Bank business models at zero interest rates*

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

[a] VU University Amsterdam and Tinbergen Institute

[b] European Central Bank, Financial Research

April 19, 2016

Abstract

We propose a novel observation-driven dynamic finite mixture model for the study of banking data. The model accommodates time-varying component means and covariance matrices, normal and t-distributed mixtures, and economic determinants of time-varying parameters. Monte Carlo experiments suggest that banks can be classified reliably into distinct components in a variety of settings. In an empirical study of 233 European banks between 2008Q1–2015Q2, we demonstrate that past crises had a differential impact on banks with different business models. In addition, changes in long-term interest rates predict features of banks' business models.

Keywords: bank business models; clustering; finite mixture model, score-driven model; low interest rates.

JEL classification: C33, G21.

---

*Author information: André Lucas, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Email: a.lucas@vu.nl. Julia Schaumburg, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Email: j.schaumburg@vu.nl. Bernd Schwaab, European Central Bank, Kaiserstrasse 29, 60311 Frankfurt, Germany, Email: bernd.schwaab@ecb.int. Lucas thanks the Dutch Science Foundation (NWO, grant VICI453-09-005) and the European Union Seventh Framework Programme (FP7-SSH/2007–2013, grant agreement 320270 - SYRTO) for financial support. Schaumburg thanks the Dutch Science Foundation (NWO, grant VENI451-15-022) for financial support. Parts of the paper were written while Schwaab was on secondment to the ECB's Single Supervisory Mechanism. We are grateful to Klaus Düllmann, Heinrich Kick, and Federico Pierobon for comments. 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.

1 Introduction

Banks are highly heterogenous, differing widely in terms of size, complexity, organization, activities, funding choices, and geographical reach. Understanding this diversity is of key importance, for example, for the study of risks originating from or acting upon the financial sector, impact assessments of newly proposed regulations and unconventional monetary policies, as well as the benchmarking of banks to appropriate peer groups for supervisory prudential purposes.[1] Ideally, the analysis of banks' business models provides insight into the overall diversity of business models, as well as the strategies adopted by individual institutions.

This paper proposes a novel observation-driven dynamic finite mixture model for the analysis of high-dimensional banking data. We first present a simple baseline dynamic mixture model for normally distributed data with time-varying component means, and subsequently consider relevant extensions to t-distributed mixture distributions, time-varying covariance matrices, and economic predictors of time-varying parameters. We then apply our modeling framework to a multivariate panel of $N = 233$ European banks between 2008Q1–2015Q2, $T = 30$, considering $D = 10$ bank-level indicator variables for J groups of similar banks. We thus track banking sector data through the 2008–2009 global financial crisis, the 2010–2012 euro area sovereign debt crisis, as well as the relatively calmer post-crises period between 2013–2015.

In our dynamic finite mixture model, all time-varying parameters are driven by the score of the mixture predictive log-likelihood. So-called Generalized Autoregressive Score (GAS) models were developed in their full generality in Creal, Koopman, and Lucas (2013); see also Harvey (2013) for a textbook treatment. In this setting, the time-varying parameters are perfectly predictable one step ahead. This feature makes the model observation-driven in the terminology of Cox (1981). The likelihood is known in closed-form through a standard prediction error decomposition, facilitating parameter estimation via likelihood-based

[1]For example, the assessment of the viability and the sustainability of a bank's business model plays a pronounced role in the European Central Bank's new Supervisory Review and Examination Process (SREP) for Significant Institutions within its Single Supervisory Mechanism; see SSM (2016).

expectation-maximization procedures.

Extensive Monte Carlo experiments suggest that our model is able to reliably classify a data set into distinct mixture components, as well as to infer the relevant component-specific time-varying parameters. In our simulations, the cluster classification is perfect for sufficiently large distances between the time-varying parameters and sufficiently informative signals relative to the variance of the noise terms.[2] This holds under correct model specification as well as under some degree of model misspecification. As the time-varying parameters become less informative, and the time-varying parameters become less distant, the share of correct classifications decreases, but generally remains high. Estimation fit, as well as the share of correct classifications, decrease further if we wrongly assume a Gaussian mixture specification when the data are generated from a mixture of fat-tailed Student's t distributions. As a result, robust models are appropriate if bank accounting ratios are fat-tailed.

We apply our model to classify European banks into six distinct business model components. We distinguish A) diversified cross-border lenders, B) small domestic lenders, C) small retail lenders, D) diversified lenders, E) wholesale/corporate lenders, and F) large universal banks. The similarities and differences between these components are discussed in detail in the main text. Based on our component mean estimates and business model classification, we confirm that the global financial crisis between 2008–2009 had a differential impact on banks with different business models, as argued in, for example, Altunbas, Manganelli, and Marques-Ibanez (2011), Beltratti and Stulz (2012), and Chiorazzo et al. (2016). We also observe such differences across business model components during the more recent euro area sovereign debt crisis between 2010–2012 for our sample of European banks. In particular, smaller domestic lenders and retail banks were relatively less affected.

Finally, we relate banks' business models to yield curve factors, specifically level and slope. The factors are extracted from AAA-rated euro area sovereign bonds based on a Svensson (1994) model. We find that, as long-term interest rates decrease, banks on average grow larger, fund themselves less with customer deposits, and hold more derivatives. Each of

[2]We use the terms 'mixture component' and 'cluster' interchangeably.

these effects – increased size, increased complexity through larger derivatives books, and less stable funding through customer deposits – are intuitive, but also potentially problematic from a financial stability perspective. In addition, we find that banks' share of net interest income to total income increases as long-term interest rates fall. Two opposing effects are at work here. First, banks funding cost also decrease, and typically do so at a faster rate than loan rates, supporting net interest income.[3] In addition, banks' long-term loans and bond holdings are worth more at lower rates. Such mark-to-market gains may be eventually realized. On the other hand, low long term interest rates squeeze net interest margins for *new* loans and bond holdings. The former effect dominates the latter in our sample. These short-term benefits, however, likely come at the expense of the long-term viability of established bank business models, and are in line with the pro's and con's of 'stealth recapitalization' as discussed in Brunnermeier and Sannikov (2015).

Allowing the level-factor impact on bank variables to differ across business model components increases the log-likelihood fit significantly, and provides additional insights. Interestingly, the response of bank size is largest for the smallest banks in our sample. Small domestic lenders (B) and small retail lenders (C) increase the size of their respective balance sheet by approximately 3% for every -100 basis points (bps) decrease in long term yields. This effect is intuitive: Small domestic banks and retail lenders are the most geographically constrained in extending new loans, implying an incentive to do more business at squeezed rates, potentially at declining underwriting standards. In addition, these lenders are highly customer deposit-funded. This feature is particularly painful at zero or negative short-term interest rates, as such rates often cannot be passed on to retail customers, for example owing to competitive pressure. Finally, small banks may currently not be as constrained by equity considerations, as such institutions were remarkably stable during the previous two crises (as argued above), and recent regulations as well as European single supervision, arguably, focus on too-individually-systemic-to-fail banks.

The two papers that are most closely related to ours are Ayadi and Groen (2015) and Catania (2016). Ayadi and Groen (2015) use cluster analysis to identify bank business

[3]Bank liabilities are typically of shorter duration than banks' interest-bearing assets.

models. Conditional on the identified clusters, the authors discuss bank profitability trends over time, study banking sector risks and their mitigation, and consider changes in banks' business models in response to new regulation. Our statistical approach is different in that our components are not identified based on single (static) cross-sections of year-end data. Instead, we consider a dynamic framework for a multivariate panel of N banks with D variables each, over $T > 1$. Catania (2016) proposes a score-driven dynamic mixture model which is similar to ours. His modeling framework is different in that the main focus is on the time series dimension, rather than on classifying a large cross-section. In addition, parameter estimation in Catania (2016) is not based on an EM algorithm, but instead relies on score-driven updates for all parameters. The advantage of our approach is that it is more likely to work well if the time dimension is short. In addition, it remains tractable when many components are considered.

We proceed as follows. Section 2 presents a baseline static and a score-driven dynamic finite mixture model. We then propose extensions to t-distributed mixture distributions, time-varying covariance matrices, and additional explanatory covariates. Section 3 discusses the outcomes of a variety of Monte Carlo simulation experiments. Section 4 applies the model to classify European financial institutions. Section 5 studies to which extent banks' business models adapt to an environment of exceptionally low interest rates. Section 6 concludes. The appendix provides further technical results.

2 Statistical model

2.1 Static finite mixture model and EM estimation

We consider multivariate panel data $\mathbf{y}_{it}$, for firms $i = 1, ..., N$, at time $t = 1, ..., T$. The data $\mathbf{y}_{it}$ are assumed to be a D-dimensional independent draw from a common mixture density $f(\mathbf{y}_{it})$ with J components. It is useful to view the data as incomplete; see, for example, McLachlan and Peel (2000). This view allows us to estimate all model parameters by maximum likelihood and a suitable expectation maximization (EM) algorithm. The complete data for each observation is

$$\mathbf{y}_{it}^c = (\mathbf{y}_{it}', \mathbf{z}_i')' \tag{1}$$

where $\mathbf{z}_i$ is a time-invariant J-dimensional selection vector that has a unit entry at position j if firm i belongs to component j and zeros elsewhere. The component indicator z_{ij} is 1 if firm i belongs to component j and zero otherwise.

