Appendix to: Dynamic Partial Correlation Models

Enzo D'Innocenzo[a], André Lucas[b]

[a] *Università di Bologna*

[b] *Vrije Universiteit Amsterdam and Tinbergen Institute*

A Proofs

Proof of Proposition 1

Define

$$\begin{aligned}
\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t} &= \hat{\boldsymbol{D}}^{-1/2}_{i,j|L_{ij}\,;\,t}\left(\boldsymbol{y}_{i,j\,;\,t} - \hat{\boldsymbol{\mu}}_{i,j|L_{ij}\,;\,t}\right), \\
\hat{\boldsymbol{D}}_{i,j|L_{ij}\,;\,t} &= \frac{\nu - 2 + \boldsymbol{y}^{\top}_{L_{ij},t}\hat{\boldsymbol{R}}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{y}_{L_{ij},t}}{\nu_{i,j|L_{ij}}}\begin{pmatrix}\hat{\boldsymbol{V}}_{i,i|L_{ij}\,;\,t} & 0 \\ 0 & \hat{\boldsymbol{V}}_{j,j|L_{ij}\,;\,t}\end{pmatrix}, \\
w_{i,j|L_{ij}\,;\,t} &= \frac{\nu_{i,j|L_{ij}} + 2}{\nu_{i,j|L_{ij}} + \hat{\boldsymbol{y}}^{\star\top}_{i,j|L_{ij}\,;\,t}\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;\,t}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}}.
\end{aligned}$$

Using standard vector derivative calculus, we have

$$\begin{aligned}
&\frac{\partial \log p(\boldsymbol{y}_{i,t}, \boldsymbol{y}_{j,t} \mid \mathcal{F}_{t-1}, \boldsymbol{y}_{L_{ij},t})}{\partial f_{i,j|L_{ij}\,;\,t}} = \\
&\quad \frac{\partial}{\partial f_{i,j|L_{ij}\,;\,t}}\left(-\tfrac{1}{2}\log|\boldsymbol{R}_{i,j|L_{ij}\,;\,t}| - \tfrac{1}{2}(\nu_{i,j|L_{ij}} + 2)\log\left(1 + \frac{\hat{\boldsymbol{y}}^{\star\top}_{i,j\,;\,t}\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;\,t}\hat{\boldsymbol{y}}^{\star}_{i,j\,;\,t}}{\nu_{i,j|L_{ij}}}\right)\right) \\
&\quad = \tfrac{1}{2}\frac{\partial\,\mathrm{vec}(\boldsymbol{R}_{i,j|L_{ij}\,;\,t})^{\top}}{\partial f_{i,j|L_{ij}\,;\,t}} \cdot \mathrm{vec}\left(w_{i,j|L_{ij}\,;\,t} \cdot \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;\,t}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}\hat{\boldsymbol{y}}^{\star\top}_{i,j|L_{ij}\,;\,t}\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;\,t} - \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;\,t}\right) \\
&\quad = \tfrac{1}{2}\frac{\partial\,\mathrm{vec}(\boldsymbol{R}_{i,j|L_{ij}\,;\,t})^{\top}}{\partial f_{i,j|L_{ij}\,;\,t}} \cdot \left(\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;\,t} \otimes \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;\,t}\right)\mathrm{vec}\left(w_{i,j|L_{ij}\,;\,t}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}\hat{\boldsymbol{y}}^{\star\top}_{i,j|L_{ij}\,;\,t} - \boldsymbol{R}_{i,j|L_{ij}\,;\,t}\right).
\end{aligned}$$

■

Proof of Proposition 2

Throughout the proof we take the symmetric matrix root.[1] Note that we have

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

$$\boldsymbol{R}^{-1/2} = \tfrac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} (1+\rho)^{-1/2} & 0 \\ 0 & (1-\rho)^{-1/2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

$$(0 \quad 1 \quad 1 \quad 0) \cdot \left(\boldsymbol{R}^{-1/2} \otimes \boldsymbol{R}^{-1/2} \right) = \frac{(-\rho \quad 1 \quad 1 \quad -\rho)}{1-\rho^2}.$$

Define

$$w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} = \frac{\nu_{i,j|L_{ij}} + 2}{\nu_{i,j|L_{ij}} + \boldsymbol{\eta}^\top_{i,j|L_{ij};t} \boldsymbol{\eta}_{i,j|L_{ij};t}},$$

and

$$\begin{aligned} s^{\boldsymbol{\eta}}_{i,j|L_{ij};t} &= \tfrac{1}{2} \dot{g}(f_{i,j|L_{ij};t}) \cdot (0 \quad 1 \quad 1 \quad 0) \cdot \left(\boldsymbol{R}^{-1/2}_{i,j|L_{ij};t} \otimes \boldsymbol{R}^{-1/2}_{i,j|L_{ij};t} \right) \cdot \operatorname{vec}\left(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \cdot \boldsymbol{\eta}_{i,j|L_{ij};t} \boldsymbol{\eta}^\top_{i,j|L_{ij};t} - \mathbf{I}_2 \right) \\ &= \tfrac{1}{2} \frac{\epsilon (1 - \rho^2_{i,j|L_{ij};t}/\epsilon^2)}{1 - \rho^2_{i,j|L_{ij};t}} \cdot (-\rho_{i,j|L_{ij};t} \quad 1 \quad 1 \quad -\rho_{i,j|L_{ij};t}) \cdot \operatorname{vec}\left(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \cdot \boldsymbol{\eta}_{i,j|L_{ij};t} \boldsymbol{\eta}^\top_{i,j|L_{ij};t} - \mathbf{I}_2 \right) \\ &= h(f_{i,j|L_{ij};t}) \cdot \left(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}_{1,t} \boldsymbol{\eta}_{2,t} - \tfrac{1}{2} g(f_{i,j|L_{ij};t})\ w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}^\top_t \boldsymbol{\eta}_t + g(f_{i,j|L_{ij};t}) \right) \end{aligned}$$

where for the second equality we used Assumption 1 so that $\dot{g}(f) = \epsilon \cdot (1 - \rho^2/\epsilon^2)$ and then defined

$$h(f_{i,j|L_{ij};t}) = \frac{\epsilon \left(1 - g(f_{i,j|L_{ij};t})^2/\epsilon^2 \right)}{1 - g(f_{i,j|L_{ij};t})^2}.$$

By Assumption 2, we look at the model as a Data Generating Process (DGP), that is, $\boldsymbol{y}_{i,j|L_{ij};t} = \boldsymbol{R}^{1/2}_{i,j|L_{ij};t} \boldsymbol{\eta}_t$, so that we can rewrite the score-driven transition equation of Proposition 1 under the DGP as

$$f_{i,j|L_{ij};t+1} = \omega_{i,j|L_{ij}} + \beta_{i,j|L_{ij}} f_{i,j|L_{ij};t} + \alpha_{i,j|L_{ij}} s^{\boldsymbol{\eta}}_{i,j|L_{ij};t}.$$

[1] Other roots can be taken as well, but typically indicate smaller stationarity regions; compare Blasques et al. (2018). Note that the Bougerol (1993) condition only provides a sufficient condition, such that we are free to take the matrix root that results in the widest region.

Note that for given $f_{i,j|L_{ij};t}$ all moments of $w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top$ and thus of the rewritten score $s^{\boldsymbol{\eta}}_{i,j|L_{ij};t}$ exist due to its uniform boundedness in $\boldsymbol{\eta}_t$. As a result, for a fixed initialization $f_{i,j|L_{ij};1}$ we directly obtain $\mathbb{E}[\log^+ |f_{i,j|L_{ij};t+1}|\ ] < \infty$. To use Theorem 3.1 of Bougerol (1993), we therefore only need to prove that the recursion is contracting on average. To do this, we note

$$\frac{\partial f_{i,j|L_{ij};t+1}}{\partial f_{i,j|L_{ij};t}} = \beta_{i,j|L_{ij}} + \alpha_{i,j|L_{ij}} \cdot \Big(\dot{h}(f_{i,j|L_{ij};t}) \cdot \Big(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_{1,t}\boldsymbol{\eta}_{2,t} - \tfrac{1}{2} g(f_{i,j|L_{ij};t})\ w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t + g(f_{i,j|L_{ij};t}) \Big) - h(f_{i,j|L_{ij};t}) \cdot \epsilon \cdot \Big(1 - g(f_{i,j|L_{ij};t})^2/\epsilon^2\Big) \cdot \Big(\tfrac{1}{2} w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t - 1\Big) \Big),$$

where

$$\dot{h}(f_{i,j|L_{ij};t}) = \frac{2\left(\epsilon^2 - 1\right) g(f_{i,j|L_{ij};t}) \left(1 - g(f_{i,j|L_{ij};t})^2/\epsilon^2\right)}{\left(1 - g(f_{i,j|L_{ij};t})^2\right)^2}.$$

such that we require

$$\mathbb{E}\left[\log \sup_f \left| \left.\frac{\partial f_{i,j|L_{ij};t+1}}{\partial f_{i,j|L_{ij};t}}\right|_{f_{i,j|L_{ij};t}=f} \right|\right] = \mathbb{E}\Bigg[\log \sup_f \Big|\beta_{i,j|L_{ij}} + \alpha_{i,j|L_{ij}} \cdot \Big(\dot{h}(f) \cdot \Big(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_{1,t}\boldsymbol{\eta}_{2,t} - \tfrac{1}{2} g(f) w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t + g(f) \Big) - h(f) \cdot \epsilon \cdot \left(1 - g(f)^2/\epsilon^2\right) \cdot \Big(\tfrac{1}{2} w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t - 1\Big) \Big)\Big|\Bigg] < 0.$$

Define $e_1 = w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_{1,t}\boldsymbol{\eta}_{2,t}$ and $e_2 = \frac{1}{2} w^{\boldsymbol{\eta}}_{i,j|L_{ij};t}\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t - 1$, then we note that we can rewrite the previous equation as

$$\mathbb{E}\left[\log \sup_{|\rho|<\epsilon} \left|\beta_{i,j|L_{ij}} + \alpha_{i,j|L_{ij}} \cdot \left(\dot{\tilde{h}}(\rho) \cdot (e_1 - \rho\, e_2) - \tilde{h}(\rho) \cdot \epsilon \cdot \left(1 - \rho^2/\epsilon^2\right) \cdot e_2 \right)\right|\right] < 0. \tag{A.1}$$

for $\tilde{h}(\rho) = \epsilon \cdot (1 - \rho^2/\epsilon^2)/(1-\rho^2)$, and $\dot{\tilde{h}}(\rho) = 2(\epsilon^2-1)\rho(1-\rho^2/\epsilon^2)/(1-\rho^2)^2$. This is clearly satisfied through Assumption 3. Theorem 3.1 of Bougerol (1993) now implies that each initialized $f_{i,j|L_{ij};t}$ converges (e.a.s.) to a unique stationary and ergodic limit sequence.

As the mappings $\rho_{i,j|L_{ij};t} = g(f_{i,j|L_{ij};t})$ are all continuously differentiable with $\sup_f \dot{g}(f) = \epsilon$, we obtain $|\hat{\rho}_{i,j|L_{ij};t} - \rho_{i,j|L_{ij};t}| = |g(\hat{f}_{i,j|L_{ij};t}) - g(f_{i,j|L_{ij};t})| \leq \epsilon \cdot |\hat{f}_{i,j|L_{ij};t} - f_{i,j|L_{ij};t}|$, such that the e.a.s.

convergence of $\hat{\rho}_{i,j|L_{ij}\,;\,t}$ follows directly from that of $\hat{f}_{i,j|L_{ij}\,;\,t}$.

The e.a.s. convergence of $\hat{\boldsymbol{R}}_t$ follows similarly by combining the e.a.s. convergence of $\hat{\rho}_{i,j|L_{ij}\,;\,t}$, the properties of the mapping from $\rho_{i,j|L_{ij}\,;\,t}$ into $\boldsymbol{R}_t$, and the fact that under Assumption 1 the correlation matrices $\boldsymbol{R}_t$ and their filtered equivalents $\hat{\boldsymbol{R}}_t$ are never singular.

