An Online Supplement for
A multiple testing approach to the regularisation of large sample correlation matrices

Natalia Bailey
Department of Econometrics and Business Statistics, Monash University

M. Hashem Pesaran
Department of Economics & USC Dornsife INET,
University of Southern California, USA, and Trinity College, Cambridge

L. Vanessa Smith
Department of Economics, University of York

October 22, 2018

Supplementary Appendix A

Technical Lemmas

A.1 Statement of technical lemmas

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

Lemma 1 *Consider the 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 $x > 0$

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

Lemma 2 *Consider the critical value function*[1]

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

where $\Phi^{-1}(.)$ *is the inverse function of the cumulative standard normal distribution,* $0 < p < 1$*,* $f(N) = c_\delta N^\delta$*, where* c_δ *and* δ *are finite positive constants, and suppose there exists finite* N_0 *such that for all* $N > N_0$

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

Then for $0 < \varkappa \leq 1$*, we have*

(a) $c_p(N) = O\left([\ln(N)]^{1/2}\right)$*;*

(b) $\exp\left[-\varkappa c_p^2(N)/2\right] = \ominus\left(N^{-\delta\varkappa}\right)$*,* $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$*;*

(c) if $\delta > 1/\varkappa$*, then* $N\exp\left[-\varkappa c_p^2(N)/2\right] \to 0$*, as* $N \to \infty$*.*

Lemma 3 *Consider the sample correlation coefficient,* $\hat{\rho}_{ij,T}$*, defined by (4), and suppose that Assumptions 1 and 2 hold,* $T = c_d N^d$*, with* $c_d > 0$*. Then, there exists* N_0 *such that for all* $N \geq N_0$*,*[2]

$$\Pr\left[\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > T^{-1/2} c_p(N)\right] \leq K e^{-\frac{1}{2}\frac{c_p^2(N)}{K_v(\boldsymbol{\theta}_{ij})}} + O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{K_v(\boldsymbol{\theta}_{ij})}\right]^{\frac{3(s-2)-1}{2}} e^{-\frac{1}{2}\frac{c_p^2(N)}{K_v(\boldsymbol{\theta}_{ij})}}\right) + O\left(T^{-(s-1)/2}\right), \tag{A.3}$$

[1] We would like to thank George Kapetanios for his help with the proof of (b) and (c) of this Lemma.

[2] To simplify the notation we have dropped the lower order terms $e^{-\frac{1}{2}\frac{c_p(N)}{\sqrt{K_v(\boldsymbol{\theta}_{ij})}}\frac{K}{\sqrt{T}}}$ and $e^{-\frac{K}{T}}$, as they do not affect the results, and can be absorbed in the remainder order term.

and

$$\Pr\left[\left|\hat{\rho}_{ij,T}\right| > T^{-1/2}c_p(N)|\rho_{ij}=0\right] \leq Ke^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{ij}}} + O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{\varphi_{ij}}\right]^{\frac{3(s-2)-1}{2}} e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{ij}}}\right) + O\left(T^{-(s-1)/2}\right), \tag{A.4}$$

where $K_v(\boldsymbol{\theta}_{ij})$ *is given by (12),* $\varphi_{ij} = E\left(y_{it}^2 y_{jt}^2 \left|\rho_{ij}=0\right.\right)$*, and*

$$c_p(N) = \Phi^{-1}\left(1-\frac{p}{2f(N)}\right) > 0, \tag{A.5}$$

with $0<p<1$*, and* $f(N) = c_\delta N^\delta$ *where* c_δ *and* δ *are finite positive constants. Further, if* $\left|\rho_{ij}\right| > c_p(N)/\sqrt{T}$ *then we have*

$$\Pr\left[\left|\hat{\rho}_{ij,T}\right| \leq T^{-1/2}c_p(N)|\rho_{ij}\neq 0\right] \leq Ke^{-\frac{T\left(\left|\rho_{ij}\right|-T^{-1/2}c_p(N)\right)^2}{2K_v(\boldsymbol{\theta}_{ij})}}\left[1+O\left(T^{\frac{2(s-2)-1}{2}}\right)\right] + O\left(T^{-(s-1)/2}\right). \tag{A.6}$$

