EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

André Lucas, Julia Schaumburg,
Bernd Schwaab

Bank business models at zero interest rates

No 2084 / June 2017

Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Abstract

We propose a novel observation-driven nite mixture model for the study of banking data. The model accommodates time-varying component means and covariance matrices, normal and Student's *t* distributed mixtures, and economic determinants of time-varying parameters. Monte Carlo experiments suggest that units of interest can be classi ed reliably into distinct components in a variety of settings. In an empirical study of 208 European banks between 2008Q1{2015Q4, we identify six business model components and discuss how their properties evolve over time. Changes in the yield curve predict changes in average business model characteristics.

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

JEL classification: G21, C33.

1 Introduction

Banks are highly heterogeneous, 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 acting upon and originating from the financial sector, for impact assessments of unconventional monetary policies and financial regulations, as well as for the benchmarking of banks to appropriate peer groups for supervisory purposes.[1] While there is broad agreement that financial institutions suffer in an environment of extremely low interest rates, see e.g. Nouy (2016), it is less clear which types of banks (business models) are affected the most. A study of banks' business models at low interest rates provides insight into the overall diversity of business models, the strategies adopted by individual institutions, and which types of banks are impacted the most by time variation in the yield curve.[2] We study these questions in a novel modeling framework.

This paper proposes an observation-driven finite mixture model for the analysis of high-dimensional banking data. The framework accommodates time-varying mean and covariance parameters and allows us to robustly cluster banks into approximately homogeneous groups. We first present a simple baseline mixture model for normally distributed data with time-varying component means, and subsequently consider extensions to time-varying covariance matrices, Student's t distributed mixture densities, and economic predictors of time-varying parameters. We apply our modeling framework to a multivariate panel of $N = 208$ European banks between 2008Q1–2015Q4, i.e. over $T = 32$ quarters, considering $D = 13$ bank-level indicator variables for J groups of similar banks. We thus track banking sector data through

[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 Signi cant Institutions within its Single Supervisory Mechanism; see SSM (2016). Similar procedures exist in other jurisdictions.

[2] An improved understanding of the nancial stability consequences of low-for-long interest rates is a top policy priority. For example, Fed Chair Yellen (2014) pointed to \... the potential for low interest rates to heighten the incentives of nancial market participants to reach for yield and take on risk, and ... the limits of macroprudential measures to address these and other nancial stability concerns." Similarly, ECB President Draghi (2016) explained that \One particular challenge has arisen across a large part of the world. That is the extremely low level of nominal interest rates. ... Very low levels are not innocuous. They put pressure on the business model[s] of nancial institutions ... by squeezing net interest income. And this comes at a time when pro tability is already weak, when the sector has to adjust to post-crisis deleveraging in the economy, and when rapid changes are taking place in regulation."

the 2008–2009 global financial crisis, the 2010–2012 euro area sovereign debt crisis, as well as the relatively calmer but persistent low-interest rate environment of the post-crises period between 2013–2015. We identify $J = 6$ business model components and discuss how these adjust to changes in the yield curve.

In our finite mixture model, all time-varying parameters are driven by the score of the local (time t) objective function using the so-called Generalized Autoregressive Score (GAS) approach developed by Creal et al. (2013); see also Harvey (2013). 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 expectation-maximization (EM) procedures. Our approach extends the standard score-driven approach of Creal et al. (2013) by using the scores of the EM-based criterion function rather than that of the usual predictive likelihood function.

Extensive Monte Carlo experiments suggest that our model is able to reliably classify units of interest into distinct mixture components, as well as to simultaneously infer the relevant component-specific time-varying parameters. In our simulations, the cluster classification is perfect for sufficiently large distances between the time-varying cluster means and sufficiently informative signals relative to the variance of the noise terms.[3] This holds under correct model specification as well as under specific forms of model mis-specification. As the simulated data become less informative or the time-varying cluster means are closer together, the share of correct classifications decreases, but generally remains high. Estimation fit and the share of correct classifications decrease further if we incorrectly assume a thin-tailed mixture specification when the data are generated by a fat-tailed mixture distribution. As a result, robust models based on fat-tailed mixtures are appropriate for the fat-tailed bank accounting ratios in our empirical sample.

We apply our model to classify European banks into distinct business model components. We distinguish A) large universal banks, including globally systemically important banks (G-SIBs), B) international diversified lenders, C) fee-based banks, D) domestic diversified

[3] We use the terms `component', `mixture component' and `cluster' interchangeably.

lenders, E) domestic retail lenders, and F) small international 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 find that the global financial crisis between 2008–2009 affected banks with different business models differently. This is in line with findings by Altunbas et al. (2011) and Chiorazzo et al. (2016), who study U.S.-based institutions.

In addition, we study how banks' business models adapt to changes in yield curve factors, specifically level and slope of the yield curve. The yield curve factors are extracted from AAA-rated euro area sovereign bonds based on a Svensson (1995) model. We find that, as long-term interest rates decrease, banks on average (across all business models) grow larger, hold more assets in trading portfolios to offset declines in loan demand, hold more sizeable derivative books, and, in some cases, increase leverage and decrease funding through customer deposits. Each of these effects – increased size, leverage, complexity, and a less stable funding base – are intuitive, but also potentially problematic from a financial stability perspective. This corroborates the unease expressed in Yellen (2014) and Draghi (2016).

From a methodological point of view, our paper also contributes to the literature on clustering of time series data. This literature can be divided into four strands. Static clustering of time series refers to a setting with fixed cluster classification, i.e., each time series is allocated to one cluster over the entire sample period. Dynamic clustering, by contrast, allows for changes in the cluster assignments over time. Each approach can be further split into whether the cluster-specific parameters are constant (static) or time-varying (dynamic).

Wang et al. (2013) is an example of static clustering with static parameters. They cluster time series into different groups of autoregressive processes, where the autoregressive parameters are constant within each cluster and cluster assignments are fixed over time.

Fruehwirth-Schnatter and Kaufmann (2008) use static clustering with elements of both static and dynamic parameters. First, they cluster time series into different groups of regression models with static parameters. Later, they generalize this to static clustering into groups of different Hidden Markov Models (HMMs), each switching between two regression models. The HMM can be regarded as a specific form of dynamic parameters for the underly-

ing regression model. Their method is used in Hamilton and Owyang (2012) to differentiate between business cycle dynamics among groups of U.S. states. Also Smyth (1996) clusters time series into groups characterized by different Hidden Markov Models.

Creal et al. (2014) is an example of dynamic clustering with static parameters. They develop a model for credit ratings based on market data. Their main objective is to classify firms into different rating categories over time. They therefore allow for transitions across clusters (dynamic clustering), while the parameters in their underlying mixture model are kept constant.

Finally, Catania (2016) is an example of dynamic clustering with dynamic parameters. He proposes a score-driven dynamic mixture model, which relies on score-driven updates of almost all parameters, allowing for time-varying parameters and changing cluster assignments and time-varying cluster assignment probabilities. Due to the high flexibility of the model, a large number of observations is required over time. The application in Catania (2016) to conditional asset return distributions typically has a sufficiently large number of observations.

Our approach falls in the category of static clustering methods with dynamic parameters. We use static clustering as banks do not tend to switch their business model frequently over short periods of time; see e.g. Ayadi and Groen (2015). Also, in contrast to the application used by for instance Catania (2016), our banking data are observed over only a moderate number of time points T, while the number of units N and the number of firm characteristics D are high. Given static clustering, the properties of bank business models are unlikely to be constant throughout the periods of market turbulence and shifts in bank regulations experienced in our sample. We therefore require the cluster components to be characterized by dynamic parameters using the score-driven framework of Creal et al. (2013).

Our paper also contributes to the literature on identifying bank business models. Roengpitya et al. (2014), Ayadi et al. (2014), and Ayadi and Groen (2015) also use cluster analysis to identify bank business 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 panel framework, which allows us to pool information over time, leading to a more accurate assessment.

We proceed as follows. Section 2 presents a static and baseline dynamic finite mixture model. We then propose extensions to incorporate time-varying covariance matrices, as well as Student's t distributed mixture distributions, and introduce model diagnostics. 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. A Web Appendix provides further technical and empirical results.