To numerically compute the supremum within the integral in Eq. (A.1), we note that the first order condition with respect to ρ boils down to solving for the roots of a 7th-degree polynomial:

$$e_2\rho^7 - 3e_2\rho^5 + e_1(\epsilon^2-1)\rho^4 + e_2(6-4\epsilon^2+\epsilon^4)\rho^3 - 3e_1(\epsilon^2-1)^2\rho^2 + e_2\epsilon^2(3\epsilon^2-4)\rho - e_1\epsilon^2(\epsilon^2-1) = 0.$$

Together with the boundary values $\rho = \pm\epsilon$, this gives at most 9 candidate points for the supremum. The polynomial roots can easily be found in standard numerical packages by supplying the coefficients of the polynomial to, for instance, `polyroot()` in R, `polynomial.polynomial.polyroots()` in Python, or `roots()` in Matlab. In practice, many of the roots are complex, such that the number of points to check for the supremum is typically even smaller. ∎

Proof of Proposition 3

Under the maintained assumptions, we can apply Proposition 2 to conclude that $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic. Using Assumption 1 and thus $g(f_{i,j|L_{ij}\,;\,t}) = \epsilon\cdot\arctan(f_{i,j|L_{ij}\,;\,t})$, we have

$$\begin{aligned}&\tfrac{1}{2}\hat{\boldsymbol{G}}_{i,j|L_{ij}\,;\,t}\boldsymbol{\mathcal{D}}_2^\top\left(\hat{\boldsymbol{R}}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\otimes\hat{\boldsymbol{R}}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right)^{-1} = \\ &\qquad \frac{1}{2}\frac{\epsilon}{1-\hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})^2}\cdot\Big(-2\hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\ ,\ 1+\hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})^2\ ,\ 1+\hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})^2\ ,\ -2\hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\Big).\end{aligned}$$

We define $\hat{\boldsymbol{y}}^\star_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})$ as

$$\begin{aligned}\hat{\boldsymbol{y}}^\star_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) &= \hat{\boldsymbol{D}}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})^{-1/2}\left(\boldsymbol{y}_{i,j\,;\,t}-\hat{\boldsymbol{\mu}}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right),\\ \hat{\boldsymbol{D}}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) &= \frac{\nu-2+\boldsymbol{y}_{L_{ij},t}^\top\hat{\boldsymbol{R}}^{-1}(\boldsymbol{\theta})_{L_{ij},L_{ij};t}\boldsymbol{y}_{L_{ij},t}}{\nu_{i,j|L_{ij}}}\begin{pmatrix}\hat{\boldsymbol{V}}_{i,i|L_{ij}\,;\,t}(\boldsymbol{\theta}) & 0\\ 0 & \hat{\boldsymbol{V}}_{j,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\end{pmatrix},\end{aligned}$$

as a perturbed bivariate data sequence, with $\hat{\boldsymbol{V}}_{i,i\,;\,t}(\boldsymbol{\theta}) = 1-\hat{\boldsymbol{R}}_{i,L_{ij};t}(\boldsymbol{\theta})\cdot\hat{\boldsymbol{R}}^{-1}_{L_{ij},L_{ij};t}(\boldsymbol{\theta})\cdot\hat{\boldsymbol{R}}_{L_{ij},i;t}(\boldsymbol{\theta})$. The perturbation is due to the initialization of the filter sequences. We can write the initialized filter recursions

$$\hat{f}_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) = \phi\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}), \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) = \omega_{i,j|L_{ij}} + \beta_{i,j|L_{ij}} \cdot \hat{f}_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) + \alpha_{i,j|L_{ij}} \cdot s_{i,j|L_{ij};t}(\boldsymbol{\theta}), \tag{A.2}$$

$$\begin{aligned}
s_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}), \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right) \\
&= J\left(\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) \cdot \Bigg(\\
&\qquad \left(1 + \hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(\hat{w}_{i,j|L_{ij};t}\hat{\boldsymbol{y}}^{\star}_{1,i,j|L_{ij};t}(\boldsymbol{\theta})\hat{\boldsymbol{y}}^{\star}_{2,i,j|L_{ij};t}(\boldsymbol{\theta}) - \hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) \\
&\qquad - \hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(\hat{w}_{i,j|L_{ij};t}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - 2\right)\Bigg),
\end{aligned} \tag{A.3}$$

$$J(\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})) = \frac{\epsilon\left(1 - \hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2/\epsilon^2\right)}{\left(1 - \hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)^2},$$

$$\begin{aligned}
\hat{w}_{i,j|L_{ij};t} &= \hat{w}_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}), \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right) \\
&= \frac{\nu_{i,j|L_{ij}} + 2}{\nu_{i,j|L_{ij}} + \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\hat{\boldsymbol{R}}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})}.
\end{aligned}$$

Also, we note that the contraction condition in equation (15) of Assumption 4 entails the following derivative

$$\begin{aligned}
s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})} &= \dot{J}\left(\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) \cdot \Bigg(\\
&\quad \left(1 + \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(w_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{1,i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^{\star}_{2,i,j|L_{ij};t}(\boldsymbol{\theta}) - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) \\
&\quad - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(\hat{w}_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - 2\right)\Bigg) \\
+ \epsilon \cdot \Bigg(& 2\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(w_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{1,i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^{\star}_{2,i,j|L_{ij};t}(\boldsymbol{\theta}) - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) \\
&\quad + \left(1 + \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(w^{f}_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{1,i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^{\star}_{2,i,j|L_{ij};t}(\boldsymbol{\theta}) - 1\right) \\
&\quad - \left(w_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - 2\right) \\
&\quad - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})w^{f}_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{1,i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^{\star}_{2,i,j|L_{ij};t}(\boldsymbol{\theta})\Bigg)
\end{aligned} \tag{A.4}$$

$$\dot{J}\left(\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) = \frac{-2\rho_{i,j|L_{ij};t}}{\epsilon\left(1 - \rho^2_{i,j|L_{ij};t}\right)^2} + \frac{4\epsilon\rho_{i,j|L_{ij};t}\left(1 - \rho^2_{i,j|L_{ij};t}/\epsilon^2\right)}{\left(1 - \rho^2_{i,j|L_{ij};t}\right)^3},$$

$$w^{f}_{i,j|L_{ij};t} = \frac{\partial w_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})} \tag{A.5}$$

$$= \frac{\partial \operatorname{vec}(\boldsymbol{R}_{i,j|L_{ij}\,;t})^\top}{\partial f_{i,j|L_{ij}\,;t}} \cdot \frac{\nu_{i,j|L_{ij}} + 2}{\left(\nu_{i,j|L_{ij}} + \boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^\top \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right)^2}$$
$$\left(\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \otimes \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right)\left(\boldsymbol{y}^\star_{i,j|L_{ij}\,;t} \otimes \boldsymbol{y}^\star_{i,j|L_{ij}\,;t}\right),$$

evaluated at some fixed point $f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) = f$.

We first *assume* that $\sup_{\theta\in\Theta} |\hat{\boldsymbol{y}}^\star_{L_{ij},L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{y}^\star_{L_{ij},L_{ij};t}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$, i.e., that $\hat{\boldsymbol{y}}^\star_{L_{ij},L_{ij};t}(\boldsymbol{\theta})$ converges uniformly e.a.s. to a unique stationary and ergodic limit $\boldsymbol{y}^\star_{L_{ij},L_{ij};t}(\boldsymbol{\theta})$, and then prove the e.a.s. convergence of $\hat{f}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ to $f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ and the existence of a log moment. The complete result then follows by induction after starting the recursion at $i - j = 1$ and noting that

$$\hat{\boldsymbol{y}}^\star_{i,i+1|L_{ij}\,;t}(\boldsymbol{\theta}) = \boldsymbol{y}^\star_{i,i+1|L_{ij}\,;t}(\boldsymbol{\theta}) = \hat{\boldsymbol{D}}_{i,i+1|L_{ij}\,;t}(\boldsymbol{\theta})^{-1/2}\boldsymbol{y}_{i,i+1\,;t} = \sqrt{\frac{(\nu-2)}{\nu}} \cdot \boldsymbol{y}_{i,i+1\,;t},$$
$$\hat{\boldsymbol{D}}_{i,i+1|L_{ij}\,;t}(\boldsymbol{\theta}) = \boldsymbol{D}_{i,i+1|L_{ij}\,;t}(\boldsymbol{\theta}) = \frac{(\nu-2)}{\nu} \cdot \mathbf{I}_2,$$

where $\hat{\boldsymbol{y}}^\star_{i,i+1\,;t}(\boldsymbol{\theta})$ is obviously stationary and ergodic due to the stationarity of $\boldsymbol{y}_{i,i+1\,;t}$.

For the remainder of the proof, we thus assume $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{\boldsymbol{y}}^\star_{L_{ij},L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{y}^\star_{L_{ij},L_{ij};t}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$. If we consider the filter recursion in (A.2) using the uninitialized stationary and ergodic $\boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ rather than the perturbed $\hat{\boldsymbol{y}}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$, we can easily see that a log moment exists for a fixed $\hat{f}_{i,j|L_{ij}\,;1}$:

$$\mathbb{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\phi\left(\hat{f}_{i,j|L_{ij}\,;1}, \boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right) - \hat{f}_{i,j|L_{ij}\,;1}\right|\right] \le$$
$$\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\left|\omega_{i,j|L_{ij}}\right| + \left|\beta_{i,j|L_{ij}} - 1\right| \cdot \left|\hat{f}_{i,j|L_{ij}\,;1}\right| + \left|\alpha_{i,j|L_{ij}}\right| \cdot \frac{K_1\epsilon}{1-\epsilon^2}, \right) \le K_2 < \infty,$$

where the K_i denote finite positive constants, and where we have used the uniform boundedness of the filtered $|\hat{\rho}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})| \le \epsilon$ via Assumption 1, as well as the uniform boundedness of the score expression $s_{i,j|L_{ij}\,;t}(\hat{f}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta})$ in $\boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ due to the analytical form of the filtered weights $\hat{w}_{i,j|L_{ij}\,;t}(\hat{f}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta})$.

Additionally, by similar arguments, we have that

$$\mathbb{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f} \left| \frac{\partial s_{i,j|L_{ij}\,;t}\left(f, \boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f} \right|\right] \le \log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\frac{K_1 2\epsilon^2}{1-\epsilon^2} + K_2\epsilon\right) \le K_3 < \infty.$$

The e.a.s. convergence of the filter that takes $\boldsymbol{y}^\star_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ as input to a unique stationary and ergodic limit then follows by Theorem 3.1 of Bougerol (1993) if we can prove that the filtering equation is contracting on average. This, however, follows immediately from Assumption 4.

The last part of the proof consists in showing that the perturbed filter recursions converge to the same limits as their unperturbed counterparts. Following Theorem 2.10 of Straumann and Mikosch (2006), this follows by showing

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| s_{i,j|L_{ij}\,;\,t}\left(\hat{f}_{i,j|L_{ij}\,;\,1}\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right) - s_{i,j|L_{ij}\,;\,t}\left(\hat{f}_{i,j|L_{ij}\,;\,1}\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right| \xrightarrow{e.a.s.} 0, \quad \text{(A.6)}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f} \left| \frac{\partial s_{i,j|L_{ij}\,;\,t}\left(f\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f} - \frac{\partial s_{i,j|L_{ij}\,;\,t}\left(f\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f}\right| \xrightarrow{e.a.s.} 0, \quad \text{(A.7)}$$

as $t \to \infty$.

To prove (A.6), note that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \hat{f}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) - \hat{\hat{f}}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right| \leq K \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right|,$$

where $K < \infty$ according to technical Lemma 1. Since $K < \infty$ and thus $\mathbb{E}[\log^{+} K] < \infty$, the desired convergence on the left hand side in (A.6) follows as an application of Lemma 2.1 of Straumann and Mikosch (2006) and the assumed e.a.s. convergence of $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})|$.

To prove (A.7), we note that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{\partial s_{i,j|L_{ij}\,;\,t}\left(f\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f} - \frac{\partial s_{i,j|L_{ij}\,;\,t}\left(f\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f}\right| \leq$$

$$K \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right|,$$

where $K < \infty$ according to technical Lemma 1. Similar as for (A.6), the result then follows as an application of Lemma 2.1 of Straumann and Mikosch (2006).