Lemma 4 *Consider the standardised sample correlation coefficient*

$$z_{ij,T} = \left[Var\left(\hat{\rho}_{ij,T}\right)\right]^{-1/2}\left[\hat{\rho}_{ij,T} - E\left(\hat{\rho}_{ij,T}\right)\right],$$

where $\hat{\rho}_{ij,T}$ *is defined by (4) and* $E\left(\hat{\rho}_{ij,T}\right)$ *and* $Var\left(\hat{\rho}_{ij,T}\right)>0$ *are given by (9) and (10), respectively. Suppose that Assumptions 1 and 2 hold,* $c_p(N) = \Phi^{-1}\left(1-\frac{p}{2f(N)}\right)$*, and condition (A.2) holds. Also let* $T = c_d N^d$*, with* $c_d > 0$*. Then, there exists* N_0 *such that for* $N \geq N_0$*, and for* $r \geq 0$*,*

$$\sup_{ij} E\left[\left|z_{ij,T}\right|^r I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}\neq 0\right.\right)\right] < K, \tag{A.7}$$

and

$$\sup_{ij} E\left[\left|z_{ij,T}\right|^r I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\left|\rho_{ij}=0\right.\right)\right] \leq Ke^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}} + O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{\varphi_{\max}}\right]^{\frac{3(s-2)-1}{2}} e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{\max}}}\right) + O\left(T^{-(s-1)/2}\right), \tag{A.8}$$

where $\varphi_{\max} = \sup_{ij}\varphi_{ij}$*,* φ_{ij} *is defined by (13), and* $s \geq 3$ *is defined by Assumption 2.*

Lemma 5 *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 $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})' \sim IIDN(\mathbf{0}, \mathbf{I}_N)$*, 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*

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

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

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

Lemma 6 *Fat-tailed shocks do not necessarily generate* $E(y_{it}^2 y_{jt}^2) > 1$*.*

A.2 Proofs of lemmas for the MT estimator

Proof of Lemma 1. Using results in Chiani et al. (2003) - eq. (5), we have

$$\operatorname{erf} c(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du \leq \exp(-x^2), \tag{A.9}$$

where $\operatorname{erf} c(x)$ is the complement of the $\operatorname{erf}(x)$ error function defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du. \tag{A.10}$$

But

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

and using (A.9) we have

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

■

Proof of Lemma 2. First note that

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

where $\Phi(x)$ is cumulative distribution function of a standard normal variate, and $\operatorname{erf}(x)$ function is defined by (A.10). Consider now the inverse complementary error function $\operatorname{erfc}^{-1}(x)$ given by

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

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

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

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

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

Therefore, for $f(N) = c_\delta N^\delta$ we have

$$c_p^2(N) \leq 2\left[\delta \ln(N) - \ln(p)\right] = O\left[\ln(N)\right],$$

which establishes part (a).

Further, by Proposition 24 of Dominici (2003) we have that

$$\lim_{N\to\infty} c_p(N)/LW\left\{\frac{1}{2\pi\left[\left(1 - \frac{p}{2f(N)}\right) - 1\right]^2}\right\}^{1/2} = 1,$$

where LW denotes the Lambert W function which satisfies $\lim_{N\to\infty} LW(N)/\left\{\ln(N) - \ln\left[\ln(N)\right]\right\} = 1$ as $N \to \infty$. We note that $\lim_{N\to\infty} \ln(N)/\left\{\ln(N) - \ln\left[\ln(N)\right]\right\} = 1$ as $N \to \infty$. So

$$\lim_{N\to\infty} \frac{LW\left\{\frac{1}{2\pi\left[\left(1-\frac{p}{2f(N)}\right)-1\right]^2}\right\}^{1/2}}{\left\{2\ln\left(\frac{\sqrt{2}f(N)}{\sqrt{\pi}p}\right)\right\}^{1/2}} = 1.$$

