Web Appendix to

Dynamic clustering of multivariate panel data*

André Lucas,[a] *Julia Schaumburg,*[a] *Bernd Schwaab,*[b]

[a] Vrije Universiteit Amsterdam and Tinbergen Institute

[b] European Central Bank, Financial Research

*Author information: André Lucas, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Email: a.lucas@vu.nl. Julia Schaumburg, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Email: j.schaumburg@vu.nl. Bernd Schwaab, European Central Bank, Sonnemannstrasse 22, 60314 Frankfurt, Germany, Email: bernd.schwaab@ecb.europa.eu. The views expressed in this paper are those of the authors and they do not necessarily reflect the views or policies of the European Central Bank or the Eurosystem.

A Derivation of the scaled scores

A.1 Time-varying mean dynamics

The scaled score for updating the time-varying j-th cluster mean $\boldsymbol{\mu}_{jt}$ is given by

$$\boldsymbol{s}_{\mu_{jt},t} = \boldsymbol{S}_{\mu_{jt},t} \cdot \boldsymbol{\nabla}_{\mu_{jt},t}, \tag{A.1}$$

where $\boldsymbol{S}_{\mu_{jt},t}$ is the scaling matrix and $\boldsymbol{\nabla}_{\mu_{jt},t}$ is the score of the predictive likelihood at time t. Starting with the score, and using the fact that $\tau_{ij,t|t-1}$ does not depend on $\boldsymbol{\mu}_{j,t}$ due to the transition probability matrix Π_t depending on the lagged cluster distances only as formulated in equations (2) and (6), we have

$$\begin{aligned}
\boldsymbol{\nabla}_{\mu_{jt},t} &= \frac{\partial \ell_t}{\partial \boldsymbol{\mu}_{jt}} = \frac{\partial \sum_{i=1}^{N} \log f\left(\boldsymbol{y}_{it}|\mathcal{F}_{t-1};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\mu}_{jt}}, \\
&= \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\mu}_{jt}} \log f\left(\boldsymbol{y}_{it}|\mathcal{F}_{t-1};\boldsymbol{\theta}\right) \\
&= \sum_{i=1}^{N} \frac{1}{f\left(\boldsymbol{y}_{it}|\mathcal{F}_{t-1};\boldsymbol{\theta}\right)} \cdot \frac{\partial}{\partial \boldsymbol{\mu}_{jt}} f\left(\boldsymbol{y}_{it}|\mathcal{F}_{t-1};\boldsymbol{\theta}\right) \\
&= \sum_{i=1}^{N} \frac{f(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1})\tau_{ij,t|t-1}}{f\left(\boldsymbol{y}_{it}|\mathcal{F}_{t-1};\boldsymbol{\theta}\right)} \cdot \frac{\frac{\partial}{\partial \boldsymbol{\mu}_{jt}} f\left(\boldsymbol{y}_{it}|\mathcal{F}_{t-1};\boldsymbol{\theta}\right)}{f(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1})\tau_{ij,t|t-1}} \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \frac{\frac{\partial}{\partial \boldsymbol{\mu}_{jt}} f\left(\boldsymbol{y}_{it}|\mathcal{F}_{t-1};\boldsymbol{\theta}\right)}{f(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1})\tau_{ij,t|t-1}} \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \frac{\frac{\partial}{\partial \boldsymbol{\mu}_{jt}} \sum_{j=1}^{J} \tau_{ij,t|t-1} f\left(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right)}{f(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1})\tau_{ij,t|t-1}} \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \frac{\frac{\partial}{\partial \boldsymbol{\mu}_{jt}} \left(\tau_{ij,t|t-1} \cdot f\left(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right)\right)}{f(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1})\tau_{ij,t|t-1}} \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \frac{\partial}{\partial \boldsymbol{\mu}_{jt}} \log \left(\tau_{ij,t|t-1} \cdot f\left(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right)\right) \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \frac{\partial}{\partial \boldsymbol{\mu}_{jt}} \log f\left(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right) \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \boldsymbol{\nabla}^{(j)}_{\mu_{jt},t}.
\end{aligned}$$

