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

Natalia Bailey[1] M. Hashem Pesaran[2] L. Vanessa Smith[3]

[1]Department of Econometrics & Business Statistics, Monash University

[2]Department of Economics & USC Dornsife INET, University of Southern California, USA, and Trinity College, Cambridge

[3]University of York

10th ECB Workshop on Forecasting Techniques

June 18-19 2018

Introduction

Improved estimation of large covariance matrices is a problem that features prominently in a number of areas of multivariate statistical analysis.

- Finance: Portfolio selection and optimisation (Ledoit and Wolf (2003))
- Risk management (Fan et al. (2008))
- Testing of capital asset pricing models (Sentana (2009); Pesaran and Yamagata (2012))
- Global macro-econometric modelling (Pesaran et al. (2004); Dees et al. (2007))
- Bio-informatics: inferring large-scale gene association networks (Carroll (2003); Schäfer and Strimmer (2005))
- Other fields: meteorology, climate research, spectroscopy, signal processing and pattern recognition.

Introduction

- Consider the estimation of an $N \times N$ dimensional population covariance matrix, $\mathbf{\Sigma}$, where $\mathbf{\Sigma} \succ 0$ and the number of cross section units, N, is large.
- Especially when N is larger than the number of observations (T) in a given sample, one suitable estimator of $\mathbf{\Sigma}$ appropriately restricts the off-diagonal elements of its sample estimate denoted by $\hat{\mathbf{\Sigma}}$.
- Three branches of literature that address this problem are:
  - Regression-based approaches: make use of suitable decompositions of $\mathbf{\Sigma}$ (e.g. Pourahmadi (1999, 2000), Rothman et al. (2010), Abadir et al. (2014)).
  - Banding or tapering methods: rely on a natural ordering among variables (e.g. Bickel and Levina (2004, 2008a), Wu and Pourahmadi (2009)).
  - Thresholding (e.g. El Karoui (2008), Cai and Liu (2011)) and Shrinkage (e.g. Ledoit and Wolf (2004, 2012)) techniques: do not make use of any ordering assumptions.

Regularisation technique: Thresholding

- Involves setting off-diagonal elements of the sample covariance matrix that are in absolute terms below a certain 'threshold' value(s), to zero.
- The selected non-zero elements of $\hat{\Sigma}$ can be:
  - set at their sample estimates $\Longrightarrow$ Hard thresholding
  - somewhat adjusted downward $\Longrightarrow$ Soft thresholding
- Universal thresholding (El Karoui, 2008; Bickel and Levina, 2008b - BL)
  - Applies the same thresholding parameter to all off-diagonal elements of the unconstrained sample covariance matrix.
- Adaptive thresholding (Cai and Liu, 2011 - CL)
  - Allows the threshold value to vary across the different off-diagonal elements of the matrix.
- Main assumption: the underlying covariance matrix is *sparse*.
  - Sparseness: 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.

This paper's contribution

- Proposes a multiple testing (*MT*) estimator as an alternative thresholding procedure
- It is employed directly to the sample correlation matrix and thus avoids the scaling problem associated with the use of covariances
- It tests the statistical significance of all pair-wise correlations and determines the thresholding parameter as part of a *MT* strategy
- Derives the rate of convergence of the error in estimation of the population correlation matrix, $\boldsymbol{R} = (\rho_{ij})$, using the *MT* method under the spectral and Frobenius norms and its support recovery (TPR,FPR, and FDR)
- Investigates its finite sample properties by use of a Monte Carlo study

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).
- The covariance matrix of $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{Nt})'$ is given by

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

  where $E(\boldsymbol{x}_t) = \boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)'$, and $\boldsymbol{\Sigma}$ is an $N \times N$ symmetric, positive definite real matrix with (i,j) element, σ_{ij}.
- We assume that x_{it} is independent over time, t.
- We consider the regularisation of the sample correlation matrix $\widehat{\mathbf{R}} = (\hat{\rho}_{ij,T})$,

$$\hat{\rho}_{ij,T} = \frac{\hat{\sigma}_{ij,T}}{\sqrt{\hat{\sigma}_{ii,T}\hat{\sigma}_{jj,T}}}, \tag{1}$$

  as an estimator of $\mathbf{R} = (\rho_{ij} = \sigma_{ij}/\sigma_{ii}^{1/2}\sigma_{jj}^{1/2})$, where

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

Assumptions

Assumption 1: The population covariance matrix, $\mathbf{\Sigma} = (\sigma_{ij})$, is sparse such that

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

is $O\left(N^{\vartheta}\right)$ for some $0 \leq \vartheta < 1/2$, where $I(A)$ is an indicator function that takes the value of 1 if A holds and zero otherwise.

Assumption 2: Let $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ with:

- mean $\mu_i = E(x_{it})$, $|\mu_i| < K$,
- variance $\sigma_{ii} = Var(x_{it})$, $0 < \sigma_{ii} < K$,
- correlation coefficient $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$, where $\left|\rho_{ij}\right| < 1$,
- $E\left|y_{it}\right|^{2s} < K < \infty$, for some positive integer $s \geq 3$.

Also, let $\boldsymbol{\xi}_{ij,t} = (y_{it}, y_{jt}, y_{it}^2, y_{jt}^2, y_{it}y_{jt})'$ such that for any $i \neq j$ the time series observations $\boldsymbol{\xi}_{ij,t}$, $t = 1, 2, \ldots, T$, are random draws from a common distribution which is absolutely continuous with non-zero density on subsets of $\mathbb{R}^5$.

Multiple Testing problem

- For a given i and j, then under $H_{0,ij} : \sigma_{ij} = 0$, $\sqrt{T}\hat{\rho}_{ij,T}$ is asymptotically distributed as $N(0,1)$ for T sufficiently large.
- To account for the multiple nature of the $N(N-1)/2$ pair-wise tests, we propose using the following critical value function for testing $\rho_{ij} = 0$

$$c_p = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right), \tag{4}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the CDF of a standard normal variate, and p is the nominal size of the test. For $f(N)$ we consider

$$f(N) = c_\delta N^\delta, \tag{5}$$

for some positive finite constants c_δ and δ.