We can now conclude that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{f}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) - f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ for all $i = 2, \ldots, N$ and $j = 1, \ldots, i-1$.

To conclude the e.a.s. convergence of $\hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})$ to its limiting process, note that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) - \rho_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right| &= \epsilon \cdot \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\tanh(\hat{f}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})) - \tanh(f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}))\right| \\ &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\hat{f}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) - f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right|,\end{aligned}$$

where the inequality follows by taking a first order Taylor series expansion.

To conclude the e.a.s. convergence of $\hat{\rho}_{i,j\,;\,t}(\boldsymbol{\theta})$ to its limiting process, note that for $i = j+1$ we have $\hat{\rho}_{i,j\,;\,t}(\boldsymbol{\theta}) = \hat{\rho}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})$, such that the result follows directly from the e.a.s. convergence of the partial correlation. For $i > j+1$, the result then follows by induction. Note that from (4) we have

$$\hat{\rho}_{i,j\,;\,t} = \hat{\boldsymbol{R}}_{i,L_{ij};t}\hat{\boldsymbol{R}}^{-1}_{L_{ij},L_{ij};t}\hat{\boldsymbol{R}}_{L_{ij},j;t} + \hat{\rho}_{i,j|L_{ij}\,;\,t}\sqrt{\hat{\boldsymbol{V}}_{i,i|L_{ij}\,;\,t}\cdot\hat{\boldsymbol{V}}_{j,j|L_{ij}\,;\,t}}, \tag{A.8}$$

where $\hat{\boldsymbol{R}}_{L_{ij},L_{ij};t}$ is never singular due to Assumption 1 and $\epsilon < 1$. This mapping is a series of products and sums of elements of $\boldsymbol{R}_{L_{ij},L_{ij};t}$ and $\boldsymbol{R}_{i,L_{ij};t}$, each term of which converges e.a.s. to its limiting process by a direct application of Lemma 2.1 of Straumann and Mikosch (2006) and Lemma TA.16 of Blasques et al. (2022). for $i > j+1$. ∎

Proof of Theorem 1

By the triangle inequality, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})]\right| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta})\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}L_T(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})]\right|. \tag{A.9}$$

To show that the first term converges almost surely to zero, we write

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{\ell}_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta})\right|, \tag{A.10}$$

and then note that by the Cesaro mean, the first term on the right hand side of inequality (A.9) converges to zero almost surely if $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\hat{\ell}_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$.

We have

$$\begin{aligned}
&2\cdot(\hat{\ell}_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}))\\
&= \log|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta})| - \log|\boldsymbol{R}_t(\boldsymbol{\theta})| + (\nu+N)\left[\log\left(1 + \frac{\boldsymbol{y}_t^\top\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right) - \log\left(1 + \frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right]\\
&\leq tr\Big(\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})\Big) + \frac{\nu+N}{\nu-2}\boldsymbol{y}_t^\top\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\Big(\boldsymbol{R}_t(\boldsymbol{\theta}) - \hat{\boldsymbol{R}}_t(\boldsymbol{\theta})\Big)\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t,
\end{aligned}$$

where the inequality follows from Theorem 11.27 in Magnus and Neudecker (2019), Lemma A.1 of Bollerslev and Wooldridge (1992) and the standard log-inequality $\log(1+x) \leq x\ \forall x > -1$. Due to Assumption 1 with $\epsilon < 1$ and the mapping between partial and Pearson correlations, we automatically have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\right\| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\lambda_1^{-1}(\boldsymbol{R}_t(\boldsymbol{\theta})) < K, \tag{A.11}$$

for some $0 < K < \infty$

For any $N \times N$ matrix $\boldsymbol{A}$ it holds that $\operatorname{tr} \boldsymbol{A} \leq N \cdot \|\boldsymbol{A}\|$. We also have $\nu > 2$ by Assumption 2, while from Proposition 3, we obtain

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta}) \right| &\leq N\frac{1}{T}\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta}) \right\| + K_1\frac{1}{T}\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta}) \right\| \|\boldsymbol{y}_t\|^2 \\ &\leq N\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\gamma_{i,j}^{-t}\right) + K_1\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\gamma_{i,j}^{-t}\right)\|\boldsymbol{y}_t\|^2,\end{aligned}$$

for some $K, c > 0$ by following similar arguments as in Hafner and Preminger (2009). Since $\gamma_{i,j} > 1$ and $\mathbb{E}[\|\boldsymbol{y}_t\|^2] < \infty$, we obtain the desired almost sure convergence

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta}) \right| \xrightarrow{a.s.} 0,$$

as $T \to \infty$, by a straightforward application of the Markov's inequality and the Borel-Cantelli Lemma.

To prove the almost sure convergence of the second term on the right hand side of inequality (A.9), we only need to show that $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\theta})|] < \infty$ such that we can apply the uniform law of large numbers for stationary and ergodic processes of Rao (1962). Using the expression for the log-likelihood function from equation (10), we have

$$\begin{aligned}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ell_t(\boldsymbol{\theta})\right|\right] \leq &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log\Gamma\left(\frac{\nu+N}{2}\right)\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\log\frac{\nu}{2}\right| + \frac{N}{2}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\log((\nu-2)\pi)\right| \\ &+ \frac{1}{2}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\log|\boldsymbol{R}_t(\boldsymbol{\theta})|\right] + \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\nu+N}{2}\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|\right] < \infty, \quad \text{(A.12)}\end{aligned}$$

where the last inequality follows as a consequence of Assumptions 2 - 4, the uniform boundedness of $\boldsymbol{R}_t(\boldsymbol{\theta})$ (being a correlation matrix), the uniform lower bound from equation A.11, and the existence of second moments of $\boldsymbol{y}_t$. As a result, we obtain

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{T}L_T(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})] \right| \xrightarrow{a.s.} 0,$$

as $T \to \infty$.

To conclude the proof, we establish identifiability: $\mathbb{E}[\ell_t(\boldsymbol{\theta}_0)] > \mathbb{E}[\ell_t(\boldsymbol{\theta})]\ \forall \boldsymbol{\theta}_0 \neq \boldsymbol{\theta}$. The proof is by contradiction. Assume there is a $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ with $\mathbb{E}[\ell_t(\boldsymbol{\theta})] = \mathbb{E}[\ell_t(\boldsymbol{\theta}_0)]$, where $\mathbb{E}[\ell_t(\boldsymbol{\theta}_0)] < \infty$ by equation (A.12). By Gibb's inequality, this implies that $\nu = \nu_0$ and $\boldsymbol{R}_t(\boldsymbol{\theta}) = \boldsymbol{R}_t(\boldsymbol{\theta}_0)$ almost surely for this specific $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. This, however, leads to a contradiction. We note that there is a one-to-one relationship between the components of the lower (or upper) triangular part of the conditional correlation matrix

$\boldsymbol{R}_t(\boldsymbol{\theta})$, and the partial conditional correlations coefficients $\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})$ for $i = 2, \ldots, N$, $j = 1, \ldots, i-1$. Therefore $\boldsymbol{R}_t(\boldsymbol{\theta}) = \boldsymbol{R}_t(\boldsymbol{\theta}_0)$ (a.s.) implies $\rho_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \rho_{i,j|L_{ij};t}(\boldsymbol{\theta}_0)$ (a.s.). This, however, cannot hold for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, because the equality $\boldsymbol{R}_t(\boldsymbol{\theta}) = \boldsymbol{R}_t(\boldsymbol{\theta}_0)$ entails that

$$
\begin{aligned}
0 = f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}_0) - f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) =& \omega_{0,i,j|L_{ij}} - \omega_{i,j|L_{ij}} + (\alpha_{0,i,j|L_{ij}} - \alpha_{i,j|L_{ij}}) s^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \\
&+ (\beta_{0,i,j|L_{ij}} - \beta_{i,j|L_{ij}}) f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0),
\end{aligned}
$$

almost surely. We thus have

$$
(\alpha_{0,i,j|L_{ij}} - \alpha_{i,j|L_{ij}}) s^{\boldsymbol{\eta}}_{i,j|L_{ij};t} = \upsilon_{i,j|L_{ij};t}
$$

where $\upsilon_{i,j|L_{ij};t}$ is an $\mathcal{F}_t$-measurable random variable. It follows that since the conditional distribution of $\upsilon_{i,j|L_{ij};t}|\mathcal{F}_t$ is not degenerate, it must be that $\alpha_{0,i,j|L_{ij}} = \alpha_{i,j|L_{ij}}$, which yields

$$
0 = \omega_{0,i,j|L_{ij}} - \omega_{i,j|L_{ij}} + (\beta_{0,i,j|L_{ij}} - \beta_{i,j|L_{ij}}) f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0).
$$

This in turn implies $\omega_{0,i,j|L_{ij}} = \omega_{i,j|L_{ij}}$ and $\beta_{0,i,j|L_{ij}} = \beta_{i,j|L_{ij}}$ by the fact that $f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0)$ has a non-degenerate distribution given the non-degenerate distribution of $s^{\boldsymbol{\eta}}_{i,j|L_{ij};t}$ and Assumptions 3 and 4. This contradicts the initial premise $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and thus proves the theorem.

The strong consistency of the the MLE $\hat{\boldsymbol{\theta}}_T$ is then guaranteed by the compactness of the parameter space $\boldsymbol{\Theta}$ and noting that all the conditions of Theorem 3.4 in White (1994) are satisfied. ■

Proof of Theorem 2

By the strong consistency established in Theorem 1 combined with Assumption 5, we have that the MLE $\hat{\boldsymbol{\theta}}_T$ lies inside an arbitrarily small neighbourhood of $\boldsymbol{\theta}_0$ for sufficiently large T. Using the first order condition for the MLE from (11) and Lemma 5, we obtain

$$
\mathbf{0} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \hat{\ell}_t(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\hat{\boldsymbol{\theta}}_T) + o_p(1),
$$

where we note the difference between the log-likelihood functions $\hat{\ell}_t(\hat{\boldsymbol{\theta}}_T)$ and $\ell_t(\hat{\boldsymbol{\theta}}_T)$, the former using the initialized filter, and the latter using its stationary and ergodic limit.

Taking a Taylor expansion, we get

$$o_p(1) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) + \frac{1}{T} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_T^{\star}) \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_T^{\star}$ lies between $\hat{\boldsymbol{\theta}}_T$ and the true $\boldsymbol{\theta}_0$. For sufficiently large T we then obtain that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) + o_p(1) = -\frac{1}{T} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_T^{\star}) \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0). \tag{A.13}$$

In Lemma 4, we prove that $T^{-1/2} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)$ obeys the central limit theorem for martingales of Billingsley (1961) and satisfies the Fisher's information matrix equality. Moreover, Lemma 6 ensures that the average $-T^{-1} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_T^{\star})$ converges to the positive definite Fisher's information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$, almost surely. Hence, as $T \to \infty$, by solving equation (A.13), we obtain by the Slutsky's Theorem (see Vaart (1998)) that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}_0)).$$

■

B Technical Lemmas

Define the operators $\nabla^{\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}}$ and $\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$, where $\boldsymbol{\theta}$ contains ν, $\omega_{i,j|L_{ij}}$, $\alpha_{i,j|L_{ij}}$, $\beta_{i,j|L_{ij}}$, for $i = 1, \ldots, N-1$ and $j = i+1, \ldots, N$. To avoid ambiguous notations, we also define $\nabla^{\nu} = \frac{\partial}{\partial \nu}$, $\nabla^{\nu\nu} = \frac{\partial^2}{\partial \nu^2}$. We use $\psi(x) = \frac{\partial}{\partial x} \log \Gamma(x)$ to denote the usual digamma function.