Hence, for any $0 < \varkappa \leq 1$,

$$\lim_{N\to\infty} \frac{\exp\left[-\varkappa c_p^2(N)/2\right]}{\exp\left[-\frac{\varkappa\left\{\left[2\ln\left(\frac{\sqrt{2}f(N)}{\sqrt{\pi}p}\right)\right]^{1/2}\right\}^2}{2}\right]} = \lim_{N\to\infty} \frac{\exp\left[-\varkappa c_p^2(N)/2\right]}{\left[f\left(N\right)\right]^{-\varkappa} \pi^{\varkappa/2} p^{\varkappa} 2^{-\varkappa/2}} = 1, \text{ as } N \to \infty,$$

and substituting $c_\delta N^\delta$ for $f(N)$ yields,

$$\lim_{N\to\infty} \frac{\exp\left[-\varkappa c_p^2\left(N\right)/2\right]}{c_\delta^{-\varkappa} N^{-\delta\varkappa}} = \frac{2^{\varkappa/2} c_\delta^{\varkappa}}{\pi^{\varkappa/2} p^{\varkappa}}. \tag{A.11}$$

It follows from (A.11) that $\exp\left[-\varkappa c_p^2\left(N\right)/2\right] = \ominus\left(N^{-\delta\varkappa}\right)$, as required. From this result it also follows that $\left[-\varkappa c_p^2\left(N\right)/2\right] = \ominus\left(-\delta\varkappa \ln N\right)$, which in turn yields $\lim_{N\to\infty} c_p^2(N)/\ln(N) = 2\delta$. This completes the proof of part (b). Finally, it readily follows from (b) that $N\exp\left[-\varkappa c_p^2\left(N\right)/2\right] = \ominus\left(N^{1-\delta\varkappa}\right)$, and therefore $N\exp\left[-\varkappa c_p^2\left(N\right)/2\right] \to 0$ when $\delta > 1/\varkappa$, as required for the proof of part (c). ■

Proof of Lemma 3. We first note that

$$\begin{aligned} & \Pr\left[\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > T^{-1/2}c_p(N)\right] \\ = \; & \Pr\left[\hat{\rho}_{ij,T} - \rho_{ij} > T^{-1/2}c_p(N)\right] + \Pr\left[-\left(\hat{\rho}_{ij,T} - \rho_{ij}\right) > T^{-1/2}c_p(N)\right] \\ = \; & \Pr\left[\hat{\rho}_{ij,T} - \rho_{ij} > T^{-1/2}c_p(N)\right] + \Pr\left[\left(\hat{\rho}_{ij,T} - \rho_{ij}\right) < -T^{-1/2}c_p(N)\right]. \end{aligned} \tag{A.12}$$

and

$$\hat{\rho}_{ij,T} - \rho_{ij} = \omega_{ij,T} z_{ij,T} + \left(\rho_{ij,T} - \rho_{ij}\right),$$

where $z_{ij,T}$ is the standardised sample correlation coefficient defined by (15), $\rho_{ij,T} = E\left(\hat{\rho}_{ij,T}\right)$ and $\omega_{ij,T}^2 = Var\left(\hat{\rho}_{ij,T}\right) > 0$ are given by (9) and (10), respectively. Hence

$$\begin{aligned} \Pr\left[\hat{\rho}_{ij,T} - \rho_{ij} > T^{-1/2}c_p(N)\right] &= \Pr\left[\omega_{ij,T} z_{ij,T} + \left(\rho_{ij,T} - \rho_{ij}\right) > T^{-1/2}c_p(N)\right] \\ &= \Pr\left(z_{ij,T} > a_{ij,T}\right), \end{aligned}$$

where

$$a_{ij,T} = \frac{T^{-1/2}c_p(N) - \left(\rho_{ij,T} - \rho_{ij}\right)}{\omega_{ij,T}}.$$