Multiple Testing (MT) estimator

- We derive conditions on δ which ensure consistent support recovery and a suitable convergence rate of the error in estimation of $\boldsymbol{R} = (\rho_{ij})$.
- We show that in both respects the choice of δ depends on the nature of dependence the pairs (y_{it}, y_{jt}), for all $i \neq j$, and on the relative rate at which N and T rise.
- The degree of dependence is defined by $K_v = \sup_{ij} K_v(\boldsymbol{\theta}_{ij})$ where $\boldsymbol{\theta}_{ij}$ is a vector of cumulants of (y_{it}, y_{jt}).
- When $\rho_{ij} = 0$ for all i and j, $i \neq j$, this parameter is given by $\varphi_{\max} = \sup_{ij} (\varphi_{ij})$ where $\varphi_{ij} = E\left(y_{it}^2 y_{jt}^2 \,|\rho_{ij} = 0\right) > 0$.
- Under independence, $\varphi_{\max} = 1$.

Multiple Testing (MT) estimator

- The multiple testing (MT) estimator of $\mathbf{R}$, $\widetilde{\boldsymbol{R}}_{MT} = (\tilde{\rho}_{ij})$, is then given by

$$\widetilde{\boldsymbol{R}}_{MT} = (\tilde{\rho}_{ij,T}) = \hat{\rho}_{ij,T} I \left[\left| \hat{\rho}_{ij,T} \right| > T^{-1/2} c_p(N) \right], \quad (6)$$

for $i = 1, 2, \ldots, N-1$, $j = i+1, \ldots, N$.

- Hence, the MT estimator of $\boldsymbol{\Sigma}$ is given by

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

where $\hat{\boldsymbol{D}} = diag(\hat{\sigma}_{11,T}, \hat{\sigma}_{22,T}, \ldots, \hat{\sigma}_{NN,T})$.

- In principle, the MT procedure can be applied to estimated residuals from a factor model, following the reasoning of Fan et al. (2011, 2013).

Theoretical properties of the MT estimator

Proposition 1

Suppose that Assumption 2 holds and $\hat{\rho}_{ij,T}$ is defined by (1). Then, for $\boldsymbol{\theta}_{ij} = (\rho_{ij}, \mu_{ij}(0,4) + \mu_{ij}(4,0), \mu_{ij}(3,1) + \mu_{ij}(1,3), \mu_{ij}(2,2))'$ with $E\left(y_{it}^{r} y_{jt}^{s}\right) = \mu_{ij}(r,s)$, for all $r, s \geq 0$,

$$\rho_{ij,T} = E\left(\hat{\rho}_{ij,T}\right) = \rho_{ij} + \frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right), \quad (7)$$

$$\omega_{ij,T}^{2} = Var\left(\hat{\rho}_{ij,T}\right) = \frac{K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right), \text{ where} \quad (8)$$

$$K_{\mathrm{m}}(\boldsymbol{\theta}_{ij}) = -\frac{1}{2}\rho_{ij}(1-\rho_{ij}^{2}) + \frac{1}{8}\rho_{ij}\left\{ \begin{array}{c} 3\left[\kappa_{ij}(4,0) + \kappa_{ij}(0,4)\right] - 4\left[\kappa_{ij}(3,1) + \kappa_{ij}(1,3)\right] \\ +2\kappa_{ij}(2,2) \end{array} \right\},$$

$$K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) = (1-\rho_{ij}^{2})^{2} + \frac{1}{4}\left\{ \begin{array}{c} \rho_{ij}^{2}\left[\kappa_{ij}(4,0) + \kappa_{ij}(0,4)\right] - 4\rho_{ij}\left[\kappa_{ij}(3,1) + \kappa_{ij}(1,3)\right] \\ +2(2+\rho_{ij}^{2})\kappa_{ij}(2,2) \end{array} \right\}.$$

Furthermore $|K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})| < K$, $K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) = \lim_{T\to\infty}\left[T Var\left(\hat{\rho}_{ij,T}\right)\right]$, and $K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) < K$.

Remarks

- The results for $E\left(\hat{\rho}_{ij,T}\right)$ and $Var\left(\hat{\rho}_{ij,T}\right)$ are established in Gayen (1951) using a bivariate Edgeworth expansion approach.
- These results hold for any law of dependence between x_{it} and x_{jt}.
- In the case where $\rho_{ij} = 0$,

$$\varphi_{ij} := K_{\text{v}}(\boldsymbol{\theta}_{ij} \,|\rho_{ij} = 0\,) = E\left(y_{it}^2 y_{jt}^2 \,|\rho_{ij} = 0\,\right) > 0, \tag{9}$$

and

$$\psi_{ij} := K_{\text{m}}(\boldsymbol{\theta}_{ij} \,|\rho_{ij} = 0\,) = -0.5\left[E\left(y_{it}^3 y_{jt} \,|\rho_{ij} = 0\,\right) + E\left(y_{it} y_{jt}^3 \,|\rho_{ij} = 0\,\right)\right]. \tag{10}$$

- When y_{it} and y_{jt} are independently distributed, then under $\rho_{ij} = 0$, we have $\varphi_{ij} = E\left(y_{it}^2\right) E\left(y_{jt}^2\right) = 1$, and $\psi_{ij} = 0$.