In case of a mixture of D-dimensional Student's t distributions, we have

$$f\left(\boldsymbol{y}_{it}|c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right)=\frac{\Gamma\left(\frac{\nu_j+D}{2}\right)}{\Gamma\left(\frac{\nu_j}{2}\right)\left(\pi\nu_j\right)^{D/2}|\Sigma_j|^{1/2}}\left(1+\frac{(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt})'\boldsymbol{\Sigma}_{jt}^{-1}(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt})}{\nu_j}\right)^{-(\frac{\nu_j+D}{2})}. \quad \text{(A.2)}$$

Taking derivatives of the log of (A.2), we obtain

$$\boldsymbol{\nabla}^{(j)}_{\mu_{jt},t}=w_{ij,t}\cdot\boldsymbol{\Sigma}_{jt}^{-1}\left(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt}\right), \quad \text{(A.3)}$$

where

$$w_{ij,t}=(1+\nu_j^{-1}\,D)\Big/\left(1+\nu_j^{-1}\left(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt}\right)'\boldsymbol{\Sigma}_{jt}^{-1}\left(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt}\right)\right). \quad \text{(A.4)}$$

Equation (A.3) contains the unscaled score. We scale the score by the weighted average of the conditional Fisher information matrices for the Gaussian setting $\nu_j^{-1}=0$, using the posterior probabilities $\tau_{ij,t|t}$ as weights; compare Lucas et al. (2019). We obtain

$$\begin{aligned}
\boldsymbol{S}^{-1}_{\mu_{jt},t} &= \sum_{i=1}^{N}\tau_{ij,t|t}\cdot\left(-\mathbb{E}\left[\frac{\partial\boldsymbol{\nabla}^{(j)}_{\mu_{jt},t}}{\partial\boldsymbol{\mu}'_{jt}}\,\middle|\,c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right]\right)\\
&= \sum_{i=1}^{N}\tau_{ij,t|t}\cdot\mathbb{E}\left[\boldsymbol{\nabla}^{(j)}_{\mu_{jt},t}\left(\boldsymbol{\nabla}^{(j)}_{\mu_{jt},t}\right)'\,\middle|\,c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right]\\
&= \sum_{i=1}^{N}\tau_{ij,t|t}\cdot\mathbb{E}\left[\boldsymbol{\Sigma}_{jt}^{-1}(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt})(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt})'\boldsymbol{\Sigma}_{jt}^{-1}\,\middle|\,c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right]\\
&= \sum_{i=1}^{N}\tau_{ij,t|t}\cdot\boldsymbol{\Sigma}_{jt}^{-1}\,\mathbb{E}\left[(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt})(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt})'\,\middle|\,c_{it}=j,\mathcal{F}_{t-1};\boldsymbol{\theta}\right]\boldsymbol{\Sigma}_{jt}^{-1}\\
&= \sum_{i=1}^{N}\tau_{ij,t|t}\cdot\boldsymbol{\Sigma}_{jt}^{-1}\,\boldsymbol{\Sigma}_{jt}\,\boldsymbol{\Sigma}_{jt}^{-1}\\
&= \sum_{i=1}^{N}\tau_{ij,t|t}\cdot\boldsymbol{\Sigma}_{jt}^{-1}, \quad \text{(A.5)}
\end{aligned}$$

where we used the fact that for the Gaussian case $w_{ij,t}=1$.