2 Statistical model

2.1 Mixture model

We consider multivariate panel data consisting of vectors $\boldsymbol{y}_{i,t} \in \mathbb{R}^{D\times 1}$ of firm characteristics for firms $i = 1, \ldots, N$ and times $t = 1, \ldots, T$, where D denotes the number of observed characteristics. We model $\boldsymbol{y}_{i,t}$ by a J-component mixture model of the form

$$\boldsymbol{y}_{i,t} = z_{i,1} \cdot \left(\mu_{1,t} + \Omega_{1,t}^{1/2} e_{i,t,1}\right) + \ldots + z_{i,J} \cdot \left(\mu_{J,t} + \Omega_{J,t}^{1/2} e_{i,t,J}\right), \tag{1}$$

where $\mu_{j,t}$ and $\Omega_{j,t}$ are the mean and covariance matrix of mixture component $j = 1, \ldots, J$ at time t, respectively, $e_{i,t,j}$ is a zero-mean, D-dimensional vector of disturbances with identity covariance matrix, and $z_{i,j}$ are unobserved indicators for the mixture component of firm i. In particular, if firm i is in mixture component j then $z_{i,j} = 1$, while $z_{i,k} = 0$ for $k \neq j$. The posterior expectations of $z_{i,j}$ given the data can be used to classify firms into specific mixture components later on. We define $\boldsymbol{z}_i = (z_{i,1}, \ldots, z_{i,J})'$ and assume $\boldsymbol{z}_i$ has a multinomial distribution with $\Pr[z_{i,j} = 1] = \pi_j \in [0, 1]$ and $\pi_1 + \ldots + \pi_J = 1$. Finally, we assume that $\boldsymbol{z}_i$ and $e_{i,t,j}$ are mutually, cross-sectionally, and serially uncorrelated. We specify the precise dynamic functional form of $\mu_{j,t}$ and $\Omega_{j,t}$ later in this section using the score-driven dynamics

of Creal et al. (2013). For the moment, it suffices to note that $\mu_{j,t}$ and $\Omega_{j,t}$ will both be functions of past data only, and therefore predetermined. Finite mixture models with static cluster-specific parameters have been widely used in the literature. For textbook treatments, see, e.g. McLachlan and Peel (2000) and Fruehwirth-Schnatter (2006).

To write down the likelihood of the mixture model in (1), we stack the observations up to time t, $\boldsymbol{y}_{i,1}, \ldots, \boldsymbol{y}_{i,t}$, into the matrix $\boldsymbol{Y}_{i,t} = (\boldsymbol{y}_{i,1} \ \cdots \ \boldsymbol{y}_{i,t})' \in \mathbb{R}^{t \times D}$. We also stack the parameters characterizing each mixture component j, such as the $\mu_{j,t}$s and $\Omega_{j,t}$s for all times t, and any remaining parameters characterizing the distribution of $e_{i,t,j}$ (such as the degrees of freedom of a Student's t), into a parameter vector $\boldsymbol{\theta}_j(\quad)$, where gathers all static parameters of the model. Note that also the multinomial probabilities π_j are functions of , i.e., $\pi_j = \pi_j(\quad)$. However, if no confusion is caused we use the short-hand notation π_j and $\boldsymbol{\theta}_j$ for $\pi_j(\quad)$ and $\boldsymbol{\theta}_j(\quad)$, respectively. The likelihood function is given by a standard prediction error decomposition as

$$\log L(\quad) = \sum_{i=1}^{N} \log \left[\sum_{j=1}^{J} \pi_j \cdot f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j) \right], \tag{2}$$

where

$$f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j) = \prod_{t=1}^{T} f_j \left(\boldsymbol{y}_{i,t} \mid \boldsymbol{Y}_{i,t-1} \ ; \ \boldsymbol{\theta}_{j,t} \right),$$

and $f_j(\boldsymbol{y}_{i,t} \mid \boldsymbol{Y}_{i,t-1}; \boldsymbol{\theta}_{j,t})$ is the conditional distribution of $\boldsymbol{y}_{i,t} = \mu_{j,t} + \Omega_{j,t}^{1/2} e_{i,t,j}$ given the past data and given the (predetermined) parameters for time t as gathered in $\boldsymbol{\theta}_{j,t}$.

Before proceeding, we note that the mixture model in (1) describes the firm characteristics using time-invariant cluster indicators $\boldsymbol{z}_i$ rather than time-varying indicators $z_{i,t}$. Our choice follows from the specific application in Section 4. Banks are unlikely to switch their business model over limited time spans such as ours. For instance, a large universal bank is unlikely to become a small retail lender from one year to the next, as strategy choices, distribution channels, brand building, and clientele formation are all slowly varying economic processes. This is why we opt for static cluster indicators. In a different empirical context, a different modeling choice might be called for. For example, Creal et al. (2014) consider corporate

credit ratings, which are much more likely to change over shorter periods of time, such that some of their specifications use time-varying cluster assignments. To explicitly check whether the assumption of fixed cluster assignments is supported by our data, we use the diagnostics developed in Section 2.5. Our findings indicate that the vast majority of banks indeed only belongs to one cluster for all time points.

Given our choice for static rather than dynamic cluster allocation, it becomes important to allow for time-variation in the cluster means $\mu_{j,t}$ (and possibly in the variances $\Omega_{j,t}$). Even though banks are less likely to switch their business model, the average characteristics of business models may change over shorter time spans, particularly if such time spans include stressful periods as is the case in our sample. This allows us to answer questions relating to how the properties of business models changed, and in particular whether some business models (and if so, which) increased their risk characteristics during the low interest rate period we study in Section 4. Such results are also important for policy makers, such as the Single Supervisory Mechanism in Europe to decide on the riskiness of banks and on adequate capital and liquidity levels for peer groups of banks.

2.2 EM estimation

As is common in the literature on mixture models, we do not estimate directly by numerically maximizing the log-likelihood function in (2). Instead we use the expectation maximization (EM) algorithm to estimate the parameters; see Dempster et al. (1977) and McLachlan and Peel (2000).[4] To write down the EM algorithm and formulate the score-driven parameter dynamics for $\mu_{j,t}$ and $\Omega_{j,t}$ later on, we define the complete data for firm i as the pair $(\boldsymbol{Y}_{i,T}, \boldsymbol{z}_i)$. If $\boldsymbol{z}_i$ is known, the corresponding complete data likelihood function is given by

$$\log L_c(\quad) = \sum_{i=1}^{N} \sum_{j=1}^{J} z_{i,j} \left[\log \pi_j + \log f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j)\right]. \tag{3}$$

[4] As pointed out by a referee, newer and faster versions of the EM algorithm are available, such as the ECM algorithm of Meng and Rubin (1993) and the ECME algorithm of Liu and Rubin (1994). All of these converge to the same optimum. Computation time for the EM was not a major issue in our setting, with the algorithm typically converging in 15 iterations. We therefore leave such extensions for future work.

Because $\boldsymbol{z}_i$ is unobserved, however, (3) cannot be maximized directly. Following Dempster et al. (1977), we instead maximize its conditional expectation over $\boldsymbol{z}_i$ given the observed data $\boldsymbol{Y}_T = (\boldsymbol{Y}_{1,T}, \ldots, \boldsymbol{Y}_{N,T})$ and some initial or previously determined parameter value $^{(k-1)}$, i.e., we maximize with respect to the function

$$\begin{aligned} Q(\quad ; \quad ^{(k-1)}) &= \mathbb{E}\left[\log L_c(\quad) \;\middle|\; \boldsymbol{Y}_T; \quad ^{(k-1)}\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{J} z_{i,j}\left[\log \pi_j + \log f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j)\right] \;\middle|\; \boldsymbol{Y}_T; \quad ^{(k-1)}\right] \\ &= \sum_{i=1}^{N}\sum_{j=1}^{J} \mathbb{P}\left[z_{i,j} = 1 \;\middle|\; \boldsymbol{Y}_T; \quad ^{(k-1)}\right]\left[\log \pi_j + \log f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j)\right]. \end{aligned} \tag{4}$$

The conditionally expected likelihood (4) can be optimized iteratively by alternately updating the conditional expectation of the component indicators $\boldsymbol{z}_i$ ('E-Step') and subsequently maximizing the remaining part of the function with respect to ('M-Step').