Proposition 2

Suppose that Assumptions 1 and 2 hold. Then, the standardised correlation coefficients, $z_{ij,T} = \left[\hat{\rho}_{ij,T} - E\left(\hat{\rho}_{ij,T}\right)\right] / \sqrt{Var\left(\hat{\rho}_{ij,T}\right)}$, for all i and j $(i \neq j)$ admit the Edgeworth expansion

$$\Pr\left(z_{ij,T} \leq x\right) = \Phi(x) + \sum_{r=1}^{s-2} T^{-r/2} g_r\left(x\right)\phi(x) + O\left[T^{-(s-1)/2}\right], \tag{11}$$

for some positive integer $s \geq 3$, where:

- $E\left(\hat{\rho}_{ij,T}\right)$ and $Var\left(\hat{\rho}_{ij,T}\right)$ are defined by (7) and (8) of Proposition 1
- $\Phi(x)$ and $\phi\left(x\right)$ are the CDF and PDF of the standard Normal $(0, 1)$
- $g_r\left(x\right)$, $r = 1, 2, \ldots, s-2$, are finite polynomials in $x \in \mathbb{R}$ x of degree $3r-1$ whose coefficients do not depend on x

Proposition 2

- Further, for all finite $s \geq 3$, and $a_T > 0$, we have

$$\Pr\left(z_{ij,T} \leq -a_T\right) \leq K e^{-\frac{1}{2}a_T^2} + O\left[T^{-\frac{(s-2)}{2}} a_T^{3(s-2)-1} e^{-\frac{1}{2}a_T^2}\right] + O\left[T^{-(s-1)/2}\right], \tag{12}$$

and

$$\Pr\left(z_{ij,T} > a_T\right) \leq K e^{-\frac{1}{2}a_T^2} + O\left[T^{-\frac{(s-2)}{2}} a_T^{3(s-2)-1} e^{-\frac{1}{2}a_T^2}\right] + O\left[T^{-(s-1)/2}\right]. \tag{13}$$

Remark

This proposition establishes a bound on the probability of $\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > T^{-1/2} c_p(N)$ without requiring sub-Gaussianity, at the expense of the additional order term, $O\left[T^{-(s-1)/2}\right]$, that describes the order of the moments of $z_{ij,T}$.

Theorem 1

(*Convergence under the spectral norm*) Consider $\hat{\rho}_{ij,T}$ defined by (1). Suppose that Assumptions 1 and 2 hold. Let

$$c_p(N) = \Phi^{-1}\left(1 - \frac{p}{2f(N)}\right), \tag{14}$$

where $0 < p < 1$. δ and d are the exponents in $f(N) = c_\delta N^\delta$, and $T = c_d N^d$, with $c_\delta, c_d > 0$. Further, suppose that there exists N_0 s.t. for all $N > N_0$,

$$1 - \frac{p}{2f(N)} > 0, \tag{15}$$

and $\rho_{\min} > c_p(N)/\sqrt{T}$, where $\rho_{\min} = \min_{ij}(\left|\rho_{ij}\right|, \rho_{ij} \neq 0)$.Consider values of δ that satisfy condition

$$\delta > \frac{2K_v}{(1-\gamma)^2}, \tag{16}$$

for some small positive constant γ, where $K_v = \sup_{ij} K_v(\boldsymbol{\theta}_{ij})$ and $K_v(\boldsymbol{\theta}_{ij})$ is defined in Proposition 1. Then for all values of $d > 4/(s-1)$,

$$\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\| = O_p\left(\frac{m_N c_p(N)}{\sqrt{T}}\right). \tag{17}$$

Remarks

In view of the conditions of Theorem 1, we note that:

- The parameter d is required to be sufficiently large such that such that $d > 4/(s-1)$ and $T^{-1/2}c_p(N) \to 0$, as $N \to \infty$.
    - This will be met if $N^{-d}\ln(N) \to 0$ as $N \to \infty$, which mathematically holds true since we must also have $d > 4/(s-1)$.
- Condition (15) is met for $\delta > 0$ and N sufficiently large.
- When the expansion rate of N is larger than T, it is required that $s > 5$, which means that x_{it} must have moments of order 10 or more.

Remarks

- Condition $\rho_{\min} > c_p(N)/\sqrt{T}$ can be written as

$$\rho_{\min}^2 > \frac{c_p^2(N)}{T} = \frac{c_p^2(N)}{c_d N^d} = c_d^{-1}\left[\frac{c_p^2(N)}{\ln(N)}\right]\left[\frac{\ln(N)}{N^d}\right].$$

  - This is satisfied for any $\delta > 0$, even if $\rho_{\min} \to 0$ with N, so long as the rate at which $\rho_{\min}$ tends to zero is slower than $\sqrt{\ln(N)/N^d}$, for some $d > 4/(s-1)$.
  - This latter condition still allows N to rise much faster than T.

- Under Gaussianity where $K_v = \sup_{ij} K_v(\theta_{ij}) = 1$, condition (16) becomes $\delta > \frac{2}{(1-\gamma)^2}$. In general,
  - The spectral norm result requires δ to be set above $2\sup_{ij} K_v(\theta_{ij})$.
  - This is larger than the value of δ required for the Frobenius norm obtained in the theorem that follows.

Theorem 2

(*Convergence under the Frobenius norm*) Consider $\hat{\rho}_{ij,T}$ defined by (1). Suppose that conditions of Theorem 1 hold, but (16) is replaced by the weaker condition on δ

$$\delta > (2-d)\,\varphi_{\max}, \tag{18}$$

where δ and d are as defined in Theorem 1 and $\varphi_{\max} = \sup_{ij}\left(\varphi_{ij}\right)$ where $\varphi_{ij} = E\left(y_{it}^2 y_{jt}^2 \left| \rho_{ij} = 0\right.\right) > 0$. Then, for $d > \max\left(\frac{2+\vartheta}{s-1}, \frac{4}{s+1}\right)$ we have

$$E\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F = O\left(\sqrt{\frac{m_N N}{T}}\right), \tag{19}$$

and

$$\left\|\widetilde{\boldsymbol{R}}_{MT} - \boldsymbol{R}\right\|_F = O_p\left(\sqrt{\frac{m_N N}{T}}\right), \tag{20}$$

where m_N is defined in Theorem 1, with $m_N = O(N^{\vartheta})$, where $0 \leq \vartheta < 1/2$.