Similarly,

$$\Pr\left[\hat{\rho}_{ij,T} - \rho_{ij} < -T^{-1/2}c_p(N)\right] = \Pr\left(z_{ij,T} < -b_{ij,T}\right),$$

where

$$b_{ij,T} = \frac{T^{-1/2}c_p(N) + \left(\rho_{ij,T} - \rho_{ij}\right)}{\omega_{ij,T}}$$

But using (9) and (10) we have (note that by assumption $\sup_{ij}\left|K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})\right|$and $\sup_{ij} K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij}) < K$)

$$a_{ij,T} = \frac{T^{-1/2}c_p(N) - \frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right)}{\sqrt{T^{-1}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) + O\left(T^{-2}\right)}} = \frac{c_p(N)}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})} + O\left(T^{-1/2}\right), \tag{A.13}$$

and

$$b_{ij,T} = \frac{T^{-1/2}c_p(N) + \frac{K_{\mathrm{m}}(\boldsymbol{\theta}_{ij})}{T} + O\left(T^{-2}\right)}{\sqrt{T^{-1}K_{\mathrm{v}}(\boldsymbol{\theta}_{ij}) + O\left(T^{-2}\right)}} = \frac{c_p(N)}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})} + O\left(T^{-1/2}\right). \tag{A.14}$$

Using the above results in (A.12) we now have

$$\begin{aligned} \Pr\left[\left|\hat{\rho}_{ij,T} - \rho_{ij}\right| > \theta\left(N,T\right)\right] &= \Pr\left(z_{ij,T} > \frac{c_p(N)}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})} + O\left(T^{-1/2}\right)\right) \\ &\quad + \Pr\left(z_{ij,T} \leq \frac{-c_p(N)}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})} + O\left(T^{-1/2}\right)\right), \end{aligned}$$

where $\theta\left(N,T\right)=T^{-1/2}c_p(N)$. Result (A.3) now follows using (17) and (18) with a_T replaced by (A.13), and ignoring the higher order terms $e^{-\frac{K}{\sqrt{T}}\frac{c_p(N)}{\sqrt{K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}}}$ and $e^{-\frac{K}{T}}$ that arise from squaring $a_T=K_{\mathrm{v}}^{-1/2}(\boldsymbol{\theta}_{ij})c_p(N)+O\left(T^{-1/2}\right)$. Result (A.4) can be obtained as a special case by setting $\rho_{ij}=0$. Finally, to establish (A.6), using similar line of reasoning as above, we first note that

$$\begin{aligned}\Pr\left[\left|\hat{\rho}_{ij,T}\right|\leq\theta\left(N,T\right)|\rho_{ij}\neq 0\right] &= \Pr\left(z_{ij,T}\leq -\frac{\sqrt{T}\rho_{ij}-c_p(N)}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})}+O\left(T^{-1/2}\right)\right) \qquad \text{(A.15)}\\ &\quad -\Pr\left(z_{ij,T}\leq -\frac{c_p(N)+\sqrt{T}\rho_{ij}}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})}+O\left(T^{-1/2}\right)\right)\\ &\leq \Pr\left(z_{ij,T}\leq -\frac{\sqrt{T}\left[\rho_{ij}-T^{-1/2}c_p(N)\right]}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})}+O\left(T^{-1/2}\right)\right).\end{aligned}$$

Suppose that $\rho_{ij}>T^{-1/2}c_p(N)>0$, then using (17) we have (again ignoring higher order terms in T^{-1})

$$\Pr\left[\left|\hat{\rho}_{ij,T}\right|\leq T^{-1/2}c_p(N)|\rho_{ij}\neq 0\right]\leq K\mathrm{e}^{-\frac{T\left(\rho_{ij}-T^{-1/2}c_p(N)\right)^2}{2K_{\mathrm{v}}(\boldsymbol{\theta}_{ij})}}\left[1+O\left(T^{\frac{2(s-2)-1}{2}}\right)\right]+O\left[T^{-(s-1)/2}\right]. \qquad \text{(A.16)}$$