In the E-Step, the conditional component indicator probabilities are updated using

$$\tau_{i,j}^{(k)} := \mathbb{P}[z_{i,j} = 1 \mid \boldsymbol{Y}_T, \quad ^{(k-1)}] = \frac{\pi_j^{(k-1)} f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j^{(k-1)})}{f(\boldsymbol{Y}_{i,T}; \quad ^{(k-1)})} = \frac{\pi_j^{(k-1)} f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j^{(k-1)})}{\sum_{h=1}^{J} \pi_h^{(k-1)} f_h(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_h^{(k-1)})}. \tag{5}$$

We again point out that the $\tau_{i,j}^{(k)}$s do not depend on time, as banks in our application in Section 4 are statically assigned to clusters. An alternative would be to use dynamic cluster assignments as in Catania (2016), in which case the densities $f_j(\boldsymbol{Y}_{i,T}; \boldsymbol{\theta}_j^{(k-1)})$ above would have to be replaced by their time t counterparts $f_j(\boldsymbol{y}_{i,t} \mid \boldsymbol{Y}_{i,t-1} \; ; \; \boldsymbol{\theta}_{j,t}^{(k-1)})$ and would result in time-specific posterior probabilities $\tau_{i,j,t}^{(k)}$; see also the diagnostic statistics introduced in Section 2.5.

Once the $\tau_{i,j}^{(k)}$s are updated, we move to the M-Step. Maximizing $Q(\quad ; \quad ^{(k-1)})$ with respect to π_j under the constraint $\pi_1 + \ldots + \pi_J = 1$, we obtain

$$\pi_j^{(k)} = \frac{1}{N}\sum_{i=1}^{N} \tau_{i,j}^{(k)}, \quad j = 1, \ldots, J. \tag{6}$$

The optimization of $Q(\quad ; \quad ^{(k-1)})$ with respect to the remaining parameters in can some-

times be done analytically, for instance in the case of the normal finite mixture model with static location $\mu_{j,t} \equiv \mu_j$ and scale $\Omega_{j,t} \equiv \Omega_j$. Otherwise, numerical maximization methods need to be used. The E-step and M-step are iterated until the difference $L(\quad^{(k+1)}) - L(\quad^{(k)})$ has converged. The EM algorithm increases the likelihood on each step, and convergence typically occurs within 15 iterations in our application. After convergence, when has been estimated, we can use the final $\tau_{i,j}^{(k)}$ to assign banks to clusters. We do so by assigning bank i to cluster j which has the highest $\tau_{i,j}^{(k)}$ across j.

2.3 Normal mixture with time-varying means

As explained in Section 2.1, it is important to allow for time-varying cluster means. We first do so for the case of a normal mixture with time varying means and constant covariance matrices. We set $f_j(\boldsymbol{y}_{i,t} \mid \boldsymbol{Y}_{i,t-1}; \boldsymbol{\theta}_{j,t}) = \phi(\boldsymbol{y}_{i,t}; \mu_{j,t}, \Omega_j)$, where $\phi(\cdot\,; \mu, \Omega)$ denotes a multivariate normal density function with mean μ and variance Ω. In this section, we introduce a version of the score-driven approach of Creal et al. (2013) to the parameter dynamics of $\mu_{j,t}$; compare also Harvey (2013) and Creal et al. (2014). Rather than using the score of the log-density as in Creal et al. (2013), however, we use the score of the EM criterion in (4) to drive the parameter dynamics. Our simulation section shows that the score-driven dynamics can fit various patterns for the cluster means, both in correctly specified and mis-specified settings.

For simplicity and parsimony, we consider the integrated score-driven dynamics as discussed in Lucas and Zhang (2016),

$$\mu_{j,t+1} = \mu_{j,t} + A_1 s_{\mu_{j,t}}, \tag{7}$$

where $A_1 = A_1(\quad)$ is a diagonal matrix that depends on the unknown parameter vector , and where $s_{\mu_{j,t}}$ is the scaled first derivative of the time t EM objective function with respect

to $\mu_{j,t}$. The score is given by

$$\begin{aligned}\nabla_{\mu_{j,t}} &= \frac{\partial}{\partial \mu_{j,t}} \left(\sum_{i=1}^{N} \sum_{j=1}^{J} \tau_{i,j}^{(k)} \left[\log \pi_j + \sum_{t=1}^{T} \log \phi(\boldsymbol{y}_{i,t}; \mu_{j,t}, \Omega_j) \right] \right) \\ &= \frac{\partial}{\partial \mu_{j,t}} \left(\sum_{i=1}^{N} \sum_{j=1}^{J} \tau_{i,j}^{(k)} \left[\log \pi_j - \tfrac{1}{2} T \log |2\pi\Omega_j| - \tfrac{1}{2} \sum_{t=1}^{T} (\boldsymbol{y}_{i,t} - \mu_{j,t})' \Omega_j^{-1} (\boldsymbol{y}_{i,t} - \mu_{j,t}) \right] \right) \\ &= \Omega_j^{-1} \sum_{i=1}^{N} \tau_{i,j}^{(k)} \left(\boldsymbol{y}_{i,t} - \mu_{j,t} \right). \end{aligned} \tag{8}$$

To scale our score for $\mu_{j,t}$, we compute the inverse of the expected negative Hessian under mixture component j. In particular, we take the derivative of (8) with respect to the transpose of $\mu_{j,t}$, switch sign, and compute the inverse, thus obtain a scaling matrix $\Omega_j / \sum_{i=1}^{N} \tau_{i,j}^{(k)}$. This yields a corresponding scaled score update of the form

$$\mu_{j,t+1} = \mu_{j,t} + A_1 \cdot \frac{\sum_{i=1}^{N} \tau_{i,j} (\boldsymbol{y}_{i,t} - \mu_{j,t})}{\sum_{i=1}^{N} \tau_{i,j}}. \tag{9}$$

This updating mechanism is highly intuitive: the component means are updated by the prediction errors for that component, accounting for the posterior probabilities that the observation was drawn from that same component. For example, if the posterior probability $\tau_{i,j}^{(k)}$ that $\boldsymbol{y}_{i,t}$ comes from component j is negligible, the update of $\mu_{j,t}$ does not depend on the observation of firm i.

We note that we do not scale the score by the inverse Fisher information matrix as suggested in for instance Creal et al. (2013). So far, there is no optimality theory for the choice of the scale for the score, and different proposals can be found in the literature. Computing the information matrix for the mixture model is hard, particularly if we take into account that also $\tau_{i,j}^{(k)}$ is a function of $\boldsymbol{y}_{i,t}$. We can show, however, that our proposed way of scaling the score collapses to the inverse information matrix if the mixture components are sufficiently far apart.

All static parameters can now be estimated using the EM-algorithm. Starting from an initial $^{(k-1)}$ and an initial mean $\mu_{j,1}^{(k-1)}$, we compute $\mu_{j,2}^{(k-1)}, \ldots, \mu_{jT}^{(k-1)}$ using the recursion

(9). We compute the posterior probabilities as

$$\tau_{i,j}^{(k)} = \frac{\pi_j^{(k-1)} \prod_{t=1}^{T} \phi\left(\boldsymbol{y}_{i,t}; \mu_{j,t}^{(k-1)}, \Omega_j^{(k-1)}\right)}{\sum_{h=1}^{J} \pi_h^{(k-1)} \prod_{t=1}^{T} \phi\left(\boldsymbol{y}_{i,t}; \mu_{ht}^{(k-1)} \Omega_h^{(k-1)}\right)}. \tag{10}$$

Next, the M-Step maximizes

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{D} \tau_{i,j}^{(k)} \left[-\tfrac{1}{2} \log(|2\pi\Omega_j|) - \tfrac{1}{2}(\boldsymbol{y}_{i,t} - \mu_{j,t})' \Omega_j^{-1} (\boldsymbol{y}_{i,t} - \mu_{j,t})\right], \tag{11}$$

with respect to A_1 and Ω_j. The initial values $\mu_{j,1}$ can also be estimated if J and D are not too large. Otherwise, the number of parameters becomes infeasible. Alternatively, one can initialize the time-varying means $\mu_{j,1}$ by the $\tau_{i,j}$-weighted average of the first cross-section(s). Given the values of J and D in our empirical study, we opt for this latter approach. We set $\mu_{j,1}$ equal to the weighted unconditional sample average in the simulation study, and to the weighted average of the first cross-section in the empirical application. Given $\mu_{j,1}$ and A_1, the optimization with respect to Ω_j can be done analytically. The optimization with respect to A_1 has to be carried out numerically.

The E-step and M-step are iterated until convergence. To start up the EM algorithm, we initialize the weights $\tau_{i,j}$ randomly. To robustify the optimization algorithm, we use a large number of random starting values and pick the highest value for the final converged criterion function.