Remarks

In view of the conditions of Theorem 2, we note that:

- Condition (18) implies that δ should be set at a sufficiently high level, determined by d (the relative expansion rates of N and T), and $\varphi_{\max}$ (the maximum degree of dependence between y_{it} and y_{jt} when $\rho_{ij} = 0$).
- The Frobenius norm result holds even if N rises faster than T.
- In the case where N and T are of the same order of magnitude ($d = 1$), and where y_{it} and y_{jt} are independently distributed when $\rho_{ij} = 0$ ($\varphi_{\max} = 1$), then the Frobenius norm results require $\delta > 1$.
- Finally, by allowing for $\varphi_{\max}$ to differ from unity our analysis applies to non-Gaussian processes.

Comparison with thresholding literature

- For the spectral norm, the term $c_p(N)$ compares to $\sqrt{\ln(N)}$ obtained in the literature for the probability order, $O_p\left(\frac{m_N\sqrt{\ln(N)}}{\sqrt{T}}\right)$, since $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$, for $\delta > 0$.
- For the Frobenius norm, the order of convergence in (19) is a slight improvement on existing rates in the thresholding literature:
  - Setting $q = 0$, convergence rate under the Fronenius norm is only obtained in BL under the Gaussianity assumption, stated as $\left\|\tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_F = O_p\left(\sqrt{\frac{m_N N \log(N)}{T}}\right)$.
  - This arises from the fact that their result is derived by explicitly using their spectral norm convergence rate.
- The moment conditions required for y_{it} in Theorems 1 and 2 depend on d.
  - The related literature requires the stronger sub-Gaussianity assumption.
  - Our conditions are comparable to the polynomial-type tail assumption considered in CL for the spectral norm result

Comparison with Bonferroni procedure

- Application of the Bonferroni procedure to this MT problem is equivalent to setting $f(N) = N(N-1)/2$. Our theoretical results suggest that:
  - this can be too conservative if $\rho_{ij} = 0$ implies that y_{it} and y_{jt} are independent,
  - but could be appropriate otherwise depending on the relative rate at which N and T rise.
- In our Monte Carlo study, we:
  - consider observations complying with $\varphi_{\max} = 1$ and $\varphi_{\max} = 1.5$, and experiment with $\delta = \{1, 2\}$.
  - present results where δ is estimated by cross validation over the range $[1 - 2.5]$.
  - find that the simulation results conform closely to our theoretical findings.

Support recovery

Consider the following statitics computed using the MT estimator, $\widetilde{\boldsymbol{R}}_{MT}$, given by (6):

1. The true positive rate,

$$TPR = \frac{\sum\sum_{i\neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} \neq 0)}{\sum\sum_{i\neq j} I(\rho_{ij} \neq 0)}, \quad (21)$$

2. the false positive rate,

$$FPR = \frac{\sum\sum_{i\neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum\sum_{i\neq j} I(\rho_{ij} = 0)}, \quad (22)$$

3. the false discovery rate,

$$FDR = \frac{\sum\sum_{i\neq j} I(\tilde{\rho}_{ij,T} \neq 0, \text{ and } \rho_{ij} = 0)}{\sum\sum_{i\neq j} I(\rho_{ij} \neq 0)}. \quad (23)$$

Theorem 3

(*Support Recovery*) Consider the true positive rate (TPR), the false positive rate (FPR) and the false discovery rate (FDR) statistics defined by (21), (22) and (23) respectively. Then for $0 \leq \vartheta < 1/2$, as $N \to \infty$ we have:

$$TPR_N \overset{a.s.}{\to} 1, \text{ for } \delta > 0, \text{ and } d > 2/(s-1)$$
$$FPR_N \overset{a.s.}{\to} 0, \text{ for } \delta > \varphi_{\max}, \text{ and } d > 2/(s-1)$$
$$FDR_N \overset{a.s.}{\to} 0, \text{ for } \delta > (2-\vartheta)\varphi_{\max}, \text{ and } d > 2\,(2-\vartheta)\,/(s-1),$$

where $\varphi_{\max} = \sup_{ij} E\left(y_{it}^2 y_{jt}^2 \,|\rho_{ij} = 0\right) > 0$, with $y_{it} = (x_{it} - \mu_i)/\sqrt{\sigma_{ii}}$ (see Assumption 2).
Further, as $N \to \infty$,

- $TPR_N \to 1$ and $FPR_N \to 0$, in probability for any $\delta > 0$ and $d > 2/(s-1)$;
- and $FDR_N \to 0$, in probability if $\delta > (1-\vartheta)\varphi_{\max}$, and $d > 2\,(1-\vartheta)\,/(s-1)$.

Small sample properties

- We consider two experiments:

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

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

- The covariances in Monte Carlo Designs A and B are examples of *exact* sparse covariance matrices

- Results are reported for $N = \{30, 100, 200\}$ and $T = 100$

Data generating process

- We generate the standardised variates, y_{it},

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

where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$, and $\boldsymbol{P}$ is the Cholesky factor associated with the choice of the correlation matrix $\boldsymbol{R} = \boldsymbol{P}\boldsymbol{P}'$.

- We consider two alternatives for the errors, u_{it}:
  - Gaussian case, $u_{it} \sim IIDN(0, 1)$ for all i and t (benchmark)
    Here, under $\rho_{ij} = 0$, $E\left(y_{it}^2 y_{jt}^2\right) = 1$ hence setting $\delta = 1$ is sufficient
  - Multivariate t-distribution case (v degrees of freedom)

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

where $\varepsilon_{it} \sim IIDN(0, 1)$, and $\chi^2_{v,t}$ is a chi-squared r.v. with $v > 4$, distributed independently of ε_{it} for all i and t.

Here $v = 8$ so that $E\left(y_{it}^6\right)$ exists (Assumption 2). Also, under $\rho_{ij} = 0$, $\varphi_{ij} = E\left(y_{it}^2 y_{jt}^2\right) = (v-2)/(v-4)$, and with $v = 8$ we have $\varphi_{\max} = 1.5$. Hence, setting $\delta = 2$ is sufficient (Theorem 2)

Data generating process

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

$$\boldsymbol{x}_t = \mathbf{a} + \boldsymbol{\gamma} f_t + \mathbf{D}^{1/2}\mathbf{y}_t, \qquad (24)$$

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

- We focus on the baseline case where $\boldsymbol{\gamma} = \mathbf{0}$ and $\mathbf{a} = \mathbf{0}$.
- We also consider the DGP that draws γ_i and a_i as $IIDN(1, 1)$ for $i = 1, 2, \ldots, N$, and generates f_t, the common factor, as a stationary AR(1) process.
- Under both settings we focus on the residuals from an OLS regression of $\boldsymbol{x}_t$ on an intercept and a factor (if needed).

