Online Appendix to: Closed-Form Multi-Factor Copula Models with Observation-Driven Dynamic Factor Loadings

Anne Opschoor[a,b], André Lucas[a,b], István Barra[a,b], Dick van Dijk[c,b]

[a] *Vrije Universiteit Amsterdam*

[b] *Tinbergen Institute*

[c] *Erasmus University Rotterdam*

Contents

A Derivations of the score

A.1 General set-up

The general set-up of the (multi) factor copulas is given by equation (2). We are interested in the score s_t, defined as

$$\boldsymbol{s}_t = \partial \log \boldsymbol{c}_t(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C)/\partial \boldsymbol{f}_t \tag{A.1}$$

where $\boldsymbol{f}_t$ holds all unique dynamic factor loadings, and with a slight abuse of notation $\boldsymbol{c}_t(\cdot)$ is related to the conditional copula density. Note that the dimension of $\boldsymbol{f}_t$ (and hence $\boldsymbol{s}_t$) depends on the chosen factor structure.

We consider a Student's t and a Gaussian copula density for $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{N,t})^\top = (T_\nu^{-1}(u_{1,t}), \ldots, T_\nu^{-1}(u_{N,t}))^\top$ for the vector of PITs $(u_{1,t}, \ldots, u_{N,t})^\top$, with $T_\nu^{-1}(\,\cdot\,)$ the inverse Student's t cdf with ν degrees of freedom, zero mean, and unit variance, where $\nu \to \infty$ for the Gaussian case. We have the following specifications:

$$\begin{aligned}
\log \boldsymbol{c}_{Stud,t}(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C) &= -\frac{1}{2}\log|\boldsymbol{R}_t| - \frac{\nu+N}{2}\log\left(1 + \frac{\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1}\boldsymbol{x}_t}{\nu-2}\right) + a_{Stud}(\nu, \boldsymbol{x}_t), && \text{(A.2)}\\
\log \boldsymbol{c}_{Gauss,t}(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C) &= -\frac{1}{2}\log|\boldsymbol{R}_t| + -\frac{1}{2}\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1}\boldsymbol{x}_t + a_{Gaus}, && \text{(A.3)}
\end{aligned}$$

where $a_{Stud}(\nu, \boldsymbol{x}_t)$ and a_{Gaus} are constants that do not depend on $\boldsymbol{R}_t$. Further, the dependence matrix $\boldsymbol{R}_t$ is modeled as

$$\boldsymbol{R}_t = \tilde{\boldsymbol{L}}_t^\top \tilde{\boldsymbol{L}}_t + \boldsymbol{D}_t, \qquad \tilde{\boldsymbol{L}}_t = \left(\tilde{\boldsymbol{\lambda}}_{1,t}, \ldots, \tilde{\boldsymbol{\lambda}}_{N,t}\right), \qquad \boldsymbol{D}_t = \operatorname{diag}\left(\sigma_{1,t}^2, \ldots, \sigma_{N,t}^2\right), \tag{A.4}$$

with

$$\tilde{\boldsymbol{\lambda}}_{i,t} = \frac{\boldsymbol{\lambda}_{i,t}}{\sqrt{1 + \boldsymbol{\lambda}_{i,t}^\top \boldsymbol{\lambda}_{i,t}}} = \boldsymbol{\lambda}_{i,t} \cdot \sigma_{it}, \qquad \sigma_{it}^2 = \frac{1}{1 + \boldsymbol{\lambda}_{i,t}^\top \boldsymbol{\lambda}_{i,t}} \tag{A.5}$$

for a vector $\boldsymbol{\lambda}_{i,t} \in \mathbb{R}^{k\times 1}$. This ensures that $x_{i,t}$ has unit variance by design.

Define $\boldsymbol{L}_t = (\boldsymbol{\lambda}_{1,t}, \ldots, \boldsymbol{\lambda}_{N,t}) \in \mathbb{R}^{k\times N}$, then $\boldsymbol{f}_t$ contains the unique factor loadings of $\boldsymbol{L}_t$. For example, for the 1-equi-factor specification, there is only one time-varying parameter, such that $\boldsymbol{f}_t$ is scalar, and $\boldsymbol{\lambda}_{i,t} = \boldsymbol{f}_t$ for all $i = 1, \ldots, N$. Also for other models, the dimension of $\boldsymbol{f}_t$ is typically much smaller than that of $\operatorname{vec}(\boldsymbol{L}_t)$ due to the factor structure and the group allocation. Using the chain rule we obtain

$$\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C)}{\partial \boldsymbol{f}_t^\top} = \frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top} \cdot \frac{\partial \operatorname{vec}(\boldsymbol{R}_t)}{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top} \cdot \frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top}. \tag{A.6}$$

The first two components in (A.6) are generic for any factor structure. The last component, by contrast, strongly depends on the factor and group structure and will be dealt with in separate subsections. A further component might be added to (A.6) in case some elements of $\boldsymbol{f}_t$ are restricted to be positive, or lie in some range. This can for instance be obtained by specifying that element of $\boldsymbol{f}_t$ as the exponential function of a new, unrestricted time varying parameter, and by taking the derivative with respect to this new parameter. The derivative of this last type of transformation can be added as a final chain rule term in (A.6) and will typically take the form of a simple, diagonal matrix.

The first component in (A.6) only depends on the conditional copula density specification. For the Student's t case, we obtain

$$\begin{aligned}
\mathrm{d} \log \boldsymbol{c}_{Stud,t}(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C) &= -\frac{1}{2} \operatorname{tr}\left(\boldsymbol{R}_t^{-1}\, \mathrm{d}\, \boldsymbol{R}_t\right) - \frac{\nu+N}{2} \frac{1}{1+\frac{\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{x}_t}{\nu-2}} \mathrm{d}\left(\frac{\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{x}_t}{\nu-2}\right) \\
&= -\frac{1}{2}\left(\operatorname{vec}(\boldsymbol{R}_t)^{-1}\right)^\top \mathrm{d} \operatorname{vec}(\boldsymbol{R}_t) + \frac{1}{2}\left(\frac{\nu+N}{\nu-2+\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{x}_t}\right) \boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} (\mathrm{d}\, \boldsymbol{R}_t) \boldsymbol{R}_t^{-1} \boldsymbol{x}_t \\
&= -\frac{1}{2}\left(\operatorname{vec}(\boldsymbol{R}_t)^{-1}\right)^\top \mathrm{d} \operatorname{vec}(\boldsymbol{R}_t) + \frac{1}{2}\left(\frac{\nu+N}{\nu-2+\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{x}_t} \boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} \otimes \boldsymbol{R}_t^{-1} \boldsymbol{x}_t\right)^\top \mathrm{d} \operatorname{vec}(\boldsymbol{R}_t) \\
&= \left(-\frac{1}{2}\left(\operatorname{vec}(\boldsymbol{R}_t)^{-1}\right)^\top + \frac{1}{2}\left(\frac{\nu+N}{\nu-2+\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{x}_t} \operatorname{vec}\left(\boldsymbol{R}_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1}\right)\right)^\top\right) \mathrm{d} \operatorname{vec}(\boldsymbol{R}_t),
\end{aligned} \tag{A.7}$$

and hence

$$\frac{\partial \log \boldsymbol{c}_{Stud,t}(\boldsymbol{x}_t; \boldsymbol{R}_t, \nu_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top} = -\frac{1}{2}\left(\operatorname{vec}(\boldsymbol{R}_t)^{-1}\right)^\top + \frac{1}{2}\left(\frac{\nu+N}{\nu-2+\boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{x}_t} \operatorname{vec}\left(\boldsymbol{R}_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1}\right)\right)^\top. \tag{A.8}$$

For the Gaussian case, we let $\nu_C \to \infty$ and obtain

$$\mathrm{d} \log \boldsymbol{c}_{Gaus,t}(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C) = \left(-\frac{1}{2}\left(\operatorname{vec}(\boldsymbol{R}_t)^{-1}\right)^\top + \frac{1}{2} \operatorname{vec}\left(\boldsymbol{R}_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t^\top \boldsymbol{R}_t^{-1}\right)^\top\right) \mathrm{d} \operatorname{vec}(\boldsymbol{R}_t), \tag{A.9}$$

such that

$$\frac{\partial \log \boldsymbol{c}_{Gaus,t}(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial\,\mathrm{vec}(\boldsymbol{R}_t)^\top} = -\frac{1}{2}\left(\mathrm{vec}(\boldsymbol{R}_t)^{-1}\right)^\top + \frac{1}{2}\,\mathrm{vec}\left(\boldsymbol{R}_t^{-1}\boldsymbol{x}_t\boldsymbol{x}_t^\top\boldsymbol{R}_t^{-1}\right)^\top. \tag{A.10}$$

For the second term in (A.6) we obtain

$$\begin{aligned}
\mathrm{d}\,\mathrm{vec}(\boldsymbol{R}_t) &= \mathrm{d}\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t^\top\tilde{\boldsymbol{L}}_t + \boldsymbol{D}_t)\\
&= \left(\tilde{\boldsymbol{L}}_t^\top \otimes \mathbf{I}_N\right)\mathrm{d}\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t^\top) + \left(\mathbf{I}_N \otimes \tilde{\boldsymbol{L}}_t^\top\right)\mathrm{d}\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t) + \mathrm{d}\,\mathrm{vec}(\boldsymbol{D}_t)\\
&= \left(\tilde{\boldsymbol{L}}_t^\top \otimes \mathbf{I}_N\right)K_{k,N}\,\mathrm{d}\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t) + \left(\mathbf{I}_N \otimes \tilde{\boldsymbol{L}}_t^\top\right)\mathrm{d}\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t) + \mathrm{d}\,\mathrm{vec}(\boldsymbol{D}_t)\\
&= \left(\left(\tilde{\boldsymbol{L}}_t^\top \otimes \mathbf{I}_N\right)K_{k,N} + \left(\mathbf{I}_N \otimes \tilde{\boldsymbol{L}}_t^\top\right)\right)\mathrm{d}\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t) + \mathrm{d}\,\mathrm{vec}(\boldsymbol{D}_t),
\end{aligned} \tag{A.11}$$

where $K_{k,N}$ is the commutation matrix, i.e., $\mathrm{vec}(A^\top) = K_{k,N}\,\mathrm{vec}(A)$ for a general $k \times N$ matrix A. We write $K_N \equiv K_{N,N}$. As a result, we obtain

$$\frac{\partial\,\mathrm{vec}(\boldsymbol{R}_t)}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top} = \left(\left(\tilde{\boldsymbol{L}}_t^\top \otimes \mathbf{I}_N\right)K_{k,N} + \left(\mathbf{I}_N \otimes \tilde{\boldsymbol{L}}_t^\top\right)\right)\cdot\frac{\partial\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t)}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top} + \frac{\partial\,\mathrm{vec}(\boldsymbol{D}_t)}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top}. \tag{A.12}$$

Let S^D be an $N^2 \times N$ selection matrix, such that for a diagonal $N \times N$ matrix A with the $N \times 1$ vector a on the diagonal we have $\mathrm{vec}(A) = S^D \cdot a$. Then

$$\frac{\partial\,\mathrm{vec}(\boldsymbol{D}_t)}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top} = S^D\;\frac{\partial \mathrm{diag}(\boldsymbol{D}_t)}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top} = -2\,S^D\;D^2\begin{pmatrix} \boldsymbol{\lambda}_{1,t}^\top & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda}_{2,t}^\top & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\lambda}_{N,t}^\top \end{pmatrix}, \tag{A.13}$$

where $\mathrm{diag}(A) \in \mathbb{R}^{N\times 1}$ holds the diagonal elements of the $N \times N$ matrix A. Similarly, we obtain

$$\frac{\partial\,\mathrm{vec}(\tilde{\boldsymbol{L}}_t)}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top} = \begin{pmatrix} \boldsymbol{Q}_{1,t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{Q}_{N,t} \end{pmatrix}, \qquad \boldsymbol{Q}_{i,t} = \frac{\mathbf{I}_k}{(1+\boldsymbol{\lambda}_{i,t}^\top\boldsymbol{\lambda}_{i,t})^{1/2}} - \frac{\boldsymbol{\lambda}_{i,t}\boldsymbol{\lambda}_{i,t}^\top}{(1+\boldsymbol{\lambda}_{i,t}^\top\boldsymbol{\lambda}_{i,t})^{3/2}}, \tag{A.14}$$

for $i = 1, \ldots, N$. Note due to the special structure in (A.14), we have that

$$\left(\mathbf{I}_N \otimes \tilde{\boldsymbol{L}}_t^\top\right) \frac{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)}{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top} = \begin{pmatrix} \tilde{\boldsymbol{L}}_t^\top \boldsymbol{Q}_{1,t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{\boldsymbol{L}}_t^\top \boldsymbol{Q}_{N,t} \end{pmatrix}. \tag{A.15}$$

We now turn to the last component in (A.6) for the factor models considered in this paper.

