL and CL procedures. The problem of $\tilde{\mathbf{\Sigma}}_{BL,\hat{C}}$ and $\tilde{\mathbf{\Sigma}}_{CL,\hat{C}}$ not being invertible in finite samples is then resolved. However, depending on the application, the selected C might not necessarily be optimal (see Appendix B for the relevant expressions). In other words, the properties of the constrained $\tilde{\mathbf{\Sigma}}_{BL,\hat{C}}$ and $\tilde{\mathbf{\Sigma}}_{CL,\hat{C}}$ can deviate noticeably from their respective unconditional versions.

[16] See Appendix B for details of the BL cross validation procedure. Further, Fang, Wang and Feng (2013) provide useful guidelines regarding the specification of various parameters used in cross-validation through an extensive simulation study.

[17] See Appendix B for details of the CL cross validation procedure.

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, $\hat{\boldsymbol{\Sigma}}$, and an identity matrix $\boldsymbol{I}_N$, and provide formulae for the appropriate weights. The LW shrinkage is expressed as

$$\hat{\boldsymbol{\Sigma}}_{LW} = \hat{\rho}_1 \boldsymbol{I}_N + \hat{\rho}_2 \hat{\boldsymbol{\Sigma}}, \tag{27}$$

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

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

and

$$\begin{aligned} \bar{b}_T^2 &= \frac{1}{T^2} \sum_{t=1}^{T} \left\| \dot{\boldsymbol{x}}_t \dot{\boldsymbol{x}}_t' - \hat{\boldsymbol{\Sigma}} \right\|_F^2 \\ &= \frac{1}{NT^2} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \dot{x}_{it}^2 \right)^2 - \frac{1}{NT} tr\left(\hat{\boldsymbol{\Sigma}}^2\right), \end{aligned}$$

with $\dot{\boldsymbol{x}}_t = (\dot{x}_{1t}, ..., \dot{x}_{Nt})'$ and $\dot{x}_{it} = (x_{it} - \bar{x}_i)$. Note that the Frobenius norm $\|\boldsymbol{A}\|_F^2 = tr(\boldsymbol{A}'\boldsymbol{A})/N$ is scaled by N which is not standard. Also, $\hat{\boldsymbol{\Sigma}}_{LW}$ is positive definite by construction. Thus, the inverse $\hat{\boldsymbol{\Sigma}}_{LW}^{-1}$ 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 $\boldsymbol{\Sigma}$, as captured by the choices of ρ_1 and ρ_2. Note however that LW do not require these paramters to add up to unity, and it is possible for the shrinkage method to place little weight on the data (ie the correlation matrix). Of particular importance is the effect that LW shrinkage has on the diagonal elements of $\hat{\boldsymbol{\Sigma}}$ 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. Further, even though shrinkage adjusts the over-dispersion of the unconstrained covariance eigenvalues, it does not correct the corresponding eigenvectors which are also inconsistent (Johnstone and Lu (2004)). But unlike the thresholding approaches considered in this paper, the LW methodology does not require $\boldsymbol{\Sigma}$ to be sparse.

6 Small sample properties

We evaluate the small sample properties of our proposed multiple testing (MT) estimator, its positive definite S-MT version and our shrinkage estimator on the sample correlation matrix by use of a Monte Carlo simulation study. For comparative purposes we also report results for the three widely used regularisation approaches described in Section 5. We consider four experiments: (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 *approximate* sparse covariance settings. The latter two are examples of *exact* 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. All experiments are based on $R = 500$ replications.

Experiment A We consider the AR(1) covariance matrix with decaying coefficient, ϕ,

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

with its inverse given by

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

The corresponding correlation matrix $\boldsymbol{R}=\left(\rho_{ij}\right)$ is given by $\boldsymbol{R} = \left(1-\phi^2\right)\boldsymbol{\Sigma}$. For this example, $\boldsymbol{\Sigma}^{-1} = \boldsymbol{Q}'\boldsymbol{Q}$, where

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

Our data generating process is then given by

$$\boldsymbol{Q}\boldsymbol{x}_t^{(r)} = \boldsymbol{\varepsilon}_t^{(r)},\ t = 1, ..., T. \tag{28}$$

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, (28) 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

$$\hat{\boldsymbol{\Sigma}}^{(r)} = T^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{x}}_t^{(r)}\dot{\boldsymbol{x}}_t^{(r)\prime}, \tag{29}$$

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

$$\hat{\boldsymbol{R}}^{(r)} = \hat{\boldsymbol{D}}^{-1/2(r)}\hat{\boldsymbol{\Sigma}}^{(r)}\hat{\boldsymbol{D}}^{-1/2(r)}, \tag{30}$$

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

$$\begin{aligned} \boldsymbol{x}_t^{(r)} &= \vartheta \boldsymbol{W}\boldsymbol{x}_t^{(r)} + \boldsymbol{\varepsilon}_t^{(r)} \\ &= (\boldsymbol{I}_N - \vartheta\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}_t^{(r)},\ t = 1, ..., T, \end{aligned} \tag{31}$$

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 $\mathbf{W}$ is unity and the strength 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 $\boldsymbol{D} = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{NN})$, and

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

Finally,

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

We generate the sample covariance and correlation matrices $\hat{\boldsymbol{\Sigma}}$ and $\hat{\mathbf{R}}$ as in experiment A using (29) and (30).

Experiment C Next we consider a banded matrix with ordering, following Model 1 of Cai and Liu (2011):

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

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 $\boldsymbol{\Sigma}$. As before, we compute the sample covariance and correlation matrices $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{R}}$ using (29) and (30).

Experiment D We analyse 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, where $N_b = \left[N^{\delta}\right]$ and set the remaining middle elements to zero. The resulting population correlation matrix $\boldsymbol{R}$ is given 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

$$\boldsymbol{x}_t^{(r)} = \boldsymbol{Q}\boldsymbol{\varepsilon}_t^{(r)},\ t = 1, ..., T. \tag{32}$$

Again, we compute the sample covariance and correlation matrices $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{R}}$ using (29) and (30).

6.1 Robustness analysis

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

1. We consider a more complex setting where $\boldsymbol{x}_t^{(r)}$ represent the error terms in a regression. 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

$$y_{it}^{(r)} = \delta_i + \gamma_i z_{it}^{(r)} + u_{it}^{(r)},\ \text{for } i = 1, 2, ..., N,\ t = 1, 2, ..., T, \tag{33}$$

   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 $i = 1, 2, ..., N$.[18]

2. 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 (29) and (30). We evaluate our results using the sample covariance matrix.[19]

[18] Note that in this instance the multiple testing approach is corrected for the degrees of freedom. Hence, as in (10) $\sqrt{T}$ is replaced by $\sqrt{T-m}$, where m is equal to the number of regressors in (33) including the intercept.

[19] We also considered using Fisher's z-transformation of the sample correlation coefficients in (30), given by:

$$Z_{ij}^{(r)} = \frac{1}{2}\ln\frac{1 + \hat{\rho}_{ij}^{(r)}}{1 - \hat{\rho}_{ij}^{(r)}},\ i, j = 1, ..., N,$$

for $r = 1, ..., R$. Results were very similar.

6.2 Evaluation metrics: Spectral / Frobenius norms and support recovery

In all experiments we apply the four regularisation techniques described above: (i) our proposed approach of multiple testing (implemented by row and on the full matrix) on $\hat{\boldsymbol{R}}$ ($\tilde{\boldsymbol{\Sigma}}_{MT_R}$, $\tilde{\boldsymbol{\Sigma}}_{MT_F}$), (ii) BL thresholding on $\hat{\boldsymbol{\Sigma}}$ ($\tilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}$), (iii) CL thresholding on $\hat{\boldsymbol{\Sigma}}$ ($\tilde{\boldsymbol{\Sigma}}_{CL,2}$, $\tilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}$), and (iv) LW shrinkage on $\hat{\boldsymbol{\Sigma}}$ ($\hat{\boldsymbol{\Sigma}}_{LW}$), with the regularised covariance matrices given in parenthesis. We also consider (v) LW shrinkage on $\hat{\boldsymbol{R}}$ ($\hat{\boldsymbol{\Sigma}}_{LW}(\xi^*)$) and (vi) shrinkage on multiple testing estimator $\tilde{\boldsymbol{R}}_{MT}$ ($\tilde{\boldsymbol{\Sigma}}_{S\text{-}MT_R}$ or $\tilde{\boldsymbol{\Sigma}}_{S\text{-}MT_F}$). These approaches are evaluated predominantly for comparison with the inverse covariance matrices. For both BL and CL thresholding procedures we further impose the FLM extension which ensures positive definiteness of the estimated matrices. [20]

Where regularisation is performed on the correlation matrix we reconstruct the corresponding covariance matrix in line with (9). We compute the spectral norm of the deviations of each of the regularised covariance matrices from their respective true $\boldsymbol{\Sigma}$ in experiments A-D:

$$\left\|\boldsymbol{A}_{\mathring{\boldsymbol{\Sigma}}}\right\| = \left\|\boldsymbol{\Sigma}-\mathring{\boldsymbol{\Sigma}}\right\|, \tag{34}$$

for $\mathring{\boldsymbol{\Sigma}}=\{\tilde{\boldsymbol{\Sigma}}_{MT_R}, \tilde{\boldsymbol{\Sigma}}_{MT_F}, \tilde{\boldsymbol{\Sigma}}_{S\text{-}MT_R}, \tilde{\boldsymbol{\Sigma}}_{S\text{-}MT_F}, \tilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}, \tilde{\boldsymbol{\Sigma}}_{CL,2}, \tilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}, \hat{\boldsymbol{\Sigma}}_{LW}\}$, where $\hat{C}$ is a constant evaluated through cross-validation - see Appendix B for details. We also evaluate the Frobenius norm of the difference displayed in (34), denoted by $\|.\|_F$. With regard to the behaviour of the inverse covariance matrices we evaluate

$$\left\|\boldsymbol{B}_{\mathring{\boldsymbol{\Sigma}}^{-1}}\right\| = \left\|\boldsymbol{\Sigma}^{-1}-\mathring{\boldsymbol{\Sigma}}^{-1}\right\|, \tag{35}$$

for $\mathring{\boldsymbol{\Sigma}}^{-1}=\{\tilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT_R}, \tilde{\boldsymbol{\Sigma}}^{-1}_{S\text{-}MT_F}, \tilde{\boldsymbol{\Sigma}}^{-1}_{BL,\hat{C}^*}, \tilde{\boldsymbol{\Sigma}}^{-1}_{CL,\hat{C}^*}, \hat{\boldsymbol{\Sigma}}_{LW}(\xi^*), \hat{\boldsymbol{\Sigma}}^{-1}_{LW}\}$, where $\hat{C}^*$ is a constant estimated through cross-validation over a reduced grid suggested by Fan, Liao and Mincheva (2013) (see Appendix B for details). Again, we also calculate the Frobenius norm of the difference displayed in (35).