Monte Carlo Design A

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

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

where $\boldsymbol{A}_1 = \boldsymbol{A} + \epsilon \boldsymbol{I}_{N/2}$, $\boldsymbol{A} = (a_{ij})_{1 \leq i,j \leq N/2}$, $a_{ij} = (1 - \frac{|i-j|}{10})_+$ with $\epsilon = \max(-\lambda_{\min}(\boldsymbol{A}), 0) + 0.01$ to ensure that $\boldsymbol{A}$ is positive definite, and $\boldsymbol{A}_2 = 4\boldsymbol{I}_{N/2}$.

- $\boldsymbol{\Sigma}$ is a two block diagonal matrix
- $\boldsymbol{A}_1$ is a banded and sparse covariance matrix
- $\boldsymbol{A}_2$ is a diagonal matrix with 4 along the diagonal
- $\boldsymbol{P}$ is obtained numerically by applying the Cholesky decomposition to the correlation matrix, $\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}^{-1/2} = \mathbf{PP}'$
- The diagonal elements of $\boldsymbol{D}$ are given by $\sigma_{ii} = 1 + \epsilon$, for $i = 1, 2, \ldots, N/2$ and $\sigma_{ii} = 4$, for $i = N/2 + 1, \ldots, N$.

Monte Carlo Design B

Here we explicitly control for the number of non-zero elements of the population correlation matrix.

- First we draw $N \times 1$ vectors $\boldsymbol{b} = (b_1, b_2, \ldots, b_N)'$ as *Uniform* $(0.7, 0.9)$ for the first and last N_b $(< N)$ elements of $\mathbf{b}$, where $N_b = \left[N^{\beta}\right]$, and set the remaining middle elements of $\mathbf{b}$ to zero.
- The resulting population correlation matrix $\boldsymbol{R}$ is defined by

$$\boldsymbol{R} = \boldsymbol{I}_N + \boldsymbol{b}\boldsymbol{b}' - \text{diag}\left(\boldsymbol{b}\boldsymbol{b}'\right).$$

- The degree of sparseness of $\boldsymbol{R}$ is determined by the value of the parameter β. We are interested in weak cross-sectional dependence, so we focus on the case where $\beta < 1/2$ and set $\beta = 0.25$.
- We set $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{R}\boldsymbol{D}^{1/2}$ where the diagonal elements of $\boldsymbol{D}$ are given by $\sigma_{ii} \sim IID\ \left(1/2 + \chi^2(2)/4\right), i = 1, 2, \ldots, N$.
- $\boldsymbol{P}$ is then obtained by applying the Cholesky decomposition to $\boldsymbol{R}$

Alternative estimators

- MT_1: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\boldsymbol{\Sigma}}_{MT}$) using $\delta = 1$ ($\widetilde{\boldsymbol{\Sigma}}_{MT_1}$)
- MT_2: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\boldsymbol{\Sigma}}_{MT}$) using $\delta = 2$ ($\widetilde{\boldsymbol{\Sigma}}_{MT_2}$)
- $MT_{\hat{\delta}}$: thresholding based on the MT approach applied to the sample correlation matrix ($\widetilde{\boldsymbol{\Sigma}}_{MT}$) using cross-validated (CV) δ ($\widetilde{\boldsymbol{\Sigma}}_{MT_{\hat{\delta}}}$)
- $BL_{\hat{C}}$: BL thresholding on the sample cov matrix using CV C ($\widetilde{\boldsymbol{\Sigma}}_{BL,\hat{C}}$)
- CL_2: CL thresholding on the sample cov matrix using the theoretical value of $C = 2$ ($\widetilde{\boldsymbol{\Sigma}}_{CL,2}$)
- $CL_{\hat{C}}$: CL thresholding on the sample cov matrix using CV C ($\widetilde{\boldsymbol{\Sigma}}_{CL,\hat{C}}$)
- $LW_{\hat{\Sigma}}$: LW shrinkage on the sample covariance matrix ($\hat{\boldsymbol{\Sigma}}_{LW_{\hat{\Sigma}}}$)

Evaluation metrics

- We compute the spectral and Frobenius norms of the deviations of each of the regularised covariance matrices from their respective true $\mathbf{\Sigma}$:

$$\left\|\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}}\right\| \text{ and } \left\|\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}}\right\|_F,$$

where $\mathring{\mathbf{\Sigma}}$ is set to one of the following $\{\widetilde{\mathbf{\Sigma}}_{MT_1}, \widetilde{\mathbf{\Sigma}}_{MT_2}, \widetilde{\mathbf{\Sigma}}_{MT_{\hat{\delta}}}, \widetilde{\mathbf{\Sigma}}_{BL,\hat{C}}, \widetilde{\mathbf{\Sigma}}_{CL,2}, \widetilde{\mathbf{\Sigma}}_{CL,\hat{C}}, \hat{\mathbf{\Sigma}}_{LW_{\hat{\Sigma}}}\}$, where $\hat{\delta}$, $\hat{C}$ is a constant evaluated through cross-validation

- 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).

- TPR and FPR are not applicable to shrinkage techniques.

Monte Carlo Simulation Results

Robustness results of MT to the choice of the p-value

We considered $p = \{0.01, 0.05, 0.10\}$, and $\delta = \{1, 2\}$.