A.2 1-Equi-Factor

In the 1-Factor equi-copula, we have $\boldsymbol{L}_t = \boldsymbol{\lambda}_t \boldsymbol{\iota}_N^\top$ and $\boldsymbol{\lambda}_t = \boldsymbol{f}_t = f_t \in \mathbb{R}$. We then have

$$\frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t} = \boldsymbol{\iota}_N. \tag{A.16}$$

The final score $\boldsymbol{s}_t$ is now obtained by combining (A.16), (A.6), (A.8) or (A.10), and (A.12)–(A.15).

A.3 1-Factor model with heterogeneous loadings

In this case we have $\boldsymbol{f}_t = (f_{t,1}, f_{t,2} \ldots, f_{t,G})^\top \in \mathbb{R}^{G \times 1}$ and $\boldsymbol{L}_t = \boldsymbol{f}_t^\top \left(S_1^{gr}\right)^\top$ with

$$S_1^{gr} = \begin{pmatrix} \boldsymbol{\iota}_{N_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\iota}_{N_G} \end{pmatrix} \in \mathbb{R}^{N \times G}, \tag{A.17}$$

where N_g for $g = 1, \ldots, G$ is the number of firms in group g. We then have

$$\frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top} = \frac{\partial \operatorname{vec}\left(\boldsymbol{f}_t^\top \left(S_1^{gr}\right)^\top\right)}{\partial \boldsymbol{f}_t^\top} = S_t^{gr} \frac{\partial \operatorname{vec}(\boldsymbol{f}_t^\top)}{\partial \boldsymbol{f}_t^\top} = S_t^{gr} \frac{\partial \operatorname{vec}(\boldsymbol{f}_t)}{\partial \boldsymbol{f}_t^\top} = S_t^{gr}. \tag{A.18}$$

A.4 2-factor model

The 2F model consists of an equi-loading vector, and a set of heterogeneous loadings. In this case we have $\boldsymbol{f}_t = (f_{t,0}, f_{t,1}, f_{t,2}, \ldots, f_{t,G})^\top \in \mathbb{R}^{(G+1)\times 1}$. Let $\delta_{i,j}$ be the kronecker delta, i.e.,

$\delta_{i,j} = 1$ if $i = j$ and zero otherwise. Also define

$$S_i^{2f} = \begin{pmatrix} \delta_{0,i}\boldsymbol{\iota}_{N_1} & \delta_{1,i}\boldsymbol{\iota}_{N_1} \\ \vdots & \vdots \\ \delta_{0,i}\boldsymbol{\iota}_{N_G} & \delta_{G,i}\boldsymbol{\iota}_{N_G} \end{pmatrix} \in \mathbb{R}^{N\times 2}. \tag{A.19}$$

Then

$$\boldsymbol{L}_t = \sum_{i=0}^{G} f_{t,i} \cdot (S_i^{2f})^\top, \tag{A.20}$$

and

$$\frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top} = \Big(\operatorname{vec}\big((S_0^{2f})^\top\big) \ , \ \operatorname{vec}\big((S_1^{2f})^\top\big) \ , \ \ldots \ , \ \operatorname{vec}\big((S_G^{2f})^\top\big) \Big) \in \mathbb{R}^{2N\times(G+1)}. \tag{A.21}$$

A.5 MF model

The MF model consists of two types of factors: an equi-factor, and G industry factors, each with a group-specific loading. In this case we have $\boldsymbol{f}_t = (f_{t,0}, f_{t,1}, f_{t,2}, \ldots, f_{t,G})^\top \in \mathbb{R}^{(G+1)\times 1}$. Define

$$S_i^{mf} = \begin{pmatrix} \delta_{0,i}\boldsymbol{\iota}_{N_1} & \delta_{1,i}\boldsymbol{\iota}_{N_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{0,i}\boldsymbol{\iota}_{N_G} & 0 & \cdots & \delta_{G,i}\boldsymbol{\iota}_{N_G} \end{pmatrix} \in \mathbb{R}^{N\times(G+1)}. \tag{A.22}$$

Then

$$\boldsymbol{L}_t = \sum_{i=0}^{G} f_{t,i} \cdot (S_i^{mf})^\top, \tag{A.23}$$

and

$$\frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top} = \Big(\operatorname{vec}\big((S_0^{mf})^\top\big) \ , \ \operatorname{vec}\big((S_1^{mf})^\top\big) \ , \ \ldots \ , \ \operatorname{vec}\big((S_G^{mf})^\top\big) \Big) \in \mathbb{R}^{(G+1)N\times(G+1)}. \tag{A.24}$$

A.6 MF LT model

For the MF LT model, we have $\boldsymbol{f}_t = (f_{t,1}, f_{t,2}, \ldots, f_{t,G(G+1)/2})^\top \in \mathbb{R}^{G(G+1)/2\times 1}$. Also define

$$S_i^{lt} = \begin{pmatrix} \delta_{1,i}\boldsymbol{\iota}_{N_1} & 0 & \cdots & 0 \\ \delta_{2,i}\boldsymbol{\iota}_{N_2} & \delta_{G+1,i}\boldsymbol{\iota}_{N_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{G,i}\boldsymbol{\iota}_{N_G} & \delta_{2G-1,i}\boldsymbol{\iota}_{N_G} & \cdots & \delta_{\frac{1}{2}G(G+1),i}\boldsymbol{\iota}_{N_G} \end{pmatrix} \in \mathbb{R}^{N\times G}. \tag{A.25}$$

Then

$$\boldsymbol{L}_t = \sum_{i=1}^{G(G+1)/2} f_{t,i} \cdot (S_i^{lt})^\top, \tag{A.26}$$

and

$$\frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top} = \Big(\operatorname{vec}\big((S_1^{lt})^\top\big) \;,\; \operatorname{vec}\big((S_2^{lt})^\top\big) \;,\; \ldots \;,\; \operatorname{vec}\big((S_{\frac{1}{2}G(G+1)}^{lt})^\top\big) \Big) \in \mathbb{R}^{G\,N\times\frac{1}{2}G(G+1)}. \tag{A.27}$$

A.7 Information matrix derivations

In our paper, we use unit scaling for the score. If one prefers to scale by a power of the inverse conditional Fisher information matrix, one can use the following results.

$$\mathcal{I}_{t-1} = \mathbb{E}_{t-1}\left[\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial \boldsymbol{f}_t}\,\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial \boldsymbol{f}_t^\top}\right]$$

$$= \frac{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top}{\partial \boldsymbol{f}_t}\cdot\frac{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top}{\partial \operatorname{vec}(\boldsymbol{L}_t)}\cdot\mathbb{E}_{t-1}\left[\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)}\,\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top}\right] \times$$
$$\times \frac{\partial \operatorname{vec}(\boldsymbol{R}_t)}{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top}\cdot\frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top}.$$

The first and last two factors in this expression are known from our score derivations. The only remaining component is the expectation in the middle. This can be computed using a similar approach as in Creal et al. (2011). Let $\tilde{\boldsymbol{x}}_t$ have a Student's t distribution with ν degrees of freedom, zero mean, and unit covariance matrix. Then

$$
\begin{aligned}
&\mathbb{E}_{t-1}\left[\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial\,\mathrm{vec}(\boldsymbol{R}_t)}\ \frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial\,\mathrm{vec}(\boldsymbol{R}_t)^\top}\right] \quad = \\
&\frac{1}{4}\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right) \quad \times \\
&\qquad \mathbb{E}_{t-1}\left[\frac{\left(\frac{\nu+N}{\nu-2}\right)^2}{(1+\frac{\boldsymbol{x}_t^\top\boldsymbol{R}_t^{-1}\boldsymbol{x}_t}{\nu-2})^2}\,\mathrm{vec}\left(\boldsymbol{R}_t^{-1/2}\boldsymbol{x}_t\boldsymbol{x}_t^\top\boldsymbol{R}_t^{-1/2}\right)\mathrm{vec}\left(\boldsymbol{R}_t^{-1/2}\boldsymbol{x}_t\boldsymbol{x}_t^\top\boldsymbol{R}_t^{-1/2}\right)^\top - \mathrm{vec}\left(\mathbf{I}_N\right)\mathrm{vec}\left(\mathbf{I}_N\right)^\top\right] \\
&\qquad \times \quad \left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)^\top \qquad = \\
&\frac{1}{4}\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)\times\mathbb{E}_{t-1}\left[\frac{\left(\frac{\nu+N}{\nu-2}\right)^2}{(1+\frac{\tilde{\boldsymbol{x}}_t^\top\tilde{\boldsymbol{x}}_t}{\nu-2})^2}\,\mathrm{vec}\left(\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top\right)\mathrm{vec}\left(\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top\right)^\top - \mathrm{vec}\left(\mathbf{I}_N\right)\mathrm{vec}\left(\mathbf{I}_N\right)^\top\right] \\
&\qquad \times \quad \left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)^\top \qquad = \\
&\frac{1}{4}\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)\mathbb{E}_{t-1}\left[\frac{\left(\frac{\nu+N}{\nu-2}\right)^2\left(\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top\otimes\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top\right)}{(1+\frac{\tilde{\boldsymbol{x}}_t^\top\tilde{\boldsymbol{x}}_t}{\nu-2})^2} - \mathrm{vec}\left(\mathbf{I}_N\right)\mathrm{vec}\left(\mathbf{I}_N\right)^\top\right]\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)^\top = \\
&\frac{1}{4}\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)\mathbb{E}_{t-1}\left[\frac{\left(\frac{\nu+N}{\nu-2}\right)^2\left(\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top\otimes\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top\right)}{(1+\frac{\tilde{\boldsymbol{x}}_t^\top\tilde{\boldsymbol{x}}_t}{\nu-2})^2}\right]\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)^\top \\
&\qquad -\frac{1}{4}\,\mathrm{vec}\left(\boldsymbol{R}_t^{-1}\right)\mathrm{vec}\left(\boldsymbol{R}_t^{-1}\right)^\top = \\
&\frac{1}{4}\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)\mathcal{G}^\star\left(\boldsymbol{R}_t^{-1/2}\otimes\boldsymbol{R}_t^{-1/2}\right)^\top - \frac{1}{4}\,\mathrm{vec}\left(\boldsymbol{R}_t^{-1}\right)\mathrm{vec}\left(\boldsymbol{R}_t^{-1}\right)^\top .
\end{aligned}
$$