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

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

Only for experiment C $\left\|\boldsymbol{\Sigma}^{-1}\right\| = O(1)$ is not satisfied, as the population covariance matrix is not invertible.

We report the averages of $\left\|\boldsymbol{A}_{\mathring{\boldsymbol{\Sigma}}}\right\|$, $\left\|\boldsymbol{A}_{\mathring{\boldsymbol{\Sigma}}}\right\|_F$, $\left\|\boldsymbol{B}_{\mathring{\boldsymbol{\Sigma}}^{-1}}\right\|$, and $\left\|\boldsymbol{B}_{\mathring{\boldsymbol{\Sigma}}^{-1}}\right\|_F$ over $R = 500$, except for the BL and CL cross-validation procedures when $N = 400$ where $R = 100$.[21]

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 (15) and (16), respectively. These are only implemented for experiments C and D. Experiments A and B are 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.

6.3 Results

We report results for the covariance matrix estimates based on the reguralisation approaches described in Sections 3 and 5. For convenience we abbreviate these as follows: MT_R and MT_F (multiple testing by row and on the full sample correlation matrix), $S\text{-}MT_R$ and $S\text{-}MT_F$ (shrinkage on

[20] We implement the method of Fan, Liao and Mincheva (2013) in our applications of Section 7. More details regading this method can be found in Section 5 and Appendix B.

[21] For the BL and CL cross-validation procedures, due to their protracted computational time, in the case of $N = 400$ we set the grid increments to 4 and reduced the number of replications to $R = 100$. The latter is in line with the BL and CL simulation specifications.

multiple testing by row and on the full correlation matrix), BL_{CV} (BL thresholding on the sample covariance matrix using cross-validation), BL_{FLM} (BL thresholding using FLM cross-validation adjustment), CL_{CV} (CL thresholding on the sample covariance matrix using cross-validation), CL_{FLM} (CL thresholding using FLM cross-validation adjustment) and $LW_{\hat{R}}$ (LW shrinkage on the sample correlation matrix) and $LW_{\hat{\Sigma}}$ (LW shrinkage on the sample covariance matrix). We employed both the Bonferroni and Holm corrections when implementing our multiple testing approach. For brevity of exposition simulation results are only reported for the latter case. Results using the Bonferroni correction were comparable for all settings and are available upon request.

First we establish robustness of our MT estimator to different levels of significance, p, used in the theoretical derivation of the parameter b_N in (8). We evaluate MT by row and on the full $\hat{\boldsymbol{R}}$ matrix at the 5% and 10% significance levels in experiments A-D. The results in Table 1 show that there is little difference in the numerical values of the average spectral and Frobenius norms of (34) between $MT_{R_{5\%}}$ (or $MT_{F_{5\%}}$) and $MT_{R_{10\%}}$ (or $MT_{F_{10\%}}$) for all N and T combinations and for all covariance matrix setups considered. Our theoretical results of Section 3.1 suggest that asymptotically, use of $f(N) = N(N-1)/2$ in multiple testing produces superior performance than when $f(N) = (N-1)$ is employed. In small samples multiple testing by row appears to perform marginally better in most cases. However, as T and N increase and depending on the experiment, performing multiple testing on the full matrix, MT_F, yields results closer to those based on by row implementation, MT_R, and even outperform them at times - see for example experiment D for $T = 100$. As results are robust to the significance level, we proceed with our analysis considering only multiple testing at the 5% level.

Tables 2-5 summarise results for experiments A to D. In all cases the top panel shows comparative results for the different regularisation estimators. The middle panel presents results for the estimated inverse matrices only for the estimators that address the issue of positive definiteness. 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 itself non-invertible and therefore results for inverse matrix estimates are not meaningful.

Starting with experiment A and focusing initially on the top panel of Table 2 results show that multiple testing and thresholding in general ourperform the shrinkage technique under both norm specifications and especially so as N increases. When T rises from 60 to 100 all regularisation measures perform better (lower values for $\|.\|$ and $\|.\|_F$) which is expected, but MT and thresholding continue to outperform shrinkage. For small N, MT_R, MT_F, BL_{CV}, CL_T and CL_{CV} behave similarly, however as N increases MT_R and MT_F outperform BL_{CV} and CL_T. In general, CL_{CV} performs better than MT_F though the difference between the two diminishes at times for large N. When the positive definite condition is imposed a clear deterioration in the value of the spectral and Frobenius norms is noticeable uniformly across estimators. However, S-MT_R and S-MT_F perform favourably relative to BL_{FLM} and CL_{FLM} across all (N, T) combinations. Finally, adaptive thresholding (CL_{CV} and CL_{FLM}) outperforms universal thresholding (BL_{CV} and BL_{FLM}), which is expected given the heteroskedasticity present in the data. Also, CL using the theoretical thresholding parameter of 2 (CL_T) produces marginally higher norms than its cross-validation based equivalent (CL_{CV}), in line with results in Cai and Liu (2011). Moving on to the middle panel of Table 2, we find that the inverse covariance matrices estimated via S-MT_R and S-MT_F perform much better than those produced using BL_{FLM} and CL_{FLM}. In fact, the average spectral norm of CL_{FLM} includes some sizeable outliers, especially for small N. Still, their more reliable Frobenius norm estimates are higher than those of the shrinkage on multiple testing estimators. Also, though $LW_{\hat{\Sigma}}$ outperforms both S-MT_R and S-MT_F for $N = \{30, 100\}$ and for both T specifications, as N rises to 200 and 400 shrinkage on thresholding appears better behaved. 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, $S\text{-}MT_R$, $S\text{-}MT_F$ and $LW_{\hat{R}}$ apply comparatively more balanced weights on $\boldsymbol{I}$ and $\hat{\mathbf{\Sigma}}$ across the range of (N, T) combinations. Finally, $S\text{-}MT_F$ marginally outperforms $S\text{-}MT_R$ when considering the regularised inverse covariance matrices.

Results for experiments B to D given in Tables 3-5 are on the whole qualitatively similar to those of experiment A, apart from some differences . The values of the spectral and Frobenius norms are lower for these experiments, particularly so for experiments B and D. This implies that the population covariance matrices on which the respective data generating processes are based are themselves better conditioned. Unlike in experiment A, both MT_R and MT_F now not only outperform BL_{CV} universally, but CL_{CV} as well. With regard to the inverse matrix comparisons, again BL_{FLM} and CL_{FLM} display outlier realisations in both cases, more so for smaller N and for both T values considered. 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 experiment D. Overall, the $S\text{-}MT$ approach remains the most appealling and the multiple testing procedure outperforms the remaining estimators across experiments. The superior properties of adaptive thresholding over universal thresholding are again visible. Finally, though $LW_{\hat{\Sigma}}$ is computationally attractive compared to the cross-validation based thresholding approaches its performance still falls short of the equally computationally appealing MT and $S\text{-}MT$. Indeed, it repeatedly shrinks the sample covariance matrix excessively towards the structured identity matrix.

Table 6 presents results for support recovery of $\mathbf{\Sigma}$ using the original multiple testing and thresholding approaches with no adjustments. Superiority of MT_R and MT_F over BL_{CV}, CL_T and CL_{CV} 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 but as N increases they drop, as expected. The only exception is BL_{CV} in experiment D, which shows improvement from $N = 30$ to $N = 100$ for both T specifications. In experiment C the TPRs are lower than in experiment D. The reason for this is that in experiment D we control explicitly for the number of non-zero elements in $\boldsymbol{R}$ and $\mathbf{\Sigma}$ and ensure that they comply to the condition set out in Theorem 2.

Finally, we comment on the results from our robustness analysis (not presented here) applied to experiments A-D. Evaluating the estimated covariance matrices based on the residuals from regression (33) in general produces similar outcomes to the main results of Tables 2-5. In the case of non-normal errors, a deterioration in the values of the average spectral and Frobenius norms is visible across all estimators and experiments. This is not surprising as these methods are based on the assumption of normality of the underlying data. However, MT and $S\text{-}MT$ still outperform the remaining estimators and appear to be more robust to non-normality than the other approaches considered. Results from the robustness analysis can be found in the accompanying supplement.

Overall, both our proposed multiple testing (MT) and shrinkage on multiple testing ($S\text{-}MT$) estimators prove to be robust to changes in the specification of the true covariance matrix $\mathbf{\Sigma}$. If the inverse covariance matrix is of interest $LW_{\hat{R}}$ and $S\text{-}MT$ are more appropriate, while MT gives better covariance matrix estimates when positive definiteness is not required. Furthermore, MT is robust to the choice of significance level p used in the calculation of b_N. Also, $S\text{-}MT$ generates covariance matrix estimates that better reflect the properties of the true covariance matrix $\mathbf{\Sigma}$ than the widely used LW shrinkage approach. Compared with shrinkage on the sample correlation matrix, the relative performance of $LW_{\hat{R}}$ and $S\text{-}MT$ appears to depend on the case studied.

7 An application to CAPM testing and portfolio optimisation

Estimation and regularisation of large covariance matrices has wide applicability in numerous fields as discussed in the introduction. In this section we focus on two applications in the area of finance that rather make use of the inverse population covariance matrix, Σ^{-1}. The first evaluates the limitations of testing a linear asset pricing model when the number of assets is substantially larger than the time dimension. The second is a typical portfolio optimisation exercise in which we apply our proposed shrinkage on multiple testing estimator. We compare our $S\text{-}MT$ estimates with those produced by the approach of Fan, Liao and Mincheva (2011, 2013), which adjusts Cai and Liu (2011) adaptive thresholding for the presence of factors in the data. We also consider the Ledoit and Wolf (2004) shrinkage approach. We do not consider here the BL methodology with the FLM adjustment for positive definiteness, as it is computationally very expensive, especially for the CAPM example given the large number of replications employed in this study. Also, drawing on the results of Section 6 it is likely that BL will underperform the FLM adaptive method since considerable heteroskedasticity exists in these settings. Further, $LW_{\hat{R}}$ largely underperforms $LW_{\hat{\Sigma}}$ for large N in these applications. Therefore, we do not report results for this estimator either but we discuss the implications of the apparent superiority of $LW_{\hat{\Sigma}}$ under specific metrics. In both applications we consider the following data generating process, which reflects the usual Fama-French (2004) model specification,

$$y_{it} = \alpha_i + \sum_{\ell=1}^{m} \beta_{\ell i} f_{\ell t} + \kappa u_{it},\ i = 1, ..., N;\ t = 1, ..., T, \tag{36}$$

where α_i is an intercept, $\beta_{\ell i}$ is the loading of asset i corresponding to factor $f_{\ell t}$, m is the number of factors, and u_{it} is the idiosyncratic error. The parameter κ controls for the relative dominance of the variance of the idiosyncratic terms over the pervasive effects.