2.4 Extensions

2.4.1 Time-varying component covariance matrices

This section derives the scaled score updates for time-varying component covariance matrices $\Omega_{j,t}$. If we also want to endow the time-varying covariance matrices with integrated score dynamics, we have

$$\Omega_{j,t+1} = \Omega_{j,t} + A_2 \, s_{\Omega_{j,t}}, \tag{12}$$

where $s_{\Omega_{j,t}}$ is again defined as the scaled first partial derivative of the expected likelihood function with respect to $\Omega_{j,t}$. Following equation (8), the unscaled score with respect to $\Omega_{j,t}$ is

$$\nabla_{\Omega_{j,t}} = \tfrac{1}{2} \sum_{i=1}^{N} \tau_{i,j}^{(k)} \Omega_{j,t}^{-1} \left[(\boldsymbol{y}_{i,t} - \mu_{j,t})(\boldsymbol{y}_{i,t} - \mu_{j,t})' - \Omega_{j,t} \right] \Omega_{j,t}^{-1}. \tag{13}$$

Taking the total differential of this expression, and subsequently taking expectations $\mathrm{E}_j[\,\cdot\,]$ conditional on mixture component j, we obtain

$$\tfrac{1}{2}\,\mathrm{E}_j \left[\sum_{i=1}^{N} \tau_{i,j}^{(k)} \left(\mathrm{d}\Omega_{j,t}^{-1} (\boldsymbol{y}_{i,t} - \mu_{j,t})(\boldsymbol{y}_{i,t} - \mu_{j,t})' \Omega_{j,t}^{-1} + \Omega_{j,t}^{-1} (\boldsymbol{y}_{i,t} - \mu_{j,t})(\boldsymbol{y}_{i,t} - \mu_{j,t})' \mathrm{d}\Omega_{j,t}^{-1} - \mathrm{d}\Omega_{j,t}^{-1} \right) \right] =$$

$$\tfrac{1}{2} \sum_{i=1}^{N} \tau_{i,j}^{(k)}\ \mathrm{d}\Omega_{j,t}^{-1} = -\left(\sum_{i=1}^{N} \tfrac{1}{2}\, \tau_{i,j}^{(k)} \right)\ \Omega_{j,t}^{-1}\ \mathrm{d}\Omega_{j,t}\ \Omega_{j,t}^{-1}. \tag{14}$$

Vectorizing (14), we obtain $-(\frac{1}{2}\sum_{i=1}^{N} \tau_{i,j}^{(k)})(\Omega_{j,t} \otimes \Omega_{j,t})^{-1} \mathrm{vec}(\mathrm{d}\Omega_{j,t})$, where $\mathrm{vec}(\cdot)$ concatenates the columns of a matrix into a column vector, and where the negative inverse of the matrix in front of $\mathrm{vec}(\mathrm{d}\Omega_{j,t})$ is our scaling matrix to correct for the curvature of the score. Multiplying the vectorized version of (13) by this scaling matrix, we obtain the scaled score

$$\mathrm{vec}(s_{\Omega_{j,t}}) = \left(\tfrac{1}{2} \sum_{i=1}^{N} \tau_{i,j}^{(k)} \right)^{-1} (\Omega_{j,t} \otimes \Omega_{j,t}) \cdot \mathrm{vec}\left(\nabla_{\Omega_{j,t}} \right) = \left(\sum_{i=1}^{N} \tau_{i,j}^{(k)} \right)^{-1} \cdot \mathrm{vec}\left(2\Omega_{j,t} \nabla_{\Omega_{j,t}} \Omega_{j,t} \right) \Leftrightarrow$$

$$s_{\Omega_{j,t}} = \frac{\sum_{i=1}^{N} \tau_{i,j}^{(k)} \left[(\boldsymbol{y}_{i,t} - \mu_{j,t})(\boldsymbol{y}_{i,t} - \mu_{j,t})' - \Omega_{j,t} \right]}{\sum_{i=1}^{N} \tau_{i,j}^{(k)}}. \tag{15}$$

The estimation of the model can be carried out using the EM algorithm as before, replacing Ω_j by $\Omega_{j,t}$ in equations (10)–(11).

2.4.2 Student's t distributed mixture

This section robustifies the dynamic finite mixture model by considering panel data that are generated by mixtures of multivariate Student's t distributions. 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 use the EM-algorithm for mixtures of Student's t distributions, we use the densities

$$f_j(\boldsymbol{y}_{i,t};\boldsymbol{\theta}_{j,t}) = \frac{\Gamma((\nu_j+D)/2)}{\Gamma(\nu_j/2)\ |\pi\nu_j\Omega_{j,t}|^{1/2}}\ \left(1+(\boldsymbol{y}_{i,t}-\mu_{j,t})'\left(\nu_j\Omega_{j,t}\right)^{-1}(\boldsymbol{y}_{i,t}-\mu_{j,t})\right)^{-(\nu+D)/2}. \quad (16)$$

Both the E-steps and the M-steps of the algorithm are unaffected save for the fact that we use Student's t rather than Gaussian densities. The main difference follows for the dynamic models, where the score steps now take a different form. Using (16), the scores for the location parameter $\mu_{j,t}$ and scale matrix $\Omega_{j,t}$ are

$$\nabla_{\mu_{j,t}} = \Omega_{j,t}^{-1}\sum_{i=1}^{N}\tau_{i,j}^{(k)}\, w_{i,j,t}\cdot(\boldsymbol{y}_{i,t}-\mu_{j,t}), \quad (17)$$

$$\nabla_{\Omega_{j,t}} = \tfrac{1}{2}\sum_{i=1}^{N}\tau_{i,j}^{(k)}\Omega_{j,t}^{-1}\left[w_{i,j,t}\cdot(\boldsymbol{y}_{i,t}-\mu_{j,t})(\boldsymbol{y}_{i,t}-\mu_{j,t})'-\Omega_{j,t}\right]\Omega_{j,t}^{-1}, \quad (18)$$

$$w_{i,j,t} = (1+\nu_j^{-1}D)\,/\left(1+\nu_j^{-1}(\boldsymbol{y}_{i,t}-\mu_{j,t})'\Omega_{j,t}^{-1}(\boldsymbol{y}_{i,t}-\mu_{j,t})\right). \quad (19)$$

The main difference between the scores of the Student's t and the Gaussian case is the presence of the weights $w_{i,j,t}$. These weights provide the model with a robustness feature: observations $\boldsymbol{y}_{i,t}$ that are outlying given the fat-tailed nature of the Student's t density receive a reduced impact on the location and volatility dynamics by means of a lower value for $w_{i,j,t}$; compare Creal et al. (2011, 2013) and Harvey (2013). We use the same scale matrices for the score as in Sections 2.3 and 2.4.1. For the location parameter, which is our main parameter of interest, the scaling matrix for the Student's t case is proportional to that for the normal, such that any differences are included in the estimation of the smoothing parameter A_1. We obtain the scaled scores

$$s_{\mu_{j,t}} = \left(\sum_{i=1}^{N}\tau_{i,j}^{(k)}\, w_{i,j,t}\cdot(\boldsymbol{y}_{i,t}-\mu_{j,t})\right)\Bigg/\left(\sum_{i=1}^{N}\tau_{i,j}^{(k)}\right), \quad (20)$$

$$s_{\Omega_{j,t}} = \left(\sum_{i=1}^{N}\tau_{i,j}^{(k)}\left(w_{i,j,t}\cdot(\boldsymbol{y}_{i,t}-\mu_{j,t})(\boldsymbol{y}_{i,t}-\mu_{j,t})'-\Omega_{j,t}\right)\right)\Bigg/\left(\sum_{i=1}^{N}\tau_{i,j}^{(k)}\right). \quad (21)$$

The intuition is the same as for the Gaussian case, except for the fact that the scaled score steps for $\mu_{j,t}$ and $\Omega_{j,t}$ are re-descending to zero and bounded, respectively, if $\boldsymbol{y}_{i,t}$ is extremely

far from $\mu_{j,t}$. Also note that for $\nu_j \to \infty$ we see in (19) that $w_{i,j,t} \to 1$, such that we recover the expressions for the Gaussian mixture model.