The conditional probability that i belongs to component j given the data is

$$\begin{aligned}\tau_{ij} &:= \mathbb{P}\left(\text{unit } i \text{ belongs to component } j \mid \mathbf{y}_{i1}, ..., \mathbf{y}_{iT}\right) \\ &= \mathbb{P}(z_{ij} = 1 | \mathbf{y}_{i1}, ..., \mathbf{y}_{iT}) \\ &= \pi_j f_j(\mathbf{Y}_i)/f(\mathbf{Y}_i),\end{aligned} \tag{2}$$

where the scalar π_j denotes the mixing proportion of component j (or, alternatively, the unconditional probability that observation i belongs to component j), $\mathbf{Y}_i$ denotes the $(T \times D)$-matrix of observations on unit i at all time points, and the last line follows from Bayes' theorem $\mathbb{P}(A|B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)/\mathbb{P}(B)$ for events A and B.

If the components f_j belong to a parametric family, the mixture density can be written as

$$f(\mathbf{Y}_i; \mathbf{\Psi}) = \sum_{j=1}^{J} \pi_j f_j(\mathbf{Y}_i; \boldsymbol{\theta}_j) \tag{3}$$

with $\mathbf{\Psi} = (\pi_1, ..., \pi_{P-1}, \boldsymbol{\xi}')'$, where $\boldsymbol{\xi}$ contains the distinct parameters in $\boldsymbol{\theta}_j$, $j = 1, ..., J$. For example, in case of normal mixture components $\boldsymbol{\xi}$ contains the component means μ_j and the distinct parameters of the covariance matrices Σ_j. Because the data are independently and identically distributed, we can express the joint density in (3) as $f_j(\mathbf{Y}_i; \boldsymbol{\theta}_j) = \prod_{t=1}^{T} f_j(\mathbf{y}_{it}; \boldsymbol{\theta}_j)$.

A direct approach to estimating the unknown parameters would be to maximize the log likelihood function of the observed data

$$\log L(\mathbf{\Psi}) = \sum_{i=1}^{N} \log \left[\sum_{j=1}^{J} \pi_j f_j(\mathbf{Y}_i; \boldsymbol{\theta}_j) \right] \tag{4}$$

with respect to π_j, $j = 1, ..., J$ and $\boldsymbol{\xi}$. This is numerically infeasible in many relevant cases. Fortunately, it is admissible to approximate the likelihood function by its conditional expectation, see Dempster, Laird, and Rubin (1977). Convergence of the conditionally expected likelihood to the complete likelihood is ensured when the parameters are updated via the EM algorithm. In addition, the conditionally expected likelihood forms a tight lower bound at the optimum.

If the component indicator variables z_{ij} were known, the likelihood function would be given by

$$\log L_c(\boldsymbol{\Psi}) = \sum_{j=1}^{J} \sum_{i=1}^{N} z_{ij} \left[\log \pi_j + \log f_j(\mathbf{Y}_i; \boldsymbol{\theta}_j)\right]. \tag{5}$$

As the z_{ij} are unobserved, however, the complete data likelihood (5) is approximated by its conditional expectation, given the observed data and some initial or previously determined value $\boldsymbol{\Psi}^{(k-1)}$ for $\boldsymbol{\Psi}$,

$$\begin{aligned} Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k-1)}) &= \mathbb{E}\left[\log L_c(\boldsymbol{\Psi}) | \mathbf{y}; \boldsymbol{\Psi} = \boldsymbol{\Psi}^{(k-1)}\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{J} \sum_{i=1}^{N} z_{ij} \left[\log \pi_j + \log f_j(\mathbf{Y}_i; \boldsymbol{\theta}_j)\right] | \mathbf{Y}_1, ..., \mathbf{Y}_n; \boldsymbol{\Psi} = \boldsymbol{\Psi}^{(k-1)}\right] \\ &= \sum_{j=1}^{J} \sum_{i=1}^{N} \mathbb{P}[z_{ij} = 1 | \mathbf{Y}_1, ..., \mathbf{Y}_n, \boldsymbol{\Psi} = \boldsymbol{\Psi}^{(k-1)}] \left[\log \pi_j^{(k-1)} + \log f_j(\mathbf{Y}_i; \boldsymbol{\theta}_j^{(k-1)})\right]. \end{aligned} \tag{6}$$

The conditionally expected likelihood (6) can be optimized iteratively by alternately updating the conditional expectation of the component indicators ('E-Step') and maximizing the remaining part of the function ('M-Step'). Starting values are required for the density parameters $\boldsymbol{\theta}_j^{(0)}$ and the mixture probabilities $\pi_j^{(0)}$, $j = 1, ..., J$.

E-Step

The conditional component indicator probabilities are updated using (2),

$$\begin{aligned}
\tau_{ij}^{(k)} &:= \mathbb{P}[z_{ij} = 1|\mathbf{Y}_1, ..., \mathbf{Y}_n, \mathbf{\Psi} = \mathbf{\Psi}^{(k-1)}] \\
&= \frac{\pi_j^{(k-1)} f_j(\mathbf{Y}_i; \boldsymbol{\theta}^{(k-1)})}{f(\mathbf{Y}_i; \mathbf{\Psi}^{(k-1)})} \\
&= \frac{\pi_j^{(k-1)} f_j(\mathbf{Y}_i; \boldsymbol{\theta}^{(k-1)})}{\sum_{h=1}^{g} \pi_h^{(k-1)} f_h(\mathbf{Y}_i; \boldsymbol{\theta}_h^{(k-1)})}.
\end{aligned} \tag{7}$$

Once $\tau_{ij}^{(k)}$, $i = 1, ..., N$, $j = 1, ..., J$, are updated, we move to the M-Step.

M-Step

The M-Step maximizes $Q(\mathbf{\Psi}; \mathbf{\Psi}^{(k)})$ with respect to $\mathbf{\Psi}$, conditional on $\tau_{ij}^{(k)}$. The updated estimates of the mixture probabilities π_j are given by

$$\pi_j^{(k)} = \frac{1}{N} \sum_{i=1}^{N} \tau_{ij}^{(k)}, \tag{8}$$

for $j = 1, ..., J$. The parameters contained in $\boldsymbol{\xi}$ are updated by maximizing the remaining part of the conditional likelihood function. Sometimes, closed from solutions are available, as in the case of the normal finite mixture model. Otherwise, numerical maximization methods can be used. The E- and M-step are iterated until the difference $L(\mathbf{\Psi}^{(k+1)}) - L(\mathbf{\Psi}^{(k)})$ has converged.

2.2 Dynamic finite mixture model

This section considers a dynamic finite mixture model for panel data $\mathbf{y}_{it}$, for firms $i = 1, ..., N$, at time $t = 1, ..., T$. We here assume that the density is a mixed normal with time-varying D-dimensional component means μ_{jt}, $j = 1, ..., J$, and fixed covariance matrices Σ_j. The component means are updated using using the scaled score of the predictive log density of $\mathbf{y}_{it}$; see Creal, Koopman, and Lucas (2013), Harvey (2013), and Creal et al. (2014). For simplicity, we consider the integrated score-driven dynamics as in Lucas and Zhang (2015),

$$\mu_{j,t+1} = A_1 S_{\mu_{jt}} \nabla_{\mu_{jt}} + \mu_{jt}, \tag{9}$$

where A_1 is an unknown diagonal matrix to be estimated, and the innovation term $S_{\mu_{jt}} \nabla_{\mu_{jt}} =: s_{\mu_{jt}}$ is the (scaled) first derivative of the conditional log-likelihood at time t with respect to μ_{jt}.

Within the k-th EM iteration step, conditional on the component indicator probabilities τ_{ij}, the score has the form

$$\begin{aligned}\nabla_{\mu_{jt}} = \frac{\partial Q(\mathbf{\Psi}; \mathbf{\Psi}^{(k)})}{\partial \mu_{jt}} &= \frac{\partial}{\partial \mu_{jt}} \left(\sum_{i=1}^{N} \sum_{j=1}^{J} \tau_{ij} \left[\log \pi_j + \log \phi(\mathbf{y}_{it}; \mu_{jt}, \Sigma_j) \right] \right) \\ &= \frac{\partial}{\partial \mu_{jt}} \left(\sum_{i=1}^{N} \sum_{j=1}^{J} \tau_{ij} \left[\log \pi_j + (c - \frac{1}{2} (\mathbf{y}_{it} - \mu_{jt})' \Sigma_j^{-1} (\mathbf{y}_{it} - \mu_{jt})) \right] \right) \\ &= \Sigma_j^{-1} \sum_{i=1}^{N} \tau_{ij} \left(\mathbf{y}_{it} - \mu_{jt} \right),\end{aligned} \tag{10}$$

where ϕ denotes the multivariate normal density, and $c = -\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma_j|)$.