7.1 Testing a linear Capital Asset Pricing Model

In our first application we set $\kappa = 1$ and determine the parameters and rhs variables in (36) following PY. These aim at approximating the conditions prevailing in the $S\&P$ 500 data set over the period of September 1989 to September 2011. We refer to Section 5 (p. 19-22) of Pesaran and Yamagata (2012) for details regarding the generation of y_{it}. When testing for market efficiency in essence the hypothesis tested is that all intercept terms α_i, $i = 1, ..., N$ are equal to zero or $H_0 : \boldsymbol{\alpha} = \mathbf{0}$, where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$. A number of tests have been developed in the literature which predominantly focus on the case where the number of assets is either limited compared to the time dimension ($N < T$) or if $N > T$ then these assets are collected in a group of porfolios to handle the issue of dimensionality - see Gibbons, Ross and Shanken (1989), Beaulieu, Dufour and Khalaf (2007), Gungor and Luger (2009, 2011), among others. Pesaran and Yamagata (2012) provide a comprehensive review of such methods and their limitations. In turn, they propose a test statistic using (36), which under normality of $\mathbf{u}_t = (u_{1t}, ..., u_{Nt})'$ can be written as

$$J(\boldsymbol{\Sigma}_u) = \frac{\left(\boldsymbol{\tau}_T' \mathbf{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \boldsymbol{\Sigma}_u^{-1} \hat{\boldsymbol{\alpha}} - N}{\sqrt{2N}} \to_d N(0,1) \text{ as } N \to \infty \text{ for any fixed } T \geq m+1, \tag{37}$$

where $\hat{\boldsymbol{\alpha}}$ are estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of ones and $\mathbf{M}_F = \mathbf{I}_T - \mathbf{F}\left(\mathbf{F}'\mathbf{F}\right)^{-1}\mathbf{F}'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{f}_t = (f_{1t}, ..., f_{mt})'$. $\boldsymbol{\Sigma}_u$ is the variance-covariance matrix of the error terms $\mathbf{u}_t$. When $\boldsymbol{\Sigma}_u$ is known this test is valid for any $T > m+1$, but if an estimator of $\boldsymbol{\Sigma}_u^{-1}$ is inserted in (37), then PY show that $J(\hat{\boldsymbol{\Sigma}}_u) \to_d N(0,1)$ only if $\frac{N \log(N)}{T} \to 0$, which requires $N < T$. To illustrate this point we repeat part of the Monte Carlo simulation in PY for $N = \{50, 100, 500\}$ and $T = \{60, 100\}$, where we plug-in $\tilde{\boldsymbol{\Sigma}}^{-1}_{\hat{u},S\text{-}MT_R}$, $\tilde{\boldsymbol{\Sigma}}^{-1}_{\hat{u},S\text{-}MT_F}$, $\tilde{\boldsymbol{\Sigma}}^{-1}_{\hat{u},FLM_{CV}}$ and $\hat{\boldsymbol{\Sigma}}^{-1}_{\hat{u},LW_{\hat{\Sigma}}}$ as estimates of $\boldsymbol{\Sigma}_u^{-1}$ in (37) and report the size and power of the test.[22] We focus on cases (ii) and (iv) of PY where the errors are cross-sectionally weakly correlated and are assumed normal - case (ii), and non-normal - case (iv).

[22] First, y_{it}, $i = 1, ..., N$, $t = 1, ..., T$ are defactored following Fan, Liao and Mincheva (2011). Then, $\hat{\boldsymbol{\Sigma}}_{\hat{u}}$ is estimated from the resulting residuals, $\hat{u}_{it}$, $i = 1, ..., N$, $t = 1, ..., T$.

As shown in Table 7, considerable size distortions are visible when either of the estimators is used for $N > T$. Size only improves as T increases. Looking at the relative performance of the four estimators there is little difference between the four approaches when $T = 60$. As T increases to 100 again results are very similar for all methods with FLM_{CV} slightly outperforming the rest of the estimators for small N, but $S\text{-}MT$ prevailing for larger N.

To overcome this problem PY propose the following simple test statistic that ignores the off-diagonal elements of $\mathbf{\Sigma}_u$,

$$J_{PY} = \frac{N^{-1/2}\sum_{i=1}^{N}\left(t_i^2 - \frac{\upsilon}{\upsilon-2}\right)}{\left(\frac{\upsilon}{\upsilon-2}\right)\sqrt{\frac{2(\upsilon-1)}{(\upsilon-4)}\left[1+(N-1)\widehat{\bar{\rho}^2}\right]}},$$

where $\upsilon = T - m - 1$, and t_i denotes the standard t-ratio of α_i in the OLS regression of individual asset returns, and

$$\widehat{\bar{\rho}^2} = \frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\hat{\rho}_{ij}^2 I\left(\upsilon\hat{\rho}_{ij}^2 \geq b_N\right), \tag{38}$$

$\hat{\rho}_{ij} = \hat{\mathbf{u}}_{i.}'\hat{\mathbf{u}}_{j.}/\sqrt{(\hat{\mathbf{u}}_{i.}'\hat{\mathbf{u}}_{i.})\left(\hat{\mathbf{u}}_{j.}'\hat{\mathbf{u}}_{j.}\right)}$, $I(.)$ is an indicator function and b_N is defined in (8).[23] Size and power for this test are summarised in Table 7. The results show that the J_{PY} controls well for size and displays high power even when $N \gg T$. For a detailed analysis of this test statistic see Pesaran and Yamagata (2012). This exercise is based on 2000 replications.

7.2 Large portfolio optimisation

Our second application focuses on the subject of optimal risk-return tradeoff in portfolio investment, analysed in the seminal work of Markowitz (1952) and further developed by Sharpe (1964), Lintner (1965) and Ross (1976) with the introduction of the capital asset pricing model and arbitrage pricing theory. Since then, Chamberlain (1983), Chamberlain and Rothschild (1983), Green and Hollifield (1992), Sentana (2004) and Pesaran and Zaffaroni (2009) among others, have considered the implications of using the factor model in finding the tangency portfolio when the number of asset returns becomes very large ($N \to \infty$). Here, we use the factor model specification given by (36) where the intercepts are generated as $\alpha_i \sim IIDN(1, 0.5^2)$, the factors as

$$f_{\ell t} \sim IIDN(0,1),\ \ell = 1,2,3;\ t = 1,...,T,$$

and the corresponding factor loadings as

$$\beta_{\ell i} \sim IIDU(\mu_{v_\ell} - 0.5, \mu_{v_\ell} + 0.5),\ \ell = 1,2,3;\ i = 1,...,N,$$

with $\mu_{v_\ell} = \sqrt{1/3}$ for all ℓ, so that $\sum_{\ell=1}^{m}\mu_{v_\ell}^2\sigma_{\ell f}^2 = 1$, where $\sigma_{\ell f}^2 = 1$ by construction.

Finally, the idiosyncratic errors u_{it} are generated to be heteroskedastic and weakly cross-sectionally dependent. Specifically, we adopt the same spatial autoregressive model (SAR) of experiment B in Section 6 to generate $\boldsymbol{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$.

Stacking over units i in (36) we have

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{B}'\mathbf{f}_t + \kappa\boldsymbol{u}_t,\ t = 1,...,T,$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$, $\boldsymbol{B} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m)'$, $\boldsymbol{\beta}_i = (\beta_{i1}, ..., \beta_{iN})'$, and $\mathbf{f}_t = (f_{1t}, ..., f_{mt})'$.[24] The population value of $\mathbf{\Sigma}_y$ is computed using

$$\mathbf{\Sigma}_y = \boldsymbol{B}'\mathbf{\Sigma}_f\,\boldsymbol{B} + \kappa^2\mathbf{\Sigma}_u = \boldsymbol{B}'\mathbf{\Sigma}_f\,\boldsymbol{B} + \kappa^2(\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{D}(\boldsymbol{I}_N - \rho\boldsymbol{W}')^{-1}$$

[23] Pesaran and Yamagata (2012) use Bonferroni by row when computing (38).

[24] The population values of σ_i^2, $\beta_{\ell i}$ for $i = 1,2,...,N$ and $\ell = 1,2,3$ are generated once and fixed throughout the replications.

where $\mathbf{D}=diag(\sigma_1^2,\sigma_2^2,...,\sigma_N^2)$ and $\boldsymbol{\Sigma}_f$ is set to its population value of $\boldsymbol{\Sigma}_f = E\left(\mathbf{f}_t\mathbf{f}_t'\right) = \boldsymbol{I}_3$. For the computation of the inverse it is convenient to use

$$\boldsymbol{\Sigma}_y^{-1} = \kappa^{-2}\boldsymbol{\Sigma}_u^{-1} - \kappa^{-4}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}'\left(\boldsymbol{I}_3 + \kappa^{-2}\boldsymbol{B}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}'\right)^{-1}\boldsymbol{B}\boldsymbol{\Sigma}_u^{-1},$$

where $\boldsymbol{\Sigma}_u^{-1} = (\boldsymbol{I}_N - \rho\boldsymbol{W}')\boldsymbol{D}^{-1}(\boldsymbol{I}_N - \rho\boldsymbol{W})$.

We consider the following combinations of sample sizes $N = \{50, 100, 200, 400\}$, $T = \{60, 100\}$, for $\kappa = \{1, 2, 3, 4\}$ and spatial autoregressive parameter $\vartheta = 0.4$. The number of replications is set to $R = 500$.