A similar result follows when $\rho_{ij}<0$. In this case we consider writing (A.15) equivalently as

$$\begin{aligned}\Pr\left[\left|\hat{\rho}_{ij,T}\right|\leq T^{-1/2}c_p(N)|\rho_{ij}\neq 0\right] &= 1-\Pr\left(z_{ij,T}>-\frac{\sqrt{T}\rho_{ij}-c_p(N)}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})}+O\left(T^{-1/2}\right)\right)\\ &\quad -1+\Pr\left(z_{ij,T}>-\frac{c_p(N)+\sqrt{T}\rho_{ij}}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})}+O\left(T^{-1/2}\right)\right)\\ &\leq \Pr\left(z_{ij,T}>\frac{\sqrt{T}\left[-\rho_{ij}+T^{-1/2}c_p(N)\right]}{K_{\mathrm{v}}^{1/2}(\boldsymbol{\theta}_{ij})}+O\left(T^{-1/2}\right)\right),\end{aligned}$$

where by assumption $-\rho_{ij}+T^{-1/2}c_p(N)>0$. Now applying (18) to the right hand side of the above yields the outcome in (A.16) with ρ_{ij} replaced by $-\rho_{ij}$ Thus the desired result (A.6) is established for positive and negative values of ρ_{ij} such that $\left|\rho_{ij}\right|-T^{-1/2}c_p(N)>0$. ■

Proof of Lemma 4. We first note that since $\inf_{ij}Var\left(\hat{\rho}_{ij,T}\right)>0$, and $\hat{\rho}_{ij,T}$ is a correlation coefficient, $\left|\hat{\rho}_{ij,T}\right|\leq 1$, there exists T_0 such that for all $T>T_0$

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

Hence, $E\left|z_{ij,T}\right|^r < K$ for any finite r. Also, by the Cauchy–Schwarz inequality

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

which establishes (A.7), as required. Similarly,

$$E\left[\left|z_{ij,T}\right|^r I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}= 0\right)\right] \leq K\Pr\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}= 0\right),$$

and using result (A.4) of Lemma 3, we have

$$\begin{aligned}
E\left[\left|z_{ij,T}\right|^r I\left(\left|\sqrt{T}\hat{\rho}_{ij,T}\right| > c_p(N)\,|\rho_{ij}= 0\right)\right] &\leq Ke^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{ij}}} + O\left(T^{-\frac{(s-2)}{2}}\left[\frac{c_p^2(N)}{\varphi_{ij}}\right]^{\frac{3(s-2)-1}{2}} e^{-\frac{1}{2}\frac{c_p^2(N)}{\varphi_{ij}}}\right) \\
&\quad + O\left(T^{-(s-1)/2}\right).
\end{aligned}$$

which establishes (A.8). ■

Proof of Lemma 5. First we note that

$$\begin{aligned}
E\left(\frac{1}{\chi^2_{\mathrm{v},t}}\right) &= \frac{1}{\mathrm{v}-2},\ Var\left(\frac{1}{\chi^2_{\mathrm{v},t}}\right) = \frac{2}{(\mathrm{v}-2)^2(\mathrm{v}-4)} \\
E\left(\frac{1}{\chi^2_{\mathrm{v},t}}\right)^2 &= \frac{2}{(\mathrm{v}-2)^2(\mathrm{v}-4)} + \left(\frac{1}{\mathrm{v}-2}\right)^2 = \frac{\mathrm{v}-2}{(\mathrm{v}-2)^2(\mathrm{v}-4)}.
\end{aligned} \tag{A.17}$$