Lemma 1. *Consider the score expression and its derivative with respect to* $f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$

$$s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}\right) = s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right),$$

$$s^{f}_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}\right) = \frac{\partial s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\ ;\ \boldsymbol{\theta})}{\partial f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})},$$

as defined in equations (A.3) *and* (A.4)*, respectively.*

Under Assumption 1, we have that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\partial s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})} \right| < \infty, \tag{B.1}$$

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\partial^2 s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \partial \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})} \right| < \infty. \tag{B.2}$$

Proof. By straightforward algebra we get that

$$\begin{aligned}
&\frac{\partial s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})} = J\left(\rho_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right) \cdot \Bigg(\\
&\left(1 + \rho_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^2\right) \left(w^{\boldsymbol{y}^{\star}}_{i,j|L_{ij}\,;t} \boldsymbol{y}^{\star}_{1,i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{2,i,j|L_{ij}\,;t}(\boldsymbol{\theta}) + w_{i,j|L_{ij}\,;t} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right) \\
&\qquad - \rho_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \left(w^{\boldsymbol{y}^{\star}}_{i,j|L_{ij}\,;t} \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^\top \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right) \Bigg)
\end{aligned}$$

$$\begin{aligned}
w^{\boldsymbol{y}^{\star}}_{i,j|L_{ij}\,;t} &= \frac{\partial w_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})} \\
&= \frac{-2\left(\nu_{i,j|L_{ij}} + 2\right)}{\left(\nu_{i,j|L_{ij}} + \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^\top \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right)^2} \cdot \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}),
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})\partial \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})} = \dot{J}\left(\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\cdot\Bigg(\\
&\left(1+\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(w^{\boldsymbol{y}^\star}_{i,j|L_{ij};t}\boldsymbol{y}^\star_{1,i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^\star_{2,i,j|L_{ij};t}(\boldsymbol{\theta}) + w_{i,j|L_{ij};t}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\\
&\qquad\qquad - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(w^{\boldsymbol{y}^\star}_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Bigg)\\
&+\epsilon\cdot\Bigg(\left(\left(2\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})-1\right)\left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\right)\\
&\qquad\left(w^{\boldsymbol{y}^\star}_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta}) + 2w_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\\
&+\left(w^{f\boldsymbol{y}}_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta}) + 2w_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Bigg)\\
&w^{f\boldsymbol{y}^\star}_{i,j|L_{ij};t} = \frac{\partial^2 w_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}),\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta});\boldsymbol{\theta}\right)}{\partial f\partial \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})}\\
&\quad = \frac{\partial\,\mathrm{vec}(\boldsymbol{R}_{i,j|L_{ij};t}(\boldsymbol{\theta}))^\top}{\partial f_{i,j|L_{ij};t}}\cdot\frac{-4(\nu_{i,j|L_{ij}}+2)}{\left(\nu_{i,j|L_{ij}}+\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)^2}\\
&\qquad\qquad \boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\otimes\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\left(\boldsymbol{y}^\star_{i,j|L_{ij};t}\otimes\boldsymbol{y}^\star_{i,j|L_{ij};t}\right)\\
&\qquad + \frac{\nu_{i,j|L_{ij}}+2}{\left(\nu_{i,j|L_{ij}}+\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)^2}\\
&\qquad\qquad \left(\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\otimes\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\left(\left(\mathbf{I}_2\otimes\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)+\left(\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\otimes\mathbf{I}_2\right)\right).
\end{aligned}
$$

By exploiting the analytical forms of the weights $w_{i,j|L_{ij};t}$, $w^{\boldsymbol{y}^\star}_{i,j|L_{ij};t}$ and $w^{f\boldsymbol{y}^\star}_{i,j|L_{ij};t}$, and the parameterization given in Assumption 1, we can show that the uniform bounds in equations (B.1) and (B.2) are easily satisfied.

In fact, one only needs to note that there exists general positive constants K_1 and K_2, such that

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})}\right| \le \frac{1}{1-\epsilon^2}\left(\left(1-\epsilon\left(1-\epsilon\right)\right)\left(K_1+2K_2\right)\right) < \infty,
$$

and also that

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})\partial \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})}\right| \le
$$

$$\frac{2\epsilon^2}{1-\epsilon^2}\left(\left(1-\epsilon\left(1-\epsilon\right)\right)\left(K_1+2K_2\right)\right)+\epsilon\cdot\left(\left(1+\left(\left(2\epsilon-1\right)\left(1-\epsilon\right)\right)\right)\left(K_1+2K_2\right)\right)<\infty,$$

■

Lemma 2. *Under the Assumptions 1-4:*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}}\hat{f}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})-\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|\xrightarrow{e.a.s.}0, \tag{B.3}$$

as $t\to\infty$.

Furthermore, under the additional Assumption 6, we have

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|^m\right]<\infty, \tag{B.4}$$

for any integer $m\geq 2$.

Proof. As in the proof of Proposition 3, to prove the uniform exponentially fast convergence in (B.3), we can show that the conditions S.1-S.3 of Theorem 2.10 in Straumann and Mikosch (2006) hold true for the first derivative processes

$$\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})=\begin{pmatrix}\nabla^{\omega}f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})\\ \nabla^{\alpha}f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})\\ \nabla^{\beta}f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})\\ \nabla^{\nu}f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})\end{pmatrix}=\boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})+X_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \tag{B.5}$$

where

$$X_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})=\beta_{i,j|L_{ij}}+\alpha_{i,j|L_{ij}}\cdot s^f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})=\beta_{i,j|L_{ij}}+\alpha_{i,j|L_{ij}}\cdot\frac{\partial s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}),\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta});\boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})}, \tag{B.6}$$

$$\boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})=\begin{pmatrix}\frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial\omega_{i,j|L_{ij}}}\\ \frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial\alpha_{i,j|L_{ij}}}\\ \frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial\beta_{i,j|L_{ij}}}\\ \frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial\nu}\end{pmatrix},$$

such that

$$\frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}} = 1, \qquad \frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}}} = s_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}),$$

$$\frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}}} = f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \qquad \frac{\partial f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})}{\partial \nu_{i,j|L_{ij}}} = \alpha_{i,j|L_{ij}} s^{\nu}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}),$$

where the term $s^{\nu}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) = s^{\nu}_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)$ is given by

$$\begin{aligned} s^{\nu}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) &= \frac{\partial s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \nu} \\ &= J\left(\rho_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right) \cdot \Bigg(\\ &\qquad \left(1 + \rho_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^2\right) \left(w^{\nu}_{i,j|L_{ij}\,;t} \boldsymbol{y}^{\star}_{1,i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{2,i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right) \\ &\qquad\qquad - \rho_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \left(w^{\nu}_{i,j|L_{ij}\,;t} \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^{\top} \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right) \Bigg). \end{aligned} \tag{B.7}$$

$$\begin{aligned} w^{\nu}_{i,j|L_{ij}\,;t} &= \frac{\partial w_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \nu} \\ &= \frac{\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^{\top} \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) - 2}{\left(\nu_{i,j|L_{ij}} + \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^{\top} \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right)^2}. \end{aligned} \tag{B.8}$$

We start by verify conditions S.1 and S.2 in Theorem 2.10 of Straumann and Mikosch (2006), which are directly implied if the following uniform bounds

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| X_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right|\right] < \infty, \qquad \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right\|\right] < \infty.$$

However, we first note that, by Proposition 3, it holds that $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| X_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right|\right] < K_3 < \infty$, and furthermore

$$\begin{aligned} \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right\|\right] \leq & 1 + \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| s_{i,j|L_{ij}\,;t}\,(\boldsymbol{\theta}) \right|\right] + \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right|\right] \\ & + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \alpha_{i,j|L_{ij}} \right| \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| s^{\nu}_{i,j|L_{ij}\,;t}\,(\boldsymbol{\theta}) \right|\right] < \infty, \end{aligned}$$

which is again implied by Proposition 3, the compactness of the parameter space, and the fact that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| s^{\nu}_{i,j|L_{ij}\,;t}\,(\boldsymbol{\theta}) \right| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| s^{\nu}_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right) \right| \leq \left(1 - \epsilon\,(1 - \epsilon)\right) \frac{K_1 \epsilon}{1 - \epsilon^2} \leq K_2 < \infty.$$

Then, conditions S.1 and S.2 in Theorem 2.10 of Straumann and Mikosch (2006) are directly satisfied.

Now, in the present case, proving that condition S.3 in Theorem 2.10 of Straumann and Mikosch (2006) is equivalent of proving that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \hat{\boldsymbol{w}}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) - \boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right\| \xrightarrow{e.a.s.} 0, \qquad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \hat{X}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) - X_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right\| \xrightarrow{e.a.s.} 0.$$

By Proposition 3, Lemma 1 and invoking again the mean value theorem, it is immediate to infer that, for a sufficiently large t, we obtain that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \hat{\boldsymbol{w}}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) - \boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right\| \leq K \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right|, \xrightarrow{e.a.s.} 0,$$

where $K < \infty$, by an application with Lemma 2.1 of Straumann and Mikosch (2006). Analogously

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \hat{X}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) - X_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right\| \leq K \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \right|, \xrightarrow{e.a.s.} 0,$$

We then conclude that S.3 is satisfied and (B.3) holds true.

Finally, we prove the existence of the integer $m \geq 1$ in (B.4), i.e., the arbitrary large number of bounded moments of the derivative processes. We remark again that we give details for the derivatives in (i). The fact that $\{\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and are stationary and ergodic implies that they admit the following almost sure representations

$$\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta}) = \boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) + \sum_{p=1}^{\infty} \left(\prod_{q=1}^{p} X_{i,j|L_{ij}\,;t-q} \right) \boldsymbol{w}_{i,j|L_{ij}\,;t-p}(\boldsymbol{\theta}),$$

Now, by Assumption 4, the compactness of the parameter space $\boldsymbol{\Theta}$, and the uniformly boundedness of the score function (and its derivative), it holds that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} \left\| \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta}) \right\| \leq K_w + \sum_{p=1}^{\infty} \gamma_{i,j|L_{ij}}^{-p} \sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} \left\| \boldsymbol{w}_{i,j|L_{ij}\,;t-p}(\boldsymbol{\theta}) \right\|,$$

where $K_w \geq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{w}_{i,j\,;t}(\boldsymbol{\theta})\|$.

Thus, the result in (B.4) can be established by repeated applications of the Minkowski and Hölder inequalities. This result follows because $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} |f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})|^m] < \infty$ with $m \geq 1$ as implied by Proposition 3, together with Proposition TA.3 of Blasques et al. (2022) to the unperturbed derivative processes $\{\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$. In fact, we only need to note that their conditions (iii) and (iv) are directly implied by the uniform bound of the score equations together with Assumption 4. ■

Lemma 3. *Under the Assumptions 1-6:*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})-\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\xrightarrow{e.a.s.}0, \tag{B.9}$$

as $t\to\infty$.

Furthermore, we also have

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|^m\right]<\infty, \tag{B.10}$$

for any integer $m\geq 1$.