As analysed in Markowitz (1952) the global optimal mean-variance portfolio is the stock portfolio that has the lowest risk and highest expected return payoff. The risk attributes are summarised in the covariance matrix $\boldsymbol{\Sigma}_y$. Simplifying the problem by assuming common mean returns equating to unity, the solution to this porfolio optimisation problem amounts to minimizing the following:

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\Sigma}_y\boldsymbol{w}, \quad \text{s.t.} \quad \boldsymbol{w}'\,\boldsymbol{e} = 1,$$

where $\boldsymbol{e}$ is an $N \times 1$ vector of ones and $\boldsymbol{w} = (w_1, ..., w_N)'$ is a vector of portfolio weights. The estimated weights of the global optimal portfolio are given by

$$\hat{\boldsymbol{w}} = \frac{\hat{\boldsymbol{\Sigma}}_y^{-1}\boldsymbol{e}}{\boldsymbol{e}'\hat{\boldsymbol{\Sigma}}_y^{-1}\boldsymbol{e}},$$

where $\hat{\boldsymbol{\Sigma}}_y = T^{-1}\sum_{t=1}^T \boldsymbol{y}_t\boldsymbol{y}_t'$. The estimated return variance of the portfolio, $\hat{\sigma}_{GOP}^2$, is then

$$\hat{\sigma}_{GOP}^2 = \frac{\boldsymbol{e}'\hat{\boldsymbol{\Sigma}}_y^{-1}\hat{\boldsymbol{\Sigma}}_y\hat{\boldsymbol{\Sigma}}_y^{-1}\boldsymbol{e}}{\left(\boldsymbol{e}'\hat{\boldsymbol{\Sigma}}_y^{-1}\boldsymbol{e}\right)^2} = \left(\boldsymbol{e}'\hat{\boldsymbol{\Sigma}}_y^{-1}\boldsymbol{e}\right)^{-1}. \tag{39}$$

Once again, using the inverse of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_y$ in (39) is problematic when $N \gg T$. We proceed to regularise $\hat{\boldsymbol{\Sigma}}_y$ using the same estimators as in the CAPM testing exercise. Given the existence of factors in (36) first we extract them via OLS following Fan, Liao and Mincheva (2011) and compute the de-factored components, $\hat{\mathbf{v}}_t = \kappa\hat{\boldsymbol{u}}_t$. Then, we regularise the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{v}} = T^{-1}\sum_{t=1}^T \hat{\mathbf{v}}_t\hat{\mathbf{v}}_t'$, and compute,

$$\mathring{\boldsymbol{\Sigma}}_y = \hat{\boldsymbol{B}}'\hat{\boldsymbol{\Sigma}}_f\hat{\boldsymbol{B}} + \mathring{\boldsymbol{\Sigma}}_{\hat{v}},$$

where $\mathring{\boldsymbol{\Sigma}}_y$ and $\mathring{\boldsymbol{\Sigma}}_{\hat{v}}$ are the regularised versions of $\hat{\boldsymbol{\Sigma}}_y$ and $\hat{\boldsymbol{\Sigma}}_{\hat{v}}$. Here $\mathring{\boldsymbol{\Sigma}}_y = \left\{\tilde{\boldsymbol{\Sigma}}_{y,S\text{-}MT_R}, \tilde{\boldsymbol{\Sigma}}_{y,S\text{-}MT_F}, \tilde{\boldsymbol{\Sigma}}_{y,FLM_{CV}}, \hat{\boldsymbol{\Sigma}}_{y,LW_{\hat{\Sigma}}}\right\}$.[25] The inverse of $\mathring{\boldsymbol{\Sigma}}_y$ is computed as

$$\mathring{\boldsymbol{\Sigma}}_y^{-1} = \left(\hat{\boldsymbol{B}}'\hat{\boldsymbol{\Sigma}}_f\hat{\boldsymbol{B}} + \mathring{\boldsymbol{\Sigma}}_{\hat{v}}\right)^{-1} = \mathring{\boldsymbol{\Sigma}}_{\hat{v}}^{-1} - \mathring{\boldsymbol{\Sigma}}_{\hat{v}}^{-1}\hat{\boldsymbol{B}}'\left(\hat{\boldsymbol{\Sigma}}_f^{-1} + \hat{\boldsymbol{B}}\mathring{\boldsymbol{\Sigma}}_{\hat{v}}^{-1}\hat{\boldsymbol{B}}'\right)^{-1}\hat{\boldsymbol{B}}\mathring{\boldsymbol{\Sigma}}_{\hat{v}}^{-1}.$$

We evaluate the following relationships across the different versions of $\mathring{\boldsymbol{\Sigma}}_y$:

1. The bias term of the estimated return variance $\hat{\sigma}_{GOP}^2$ of the portfolio

$$\frac{1}{R}\sum_{r=1}^R\left(\sigma_{GOP}^2 - \hat{\sigma}_{GOP}^{2\,(r)}\right) = \frac{1}{R}\sum_{r=1}^R\left(\frac{1}{\boldsymbol{e}'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{e}} - \frac{1}{\boldsymbol{e}'\mathring{\boldsymbol{\Sigma}}_y^{-1\,(r)}\boldsymbol{e}}\right).$$

[25] The equivalent regularisation estimators of $\hat{\boldsymbol{\Sigma}}_{\hat{v}}$ are $\mathring{\boldsymbol{\Sigma}}_{\hat{v}} = \left\{\tilde{\boldsymbol{\Sigma}}_{\hat{v},S\text{-}MT_R}, \tilde{\boldsymbol{\Sigma}}_{\hat{v},S\text{-}MT_F}, \tilde{\boldsymbol{\Sigma}}_{\hat{v},FLM_{CV}}, \hat{\boldsymbol{\Sigma}}_{\hat{v},LW_{\hat{\Sigma}}}\right\}$.

2. The corresponding root-mean-square error (RMSE)

$$\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\sigma^2_{GOP}-\hat{\sigma}^{2\,(r)}_{GOP}\right)^2}=\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{\boldsymbol{e}'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{e}}-\frac{1}{\boldsymbol{e}'\mathring{\hat{\boldsymbol{\Sigma}}}_y^{-1\,(r)}\boldsymbol{e}}\right)^2}.$$

3. The RMSE of the Euclidean norm of the optimal portfolio weights

$$\frac{1}{R}\sum_{r=1}^{R}\left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(w_i-\mathring{\hat{w}}_i^{(r)}\right)^2}\right],$$

where

$$\boldsymbol{w}=\frac{\boldsymbol{\Sigma}_y^{-1}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{e}}\text{ and }\mathring{\hat{\boldsymbol{w}}}=\frac{\mathring{\hat{\boldsymbol{\Sigma}}}_y^{-1}\boldsymbol{e}}{\boldsymbol{e}'\mathring{\hat{\boldsymbol{\Sigma}}}_y^{-1}\boldsymbol{e}}.$$

4. The average of norms - both spectral and Frobenius - for all $\mathring{\hat{\boldsymbol{\Sigma}}}_y^{-1}$, given by

$$\frac{1}{R}\sum_{r=1}^{R}\left\|\boldsymbol{\Sigma}_y^{-1}-\mathring{\hat{\boldsymbol{\Sigma}}}_y^{-1\,(r)}\right\|.$$

We also report the average shrinkage parameter estimates corresponding to each $\mathring{\hat{\boldsymbol{\Sigma}}}_{\hat{v}}^{-1}$ estimator.

In Table 8 we only report results for $\kappa=1$. Those for $\kappa=2,3,4$, where increased dominance of the error term is considered, can be found in the accompanying supplement. As noted in Pesaran and Zaffaroni (2009) the mean-variance efficient portfolio (MV) is a function of the inverse of the variance matrix of the asset returns. However, the workings of this relationship are considerably complex and assessment of the performance of the estimated MV portfolio is not always clear cut. For this reason we consider the use of more than one statistical measures.

We first look at the bias and RMSE of the estimated return variance $\hat{\sigma}^2_{GOP}$ of the global optimal portfolio when using the different regularisation techniques. Using $S\text{-}MT$ at the 5% or 10% significance level appears to produce more accurate $\hat{\sigma}^2_{GOP}$ estimates compared with FLM_{CV} for small N, though FLM_{CV} does better as N increases for both $T=\{60,100\}$ specifications. In this case, it is $S\text{-}MT_F$ that delivers lower bias compared with $S\text{-}MT_R$. Further, $LW_{\hat{\Sigma}}$ persistently generates the most accurate $\hat{\sigma}^2_{GOP}$ estimates out of all other estimators, however at the same time it shrinks the sample covariance matrix considerably more than either of the four versions of $S\text{-}MT$ for all N and T, as shown in the last column of Table 8. This is in line with the findings of Section 6. This extended shrinkage, where in fact the variance components of the estimated covariance matrix become a fraction of their original values, is precisely what causes the $LW_{\hat{\Sigma}}$ method to work so well. In the context of portfolio optimisation this implies that the reduced overall risk estimated is not due to selecting the optimal cross-return fluctuations while disregarding less important asset return co-movements, but rather a result of artificially dampening the volatilities of individual asset returns. Though desirable, results using this metric should be treated with caution. Turning to the RMSE of the portfolio weights these appear similar for all estimators. From the fifth column of Table 8 it is evident that $S\text{-}MT$ produces marginally more accurate estimates than the other methods across all (N,T) combinations. Finally, the spectral and Frobenius norm results for the inverse variance matrix of the optimal portfolio estimates follow the ranking shown in Section 6.

8 Concluding Remarks

This paper considers the issue of regularising large covariance matrices particularly in the case where the cross-sectional dimension N of the data under consideration exceeds the time dimension, T, and the sample variance-covariance matrix, $\hat{\mathbf{\Sigma}}$, becomes non-invertible. 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 convergence rate of $\sqrt{\frac{m_N N}{T}}$ under the Frobenius norm, where m_N is bounded in N. 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 $\mathbf{\Sigma}$ over time. If regularisation of both $\hat{\mathbf{\Sigma}}$ and its inverse is of interest then we recommend shrinkage applied to the sample correlation matrix. This method can also be used for supplementary regularisation of our multiple testing estimator and ensures its invertibility.

Monte Carlo simulation findings are supportive of the theoretical properties of MT. They show favourable performance of both versions of the MT estimator compared with a number of key regularisation techniques in the literature as well as their robustness to different covariance matrix settings and deviations from the main assumptions of the underlying theory. The challenges of testing a capital asset pricing model and estimating a global optimal portfolio when the number of assets is large are explored in two empirical applications of the MT method among other regularisation approaches.