Inserting (A.5) and (A.3) into (A.1) yields the scaled score

$$
\begin{aligned}
\boldsymbol{s}_{\mu_{jt},t} &= \boldsymbol{S}_{\mu_{jt},t} \cdot \boldsymbol{\nabla}_{\mu_{jt},t} \\
&= \left(\sum_{i=1}^{N} \tau_{ij,t|t} \cdot \boldsymbol{\Sigma}_{jt}^{-1} \right)^{-1} \sum_{i=1}^{N} \tau_{ij,t|t} \cdot w_{ij,t} \cdot \boldsymbol{\Sigma}_{jt}^{-1} \left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt} \right) \\
&= \boldsymbol{\Sigma}_{jt} \boldsymbol{\Sigma}_{jt}^{-1} \left(\sum_{i=1}^{N} \tau_{ij,t|t} \right)^{-1} \sum_{i=1}^{N} \tau_{ij,t|t} \cdot w_{ij,t} \cdot \left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt} \right) \\
&= \frac{\sum_{i=1}^{N} \tau_{ij,t|t} \cdot w_{ij,t} \cdot \left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt} \right)}{\sum_{i=1}^{N} \tau_{ij,t|t}}.
\end{aligned}
$$

Transition equation (12) now follows directly.

A.2 Time-varying covariance matrix dynamics

The scaled score for the time-varying cluster covariance matrix parameters is

$$
\boldsymbol{s}_{\Sigma_{jt},t} = \boldsymbol{S}_{\Sigma_{jt},t} \cdot \boldsymbol{\nabla}_{\Sigma_{jt},t}, \tag{A.6}
$$

where $\boldsymbol{S}_{\Sigma_{jt},t}$ is the scaling matrix and $\boldsymbol{\nabla}_{\Sigma_{jt},t}$ is the score. The score is given by

$$
\boldsymbol{\nabla}_{\Sigma_{jt},t} = \frac{\partial \ell_t}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_{jt})} = \frac{\partial \left[\sum_{i=1}^{N} \ln \left(f \left(\boldsymbol{y}_{it} | \mathcal{F}_{t-1}; \boldsymbol{\theta} \right) \right) \right]}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_{jt})},
$$

where we can take the derivatives with respect to a general matrix $\boldsymbol{\Sigma}_{jt}$ rather than a symmetric matrix. Using the arguments in Proposition 3 of Opschoor et al. (2018), this gives the same steps for the free elements in $\boldsymbol{\Sigma}_{jt}$.

The initial derivations follow the same steps as for the time-varying mean; see Web Appendix A.1. Leaving these steps out, taking the log of (A.2) and omitting the terms that do not depend on Σ_{jt}, we arrive at

$$
\boldsymbol{\nabla}_{\Sigma_{jt},t} = \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \left(-\frac{\partial}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_{jt})} \frac{1}{2} \ln |\boldsymbol{\Sigma}_{jt}| - \frac{\partial}{\partial \mathrm{vec}(\boldsymbol{\Sigma}_{jt})} \left[\left(\frac{\nu_j + D}{2} \right) \ln \left(1 + \frac{(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt})' \boldsymbol{\Sigma}_{jt}^{-1} (\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt})}{\nu_j} \right) \right] \right).
$$

Following Abadir and Magnus (2005) for the derivative of the log of the determinant of the covariance matrix, and for the derivative of a matrix inside a quadratic form, and using $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, we obtain