The inverse conditional Fisher information matrix is an appropriate choice for $S_{\mu_{jt}}$ in (9), see Creal et al. (2013). Consequently, we compute the expectation of the matrix of second derivatives of the objective function with respect to μ_{jt},

$$S_{\mu_{jt}} = \mathcal{I}^{-1} = -\mathbb{E} \left(\frac{\partial Q(\mathbf{\Psi}; \mathbf{\Psi}^{(k)})}{\partial \mu_{jt} \mu_{jt}'} \right)^{-1} = \left(\sum_{i=1}^{N} \tau_{ij} \right)^{-1} \Sigma_j. \tag{11}$$

This expression yields an explicit updating scheme for the time-varying component means,

$$\begin{aligned}\mu_{j,t+1} &= A_1 S_{\mu_{jt}} \nabla_{\mu_{jt}} + \mu_{jt} \\ &= A_1 \cdot \frac{\sum_{i=1}^{N} \tau_{ij} (\mathbf{y}_{it} - \mu_{jt})}{\sum_{i=1}^{N} \tau_{ij}} + \mu_{jt}.\end{aligned} \tag{12}$$

All model parameters can be estimated by a suitable EM algorithm. The E-Step is performed, as above, by calculating the conditional component indicator probabilities,

$$\tau_{ij}^{(k)} = \frac{\pi_j^{(k-1)} \phi_j \left(\mathbf{Y}_i; A_m^{(k-1)}, \mu_{j0}^{(k-1)}, \Sigma_j^{(k-1)} \right)}{\sum_{h=1}^{g} \pi_h^{(k-1)} \phi_h \left(\mathbf{Y}_i; A_1^{(k-1)}, \mu_{h0}^{(k-1)}, \Sigma_h^{(k-1)} \right)}, \tag{13}$$

where

$$\phi_j \left(\mathbf{Y}_i; A_1^{(k-1)}, \mu_{j0}^{(k-1)}, \Sigma_j^{(k-1)} \right) = \prod_{t=1}^{T} \phi_j \left(\mathbf{y}_{it}; A_1^{(k-1)}, \mu_{j0}^{(k-1)}, \Sigma_j^{(k-1)} \right),$$

and initial means $\mu_{j0}^{(k-1)}$. The M-Step maximizes the conditional likelihood given $\tau_{ij}^{(k)}$,

$$A_1^{(k)}, \mu_{0,1}^{(k)}, ..., \mu_{0,J}^{(k)}, \Sigma_{0,1}^{(k)}, ..., \Sigma_{0,J}^{(k)} = \text{argmax}_{A_1, \mu_{0,1}, ..., \mu_{0,J}, \Sigma_{0,1}, ..., \Sigma_{0,J}} \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{P} \tau_{ij}^{(k)} \left[-\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma_j|) - \frac{1}{2} (\mathbf{y}_{it} - \mu_{jt})' \Sigma_j^{-1} (\mathbf{y}_{it} - \mu_{jt}) \right] \right), \tag{14}$$

where μ_{jt} has the recursive structure (12). Again, the E-step and M-step are iterated until convergence.

2.3 Extensions

2.3.1 t-distributed mixture

This section robustifies the dynamic finite mixture model by considering panel data that are generated by mixtures of multivariate t-distributions.[4] Assuming a multivariate normal mixture is not always appropriate. For example, extreme tail observations can easily occur in the analysis of accounting ratios when the denominator is close to zero, implying pronounced changes from negative to positive values.

To robustly infer all model parameters, we use the fact that the probability density function (pdf) of a t-distributed random variable Y with ν degrees of freedom can be represented equivalently as the pdf of a random variable $Y = \mu + U^{-\frac{1}{2}} X$, where $X \sim N(0, \Sigma)$ is a D-dimensional Gaussian random vector, and where U follows a Gamma distribution

[4] For a textbook treatment of the static finite mixture model of multivariate t-distributions; see McLachlan and Peel (2000) Chapter 7. For an EM algorithm for the estimation of models with time-varying volatilities and correlations for one-component elliptical distributions, see also McNeil, Frey, and Embrechts (2005) and Zhang, Creal, Koopman, and Lucas (2011).

$U \sim \text{Gam}(\frac{\nu}{2}, \frac{\nu}{2})$ with pdf $\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \exp(-\frac{\nu}{2}u)$ where Γ denotes the Gamma function; see, for instance, Rao (2001) and McNeil et al. (2005). As a result, the finite mixture of t-distributions with time-varying component means and scale matrices

$$f(\mathbf{Y}_i; \mathbf{\Psi}) = \sum_{j=1}^{J} \pi_j f(\mathbf{Y}_i; \mu_{j1}, ..., \mu_{jT}, \Sigma_{j1}, ..., \Sigma_{jT}, \nu_j),$$

can be written in terms of unobserved time-invariant cluster indicators z_{ij}, and a second set of unobserved data $u_{1t}, ..., u_{Nt}$ for $t = 1, ..., T$ from the conditional distribution

$$U_{it}|z_{ij} = 1 \overset{iid}{\sim} \text{Gam}\left(\frac{\nu_j}{2}, \frac{\nu_j}{2}\right), \tag{15}$$

where $i = 1, ..., N$, $j = 1, ..., J$.

If both $u_{1t}, ..., u_{Nt}$ and z_{ij} were known for all t, the complete data log-likelihood function would consist of three parts

$$\log L_c = \log L_{1c} + \log L_{2c} + \log L_{3c}, \tag{16}$$

where

$$L_{1c} = \sum_{i=1}^{N} \sum_{j=1}^{J} z_{ij} [\log \pi_j]$$

$$L_{2c} = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{J} z_{ij} \left[-\log \Gamma\left(\frac{\nu_j}{2}\right) + \frac{\nu_j}{2} \log\left(\frac{\nu_j}{2}\right) - \log u_{it} + \frac{\nu_j}{2}(\log u_{it} - u_{it}) \right]$$

$$L_{3c} = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{J} z_{ij} \left[-\frac{D}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_{jt}| + \frac{D}{2} \log u_{it} - \frac{1}{2} u_{it} (\mathbf{y}_{it} - \mu_{jt})' \Sigma_{jt}^{-1} (\mathbf{y}_{it} - \mu_{jt}) \right].$$

We do not observe u_{it} and z_{ij} in practice, however. Consequently, we again approximate the complete data likelihood by its conditional expectation. An EM algorithm can be used to estimate this model as well. Appendix A provides the respective detailed expressions.

The scaled score function for the GAS update of component mean j becomes

$$s_{\mu_{jt}} = \frac{\sum_{i=1}^{N} \tau_{ij}^{(k)} u_{ijt}^{(k)} (\mathbf{y}_{it} - \mu_{jt})}{\sum_{i=1}^{N} \tau_{ij}^{(k)} u_{ijt}^{(k)}}, \tag{17}$$

where the additional scaling parameter corresponds to the conditional expectation of U_{it} given the data, $\mathbf{\Psi} = \mathbf{\Psi}^{(k-1)}$, and $z_{ij} = 1$, and is given by

$$u_{ijt}^{(k)} = \frac{1 + D/\nu_j^{(k-1)}}{1 + \sum_{t=1}^{T} (\mathbf{y}_{it} - \mu_{jt}^{(k-1)})' (\Sigma_{jt}^{(k-1)})^{-1} (\mathbf{y}_{it} - \mu_{jt}^{(k-1)}) / \nu_j^{(k-1)}}. \tag{18}$$

If the data are fat tailed, i.e. when ν is low, then $u_{ijt}^{(k)}$ takes lower values as well. Equation (17) suggests that this implies less volatile parameter updates over time, as outliers are automatically downweighted. On the other hand, for $\nu \to \infty$, u_{ijt} converge to one, and we recover the expressions for the Gaussian mixture model. The robustly scaled score for the component-specific scale matrices is given by

$$s_{\Sigma_{jt}} = \frac{1}{\sum_{i=1}^{N} \tau_{ij}^{(k)}} \left(\sum_{i=1}^{N} \tau_{ij}^{(k)} \left(u_{ijt}^{(k)} (\mathbf{y}_{it} - \mu_{jt})(\mathbf{y}_{it} - \mu_{jt})' - \Sigma_{jt} \right) \right). \tag{19}$$

2.3.2 Time-varying component covariance matrices

This section derives the scaled score updates for time-varying component covariance matrices Σ_{jt}. These matrices are also assumed to follow integrated GAS dynamics, and evolve over time as

$$\Sigma_{j,t+1} = A_2 S_{\Sigma_{jt}} \cdot \nabla_{\Sigma_{jt}} + \Sigma_{jt}, \tag{20}$$

where the scaled score $S_{\Sigma_{jt}} \cdot \nabla_{\Sigma_{jt}} =: s_{\Sigma_{jt}}$ is again defined as the first derivative of the expected likelihood function with respect to Σ_{jt}, scaled by the inverse of the conditional Fisher information matrix. In this case, the scaled score for a given component covariance matrix is given by

$$s_{\Sigma_{jt}} = S_{\Sigma_{jt}} \cdot \nabla_{\Sigma_{jt}} = \frac{1}{\sum_{i=1}^{N} \tau_{ij}} \cdot \sum_{i=1}^{N} \tau_{ij} \left[(\mathbf{y}_{it} - \mu_{jt})(\mathbf{y}_{it} - \mu_{jt})' - \Sigma_{jt} \right]. \tag{21}$$

The estimation of the model can be carried out using the EM algorithm as above, replacing Σ_j by Σ_{jt} in equations (13) – (14).

2.3.3 Explanatory covariates

This section extends the score-driven dynamics for the time-varying mean to include contemporaneous or lagged variables of economic variables as additional conditioning variables. For example, a particularly low interest rate environment may push financial institutions to grow larger and take riskier bets. Additional term structure factors as conditioning variables, for example, allow us to quantify such effects for different groups of banks.

A score-driven updating scheme with additional explanatory covariates builds on (12) and evolves over time as

$$\mu_{j,t+1} = \mu_{jt} + A_1 \cdot \frac{\sum_{i=1}^{N} \tau_{ij} (\mathbf{y}_{it} - \mu_{jt})}{\sum_{i=1}^{N} \tau_{ij}} + B_1 \cdot W_t, \tag{22}$$

where B_1 is a matrix of unknown coefficients to be estimated, and W_t contains economic covariates of interest. We consider the extended specification (22) in Section 5. The score-driven dynamics for time-varying component covariance matrices could be extended along similar lines.