- For the average spectral and Frobenius norms (over 2000 replications) we have:
  - They are not much affected by the choice of p when setting $\delta = 1, 2$ or $\hat{\delta}$, given that the effective p-value amounts to $2p/N^{\delta}$ which is very small.
  - This is irrespective of whether the observations are drawn from a Gaussian or a multivariate t distribution.
- The results confirm our theoretical finding that:
  - In the case of Gaussian observations ($\varphi_{\max} = 1$) the scaling factor $\delta = 1$ is likely to perform better as compared to $\delta = 2$.
  - The reverse is true if the observations are multivariate t distributed and the scaling factor $\delta = 2$ is to be preferred.
  - The average estimates of $\hat{\delta}$ are also indicative that a higher value of δ is required when observations are multivariate t distributed.
- Overall, the simulation results support using a sufficiently high value of δ (say around 2) or its estimate, $\hat{\delta}$, obtained by cross validation.

Table: Spectral and Frobenius norm losses for the MT estimator using significance levels $p = \{0.01, 0.05, 0.10\}$ and $\delta = \{1, 2, \hat{\delta}\}$

Monte Carlo design A									
	$\delta = 1$			$\delta = 2$			$\hat{\delta}$		
N\p	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
$\mathbf{u}_{it}\sim$ Gaussian									
Spectral norm									
30	1.70(0.49)	1.68(0.49)	1.71(0.49)	1.89(0.51)	1.79(0.50)	1.75(0.50)	1.71(0.49)	1.68(0.49)	1.69(0.49)
100	2.61(0.50)	2.51(0.50)	2.50(0.50)	3.11(0.50)	2.91(0.50)	2.84(0.50)	2.62(0.50)	2.52(0.50)	2.51(0.50)
200	3.04(0.48)	2.92(0.49)	2.89(0.49)	3.67(0.47)	3.46(0.47)	3.37(0.47)	3.05(0.48)	2.93(0.49)	2.90(0.49)
Frobenius norm									
30	3.17(0.45)	3.14(0.50)	3.20(0.53)	3.49(0.42)	3.32(0.43)	3.26(0.43)	3.19(0.44)	3.13(0.48)	3.16(0.52)
100	6.67(0.45)	6.51(0.51)	6.60(0.55)	7.75(0.40)	7.34(0.41)	7.17(0.42)	6.70(0.45)	6.52(0.50)	6.57(0.54)
200	9.87(0.46)	9.60(0.53)	9.73(0.58)	11.76(0.40)	11.15(0.41)	10.89(0.42)	9.91(0.46)	9.62(0.52)	9.69(0.57)
$\mathbf{u}_{it}\sim$ multivariate $t-$ distributed with 8 degrees of freedom									
Spectral norm									
30	2.26(1.08)	2.42(1.20)	2.55(1.26)	2.29(0.90)	2.24(0.99)	2.24(1.03)	2.23(0.95)	2.32(1.04)	2.39(1.08)
100	3.85(4.84)	4.20(5.28)	4.46(5.48)	3.78(3.78)	3.71(4.12)	3.71(4.27)	3.67(3.81)	3.83(4.11)	3.93(4.21)
200	4.49(3.46)	5.04(4.34)	5.44(4.77)	4.26(1.80)	4.20(2.21)	4.19(2.37)	4.20(2.43)	4.45(2.78)	4.57(2.94)
Frobenius norm									
30	4.06(1.14)	4.35(1.32)	4.60(1.40)	4.12(0.90)	4.04(1.00)	4.03(1.06)	4.03(1.00)	4.19(0.13)	4.32(1.19)
100	8.88(5.17)	9.75(5.67)	10.49(5.87)	9.04(4.04)	8.80(4.40)	8.74(4.57)	8.65(4.16)	9.09(4.48)	9.41(4.59)
200	12.96(4.23)	14.50(5.41)	15.81(5.95)	13.25(2.10)	12.85(2.54)	12.71(2.76)	12.57(2.97)	13.25(3.48)	13.73(3.67)

	Cross validated values of δ		
N\p	0.01	0.05	0.10
$\mathbf{u}_{it}\sim$ Gaussian			
30	1.08(0.11)	1.10(0.12)	1.12(0.13)
100	1.04(0.06)	1.05(0.07)	1.06(0.08)
200	1.03(0.05)	1.03(0.06)	1.04(0.06)
$\mathbf{u}_{it}\sim$ multivariate t-distr. with 8 dof			
30	1.13(0.18)	1.19(0.22)	1.25(0.25)
100	1.12(0.18)	1.18(0.22)	1.23(0.25)
200	1.15(0.20)	1.20(0.23)	1.24(0.25)

Notes: The MT approach is implemented using $\delta = 1$, $\delta = 2$, and $\hat{\delta}$, computed using cross-validation. Norm losses and estimates of δ, $\hat{\delta}$, are averages over 2,000 replications. Simulation standard deviations are given in the parentheses.

Table: Spectral and Frobenius norm losses for the MT estimator using significance levels $p = \{0.01, 0.05, 0.10\}$ and $\delta = \{1, 2, \hat{\delta}\}$