2.4.3 Explanatory covariates

The score-driven dynamics for component-specific time-varying parameters can be extended further to include contemporaneous or lagged economic variables as additional conditioning variables. For example, a particularly low interest rate environment may push financial institutions, overall or in part, to take more risk or change their asset composition; see e.g. Hannoun (2015), Abbassi et al. (2016), and Heider et al. (2017). Using additional yield curve-related conditioning variables allows us to incorporate and test for such effects. Let X_t be a vector of observed covariates, and $B_j = B_j(\quad)$ a matrix of unknown coefficients that need to be estimated. In the case of a Student's t distributed mixture, the score-driven updating scheme then changes slightly to

$$\mu_{j,t+1} = \mu_{j,t} + A_1 \cdot \frac{\sum_{i=1}^{N} \tau_{i,j}^{(k)} w_{i,j,t} (\boldsymbol{y}_{i,t} - \mu_{j,t})}{\sum_{i=1}^{N} \tau_{i,j}^{(k)}} + B_j \cdot X_t. \tag{22}$$

Again, in the case of a Gaussian mixture, $w_{i,j,t} = 1$. The covariates can also be made firm and cluster component specific, i.e., $X_{i,j,t}$.

2.5 Diagnostics: Stability of cluster allocation over time

The assumption that component membership is time-invariant implies that pooling information over $t = 1, \ldots, T$ is optimal. This is of substantial help to robustly classify each unit i. Although our sample covers only 32 quarters (8 years), it is clear that switches in component membership become more likely as the sample period grows. In such a case, we have to trade off estimation efficiency against estimation bias.

To check whether component probabilities $\tau_{i,j}$ are time-varying, we consider the point-

in-time diagnostic statistic

$$\hat{\tau}_{ij|t} = \frac{\hat{\pi}_j f_j \left(\boldsymbol{y}_{i,t} \middle| \boldsymbol{Y}_{i,t-1}; \boldsymbol{\theta}_{j,t}(\hat{}) \right)}{\sum_{h=1}^{J} \hat{\pi}_h f_h \left(\boldsymbol{y}_{i,t} \middle| \boldsymbol{Y}_{i,t-1}; \boldsymbol{\theta}_{h,t}(\hat{}) \right)}, \tag{23}$$

which can be viewed as the time t posterior probability that firm i belongs to cluster component j, computed using the estimates under the null of time-invariant cluster assignments. A filtered counterpart using information from time 1 to t can be constructed by replacing $f_j(\boldsymbol{y}_{i,t} \mid \boldsymbol{Y}_{i,t-1}; \boldsymbol{\theta}_{j,t}(\hat{}))$ by $\prod_{s=1}^{t} f_j(\boldsymbol{y}_{i,s} \mid \boldsymbol{Y}_{i,s-1}; \boldsymbol{\theta}_{j,s}(\hat{}))$. If $\hat{\tau}_{i,j|t}$ is close to 1 or 0 for all t for a specific (i, j), firm i is unlikely to have switched clusters. Otherwise, switches may be a concern. We discuss time series plots of $\tau_{i,j|t}$ for diagnostic purposes in our application in Section 4.

3 Simulation study

3.1 Simulation design

This section investigates the ability of our score-driven dynamic mixture model to simultaneously *i*) correctly classify a data set into distinct components, and *ii*) recover the dynamic cluster means over time. In addition, we investigate the performance of several model selection criteria from the literature in detecting the correct model when the number of clusters is unknown. In all cases, we pay particular attention to the sensitivity of the EM algorithm to the (dis)similarity of the clusters, the number of units per cluster, and the impact of model misspecification.

We simulate from a mixture of dynamic bivariate densities. These densities are composed of sinusoid mean functions and i.i.d. disturbance terms that are drawn from a bivariate Gaussian distribution or a bivariate Student's t distribution with five or three degrees of freedom. The covariance matrices are chosen to be time-invariant identity matrices.

The sample sizes are chosen to resemble typical sample sizes in studies of banking data. We thus keep the number of time points small to moderate, considering $T \in \{10, 30\}$, and

set the number of cross-sectional units equal to $N = 100$ or to $N = 400$. The number of clusters used to generate the data is fixed at $J = 2$ throughout. In our first set of simulation results in Section 3.2, we assume $J = 2$ is known. In a second set of simulations, we do not assume to know the number of clusters, but determine it using different model selection criteria. To save space in the main text, the description of these criteria has been moved to Web Appendix A, together with the outcomes of these simulations.

In our baseline setting, visualized in Figure 1, we generate data from two clusters located around means that move in two non-overlapping circles over time. Across our different simulation designs, the data have different signal-to-noise ratios in the sense that the radius of the circles is large or small relative to the variance of the error terms. In addition, we also consider two more challenging settings where the two circles overlap completely: the circles have the same center, but differ in the orientation of the time-varying mean component (clockwise vs. counterclockwise). Again, we consider circles with a large and small radius, respectively, while keeping the variance of the error terms fixed and thus changing the signal-to-noise ratio in the simulation set-up.

Finally, we investigate the impact of two types of model misspecification. First, we incorrectly assume a Gaussian mixture in the estimation process when the data are generated by a mixture of Student's t densities with five degrees of freedom ($\nu = 5$). Alternatively, we simulate from a $t(3)$-mixture, but fix the degrees of freedom parameter to five in the estimation. In both cases, we check the effect of mis-specifying the tail behavior of the mixture distribution. In total, we consider 96 different simulation settings.

3.2 Simulation results for classi cation and tracking

Using the score-driven model set-up and EM estimation methodology from Section 2, we classify the data points and estimate the component parameters from the simulated data. The static parameters to be estimated include the distinct entries of the covariance matrices, and the diagonal elements of the smoothing matrix A_1, which, for simplicity, we assume to be equal across dimensions and components, i.e. $A_1 = a_1 \mathrm{I}_D$.

Figure 1 illustrates our simulation setup with two examples. The data generating pro-

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 (distance=8). The upper panel corresponds to the simulation setup with radius 4, while the lower panel depicts the mean circles with radius 1.

cesses are plotted as solid black lines. In each panel, the true process is compared to the pointwise median of the estimated paths over simulation runs (solid red line), as well as the filtered mean estimates for each simulation run (green triangles). The actual observations are dispersed much more widely around the black circles, as for each point on the circle they are drawn from the bivariate standard normal distribution, thus ranging from approximately $\mu_{j,t} - 2.5$ to $\mu_{j,t} + 2.5$ with 99% probability. Our methodology allocates each data point to its correct component, and in addition tracks the dynamic mean processes accurately.

Table 1 presents mean squared error (MSE) statistics as our main measure of estimation fit. 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 $N_j = 100/2 = 50$ units per component. The bottom panel of Table 1 presents the same information for $N_j = 400/2 = 200$ 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 (i.e., a large circle radius) and a large distance between the unconditional means, the cluster classification is close to perfect, both under

Table 1: Simulation outcomes

Mean squared error (MSE) and average percentage of correct classi cation for each cluster (% C1 and % C2) across simulation runs. Considered sample sizes are $N = 100; 400$ and $T = 10; 30$. Radius (rad.) refers to the radius of the true mean circles and is a measure of the signal-to-noise ratio. Distance (dist.) is the distance between circle centers and measures the distinctness of clusters. Correct speci cation refers to the case of simulating from a normal mixture and estimating assuming a mixture of normal distributions. In the case of misspeci cation 1, data are simulated from a t-mixture with 5 degrees of freedom, but the model is estimated assuming normal mixtures. In the case of misspeci cation 2, a $t(3)$-mixture is used to simulate the data while in the estimation, a xed value = 5 is assumed.