3 Simulation study

3.1 Simulation design

This section investigates the ability of our dynamic mixture GAS model to *i*) correctly classify a data set into distinct components, and *ii*) recover the dynamic cluster means over time. We pay particular attention to the sensitivity of the EM algorithm to the number of units per cluster, the distinctiveness of the clusters, and the impact of model misspecification.

We simulate from a mixture of dynamic bivariate densities. These densities are composed of sinusoid mean functions, as well as i.i.d. disturbance terms that are drawn from a bivariate Gaussian distribution or a bivariate Student's t-distribution with five degrees of freedom.

The covariance matrices are chosen to be time-invariant identity matrices. The smoothing parameter A_1 is common to both clusters.

Visually, the simulated processes correspond to two data clouds. Each data cloud moves in circles according to their respective time-varying mean. To investigate the strengths and potential weaknesses of our method, we alter the characteristics of these circles. In total, we consider simulations from 16 different data generating processes.

The sample sizes are chosen to resemble our empirical application in Section 4. We thus keep the number of time points fixed at $T = 30$, and set the number of cross-sectional units equal to $N = 100$ or to $N = 400$. In our baseline setting, the two circles do not overlap. In addition, the data have a fairly large signal-to-noise ratios, in the sense that the radius is large relative to the variance of the error terms.

In other, more challenging settings, the circles overlap completely. In this setting the circles have the same center, and only differ in the orientation of the component means over time (clockwise vs. counterclockwise). Again, we decrease the radii of the circles while leaving the variance of the error terms unchanged. Finally, we investigate the impact of a particular type of model mis-specification by assuming a Gaussian mixture in the estimation procedure, although the data were generated from a mixture of Student's t-densities with five degrees of freedom.

Figure 1 illustrates our simulation setup with two examples. The data generating processes are plotted as a solid black line, together with the pointwise median of the estimated paths (over simulation runs, solid red line), and the filtered mean estimates across simulation runs (green triangles).

3.2 Simulation results

This section discusses our main simulation outcomes. Using our methodology, we estimate the component parameters from the simulated data. Parameters to be estimated include the initial values for the component mean processes, the distinct entries of the covariance matrices, and the smoothing parameter A_1.

We rely on mean squared error (MSE) statistics as our main measure of estimation fit.

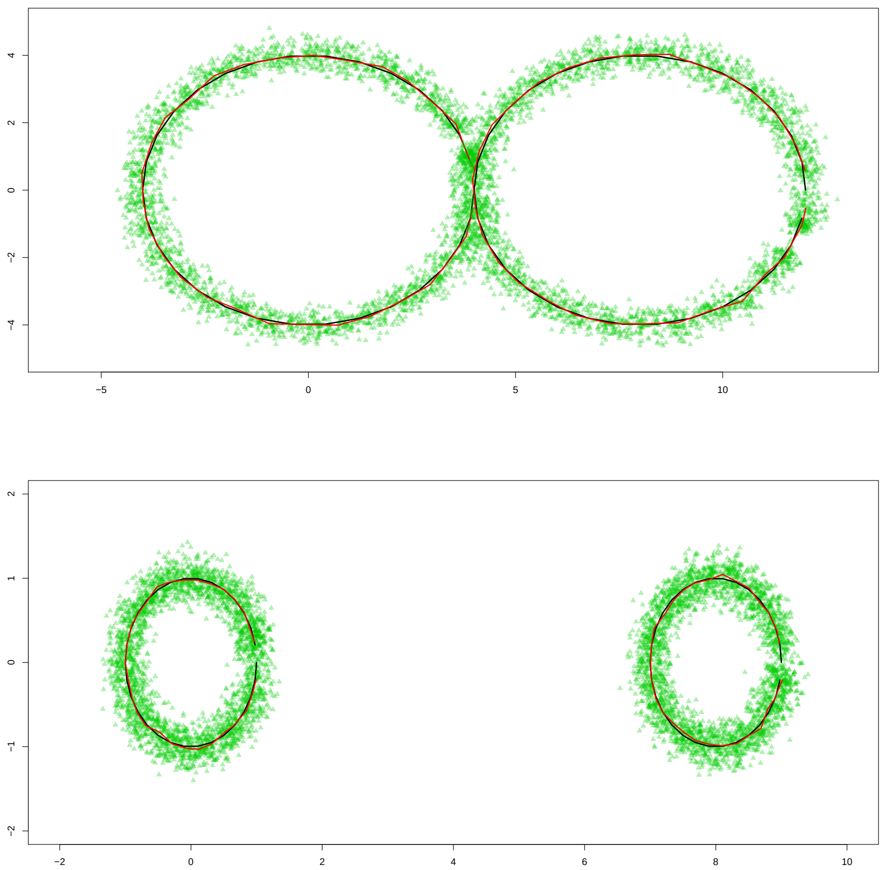

Figure 1: True mean processes (black) together with median filtered means over 100 simulation runs (red) and the filtered means (green triangles). Both panels correspond to simulation setups under correct specification with circle centers that are 8 units apart. The upper panel corresponds to the simulation setup with radius 4, while the lower panel depicts the mean circles with radius 1.

MSE statistics for time-varying component means are computed as the squared deviation of the estimated means from their true counterparts, averaged over time and simulation runs.

The top panel of Table 1 contains MSE statistics for eight simulation settings. Each of these settings considers $N = 100$ units per component. The bottom panel of Table 1 presents the same information for $N = 400$ units per component. In each case, we also report the proportion of correctly classified data points, averaged across simulation runs.

Not surprisingly, the performance of our estimation methodology depends on the simulation settings. For a high signal to-noise ratio, and a large distance between the unconditional means, the cluster classification is perfect both under correct specification and model mis-

Table 1: Simulation outcomes

Mean squared error (MSE) and average percentage of correct classification (%) across simulation runs. The top panel considers $N = 100$, while the bottom table corresponds to $N = 400$. Radius refers to the radius of the true mean circles and is a measure of the signal-to-noise ratio. Distance is the distance between circle centers and measures the distinctness of clusters.

correct specification					
		cluster 1		cluster 2	
radius	distance	MSE	%	MSE	%
4	8	0.21	100	0.21	100
4	0	0.2	100	0.21	100
1	8	0.04	100	0.04	100
1	0	0.04	99.97	0.04	99.96
misspecification					
		cluster 1		cluster 2	
radius	distance	MSE	%	MSE	%
4	8	0.25	100	0.25	100
4	0	0.25	100	0.25	100
1	8	0.05	100	0.05	100
1	0	0.06	98.56	0.06	97.51

correct specification					
		cluster 1		cluster 2	
radius	distance	MSE	%	MSE	%
4	8	0.14	100	0.14	100
4	0	0.12	100	0.12	100
1	8	0.02	100	0.02	100
1	0	0.02	100	0.02	99.99
misspecification					
		cluster 1		cluster 2	
radius	distance	MSE	%	MSE	%
4	8	0.16	100	0.16	100
4	0	0.16	100	0.16	100
1	8	0.03	100	0.03	100
1	0	0.05	97.94	0.05	96.33

specification. As the distance between means and the circle radii decrease, the share of correct classifications decreases as well. Estimation fit and the share of correct classification decrease further if we (wrongly) assume a Gaussian mixture although the data are generated from a mixture of fat-tailed Student's t distributions. Interestingly, in the case of large radii, the distance between circles is irrelevant for estimation fit and classification ability.

4 Bank business models at low interest rates

4.1 Data

The sample under study consists of $N = 233$ European banks, for which we consider quarterly bank-level accounting data from SNL Financial between 2008Q1 – 2015Q2. This implies $T = 30$. We assume that differences in bank business models can be characterized along six dimensions: size, complexity, activities, geographical reach, funding structure, and ownership. We select a parsimonious set of $D = 10$ indicators from these six categories. Table 2 lists the respective indicators.

Our panel data is unbalanced. As a result, missing values need to be taken into account. Missing values routinely occur because some banks are reporting at a quarterly frequency, while others report at an annual or semi-annual frequency. We remove such missing values by substituting the most recently available observation for that variable (backfilling). If variables are missing in the beginning of the sample, we use the most adjacent future value.[5] In the cross-section, we require a complete set of indicators for each bank; see Table 2.

We consider banks at their highest level of consolidation. In addition, we include large subsidiaries of bank holding groups in our analysis provided that a complete set of data is available in the cross-section. Most banks are located in the euro area (55%) and the European Union (EU, 66%). European non-EU banks are located in Switzerland (12%), Norway (10%), and other countries (12%).

[5]We experimented with different approaches to the treatment of missing values in our array of panel data, including using the EM algorithm of Stock and Watson (2002) to infer missing data at the indicator level. However, advanced econometric approaches to the treatment of missing values do not work robustly in our setting of $T = 30$, of which typically only a subset are observed.

Table 2: Indicator variables

Bank-level panel data variables for the empirical analysis. We consider J=10 indicator variables covering six different categories. The third column explains which transformation is applied to each indicator before the statistical analysis. $\Phi^{-1}(\cdot)$ denotes the inverse Probit transform.