$$
\begin{aligned}
\boldsymbol{\nabla}_{\Sigma_{jt},t} &= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \text{vec}\left(-\tfrac{1}{2}\left(\boldsymbol{\Sigma}_{jt}^{-1}\right)' + \tfrac{1}{2}\left(\boldsymbol{\Sigma}_{jt}^{-1}\right)' w_{ij,t}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)'\left(\boldsymbol{\Sigma}_{jt}^{-1}\right)'\right) \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \text{vec}\left(-\tfrac{1}{2}\left(\boldsymbol{\Sigma}_{jt}'\right)^{-1} + \tfrac{1}{2}\left(\boldsymbol{\Sigma}_{jt}'\right)^{-1} w_{ij,t}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)'\left(\boldsymbol{\Sigma}_{jt}'\right)^{-1}\right) \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \text{vec}\left(-\tfrac{1}{2}\boldsymbol{\Sigma}_{jt}^{-1} + \tfrac{1}{2}\boldsymbol{\Sigma}_{jt}^{-1} w_{ij,t}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)'\boldsymbol{\Sigma}_{jt}^{-1}\right) \\
&= \tfrac{1}{2}\sum_{i=1}^{N} \tau_{ij,t|t} \cdot \text{vec}\left(\boldsymbol{\Sigma}_{jt}^{-1}\left(w_{ij,t}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)' - \boldsymbol{\Sigma}_{jt}\right)\boldsymbol{\Sigma}_{jt}^{-1}\right) \\
&= \tfrac{1}{2}\left(\boldsymbol{\Sigma}_{jt} \otimes \boldsymbol{\Sigma}_{jt}\right) \cdot \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \text{vec}\left(w_{ij,t}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)' - \boldsymbol{\Sigma}_{jt}\right), \qquad \text{(A.7)}
\end{aligned}
$$

where the robustness weight $w_{ij,t}$ is defined in (A.4)

Next, we derive the scaling matrix, which we take as the weighted average of Fisher information matrices given $\nu_j^{-1} = 0$ for all j. We have

$$
\begin{aligned}
\boldsymbol{S}_{\Sigma_{jt},t}^{-1} &= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \mathbb{E}\left[\boldsymbol{\nabla}_{\Sigma_{jt},t}\boldsymbol{\nabla}_{\Sigma_{jt},t}' \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right] \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \left(-\mathbb{E}\left[\frac{\partial \boldsymbol{\nabla}_{\Sigma_{jt},t}}{\partial \text{vec}(\boldsymbol{\Sigma}_{jt})'} \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right]\right) \\
&= \sum_{i=1}^{N} \tau_{ij,t|t} \cdot \left(-\mathbb{E}\left[\frac{\partial}{\partial \text{vec}(\boldsymbol{\Sigma}_{jt})'}\tfrac{1}{2}\text{vec}\left(\boldsymbol{\Sigma}_{jt}^{-1}\left(\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)' - \boldsymbol{\Sigma}_{jt}\right)\boldsymbol{\Sigma}_{jt}^{-1}\right) \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right]\right) \\
&= -\tfrac{1}{2}\sum_{i=1}^{N} \tau_{ij,t|t} \cdot \mathbb{E}\left[\frac{\partial}{\partial \text{vec}(\boldsymbol{\Sigma}_{jt})'}\text{vec}\left(\boldsymbol{\Sigma}_{jt}^{-1}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)'\boldsymbol{\Sigma}_{jt}^{-1} - \boldsymbol{\Sigma}_{jt}^{-1}\right) \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right] \\
&= -\tfrac{1}{2}\sum_{i=1}^{N} \tau_{ij,t|t} \cdot \Bigg\{\mathbb{E}\left[-\left(\mathrm{I} \otimes \boldsymbol{\Sigma}_{jt}^{-1}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)'\right)\left(\boldsymbol{\Sigma}_{jt}^{-1} \otimes \boldsymbol{\Sigma}_{jt}^{-1}\right) \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right] + \\
&\qquad \mathbb{E}\left[-\left(\boldsymbol{\Sigma}_{jt}^{-1}\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{jt}\right)' \otimes \mathrm{I}\right)\left(\boldsymbol{\Sigma}_{jt}^{-1} \otimes \boldsymbol{\Sigma}_{jt}^{-1}\right) \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right] - \\
&\qquad \mathbb{E}\left[-\left(\boldsymbol{\Sigma}_{jt}^{-1} \otimes \boldsymbol{\Sigma}_{jt}^{-1}\right) \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right]\Bigg\} \\
&= \tfrac{1}{2}\sum_{i=1}^{N} \tau_{ij,t|t} \cdot \mathbb{E}\left[\left(\boldsymbol{\Sigma}_{jt}^{-1} \otimes \boldsymbol{\Sigma}_{jt}^{-1}\right) \,\middle|\, c_{it} = j, \mathcal{F}_{t-1}; \boldsymbol{\theta}\right] = \tfrac{1}{2}\sum_{i=1}^{N} \tau_{ij,t|t} \cdot \left(\boldsymbol{\Sigma}_{jt}^{-1} \otimes \boldsymbol{\Sigma}_{jt}^{-1}\right),
\end{aligned}
$$