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

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

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.17) we have

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

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

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

Hence

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

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 $E(y_{it}^2 y_{jt}^2) = (\mathrm{v}-2)/(\mathrm{v}-4)$. Also

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

■

Proof of Lemma 6. Consider the data generating process $\mathbf{y}_t = \mathbf{P}\mathbf{u}_t$ where the elements of $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$, 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 $E(u_{it}) = 0$, $E(u_{it}^2) = 1$, as well as $E(u_{it}^2 u_{jt}^2) = E(u_{it}^2)E(u_{jt}^2) = 1$, and $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{I}_N$. Let $\mathbf{p}_i'$ be the i^{th} row of $\mathbf{P}$ and note that

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

Also

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

But

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

and hence

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

Therefore, fat-tailed shocks do not necessarily generate $\mu_{ij}(2,2) = E\left(y_{it}^2 y_{jt}^2\right) > 1$. ■

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

$$\widetilde{\mathbf{\Sigma}}_{BL,C} = \left(\hat{\sigma}_{ij} I \left[|\hat{\sigma}_{ij}| \geq C \sqrt{\frac{\ln(N)}{T}} \right] \right), \; i = 1, 2, \ldots, N-1, \; j = i+1, i+2, \ldots, N \quad \text{(B.18)}$$

where $I(.)$ is an indicator function and C is a positive constant which is unknown. The choice of thresholding function - $I(.)$ - implies that (B.18) implements 'hard' thresholding. The consistency rate of the BL estimator is $m_N \sqrt{\frac{\ln(N)}{T}}$ 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 rooted in the specification chosen for the implementation of CV.[3] 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

$$\left\| \widetilde{\mathbf{\Sigma}}_{BL,C} - \widetilde{\mathbf{\Sigma}}_{BL,0} \right\| \leq \epsilon \text{ and } \lambda_{\min}(\mathbf{\Sigma}) > \epsilon, \quad \text{(B.19)}$$

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

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,

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

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

$$\widetilde{\mathbf{\Sigma}}_{CL,C} = \left(\hat{\sigma}_{ij} s_{\tau_{ij}} \left[|\hat{\sigma}_{ij}| \geq \tau_{ij}\right]\right),\ i = 1, 2, \ldots, N-1,\ j = i+1, i+2, \ldots, N \tag{B.20}$$

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 $\omega_T = C\sqrt{\ln(N)/T}$, 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 $C_0 m_N \sqrt{\ln(N)/T}$ 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.19), Fan et al. (2013) - FLM - 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

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 $C_{\min} = \left|\min_{ij} \hat{\sigma}_{ij}\right| \sqrt{\frac{T}{\ln N}}$ and $C_{\max} = \left|\max_{ij} \hat{\sigma}_{ij}\right| \sqrt{\frac{T}{\ln N}}$ 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 for $c = 1$. In each point of the respective ranges, c, we use x_{it}, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$ and select the $N \times 1$ column vectors $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$, $t = 1, 2, \ldots, T$ which we randomly reshuffle over the t-dimension. This gives rise to a new set of $N \times 1$ column vectors $\boldsymbol{x}_t^{(s)} = \left(x_{1t}^{(s)}, x_{2t}^{(s)}, \ldots, x_{Nt}^{(s)}\right)'$ 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 et al. (2013)). In each shuffle $s = 1, 2, \ldots, S$, we divide $\boldsymbol{x}^{(s)} = \left(\boldsymbol{x}_1^{(s)}, \boldsymbol{x}_2^{(s)}, \ldots, \boldsymbol{x}_T^{(s)}\right)$ 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 $T_1 = T\left(1 - \frac{1}{\ln(T)}\right)$ and $T_2 = \frac{T}{\ln(T)}$. 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 $\hat{\sigma}_{2,ij}^{(s)} = T_2^{-1} \sum_{t=T_1+1}^{T} x_{it}^{(s)} x_{jt}^{(s)}$, $i, j = 1, 2, \ldots, N$, denote the sample covariance matrices generated using T_1 and T_2 respectively, for each shuffle s. We threshold $\hat{\mathbf{\Sigma}}_1^{(s)}$ as in (B.18), (B.20) using $I(.)$ as the thresholding function, where for CL 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