Category	Variable	Transformation
Size	1. Total assets	$\ln \text{Total Assets}$
Complexity	2. Net loans to assets	$\Phi^{-1}\left(\frac{\text{Loans}}{\text{Assets}}\right)$
	3. Risk mix	$\ln\left(\frac{\text{Market Risk+Operational Risk}}{\text{Credit Risk}}\right)$
	4. Derivatives held for trading	$\ln(1+\text{DHT})$
Activities	5. Share of net interest income	$\frac{\text{Net interest income}}{\text{Operating revenue}}$
	6. Share of net fees & commission income	$\frac{\text{Net fees and commissions}}{\text{Operating income}}$
	7. Share of trading income	$\frac{\text{Trading income}}{\text{Operating income}}$
Geography	8. Domestic loans ratio	$\Phi^{-1}\left(\frac{\text{Domestic loans}}{\text{Loans}}\right)$
Funding	9. Loan-to-deposits ratio	$\frac{\text{Loans}}{\text{Deposits}}$
Ownership	10. Ownership	categorial

Note: Total Assets are all assets owned by the company (SNL id 131929). Net loans to assets are loans and finance leases, net of loan-loss reserves, as a percentage of all assets owned by the bank (226933). Risk mix is a function of Market Risk, Operational Risk, and Credit Risk (248881, 248882, 248880, respectively), and are as reported by the company. Derivatives held for trading are derivatives with positive replacement values not identified as hedging or embedded derivatives (224997). P&L variables are expressed as percentages of operating revenue (248959) or operating income (248961, 249289). Domestic loans are in percent of total loans by geography (226960). The loans-to-deposits ratio are loans held for investment, before reserves, as a percent of total deposits, the latter comprising both retail and commercial deposits (248919). Ownership combines information on ownership structure (131266) with information on whether a bank is listed at a stock exchange (255389). Ownership structure distinguishes stock corporations, mutual banks, co-op banks, and government ownership. Stock corporations can be listed or non-listed.

Table 2 also report the data transformation used in the applied modeling. For example, some ratios take values within the unit interval. We map these into approximately normally distributed variables by mapping them through an inverse Probit transform. In addition, we take natural logarithms of large numbers such as total assets and derivatives held for trading. The variable 'ownership' combines information on ownership structure with information on whether a bank is listed at a stock exchange. Ownership structure distinguishes stock corporations, mutual banks, co-op banks, and government ownership. Stock corporations can be listed or non-listed. We add a standard deviation of noise to the resulting categorical variable to make it continuous.

4.2 Model selection

This section motivates the model specification employed in our empirical analysis. We first discuss our choice of number of clusters. We then determine the parametric distribution, pooling restrictions, and choice of covariance matrix dynamics.

Table 3 presents likelihood-based information criteria as well as non-parametric cluster validation indexes for different values of $J = 2, \ldots, 10$. Different criteria point towards different numbers of relevant components. Interestingly, almost any choice $J \leq 6$ can be supported by some criterion. We therefore turn to additional considerations. First, experts consider up to eleven different bank business models in practise; see, for example, Bankscope (2012). Second, the degree of homogeneity in the resulting peer groups may be more important than model parsimony if the model is used for benchmarking purposes that are not related to forecasting future banking sector trends. The latter argument suggests putting particular emphasis on the within-cluster sum of squared errors (SSE) as scaled by the Hardigan index (HDG). For these considerations in mind, and to be conservative, we select $J = 6$ components for our subsequent empirical analysis.

In addition, we chose to work with a t-distributed dynamic finite mixture model instead of a Gaussian one; see Section 2.3.1). We also adopt dynamic cluster-specific covariance matrices Σ_{jt}, as specified in Section 2.3.2. Finally, we restrict $A_1 = a_1 \cdot I_D$, where a_1 is a single parameter and I_D is the D-dimensional identity matrix.

Table 3: Information criteria

We report (likelihood-based) information criteria and (non-parametric) cluster validation indexes for different values of $J = 2, \ldots, 10$. The top panel refers to a model specification with time-invariant component variance matrices Σ_j with ν estimated as a free parameter. The bottom panel refers to a specification with dynamic Σ_{jt}, with ν fixed at five. loglik is the maximum value of the log-likelihood. AICc is the likelihood-based AIC criterion, corrected by a finite sample adjustment; see Hurvich and Tsai (1989). AICk is a non-parametric AIC as suggested for k-means clustering; see Peel and McLachlan (2000). BNg1 and BNg2 are panel information criteria as derived for approximate dynamic factor models; see Bai and Ng (2002). CHI, Silh., and DBI refer to the Calinski-Harabasz index, average Silhouette, and Davies-Boulder index, respectively. SSE is the sum over within-component sum of squared errors. HGN is the Hardigan index which scales the increments in SSE by an adjustment term; see Peel and McLachlan (2000).

Σ_j static, estimated df.

J	loglik	AICc	AICk	BNg1	BNg2	CHI	Silh.	DBI	SSE	HGN
2	-49098.2	98468.9	5887.3	1.001	1.011	159.5	**0.32**	**0.98**	4955.3	83.6
3	-43811.6	88035.1	**5032.3**	0.815	**0.828**	**170.7**	0.24	1.21	3634.3	30.4
4	-39829.4	80211.7	5072.2	**0.814**	0.832	135.3	0.16	1.54	3208.2	12.1
5	-36849.3	74395.9	5377.1	0.885	0.908	107.0	0.16	1.64	3047.1	**-6.9**
6	-34712.7	70725.9	5938.4	1.040	1.067	81.9	0.09	1.83	3142.4	4.7
7	-33264.9	67523.9	6340.0	1.142	1.174	67.5	0.09	2.16	3078.0	16.1
8	-31412.9	63973.2	6600.1	1.196	1.233	64.7	0.07	2.25	2872.1	12.1
9	-30059.1	61421.4	6918.7	1.267	1.308	60.0	0.03	2.31	**2724.7**	-1.5
10	**-28751.2**	**58965.1**	7403.5	1.398	1.443	54.0	0.03	2.40	2743.5	-

Σ_{jt} dynamic, $\nu = 5$

J	loglik	AICc	AICk	BNg1	BNg2	CHI	Silh.	DBI	SSE	HGN
2	-40840.0	81951.9	5650.2	0.952	0.962	**176.4**	**0.34**	**1.02**	4718.2	41.8
3	-35388.0	71187.9	5390.5	0.909	0.922	135.6	0.22	1.28	3992.5	38.1
4	-30642.1	61837.1	**5287.2**	**0.878**	**0.897**	127.2	0.16	1.69	3423.2	10.8
5	-26773.7	54244.7	5598.0	0.955	0.978	99.8	0.10	1.98	3268.0	**8.1**
6	-23854.6	48553.5	5952.0	1.044	1.071	83.7	0.09	2.18	3156.0	**-16.8**
7	-20700.0	-87919.7	6671.0	1.244	1.276	61.5	0.06	2.66	3409.0	22.4
8	-18199.3	-109682.0	6828.8	1.273	1.310	60.7	0.02	2.75	3100.8	1.4
9	-13699.1	-162654.7	7276.2	1.390	1.431	51.9	0.01	2.56	**3082.2**	-5.6
10	**-6782.8**	**-195931.4**	7821.5	1.539	1.585	46.2	-0.01	2.78	3161.5	-

Table 4: Model specification

We report log-likelihoods and differences in log-likelihoods for a set of different model specifications. The estimates are based on $J = 6$.

Density	ν	value	A_1	Σ_j/Σ_{jt}	loglik	d(loglik)
N	-	∞	scalar	static	-36290.1	
t	fixed	10	scalar	static	-34952.2	-1337.9
t	fixed	5	scalar	static	-34726.9	-1563.2
t	fixed	5	vector	static	-34723.6	-3.3
t	est	6.8	scalar	static	-34712.7	-10.9
t	est	6.8	vector	static	-34709.6	-3.1
N	-	∞	scalar	dynamic	-30549.5	-4160.1
t	fixed	10	scalar	dynamic	-23854.6	-6694.9
t	**fixed**	**5**	**scalar**	**dynamic**	**-23577.3**	**-277.3**
t	est	5.9	scalar	dynamic	-23574.4	-2.9

Table 4 supports these empirical choices. Commencing from the top row to the bottom, the log-likelihood improves significantly as we move from the Gaussian to a t-distributed finite mixture model specification, while keeping Σ_j static. The data appears to favor a t-distributed model with low degrees of freedom, even after conditioning on component membership. Pooling the diagonal elements of A_1 into a single parameter decreases the likelihood fit only insignificantly. We therefore adopt this restriction, saving nine parameters to be estimated. The adoption of dynamic covariance matrices Σ_{jt} leads to further large improvements in log-likelihood fit.

4.3 Component analysis

This section studies the different business models implied by the $J = 6$ different component densities.

Figure 2 plots the filtered mean estimates for each business model component and indicator variable. Visually, the six component means differ frequently and substantially, for each indicator. For example, banks are impacted differently by the global financial crisis between 2008–2009. At that time, component means for the share of trading income to operating income are negative for approximately half of the components, positive for the other components, and in all cases are substantially different from their respective historical