where we used again vec $(ABC) = (C' \otimes A)\,\text{vec}\,(B)$, and $\partial\text{vec}(A^{-1})/\partial\text{vec}(A)' = -((A')^{-1} \otimes A^{-1})$ for a general matrix A. Pre-multiplying the score by the scaling matrix, we obtain the scaled score

$$\begin{aligned}
\boldsymbol{s}_{\Sigma_{jt},t} &= \left(\frac{1}{2}\sum_{i=1}^{N}\tau_{ij,t|t}\cdot\left(\boldsymbol{\Sigma}_{jt}^{-1}\otimes\boldsymbol{\Sigma}_{jt}^{-1}\right)\right)^{-1}\times \\
&\qquad \left(\frac{1}{2}\left(\boldsymbol{\Sigma}_{jt}\otimes\boldsymbol{\Sigma}_{jt}\right)\cdot\sum_{i=1}^{N}\tau_{ij,t|t}\cdot\text{vec}\left(w_{ij,t}\left(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt}\right)'-\boldsymbol{\Sigma}_{jt}\right)\right) \\
&= \frac{\sum_{i=1}^{N}\tau_{ij,t|t}\cdot\text{vec}\left(w_{ij,t}\left(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt}\right)\left(\boldsymbol{y}_{it}-\boldsymbol{\mu}_{jt}\right)'-\boldsymbol{\Sigma}_{jt}\right)}{\sum_{i=1}^{N}\tau_{ij,t|t}}. \qquad \text{(A.8)}
\end{aligned}$$

Now transition equation (21) follows directly.

B Sketch of $k-$means algorithm

The cluster probabilities $\tau_{ij,1|1}$, the cluster means $\boldsymbol{\mu}_{j1}$, and the cluster covariance matrices $\boldsymbol{\Sigma}_{j1}$ need to be initialized to start the filtering recursions as derived in Section 2 and Web Appendix A. In principle, we can initialize by any cross-sectional clustering algorithm using data of $t = 1$ only, $\boldsymbol{y}_{i1}$, for $i = 1, \ldots, N$. We initialize by k-means in our simulation study for simplicity; see Hartigan and Wong (1979). The k-means algorithm allocates N observations in D dimensions to $k = J$ clusters such that the within-cluster sum of squares is minimized. We sketch the steps below for completeness and ease of reference.

1. **Initialization:** initialize random centers for the J clusters in D dimensions.

2. **Assignment:** assign each observation, for a total of N observations, to the closest cluster according to Euclidean distance. $\tau_{ij,1|0} = \begin{cases} 1 \text{ for } \min_j \sqrt{(\boldsymbol{y}_{i1} - \boldsymbol{\mu}_{j1})' (\boldsymbol{y}_{i1} - \boldsymbol{\mu}_{j1})} \\ 0 \text{ else} \end{cases}$.

3. **Update:** recalculate the cluster centers as the average of the observations assigned to that cluster $\boldsymbol{\mu}_{j1} = \frac{\sum_{i=1}^{N} \tau_{ij,1|0} \cdot \boldsymbol{y}_{i1}}{\sum_{i=1}^{N} \tau_{ij,1|0}}$.

4. **Convergence 2:** return to step 2, and repeat until convergence of within-cluster sum of squared errors.

5. **Convergence 1:** return to step 1, and repeat 10 times for different initial random centers. Chose the one with minimal within-cluster sum of squared errors.

6. **Order** the clusters, e.g. in terms of declining averages for the first variable.