Proof. To prove this Lemma, we can show again that the conditions S.1-S.3 in Theorem 2.10 of Straumann and Mikosch (2006) for the second derivative processes hold true for the second derivative processes

$$\begin{aligned}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})=&\begin{pmatrix}\nabla^{\omega\omega}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) & \nabla^{\omega\alpha}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) & \nabla^{\omega\beta}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) & \nabla^{\omega\nu}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})\\ \star & \nabla^{\alpha\alpha}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) & \nabla^{\alpha\beta}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) & \nabla^{\alpha\nu}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})\\ \star & \star & \nabla^{\beta\beta}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) & \nabla^{\beta\nu}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})\\ \star & \star & \star & \nabla^{\nu\nu}f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})\end{pmatrix}\\ =&\boldsymbol{W}_{i,j|L_{ij};t}(\boldsymbol{\theta})+X_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta}),\end{aligned}$$

with $X_{i,j|L_{ij};t}(\boldsymbol{\theta})$ as defined in equation (B.6) and

$$\boldsymbol{W}_{i,j|L_{ij};t}(\boldsymbol{\theta})=\begin{pmatrix}\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega^2_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega_{i,j|L_{ij}}\partial\alpha_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega_{i,j|L_{ij}}\partial\beta_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega_{i,j|L_{ij}}\partial\nu_{i,j|L_{ij}}}\\ \star & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\alpha^2_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\alpha_{i,j|L_{ij}}\partial\beta_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\alpha_{i,j|L_{ij}}\partial\nu_{i,j|L_{ij}}}\\ \star & \star & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\beta^2_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\beta_{i,j|L_{ij}}\partial\nu_{i,j|L_{ij}}}\\ \star & \star & \star & \frac{\partial^2 f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial\nu^2}\end{pmatrix},$$

such that

$$\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega^2_{i,j|L_{ij}}}=\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega_{i,j|L_{ij}}\partial\alpha_{i,j|L_{ij}}}=\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega_{i,j|L_{ij}}\partial\beta_{i,j|L_{ij}}}=\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\omega_{i,j|L_{ij}}\partial\nu_{i,j|L_{ij}}}=0,$$

$$\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial\alpha^2_{i,j|L_{ij}}}=s^f_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^\alpha f_{i,j|L_{ij};t}(\boldsymbol{\theta})+\alpha_{i,j|L_{ij}}s^{ff}_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^\alpha f_{i,j|L_{ij};t}(\boldsymbol{\theta})^2$$

$$
\begin{aligned}
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \beta^2_{i,j|L_{ij}}} &= 2\nabla^{\beta} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) + \alpha_{i,j|L_{ij}} s^{ff}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\beta} f_{i,j|L_{ij};t}(\boldsymbol{\theta})^2 \\
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \nu^2_{i,j|L_{ij}}} &= \alpha_{i,j|L_{ij}} \left(s^{\nu\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + s^{ff}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\nu} f_{i,j|L_{ij};t}(\boldsymbol{\theta})^2 \right),
\end{aligned}
$$

and moreover, we have

$$
\begin{aligned}
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}} \partial \beta_{i,j|L_{ij}}} &= \left(1 + \alpha_{i,j|L_{ij}} s^{ff}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\beta} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) \nabla^{\alpha} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \\
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}} \partial \nu_{i,j|L_{ij}}} &= s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + \alpha_{i,j|L_{ij}} s^{f\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\nu} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\alpha} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \\
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}} \partial \nu_{i,j|L_{ij}}} &= \nabla^{\nu} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) + \alpha_{i,j} s^{f\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\nu} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\beta} f_{i,j|L_{ij};t}(\boldsymbol{\theta}),
\end{aligned}
$$

where

$$
\begin{aligned}
s^{ff}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})^2}, \\
s^{\nu\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \nu^2_{i,j|L_{ij};t}}, \\
s^{f\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \partial \nu}.
\end{aligned}
$$

From this formulas and Proposition 3.4 of Blasques et al. (2022) it is obvious that the same arguments discussed in Lemma 2 apply sequentially, yielding the desired results in (B.9) and (B.10). ■

Lemma 4. *Under Assumption 1-6, the process* $\{\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$ *is a square integrable martingale difference, that is,* $\mathbb{E}[\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) | \mathcal{F}_{t-1}] = \mathbf{0}$ *and* $\mathbb{E}[(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))^{\top}] < \infty$.

Moreover, we have that

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) \Rightarrow \mathcal{N}\left(\mathbf{0}, \mathbb{E}[(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))^{\top}]\right).
$$

Proof. To show the zero mean property of the score vector, we take term-wise derivatives of the log-likelihood function $\ell_t(\boldsymbol{\theta})$ in (10) for each couple of indices (i, j), in order to obtain the following score vector:

$$
\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) = \begin{pmatrix} \nabla^{\nu} \ell_t(\boldsymbol{\theta}) \\ \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) s_{i,j|L_{ij};t}(\boldsymbol{\theta}) \end{pmatrix},
$$

where

$$\nabla^{\nu}\ell_t(\boldsymbol{\theta}) = \frac{1}{2}\left[\psi\left(\frac{\nu+N}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{N}{\nu-2} - \frac{\boldsymbol{y}_t^\top \boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2} w_t - \log\left(1 + \frac{\boldsymbol{y}_t^\top \boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right],$$

$$w_t = w_t\left(\boldsymbol{R}_t(\boldsymbol{\theta})\,,\,\boldsymbol{y}_t\,;\,\boldsymbol{\theta}\right) = \frac{\nu+N}{\nu-2+\boldsymbol{y}_t^\top \boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t},$$

and with $s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}\right) = s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\,,\,\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\,;\,\boldsymbol{\theta}\right)$ as defined in equation (A.3), respectively.

Now, by a straightforward application of the conditional expectation we obtain

$$\mathbb{E}\left[\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\left.\begin{pmatrix}\nabla^{\nu}\ell_t(\boldsymbol{\theta}_0)\\ \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}_0)s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}_0\right)\end{pmatrix}\right|\mathcal{F}_{t-1}\right] = \begin{pmatrix}0\\0\end{pmatrix},$$

where the last equality follow because the derivatives $\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}_0)$ are $\mathcal{F}_{t-1}$-measurable, whereas the conditional expectations of $\nabla^{\nu}\ell_t(\boldsymbol{\theta}_0)$, and $s_{i,j|L_{ij}\,;t}(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}_0)\,,\,\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}_0)\,;\,\boldsymbol{\theta}_0)$ are obviously zero almost surely, since, by Assumption 2, they are the terms of the conditional score vector of the multivariate Student t density function evaluated at the true parameter vector $\boldsymbol{\theta}_0$. On the other hand, to show that $\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)$ is square integrable, it suffices to prove that the derivatives of the log-likelihood have a uniformly bounded second moment, that is

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|^2\right] < \infty. \tag{B.11}$$

An application of the Cauchy-Schwartz inequality, we can show that

$$\begin{aligned}
\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|^2\right] &\leq \mathbb{E}\left[\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right| + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}\right)\right\|\right)^2\right]\\
\leq \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right|^2\right] &+ 2\left(\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right|^2\right]\cdot\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|^2\cdot\left(\frac{K_1\epsilon}{1-\epsilon^2}\right)^2\right]\right)^{1/2}\\
&+ \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|^2\cdot\left(\frac{K_1\epsilon}{1-\epsilon^2}\right)^2\right],
\end{aligned}$$

where the last inequality follows by the arguments discussed in Proposition 3 since the uniform boundedness of the score $s_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) = s_{i,j|L_{ij}\,;t}(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta})$ implies the existence of an arbitrary large number of bounded moments. Hence

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|s_{i,j|L_{ij};t}\left(\boldsymbol{\theta}\right)\right\|^2\right] = \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right\|^2\right] \le \left(\frac{K_1\epsilon}{1-\epsilon^2}\right)^2 < \infty.$$

Moreover, from Lemma 2 it also holds that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|^2\right] < \infty.$$

Now, by the compactness of the parameter space $\boldsymbol{\Theta}$, we can also show that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right|^2\right]\\
\le&\frac{1}{4}\mathbb{E}\left[\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\psi\left(\frac{\nu+N}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{N}{\nu-2}-\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2}w_t\right|+\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|\right)^2\right]\\
\le&\frac{1}{4}\mathbb{E}\left[|K_\nu-K_1|^2\right]+\frac{1}{2}\left(|K_\nu-K_1|\cdot\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|\right]\right)\\
&\qquad+\frac{1}{4}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|^2\right],
\end{aligned}$$

where the second inequality holds because from the compactness of the parameter space $\boldsymbol{\Theta}$ with $2<\nu<\infty$, $\exists K_\nu \ge \psi\left(\frac{\nu+N}{2}\right)+\psi\left(\frac{\nu}{2}\right)+\frac{N}{\nu-2}$, together with the the analytical form of the weights w_t, which implies that $\exists K_1>0$ such that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2}w_t\right|^2\right] \le K_1 < \infty.$$

Moreover, it is obvious that from the second moment bound $\mathbb{E}[\|\boldsymbol{y}_t\|^2]<\infty$ and the lower bound in (A.11) we also have that $\exists K_2>0$ such that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|^2\right] \le K_2 < \infty,$$

by virtue of the inequality $\log(1+x)\le x\ \forall x\ge -1$. By collecting all the results obtained above, we conclude that (B.11) holds true.

Finally, we simply note that the Fisher's information equality $\mathbb{E}[(\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0))(\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0))^\top]=\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$

follows by standard arguments since by Assumption 2 the $\ell_t(\boldsymbol{\theta}_0)$ is the true conditional log-density of the Student's t distribution. This concludes the proof.

■

Lemma 5. *Under Assumptions 1-4,*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{T}\nabla^{\boldsymbol{\theta}} \hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}\nabla^{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}) \right\| \xrightarrow{a.s.} 0, \tag{B.12}$$

as $T \to \infty$.

Proof. An application of the triangle inequality yields

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{1}{T}\nabla^{\boldsymbol{\theta}} \hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}\nabla^{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}) \right\| \leq \frac{1}{T}\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^{\boldsymbol{\theta}} \hat{\ell}_t(\boldsymbol{\theta}) - \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) \right\| \leq \frac{1}{T}\sum_{t=1}^{T} \Big(I + II \Big),$$

with

$$I := \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Bigg\| \nabla^{\boldsymbol{\theta}} \hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;,\; \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;;\; \boldsymbol{\theta} \right) \\ - \; \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;,\; \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;;\; \boldsymbol{\theta} \right) \Bigg\|,$$

and

$$II := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^{\nu} \hat{\ell}_t(\boldsymbol{\theta}) - \nabla^{\nu} \ell_t(\boldsymbol{\theta}) \right\|.$$

As a first step, we focus on I. We recognize that each term of

$$\nabla^{\boldsymbol{\theta}} \hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;,\; \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;;\; \boldsymbol{\theta} \right)$$

is a continuous function of $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and its derivatives. In contrast, the terms in

$$\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;,\; \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;;\; \boldsymbol{\theta} \right)$$

are continuous functions of the stationary counterparts, i.e. $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and its derivatives. Therefore, by means of elementary decomposition, we can write

$$I \leq \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^{\boldsymbol{\theta}} \hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \;-\; \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \right\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \right\| \right) \tag{B.13}$$

$$\times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right) - s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right|$$
$$+ \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^{\boldsymbol{\theta}} \hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ -\ \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right|.$$

Now, we note that in view of Proposition 3 and Lemma 2, we can easily show that both the first and the second addends of the inequality (B.13) vanish *e.a.s.* as $t \to \infty$, as implied by Lemma 2.1 in Straumann and Mikosch (2006).

Therefore, there exists some finite constant $K_I > 0$ such that

$$I \leq K_I \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \gamma_{i,j}^{-t}$$

and since for $\gamma_{i,j}^{-t} < 1\ \forall t \in \mathbb{N}$, we obtain that $I \xrightarrow{e.a.s.} 0$, as $t \to \infty$.

As concerns II, we have that

$$\nabla^{\nu}\hat{\ell}_t(\boldsymbol{\theta}) - \nabla^{\nu}\ell_t(\boldsymbol{\theta}) = \frac{1}{2}\Bigg[\frac{\boldsymbol{y}_t^{\top}\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2} w_t - \frac{\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2}\hat{w}_t$$
$$+ \log\left(1 + \frac{\boldsymbol{y}_t^{\top}\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right) - \log\left(1 + \frac{\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\Bigg]. \tag{B.14}$$

We can then combine the facts that: (i) $0 < \nu < \infty$ by Assumption 2, (ii) the lower bound obtained in (A.11) and (iii) the uniform bound $\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}} |w_t| \leq 1$, in order to see that for the first added in squared brackets of the right hand side of equation (B.14), it holds that

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{\boldsymbol{y}_t^{\top}\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2}\hat{w}_t - \frac{\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2} w_t\right| &\leq c_\nu \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| tr\Big(\boldsymbol{y}_t\boldsymbol{y}_t^{\top}(\boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) - \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}))\Big)\right| \\
&= c_\nu \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| tr\Big(\boldsymbol{y}_t\boldsymbol{y}_t^{\top}\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})(\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta}))\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\Big)\right| \\
&\leq c_\nu \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\right\| \left\|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})\right\| \left\|\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\right\| \|\boldsymbol{y}_t\boldsymbol{y}_t^{\top}\| \\
&\leq c_\nu K \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})\right\| \|\boldsymbol{y}_t\|^2.
\end{aligned}$$

Moreover, since $\log x \leq x - 1\ \forall x \geq 1$, the same result holds for the second added in squared brackets of the right hand side of equation (B.14), in fact

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \log\left(1 + \frac{\boldsymbol{y}_t^{\top}\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right) - \log\left(1 + \frac{\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right| &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{\boldsymbol{y}_t^{\top}\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2} - \frac{\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right| \\
&\leq c_\nu K \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})\right\| \|\boldsymbol{y}_t\|^2,
\end{aligned}$$

for some $K, c_\nu > 0$. We can now recall that the conditional correlation matrix $\hat{\boldsymbol{R}}_t(\boldsymbol{\theta})$ is a continuous function of each $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, whereas $\boldsymbol{R}_t(\boldsymbol{\theta})$ is a continuous function of each of the stationary counterpart $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$. Therefore, by Proposition 3 it holds that

$$II \leq 2c_\nu K \|\boldsymbol{y}_t\|^2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \gamma_{i,j}^{-t},$$

and since $\gamma_{i,j}^{-t} < 1$, $\forall t \in \mathbb{N}$ and $\mathbb{E}[\|\boldsymbol{y}_t\|^2] < \infty$ we obtain that $II \xrightarrow{e.a.s.} 0$, as $t \to \infty$.