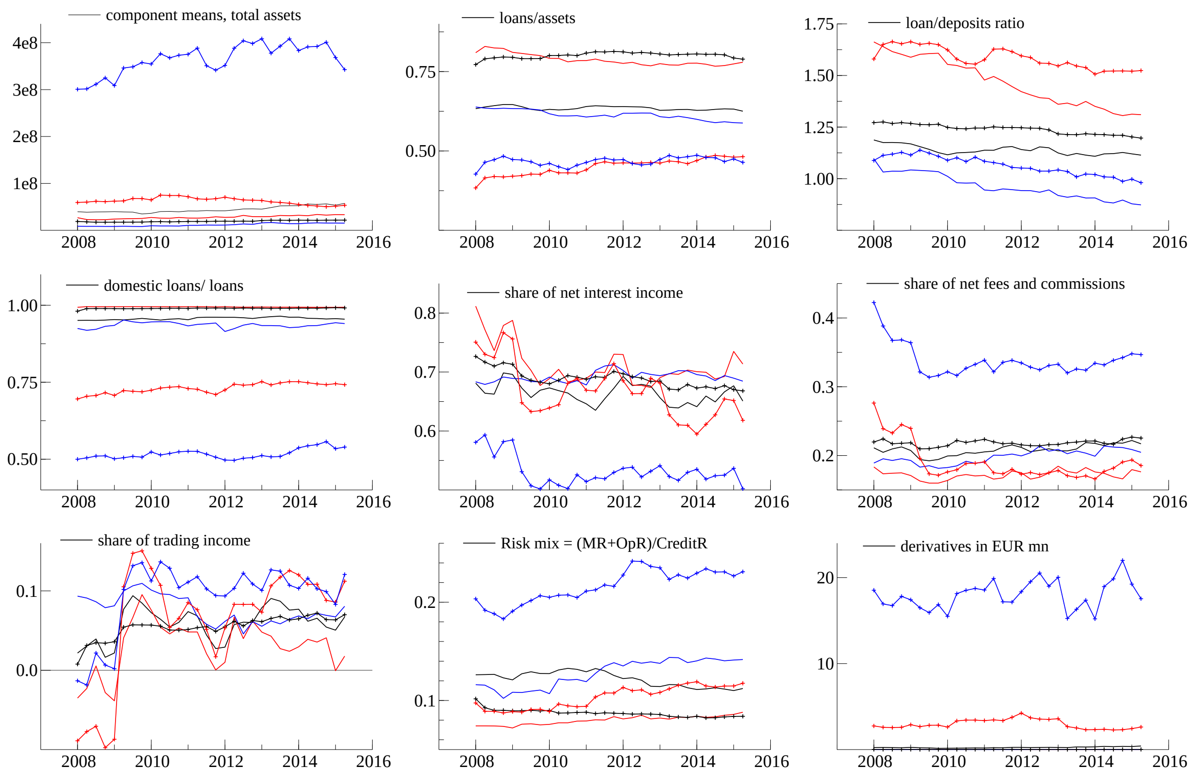

Figure 2: Time-varying component means

Filtered component means for nine indicator variables; see Table 2. The ownership variable is omitted since it is time-invariant. Mean estimates are based on a t-mixture model with $J = 6$ components and dynamic component covariance matrices Σ_{jt}. We fixed $\nu = 5$ to achieve outlier-robust results. We distinguish diversified x-border lenders (black solid line), small domestic lenders (red solid line), small retail lenders (blue solid line), diversified lenders (black crossed line), wholesale/corporate lenders (red crossed line), and large universal banks (blue crossed line).

averages. Large universal banks are most affected. By contrast, small domestic lenders and retail lenders experience less fluctuation, particularly in the share of income sources.

We assign labels to the identified components for ease of later reference. The assignment of these labels is based on Figure 2 and 3, as well as the identity of the firms in each component. The labeling is approximately in line with business models as listed in SSM (2016).

(A) diversified cross-border lender (black solid line)

(B) small domestic lender (red solid line)

(C) small retail lender (blue solid line)

Figure 3: Box plots for indicator variables
We report box plots for nine indicator variables, for six business model components; see Table 2 for the description. Each variable refers the time series average over $T = 30$ quarters. In each panel, we distinguish (from left to right): diversified x-border lenders, small domestic lenders, small retail lenders, diversified lenders, wholesale/corporate lenders, and large universal banks.

(D) diversified lender (black crossed line)

(E) wholesale/corporate lender (red crossed line)

(F) large universal bank (blue crossed line)

Large universal banks stand out in Figures 2 and 3. These banks are the largest firms (with up to approximately €2.5 trn in total assets per firm), the most international in terms of cross-border lending, with significant shares of trading as well as net fees & commission income, and large derivative positions. Wholesale/corporate lenders are the second-largest group in terms of total assets. These lenders are mostly non-deposit (wholesale) funded, obtain approximately two-thirds of their income from interest-bearing assets, hold a mediumsized derivatives book, and are second in terms of cross-border activities.

Diversified cross-border lenders (black solid line) are the third-largest group. They differ from diversified lenders (black crossed line) in that the latter are purely domestic in terms of their geographical orientation. The latter also tend to hold more loans (as a share of total assets), and tend to be slightly more deposit funded.

Small retail lenders and small domestic lenders are the smallest group of banks in terms of total assets. They are, however, numerous. Both types of banks are approximately similar in terms of size, absence of derivative holdings, and income composition. They differ in terms of size (the domestic lenders are somewhat larger), propensity to give loans (particularly the loans-to-assets ratio), and their tendency to be funded by customer deposits. Small domestic lenders are entirely local in terms of geographical orientation, while retail lenders have some cross-border exposures.

Figure 4 reports the filtered time-varying standard deviations around these time-varying means. The standard deviations are approximately stable over time, with the possible exceptions of 'derivatives held for trading' and size (ln total assets). Despite these stable second moments, allowing for time-varying component covariance matrices still results in significant increases in log-likelihood; see Table 4. This increase in fit suggests that higher-order moments may play a role. In addition, the off-diagonal elements of Σ_{jt} are mildly time-varying.

5 Term structure factors as explanatory covariates

This section studies to which extent banks' business models adapt to an environment of exceptionally low interest rates. To this purpose we consider the extended model as discussed in Section 2.3.3, adding two term structure factors, level and slope, as additional economic covariates. The extended specification allows us to quantify whether term structure factors contribute to explaining banks' business models.

Level and slope factors refer to the yield curve of AAA-rated euro area government bonds from 2008Q1 to 2015Q2. The factor estimates are based on a Svensson (1994) model, and are

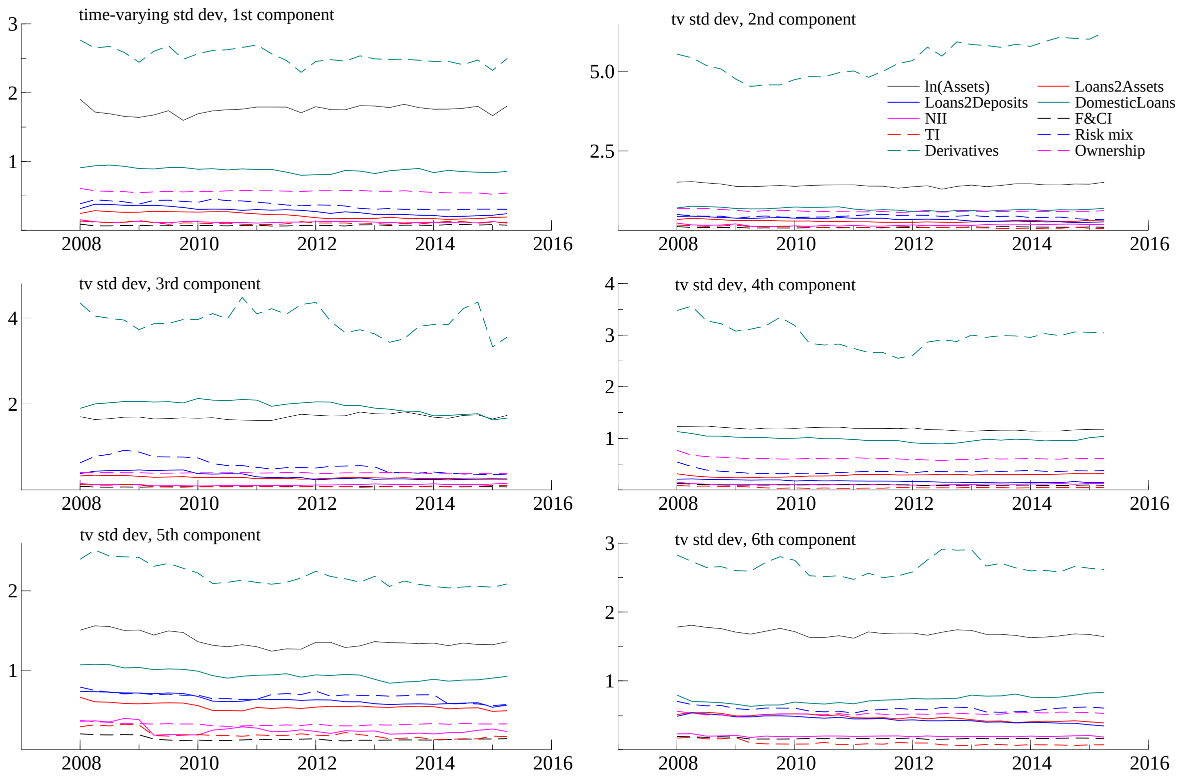

Figure 4: Time-varying standard deviations

Filtered time-varying standard deviations around time-varying means as graphed in Figure 2. Each panel refers to a component $j = 1, \ldots, 6$, and contains $D = 10$ standard deviation estimates over time. Mean and standard deviation estimates are based on a t-mixture model with six components and dynamic covariance matrices Σ_{jt}. We fixed $\nu = 5$ to achieve outlier-robust results.