We index $\mathcal{G}^\star$ as (i,j,k,ℓ) according to the $\tilde{x}_{i,t}\tilde{x}_{j,t}\otimes\tilde{x}_{k,t}\tilde{x}_{\ell,t}$ element of $\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top\otimes\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t^\top$. Using our Kronecker delta notation again, we have

$$
\mathcal{G}^\star_{i,j,k,\ell} = \frac{(\nu+N)}{(\nu+2+N)}\cdot\left(\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k}\right)
$$

To see this, we first note

$$
\begin{aligned}
&\mathbb{E}_{t-1}\left[\frac{\left(\frac{\nu+N}{\nu-2}\right)^2\left(\tilde{x}_{i,t}\tilde{x}_{j,t}\tilde{x}_{k,t}\tilde{x}_{\ell,t}\right)}{(1+\frac{\tilde{\boldsymbol{x}}_t^\top\tilde{\boldsymbol{x}}_t}{\nu-2})^2}\right] = \\
&\quad \int \frac{\left(\frac{\nu+N}{\nu-2}\right)^2\left(\tilde{x}_{i,t}\tilde{x}_{j,t}\tilde{x}_{k,t}\tilde{x}_{\ell,t}\right)}{\left(1+\frac{\tilde{\boldsymbol{x}}_t^\top\tilde{\boldsymbol{x}}_t}{\nu-2}\right)^2}\cdot\frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left((\nu-2)\pi\right)^{N/2}}\frac{1}{\left(1+\frac{\tilde{\boldsymbol{x}}_t^\top\tilde{\boldsymbol{x}}_t}{\nu-2}\right)^{(\nu+N)/2}}\,\mathrm{d}\,\tilde{\boldsymbol{x}}_t = \\
&\quad \frac{\left(\frac{\nu+N}{\nu-2}\right)^2\Gamma\left(\frac{\nu+N}{2}\right)\Gamma\left(\frac{\nu+4}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu+4+N}{2}\right)}\int\frac{\Gamma\left(\frac{\nu+4+N}{2}\right)}{\Gamma\left(\frac{\nu+4}{2}\right)\left((\nu-2)\pi\right)^{N/2}}\ \frac{\left(\tilde{x}_{i,t}\tilde{x}_{j,t}\tilde{x}_{k,t}\tilde{x}_{\ell,t}\right)}{\left(1+\frac{\tilde{\boldsymbol{x}}_t^\top\tilde{\boldsymbol{x}}_t}{\nu-2}\right)^{(\nu+4+N)/2}}\,\mathrm{d}\,\tilde{\boldsymbol{x}}_t. \qquad \text{(A.28)}
\end{aligned}
$$

The latter integral gives the 4th order and 2nd order cross moments of a Student's t random variable with $\nu+4$ degrees of freedom and scaling matrix $(\nu-2)(\nu+4)^{-1}\mathbf{I}_N$. Using the construction of a Student's t random variable as the ratio of a vector normal with mean zero by a square root of a $\chi^2_{\nu+4}/(\nu+4)$ random variable, we obtain

$$
\begin{aligned}
&\int \frac{\Gamma\left(\frac{\nu+4+N}{2}\right)}{\Gamma\left(\frac{\nu+4}{2}\right)((\nu-2)\pi)^{N/2}} \frac{(\tilde{x}_{i,t}\tilde{x}_{j,t}\tilde{x}_{k,t}\tilde{x}_{\ell,t})}{\left(1+\frac{\tilde{\boldsymbol{x}}_t^\top \tilde{\boldsymbol{x}}_t}{\nu-2}\right)^{(\nu+4+N)/2}} \,\mathrm{d}\,\tilde{\boldsymbol{x}}_t \\
&\quad = (\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k}) \frac{(\nu-2)^2}{(\nu+4)^2} \times \int \frac{(\nu+4)^2 z^{-2}}{\Gamma\left(\frac{\nu+4}{2}\right)2^{(\nu+4)/2}} z^{0.5(\nu+4)-1}\exp(-z/2)\,\mathrm{d}\,z \\
&\quad = (\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k}) \frac{(\nu-2)^2\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}}{\Gamma\left(\frac{\nu+4}{2}\right)2^{(\nu+4)/2}} \times \int \frac{z^{0.5\nu-1}\exp(-z/2)}{\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}}\,\mathrm{d}\,z \\
&\quad = (\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k}) \frac{(\nu-2)^2\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}}{\Gamma\left(\frac{\nu+4}{2}\right)2^{(\nu+4)/2}} \\
&\quad = (\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k}) \frac{(\nu-2)^2}{(\nu+2)\nu}.
\end{aligned} \tag{A.29}
$$

Combining (A.28) and (A.29), we obtain

$$
\begin{aligned}
\mathcal{G}^\star_{i,j,k,\ell} &= \frac{\left(\frac{\nu+N}{\nu-2}\right)^2 \Gamma\left(\frac{\nu+N}{2}\right)\Gamma\left(\frac{\nu+4}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu+4+N}{2}\right)} \cdot \frac{(\nu-2)^2}{(\nu+2)\nu} \cdot (\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k}) \\
&= \frac{\left(\frac{\nu+N}{\nu-2}\right)^2 (\nu+2)\nu}{(\nu+2+N)(\nu+N)} \cdot \frac{(\nu-2)^2}{(\nu+2)\nu} \cdot (\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k}) \\
&= \frac{(\nu+N)}{(\nu+2+N)} \cdot (\delta_{i,j}\delta_{k,\ell}+\delta_{i,k}\delta_{j,\ell}+\delta_{i,\ell}\delta_{j,k})
\end{aligned}
$$

Combining all results, we have the conditional Fisher information matrix

$$
\begin{aligned}
\mathcal{I}_{t-1} = \frac{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top}{\partial \boldsymbol{f}_t} \cdot \frac{\partial\,\mathrm{vec}(\boldsymbol{R}_t)^\top}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)} \cdot \left(\boldsymbol{R}_t^{-1/2} \otimes \boldsymbol{R}_t^{-1/2}\right) \times \\
\left(\mathcal{G}^\star - \mathrm{vec}\left(\mathbf{I}_N\right)\mathrm{vec}\left(\mathbf{I}_N\right)^\top\right) \times \\
\left(\boldsymbol{R}_t^{-1/2} \otimes \boldsymbol{R}_t^{-1/2}\right)^\top \cdot \frac{\partial\,\mathrm{vec}(\boldsymbol{R}_t)}{\partial\,\mathrm{vec}(\boldsymbol{L}_t)^\top} \cdot \frac{\partial\,\mathrm{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top}.
\end{aligned}
$$

We obtain the Gaussian case by letting $\nu \to \infty$, such that $\mathcal{G}^{\star}_{i,j,k,\ell}$ collapses to $\delta_{i,j}\delta_{k,\ell} + \delta_{i,k}\delta_{j,\ell} + \delta_{i,\ell}\delta_{j,k}$.

B Sample composition

B.1 Sample composition

Table B.1: Selected S&P500 constituents

This table lists ticker symbols of the 100 stocks in our dataset. All stocks are included in the S&P 500 index. Tickers are grouped per industry.

Ind Nr.	Industry	# Comp.	Tickers
1	Capital Goods	10	AA,BA,CAT,HON,F,NOC,UTX,A,IR,GD
2	Financials	19	AXP,JPM,AIG,BAC,C,KEY,MTB,COF,USB, BBT,STI,WFC,GS,MS,MMC,HIG,PNC, XL,MCO
3	Energy	12	GE,XOM,BHI,MUR,SLB,CVX,HAL,OXY, APC,SU,CNX,PXD
4	Consumer Services	14	HD,MCD,WMT,TGT,BXP,DIS,JCP,NLY, ANF,EQR,WY,RCL,WSM,TV
5	Consumer Non-Durables	9	KO,MO,SYY,PEP,CL,AVP,GIS,CPB,EL
6	Health Care	11	PFE,ABT,BAX,JNJ,LLY,THC,MMM,MRK,BMY, MDT,CI
7	Public Utilities	7	AEP,AEE,DUK,SO,WMB,VZ,EXC
8	Technology	5	IBM,DOV,HPQ,TSM,CSC
9	Basic Industries	9	PG,DD,FLR,DOW,AES,IP,ATI,LPX,POT
10	Transportation	4	LUV,UPS,NSC,FDX

B.2 Full simulation results

This supplementary appendix presents the full details of the three Monte Carlo experiments from Section 3.

In the first experiment, we simulate $N = 100$ dimensional time series of length $T = 500$ or $1,000$ with $G = 10$ equally sized groups holding $N/G = 10$ individual cross-sectional units each. These sizes roughly correspond to the data dimensions in our empirical application.

As our data-generating process (DGP), we take the Multi-Factor copula

$$x_{i,t} = \sqrt{\zeta_t}\left(\tilde{\boldsymbol{\lambda}}_{i,t}^{\top}\boldsymbol{z}_t + \sigma_{i,t}\epsilon_{i,t}\right), \tag{B.1}$$

$$\tilde{\boldsymbol{L}}_t = \begin{pmatrix} \tilde{\lambda}^{eq}_{1,t} & \tilde{\lambda}^{gr,f}_{2,1,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\lambda}^{eq}_{1,t} & 0 & \cdots & \tilde{\lambda}^{gr,f}_{2,G,t} \end{pmatrix} \otimes \boldsymbol{\iota}_{N/G}$$

$$f^{eq}_{1,t+1} = \omega^{eq} + A^{eq}\, s^{eq}_t + B\, f^{eq}_{1,t}, \tag{B.2}$$

$$f^{gr,f}_{2,g,t+1} = \omega^{gr,f}_g + A^{gr,f}\, s^{gr,f}_{g,t} + B\, f^{gr,f}_{2,g,t}, \qquad g = 1,\ldots,G, \tag{B.3}$$

with $\otimes$ denoting the Kronecker product, and where $\boldsymbol{z}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_{G+1})$, $\epsilon_{i,t} \sim \mathrm{N}(0,1)$, and $\zeta_t \sim \text{Inv-Gamma}\left(\frac{1}{2}\nu_C, \frac{1}{2}\nu_C\right)$, and $f^{eq}_{1,t+1}$ together with $f^{gr,f}_{2,g,t+1}$ form the vectors $\boldsymbol{\lambda}_{i,t}$, which are finally scaled into $\tilde{\boldsymbol{\lambda}}_{i,t}$ using (3). The expressions for the scores s^{eq}_t and $s^{gr,f}_{g,t}$ can be found in Supplementary Appendix A.

The Multi-Factor copula model has two different types of factor loadings, each with its own score-driven dynamics: one $f^{eq}_{1,t}$ for the common equi-factor, and G different $f^{gr,f}_{2,g,t}$s for each of the group-specific factors. Each of these 11 loadings has its own intercept. We use a pooled persistence parameter B common to all factor loadings, and type-specific score parameters A^{eq} and $A^{gr,f}$.

Guided by the empirical application, we set $\omega^{eq} = 0.07$ and let ω_g be equally spaced on the interval $[0.01\,, 0.07]$. For the Gaussian copulas, we set $A^{eq} = 0.0085$ and $A^{gr,f} = 0.0095$, while for the t-copula these parameters equal 0.015 and 0.01, respectively. For the copula's tail behavior, we use $\nu_C \in \{35, \infty\}$, where $\nu_C \to \infty$ corresponds to the Gaussian factor copula. Finally, in line with our empirical results later on we set $B = 0.87$ for normally distributed factors ($\nu_C \to \infty$) and $B = 0.92$ for the Student's t case ($\nu_C = 35$).