$N = 100$							
correct speci cation							
			T=10			T=30	
rad.	dist.	MSE	% C1	% C2	MSE	% C1	% C2
4	8	0.34	100	100	0.36	100	100
4	0	0.34	100	100	0.36	100	100
1	8	0.04	100	100	0.04	100	100
1	0	0.04	99.18	99.28	0.04	100	99.99
misspeci cation 1							
			T=10			T=30	
rad.	dist.	MSE	% C1	% C2	MSE	% C1	% C2
4	8	0.34	100	100	0.37	100	100
4	0	0.35	100	100	0.37	100	100
1	8	0.05	100	99.98	0.05	100	100
1	0	0.12	89.31	84.67	0.08	96.62	93.8
misspeci cation 2							
			T=10			T=30	
rad.	dist.	MSE	% C1	% C2	MSE	% C1	% C2
4	8	0.42	100	100	0.46	100	100
4	0	0.42	100	100	0.46	100	100
1	8	0.05	100	100	0.05	100	100
1	0	0.07	94.14	93.78	0.05	99.7	99.62

$N = 400$							
correct speci cation							
			T=10			T=30	
rad.	dist.	MSE	% C1	% C2	MSE	% C1	% C2
4	8	0.32	100	100	0.34	100	100
4	0	0.32	100	100	0.34	100	100
1	8	0.02	100	100	0.03	100	100
1	0	0.04	98.38	98.31	0.03	99.98	99.99
misspeci cation 1							
			T=10			T=30	
rad.	dist.	MSE	% C1	% C2	MSE	% C1	% C2
4	8	0.32	100	100	0.35	100	100
4	0	0.32	100	100	0.35	100	100
1	8	0.03	100	100	0.03	100	100
1	0	0.06	94.16	91.68	0.03	99.71	99.61
misspeci cation 2							
			T=10			T=30	
rad.	dist.	MSE	% C1	% C2	MSE	% C1	% C2
4	8	0.41	100	100	0.44	100	100
4	0	0.41	100	100	0.44	100	100
1	8	0.03	100	100	0.04	100	100
1	0	0.05	95.03	95.18	0.06	97.74	97.78

correct specification and model misspecification. Interestingly, the distance between circles is irrelevant for estimation fit and classification ability in the case of large radii (signal-to-noise ratios).

As the distance between means and the circle radii decrease, the shares of correct classifications decrease as well. Both estimation fit and share of correct classification decrease further if we assume a Gaussian mixture although the data are generated from a mixture of fat-tailed Student's t distributions. This indicates a sensitivity to outliers when assuming a Gaussian mixture in the case of fat-tailed data. Incorrectly assuming five degrees of freedom when the data are generated by a $t(3)$-mixture, on the other hand, leads to little bias. Consequently, a misspecified t-mixture model allows us to obtain more robust estimation and classification results than a Gaussian mixture model when the data are fat-tailed. This is due to the fact that also the $t(5)$ based score-dynamics for $\mu_{j,t}$ already discount the impact of outlying observations, although not as strictly as the score-based dynamics of a Student's $t(3)$ distribution.

In our empirical study of banking data, the number of clusters, i.e. bank business models, is unknown a priori. A number of model selection criteria and so-called cluster validation indices have been proposed in the literature, and comparative studies have not found a dominant criterion that performs best in all settings; see e.g. Milligan and Cooper (1985) and de Amorim and Hennig (2015). We therefore run an additional simulation study to see which model selection criteria are suitable to choose the optimal number of components in our multivariate panel setting. We refer to Web Appendix A for the results. The Davies-Bouldin index (DBI; see Davies and Bouldin (1979)), the Calinski-Harabasz index (CHI; see Calinski and Harabasz (1974)), and the average Silhouette index (SI; see de Amorim and Hennig (2015)) perform well.

4 Bank business models

4.1 Data

The sample under study consists of $N = 208$ European banks, for which we consider quarterly bank-level accounting data from SNL Financial between 2008Q1 – 2015Q4. This implies $T = 32$. We assume that differences in banks' business models can be characterized along six dimensions: size, complexity, activities, geographical reach, funding strategies, and ownership structure. We select a parsimonious set of $D = 13$ indicators from these six categories. We consider banks' total assets, leverage with respect to CET1 capital [size], net loans to assets ratio, risk mix, assets held for trading, derivatives held for trading [complexity], share of net interest income, share of net fees & commissions income, share of trading income, ratio of retail loans to total loans [activities], ratio of domestic loans to total loans [geography], loans to deposits ratio [funding], and an ownership index [ownership].

We refer to Web Appendix B for a detailed discussion of our data, including data transformations and SNL Financial field keys. Web Appendix B also discusses our treatment of missing observations and banks' country location.

4.2 Model selection

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

Table 2 presents likelihood-based and distance-based information criteria, as well as different cluster validation indices for different values of $J = 2, \ldots, 10$. The log-likelihood fit increases monotonically with the number of clusters. Likelihood-based information criteria turn out to be sensitive to the specification of the penalty term. They either select the maximum number (AICc, BIC) or minimum number (AICk, BaiNg2) of components; see Web Appendix A for definitions of the different criteria. Distance-based cluster validation indices such as the CHI, DBI, and SI suggest $J = 6$. Each of these take a local maxium/minimum at

Table 2: Information criteria

Likelihood- and distance-based information criteria as well as cluster-validation indices for di erent values of $J = 2;:::;10$. The panel refers to a model speci cation with time-varying component means $_{j,t}$, time-invariant $_j$, and estimated as a free parameter. Each statistic is the maximum (respectively minimum) obtained from 5,000 random starting values for the model parameters. We refer to Web Appendix A for exact formulas and literature references for all reported selection criteria. The average Silhouette index (SI) is multiplied by 100. The top three suggested values are printed in bold.

J	loglik	AICc	BIC	AICk	BaiNg2	CHI	DBI	SI
2	1134.4	-1828.8	-393.8	**2525.4**	**-0.218**	**12.71**	3.61	**11.0**
3	9159.9	-17652.0	-15520.5	**2973.5**	**-0.074**	**9.00**	4.03	1.3
4	13700.4	-26497.4	-23677.2	**3243.5**	**-0.038**	**8.08**	3.79	1.4
5	17055.1	-32963.1	-29462.2	3615.3	0.059	7.42	3.84	1.1
6	19812.8	-38226.5	-34053.3	3990.6	0.157	7.90	**3.18**	**1.9**
7	21522.2	-41384.0	-36547.6	4085.8	0.040	7.25	3.27	1.1
8	**25142.2**	**-48353.6**	**-42863.3**	4576.8	0.227	7.57	**2.86**	1.1
9	**27341.3**	**-52471.3**	**-46337.1**	5035.6	0.386	7.43	**2.85**	**1.8**
10	**29883.8**	**-57265.3**	**-50497.7**	5281.3	0.369	6.31	3.23	0.5

this value. In practice, experts consider between five and up to more than ten different bank business models; see, for example, Ayadi et al. (2014) and Bankscope (2014, p. 299). With these considerations in mind, to be conservative and in line with Table 2 and the simulation results as reported in Web Appendix A, we choose $J = 6$ components for our subsequent empirical analysis.

Table 3: Model specification

Log-likelihoods and di erences in log-likelihoods for di erent model speci cations. The estimates are conditional on the same (optimal) allocation of banks to $J = 6$ components; cluster validation indices are not presented for this reason.

Density		value	A_1	$_j$; $_{j,t}$	loglik	loglik
N	-	∞	scalar	static	17,131.4	
t	xed	$\equiv 5$	scalar	static	20,485.3	3,353.9
t	xed	$\equiv 5$	vector	static	20,499.6	14.3
t	est	6.6	scalar	static	20,617.1	117.5
t	est	6.6	vector	static	20,631.2	14.1
N	-	∞	scalar	dynamic	21,305.9	674.7
t	xed	$\equiv 10$	scalar	dynamic	28,606.0	7,300.1
t	**est**	**5.7**	**scalar**	**dynamic**	**29,190.3**	**584.3**

Table 3 motivates our additional empirical choices. We estimate a range of models with

varying degrees of flexibility: normal versus Student's t, static versus dynamic covariance matrices, and a scalar versus a diagonal A_1. We observe two large likelihood improvements. First, allowing for fat-tailed rather than Gaussian mixtures increases the likelihood by more than 3,400 points for the static covariance case, and more than 7,800 points for dynamic covariance matrices. Second, allowing covariance matrices to be dynamic increases the likelihood more than 4,100 points for Gaussian mixtures, and more than 9,500 points for the Student's t case. Allowing A_1 to be diagonal only results in a minor likelihood increase, and the diagonal elements are all quite similar. We therefore adopt a Student's t model with scalar A_1, estimated degrees of freedom ν, and dynamic covariance matrices $\Omega_{j,t}$ as our main empirical specification. The autoregressive matrices are given by $A_1 = a_1 \cdot \mathrm{I}_D$, and $A_2 = a_2 \cdot \mathrm{I}_D$. Unknown parameters to be estimated in the M-step are therefore $\Theta = (a_1, a_2, \nu)'$. Using this parameter specification, we combine model parsimony with the ability to study a high-dimensional array of data.