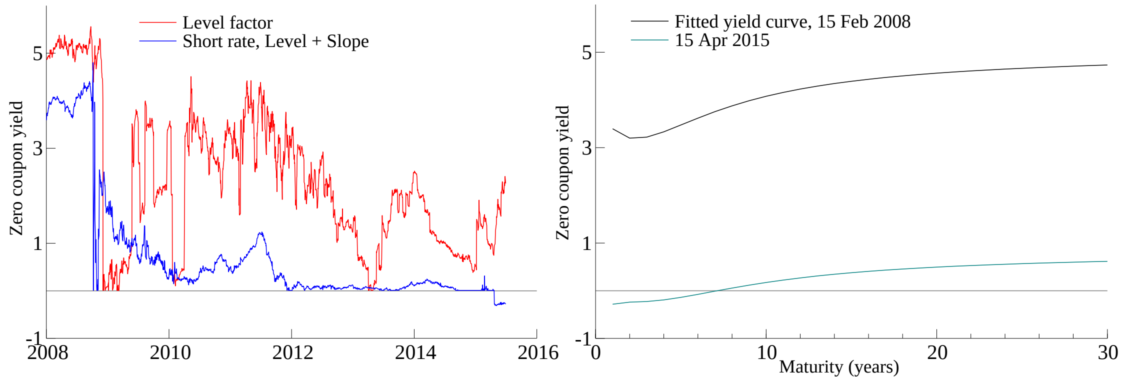

Figure 5: Yield curve factors and estimates

The left panel plots the level factor from a Svensson (1994) model, along with the model-implied short rate as given by the sum of level and slope. Yield curve factors refer to AAA-rated euro area government bonds from 2008Q1 to 2015Q2. The right panel plots two fitted yield curves on two dates, in mid-2008Q1 and mid 2015Q2, based on all factors.

publicly available from the ECB's website. We take these euro area yields as representative of low-risk European yields more generally.[6] The left panel of Figure 5 plots the level factor, and the short rate as implied by the sum of level and slope. The slope factor fluctuates relatively steadily around a value of -2 in our sample, and is not plotted. The right panel of Figure 5 plots two yield curves on two dates, in mid-2008Q1 and mid 2015Q2, as implied by all four factors. AAA-rated government bond yields dropped sharply in late 2008, from approximately 5% to below 1%. Long-term yields rose again during the euro area sovereign debt crisis, up to a peak of 4% between 2010–2012, and steady declined to close to zero between 2012–2015.

Table 5 presents the parameter estimates of two GAS-X specifications. We set $A_1 = A_2 = 1$. As a result, only the term structure factors predict the component means.

We focus on the level factor specification first. All signs are fairly intuitive. First, as long-term interest rates decrease, banks on average grow larger. This is in line with an incentive to extend the balance sheet to offset squeezed net interest margins for new loans and investments. Second, banks fund themselves less with customer deposits (a negative coefficient on the loans-to-deposits ratio), as deposit rates do not drop as much as long term rates, and do not go negative. Third, banks appear to hold more derivatives. We conjecture that most 'derivatives held for trading' are interest rate swaps, and are either traded on behalf of the bank or its clients to hedge against future moves in the term structure of interest rates. Each of these effects - increased size, increased complexity through larger derivatives books, and less stable funding through customer deposits – are potentially problematic and need to be assessed from a financial stability perspective.

In addition, declining long-term interest rates predict a declining share of net fees & commission income to total operating income in the next quarter. This is, again, intuitive. The drop in long term interest rates is in part a consequence of the global and euro area economic malaise during a large part of our sample. Such an environment is not conducive to charging high asset management fees, obtaining commissions from equity and bond placements, or

[6] The evolution of low-risk euro area yields are directly relevant for the euro area subset of banks in our sample. In addition, European sovereign bond yields are highly correlated across borders, and may in part reflect global developments as well; see, for instance, ECB (2013) and Lucas et al. (2014).

Table 5: GAS-X parameter estimates

We report parameter estimates for model specification (22), where term structure factors serve as additional conditioning variables. Standard errors are obtained from the numerical second derivative of the likelihood function.

		GAS-X			
		par	t-val	par	t-val
A_1		1.00	-	1.00	-
A_2		1.00	-	1.00	-
B_1 level	ln(TA)	**-2.27**	**-3.42**	-0.94	-1.05
	L2Assets	-0.24	-1.69	-0.16	-0.80
	L2Deposits	**-0.71**	**-5.42**	-0.31	-1.62
	Domestic	0.48	1.36	0.00	0.00
	NII	**-0.36**	**-5.71**	**-0.24**	**-2.79**
	F&CI	**0.22**	**5.24**	**0.17**	**2.98**
	TI	0.07	1.53	0.05	0.70
	Risk Mix	0.16	0.75	-0.06	-0.21
	Derivatives	**-3.31**	**-2.80**	-0.52	-0.32
	Ownership	0.00	-0.20	-0.06	-0.84
B_1 slope	ln(TA)			0.02	0.25
	L2Assets			0.00	0.01
	L2Deposits			0.02	1.38
	Domestic			-0.02	-0.35
	NII			0.01	1.06
	F&CI			0.00	-0.65
	TI			0.00	-0.58
	Risk Mix			-0.03	-1.00
	Derivatives			0.15	0.89
	Ownership			0.00	0.13
mlik		-23319.2		-23198.0	

providing advisory services.

Finally, as long-term interest rates fall, banks' share of net interest income increases. Two effects may be at work here. First, banks' long-term loans and bond holdings are worth more at lower rates. This may lead to mark-to-market gains, which are eventually realized. In addition, banks funding cost also decrease, and at a faster rate than long-term loan rates. On the other hand, low long term interest rates squeeze net interest margins for *new* loans and bond holdings. Our results suggest that the former effect dominates the latter in our sample. These short-term benefits (in line with a 'stealth recapitalization' of the banking sector) may, however, come at the expense of the long-term viability of established business models; see Brunnermeier and Sannikov (2015).

The right hand column of Table 5 presents parameter estimates associated with both level and slope factors. The signs of the significant level factor coefficients remain unchanged when the slope factor is included. However, the slope factor does not appear to contribute much explanatory power for banks' business models. The respective coefficient t-values are insignificant. This is not too surprising. Short-term rates are low and close to zero during most of our sample. A one (level) factor specification for the term structure appears to be sufficient to capture both the decline in long term rates as well as the overall flattening of the curve.

Table 6 presents component-specific level factor GAS-X coefficients. We focus on three findings. First, allowing the level-factor impacts to be different across business model components increases the log-likelihood fit significantly, despite adding $5 \times 9 = 45$ coefficients. On the other hand, it also implies less precise estimates and diminished statistical power. Most coefficients now have t-values below 1.64 in absolute value. The significant entries are in line with the estimates reported in Table 5: As long-term interest rates fall, the share of net interest income to total income increases, across all business model components. Similarly, the share of fees & commission income falls if long-term rates decline, again in line with our previous findings.

Second, the response of bank size (total assets) is largest for the smallest banks. Small small domestic lenders (B) and retail lenders (C) increase the size of their balance sheet by

Table 6: Component-specific GAS-X estimates

We report parameter estimates for model specification (22), where the level factor serves as an additional conditioning variable. Level factor sensitivities are component-specific. The ownership variable is time-invariant, and therefore does not have a coefficient. Parameters with t-values larger than 1.64 are in bold. Standard errors are obtained from the numerical second derivative of the likelihood function.

GAS-X, component-specific impact						
		par	t-val		par	t-val
ln(TA)	A	-0.82	-0.39	D	-1.59	-0.85
L2Assets		0.09	0.37		-0.47	-1.15
L2Deposits		0.05	0.15		0.07	0.26
Domestic		-0.04	-0.04		-0.83	-0.54
NII		-0.05	-0.32		**-0.36**	**-2.46**
F&CI		0.02	0.17		**0.24**	**1.78**
TI		-0.02	-0.18		-0.02	-0.46
Risk Mix		0.31	0.70		0.62	1.16
Derivatives		-0.69	-0.23		0.57	0.13
Ownership		-	-		-	-
ln(TA)	B	-3.30	-1.45	E	-1.23	-0.67
L2Assets		-0.20	-0.42		-0.32	-0.43
L2Deposits		-0.72	-1.56		0.23	0.26
Domestic		-0.71	-0.63		0.18	0.14
NII		**-0.56**	**-2.45**		-0.09	-0.25
F&CI		**0.36**	**3.38**		0.04	0.26
TI		0.07	0.46		-0.08	-0.32
Risk Mix		0.12	0.17		-1.08	-1.13
Derivatives		4.09	0.48		-2.12	-0.70
Ownership		-	-		-	-
ln(TA)	C	-2.83	-0.94	F	-2.49	-0.92
L2Assets		0.49	0.97		0.15	0.21
L2Deposits		0.25	0.48		0.03	0.04
Domestic		-1.93	-0.55		-0.56	-0.47
NII		-0.16	-0.83		-0.02	-0.08
F&CI		0.02	0.17		0.02	0.09
TI		0.07	0.49		0.02	0.17
Risk Mix		-0.79	-0.85		-0.35	-0.37
Derivatives		3.10	0.42		-3.55	-0.87
Ownership		-	-		-	-
mlik			-22727,50			

approximately 3% for every -100 basis points (bps) drop in long term yields. Small domestic lenders are the most geographically constrained in terms of lending, and are highly deposit-funded; see Figure 2. In addition, small domestic lenders may not be as strictly supervised, as such institutions were remarkably stable during the previous crises between 2008–2009 and 2010–2012.

Finally, the largest banks tend to increase their derivatives positions as long-term rates fall. The point estimates for wholesale/corporate lenders (E) and large universal banks (F) suggest that a -100 bps decrease in rates increases the size of their derivatives book by 2–4%. This increase likely reflects an incentive by either the bank or its customers to hedge against sharply rising future rates. It is also consistent with a smaller size response of these firms compared to the smaller lenders (B) and (C) which may be hedged to a lesser extent.

6 Conclusion

We proposed a novel score-driven dynamic finite mixture model for the study of banking data, accommodating time-varying component means and covariance matrices, normal and t-distributed mixtures, and term structure factors as economic determinants of time-varying parameters. In an empirical study of European banks, we classified numerous institutions into distinct business model components. Our results suggest that the global financial crisis and the euro area sovereign debt crisis had a differential impact on banks with different business models. In addition, banks' business models adapt over time to changes in long-term interest rates.