Table B.2 presents the results based on 1,000 replications. All parameters are estimated near their true values. The standard deviations decrease in T. We also observe that the mean of the estimated standard errors over all simulation runs matches closely the Monte-Carlo standard error of the estimates, indicating that computed standard errors fairly reflect estimation uncertainty. Overall, we conclude that the parameters of the Gaussian and Student's t factor copulas with score-driven dynamic factor loadings can be accurately estimated if the model is correctly specified.

In the second Monte Carlo experiment, we investigate the two-step approach of estimating the copula parameters of the Multi-Factor LT model. For this study, we simulate 1,000 time-series of length $T = 1,000$ and dimension $N = 100$ with $G = 10$ equally sized groups holding $N/G = 10$ assets using the MF-LT model with Normal and Student's $t(35)$ distributed errors. Based on empirical parameter estimates, we set A and B equal to 0.015 and 0.97 respectively and allow for $(10 \times 11)/2 = 55$ different ω parameters, ranging from -0.10 to 0.9. Table B.3 presents the results based on 1,000 replications. The Monte Carlo averages of almost all parameters again lie close to their true values. Note that the standard deviations of moment-based estimators for ω are considerably higher than the standard errors of the ML estimators for A,B, and ν_C. Using the two-step estimator thus implies a huge computational gain, but at the cost of some efficiency loss. The average estimated standard errors for A, B for the

Table B.2: Monte Carlo results of parameter estimates of the Multi-Factor-Copula
This table provides Monte Carlo results of parameter estimates using the multi-factor (MF) Gaussian and t-copula model as given in (B.1)–(B.3). $B(N)$ and $B(t)$ denote the value of B in case of the Gaussian (N) and Student's t (t) factor copula model, respectively. The table reports the mean and standard deviation of the estimated coefficients, as well as the mean of the computed standard error. Results are based on 1,000 Monte Carlo replications.

Panel A: T = 500							
		MF N			MF t		
Coef.	True	mean	std	mean(s.e.)	mean	std	mean(s.e.)
ω_{eq}	0.0700	0.0761	0.0145	0.0159	0.0752	0.0121	0.0127
ω_1	0.0100	0.0094	0.0074	0.0097	0.0097	0.0050	0.0051
ω_2	0.0167	0.0172	0.0073	0.0076	0.0179	0.0044	0.0042
ω_3	0.0233	0.0251	0.0070	0.0073	0.0250	0.0049	0.0049
ω_4	0.0300	0.0326	0.0076	0.0081	0.0323	0.0060	0.0059
ω_5	0.0367	0.0400	0.0090	0.0093	0.0395	0.0068	0.0070
ω_6	0.0433	0.0471	0.0097	0.0106	0.0468	0.0079	0.0081
ω_7	0.0500	0.0542	0.0108	0.0120	0.0540	0.0088	0.0092
ω_8	0.0567	0.0617	0.0125	0.0134	0.0612	0.0100	0.0104
ω_9	0.0633	0.0691	0.0135	0.0149	0.0684	0.0111	0.0115
ω_{10}	0.0700	0.0763	0.0150	0.0162	0.0755	0.0121	0.0127
$A_{eq}(N)$	0.0085	0.0085	0.0012	0.0011			
$A_{gr,f}(N)$	0.0095	0.0089	0.0027	0.0025			
$A_{eq}(t)$	0.0150				0.0146	0.0027	0.0026
$A_{gr,f}(t)$	0.0100				0.0090	0.0025	0.0023
$B(N)$	0.8700	0.8584	0.0269	0.0296			
$B(t)$	0.9200				0.9136	0.0137	0.0144
ν_C	35.000				35.445	2.828	2.702
Panel B: T = 1000							
ω_{eq}	0.0700	0.0739	0.0119	0.0134	0.0744	0.0115	0.0112
ω_1	0.0100	0.0094	0.0058	0.0068	0.0104	0.0033	0.0032
ω_2	0.0167	0.0173	0.0050	0.0054	0.0176	0.0036	0.0033
ω_3	0.0233	0.0246	0.0051	0.0057	0.0248	0.0043	0.0042
ω_4	0.0300	0.0314	0.0059	0.0065	0.0319	0.0052	0.0052
ω_5	0.0367	0.0387	0.0069	0.0077	0.0389	0.0063	0.0061
ω_6	0.0433	0.0458	0.0078	0.0088	0.0461	0.0073	0.0071
ω_7	0.0500	0.0529	0.0088	0.0100	0.0531	0.0084	0.0081
ω_8	0.0567	0.0599	0.0100	0.0112	0.0602	0.0094	0.0091
ω_9	0.0633	0.0667	0.0110	0.0123	0.0673	0.0103	0.0101
ω_{10}	0.0700	0.0738	0.0121	0.0136	0.0744	0.0115	0.0112
$A_{eq}(N)$	0.0085	0.0085	0.0009	0.0008			
$A_{gr,f}(N)$	0.0095	0.0093	0.0018	0.0018			
$A_{eq}(t)$	0.0150				0.0149	0.0020	0.0019
$A_{gr,f}(t)$	0.0100				0.0096	0.0016	0.0016
$B(N)$	0.8700	0.8626	0.0221	0.0248			
$B(t)$	0.9200				0.9149	0.0129	0.0126
ν_C	35.00				35.1760	1.8629	1.8821

Table B.3: Monte Carlo results of parameter estimates of the MF-LT model

This table provides Monte Carlo results of parameter estimates using the multi-factor (MF) LT Gaussian and t-copula model with a loading matrix given in (6). The table reports the mean and standard deviation based on 1,000 Monte Carlo replications. Since we have 55 different values of ω, we only report $\omega_1, \omega_4, \dots, \omega_{55}$ in addition to A, B and ν_C.

		MF-LT N		MF-LT t	
Coef.	True	mean	std	mean	std
ω_1	0.893	0.884	0.0530	0.886	0.0539
ω_4	0.621	0.606	0.0538	0.607	0.0541
ω_7	0.560	0.550	0.0526	0.551	0.0511
ω_{10}	0.845	0.832	0.0524	0.830	0.0529
ω_{13}	0.187	0.185	0.0485	0.186	0.0495
ω_{16}	0.146	0.144	0.0491	0.149	0.0490
ω_{19}	0.119	0.122	0.0478	0.127	0.0487
ω_{22}	0.009	0.012	0.0473	0.013	0.0475
ω_{25}	0.003	0.007	0.0492	0.005	0.0477
ω_{28}	0.310	0.385	0.0434	0.384	0.0397
ω_{31}	0.163	0.127	0.0508	0.127	0.0511
ω_{34}	0.156	0.119	0.0483	0.123	0.0475
ω_{37}	0.250	0.229	0.0477	0.230	0.0479
ω_{40}	0.026	0.041	0.0466	0.042	0.0458
ω_{43}	0.011	0.011	0.0453	0.007	0.0479
ω_{46}	0.591	0.626	0.0447	0.622	0.0435
ω_{49}	-0.020	-0.009	0.0466	-0.008	0.0457
ω_{52}	0.016	0.017	0.0491	0.018	0.0493
ω_{55}	0.347	0.504	0.0436	0.497	0.0431
A	0.015	0.016	0.0006	0.016	0.0007
B	0.970	0.970	0.0025	0.970	0.0024
ν_C	30.00			35.06	1.862

MF-LT N and for the MF-LT t model again lie close to their Monte-Carlo counterparts, such that standard errors correctly reflect the estimation uncertainty. We further note that the assumed distribution does not have a large impact on the moment estimator of ω.

C Recap of benchmark MGARCH models

In this appendix we give a brief recap of the MGARCH benchmark models we use, in particular the cDCC model (Engle, 2002) (with the correction of Aielli, 2013) and the (Block) DECO model of Engle and Kelly (2012) in high dimensions. To maintain a fair comparison between both classes of models, we also cast the MGARCH models into a copula framework. Hence the innovations in these models are $x_{i,t} = P^{-1}(u_{i,t})$, with $u_{i,t}$ estimated in a first step by the same marginals, and $P^{-1}(\,\cdot\,)$ the inverse marginal CDF corresponding to the copula specification at hand.

The cDCC model is given by

$$\begin{aligned} \boldsymbol{Q}_{t+1} &= \boldsymbol{\Omega} + A\,\boldsymbol{Q}_t^* \boldsymbol{x}_t \boldsymbol{x}_t^\top \boldsymbol{Q}_t^* + B\,\boldsymbol{Q}_t \\ \boldsymbol{R}_t^{cDCC} &= \boldsymbol{Q}_t^{*-1} \boldsymbol{Q}_t \boldsymbol{Q}_t^{*-1} \end{aligned} \tag{C.1}$$

with $\boldsymbol{Q}_t^*$ a diagonal matrix with entries $q_{ii,t}$, A and B scalars and $\boldsymbol{\Omega}$ a $N \times N$ matrix. The DECO model assumes that the dependence between all assets is the same (equi-dependence) and takes the average of all pairwise DCC correlations:

$$\boldsymbol{R}_t^{DECO} = \rho_t \boldsymbol{J}_{N\times N} + (1-\rho_t)\boldsymbol{I}_N \tag{C.2}$$

$$\rho_t = \frac{1}{N(N-1)}(\boldsymbol{\iota}^\top \boldsymbol{R}_t^{cDCC} \boldsymbol{\iota} - N) \tag{C.3}$$

where $\boldsymbol{J}_{N\times N}$ denotes a $N \times N$ matrix of ones. As noted earlier, the DECO model corresponds to a one-factor model, though the DECO and score-driven dynamics are different.

A third variant is the Block DECO model that allows for different intra-block correlations $\rho_{g,g}$, and inter-block correlations $\rho_{g,h}$ with $g \neq h$. Similar to the multi-factor models, the size of each block may differ. The Block DECO correlation matrix is defined as

$$\begin{aligned} \boldsymbol{R}_t^{BL-DECO} &= \begin{pmatrix} (1-\rho_{1,1,t})\boldsymbol{I}_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (1-\rho_{G,G,t})\boldsymbol{I}_{n_G} \end{pmatrix} \\ &+ \begin{pmatrix} \rho_{1,1,t}\boldsymbol{J}_{n_1\times n_1} & \cdots & \rho_{1,G,t}\boldsymbol{J}_{n_1\times n_G} \\ \vdots & \ddots & \vdots \\ \rho_{1,G,t}\boldsymbol{J}_{n_G\times n_1} & \cdots & \rho_{G,G,t}\boldsymbol{J}_{n_G\times n_G} \end{pmatrix}. \end{aligned} \tag{C.4}$$

The Block DECO model allows for G distinct within-group correlations $\rho_{g,g,t}$, $g = 1, \ldots, G$ as well as for $G(G-1)/2$ unique (off-diagonal) between-group correlations $\rho_{g,h,t}$ for $g \neq h$. The dynamic correlations are computed as

$$\rho_{g,g,t} = \frac{1}{n_g(n_g - 1)} \sum_{i \in g, j \in g, i \neq j} \frac{q_{i,j,t}}{\sqrt{q_{i,i,t} q_{j,j,t}}}, \tag{C.5}$$

$$\rho_{g,h,t} = \frac{1}{n_g n_h} \sum_{i \in g, j \in h} \frac{q_{i,j,t}}{\sqrt{q_{i,i,t} q_{j,j,t}}}, \quad g \neq h, \tag{C.6}$$

where $q_{i,j,t}$ is the i,j-th element of the matrix $\boldsymbol{Q}_t$ from the cDCC model in (C.1). Put differently, the correlations of the Block DECO model are obtained by averaging all DCC correlations within each block.