Web Appendix C presents the estimated diagnostic statistics as defined in Section 2.5. Specifically, we report all time t posterior probabilities that unit i belongs to cluster component j for all 208 banks in our sample. In addition, we present histograms of the maximum average point-in-time and filtered component probabilities. The plots indicate that there is one most suitable business model for almost all banks in our data.

Web Appendix D compares our cluster allocation outcomes with a 2016 supervisory ECB/SSM bank survey ('thematic review') that asked a subset of banks in our sample which *other* banks they consider to follow a similar business model. We find that our classification outcomes for these banks approximately, but not perfectly, correspond to bank managements' own views.

Web Appendix E studies to which extent the clustering outcomes change by leaving out variables $d = 1, \ldots, D$ one-at-a-time and then re-estimating the model. All variables turn out to be important in the sense that they have a substantial influence on the clustering outcome. In addition, the clustering outcomes are not dominated by a single variable, such as for instance total assets.

Figure 2: Time-varying component medians

Filtered component medians for twelve indicator variables; see Table B.1. The component medians coincide with the component means unless the variable is transformed; see the last column of Table B.1 in Web Appendix B. The ownership variable is omitted since it is time-invariant. The component mean estimates are based on a t-mixture model with $J = 6$ components and time-varying component means $_{j,t}$ and covariance matrices $_{j,t}$. We distinguish large universal banks, including G-SIBs (black line), international diversi ed lenders (red line), fee-focused lenders (blue line), domestic diversi ed lenders (green dashed line), domestic retail lenders (purple dashed line), and small international banks (light-green dashed line).

4.3 Discussion of bank business models

This section studies the different business models implied by the $J = 6$ different component densities. Specifically, we assign labels to the identified components to guide intuition and for ease of reference. These labels are chosen in line with Figure 2, Figure F.1 in Web Appendix D, and the identities of the firms in each component. In addition, our labeling is approximately in line with the examples listed in SSM (2016, p.10).

Figure 2 plots the component median estimates for each indicator variable and business model component (except ownership, which is time-invariant). Web Appendix F presents additional figures, such as box plots of the time series averages of each variable for each business model group. In addition, Web Appendix F presents the filtered component-specific time-varying standard deviations $\sqrt{\Omega_{j,t}(d,d)}$ for variables $d = 1, \ldots, D-1$. The standard deviations tend to decrease over time starting from the high dispersion observed during the financial crisis (2008–2009). The standard deviation estimates also differ across business models.

We distinguish

(A) **Large universal banks, including G-SIBs** (14.9% of firms; comprising e.g. Barclays plc, Credit Agricole SA, Deutsche Bank AG.)

(B) **International diversi ed lenders** (11.1% of firms; e.g. ABN Amro NV, BBVA SA, Confederation Nationale du Credit Mutuel SA.)

(C) **Fee-focused bank** (15.9% of firms; e.g. Monte Dei Paschi di Sienna, Banco Populare, Bankinter SA.)

(D) **Domestic diversi ed lenders** (26.9 % of firms; e.g. Aareal Bank AG, Abanca Corporacion Bancaria SA, Alpha Bank SA.)

(E) **Domestic retail lenders** (17.8% of firms; e.g. Alandsbanken Abp, Berner Kantonalbank, Newcastle Building Society.)

(F) **Small international banks** (13.5% of firms; e.g. Alpha Bank Skopje, AS Citadele Banka, AS SEB Pank.)

Large universal banks, including G-SIBs (black line) stand out as the largest institutions, with up to €2 trn in total assets per firm for globally significantly important banks. Approximately 60% of operating revenue tends to come from interest-bearing assets such as loans and securities holdings. This leaves net fees & commissions as well as trading income as significant other sources. Large universal banks are the most leveraged at any time between 2008Q1–2015Q4, even though leverage, i.e., total assets to CET1 capital, decrease by more than a third from pre-crisis levels, from approximately 45 to below 30; see Figure 2. Large universal banks hold significant trading and derivative books, both in absolute terms and relative to total assets. Naturally, such large banks engage in significant cross-border activities, including lending (between 40-50% of loans are cross-border loans).

International diversi ed lenders (red line) are second in terms of firm size, with total assets ranging between approximately €100 – 500 bn per firm. As the label suggests, such banks lend significantly across borders and to both retail and corporate clients. The share of non-domestic loans to total loans is approximately 30%, and the share of retail loans ranges between approximately 20–60%. International diversified lenders also serve their corporate customers by trading securities and derivatives on their behalf, resulting in significant trading and derivatives books. In addition, such banks tend to be non-deposit funded, as indicated by a high loans-to-deposits ratio between 100 and 200%.

Fee-focused banks (blue line) achieve most of their income from net fees and commissions (approximately 30%). This group contains banks that focus on fee-based commercial banking activities, such as transaction banking services, trade finance, credit lines, advisory services, and guarantees. In addition, however, this component appears to contain 'weak' banks which do not generate much income in their traditional lines of business. The component mean for net interest income is low, raising the share of net fees and commissions. Fee-focused banks tend to exhibit a relatively high loans-to-assets ratio of approximately 70%, and also tend to focus on domestic loans (approximately 80%). Median total assets are typically below 100 bn per firm.

Domestic diversi ed lenders (green dashed line) are relatively numerous, comprising approximately 27% of firms, and are of moderate size. Total assets are typically below €50

bn per firm. Domestic diversified lenders tend to be well capitalized, as implied by relatively low leverage ratios (of typically less than 20). Trading and derivatives books are small. Lending is split approximately evenly between corporate and retail clients. Non-domestic loans are typically below 20%.

Finally, **domestic retail lenders** and **small international banks** are the smallest firms, with typically less than €25 bn in total assets. Domestic retail lenders and small international banks have much in common. Both types of banks display low leverage, suggesting they are well capitalized. The relatively largest part of their risk is credit risk (risk mix). Neither group holds significant amounts of securities or derivatives in trading portfolios. Approximately two-thirds of their income comes from interest-bearing assets, making it the dominant source of income.

Domestic retail lenders differ from small international banks in two ways: asset composition and geographical focus. Domestic retail lenders focus almost exclusively on loans (as indicated by a high loans-to-assets ratio) and domestic retail clients. By contrast, small international banks own substantial non-loan assets, and also serve non-domestic and non-retail (corporate) clients. The loans-to-deposits ratio is low for small international banks, at approximately one.

Figure 2 can also be used to discuss bank heterogeneity during the great financial crisis between 2008–2010 and the euro area sovereign debt crisis between 2010–2012, as well as overall banking sector trends during our sample. We refer the interested reader to Web Appendix G.

5 Bank business models and the yield curve

This section studies the extent to which banks adapt their business models to changes in the yield curve. We first review European interest rate developments before discussing parameter estimates.

Figure 3: Yield curve and factor plots

All yield curve and factor plots refer to AAA-rated euro area government bonds, and are based on a Svensson (1995) four-factor model. Yield factor estimates are taken from the ECB. The left panel plots tted Svensson yield curves on four dates { mid-2008Q1, mid-2010Q1, mid-2015Q1, and mid-2015Q4, for maturities between one and 20 years, and based on all yield curve factors. The right panel plots the level factor estimate, along with the model-implied short rate (given by the sum of the level and slope factor).

5.1 Low interest rates

Figure 3 plots fitted zero-coupon yield curves for maturities between one and twenty years at different times during our sample (left panel). European government bond yields experienced a pronounced downward shift during our sample, ultimately reaching ultra-low and in part negative values. The yield curve factors underlying the yield curve estimates are based on a Svensson (1995) four-factor model and are extracted daily from market prices of AAA-rated sovereign bonds issued by euro area governments. The yield curve factor estimates can be obtained from the ECB's website.

Figure 3 also plots the level factor, along with the implied short rate (right panel). The slope factor fluctuates around a value of approximately -2 in our sample, and is not reported. Long-term yields increase up to approximately 4% between 2009–2011 following an initial sharp drop during the global financial crisis. Between 2013–2015, nominal yields decline to historically low levels. In 2015, European 10-year rates are often below 1%. Short-term rates become negative in 2015 following a cut of the ECB's deposit facility rate to negative values. Low nominal interest rates do not necessarily only reflect unconventional monetary policies, including the ECB's Public Sector Purchase Programme (PSPP; or "Quantitative Easing"). Decreasing inflation rates, inflation risk premia, demographic factors, and an imbalance between global saving and investment likely also play a role; see e.g. Draghi (2016).