In conclusion, the uniform convergence in (B.12) holds true. ∎

Lemma 6. *Under Assumptions 1-6,*

$$-\frac{1}{T}\sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\hat{\boldsymbol{\theta}}_T) \xrightarrow{a.s.} -\mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)] = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0), \tag{B.15}$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is positive definite.

Proof. First, we establish the almost sure convergence in (B.15), by proving that the second derivatives of the log-likelihood function has a uniformly bounded moment, that is

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) \right\| \right] < \infty. \tag{B.16}$$

Then, analogously to the Proof of Thoerem 1, we apply again the uniform law of large numbers for stationary and ergodic processes of Rao (1962).

Taking term-wise second derivatives of the log-likelihood function $\ell_t(\boldsymbol{\theta})$ in (10) for each couple of indices (i, j), we obtain the following Hessian matrix:

$$\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) = \begin{pmatrix} \nabla^{\nu\nu}\ell_t(\boldsymbol{\theta}) & \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \\ \star & \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \left(s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\boldsymbol{\theta}} f^{\top}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + s_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \right) \end{pmatrix},$$

where

$$\nabla^{\nu\nu}\ell_t(\boldsymbol{\theta}) = \frac{1}{4}\left[\psi\prime\left(\frac{\nu+N}{2}\right) - \psi\prime\left(\frac{\nu}{2}\right) + \frac{2N}{(\nu-2)^2} + \frac{4\boldsymbol{y}_t^\top \boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^3} w_t + \frac{2N\boldsymbol{y}_t^\top \boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^3(\nu+N)^2} w_t^3 - \frac{2(\nu+N)\boldsymbol{y}_t^\top \boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2} w_t \right],$$

with w_t as defined in Lemma 4, $s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, $s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and $s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, are as defined in equations

(A.3), (A.4) and (B.7), respectively. Now, by elementary calculations, we note that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|\right] \leq \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\nu\nu}\ell_t(\boldsymbol{\theta})\right\|\right] + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\right]$$
$$+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}}f^{\top}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + s_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\right].$$

In view of the the uniform boundedness properties of the score expression and the results obtained in Proposition 3, and Lemmas 2 and 3 it is clear that all the elements of the Hessian matrix are still polynomials of uniformly bounded random variables, with an arbitrary large number of finite moments. Thus, repeated applications of the Cauchy-Schwartz inequality and the Minkowski inequality, and some straightforward calculations, allow us to conclude that also the Hessian matrix is uniformly bounded, and hence (B.16) holds true. Therefore, a straightforward application of the uniform law of large number for stationary and ergodic sequences of Rao (1962) give us the desired almost sure convergence in (B.15).

Second, we show that $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is positive definite. To do so, we note that the strong consistency of the MLE $\hat{\boldsymbol{\theta}}_T$ established in Theorem 1, implies that as $T\to\infty$, $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ and hence $\hat{\boldsymbol{\theta}}_T \in V(\boldsymbol{\theta}_0)$ almost surely, where $V(\boldsymbol{\theta}_0)$ denotes a neighbourhood of $\boldsymbol{\theta}_0$.

We thus have that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)]\right\| \leq \left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0) - \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)]\right\|$$
$$+\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}) - \frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)\right\|.$$

However, since $\{\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is stationary and ergodic, it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)] = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0),$$

and therefore, by the uniform law of large numbers of Rao (1962), $\exists\delta>0$ such that

$$\lim_{T\to\infty}\left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)]\right\| \leq \delta.$$

As the constant $\delta > 0$ can be chosen as small as we want, we conclude that the almost sure convergence in (B.15) holds true.

In conclusion, it remains to be shown that $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is invertible. As argued by Darolles et al. (2018)

in their proof of Theorem 4.3, if $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ were not invertible, then there would exists some vector $\boldsymbol{\lambda} \in \mathbb{R}^d$, where d denotes the dimension of the compact parameter space $\boldsymbol{\theta}$, such that $\boldsymbol{\lambda}^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)\boldsymbol{\lambda} = 0$ with $\boldsymbol{\lambda} \neq \mathbf{0}$. We recall that in our setting of a correctly specified model the usual Bartlett identities hold and the Fisher information matrix equals the covariance matrix of the score vector: $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0) = -\mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)] = \mathbb{E}[(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))^\top]$. If $\exists \boldsymbol{\lambda} \neq \mathbf{0}$ such that $\boldsymbol{\lambda}^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)\boldsymbol{\lambda} = 0$, it must then also hold that the same $\boldsymbol{\lambda}$ makes the covariance matrix of the likelihood scores singular. Given that the score has expectation zero, it must then hold that for this linear combination $\boldsymbol{\lambda}^\top \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) \equiv 0$, or, more precisely, that

$$\boldsymbol{\lambda}^\top \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) = \boldsymbol{\lambda}^\top \nabla^{\nu} \ell_t(\boldsymbol{\theta}_0) + \boldsymbol{\lambda}^\top \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \nabla^{\theta} f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0) s_{i,j|L_{ij};t}(\boldsymbol{\theta}_0) = 0$$

almost surely, which given the non-off-setting expressions for the log-likelihood derivatives implies that each of the terms must be zero almost surely. As both $\boldsymbol{\lambda}^\top \nabla^{\nu} \ell_t(\boldsymbol{\theta}_0)$ and the score $s_{i,j|L_{ij};t}(\boldsymbol{\theta}_0)$ are non-degenerate random variables and the derivatives $\nabla^{\theta} f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0)$ also converge to a non-degenerate stochastic process, the log-likelihood derivative $\boldsymbol{\lambda}^\top \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)$ is a non-degenerate random variable, such that the above equality can only hold if $\boldsymbol{\lambda} = \mathbf{0}$. This provides a contradiction and thus proves the result.

■

C Additional empirical results

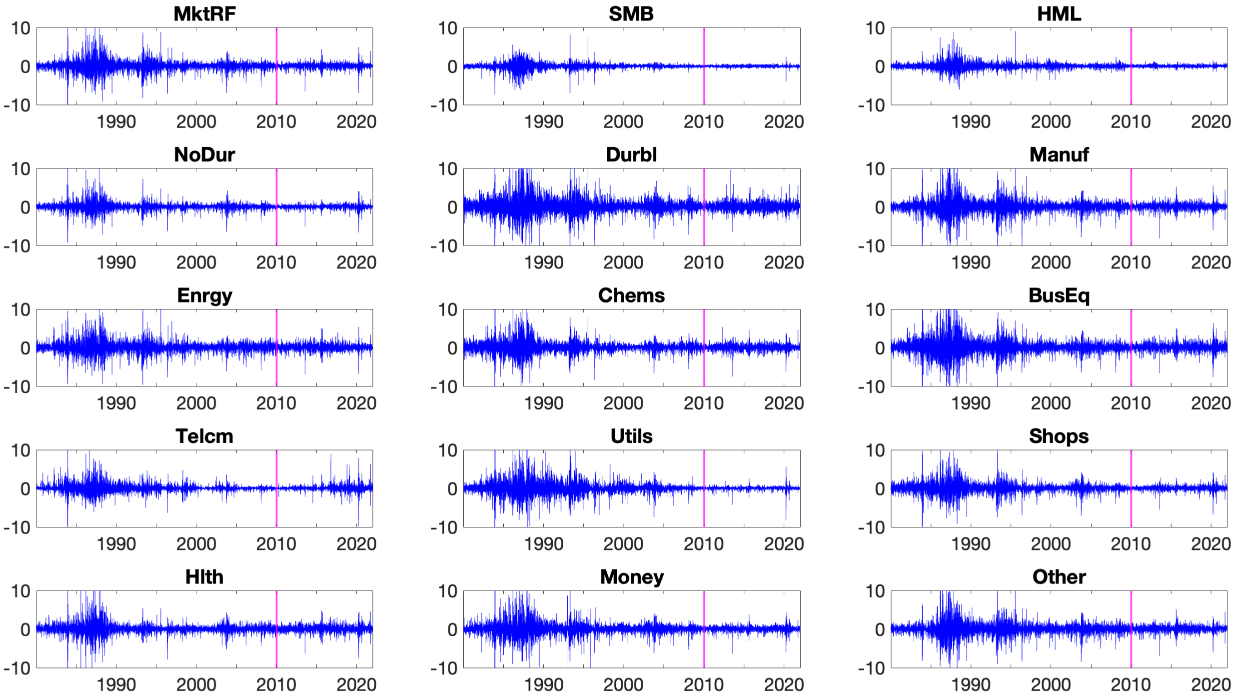


Figure C.1: Daily returns on the three main risk factors and the twelve industry portfolios

Note: The period is 03 January 1980 to 31 December 2021. The vertical lines indicate the 4th of January 2010, i.e. the first trading day of 2010 and the start of the out-of-sample period.

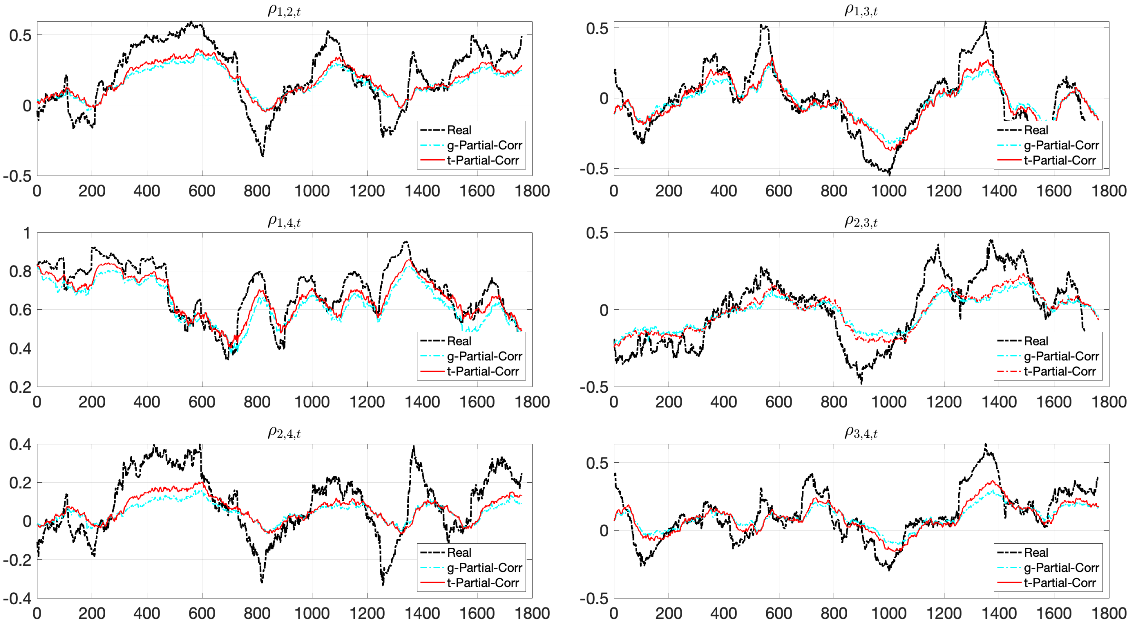
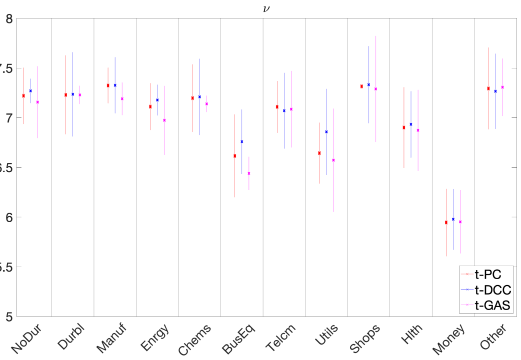
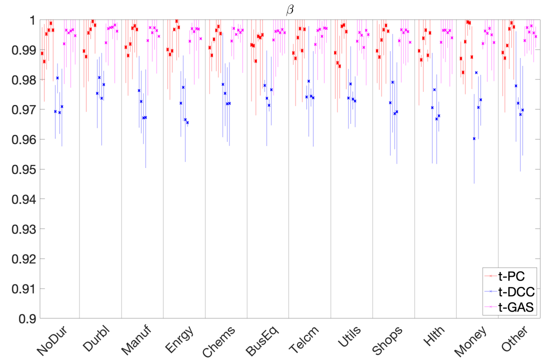