Similar to the multi-factor dynamic copula models, the Block DECO model allows for different within-group and between-group correlations. This model comes with an additional flexibility: via the matrix Ω each between-group correlation has its own intercept, while in the factor copula approach the between-group correlations are spanned by a smaller set of parameters. This flexibility comes at two important costs. First, it is hard to impose ex-ante that the dynamic correlations from the Block DECO give rise to a positive definite correlation matrix. Though in practice a maximum likelihood type approach will steer the parameters away from a region where the predicted dependence matrix is indefinite, this is not guaranteed by the structure of the model. By contrast, the factor copula models with score-driven dynamics automatically ensure a positive semi-definite correlation matrix at all times, which is particularly relevant when using the model for forecasting. Second, the Block DECO model averages DCC correlations, which means that it relies heavily on the A and B parameters from the cDCC model and its unconditional $N \times N$ intercept $\boldsymbol{\Omega}$.

D Models for the marginals

In our main analysis, we use the univariate t-GAS volatility model of Creal et al. (2011, 2013) for the marginal distributions. That is, we assume a Student's t distribution for the individual returns $y_{i,t}$ with ν_i degrees of freedom with the following return and volatility dynamics (omitting the subscript i for the sake of exposition)

$$y_t = \phi_0 + \sum_{j=1}^{p} \phi_j y_{t-j} + \epsilon_t, \qquad \epsilon_t \sim t(0, h_t, \nu), \tag{D.1}$$

$$h_{t+1} = \omega + \alpha\,(w_t \epsilon_t^2 - h_t) + \beta\, h_t, \qquad w_t = \frac{\nu + 1}{\nu - 2 + {h_t}^{-1} \epsilon_t^2}, \tag{D.2}$$

with h_t the conditional variance at time t. This model updates the conditional variance by the (scaled) score, i.e., the partial derivative of the log Student's t density with respect to the variance h_t. We follow Creal et al. (2011, 2013) and scale the score by the inverse conditional Fisher information matrix. The interpretation of the scaled score is highly intuitive in this model: Large values of ϵ_t^2 are downweighted by w_t, since possible outliers (jumps) might not only be attributed to an increase in variance, but also to the fat-tailed nature of the return data. The estimation results for the marginal models are summarized in Table D.1.

Table D.1: Marginal distribution parameter estimates

This table reports summaries of the maximum likelihood parameter estimates of the t-GAS volatility models in (D.1)-(D.2) for 100 daily time series of equity returns. The columns present the mean and quantiles of the cross-sectional distribution of each parameter. Data are observed over the period January 2, 2001 until December 31, 2014 ($T = 3,521$ trading days).

	Mean	5%	25%	Med	75%	95%
ϕ_0	0.027	-0.030	0.010	0.025	0.046	0.091
ϕ_1	-0.009	-0.049	-0.027	-0.008	0.008	0.026
ϕ_2	-0.012	-0.044	-0.028	-0.011	0.001	0.020
ω	0.025	0.009	0.014	0.021	0.029	0.060
α	0.091	0.062	0.077	0.088	0.104	0.129
β	0.991	0.983	0.988	0.992	0.995	0.998
ν	8.22	5.53	6.77	8.21	9.25	11.41
KS test for Student's t dist of std. residuals						
Number of rejections		5				

As a robustness check, we also considered marginal distributions based on a GARCH model with a skewed Student's t distribution for the innovations. The specification of that model is

$$\begin{aligned} y_t &= \phi_0 + \sum_{j=1}^{p} \phi_j y_{t-j} + \epsilon_t, \qquad \epsilon_t \sim St(0, h_t, \nu, \lambda), && \text{(D.3)} \\ h_{t+1} &= \omega + \alpha\, \epsilon_t^2 + \beta\, h_t, && \text{(D.4)} \end{aligned}$$

where the pdf of the skewed Student's t distribution of Hansen (1994) for a zero mean variable $z_t = (y_t - \mu_t)/\sqrt{h_t}$ with $\mu_t = \mathbb{E}_{t-1}[y_t]$ is given by

$$f(z_t; \lambda, \nu) = \begin{cases} bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz_t+a}{1-\lambda}\right)^2\right)^{-\frac{\nu+1}{2}} & \text{if} \quad z_t < -\frac{a}{b} \\ bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz_t+a}{1+\lambda}\right)^2\right)^{-\frac{\nu+1}{2}} & \text{if} \quad z_t \geq -\frac{a}{b} \end{cases} \tag{D.5}$$

with

$$a = 4\lambda c \frac{\nu-2}{\nu-1}, \quad b^2 = 1 + 3\lambda^2 - a^2, \text{and} \quad c = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}$$

such that $f(y_t|\mu_t, h_t, \nu, \lambda) = 1/h_t f(z_t; \lambda, \nu)$. Further, λ is the skewness parameter and ν again represents the degrees of freedom. A (positive) negative value of λ indicates (positive) negative skewness.

The results for these marginals are given in Table D.2 and result in qualitatively similar conclusions as the main results in Table 4 (see Appendix E).

Table D.2: Marginal distribution parameter estimates (skewed Student's t distribution)

This table reports summaries of the maximum likelihood parameter estimates of the GARCH skewed Student's t volatility models in (D.3)-(D.4) for 100 daily time series of equity returns. The columns present the mean and quantiles of the cross-sectional distribution of each parameter. Data are observed over the period January 2, 2001 until December 31, 2014 ($T = 3,521$ trading days).

	Mean	5%	25%	Med	75%	95%
ϕ_0	0.025	-0.026	0.008	0.022	0.041	0.095
ϕ_1	-0.009	-0.051	-0.030	-0.007	0.009	0.025
ϕ_1	-0.013	-0.043	-0.029	-0.013	-0.001	0.018
ω	0.026	0.008	0.014	0.021	0.029	0.064
α	0.067	0.043	0.055	0.063	0.079	0.098
β	0.924	0.887	0.909	0.928	0.939	0.950
ν	7.84	5.28	6.47	7.74	8.92	11.13
λ	-0.011	-0.071	-0.037	-0.010	0.013	0.055

KS test for Student's t dist of std. residuals	
Number of rejections	2

E Additional In-sample Factor Copula results

This appendix provides supplementary results for multi-factor copula models with respect to three issues: 1) the non-reported different ω_i of our (multi-) factor copulas listed in Table 4, 2) parameter estimates of all factor models when the PITs are based on skewed Student's t GARCH model and 3) the sensitivity of the MF-LT t model with respect to the ordering of the industries.

The intercepts reported in Table E.1 can be further interpreted. For instance, the industry intercepts of the MF model show that the within Financial and Energy correlations are unconditionally much higher than for example within Capital Goods and Basic Industries correlations. This holds for both the Gaussian and t copula models.

The second part of this appendix holds a robustness check with respect to the assumed marginal distributions. As mentioned in Appendix D, in our main analysis we use the univariate t-GAS volatility model of Creal et al. (2011, 2013) for the marginal distributions. In the same appendix, we also presented the results for skewed Student's t GARCH marginals. Table E.2 contains the results for all Factor and MGARCH Copula models if the PITs of the skewed t GARCH marginals are used in the copula analysis. The results confirm the analysis of the main text, and the statistical ordering of the different copula specifications.

The third part of this appendix contains a robustness check with respect to the ordering of the industries when estimating the MF-LT model. We re-estimated the MF-LT t model for 50 different random orderings to investigate the model's sensitivity to this. Table E.3 presents the average, minimum and maximum values of the estimates of A, B, and ν, as well as the log-likelihood value.

We find that the estimated parameters are very stable with respect to the ordering chosen. There appears to be some limited variation ($< 1\%$) in the maximized log-likelihood value, such that some further small gains in likelihood might be possible by optimizing over the ordering of the industries.

To conclude, Figure E.1 shows the fitted within and between dependencies of Capital Goods, Financials and Health companies according to our empirical specification against a randomly chosen alternative ordering. Again, the differences are hardly noticeable. We conclude that the ordering of the groups does not materially affect our results.

Table E.1: Intercept parameter estimates of (multi-) Factor Copula models

This table reports the maximum likelihood parameter estimation results of intercepts ω_i for the 1F-Group, 2F and MF copula models listed in Table 4. The different intercepts correspond to the Capital Goods, Finance, Energy, Consumer Services, Consumer Non-Durables, Health Care, Public Utilities, Technology, Basic Industries, and Transportation industries respectively. We also list again the estimated B parameter. Panel A (B) corresponds with the Gaussian (t) copula likelihood. Data are observed over the period January 2, 2001 until December 31, 2014 ($T = 3,521$ trading days).

	1F-Group		2F		MF	
Parameter	$\hat{\omega}_g$	s.e.	$\hat{\omega}_g$	s.e.	$\hat{\omega}_g$	s.e.
Panel A: Gaussian factor copulas						
ω_{eq}			0.047	(0.005)	0.042	(0.005)
$\omega_{CapGoods}$	0.025	(0.005)	0.026	(0.004)	0.011	(0.001)
ω_{Fin}	0.030	(0.006)	0.054	(0.002)	0.056	(0.008)
ω_{Energy}	0.020	(0.005)	0.011	(0.003)	0.055	(0.007)
$\omega_{ConsSer}$	0.020	(0.004)	0.017	(0.007)	0.021	(0.003)
$\omega_{ConsNon-Dur}$	0.016	(0.004)	0.013	(0.005)	0.033	(0.005)
ω_{Health}	0.018	(0.004)	0.005	(0.007)	0.029	(0.004)
$\omega_{PublUtil}$	0.017	(0.004)	0.012	(0.003)	0.041	(0.005)
ω_{Tech}	0.023	(0.005)	0.011	(0.002)	0.021	(0.003)
$\omega_{BasicInd}$	0.022	(0.005)	0.009	(0.009)	0.008	(0.001)
$\omega_{Transport}$	0.024	(0.005)	0.014	(0.005)	0.038	(0.006)
B	0.970	(0.006)	0.941	(0.004)	0.930	(0.009)
Panel B: t factor copulas						
ω_{eq}			0.004	(0.002)	0.033	(0.002)
$\omega_{CapGoods}$	0.012	(0.001)	0.002	(0.001)	0.015	(0.001)
ω_{Fin}	0.014	(0.002)	0.006	(0.002)	0.034	(0.002)
ω_{Energy}	0.010	(0.001)	-0.001	(0.001)	0.033	(0.002)
$\omega_{ConsSer}$	0.009	(0.001)	0.002	(0.001)	0.011	(0.001)
$\omega_{ConsNon-Dur}$	0.008	(0.001)	0.001	(0.001)	0.018	(0.001)
ω_{Health}	0.009	(0.001)	0.002	(0.001)	0.017	(0.001)
$\omega_{PublUtil}$	0.008	(0.001)	0.000	(0.000)	0.026	(0.001)
ω_{Tech}	0.011	(0.001)	0.003	(0.001)	0.012	(0.001)
$\omega_{BasicInd}$	0.010	(0.001)	0.002	(0.001)	0.008	(0.001)
$\omega_{Transport}$	0.011	(0.001)	0.003	(0.001)	0.020	(0.001)
B	0.986	(0.001)	0.993	(0.002)	0.957	(0.002)

Table E.2: Parameter estimates of the full sample based on skewed Student's t errors

This table reports maximum likelihood parameter estimates of various factor copula models, the (block) DECO model of Engle and Kelly (2012) and the cDCC model of Engle (2002), applied to daily equity returns of 100 assets listed at the S&P 500 index. The marginals are modeled assuming a skewed t GARCH model. We consider five different factor copula models, see Table 1 for the definition of their abbreviations. Panel A.1 presents the factor models with a Gaussian copula density, Panel A.2 presents the parameter estimates corresponding with the t-factor copula. Panel B.1 and B.2 present the estimates of the MGARCH class of models. In case of the cDCC and Block DECO models, the table shows parameters estimates obtained by the Composite Likelihood (CL) method. Standard errors are provided in parenthesis and based on the (sandwich) robust covariance matrix estimator. We report the copula log-likelihood,the Akaike Information Criteria (AIC) as well as the number of estimated parameters for all models. The sample comprises daily returns from January 2, 2001 until December 31, 2014 (3,521 observations).