7. **Calculate initial covariance matrices:** estimate covariance matrix from the observations that were assigned to each cluster. $\boldsymbol{\Sigma}_{j1} = \frac{\sum_{i=1}^{N} \tau_{ij,1|0} \cdot (\boldsymbol{y}_{i1} - \boldsymbol{\mu}_{j1})(\boldsymbol{y}_{i1} - \boldsymbol{\mu}_{j1})'}{\sum_{i=1}^{N} \tau_{ij,1|0}}$.

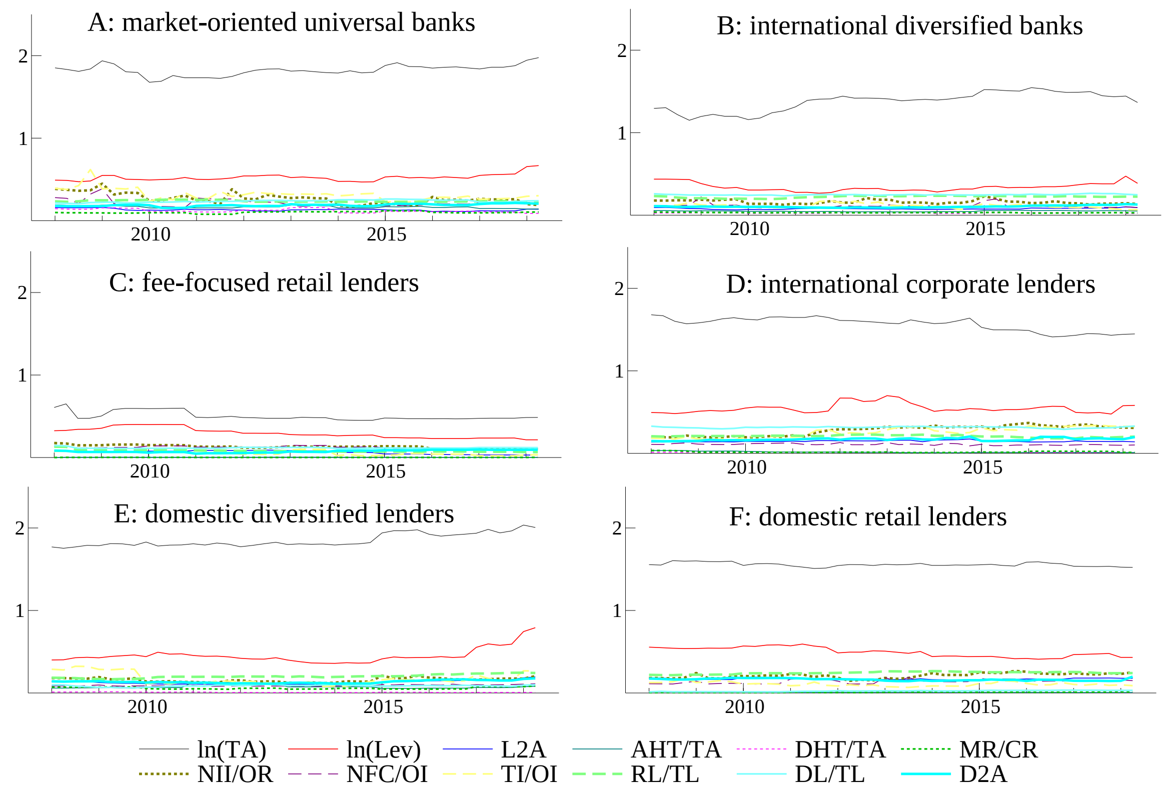
C Additional results

C.1 Estimated cluster standard deviations

Figure C.1 plots the filtered component-specific time-varying standard deviations $\hat{\sigma}_{j,t|t}(d) = \left(\hat{\Sigma}_{j,t|t}(d,d)\right)^{\frac{1}{2}}$ for variables $d = 1, \ldots, 12$. The first two variables, log total assets and log leverage, are the most dispersed across banks within each group A to F. Other variables, such as the share of assets held for trading, and the share of derivatives held for trading, are the least dispersed, particularly for banks in groups C to F.