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

Table 1: Multipe testing (MT) estimator

Normally distributed errors

5% and 10% significance level - averages over 500 replications

	$N=30$		$N=100$		$N=200$		$N=400$	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Experiment A							
	$T=60$							
$MT_{R_{5\%}}$	4.426	7.924	5.691	16.254	6.112	23.981	6.481	35.128
$MT_{F_{5\%}}$	5.187	9.117	6.614	19.059	7.163	28.482	7.735	42.602
$MT_{R_{10\%}}$	4.241	7.646	5.495	15.733	5.937	23.287	6.328	34.181
$MT_{F_{10\%}}$	5.007	8.825	6.467	18.612	6.994	27.823	7.538	41.455
	$T=100$							
$MT_{R_{5\%}}$	3.492	6.249	4.540	12.863	4.941	18.999	5.341	27.872
$MT_{F_{5\%}}$	4.025	7.138	5.384	15.258	5.891	23.010	6.296	34.339
$MT_{R_{10\%}}$	3.373	6.044	4.395	12.487	4.791	18.486	5.192	27.167
$MT_{F_{10\%}}$	3.887	6.911	5.237	14.816	5.757	22.382	6.193	33.529
	Experiment B							
	$T=60$							
$MT_{R_{5\%}}$	1.419	3.334	1.634	6.477	2.012	10.093	2.170	14.753
$MT_{F_{5\%}}$	1.557	3.926	1.755	7.407	2.098	11.187	2.243	15.997
$MT_{R_{10\%}}$	1.378	3.180	1.613	6.241	1.988	9.795	2.153	14.400
$MT_{F_{10\%}}$	1.525	3.790	1.731	7.313	2.095	11.131	2.242	15.964
	$T=100$							
$MT_{R_{5\%}}$	1.029	2.303	1.284	4.559	1.618	7.256	1.749	10.835
$MT_{F_{5\%}}$	1.220	2.858	1.522	6.160	1.938	9.894	2.046	14.659
$MT_{R_{10\%}}$	1.001	2.238	1.251	4.365	1.565	6.916	1.712	10.332
$MT_{F_{10\%}}$	1.169	2.690	1.494	5.901	1.906	9.588	2.022	14.335
	Experiment C							
	$T=60$							
$MT_{R_{5\%}}$	2.194	4.061	3.299	8.479	3.866	12.568	4.299	18.400
$MT_{F_{5\%}}$	2.394	4.336	3.949	9.650	4.725	14.710	5.333	22.164
$MT_{R_{10\%}}$	2.214	4.131	3.257	8.531	3.794	12.604	4.215	18.417
$MT_{F_{10\%}}$	2.325	4.236	3.811	9.394	4.577	14.320	5.182	21.594
	$T=100$							
$MT_{R_{5\%}}$	1.686	3.110	2.518	6.435	2.876	9.483	3.221	13.884
$MT_{F_{5\%}}$	1.730	3.199	2.836	7.082	3.307	10.716	3.774	16.110
$MT_{R_{10\%}}$	1.729	3.189	2.513	6.549	2.857	9.605	3.193	14.023
$MT_{F_{10\%}}$	1.704	3.152	2.762	6.920	3.227	10.465	3.688	15.745
	Experiment D							
	$T=60$							
$MT_{R_{5\%}}$	0.656	1.197	1.061	2.185	0.998	2.980	1.400	4.343
$MT_{F_{5\%}}$	0.728	1.245	1.513	2.512	1.487	3.195	2.416	4.818
$MT_{R_{10\%}}$	0.678	1.259	1.054	2.293	1.014	3.163	1.357	4.568
$MT_{F_{10\%}}$	0.687	1.200	1.394	2.397	1.366	3.106	2.248	4.705
	$T=100$							
$MT_{R_{5\%}}$	0.488	0.902	0.763	1.653	0.730	2.310	0.920	3.331
$MT_{F_{5\%}}$	0.468	0.852	0.798	1.589	0.743	2.154	1.120	3.224
$MT_{R_{10\%}}$	0.512	0.962	0.787	1.775	0.766	2.493	0.951	3.608
$MT_{F_{10\%}}$	0.467	0.853	0.780	1.574	0.723	2.141	1.054	3.178

MT_R=Multiple testing by row, MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both estimators use Holm method at 5% and 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

	$N=30$		$N=100$		$N=200$		$N=400$	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Sparse covariance matrix $\mathbf{\Sigma}$							
	$T=60$							
MT_R	4.426	7.924	5.691	16.254	6.112	23.981	6.481	35.128
MT_F	5.187	9.117	6.614	19.059	7.163	28.482	7.735	42.602
$S\text{-}MT_R$	5.784	8.736	7.170	18.475	7.572	27.568	7.882	40.656
$S\text{-}MT_F$	6.450	9.898	7.774	20.867	8.172	31.149	8.506	46.079
BL^*_{CV}	4.284	7.497	5.648	16.028	6.384	24.347	6.963	36.414
BL^*_{FLM}	8.543	14.503	9.142	27.137	9.223	38.570	9.267	54.679
CL_T	5.566	9.705	7.537	21.611	8.263	33.149	8.729	49.729
CL^*_{CV}	4.088	7.339	5.228	15.610	5.785	23.612	6.274	35.382
CL^*_{FLM}	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	3.492	6.249	4.540	12.863	4.941	18.999	5.341	27.872
MT_F	4.025	7.138	5.384	15.258	5.891	23.010	6.296	34.339
$S\text{-}MT_R$	4.763	7.048	6.197	15.374	6.675	23.267	7.066	34.772
$S\text{-}MT_F$	5.460	8.102	6.884	17.646	7.367	26.768	7.736	40.089
BL^*_{CV}	3.336	5.829	4.383	12.439	4.893	18.775	5.496	28.182
BL^*_{FLM}	8.527	14.450	9.114	27.043	9.187	38.438	9.228	54.503
CL_T	4.140	7.336	5.695	16.169	6.323	24.760	6.931	37.571
CL^*_{CV}	3.247	5.757	4.144	12.227	4.585	18.407	5.000	27.459
CL^*_{FLM}	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
	Inverse of sparse covariance matrix $\mathbf{\Sigma^{-1}}$							
	$T=60$							
$S\text{-}MT_R$	4.065	5.261	4.747	10.269	4.994	15.033	5.174	21.862
$S\text{-}MT_F$	4.101	5.023	4.457	9.992	4.559	15.130	4.719	22.754
BL^*_{FLM}	5.683	7.348	5.868	13.663	5.941	19.403	6.002	27.487
CL^*_{FLM}	2.5E+02	8.723	1.2E+02	14.302	6.298	19.404	7.520	27.514
$LW_{\hat{R}}$	2.216	3.920	3.421	9.028	3.818	13.865	3.995	20.560
$LW_{\hat{\Sigma}}$	2.523	4.187	4.038	10.674	4.666	16.953	5.074	25.610
	$T=100$							
$S\text{-}MT_R$	3.505	5.053	4.302	10.068	4.612	14.825	4.862	21.651
$S\text{-}MT_F$	4.053	5.190	4.731	9.987	4.969	14.637	5.134	21.425
BL^*_{FLM}	29.820	7.590	5.822	13.731	5.879	19.496	5.925	27.623
CL^*_{FLM}	7.1E+03	13.561	6.9E+03	18.230	32.454	19.744	4.1E+02	29.356
$LW_{\hat{R}}$	1.712	3.368	3.042	8.254	3.601	13.124	3.896	19.965
$LW_{\hat{\Sigma}}$	1.927	3.480	3.511	9.463	4.285	15.764	4.846	24.669
	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 $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T=60$							
$S\text{-}MT_R$	0.389	0.611	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
$LW_{\hat{R}}$	0.157	0.843	0.306	0.694	0.377	0.623	0.425	0.575
$LW_{\hat{\Sigma}}$	0.443	0.770	0.898	0.534	1.202	0.377	1.458	0.244
	$T=100$							
$S\text{-}MT_R$	0.309	0.691	0.400	0.600	0.445	0.555	0.483	0.517
$S\text{-}MT_F$	0.352	0.648	0.435	0.565	0.480	0.520	0.522	0.478
$LW_{\hat{R}}$	0.109	0.891	0.248	0.752	0.331	0.669	0.396	0.604
$LW_{\hat{\Sigma}}$	0.298	0.846	0.678	0.650	0.988	0.491	1.296	0.333

* For $N=400$ and $T=60,100$ replications are set to 100. For all other N,T combinations replications are set to 500.

MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use Holm method at 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.

CV uses cross-validation parameter; FLM uses Fan, Liao and Michela grid adjustment; T uses theoretical parameter.

LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on sample covariance matrix; $\hat{R}$ on 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

	$N=30$		$N=100$		$N=200$		$N=400$	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Sparse covariance matrix $\mathbf{\Sigma}$							
	$T=60$							
MT_R	1.419	3.334	1.634	6.477	2.012	10.093	2.170	14.753
MT_F	1.557	3.926	1.755	7.407	2.098	11.187	2.243	15.997
$S\text{-}MT_R$	1.435	3.323	1.657	6.388	2.003	9.859	2.144	14.428
$S\text{-}MT_F$	1.559	3.876	1.759	7.360	2.093	11.136	2.238	15.964
BL^*_{CV}	1.615	3.941	1.983	7.625	2.106	11.250	2.277	16.048
BL^*_{FLM}	1.599	4.095	1.978	7.639	2.103	11.251	2.267	16.060
CL_T	1.571	3.974	1.894	7.505	2.093	11.182	2.242	15.986
CL^*_{CV}	1.436	3.361	1.900	7.214	2.089	11.129	2.239	15.929
CL^*_{FLM}	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	1.029	2.303	1.284	4.559	1.618	7.256	1.749	10.835
MT_F	1.220	2.858	1.522	6.160	1.938	9.894	2.046	14.659
$S\text{-}MT_R$	1.141	2.531	1.409	4.964	1.701	7.632	1.814	11.215
$S\text{-}MT_F$	1.290	2.960	1.575	6.197	1.952	9.812	2.053	14.530
BL^*_{CV}	1.214	2.705	1.574	5.843	1.911	9.915	2.145	15.584
BL^*_{FLM}	1.193	2.718	1.543	6.145	1.919	10.161	2.148	15.649
CL_T	1.249	2.961	1.553	6.401	1.970	10.214	2.086	15.020
CL^*_{CV}	1.034	2.334	1.295	4.587	1.628	7.423	1.860	11.911
CL^*_{FLM}	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
	Inverse of sparse covariance matrix $\mathbf{\Sigma^{-1}}$							
	$T=60$							
$S\text{-}MT_R$	1.963	3.373	2.652	6.891	3.259	10.148	3.691	14.938
$S\text{-}MT_F$	2.581	3.923	3.157	8.028	3.723	11.584	4.078	16.669
BL^*_{FLM}	1.4E+04	19.315	58.881	9.377	3.9E+03	15.321	14.009	17.017
CL^*_{FLM}	2.1E+04	33.982	2.4E+04	23.651	44.094	12.593	16.774	17.064
$LW_{\hat{R}}$	1.969	3.539	4.809	8.773	6.958	13.956	8.767	20.919
$LW_{\hat{\Sigma}}$	2.971	3.874	3.715	8.438	4.932	12.850	5.832	18.870
	$T=100$							
$S\text{-}MT_R$	1.294	2.647	1.889	5.436	2.435	8.034	2.854	11.933
$S\text{-}MT_F$	1.760	3.082	2.636	6.868	3.273	10.351	3.763	15.436
BL^*_{FLM}	5.0E+03	23.048	4.2E+03	24.145	2.7E+04	30.297	43.825	17.318
CL^*_{FLM}	3.0E+05	65.501	1.9E+05	1.0E+02	2.2E+07	3.6E+02	2.2E+03	31.662
$LW_{\hat{R}}$	1.333	2.982	2.805	7.349	4.381	12.101	5.719	18.967
$LW_{\hat{\Sigma}}$	2.338	3.374	3.406	7.993	4.735	12.515	5.744	18.663
	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
$LW_{\hat{R}}$	0.341	0.659	0.436	0.564	0.461	0.539	0.474	0.526
$LW_{\hat{\Sigma}}$	0.591	0.517	0.871	0.257	1.011	0.162	1.086	0.105
	$T=100$							
$S\text{-}MT_R$	0.362	0.638	0.418	0.582	0.416	0.584	0.408	0.592
$S\text{-}MT_F$	0.349	0.651	0.381	0.619	0.355	0.645	0.331	0.669
$LW_{\hat{R}}$	0.288	0.712	0.412	0.588	0.450	0.550	0.470	0.530
$LW_{\hat{\Sigma}}$	0.449	0.635	0.770	0.348	0.946	0.221	1.055	0.137

* For $N=400$ and $T=60,100$ replications are set to 100. For all other N,T combinations replications are set to 500.

MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use Holm method at 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.

CV uses cross-validation parameter; FLM uses Fan, Liao and Michela grid adjustment; T uses theoretical parameter.

LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on sample covariance matrix; $\hat{R}$ on sample correlation matrix.

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

Experiment C[1] - normally distributed errors

Averages over 500 replications

	$N=30$		$N=100$		$N=200$		$N=400$	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Sparse covariance matrix $\mathbf{\Sigma}$							
	$T=60$							
MT_R	2.194	4.061	3.299	8.479	3.866	12.568	4.299	18.400
MT_F	2.394	4.336	3.949	9.650	4.725	14.710	5.333	22.164
$S\text{-}MT_R$	3.377	4.837	5.995	11.621	6.638	17.474	7.043	25.812
$S\text{-}MT_F$	4.148	5.497	6.599	12.645	7.211	18.983	7.609	28.103
BL^*_{CV}	7.040	8.795	8.755	17.234	8.961	24.701	9.031	35.161
BL^*_{FLM}	7.091	8.804	8.755	17.233	8.961	24.701	9.031	35.172
CL_T	2.661	4.641	5.138	11.183	6.477	17.786	7.468	27.640
CL^*_{CV}	2.381	4.394	3.574	9.404	4.316	14.278	5.024	21.375
CL^*_{FLM}	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	3.110	2.518	6.435	2.876	9.483	3.221	13.884
MT_F	1.730	3.199	2.836	7.082	3.307	10.716	3.774	16.110
$S\text{-}MT_R$	2.431	3.610	5.079	9.597	5.734	14.752	6.221	22.258
$S\text{-}MT_F$	3.050	4.118	5.768	10.753	6.378	16.357	6.821	24.500
BL^*_{CV}	5.118	7.511	8.747	16.895	8.946	24.243	9.014	34.528
BL^*_{FLM}	7.082	8.609	8.747	16.898	8.946	24.241	9.014	34.534
CL_T	1.781	3.279	3.084	7.534	3.786	11.748	4.585	18.160
CL^*_{CV}	1.738	3.230	2.634	6.816	3.002	10.180	3.395	15.206
CL^*_{FLM}	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 $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
	$T=60$							
$S\text{-}MT_R$	0.381	0.619	0.562	0.438	0.603	0.397	0.633	0.367
$S\text{-}MT_F$	0.469	0.531	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
	$T=100$							
$S\text{-}MT_R$	0.263	0.737	0.481	0.519	0.532	0.468	0.572	0.428
$S\text{-}MT_F$	0.347	0.653	0.543	0.457	0.585	0.415	0.618	0.382
$LW_{\hat{\Sigma}}$	0.744	0.700	1.373	0.445	1.741	0.297	2.024	0.183

[1] Note that the population covariance matrix $\mathbf{\Sigma}$ does not have an inverse hence results relating to matrix inverses have no meaning.

* For $N=400$ and $T=60,100$ replications are set to 100. For all other N,T combinations replications are set to 500.

MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use Holm method at 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.

CV uses cross-validation parameter; FLM uses Fan, Liao and Michela grid adjustment; T uses theoretical parameter.

LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on sample covariance matrix.

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

Experiment D - normally distributed errors

Averages over 500 replications

	$N=30$		$N=100$		$N=200$		$N=400$	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
	Sparse covariance matrix $\mathbf{\Sigma}$							
	$T=60$							
MT_R	0.656	1.197	1.061	2.185	0.998	2.980	1.400	4.343
MT_F	0.728	1.245	1.513	2.512	1.487	3.195	2.416	4.818
$S\text{-}MT_R$	0.782	1.309	1.538	2.471	1.302	3.041	1.961	4.435
$S\text{-}MT_F$	0.889	1.391	2.026	2.877	1.946	3.399	3.050	5.041
BL^*_{CV}	1.436	1.931	2.635	3.512	2.735	3.985	3.722	5.566
BL^*_{FLM}	1.512	2.016	3.336	4.072	2.744	3.987	3.730	5.557
CL_T	0.847	1.389	2.055	3.054	1.976	3.550	3.088	5.218
CL^*_{CV}	0.925	1.478	1.854	2.939	2.328	3.761	3.362	5.372
CL^*_{FLM}	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
$S\text{-}MT_R$	0.647	1.056	1.415	2.094	1.083	2.422	1.364	3.408
$S\text{-}MT_F$	0.646	1.040	1.457	2.105	1.193	2.423	1.877	3.612
BL^*_{CV}	0.879	1.308	1.237	2.120	2.544	3.508	3.526	4.909
BL^*_{FLM}	1.133	1.573	3.328	3.915	2.727	3.617	3.696	4.989
CL_T	0.485	0.875	0.948	1.738	0.923	2.309	1.595	3.589
CL^*_{CV}	0.496	0.917	0.812	1.718	1.141	2.533	2.445	4.258
CL^*_{FLM}	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
	Inverse of sparse covariance matrix $\mathbf{\Sigma^{-1}}$							
	$T=60$							
$S\text{-}MT_R$	4.756	2.905	15.425	6.136	13.367	6.044	14.038	7.853
$S\text{-}MT_F$	5.282	3.031	17.857	6.501	16.941	6.540	18.114	8.513
BL^*_{FLM}	7.1E+02	7.034	46.674	8.388	26.348	7.707	24.963	9.503
CL^*_{FLM}	9.3E+04	21.119	29.780	8.096	34.349	7.834	45.816	9.851
$LW_{\hat{R}}$	5.187	4.452	15.736	12.584	15.080	19.470	18.160	30.113
$LW_{\hat{\Sigma}}$	12.420	4.558	31.907	8.771	31.988	9.478	31.854	12.568
	$T=100$							
$S\text{-}MT_R$	4.532	2.684	15.398	5.866	12.793	5.364	11.034	6.434
$S\text{-}MT_F$	4.526	2.665	15.673	5.882	13.853	5.444	14.398	6.900
BL^*_{FLM}	1.7E+04	19.022	2.7E+02	8.880	48.354	7.690	26.695	8.897
CL^*_{FLM}	4.5E+02	6.177	8.1E+02	9.214	1.9E+02	8.419	40.085	9.033
$LW_{\hat{R}}$	4.850	3.720	16.168	10.239	14.347	16.032	13.104	26.403
$LW_{\hat{\Sigma}}$	10.861	4.240	30.981	8.611	31.783	9.400	31.841	12.526
	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	0.405	0.595	0.399	0.601	0.454	0.546
$S\text{-}MT_F$	0.424	0.576	0.496	0.504	0.540	0.460	0.614	0.386
$LW_{\hat{R}}$	0.423	0.577	0.467	0.533	0.483	0.517	0.485	0.515
$LW_{\hat{\Sigma}}$	0.579	0.375	0.735	0.180	0.842	0.091	0.871	0.067
	$T=100$							
$S\text{-}MT_R$	0.352	0.648	0.394	0.606	0.364	0.636	0.334	0.666
$S\text{-}MT_F$	0.353	0.647	0.402	0.598	0.401	0.599	0.459	0.541
$LW_{\hat{R}}$	0.392	0.608	0.460	0.540	0.485	0.515	0.489	0.511
$LW_{\hat{\Sigma}}$	0.473	0.492	0.682	0.244	0.826	0.115	0.868	0.077

* For $N=400$ and $T=60,100$ replications are set to 100. For all other N,T combinations replications are set to 500.

MT_R=Multiple testing by row; MT_F=Multiple testing on full $\hat{\mathbf{R}}$ matrix. Both use Holm method at 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.

CV uses cross-validation parameter; FLM uses Fan, Liao and Michela grid adjustment; T uses theoretical parameter.

LW=Ledoit and Wolf shrinkage: $\hat{\Sigma}$ on sample covariance matrix; $\hat{R}$ on sample correlation matrix.

Table 6: Comparison of $\boldsymbol{\Sigma}$ support recovery between MT and thresholding estimators

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	0.729	0.002	0.591	0.000	0.553	0.000	0.522	0.000
MT_F	0.623	0.000	0.456	0.000	0.402	0.000	0.357	0.000
BL_{CV}^*	0.013	0.002	0.000	0.000	0.000	0.000	0.000	0.000
CL_T	0.584	0.000	0.370	0.000	0.286	0.000	0.215	0.000
CL_{CV}^*	0.710	0.005	0.576	0.002	0.528	0.001	0.478	0.000
	$T=100$							
MT_R	0.813	0.002	0.700	0.000	0.669	0.000	0.641	0.000
MT_F	0.739	0.000	0.597	0.000	0.553	0.000	0.515	0.000
BL_{CV}^*	0.324	0.048	0.000	0.000	0.000	0.000	0.000	0.000
CL_T	0.729	0.000	0.566	0.000	0.506	0.000	0.453	0.000
CL_{CV}^*	0.781	0.002	0.686	0.001	0.655	0.001	0.623	0.000
	Experiment D							
	$T=60$							
MT_R	0.975	0.001	0.972	0.000	0.941	0.000	0.895	0.000
MT_F	0.869	0.000	0.833	0.000	0.649	0.000	0.469	0.000
BL_{CV}^*	0.187	0.001	0.325	0.000	0.009	0.000	0.006	0.000
CL_T	0.753	0.000	0.607	0.000	0.375	0.000	0.214	0.000
CL_{CV}^*	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_{CV}^*	0.686	0.002	0.852	0.001	0.101	0.000	0.051	0.000
CL_T	0.981	0.000	0.950	0.000	0.886	0.000	0.749	0.000
CL_{CV}^*	0.994	0.002	0.986	0.001	0.790	0.000	0.469	0.000

* For $N=400$ and $T=60,100$ replications are set to 100. For all other N,T combinations replications are set to 500.

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

Both MT estimators use Holm method at 5% significance level.

BL_{CV}=Bickel and Levina universal thresholding using cross-validation parameter.

CL_T= Cai and Liu adaptive thresholding using theoretical parameter.

CL_{CV}= Cai and Liu adaptive thresholding using cross-validation parameter.