Model	ω^{eq}	A^{eq}	A^{gr}	B	ν	LogL	AIC	♯ para
Panel A.1: Gaussian factor copulas								
1F-Equi	0.018	0.005		0.973		66,055	-132,105	3
	(0.002)	(0.000)		(0.003)				
1F-Group			0.007	0.969		68,221	-136,419	12
			(0.001)	(0.009)				
2F	0.058	0.006	0.008	0.913		73,380	-146,733	14
	(0.007)	(0.000)	(0.001)	(0.010)				
MF	0.080	0.007	0.005	0.896		82,329	-164,630	14
	(0.004)	(0.000)	(0.001)	(0.005)				
MF-LT		0.009		0.962		**83,401**	**-166,688**	57
		(0.001)		(0.006)				
Panel A.2: t-factor copulas								
1F-Equi	0.060	0.012		0.920	34.39	69,790	-139,571	4
	(0.010)	(0.001)		(0.013)	(1.29)			
1F-Group			0.004	0.986	30.12	72,420	-144,815	13
			(0.000)	(0.000)	(0.99)			
2F	0.035	0.011	0.011	0.946	36.80	76,607	-153,184	15
	(0.004)	(0.001)	(0.001)	(0.005)	(1.52)			
MF	0.070	0.014	0.010	0.909	42.83	84,804	-169,578	15
	(0.002)	(0.001)	(0.001)	(0.003)	(1.48)			
MF-LT		0.004		0.991	34.30	**86,603**	**-173,091**	58
		(0.000)		(0.002)	(1.24)			
Panel B.1: Gaussian copula-MGARCH models								
cDCC (CL)		0.017		0.967		76,210	-142,515	4,952
		(0.001)		(0.003)				
DECO		0.031		0.957		65,034	-120,165	4,952
		(0.003)		(0.005)				
Block DECO		0.030		0.956		**83,306**	**-156,707**	4,952
		(0.002)		(0.003)				
Panel B.2: t copula-MGARCH models								
cDCC (CL)		0.018		0.967	13.92	84356	-158,807	4,953
		(0.001)		(0.003)	(0.56)			
DECO		0.038		0.949	29.97	69630	-129,354	4,953
		(0.003)		(0.005)	(1.13)			
Block DECO		0.031		0.955	21.82	**86,450**	**-162,995**	4,953
		(0.002)		(0.003)	(0.57)			

Table E.3: Estimated parameters of the MF-LT t model

This table contains summary statistics of the estimate parameters of a MF-LT t model with 50 different random ordering of groups. We show the average, minimum and maximum values of the parameters and the maximized log-likelihood over the 50 generated random orderings. The first row (*current*) corresponds with the ordering used in the paper. Results are based on the full sample.

	A	B	ν	LogL
current	0.004	0.990	36.22	86,433
mean	0.005	0.990	36.35	86,463
min	0.004	0.988	35.82	86,314
max	0.006	0.992	36.79	86,551

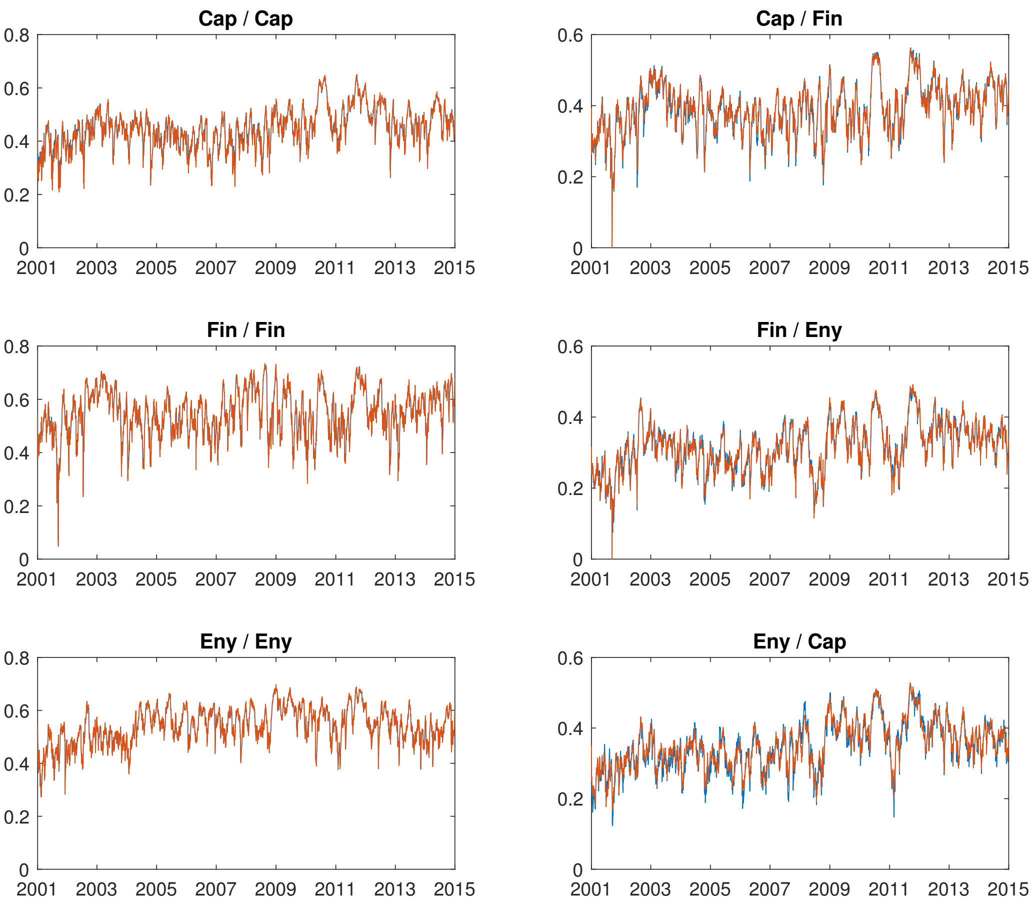

Figure E.1: Fitted dependencies of the MF-LT t model according to two different orders

This figure shows within and between dependencies of Financials, Capital Goods and Energy according to the MF-LT model. The red line is based on the group ordering used in the paper, while the blue line corresponds with a randomly chosen group ordering. The sample spans the period from January 2, 2001 until December 31, 2014 ($T = 3,521$ days).

F Further derivations for numerical implementation

F.1 General shortcuts

The construction of the score in the Matlab code has a number of numerical efficiency enhancements compared to the notationally simple set-up from Appendix A. In particular, we use several matrix algebraic identities to speed up the likelihood calculations.

In general we use formulas (A.1)–(A.11). We then deviate from Appendix A in three ways:

1. With respect to (A.11), we directly derive

$$\frac{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)}{\partial \operatorname{vec}(\boldsymbol{f}_t)^\top}, \qquad \frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\boldsymbol{f}_t)^\top},$$

rather than

$$\frac{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)}{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top} \cdot \frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top}, \qquad \frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top} \cdot \frac{\partial \operatorname{vec}(\boldsymbol{L}_t)}{\partial \boldsymbol{f}_t^\top}.$$

This implies that we need not explicitly derive

$$\frac{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)}{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top}, \qquad \frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\boldsymbol{L}_t)^\top},$$

and hence we do not use (A.13)–(A.27). This also means that (A.6) changes into

$$\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C)}{\partial \boldsymbol{f}_t^\top} = \frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top} \cdot \frac{\partial \operatorname{vec}(\boldsymbol{R}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top} \cdot \frac{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)}{\partial \boldsymbol{f}_t^\top}. \tag{F.1}$$

For the 1-factor model with an equi-loading, $\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)/\partial \boldsymbol{f}_t^\top$ and $\partial \operatorname{vec}(\boldsymbol{D}_t)/\partial \operatorname{vec}(\boldsymbol{f}_t)^\top$ are relatively easy to derive at once. For the 1F-Gr, 2F, MF and MF-LT models, this is more involved. We provide more detailed information in the next sections.

2. Let $K_{k,N}$ be a commutation matrix such that $K_{k,N} \operatorname{vec}(\tilde{\boldsymbol{L}}) = \operatorname{vec}(\tilde{\boldsymbol{L}}^\top)$ for a general matrix $\tilde{\boldsymbol{L}} \in \mathbb{R}^{k\times N}$. We use K_N for $K_{N,N}$. Let us define $\boldsymbol{B}_t$ to rewrite (A.10) as

$$\operatorname{vec}(\boldsymbol{B}_t) = \frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t; \boldsymbol{R}_t, \boldsymbol{\psi}_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)}.$$

Then we have

$$\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top}\cdot\frac{\partial \operatorname{vec}(\boldsymbol{R}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top} = \operatorname{vec}(\boldsymbol{B}_t)^\top\left(\tilde{\boldsymbol{L}}_t^\top\otimes \mathbf{I}_N\right)K_{k,N}+\operatorname{vec}(\boldsymbol{B}_t)^\top\left(\mathbf{I}_N\otimes\tilde{\boldsymbol{L}}_t^\top\right) + \operatorname{vec}(\boldsymbol{B}_t)^\top\frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}$$

Using the rule $\operatorname{vec}(\boldsymbol{B})^\top(\boldsymbol{C}\otimes\boldsymbol{A}^\top)=\operatorname{vec}\left(\boldsymbol{ABC}\right)^\top$ for arbitrarily chosen matrices $\boldsymbol{C}$ and $\boldsymbol{A}$, provided that the matrix product $\boldsymbol{ABC}$ exists, and using $\operatorname{vec}(\boldsymbol{L}^\top)=K_{k,N}\operatorname{vec}(\boldsymbol{L})$ for $\boldsymbol{L}\in\mathbb{R}^{k\times N}$ and $K_{k,N}=K_{N,k}^\top$, we obtain

$$\begin{aligned}
\frac{\partial \log \boldsymbol{c}_t(\boldsymbol{x}_t;\boldsymbol{R}_t,\boldsymbol{\psi}_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top}\cdot\frac{\partial \operatorname{vec}(\boldsymbol{R}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top} &= \\
&= \operatorname{vec}(\boldsymbol{B}_t\tilde{\boldsymbol{L}}_t^\top)^\top K_{k,N}+\operatorname{vec}(\tilde{\boldsymbol{L}}_t\boldsymbol{B}_t)^\top+\operatorname{vec}(\boldsymbol{B}_t)^\top\frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}\\
&= \operatorname{vec}(\boldsymbol{B}_t\tilde{\boldsymbol{L}}_t^\top)^\top K_{N,k}^\top+\operatorname{vec}(\tilde{\boldsymbol{L}}_t\boldsymbol{B}_t)^\top+\operatorname{vec}(\boldsymbol{B}_t)^\top\frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}\\
&= \operatorname{vec}(\tilde{\boldsymbol{L}}_t\boldsymbol{B}_t^\top)^\top+\operatorname{vec}(\tilde{\boldsymbol{L}}_t\boldsymbol{B}_t)^\top+\operatorname{vec}(\boldsymbol{B}_t)^\top\frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}\\
&= \operatorname{vec}(\tilde{\boldsymbol{L}}_t\boldsymbol{B}_t)^\top+\operatorname{vec}(\tilde{\boldsymbol{L}}_t\boldsymbol{B}_t)^\top+\operatorname{vec}(\boldsymbol{B}_t)^\top\frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}\\
&= 2\operatorname{vec}(\tilde{\boldsymbol{L}}_t\boldsymbol{B}_t)^\top+\operatorname{vec}(\boldsymbol{B}_t)^\top\frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top} \qquad \text{(F.2)}
\end{aligned}$$