Figure C.1: Time-varying standard deviations

Filtered time-varying standard deviations $\hat{\sigma}_{j,t|t}(d) = \left(\hat{\Sigma}_{j,t|t}(d,d)\right)^{\frac{1}{2}}$ for variables $d = 1, \ldots, 12$. Each panel contains 12 standard deviation estimates over time, corresponding to the variables listed in Table 3. The standard deviation estimates refer to model specification M4 in Table 4.

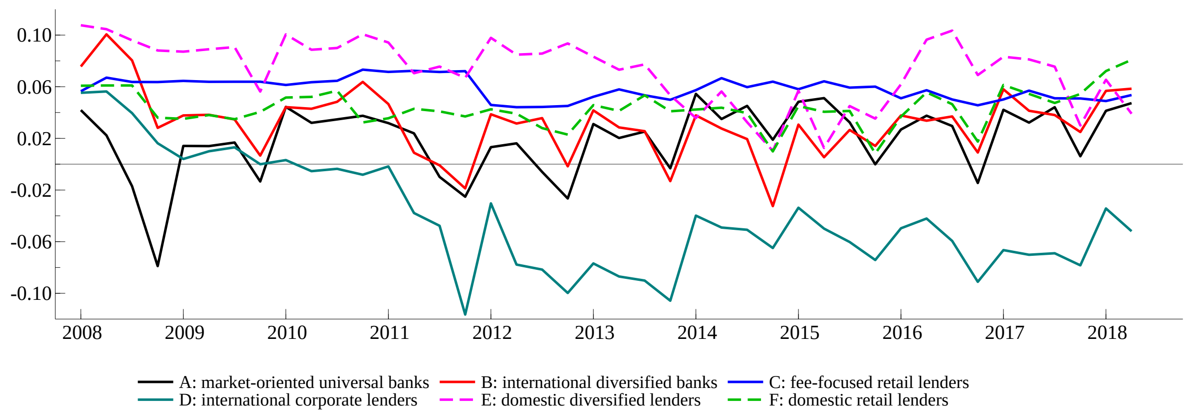


C.2 Bank profitability

The cluster transitions underlying Figures 2 – 4 are in part explained by differences in bank profitability. Figure C.2 below plots the return on equity (ROE) per cluster over time. Bank-specific observations ROE_{it} are weighted by the conditional probability $\tau_{ij,t|t}$ that bank i belongs to cluster j; see Section 4.2. ROE is not used as an input variable for the clustering; see Table 3. European banks' ROE tend to vary between approximately -2% and 12% over time. Banks assigned to cluster D (the "international corporate lenders") are an exception. Their ROE turns negative at onset of the euro area sovereign debt crisis in mid-2010, and remains negative until the end of the sample, adding to the move out of D to other business models.

Figure C.2: Bank profitability

Return on equity (ROE) per cluster. Bank-specific observations ROE_{it}s are weighted by the conditional probability $\tau_{ij,t|t}$ that bank i belongs to cluster j.