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

Then (B.20) 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 and CL,

$$\hat{G}(c) = \frac{1}{S}\sum_{s=1}^{S}\left\|\widetilde{\mathbf{\Sigma}}_1^{(s)}(c) - \hat{\mathbf{\Sigma}}_2^{(s)}\right\|_F^2, \tag{B.21}$$

for each c and

$$\hat{C} = \arg\min_{C_{\min}\leq c\leq C_{\max}} \hat{G}(c). \tag{B.22}$$

If several values of c attain the minimum of (B.22), then $\hat{C}$ is chosen to be the smallest one. The final estimator of the covariance matrix is then given by $\widetilde{\mathbf{\Sigma}}_{\hat{C}}$. The thresholding approach does not necessarily ensure that the resultant estimate, $\widetilde{\mathbf{\Sigma}}_{\hat{C}}$, is positive definite. To ensure that the threshold estimator is positive definite FLM propose setting a lower bound on the cross-validation grid for the search of C such that $\lambda_{\min}\left(\widetilde{\mathbf{\Sigma}}_{\hat{C}}\right) > 0$ - see Fryzlewicz (2013). Therefore, for BL and CL we modify (B.22) so that

$$\hat{C}^{*} = \arg\min_{C_{pd}+\epsilon\leq c\leq C_{\max}} \hat{G}(c), \tag{B.23}$$

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

B.4 Ledoit and Wolf (LW) shrinkage

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

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

with the estimated weights given by

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

where

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

and

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

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

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

with $\dot{\boldsymbol{x}}_t = (\dot{x}_{1t}, \dot{x}_{2t}, \ldots, \dot{x}_{Nt})'$ 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.

Supplementary Appendix C

Shrinkage on MT estimator (S-MT)

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

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

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

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

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

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

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

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

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

where

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

Hence, ξ^* is obtained as the solution of

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

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

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

or

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

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

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

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 $\hat{\mathbf{R}}_0$:

$$E\left\|\hat{\mathbf{R}}_0-\mathbf{R}\right\|_F^2=\sum_{i\neq j}\sum E\left(\hat{\rho}_{ij}-\rho_{ij}\right)^2+\theta^2\sum_{i\neq j}\sum E\left(\hat{\rho}_{ij}^2\right)-2\theta\sum_{i\neq j}\sum E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right], \quad \text{(C.25)}$$

is given by

$$\theta^*=\frac{\sum_{i\neq j}\sum E\left[\hat{\rho}_{ij}\left(\hat{\rho}_{ij}-\rho_{ij}\right)\right]}{\sum_{i\neq j}\sum E\left(\hat{\rho}_{ij}^2\right)}=1-\frac{\sum_{i\neq j}\sum E\left(\hat{\rho}_{ij}\rho_{ij}\right)}{\sum_{i\neq j}\sum E\left(\hat{\rho}_{ij}^2\right)}, \quad \text{(C.26)}$$

with

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

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

References

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

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

Chiani, M., D. Dardari, and M. K. Simon (2003). New exponential bounds and approximations for the computation of error probability in fading channels. IEEE Transactions on Wireless Communications 2(4), 840-845.

Dominici, D. E. (2003). The inverse of the cumulative standard normal probability function. Integral Transforms and Special Functions 14, 281-292.

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

Fang, Y., B. Wang, and Y. Feng (2016). Tuning parameter selection in regularised estimations of large covariance matrices. Journal of Statistical Computation and Simulation 86, 494-509.

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

Ledoit, O. and M.Wolf (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. Statistica Neerlandica 32, 201-210.

Schäfer, J. and K. Strimmer (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional Genomics. Statistical Applications in Genetics and Molecular Biology 4, Article 32.