3. Finally, we simplify the last term of (F.2). Given that $\boldsymbol{D}_t$ is a $N\times N$ diagonal matrix with diagonal vector $\boldsymbol{\sigma}_t^2$ we have $\operatorname{vec}(\boldsymbol{D}_t)=S^D\boldsymbol{\sigma}_t^2$ with S^D a $N^2\times N$ selection matrix such that element$(i+(i-1)N,i)$ equals 1. This implies that

$$\frac{\partial \operatorname{vec}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}=S^D\frac{\partial \operatorname{diag}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}=S^D\frac{\partial \boldsymbol{\sigma}_t^2}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}.$$

The last term in (F.2) now equals

$$\operatorname{vec}(\boldsymbol{B}_t)^\top S^D\frac{\partial \operatorname{diag}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}=\operatorname{diag}(\boldsymbol{B}_t)^\top\frac{\partial \operatorname{diag}(\boldsymbol{D}_t)}{\partial \operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top}. \qquad \text{(F.3)}$$

The benefit is twofold: computing $\operatorname{diag}(\boldsymbol{B}_t)$ is much faster than computing $\operatorname{vec}(\boldsymbol{B}_t)^\top S^D$. Second, the matrix $\partial\operatorname{diag}(\boldsymbol{D}_t)/\partial\operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top$ is much smaller than $\partial\operatorname{vec}(\boldsymbol{D}_t)/\partial\operatorname{vec}(\tilde{\boldsymbol{L}}_t)^\top$.

F.2 1-Factor-equi model

We have

$$\boldsymbol{R}_t = \tilde{\boldsymbol{\lambda}}_t^2 \boldsymbol{\iota}_N \boldsymbol{\iota}_N^\top + \left(1 - \tilde{\boldsymbol{\lambda}}_t^2\right) \mathbf{I}_N = \frac{f_t^2}{1+f_t^2} \boldsymbol{\iota}_N \boldsymbol{\iota}_N^\top + \frac{1}{1+f_t^2} \mathbf{I}_N,$$

where $\boldsymbol{\iota}_N \in \mathbb{R}^{N\times 1}$ is a vector of ones. As a result,

$$\begin{aligned}
\frac{\partial \log \boldsymbol{c}_{Stud,t}(\boldsymbol{x}_t; \boldsymbol{R}_t, \nu_C)}{\partial \operatorname{vec}(\boldsymbol{R}_t)^\top} \cdot \frac{\partial \operatorname{vec}(\boldsymbol{R}_t)}{\partial f_t} &= \operatorname{vec}(\boldsymbol{B}_t)^\top \cdot \operatorname{vec}\left(\boldsymbol{\iota}_N \boldsymbol{\iota}_N^\top - \mathbf{I}_N\right) \frac{\partial \operatorname{vec}(\tilde{\boldsymbol{\lambda}}_t^2)}{\partial f_t} \\
&= \frac{2 f_t}{(1+f_t^2)^2} \cdot \operatorname{tr}\left(\boldsymbol{B}_t^\top \cdot \left(\boldsymbol{\iota}_N \boldsymbol{\iota}_N^\top - \mathbf{I}_N\right)\right) \\
&= \frac{2 f_t}{(1+f_t^2)^2} \cdot \left(\boldsymbol{\iota}_N^\top \boldsymbol{B}_t \boldsymbol{\iota}_N - \operatorname{tr}\left(\boldsymbol{B}_t\right)\right). \qquad \text{(F.4)}
\end{aligned}$$

F.3 1-Factor-Gr model

We have

$$\begin{aligned}
\boldsymbol{R}_t &= \tilde{\boldsymbol{L}}_t^\top \tilde{\boldsymbol{L}}_t + \boldsymbol{D}_t, \\
\tilde{\boldsymbol{L}}_t^\top &= S^L \tilde{\boldsymbol{\lambda}}_t^{gr}, \quad \tilde{\boldsymbol{\lambda}}_t^{gr} = \left(\tilde{\lambda}_{1,t}, \ldots, \tilde{\lambda}_{G,t}\right)^\top \\
\operatorname{diag} \boldsymbol{D}_t &= S^L \operatorname{diag} \boldsymbol{\sigma}_t^{2,gr}, \quad \boldsymbol{\sigma}_t^{2,gr} = \left(\sigma_{1,t}^2, \ldots, \sigma_{G,t}^2\right)^\top,
\end{aligned}$$

with S^L a $N \times G$ selection matrix such that $S^L_{i,g} = 1$ if asset i belongs to group g and 0 elsewhere. Hence $\tilde{\boldsymbol{\lambda}}_t^{gr}$ and $\boldsymbol{\sigma}_t^{2,gr}$ are vectors containing the unique values of $\tilde{\boldsymbol{L}}_t^\top$ and $\operatorname{diag}(\boldsymbol{D}_t)$ respectively. Finally, we have $\boldsymbol{f}_t^\top = \left(f_{1,t}, \ldots, f_{G,t}\right)$ and $\tilde{\lambda}_{g,t} = f_{g,t} / \sqrt{1 + f_{g,t}^2}$.

As a result,

$$\frac{\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t}{\partial \boldsymbol{f}_t^\top} = S^L \operatorname{diag}\left(\frac{\partial \tilde{\lambda}_{1,t}}{\partial f_{1,t}}, \ldots, \frac{\partial \tilde{\lambda}_{G,t}}{\partial f_{G,t}}\right) \qquad \text{(F.5)}$$

$$\frac{\partial \operatorname{diag} \boldsymbol{D}_t}{\partial \boldsymbol{f}_t^\top} = S^L \operatorname{diag}\left(\frac{\partial \sigma_{1,t}^2}{\partial f_{1,t}}, \ldots, \frac{\partial \sigma_{G,t}^2}{\partial f_{G,t}}\right) \qquad \text{(F.6)}$$

with

$$\frac{\partial \tilde{\lambda}_{g,t}}{\partial f_{g,t}} = \frac{1}{(1+f_{g,t}^2)^{3/2}}, \quad \frac{\partial \sigma_{g,t}^2}{\partial f_{g,t}} = \frac{-2 f_{g,t}}{(1+f_{g,t}^2)^2}.$$

The total score in the code is now obtained by combining (F.2), (F.3), (F.5) and (F.6).

F.4 2-Factor model

We have

$$\begin{aligned}
\boldsymbol{R}_t &= \tilde{\boldsymbol{L}}_t^\top \tilde{\boldsymbol{L}}_t + \boldsymbol{D}_t, \\
\tilde{\boldsymbol{L}}_t^\top &= \begin{pmatrix} \tilde{\boldsymbol{\lambda}}_{1,t} & S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr} \end{pmatrix}, \quad \tilde{\boldsymbol{\lambda}}_{2,t}^{gr} = \left(\tilde{\lambda}_{2,1,t}, \ldots, \tilde{\lambda}_{2,G,t}\right)^\top \\
\operatorname{diag} \boldsymbol{D}_t &= S^L \operatorname{diag} \boldsymbol{\sigma}_t^{2,gr}, \quad \boldsymbol{\sigma}_t^{2,gr} = \left(\sigma_{1,t}^2, \ldots, \sigma_{G,t}^2\right)^\top,
\end{aligned}$$

with S^L a $N \times G$ selection matrix defined earlier. Further, define $\boldsymbol{f}_{2,t}^{gr} = \left(f_{2,1,t}, \ldots, f_{2,G,t}\right)^\top$ such that $\boldsymbol{f}_t = \begin{pmatrix} f_{1,t} & (\boldsymbol{f}_{2,t}^{gr})^\top \end{pmatrix}^\top$. Moreover, $\tilde{\boldsymbol{\lambda}}_{1,t}$ is an $N \times 1$ vector with $\tilde{\lambda}_{1,i,t} = f_{1,t}/(1 + f_{1,t}^2 + (S_{i,\cdot}^L \boldsymbol{f}_{2,t}^{gr})^2)^{1/2}$ for $i = 1, \ldots, N$ with $S_{i,\cdot}^L$ the i-th row of the matrix S^L; and $\tilde{\lambda}_{2,g,t} = f_{2,g,t}/(1 + f_{1,t}^2 + (S_{i,\cdot}^L \boldsymbol{f}_{2,t}^{gr})^2)^{1/2}$. It is convenient to define the two vectors $\boldsymbol{f}_{t,de}$ and $\boldsymbol{f}_{t,degr}$ as

$$\boldsymbol{f}_{t,de} = \boldsymbol{\iota}_N + f_{1,t}^2 \boldsymbol{\iota}_N + S^L(\boldsymbol{f}_{2,t}^{gr} \odot \boldsymbol{f}_{2,t}^{gr}), \tag{F.7}$$

$$\boldsymbol{f}_{t,degr} = \boldsymbol{\iota}_G + f_{1,t}^2 \boldsymbol{\iota}_G + (\boldsymbol{f}_{2,t}^{gr} \odot \boldsymbol{f}_{2,t}^{gr}), \tag{F.8}$$

where $\odot$ is the (element-wise) Hadamard product. The $2N \times (G+1)$ matrix $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \boldsymbol{f}_t^\top$ consists of the following four building blocks:

$$\begin{aligned}
\frac{\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t}}{\partial f_{1,t}} &= \left(1/\sqrt{f_{1,t,de}}, \ldots, 1/\sqrt{f_{N,t,de}}\right)^\top - \left(f_{1,t}^2/f_{1,t,de}^{3/2}, \ldots, f_{1,t}^2/f_{N,t,de}^{3/2}\right)^\top \\
&= S^L \left(\operatorname{diag}\left(1/\sqrt{f_{1,t,degr}}, \ldots, 1/\sqrt{f_{G,t,degr}}\right)\right. \\
&\qquad \left. - \operatorname{diag}\left(f_{1,t}^2/f_{1,t,degr}^{3/2}, \ldots, f_{1,t}^2/f_{G,t,degr}^{3/2}\right)\right)
\end{aligned} \tag{F.9}$$

$$\begin{aligned}
\frac{\partial \operatorname{vec} S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}}{\partial (\boldsymbol{f}_{2,t}^{gr})^\top} &= S^L \left(\operatorname{diag}\left(1/\sqrt{f_{1,t,degr}}, \ldots, 1/\sqrt{f_{G,t,degr}}\right)\right. \\
&\qquad \left. - \operatorname{diag}\left(f_{2,1,t}^2/f_{1,t,degr}^{3/2}, \ldots, f_{2,G,t}^2/f_{G,t,degr}^{3/2}\right)\right)
\end{aligned} \tag{F.10}$$

$$\frac{\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t}}{\partial (\boldsymbol{f}_{2,t}^{gr})^\top} = S^L \operatorname{diag}\left(- f_{1,t}\left(f_{2,1,t}/f_{1,t,degr}^{3/2}, \ldots, f_{2,G,t}/f_{G,t,degr}^{3/2}\right)^\top\right) \tag{F.11}$$

$$\frac{\partial \operatorname{vec} S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}}{\partial f_{1,t}} = S^L \left(-f_{1,t}\left(f_{2,1,t}/f_{1,t,degr}^{3/2}, \ldots, f_{2,G,t}/f_{G,t,degr}^{3/2}\right)^\top\right) \tag{F.12}$$

More specifically, these building blocks are inserted as follows:

$$\left(\frac{\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t}{\partial \boldsymbol{f}_t^\top}\right)_{2i-1,\cdot} = \left(\frac{\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t}}{\partial f_{1,t}} \quad \frac{\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t}}{\partial (\boldsymbol{f}_{2,t}^{gr})^\top}\right)_{i,\cdot} \tag{F.13}$$

$$\left(\frac{\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t}{\partial \boldsymbol{f}_t^\top}\right)_{2i,\cdot} = \left(\frac{\partial \operatorname{vec} S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}}{\partial f_{1,t}}, \quad \frac{\partial \operatorname{vec} S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}}{\partial (\boldsymbol{f}_{2,t}^{gr})^\top}\right)_{i,\cdot} \tag{F.14}$$

for $i = 1, \ldots, N$. In words, the odd rows of $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \boldsymbol{f}_t^\top$ are filled with (F.13), while its even rows are filled with (F.14).