Table 7: Size and power of $J(\hat{\Sigma}_{\hat{u}})$ test in the case of models with three factors

		normal errors						non-normal errors					
		Size	Power	Size	Power	Size	Power	Size	Power	Size	Power	Size	Power
	(T,N)	50		100		500		50		100		500	
PY	60	0.054	0.660	0.063	0.786	0.058	0.984	0.056	0.669	0.063	0.811	0.064	0.991
	100	0.065	0.894	0.056	0.970	0.054	1.000	0.066	0.881	0.068	0.974	0.056	1.000
$S\text{-}MT_R$	60	0.156	0.816	0.203	0.941	0.536	1.000	0.155	0.828	0.217	0.945	0.559	1.000
	100	0.107	0.940	0.128	0.991	0.255	1.000	0.112	0.928	0.141	0.991	0.299	1.000
$S\text{-}MT_F$	60	0.154	0.813	0.200	0.939	0.526	1.000	0.156	0.824	0.218	0.945	0.544	1.000
	100	0.106	0.941	0.127	0.992	0.235	1.000	0.113	0.927	0.140	0.989	0.285	1.000
FLM_{CV}	60	0.192	0.834	0.252	0.954	0.551	1.000	0.184	0.813	0.226	0.951	0.563	1.000
	100	0.116	0.951	0.149	0.993	0.287	1.000	0.135	0.945	0.153	0.993	0.307	1.000
$LW_{\hat{\Sigma}}$	60	0.166	0.695	0.223	0.827	0.577	0.994	0.158	0.683	0.241	0.821	0.566	0.997
	100	0.119	0.865	0.168	0.956	0.297	1.000	0.121	0.864	0.161	0.952	0.321	0.999

PY= Pesaran and Yamagata, FLM=Fan, Liao and Mincheva, LW=Ledoit and Wolf.

$S\text{-}MT_R$ and $S\text{-}MT_F$ stand for Shrinkage on Multiple Testing by row and on the full sample correlation matrix.

$S\text{-}MT_R$ and $S\text{-}MT_F$ are evaluated at 5% significance level.

FLM approach uses cross validation to evaluate the thresholding parameter. LW method is applied to the sample covariance matrix.

Errors are weakly cross-sectionally dependent. Sparseness of Σ_u is defined as in Table 3 of PY(2012) with $\delta_b = 1/4$.

Size: $\alpha_i = 0$ for all $i = 1, ..., p$. Power: $\alpha_i \sim$IIDN(0,1) for $i = 1, 2, ..., p_\alpha$, $p_\alpha = [p^{0.8}]$, otherwise $\alpha_i = 0$. Replications are set to 2000.

Table 8: Deviation of estimated Global Optimal Portfolio from true portfolio

Averages over 500 replications

		$(\sigma^2_{GOP} - \hat{\sigma}^2_{GVO})$ Bias (x100)	RMSE (x100)	Weights RMSE (x100)	Norms Spectral	Frobenius	Shrinkage Parameters on $\mathbf{I}$	on $\hat{\mathbf{R}}/\hat{\mathbf{\Sigma}}$
$T=60$, $k=1$								
$N=50$	$S\text{-}MT_{R_{5\%}}$	3.307	4.369	3.988	1.442	4.072	0.384	0.616
	$S\text{-}MT_{F_{5\%}}$	3.027	4.215	4.306	1.637	4.767	0.292	0.708
	$S\text{-}MT_{R_{10\%}}$	3.327	4.375	3.903	1.377	3.895	0.401	0.599
	$S\text{-}MT_{F_{10\%}}$	3.123	4.256	4.253	1.610	4.653	0.308	0.692
	FLM_{CV}	3.966	5.736	5.308	5.242	8.544	-	-
	$LW_{\hat{\Sigma}}$	1.811	3.548	4.364	1.774	4.851	0.720	0.422
$N=100$	$S\text{-}MT_{R_{5\%}}$	1.737	1.922	2.002	1.652	6.649	0.379	0.621
	$S\text{-}MT_{F_{5\%}}$	1.627	1.833	2.156	1.803	7.642	0.293	0.707
	$S\text{-}MT_{R_{10\%}}$	1.737	1.923	1.964	1.607	6.395	0.400	0.600
	$S\text{-}MT_{F_{10\%}}$	1.663	1.864	2.139	1.789	7.535	0.298	0.702
	FLM_{CV}	1.725	2.070	2.425	4.764	10.799	-	-
	$LW_{\hat{\Sigma}}$	0.777	1.191	2.184	2.015	8.019	0.811	0.304
$N=200$	$S\text{-}MT_{R_{5\%}}$	0.771	0.886	1.232	1.704	9.404	0.365	0.635
	$S\text{-}MT_{F_{5\%}}$	0.584	0.749	1.323	1.893	10.641	0.282	0.718
	$S\text{-}MT_{R_{10\%}}$	0.801	0.913	1.208	1.671	9.089	0.388	0.612
	$S\text{-}MT_{F_{10\%}}$	0.602	0.763	1.318	1.873	10.566	0.284	0.716
	FLM_{CV}	0.626	0.864	1.391	3.143	11.946	-	-
	$LW_{\hat{\Sigma}}$	0.199	0.575	1.403	2.267	11.893	1.005	0.218
$N=400$	$S\text{-}MT_{R_{5\%}}$	0.357	0.387	0.607	1.884	14.898	0.354	0.646
	$S\text{-}MT_{F_{5\%}}$	0.281	0.322	0.641	1.980	16.365	0.316	0.684
	$S\text{-}MT_{R_{10\%}}$	0.374	0.401	0.597	1.852	14.483	0.378	0.622
	$S\text{-}MT_{F_{10\%}}$	0.285	0.325	0.640	1.976	16.322	0.300	0.700
	FLM_{CV}	0.292	0.342	0.660	2.981	17.382	-	-
	$LW_{\hat{\Sigma}}$	0.100	0.204	0.685	2.348	18.485	0.994	0.153
$T=100$, $k=1$								
$N=50$	$S\text{-}MT_{R_{5\%}}$	1.068	2.611	3.054	1.184	3.158	0.393	0.607
	$S\text{-}MT_{F_{5\%}}$	1.171	2.747	3.411	1.433	3.882	0.363	0.637
	$S\text{-}MT_{R_{10\%}}$	1.070	2.579	3.011	1.127	3.056	0.398	0.602
	$S\text{-}MT_{F_{10\%}}$	1.183	2.771	3.325	1.387	3.713	0.372	0.628
	FLM_{CV}	5.775	7.150	6.465	11.684	14.623	-	-
	$LW_{\hat{\Sigma}}$	0.071	2.592	3.746	1.664	4.466	0.608	0.526
$N=100$	$S\text{-}MT_{R_{5\%}}$	0.712	0.989	1.531	1.397	5.200	0.410	0.590
	$S\text{-}MT_{F_{5\%}}$	0.812	1.096	1.800	1.661	6.613	0.364	0.636
	$S\text{-}MT_{R_{10\%}}$	0.691	0.976	1.493	1.349	5.029	0.418	0.582
	$S\text{-}MT_{F_{10\%}}$	0.803	1.085	1.755	1.624	6.376	0.375	0.625
	FLM_{CV}	2.210	2.785	3.202	17.267	23.390	-	-
	$LW_{\hat{\Sigma}}$	-0.010	0.812	1.959	1.954	7.671	0.737	0.384
$N=200$	$S\text{-}MT_{R_{5\%}}$	0.196	0.403	0.940	1.472	7.385	0.407	0.593
	$S\text{-}MT_{F_{5\%}}$	0.060	0.381	1.132	1.719	9.601	0.345	0.655
	$S\text{-}MT_{R_{10\%}}$	0.207	0.409	0.917	1.421	7.108	0.417	0.583
	$S\text{-}MT_{F_{10\%}}$	0.085	0.388	1.109	1.691	9.323	0.356	0.644
	FLM_{CV}	0.506	1.332	1.671	14.199	23.706	-	-
	$LW_{\hat{\Sigma}}$	-0.467	0.649	1.281	2.248	11.709	0.968	0.269
$N=400$	$S\text{-}MT_{R_{5\%}}$	0.109	0.162	0.464	1.673	11.803	0.395	0.605
	$S\text{-}MT_{F_{5\%}}$	0.024	0.133	0.563	1.907	15.406	0.314	0.686
	$S\text{-}MT_{R_{10\%}}$	0.111	0.162	0.450	1.617	11.331	0.408	0.592
	$S\text{-}MT_{F_{10\%}}$	0.037	0.134	0.555	1.891	15.086	0.325	0.675
	FLM_{CV}	0.068	0.380	0.710	9.440	24.824	-	-
	$LW_{\hat{\Sigma}}$	-0.261	0.299	0.635	2.365	18.550	1.004	0.170

FLM=Fan, Liao and Mincheva and LW=Ledoit and Wolf.

$S\text{-}MT_R$ and $S\text{-}MT_F$ stand for shrinkage on multiple testing by row and on the full sample correlation matrix.

$S\text{-}MT_R$ and $S\text{-}MT_F$ are evaluated at 5% and 10% significance levels.

FLM approach uses cross validation to evaluate the thresholding parameter. LW method is applied to the sample covariance matrix.

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, Dardari and Simon (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*

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

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, Dardari and Simon (2003, p.840).we have

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

where $\mathrm{erf}\,c(x)$ is the complement of the $\mathrm{erf}(x)$ function defined by (A.4). But

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

and using (A.6) we have

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

■

Lemma 4 *(i) Under assumption 1,*

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

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

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

(ii) Under assumptions 1 and 2,

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

Proof. *(i).Under (11) of assumption 1*

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

The required result follows trivially,

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

(ii). From part (i) it follows that

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

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

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

Hence,

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

■

A.2 Proofs of theorems for MT estimator

Proof of Theorem 1. Consider

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

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

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

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

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

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 (12) and (13) 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 (12) and (13) 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

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

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

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

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

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

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)

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

Overall, the order of the final term is given by

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

Putting the results for all the three terms together we now have

$$E\left\|\tilde{\boldsymbol{R}} - \boldsymbol{R}\right\|_F^2 = O\left(\frac{m_N N}{T}\right), \text{ if } N b_N \exp\left(\frac{-1}{2} b_N^2\right) = O(1).$$

From (A.3) of Lemma 2 setting $b_N = \sqrt{2\left[\ln f(N) - \ln(p)\right]}$ we have that

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

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

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

$$FPR = \frac{\sum\sum_{i \neq j} I\left(\left|\hat{\rho}_{ij}\right| > \frac{b_N}{\sqrt{T}} |\rho_{ij} = 0\right)}{N(N - m_N - 1)}. \tag{A.10}$$

Taking the expectation of (A.10) we have

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

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 (12) and (13) 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, $E\,|FPR| = \frac{p}{f(N)} \to 0$ 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 resultis established. This holds irrespective of the order by which N and $T \to \infty$.