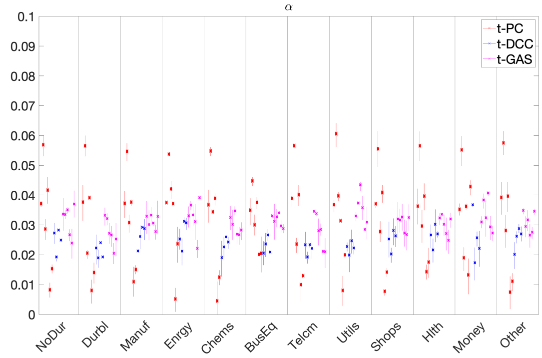
Figure C.2: Comparison of the mean of the Monte Carlo simulation of the filtered conditional correlation coefficients provided by the PCorrmodel with different distributional assumption, Gaussian and Student's t, and with Student's t DGP with $\nu = 7$.



(a) MLE estimates of ν



(b) MLE estimates of $\beta^{PCorr}_{i,j|L_{ij}}$, β^{DCC}_{i}, and $\beta^{GAS}_{i,j}$



(c) MLE estimates of $\alpha^{PCorr}_{i,j|L_{ij}}$, α^{DCC}_{i}, and $\alpha^{GAS}_{i,j}$

Figure C.3: Parameter estimates of all correlation models across industries

Note: the top-left panel for $\alpha_{i,j}$ has 12 vertical areas, each corresponding to an industry. The red left six lines in each band provide the parameter estimates (as a point) and their confidence intervals (as a line) for the PCorr model in the order of our decomposition L_{ij}, i.e., $(i,j) = (SMB, HML), (MKT, SMB), (IND, MKT), (MKT, HML \mid SMB), (IND, SMB \mid MKT), (IND, HML \mid MKT, SMB)$ which indexes along each lower sub-diagonal of $\boldsymbol{R}_t$, starting from the first sub-diagonal. The next 4 blue lines indicate the estimates and confidence intervals for the t-cDCC model, followed by the estimates of of the t-GAS model (in the same order as for the PCorr model). The estimates for the $\beta_{i,j}$ in the t-cDCC appear slightly lower than for the other models, but one should measure persistence for the t-cDCC using a composite of $\alpha_{i,i}$ and $\beta_{i,i}$ rather than $\beta_{i,j}$ alone as in the PCorr and t-GAS model. Adding $\alpha_{i,i}$ and $\beta_{i,i}$ together, the estimates for persistence are again close across all models.

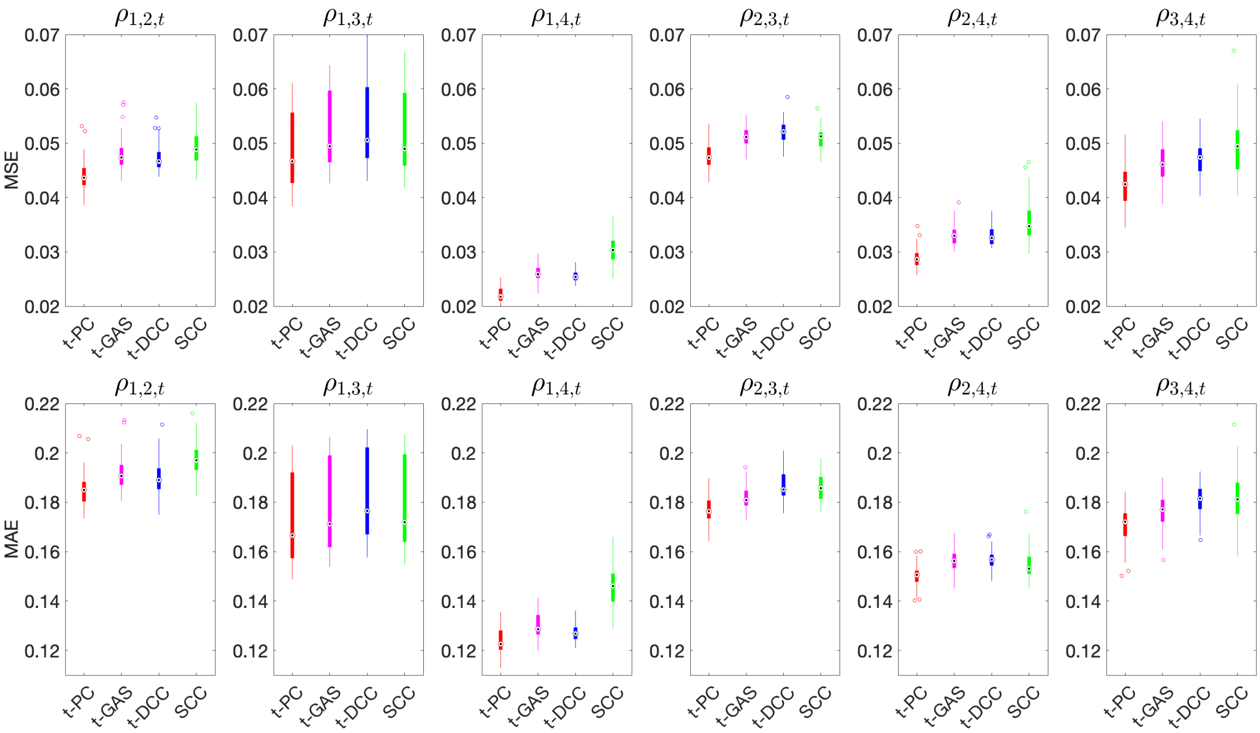


Figure C.4: Comparison of the box-plots of the block-bootstrap Monte Carlo simulation of the filtered conditional correlation coefficients with Student's t DGP with $\nu = 7$.

Table C.1: Descriptive statistics of the daily returns of the three main risk factors and the twelve US industry portfolios over the full data period, 03 January 1980 to 31 December 2021.

Series	Mean	Max	Min	Std	Skewness	Kurtosis
Mkt-RF	0.0349	15.7600	-12.0100	1.1536	0.2017	19.7633
SMB	0.0023	8.1800	-7.2700	0.6542	-0.5691	23.3088
HML	0.0157	9.0400	-6.0200	0.6796	1.0057	20.7701
NoDur	0.0341	13.9700	-9.2500	0.8607	0.2158	25.0626
Durl	0.0583	27.0600	-16.7000	1.7385	0.6772	19.3277
Manuf	0.0434	23.4000	-11.4900	1.4425	0.6376	23.1950
Enrgy	0.0420	17.2900	-9.6100	1.2444	0.4425	15.5661
Chems	0.0493	18.5000	-19.1100	1.2864	0.0229	25.6146
BusEq	0.0528	22.3900	-16.7500	1.6460	0.3410	18.1912
Telcm	0.0369	15.9800	-12.8800	0.9383	0.6995	29.5993
Utils	0.0377	17.9200	-15.2600	1.3417	0.3484	21.1585
Shops	0.0386	17.8600	-11.7900	1.1285	0.2282	21.2056
Health	0.0442	12.6200	-14.4000	1.1300	-0.1520	24.8621
Money	0.0381	19.7100	-17.2300	1.3972	0.3297	26.2926
Other	0.0331	17.5800	-11.1600	1.3642	0.3269	16.9240

Table C.2: *MSE* and *MAE* simulation results

Note: the labels PCorr, t-GAS and t-DCC indicate the new score-driven partial correlation model discussed in Section 2, the Student's t GAS model of Creal et al. (2011) with hypersphere parameterization, and the t-cDCC model of Engle (2002) with a multivariate Student's t log-likelihood, respectively. Results are based on 300 Monte Carlo experiments with sample size $T = 1000$ and $N = 4$. True correlation paths used in the data generating process are given from 100-day rolling window estimates of empirical correlation matrices of the series (HML, SMB, Mkt - RF, BusEq).

	PCorr	t-GAS	t-DCC	SCC	PCorr	t-GAS	t-DCC	SCC
	Gaussian				**Student t_7**			
$\rho_{1,2;t}$								
MSE	**0.0203**	0.0281	0.0310	0.0276	**0.0231**	0.0292	0.0367	0.0319
MAE	**0.1143**	0.1245	0.1355	0.1210	**0.1244**	0.1256	0.1450	0.1358
$\rho_{1,3;t}$								
MSE	**0.0236**	0.0305	0.0309	0.0321	**0.0260**	0.0291	0.0376	0.0325
MAE	**0.1232**	0.1337	0.1361	0.1351	**0.1283**	0.1253	0.1472	0.1326
$\rho_{1,4;t}$								
MSE	**0.0067**	0.0154	0.0170	0.0163	**0.0080**	0.0160	0.0205	0.0213
MAE	**0.0640**	0.0787	0.0858	0.0836	**0.0725**	0.0803	0.0957	0.1023
$\rho_{2,3;t}$								
MSE	**0.0186**	0.0261	0.0296	0.0288	**0.0195**	0.0259	0.0508	0.0286
MAE	**0.1072**	0.1193	0.1309	0.1291	**0.1116**	0.1146	0.1729	0.1237
$\rho_{2,4;t}$								
MSE	**0.0163**	0.0221	0.0267	0.0466	**0.0186**	0.0226	0.0304	0.0285
MAE	**0.1058**	0.1069	0.1227	0.1792	**0.1142**	**0.1142**	0.1267	0.1217
$\rho_{3,4;t}$								
MSE	**0.0186**	0.0267	0.0270	0.0393	**0.0197**	0.0257	0.0332	0.0325
MAE	**0.1072**	0.1210	0.1229	0.1547	**0.1125**	0.1138	0.1342	0.1316
$\overline{MSE}$	**0.0174**	0.0248	0.0270	0.0318	**0.0192**	0.0247	0.0349	0.0292
$\overline{MAE}$	**0.1106**	0.1140	0.1223	0.1338	**0.1106**	0.1123	0.1370	0.1246

Table C.3: *MSE* and *MAE* for a block-bootstrap simulation

Note: we take the rolling window estimated correlation paths from Table C.2 and block-bootstrapping (block length 50) these paths. For each of the 300 bootstraps, we generate return data obeying the bootstrapped correlation path. Then we estimate 8 models and filter the correlations, computing time averages of MSE and MAE, and the quartiles across bootstraps.