Finally, the $N \times (G+1)$ matrix $\partial \operatorname{diag} \boldsymbol{D}_t / \partial \boldsymbol{f}_t^\top$ is given by

$$\frac{\partial \operatorname{vec} \boldsymbol{D}_t}{\partial \boldsymbol{f}_t^\top} = -2\left(\left(f_{1,t}/f_{1,t,de}^2, \ldots, f_{1,t}/f_{N,t,de}^2\right)^\top \quad S^L \operatorname{diag}\left(f_{2,1,t}/f_{1,t,degr}^2, \ldots, f_{2,G,t}/f_{G,t,degr}^2\right)\right). \tag{F.15}$$

F.5 MF model

We have

$$\begin{aligned}
\boldsymbol{R}_t &= \tilde{\boldsymbol{L}}_t^\top \tilde{\boldsymbol{L}}_t + \boldsymbol{D}_t, \\
\tilde{\boldsymbol{L}}_t^\top &= \left(\tilde{\boldsymbol{\lambda}}_{1,t} \quad S^L \operatorname{diag} \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}\right), \quad \tilde{\boldsymbol{\lambda}}_{2,t}^{gr} = \left(\tilde{\lambda}_{2,1,t}, \ldots, \tilde{\lambda}_{2,G,t}\right)^\top \\
\operatorname{diag} \boldsymbol{D}_t &= S^L \operatorname{diag} \boldsymbol{\sigma}_t^{2,gr}, \quad \boldsymbol{\sigma}_t^{2,gr} = \left(\sigma_{1,t}^2, \ldots, \sigma_{G,t}^2\right)^\top,
\end{aligned}$$

with S^L a $N \times G$ selection matrix defined earlier. Note that this model differs only slightly from the 2F model: here we have $S^L \operatorname{diag} \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}$ instead of $S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}$. Hence the industry loadings are now allocated over G columns instead of only 1 column such that we have $G+1$ different factors.

The vectors $\boldsymbol{f}_{2,t}^{gr}$, $\boldsymbol{f}_t$ and $\tilde{\boldsymbol{\lambda}}_{1,t}$ are exactly similar as defined in the 2F model. This also holds for $\boldsymbol{f}_{t,de}$ and $\boldsymbol{f}_{t,degr}$ as defined in (F.7) and (F.8) respectively. Moreover, we can use the three building blocks $\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t} / \partial f_{1,t}$ $\partial \operatorname{vec} S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr} / \partial (\boldsymbol{f}_{2,t}^{gr})^\top$ and $\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t} / \partial (\boldsymbol{f}_{2,t}^{gr})^\top$ of (F.9), (F.10) and (F.11).

As a result, the $(G+1)N \times (G+1)$ matrix $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \boldsymbol{f}_t^\top$ has to be filled as follows:

$$\left(\frac{\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t}{\partial \boldsymbol{f}_t^\top} \right)_{(i-1)(G+1)+1,\cdot} = \left(\frac{\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t}}{\partial f_{1,t}} \quad \frac{\partial \operatorname{vec} \tilde{\boldsymbol{\lambda}}_{1,t}}{\partial (\boldsymbol{f}_{2,t}^{gr})^\top} \right)_{i,\cdot} \tag{F.16}$$

$$\left(\frac{\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t}{\partial \boldsymbol{f}_t^\top} \right)_{(i-1)(G+1)+g_i+1\ ,\ \cdot} = \left(\frac{\partial \operatorname{vec} S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}}{\partial f_{1,t}} \quad \frac{\partial \operatorname{vec} S^L \tilde{\boldsymbol{\lambda}}_{2,t}^{gr}}{\partial (\boldsymbol{f}_{2,t}^{gr})^\top} \right)_{i\ ,\ \cdot} \tag{F.17}$$

for $i = 1, \ldots, N$, and g_i denoting the group number of asset i. All other elements of $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \boldsymbol{f}_t^\top$ are equal to zero. In words: starting from row 1, every $(G+1)$st row of $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \boldsymbol{f}_t^\top$ is filled by (F.13). For the blocks between these rows, only the row corresponding the group of the ith asset has a non-zero row equal to the $2i$th row of the derivative from the 2F-model's derivative.

Finally, the $N \times (G+1)$ matrix $\partial \operatorname{diag} \boldsymbol{D}_t / \partial \boldsymbol{f}_t^\top$ does also not change and is given by (F.15).

F.6 MF-LT model

We have

$$\begin{aligned}
\boldsymbol{R}_t &= \tilde{\boldsymbol{L}}_t^\top \tilde{\boldsymbol{L}}_t + \boldsymbol{D}_t, \\
\tilde{\boldsymbol{L}}_t^\top &= \begin{pmatrix} \tilde{\lambda}_{t,1} \boldsymbol{\iota}_{N_1} & 0 & \cdots & 0 \\ \tilde{\lambda}_{t,2} \boldsymbol{\iota}_{N_2} & \tilde{\lambda}_{t,G+1} \boldsymbol{\iota}_{N_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\lambda}_{t,G} \boldsymbol{\iota}_{N_G} & \tilde{\lambda}_{t,2G-1} \boldsymbol{\iota}_{N_G} & \cdots & \tilde{\lambda}_{t,\frac{1}{2}G(G+1)} \boldsymbol{\iota}_{N_G} \end{pmatrix} \\
\operatorname{diag} \boldsymbol{D}_t &= S^L \operatorname{diag} \boldsymbol{\sigma}_t^{2,gr}, \quad \boldsymbol{\sigma}_t^{2,gr} = \left(\sigma_{1,t}^2, \ldots, \sigma_{G,t}^2 \right)^\top,
\end{aligned} \tag{F.19}$$

with N_g the number of assets in industry g,$(g = 1, \ldots, G)$ and $\boldsymbol{f}_t = (f_{t,1}, f_{t,2}, \ldots, f_{t,G(G+1)/2})^\top \in \mathbb{R}^{G(G+1)/2 \times 1}$. We define $\tilde{\boldsymbol{\lambda}}_t = (\tilde{\lambda}_{t,1}, \tilde{\lambda}_{t,2}, \ldots, \tilde{\lambda}_{t,G(G+1)/2})^\top$

We first decompose $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \boldsymbol{f}_t$ into

$$\frac{\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t}{\partial \boldsymbol{f}_t^\top} = \frac{\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t}{\partial \tilde{\boldsymbol{\lambda}}_t^\top} \frac{\partial \tilde{\boldsymbol{\lambda}}_t}{\partial \boldsymbol{f}_t^\top} \tag{F.20}$$

where $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \tilde{\boldsymbol{\lambda}}_t^\top$ is a $NG \times \frac{1}{2}G(G+1)$ matrix of zeros and ones and $\partial \tilde{\boldsymbol{\lambda}}_t / \partial \boldsymbol{f}_t^\top$ a $\frac{1}{2}G(G+1) \times \frac{1}{2}G(G+1)$ square matrix.

We refer to the Matlab code how to construct $\partial \operatorname{vec} \tilde{\boldsymbol{L}}_t / \partial \tilde{\boldsymbol{\lambda}}_t^\top$. Here we continue with some aspects about $\partial \tilde{\boldsymbol{\lambda}}_t / \partial \boldsymbol{f}_t^\top$.

Note that each asset within a certain industry has a different number of factor loadings. For example, an asset from to industry 1, only has 1 factor loading $\tilde{\lambda}_1$, while an asset from industry G has G factor loadings $\tilde{\lambda}_G, \tilde{\lambda}_{2G-1}, \ldots, \tilde{\lambda}_{G(G+1)/2}$.

Setting $N_g = 1$ for each g in (F.19), we obtain a $G \times G$ matrix $\tilde{\boldsymbol{L}}_t^{gr}$ with unique elements $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{G(G+1)/2})$. Denote $\tilde{\boldsymbol{\lambda}}_{g,t}$ as column g from this matrix $\tilde{\boldsymbol{L}}_t^{gr}$. Then element g of the unique group specific $G \times 1$ denomimator vector $\boldsymbol{f}_{t,degr}$ is defined as

$$f_{t,g,degr} = 1 + \tilde{\boldsymbol{\lambda}}_{g,t}^\top \tilde{\boldsymbol{\lambda}}_{g,t} \tag{F.21}$$

The square matrix $\partial \tilde{\boldsymbol{\lambda}}_t / \partial \boldsymbol{f}_t^\top$ now consists of two building blocks. Suppose $\tilde{\lambda}_{t,j}$, $j = 1, \ldots, G(G+1)/2$, belongs to column g of $\tilde{\boldsymbol{L}}_t^{gr}$ Then

$$\frac{\partial \tilde{\lambda}_{t,j}}{\partial f_{t,j}} = \frac{1}{\sqrt{f_{t,g,degr}}} - \frac{f_{t,j}^2}{f_{t,g,degr}^{3/2}} \tag{F.22}$$

for all values of j.

Further we know that column g of $\tilde{\boldsymbol{L}}_t^{gr}$ contains $(g-1)$ other values of $\tilde{\lambda}_{t,j}$. Let indices $j, o = 1, \ldots, G(G+1)/2$ with $o \neq j$ both come from column g of $\tilde{\boldsymbol{L}}_t^{gr}$. The second building block now becomes

$$\frac{\partial \tilde{\lambda}_{t,j}}{\partial f_{t,o}} = -\frac{f_{t,o} f_{t,j}}{f_{t,g,degr}^{3/2}}. \tag{F.23}$$

Finally, we have

$$\frac{\partial \operatorname{diag} \boldsymbol{D}_t}{\partial \boldsymbol{f}_t^\top} = S^L \frac{\partial \boldsymbol{\sigma}_t^{2,gr}}{\partial \boldsymbol{f}_t^\top} \tag{F.24}$$

Similar as before, suppose that $f_{t,j}$ belongs to column g of $\tilde{\boldsymbol{L}}_t^{gr}$. Then $\sigma_{g,t}^2$ depends on exactly g different values of $f_{t,j}$. Now we have

$$\frac{\partial \sigma_{g,t}^2}{\partial f_{t,j}} = \frac{-2 f_{t,j}}{f_{t,g,degr}^2}. \tag{F.25}$$

for $g = 1, \ldots, G$, and zero otherwise.

References

Aielli, G. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business and Economic Statistics 31*, 282–299.

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

Creal, D., S. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*, 777–795.

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

Engle, R. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business Economic Statistics 30*, 212–228.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review 35*, 705–730.