Consider next the TPR statistic given by (15) 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

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 optimal 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

$$N^{-1}E\left\|\hat{\boldsymbol{R}}_{LW}-\boldsymbol{R}\right\|_F^2=N^{-1}\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}-\rho_{ij}\right)^2+\xi^2N^{-1}\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)-2\xi N^{-1}\sum\sum_{i\neq j}E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right], \tag{A.11}$$

and following Ledoit and Wolf (2003,2004) and Schäfer and Strimmer (2005) the optimal value of ξ that minimizes (A.11) is given by

$$\xi^*=\frac{\sum\sum_{i\neq j}E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right]}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}=1-\frac{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}. \tag{A.12}$$

Using (12) of Assumption 1 we have that

$$b_{ij}=E(\hat{\rho}_{ij})-\rho_{ij}=-\frac{\rho_{ij}(1-\rho_{ij}^2)}{2T}+\frac{G(\rho_{ij})}{T^2}. \tag{A.13}$$

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

$$1-\xi^*=\frac{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}=\frac{\sum\sum_{i\neq j}\rho_{ij}(b_{ij}+\rho_{ij})}{\sum\sum_{i\neq j}Var\left(\hat{\rho}_{ij}\right)+\sum\sum_{i\neq j}(b_{ij}+\rho_{ij})^2}. \tag{A.14}$$

Substituting for (13) of Assumption 1 and (A.13) in (A.14) yields

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

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 $\lim_{T\to\infty}(\hat{\xi}^*)=0$ 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.12) in (A.11) we have that

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

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

$$N^{-1}E\left\|\hat{\boldsymbol{R}}_{LW}-\boldsymbol{R}\right\|_F^2=N^{-1}\frac{\begin{array}{c}\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)-2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)+\sum\sum_{i\neq j}\rho_{ij}^2\right]\\-\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)\right]^2-\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\right]^2+2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\end{array}}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}.$$

Hence,

$$\begin{aligned}N^{-1}E\left\|\hat{\boldsymbol{R}}_{LW}-\boldsymbol{R}\right\|_F^2 &= N^{-1}\frac{\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)-\left[\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}\rho_{ij}\right)\right]^2}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}\\
&= N^{-1}\frac{\sum\sum_{i\neq j}\rho_{ij}^2[\sum\sum_{i\neq j}Var\left(\hat{\rho}_{ij}\right)+\sum\sum_{i\neq j}(b_{ij}+\rho_{ij})^2]-\left[\sum\sum_{i\neq j}\rho_{ij}(b_{ij}+\rho_{ij})\right]^2}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}\\
&= N^{-1}\frac{\begin{array}{c}\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}Var\left(\hat{\rho}_{ij}\right)+\left[\sum\sum_{i\neq j}\rho_{ij}^2\right]^2+\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}b_{ij}^2+2\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}b_{ij}\rho_{ij}\\-\left[\sum\sum_{i\neq j}b_{ij}\rho_{ij}\right]^2-\left[\sum\sum_{i\neq j}\rho_{ij}^2\right]^2-2\left[\sum\sum_{i\neq j}b_{ij}\rho_{ij}\right]\left[\sum\sum_{i\neq j}\rho_{ij}^2\right]\end{array}}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}\\
&= N^{-1}\frac{\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}Var\left(\hat{\rho}_{ij}\right)+\sum\sum_{i\neq j}\rho_{ij}^2\sum\sum_{i\neq j}b_{ij}^2-\left[\sum\sum_{i\neq j}b_{ij}\rho_{ij}\right]^2}{\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)}.\end{aligned}$$

For $E\left(\hat{\rho}_{ij}^2\right)$, using (12) and (13) of Assumption 1, we have that

$$\begin{aligned}E\left(\hat{\rho}_{ij}^2\right) &= Var(\hat{\rho}_{ij})+[E\left(\hat{\rho}_{ij}\right)]^2\\
&= \rho_{ij}^2+\frac{\rho_{ij}^2(1-\rho_{ij}^2)^2}{4T^2}+\frac{\rho_{ij}^2(1-\rho_{ij}^2)}{T}+\frac{(1-\rho_{ij}^2)^2}{T}+O(\frac{1}{T^2}),\end{aligned}$$

and since

$$\begin{aligned}\sum\sum_{i\neq j}\rho_{ij}^2 &= O(m_N N),\sum\sum_{i\neq j}E\left(\hat{\rho}_{ij}^2\right)=O(m_N N)+O(\frac{N(N-1)}{T}),\\
\sum\sum_{i\neq j}b_{ij}^2 &= O(\frac{m_N N}{T^2}),\quad \sum\sum_{i\neq j}b_{ij}\rho_{ij}=O(\frac{m_N N}{T}),\end{aligned}$$

it follows from the above results that

$$N^{-1}E\left\|\hat{\boldsymbol{R}}_{LW}^{*}(\xi^*)-\mathbf{R}\right\|_F^2=O\left(\frac{N}{T}\right),$$

which is in line with the result found 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

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

with $\boldsymbol{A}=\boldsymbol{R}_0^{-1}$ and $\boldsymbol{B}\left(\lambda\right)=\tilde{\boldsymbol{R}}_{MT}^{-1}(\lambda)$. 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

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

or

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

which is generally non-zero. Also, $\lambda = 0$ is ruled out, since $\tilde{\boldsymbol{R}}(0) = \tilde{\boldsymbol{R}}$ 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 $\tilde{\boldsymbol{R}}(\lambda_0)$ is non-singular. First, we can compute λ_0 by implementing naive shrinkage as an initial estimate:

$$
\tilde{\boldsymbol{R}}(\lambda_0) = \lambda_0 \boldsymbol{I}_N + (1-\lambda_0)\tilde{\boldsymbol{R}}.
$$

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

$$
\lambda_0 = \max\left(\frac{0.01-\lambda_{\min}\left(\tilde{\boldsymbol{R}}\right)}{1-\lambda_{\min}\left(\tilde{\boldsymbol{R}}\right)}, 0\right).
$$

Here, $\lambda_{\min}(\boldsymbol{A})$ stands for the minimum eigenvalue of matrix $\boldsymbol{A}$. If $\tilde{\boldsymbol{R}}$ is already positive definite and $\lambda_{\min}\left(\tilde{\boldsymbol{R}}\right) > 0$, then λ_0 is automatically set to zero. Conversely, if $\lambda_{\min}\left(\tilde{\boldsymbol{R}}\right) \leq 0$, then λ_0 is set to the smallest possible value that ensures positivity of $\lambda_{\min}\left(\tilde{\boldsymbol{R}}(\lambda_0)\right)$.

Second, we implement the optimisation procedure. In our simulation study and empirical applications we employ a grid search for $\lambda^* = \{\lambda : \lambda_0 \leq \lambda \leq 1\}$ 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 (24) or (26) where both $\hat{\theta}_{ij}$ and ω_T are adjusted to

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

and

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

Then (26) becomes

$$
\tilde{\boldsymbol{\Sigma}}_1^{(s)}(c) = \left(\hat{\sigma}_{1,ij}^{(s)} I\left[\left|\hat{\sigma}_{1,ij}^{(s)}\right| \geq \tau_{1,ij}^{(s)}(c)\right]\right), \tag{B.15}
$$

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,

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

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 $\tilde{\mathbf{\Sigma}}_{\hat{C}}$. The thresholding approach does not necessarily ensure that the resultant estimate, $\tilde{\mathbf{\Sigma}}_{\hat{C}}$, 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 $\lambda_{\min}\left(\tilde{\mathbf{\Sigma}}_{\hat{C}}\right) > 0$. Therefore, we modify (B.17) so that

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

where C_{pd} is the lowest c such that $\lambda_{\min}\left(\tilde{\mathbf{\Sigma}}_{C_{pd}}\right) > 0$. 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. (2012). Design-free estimation of large variance matrices. Working paper, Imperial College London.

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.

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.

Banerjee, O., Ghaoui, L. E. and D'Aspermont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. Journal of Machine Learning Research, 9, 485-516.

Barndor-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.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2007). Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: an exact simulation-based approach. Journal of Business and Economic Statistics, 25, 398-410.

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. 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, no. 494, 594-607.

Cai, T. T., and Zhou, H. H. (2009). Minimax estimation of large covariance matrices under ℓ_1 norm. Technical report, Department of Statistics, University of Pennsylvania.

Cai, T. T., and Zhou, H. H. (2010). Optimal rates of convergence for sparse covariance matrix estimation. Technical report, Department of Statistics, University of Pennsylvania.

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

Chamberlain, G. (1983). Funds, factors and diversification in arbitrage pricing models. Econometrica, 51, 1305-1324.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 51, 1281-1304.

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.

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.

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.

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. Annals of Statistics, 13, 1581-1591.

Donoho, D. L., Johstone, 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. (1975). Biased versus unbiased estimation. Advanced Mathematics, 16, 259-277.

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

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.

Fama, E. F. and French, K.R. (2004). The capital asset pricing model: Theory and evidence. Journal of Economic Perspectives, 18, 25-46.

Fan, J. (1997). Comments on 'Wavelets in statistics: A review' by A. Antoniadis. Journal of Italian Statistical Association, 6, 131-139.

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. and Li, R. (2001). Variable selection via nonconcave penalised likelihood and its oracle properties. Journal of American Statistical Association, 96, 1348-1360.

Fan, J., Liao, Y. and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. 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 coefficient 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.

Gibbons, M. R., Ross, S. A., Shanken, J. (1989). A test of the efficiency of a given portfolio. Econometrica, 57, 1121-1152.

Green, R. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? Journal of Finance, 47, 1785-1809.

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

Gungor, S. and Luger, R. (2009). Exact distribution-free tests of mean-variance efficiency. Journal of Empirical Finance, 16, 816-829.

Gungor, S. and Luger, R. (2011). Testing linear factor pricing models with large cross-sections: A distribution free approach. Georgia State University, mimeo.

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

Haff, L. R. (1991). The variational form of certain Bayes estimators. 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, no. 2, 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.

Johnstone, I. M. and Lu, A. Y. (2004). Sparse principal components analysis. Unpublished manuscript.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrices estimation. 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.

Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics, 47, 13-37.

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.

Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7, no. 1, 77-91.

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

Pesaran, M. H. (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 and Economic Statistics, 22, 129-162.

Pesaran, M. H. and Yamagata, T. (2012). Testing CAPM with a large number of assets. http://ssrn.com/abstract=2020423, AFA 2013 San Diego Meetings Paper.

Pesaran, M. H. and Zaffaroni, P. (2009). Optimal asset allocation with factor models for large portfolios. SSRN working paper.

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, no. 3, 369-387.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13, 341-360.

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., Levina, E. and Zhu, J. (2009). Generalised thresholding of large covariance matrices. Journal of American Statistical Association, 104, no. 485, 177-186.

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, no. 3, 539-550.

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. (2004). Factor representing portfolios in large asset markets. Journal of Econometrics, 119, 257-289.

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

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

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium. Journal of Finance, 19, no. 3, 425-442.

Soper, H. E. (1913). On the probable error of the correlation coefficient to a second approximation. Biometrika, 9, no. 1-2, 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.

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 Sininca, 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, no. 4, 431-442.