	PCorr	t-GAS	t-DCC	SCC	PCorr	t-GAS	t-DCC	SCC
	Gaussian				**Student t_7**			
$\rho_{1,2;t}$								
BM-MSE	**0.0511**	0.0535	0.0550	0.0539	**0.0440**	0.0480	0.0475	0.0491
BQ1-MSE	**0.0496**	0.0517	0.0532	0.0520	**0.0423**	0.0460	0.0456	0.0468
BQ3-MSE	**0.0526**	0.0547	0.0562	0.0561	**0.0454**	0.0491	0.0484	0.0513
BM-MAE	**0.1955**	0.2029	0.2014	0.1991	**0.1853**	0.1916	0.1895	0.1973
BQ1-MAE	**0.1918**	0.1986	0.1974	0.1950	**0.1803**	0.1871	0.1853	0.1930
BQ2-MAE	**0.1976**	0.2057	0.2030	0.2015	**0.1883**	0.1951	0.1937	0.2012
$\rho_{1,3;t}$								
BM-MSE	**0.0638**	0.0657	0.0673	0.0663	**0.0487**	0.0523	0.0534	0.0521
BQ1-MSE	**0.0616**	0.0636	0.0651	0.0641	**0.0427**	0.0465	0.0473	0.0459
BQ3-MSE	**0.0665**	0.0687	0.0702	0.0684	**0.0556**	0.0597	0.0603	0.0593
BM-MAE	**0.2055**	0.2121	0.2100	0.2085	**0.1740**	0.1787	0.1826	0.1796
BQ1-MAE	**0.2001**	0.2065	0.2047	0.2036	**0.1574**	0.1618	0.1670	0.1640
BQ2-MAE	**0.2124**	0.2167	0.2155	0.2137	**0.1921**	0.1990	0.2022	0.1994
$\rho_{1,4;t}$								
BM-MSE	**0.0229**	0.0255	0.0270	0.0256	**0.0218**	0.0261	0.0255	0.0303
BQ1-MSE	**0.0224**	0.0250	0.0265	0.0249	**0.0210**	0.0252	0.0248	0.0286
BQ3-MSE	**0.0236**	0.0262	0.0277	0.0264	**0.0232**	0.0271	0.0262	0.0321
BM-MAE	**0.1254**	0.1323	0.1317	0.1293	**0.1236**	0.1298	0.1273	0.1456
BQ1-MAE	**0.1236**	0.1298	0.1295	0.1271	**0.1203**	0.1265	0.1246	0.1398
BQ2-MAE	**0.1278**	0.1350	0.1342	0.1313	**0.1281**	0.1344	0.1293	0.1511
$\rho_{2,3;t}$								
BM-MSE	**0.0537**	0.0546	0.0561	0.0554	**0.0476**	0.0512	0.0520	0.0510
BQ1-MSE	**0.0492**	0.0504	0.0519	0.0517	**0.0460**	0.0499	0.0506	0.0494
BQ3-MSE	**0.0570**	0.0587	0.0602	0.0589	**0.0492**	0.0524	0.0534	0.0520
BM-MAE	**0.1956**	0.2019	0.2003	0.1992	**0.1772**	0.1820	0.1866	0.1857
BQ1-MAE	**0.1852**	0.1918	0.1899	0.1873	**0.1735**	0.1787	0.1828	0.1814
BQ2-MAE	**0.2060**	0.2122	0.2098	0.2094	**0.1807**	0.1847	0.1913	0.1902
$\rho_{2,4;t}$								
BM-MSE	**0.0330**	0.0353	0.0368	0.0498	**0.0476**	0.0512	0.0520	0.0510
BQ1-MSE	**0.0311**	0.0337	0.0352	0.0447	**0.0460**	0.0499	0.0506	0.0494
BQ3-MSE	**0.0353**	0.0370	0.0385	0.0534	**0.0492**	0.0524	0.0534	0.0520
BM-MAE	**0.1557**	0.1631	0.1614	0.1826	**0.1772**	0.1820	0.1866	0.1857
BQ1-MAE	**0.1505**	0.1588	0.1571	0.1759	**0.1735**	0.1787	0.1828	0.1814
BQ2-MAE	**0.1609**	0.1683	0.1670	0.1899	**0.1807**	0.1847	0.1913	0.1902
$\rho_{3,4;t}$								
BM-MSE	**0.0318**	0.0339	0.0354	0.0369	**0.0424**	0.0464	0.0471	0.0499
BQ1-MSE	**0.0306**	0.0327	0.0342	0.0345	**0.0394**	0.0439	0.0449	0.0452
BQ3-MSE	**0.0330**	0.0350	0.0365	0.0385	**0.0447**	0.0489	0.0490	0.0524
BM-MAE	**0.1420**	0.1491	0.1475	0.1442	**0.1704**	0.1762	0.1810	0.1815
BQ1-MAE	**0.1363**	0.1443	0.1420	0.1390	**0.1663**	0.1721	0.1772	0.1752
BQ2-MAE	**0.1480**	0.1534	0.1528	0.1502	**0.1755**	0.1810	0.1854	0.1880

Table C.4: Out-of-sample results of the four correlation models

This table contains the estimates of a_1^{Mod} for $Mod \in \{PCorr, t-GAS, t-cDCC\}$ in the regression model $r_{i,t} = a_0^{Mod} + a_1^{Mod} \cdot \hat{r}_{i,t}^{Mod} + u_{i,t}$, where $\hat{r}_{i,t}^{Mod}$ is obtained (recursively) using one-year-ahead estimates of $\boldsymbol{R}_t$ and $\gamma_{MKT,t}$, $\gamma_{SMB,t}$, and $\gamma_{HML,t}$ as in (18). A $*$, $**$, or $***$ indicates rejection of $H_0 : a_0^{Mod} = 0, a_1^{Mod} = 1$, at the 10%, 5%, and 1% significance level, respectively. The MCS column indicates whether the model lies in the 95% model confidence set of Hansen et al. (2011) based on tracking error MSE. Results are similar for the 99% MCS.

	PCorr		t-GAS		t-cDCC		SCC	
	$\hat{a}_1^{PCorr}$	MCS	$\hat{a}_1^{t-GAS}$	MCS	$\hat{a}_1^{t-cDCC}$	MCS	$\hat{a}_1^{SCC}$	MCS
NoDur	1.013	✓	0.987		0.966 ***		0.985 **	
	(0.013)		(0.013)		(0.013)		(0.013)	
Durbl	1.018	✓	0.956 **		0.984		0.962 ***	
	(0.013)		(0.013)		(0.012)		(0.013)	
Manuf	1.012	✓	1.001		0.975 ***		0.998	
	(0.007)		(0.007)		(0.007)		(0.007)	
Enrgy	1.053 **	✓	1.005		0.967 ***		0.974 ***	
	(0.023)		(0.015)		(0.013)		(0.014)	
Chems	1.002	✓	0.981		0.965 ***		0.966 ***	
	(0.011)		(0.012)		(0.011)		(0.011)	
BusEq	0.913 ***	✓	0.886 ***		0.861 ***		0.874 ***	
	(0.006)		(0.007)		(0.006)		(0.007)	
Telcm	1.000	✓	0.975		0.945 ***		0.959 ***	
	(0.014)		(0.014)		(0.013)		(0.014)	
Utils	1.050	✓	0.957		0.990	✓	0.959 **	
	(0.032)		(0.023)		(0.020)		(0.023)	
Shops	0.987	✓	0.983		0.946 ***		0.969 *	
	(0.009)		(0.009)		(0.009)		(0.009)	
Health	1.009	✓	0.996 *		0.958 ***		0.990 *	
	(0.011)		(0.012)		(0.010)		(0.011)	
Money	0.986 *	✓	0.982 **	✓	0.928 ***		0.965 ***	
	(0.006)		(0.006)		(0.007)		(0.007)	
Other	1.011	✓	1.010 *		0.974 ***		1.004	
	(0.006)		(0.006)		(0.006)		(0.006)	

Table C.5: In-sample performance in 23-dimensional application

In the top-part of the table, we report the Diebold-Mariano statistics and MCS based on the predictive log-likelihood for the full 23-variate system. In the bottom part of the table, we report the asset pricing implications for each of the individual stocks based on the tracking errors. Here, the Diebold-Mariano t statistics are reported based on the tracking error MSE and MAE from equations (17) – (18). The MCS columns indicate whether the model is selected for the 95% model confidence set based on MSE. The PCorr, t-cDCC, and SCC model contain 760/47/1265 parameters, respectively, and take 2116/1019/2309 seconds to estimate. The sample is 1992–2022.

	PCorr-Full vs PCorr-EbE		PCorr-Full vs t-cDCC		PCorr-Full vs SCC		MCS			
							PCorr-Full	PCorr-EbE	t-cDCC	SCC
	DM_{PLL}		DM_{PLL}		DM_{PLL}					
ALL	10.48***		5.97***		53.22***		✓			
	DM_{MSE}	DM_{MAE}	DM_{MSE}	DM_{MAE}	DM_{MSE}	DM_{MAE}				
APPL	-4.43***	-3.72***	-3.00***	-3.35***	-3.92***	-3.61***	✓			
AXP	-1.51*	-0.80	-1.58*	-1.32*	-2.63***	-3.64***	✓	✓		
BA	0.23	1.68**	-2.02***	-0.38	-2.21***	-1.68**	✓	✓		
CAT	0.04	1.22	-3.73***	-4.41***	-4.06***	-4.82***	✓	✓		
CSCO	-1.17	0.45	-3.00***	-3.04***	-3.30***	-3.11***	✓	✓		
DOW	0.49	-1.15	-2.29***	-2.15***	-2.45***	-2.34***	✓	✓		
HD	-2.79***	-1.84**	-1.11	-3.46***	-1.82**	-4.08***	✓			
IBM	-0.82	-0.03	-4.12***	-4.33***	-3.91***	-4.31***	✓	✓		
INTC	-0.59	0.78	-5.97***	-6.15***	-3.47***	-4.67***	✓	✓		
JNJ	0.52	0.55	-1.94**	-1.92**	-3.57***	-3.76***	✓	✓		
JPM	-1.05	-1.25	-1.72**	-1.49*	-0.07	-0.40	✓	✓		✓
KO	-1.65**	-1.20	-1.27*	-2.16***	-2.30***	-2.79***	✓	✓		
MCD	-0.37	-0.64	-0.94	-0.64	-1.39*	-1.24*	✓	✓		
MMM	-1.32*	-0.43	-1.78**	-2.54***	-3.10***	-3.65***	✓	✓		
MRK	-0.46	0.29	-3.74***	-4.92***	-4.20***	-4.74***	✓	✓		
PFE	-1.43*	-1.23*	-1.91**	-2.35***	-1.45*	-2.49***	✓			
PG	-2.30***	-1.56*	-0.40	-0.84	-3.26***	-4.18***	✓			
UTX	-1.89**	-2.28***	-1.88***	-2.63***	-0.76	-0.44	✓			✓
V	1.01	-1.37	-3.49***	-3.75***	-1.31*	-2.18***	✓	✓		✓
WMT	0.51	0.86	-0.62	-1.16	-1.01	-1.25	✓	✓	✓	✓

D Additional simulation results

Table D.1: *MSE* and *MAE* simulation results

Note: the labels PCorr, t-GAS and t-DCC indicate the new score-driven partial correlation model discussed in Section 2, the Student's t GAS model of Creal et al. (2011) with hypersphere parameterization, and the t-cDCC model of Engle (2002) with a multivariate Student's t log-likelihood, respectively. Results are based on 300 Monte Carlo (top panel) or bootstrap (bottom panel) experiments with sample size $T = 2000$ and $N = 4$. The true correlation path used in the data generating process is given by a high constant equicorrelation matrix regime with $\rho = 0.9$ that shifts to a lower constant equicorrelation regime with $\rho = 0.4$. Results are averaged over time, replications, and across pairs (i, j), $1 \leq i < j \leq 4$.

	PCorr	t-GAS	t-DCC	SCC	PCorr	t-GAS	t-DCC	SCC
	Gaussian				**Student t_7**			
	Regime switching constant equicorrelation path							
$\overline{MSE}$	**0.0048**	0.0053	0.0074	0.0140	**0.0074**	0.0085	0.0093	0.0132
$\overline{MAE}$	**0.0411**	0.0428	0.0555	0.0850	**0.0633**	0.0670	0.0685	0.0865

References

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society 12*(5), 788–792.

Blasques, F., A. Lucas, and E. Silde (2018). A stochastic recurrence equations approach for score driven correlation models. *Econometric Reviews 37*(2), 166–181.

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics 227*(2), 325–346.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews 11*(2), 143–172.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*(4), 552–563.

Darolles, S., C. Francq, and S. Laurent (2018). Asymptotics of cholesky garch models and time-varying conditional betas. *Journal of Econometrics 204*(2), 223–247.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Hafner, C. M. and A. Preminger (2009). Asymptotic theory for a factor garch model. *Econometric Theory 25*(2), 336–363.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Magnus, J. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics 33*(2), 659–680.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Vaart, A. W. v. d. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

White, H. (1994). *Estimation, Inference and Specification Analysis.* Cambridge University Press.