	Monte Carlo design B								
	$\delta = 1$			$\delta = 2$			$\hat{\delta}$		
N\p	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
	$u_{it}\sim$ Gaussian								
	Spectral norm								
30	0.48(0.16)	0.50(0.16)	0.53(0.16)	0.50(0.20)	0.49(0.18)	0.48(0.17)	0.48(0.17)	0.49(0.16)	0.49(0.16)
100	0.75(0.34)	0.76(0.32)	0.78(0.31)	0.89(0.43)	0.81(0.39)	0.79(0.37)	0.76(0.35)	0.76(0.34)	0.76(0.34)
200	0.71(0.22)	0.74(0.20)	0.77(0.20)	0.85(0.33)	0.78(0.28)	0.75(0.26)	0.72(0.24)	0.72(0.22)	0.72(0.22)
	Frobenius norm								
30	0.87(0.17)	0.91(0.18)	0.97(0.19)	0.89(0.20)	0.87(0.17)	0.86(0.17)	0.86(0.17)	0.88(0.17)	0.88(0.17)
100	1.56(0.24)	1.66(0.24)	1.77(0.24)	1.67(0.34)	1.60(0.29)	1.58(0.27)	1.56(0.25)	1.58(0.24)	1.58(0.25)
200	2.16(0.18)	2.32(0.20)	2.50(0.21)	2.25(0.24)	2.19(0.21)	2.16(0.20)	2.15(0.18)	2.18(0.19)	2.18(0.20)
	$u_{it}\sim$ multivariate $t-$ distributed with 8 degrees of freedom								
	Spectral norm								
30	0.70(0.39)	0.78(0.43)	0.84(0.45)	0.67(0.33)	0.67(0.35)	0.67(0.37)	0.67(0.33)	0.68(0.35)	0.68(0.36)
100	1.16(0.97)	1.32(1.10)	1.42(1.18)	1.15(0.75)	1.11(0.80)	1.10(0.83)	1.10(0.72)	1.10(0.77)	1.11(0.80)
200	1.36(1.73)	1.65(2.05)	1.83(2.20)	1.14(1.03)	1.13(1.21)	1.14(1.28)	1.16(1.06)	1.19(1.20)	1.20(1.27)
	Frobenius norm								
30	1.23(0.42)	1.40(0.48)	1.53(0.51)	1.15(0.35)	1.16(0.38)	1.17(0.39)	1.17(0.36)	1.19(0.38)	1.20(0.39)
100	2.39(1.12)	2.90(1.31)	3.25(1.40)	2.17(0.77)	2.15(0.86)	2.16(0.90)	2.17(0.76)	2.22(0.85)	2.24(0.89)
200	3.57(2.13)	4.52(2.54)	5.18(2.72)	2.97(1.21)	2.98(1.43)	3.01(1.53)	3.06(1.27)	3.17(1.48)	3.21(1.57)

	Cross validated values of δ		
N\p	0.01	0.05	0.10
	$u_{it}\sim$ Gaussian		
30	1.27(0.27)	1.46(0.35)	1.61(0.36)
100	1.25(0.24)	1.43(0.31)	1.56(0.32)
200	1.23(0.22)	1.36(0.26)	1.49(0.27)
	$u_{it}\sim$ multivariate t-distr. with 8 dof		
30	1.45(0.38)	1.72(0.39)	1.87(0.35)
100	1.59(0.41)	1.76(0.40)	1.85(0.37)
200	1.68(0.44)	1.78(0.41)	1.85(0.39)

Notes: The MT approach is implemented using $\delta = 1$, $\delta = 2$, and $\hat{\delta}$, computed using cross-validation. Norm losses and estimates of δ, $\hat{\delta}$, are averages over 2,000 replications. Simulation standard deviations are given in the parentheses.

Norm comparisons of MT, BL, CL and LW estimators

- Norm comparisons for Monte Carlo designs A and B consider norm losses averaged over 100 replications, due to the implementation of the cross-validation procedure used in MT, BL and CL thresholding.
  - For the *MT* estimator we use $p = 0.05$ and scaling factor using $\delta = 2$ and $\hat{\delta}$
- We consider the threshold estimators, the two versions of *MT* (MT_2 and $MT_{\hat{\delta}}$) and *CL* (CL_2 and $CL_{\hat{C}}$) estimators, and *BL*.
- *MT* and *CL* estimators (both versions) dominate the *BL* estimator in every case, without any exceptions and for both designs.
- The same is also true if we compare *MT* and *CL* estimators to the *LW* shrinkage estimator.
- The *MT* estimator (both versions) outperforms the *CL* estimator, especially under multivariate t distributed observations.

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

	N=30		N=100		N=200	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
$\mathbf{u}_{it} \sim$ Gaussian						
	Error matrices $(\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}})$					
MT_2	1.85(0.53)	3.38(0.40)	2.83(0.50)	7.29(0.42)	3.45(0.43)	11.17(0.38)
$MT_{\hat{\delta}}$	1.75(0.55)	3.21(0.49)	2.44(0.50)	6.48(0.50)	2.95(0.45)	9.65(0.48)
$BL_{\hat{C}}$	5.30(2.16)	7.61(1.23)	8.74(0.06)	16.90(0.10)	8.94(0.04)	24.26(0.13)
CL_2	1.87(0.55)	3.39(0.44)	2.99(0.49)	7.57(0.44)	3.79(0.47)	11.88(0.42)
$CL_{\hat{C}}$	1.82(0.58)	3.33(0.56)	2.54(0.50)	6.82(0.51)	3.02(0.46)	10.22(0.59)
$LW_{\hat{\Sigma}}$	2.99(0.47)	6.49(0.29)	5.20(0.34)	16.70(0.19)	6.28(0.20)	26.84(0.14)
$\mathbf{u}_{it} \sim$ multivariate $t-$ distributed with 8 degrees of freedom						
	Error matrices $(\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}})$					
MT_2	2.17(0.72)	4.02(0.88)	3.44(0.98)	8.52(1.17)	4.00(0.83)	12.79(1.66)
$MT_{\hat{\delta}}$	2.27(0.88)	4.20(1.11)	3.59(1.39)	8.76(1.65)	4.32(1.53)	13.28(2.83)
$BL_{\hat{C}}$	6.90(0.82)	8.75(0.55)	8.74(0.10)	17.26(0.30)	9.00(0.42)	24.93(1.02)
CL_2	2.55(0.93)	4.53(1.00)	4.63(1.11)	10.35(1.48)	5.92(0.81)	16.43(1.74)
$CL_{\hat{C}}$	2.27(0.76)	4.24(0.94)	3.85(1.51)	9.44(2.33)	5.04(2.04)	15.65(4.71)
$LW_{\hat{\Sigma}}$	3.35(0.51)	7.35(0.50)	5.67(0.46)	18.04(0.45)	6.60(0.43)	28.18(0.53)