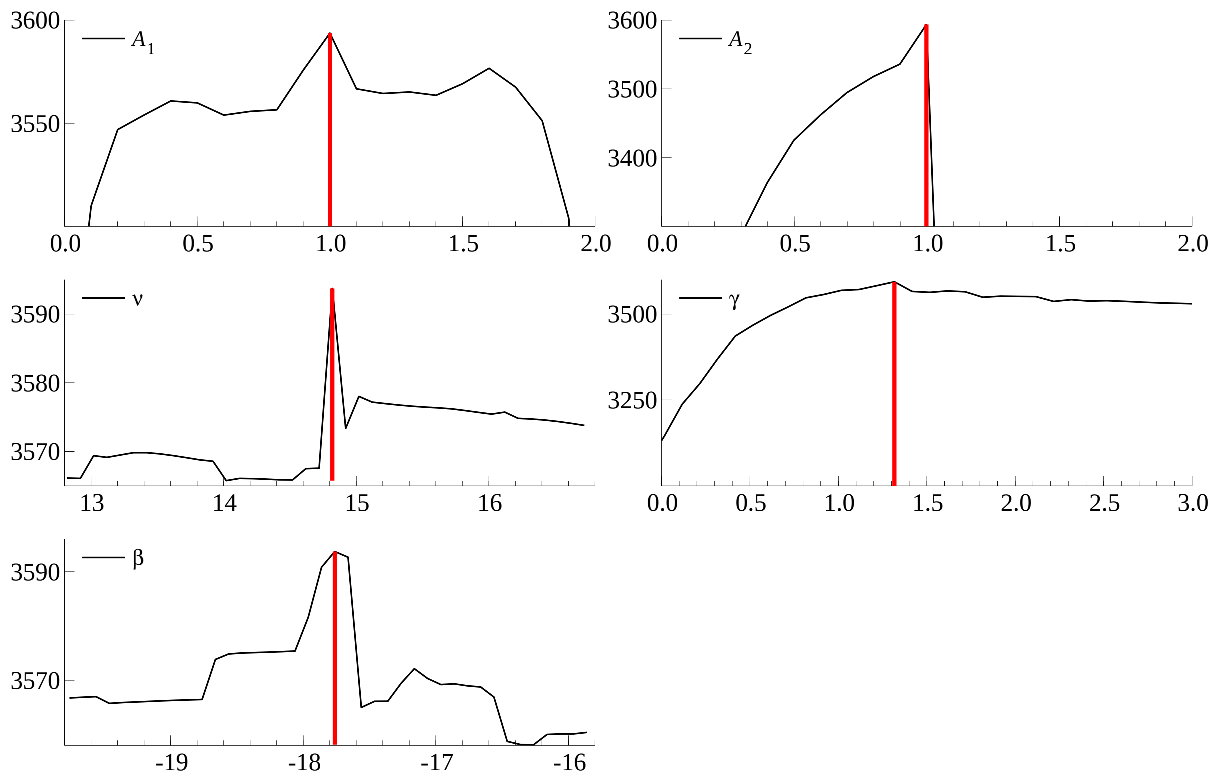


C.3 Likelihood slices

Mixture time series models such as ours can imply incalcitrant log-likelihood surfaces. Robust optimization methods such as simulated annealing (see, e.g., Goffe et al., 1994) can have an advantage over repeatedly re-running standard gradient-based optimization methods such as MaxBFGS in such cases. Figure C.3 plots directional slices through the log-likelihood function evaluated at the global optimum. Several local maxima are visible in which standard gradient-based methods are at risk of getting stuck.

Figure C.3: Log-likelihood slices

We report directional slices through the log-likelihood function (24) evaluated at the global optimum for $\theta = (A_1, A_2, \nu, \gamma, \beta)' \in \mathbb{R}^5$.



C.4 Filtered cluster probabilities and computer code

Most banks $i = 1, \dots, 299$ are allocated fairly unequivocally to one cluster j at any time t. A file containing banks' filtered membership probabilities $\hat{\tau}_{ij,t|t}$ is available from the authors. Computer code will be made available at https://www.gasmodel.com/code.htm.

References

Abadir, K. and J. Magnus (2005). *Matrix Algebra*. Cambridge University Press.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics 60*(1-2), 65–99.

Hartigan, J. A. and M. A. Wong (1979). A k-means clustering algorithm. *Applied Statistics 28*(1), 100–108.

Lucas, A., J. Schaumburg, and B. Schwaab (2019). Bank business models at zero interest rates. *Journal of Business & Economic Statistics 37*(3), 542–555.

Opschoor, A., A. Lucas, P. Januw, and D. J. van Dijk (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business and Economic Statistics 36*(4), 643–657.