Appendix: EM algorithm for t-distributed panel data

This appendix discusses parameter estimation in the context of a dynamic finite mixture panel data model with t-distributed data. We build on Chapter 7 of Peel and McLachlan (2000) who consider the case of constant component mean and variance parameters.

E-Step

Taking the conditional expectation of the complete data likelihood (16) requires calculating

$$\mathbb{E}[Z_{ij}|\mathbf{Y}_1,...,\mathbf{Y}_T,\mathbf{\Psi}=\mathbf{\Psi}^{(k)}] =: \tau_{ij}^{(k-1)},$$
$$\mathbb{E}[U_{it}|\mathbf{Y}_1,...,\mathbf{Y}_T,z_{ij}=1,\mathbf{\Psi}=\mathbf{\Psi}^{(k-1)}] =: u_{ijt}^{(k)}, \text{ and}$$
$$\mathbb{E}[\log U_{it}|\mathbf{Y}_1,...,\mathbf{Y}_T,z_{ij}=1,\mathbf{\Psi}=\mathbf{\Psi}^{(k-1)}],$$

where $i=1,...,N$, and $\mathbf{\Psi}=(\pi_1,...,\pi_{J-1},A_1,A_2,\boldsymbol{\theta}'_{10},...,\boldsymbol{\theta}'_{J0},\nu_1,...,\nu_J)'$, as well as $\boldsymbol{\theta}_{j0}$, $j=1,...,J$ collect the initial mean and covariance matrix parameters.

We provide closed-form solutions for the above three expectations. The cluster probabilities can be updated using Bayes' rule,

$$\tau_{ij}^{(k)} = \frac{\pi_j^{(k-1)} f_j\left(\mathbf{Y}_i; A_1^{(k-1)}, \mu_{j0}^{(k-1)}, \Sigma_{j0}^{(k-1)}, \nu_j^{(k-1)}\right)}{\sum_{h=1}^{J} \pi_h^{(k-1)} f_h\left(\mathbf{Y}_i; A_1^{(k-1)}, \mu_{h0}^{(k-1)}, \Sigma_{h0}^{(k-1)}, \nu_h^{(k-1)}\right)}, \tag{A1}$$

where

$$f_j\left(\mathbf{Y}_i; A_1^{(k-1)}, \mu_{j0}^{(k-1)}, \Sigma_{0j}^{(k-1)}, \nu_j^{(k-1)}\right) = \prod_{t=1}^{T} f_j\left(\mathbf{y}_{it}; A_1^{(k-1)}, \mu_{j0}^{(k-1)}, \Sigma_{0j}^{(k-1)}, \nu_j^{(k-1)}\right).$$

The conditional expectation of U_{it} is given in equation (18). The conditional expectation of the logarithm of U_{it} is

$$\mathbb{E}[\log U_{it}|\mathbf{Y}_1,...,\mathbf{Y}_T,z_{ij}=1,\mathbf{\Psi}=\mathbf{\Psi}^{(k-1)}] = \log u_{ijt}^{(k)} + \left[\psi\left(\frac{\nu_j^{(k-1)}+D}{2}\right) - \log\left(\frac{\nu_j^{(k-1)}+D}{2}\right)\right], \tag{A2}$$

where ψ denotes the digamma function; see Peel and McLachlan (2000).

Combining (A1) – (A2), we obtain the conditionally expected log likelihood function in three parts

$$Q(\mathbf{\Psi};\mathbf{\Psi}^{(k-1)}) = Q_1(\pi_1,...,\pi_J;\mathbf{\Psi}^{(k-1)}) + Q_2(\nu_1,...,\nu_J;\mathbf{\Psi}^{(k-1)}) + Q_3(A_1,A_2,\boldsymbol{\theta}_{10},...,\boldsymbol{\theta}_{J0};\mathbf{\Psi}^{(k-1)}),$$

where

$$Q_1(\pi_1,...,\pi_J;\mathbf{\Psi}^{(k-1)}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\tau_{ij}^{(k)}\log\pi_j,$$

$$Q_2(\nu_1,...,\nu_J;\mathbf{\Psi}^{(k-1)}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\tau_{ij}^{(k)}Q_{2i}(\nu_j;\mathbf{\Psi}^{(k-1)}),$$

$$Q_3(A_1,A_2,\boldsymbol{\theta}_{10},...,\boldsymbol{\theta}_{J0};\mathbf{\Psi}^{(k-1)}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\tau_{ij}^{(k)}Q_{3i}(A_1,A_2,\boldsymbol{\theta}_{j0};\mathbf{\Psi}^{(k-1)}),$$

with

$$\begin{aligned} Q_{2i}(\nu_j;\mathbf{\Psi}^{(k-1)}) = &-T\log\Gamma\left(\frac{\nu_j}{2}\right)+\frac{T\nu_j}{2}\log\left(\frac{\nu_j}{2}\right) \\ &-\sum_{t=1}^{T}\left[\log u_{ijt}^{(k)}+\psi\left(\frac{\nu_j^{(k-1)}+D}{2}\right)-\log\left(\frac{\nu_j^{(k-1)}+D}{2}\right)\right] \\ &+\frac{\nu_j}{2}\left[\sum_{t=1}^{T}\left(\log u_{ijt}^{(k)}-u_{ijt}^{(k)}\right)+\psi\left(\frac{\nu_j^{(k-1)}+D}{2}\right)-\log\left(\frac{\nu_j^{(k-1)}+D}{2}\right)\right] \end{aligned} \tag{A3}$$

and

$$Q_{3i}(A_1,A_2,\boldsymbol{\theta}_{j0};\mathbf{\Psi}^{(k-1)}) = \sum_{t=1}^{T}-\frac{1}{2}D\log(2\pi)-\frac{1}{2}\log|\Sigma_{jt}|+\frac{1}{2}D\log u_{ijt}^{(k)}-\frac{1}{2}u_{ijt}^{(k)}(\mathbf{y}_{it}-\mu_{jt})'\Sigma_{jt}^{-1}(\mathbf{y}_{it}-\mu_{jt}) \tag{A4}$$

M-Step

The mixing probabilities continue to updated as the average of the component membership probabilities, $\pi_j^{(k)} = \frac{1}{N}\sum_{i=1}^{N}\tau_{ij}^{(k)}$. The dynamics for μ_{jt} and Σ_{jt} change to account for the fat tails. Taking derivatives of (A4) with respect to μ_{jt} and Σ_{jt}, we obtain (17) and (19).

Parameters A_1, A_2 and $\boldsymbol{\theta}_{j0}$, $j = 1,\ldots,J$ can be obtained by numerically maximizing (A4), in combination with (12), (20), (17) and (19). The degrees of freedom parameters are estimated by maximizing (A3) with respect to $\nu_1,...,\nu_J$. As before, the E-step and M-step are iterated until convergence.

References

Altunbas, Y., S. Manganelli, and D. Marques-Ibanez (2011). Bank risk during the financial crisis - Do business models matter? *ECB working paper No. 1394*.

Ayadi, R., E. Arbak, and W. P. de Groen (2014). Business models in European banking: A pre- and post-crisis screening. *CEPS discussion paper*, 1–104.

Ayadi, R. and W. P. D. Groen (2015). Bank business models monitor Europe. *CEPS working paper*, 0–122.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometica 70(1)*, 191–221.

Beltratti, A. and R. M. Stulz (2012). The credit crisis around the globe: Why did some banks perform better? *Journal of Financial Economics 105(1)*, 1–17.

Brunnermeier, M. K. and Y. Sannikov (2015). The i theory of money. *Princeton university, Unpublished working paper*.

Catania, L. (2016). Dynamic adaptive mixture models. *University of Rome Tor Vergata, unpublished working paper*.

Chiorazzo, V., V. D'Apice, R. DeYoung, and P. Morelli (2016). Is the traditional banking model a survivor? *Unpublished working paper*, 1–44.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Creal, D., S. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). An observation driven mixed measurement dynamic factor model with application to credit risk. *The Review of Economics and Statistics 96(5)*, 898–915.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B 39*, 1–38.

ECB (2013). European Central Bank Financial Integration Report 2012.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails, with applications to financial and economic time series.* Number 52. Cambridge University Press.

Hurvich, C. M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika 76*, 297307.

Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics 32 (2)*, 271–284.

Lucas, A. and X. Zhang (2015). Score driven exponentially weighted moving average and value-at-risk forecasting. *International Journal of Forecasting forthcoming.*

McLachlan, G. and D. Peel (2000). *Finite Mixture Models.* Wiley.

McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative risk management: Concepts, techniques and tools.* Princeton Univ Pr.

Peel, D. and G. J. McLachlan (2000). Robust mixture modelling using the t distribution. *Statistics and Computing 10*, 339–348.

Poengpitya, R., N. Tarashev, and K. Tsatsaronis (2014). Bank business models. *BIS Quarterly Review*, 55–65.

Rao, C. R. (2001). *Linear Statistical Inference and its Applications.* Wiley-Interscience.

SSM (2016). SSM SREP methodology booklet. *available at www.bankingsupervision.europa.eu, accessed on 14 April 2016.*, 1–36.

Stock, J. and M. Watson (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics 20(2)*, 147–162.

Svensson, L. E. O. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992 to 1994. *NBER Working Paper No. 4871*, 1–49.

Zhang, X., D. Creal, S. Koopman, and A. Lucas (2011). Modeling dynamic volatilities and correlations under skewness and fat tails. *Tinbergen Institute Discussion Paper*.