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

	N=30		N=100		N=200	
	Norms		Norms		Norms	
	Spectral	Frobenius	Spectral	Frobenius	Spectral	Frobenius
			$u_{it}\sim$ Gaussian			
	Error matrices $(\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}})$					
MT_2	0.49(0.18)	0.89(0.19)	0.87(0.37)	1.63(0.28)	0.73(0.24)	2.15(0.19)
$MT_{\hat{\delta}}$	0.48(0.14)	0.89(0.16)	0.79(0.31)	1.57(0.23)	0.67(0.18)	2.15(0.17)
$BL_{\hat{C}}$	0.91(0.50)	1.35(0.43)	1.40(0.95)	2.25(0.78)	2.53(0.55)	3.49(0.32)
CL_2	0.49(0.17)	0.90(0.18)	1.00(0.48)	1.77(0.44)	0.90(0.37)	2.30(0.30)
$CL_{\hat{C}}$	0.49(0.15)	0.92(0.17)	0.83(0.31)	1.71(0.28)	1.14(0.83)	2.54(0.58)
$LW_{\hat{\Sigma}}$	1.05(0.13)	2.07(0.10)	2.95(0.26)	4.47(0.09)	2.46(0.06)	6.01(0.03)
			$u_{it}\sim$ multivariate $t-$ distributed with 8 degrees of freedom			
	Error matrices $(\mathbf{\Sigma}-\mathring{\mathbf{\Sigma}})$					
MT_2	0.64(0.24)	1.12(0.24)	1.05(0.45)	2.13(0.49)	1.29(2.32)	3.15(2.66)
$MT_{\hat{\delta}}$	0.66(0.25)	1.15(0.26)	1.03(0.42)	2.17(0.53)	1.30(1.90)	3.29(2.22)
$BL_{\hat{C}}$	1.36(0.40)	1.84(0.35)	2.70(0.94)	3.58(0.74)	2.70(0.29)	4.08(0.67)
CL_2	0.71(0.29)	1.21(0.30)	1.69(0.70)	2.73(0.70)	1.62(0.57)	3.31(0.65)
$CL_{\hat{C}}$	0.80(0.39)	1.33(0.39)	2.03(1.08)	3.07(0.90)	2.19(0.78)	3.72(0.62)
$LW_{\hat{\Sigma}}$	1.13(0.15)	2.25(0.11)	3.14(0.21)	4.68(0.11)	2.52(0.08)	6.18(0.13)

Support recovery statistics

- We report the true positive and false positive rates (TPR and FPR) for the support recovery of $\mathbf{\Sigma}$ using the MT and thresholding estimators.
  - In the comparison set we include the *MT* estimator for the three choices of the scaling factor, $\delta = 1$, $\delta = 2$ and $\hat{\delta}$, computed at $p = 0.05$.
- Results show that the *FPR* values of all estimators are very close to zero, so any comparisons of different estimators must be based on the *TPR* values.
- We find that, *TPR* values of $\widetilde{\mathbf{\Sigma}}_{MT_1}$ are closer to unity as compared to the *TPR* values obtained for $\widetilde{\mathbf{\Sigma}}_{MT_2}$, in line with Theorem 3. This result is confirmed when using cross-validated δ.
- Turning to a comparison with other estimators, we find that the *MT* and *CL* estimators perform substantially better than the *BL* estimator.
- Allowing for t-distributed errors causes the support recovery performance of $BL_{\hat{C}}$, CL_2 and $CL_{\hat{C}}$ to deteriorate noticeably while MT_1 and MT_2 remain remarkably stable.
- *TPR* values are high for design B, since there we explicitly control for the number of non-zero elements in $\mathbf{\Sigma}$, and ensure that conditions of Theorem 3 are met.

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

N		MT_1	MT_2	$MT_{\hat{\delta}}$	$BL_{\hat{C}}$	CL_2	$CL_{\hat{C}}$	N		MT_1	MT_2	$MT_{\hat{\delta}}$	$BL_{\hat{C}}$	CL_2	$CL_{\hat{C}}$
		Monte Carlo design A								Monte Carlo design B					
								$u_{it}\sim$ Gaussian							
30	TPR	0.80	0.71	0.79	0.29	0.72	0.78	30	TPR	1.00	0.98	1.00	0.64	0.98	1.00
	FPR	0.00	0.00	0.00	0.04	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
100	TPR	0.69	0.57	0.69	0.00	0.56	0.68	100	TPR	1.00	0.98	1.00	0.80	0.94	0.99
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
200	TPR	0.66	0.53	0.66	0.00	0.50	0.65	200	TPR	1.00	0.96	0.99	0.11	0.88	0.78
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
								$u_{it}\sim$ multivariate $t-$distributed with 8 degrees of freedom							
30	TPR	0.80	0.72	0.79	0.03	0.62	0.74	30	TPR	1.00	0.98	0.99	0.26	0.89	0.82
	FPR	0.01	0.00	0.00	0.00	0.00	0.00		FPR	0.01	0.00	0.00	0.00	0.00	0.00
100	TPR	0.69	0.58	0.67	0.00	0.43	0.57	100	TPR	1.00	0.97	0.98	0.27	0.70	0.57
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00
200	TPR	0.66	0.53	0.64	0.00	0.35	0.47	200	TPR	0.99	0.93	0.95	0.05	0.57	0.30
	FPR	0.00	0.00	0.00	0.00	0.00	0.00		FPR	0.00	0.00	0.00	0.00	0.00	0.00

TPR is the true positive and FPR is the false positive rates defined in Theorem 4. *MT* estimators are computed with $p = 0.05$.

Conclusion

- This paper considers regularisation of large covariance matrices particularly when the cross-sectional dimension N exceeds the time dimension T in a given sample.
- Proposes a multiple testing (MT) estimator as an alternative thresholding procedure, the properties of which are governed by:
  - the maximum degree of dependence of the underlying observations
  - the relative expansion rates of N and T
- The MT estimator has a convergence rate of $m_N c_p(N) T^{-1/2}$ under the spectral norm and $(m_N N/T)^{1/2}$ under the Frobenius norm (m_N is a function of N):
  - These rates are comparable to those established in the literature.
  - These results are valid under both Gaussian and non-Gaussian assumptions.
- Small sample results show that:
  - In terms of spectral and Frobenius norm losses, the MT estimator is reasonably robust to the choice of p in the threshold criterion.
  - For support recovery, better results are obtained when p is scaled by $f(N) = N^{\delta}$, where $\delta = 1$.