Supplementary appendix to: A multiple testing approach to the regularisation of large sample correlation matrices

Natalia Bailey
Queen Mary, University of London

M. Hashem Pesaran
USC and Trinity College, Cambridge

L. Vanessa Smith
University of York

November 5, 2015

Appendix A Technical Lemmas

A.1 Statement of technical lemmas

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

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

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

for some small positive ϵ.

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

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

and

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

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

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

and as N and $T\to\infty$

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

Then

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

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

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

as N and $T\to\infty$.

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

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

for $s=0,1,2,...$, where

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

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

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

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

Then for $x > 0$

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

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

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

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

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

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

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

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

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

Lemma 7 *Consider the data generating process*

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

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

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

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

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

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

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

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

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

A.2 Proofs of lemmas for the MT estimator

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

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

we have

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

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

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

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

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

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

as required. ■

Proof of Lemma 3. First note that

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

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

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

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

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

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

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

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

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

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

■

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

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

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

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

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

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

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

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

as required. ■

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

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

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

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

and using (A.17) we have

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

■

Proof of Lemma 6. We first note that

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

where

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

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

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

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

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

where

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

Hence,

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

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

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

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

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

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

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

Substituting the above results in (A.20) yields

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

or

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

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

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

and

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

and by Lemma 5 we have

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

Similarly,

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

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

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

or[2]

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

as required.

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

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

where

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

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

[2] Note that

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

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

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

Hence

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

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

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

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

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

and hence

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

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

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

■

Proof of Lemma 7. We first note that

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

Then

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

and

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

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

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

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

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

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

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

Hence

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

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

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

■

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

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

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

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

Also

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

But

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

and hence

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

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

Appendix B An overview of key regularisation techniques

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

B.1 Bickel-Levina (BL) thresholding

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

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

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

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

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

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

B.2 Cai and Liu (CL) thresholding

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

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

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

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

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

B.3 Cross-validation for BL and CL

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

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

and

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

Then (B.31) becomes

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

for each c, where

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

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

The following expression is computed for BL or CL,

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

for each c and

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

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

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

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

B.4 Ledoit and Wolf (LW) shrinkage

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

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

with the estimated weights given by

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

where

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

and

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

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

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

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

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

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

Appendix C Shrinkage on MT estimator (S-MT)

Recall the shrinkage on the multiple testing estimator (*S-MT*) expression displayed in Section 3.1,

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

where the $N \times N$ identity matrix $\boldsymbol{I}_N$ is set as benchmark target, the shrinkage parameter is denoted by $\xi \in (\xi_0, 1]$, and ξ_0 is the minimum value of ξ that produces a non-singular $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi_0)$ matrix. Note that shrinkage is deliberately implemented on the correlation matrix $\widetilde{\boldsymbol{R}}_{MT}$ rather than on $\widetilde{\boldsymbol{\Sigma}}_{MT}$. In this way we ensure that no shrinkage is applied to the variances. Further, shrinkage is applied to the non-zero elements of $\widetilde{\boldsymbol{R}}_{MT}$, and as a result the shrinkage estimator, $\widetilde{\boldsymbol{R}}_{S\text{-}MT}$, also consistently recovers the support of $\mathbf{R}$, since it has the same support recovery property as $\widetilde{\boldsymbol{R}}_{MT}$. With regard to the calibration of the shrinkage parameter, ξ, we solve the following optimisation problem

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

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

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

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

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

where

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

Hence, ξ^* is obtained as the solution of

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

where $f(\xi)$ is an analytic differentiable function of ξ for values of ξ close to unity, such that $\boldsymbol{B}(\xi)$ exists.

The resulting $\widetilde{\boldsymbol{R}}_{S\text{-}MT}(\xi^*)$ is guaranteed to be positive definite since

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

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

C.1 Derivation of S-MT shrinkage parameter

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

Abstracting from the subscripts, note that

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

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

or

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

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

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

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

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

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

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

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

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

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

Implementation of the above procedure requires the use of a suitable reference matrix $\boldsymbol{R}_0$. Our experimentations suggested that the shrinkage estimator of Ledoit and Wolf (2004, LW) applied to the correlation matrix is likely to work well in practice, and is to be recommended. Schäfer and Strimmer (2005) consider LW shrinkage on the correlation matrix. In our application we also take account of the small sample bias of the correlation coefficients in what follows. We set as reference matrix $\mathbf{R}_0$ the shrinkage estimator of LW applied to the sample correlation matrix:

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

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

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

is given by

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

with

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

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

Appendix D Additional Monte Carlo simulation results

D.1 Approximately sparse covariance matrix specifications

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

D.2 Additional results

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

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

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

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

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

D.2.3 Norm comparisons of inverse estimators

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

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

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

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

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

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

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

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

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

See the note to Table D2.

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

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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