Table 4: Factor sensitivity estimates

Fixed e ects panel regression estimates for annual changes in bank-level accounting data $_4y_{it}(d)$, d = 1;:::;12, on a constant and contemporaneous and one-year lagged annual changes in yield curve factors level and slope. Speci cally, $_4y_{it}(d) = b_1\ _4\text{level}_t + b_2\ _4\text{slope}_t + b_3\ _4\text{level}_{t-4} + b_4\ _4\text{slope}_{t-4}$ + constant + xed e ects$_i$ + $_{it}$. Dependent variables are as listed in Table B.1 of the Web Appendix. Standard errors are Driscoll-Kraay standard errors with three lags; stars denote signi cance at a 10%, 5%, and 1% level.

	$\Delta_4\ln(\text{TA}_t)$	$\Delta_4\ln(\text{Lev}_t)$	$\Delta_4(\text{TL/TA})_t$	$\Delta_4\ln(\text{RM}_t)$	$\Delta_4(\text{AHFT/TA})_t$	$\Delta_4(\text{DHFT/TA})_t$
Δ_4 level$_t$	-0.0508***	-0.0260**	1.824***	0.00586	-0.00688**	-0.0111***
	(0.0118)	(0.0116)	(0.292)	(0.0125)	(0.00327)	(0.00354)
Δ_4 slope$_t$	-0.0514***	-0.0267**	1.470***	-0.00336	-0.00489	-0.0105***
	(0.0123)	(0.0112)	(0.303)	(0.0141)	(0.00304)	(0.00339)
Δ_4 level$_{t-4}$	-0.0169***	0.00685	0.213**	-0.00465	0.000441	0.00237
	(0.00422)	(0.00490)	(0.0984)	(0.00527)	(0.00133)	(0.00159)
Δ_4 slope$_{t-4}$	-0.0346***	-0.00777	0.118	-0.00742	-0.000742	-0.000591
	(0.00540)	(0.00567)	(0.143)	(0.00645)	(0.00153)	(0.00179)
constant	0.0104*	-0.0425***	0.236	0.0201***	-0.00238*	-0.00107
	(0.00578)	(0.00344)	(0.138)	(0.00541)	(0.00127)	(0.00140)
Observations	3,064	2,640	2,902	2,179	2,285	2,286
Number of groups	208	206	208	203	208	208
Bank FE	YES	YES	YES	YES	YES	YES
Within R^2	0.0508	0.00758	0.0251	0.00218	0.0207	0.0656

	$\Delta_4(\text{NII/OR})_t$	$\Delta_4(\text{NFC/OI})_t$	$\Delta_4(\text{TI/OI})_t$	$\Delta_4(\text{RL/TL})_t$	$\Delta_4(\text{DL/TL})_t$	$\Delta_4(\text{L/D})_t$
Δ_4 level$_t$	1.931	0.371	-0.0509	0.00562***	1.225***	-0.230
	(2.040)	(0.664)	(1.871)	(0.00194)	(0.291)	(0.498)
Δ_4 slope$_t$	2.712	1.606**	1.041	0.00615***	1.182***	-1.047
	(1.868)	(0.740)	(1.610)	(0.00174)	(0.395)	(0.700)
Δ_4 level$_{t-4}$	-1.160**	-1.293***	0.686	-0.00210***	-0.200	0.277
	(0.536)	(0.414)	(0.433)	(0.000714)	(0.196)	(0.279)
Δ_4 slope$_{t-4}$	-2.191***	-1.335**	1.243**	0.000269	-0.111	0.0282
	(0.631)	(0.572)	(0.509)	(0.00102)	(0.189)	(0.349)
constant	0.118	-0.0291	-0.0375	0.00728***	0.362	-1.973***
	(0.839)	(0.313)	(0.711)	(0.00117)	(0.212)	(0.362)
Observations	2,836	2,827	2,737	1,895	1,498	2,417
Number of groups	208	208	208	181	172	207
Bank FE	YES	YES	YES	YES	YES	YES
Within R^2	0.00287	0.00391	0.00745	0.00578	0.00633	0.00426

5.2 Fixed e ects panel regression results

Table 4 presents fixed effects panel regression estimates of bank-level accounting variables $\Delta_4 y_{i,t}(d)$, $d = 1, \ldots, 12$, on a constant and contemporaneous as well as one-year lagged changes in two yield curve factors, level and slope. We consider four-quarter differences since most banks report at an annual frequency. Table 4 pools bank data across business model components. Table 5 reports the regression coefficients for the one-year changes in the yield curve level, pooled as well as disaggregated across business model clusters. Web Appendix H reports all estimates for each variable and business model group.

We discuss five findings. First, as long-term interest rates decrease, banks on average grow larger in terms of total assets, by approximately 5% in response to a 100 bps drop in the level factor. The coefficient estimates for short term rates and lagged changes in yields are negative as well. This finding is in line with banks' incentive to extend the balance

Table 5: Sensitivities to changes in the term structure level ($\Delta_4 l_t$)

Yield level factor sensitivities for bank-level accounting data, controlling for changes in the slope factor and lagged changes in the yield curve. Factor sensitivity parameters are reported as pooled across all business models (column 2: All) as well as disaggregated across business model components A { F (columns 3 to 8). Estimates are obtained by xed e ects panel regression. Standard errors are Driscoll-Kraay standard errors with three lags.

Dependent variable	All	A	B	C	D	E	F
Δ_4 ln(TA)$_t$	-0.0508***	-0.126***	-0.0612***	-0.0501***	-0.0286***	-0.0375	-0.0301
	(0.0118)	(0.0216)	(0.00568)	(0.0150)	(0.00809)	(0.0432)	(0.0253)
Δ_4 ln(Lev)$_t$	-0.0260**	-0.0678*	-0.0358	0.0107	0.00223	-0.0745***	-0.0474
	(0.0116)	(0.0354)	(0.0235)	(0.0177)	(0.0204)	(0.0255)	(0.0341)
Δ_4 (TL/TA)$_t$	1.824***	3.391***	2.236***	2.495***	1.204***	1.446**	0.0682
	(0.292)	(0.666)	(0.394)	(0.234)	(0.258)	(0.611)	(0.619)
Δ_4 ln(RM)$_t$	0.00586	-0.0807	-0.101***	0.0567	0.0257*	0.0206	0.0850*
	(0.0125)	(0.0702)	(0.0254)	(0.0364)	(0.0143)	(0.0304)	(0.0462)
Δ_4 (AHFT/TA)$_t$	-0.00688**	-0.0290**	-0.00814***	-0.00926***	0.00168	0.000203	0.00138***
	(0.00327)	(0.0115)	(0.00203)	(0.00310)	(0.00168)	(0.000869)	(0.000416)
Δ_4 (DHFT/TA)$_t$	-0.0111***	-0.0474***	-0.0165***	-0.0107***	-0.00197	0.00135**	0.000454
	(0.00354)	(0.0135)	(0.00412)	(0.00164)	(0.00138)	(0.000554)	(0.000306)
Δ_4 (NII/OR)$_t$	1.931	-6.797	4.614	5.992**	2.985	0.184	-2.010
	(2.040)	(4.742)	(5.120)	(2.195)	(1.753)	(4.745)	(3.212)
Δ_4 (NFC/OI)$_t$	0.371	1.447	1.703	0.234	-0.986	0.184	2.591
	(0.664)	(0.954)	(1.767)	(0.728)	(0.987)	(1.631)	(2.216)
Δ_4 (TI/OI)$_t$	-0.0509	12.40***	0.623	-1.792	-1.755	-4.171	1.448
	(1.871)	(4.188)	(2.611)	(2.585)	(1.725)	(3.010)	(1.518)
Δ_4 (RL/TL)$_t$	0.00562***	0.00129	0.00252	0.0108***	0.00141	0.0113***	-0.00195
	(0.00194)	(0.00316)	(0.00480)	(0.00351)	(0.00373)	(0.00398)	(0.00828)
Δ_4 (DL/TL)$_t$	1.225***	1.158*	3.736***	1.418***	0.384	0.0161	2.581*
	(0.291)	(0.563)	(1.009)	(0.374)	(0.447)	(0.205)	(1.466)
Δ_4 (L/D)$_t$	-0.230	-1.754	-3.275*	2.924	-1.534	0.347	0.620
	(0.498)	(1.837)	(1.705)	(2.080)	(1.510)	(1.642)	(1.462)

sheet to offset squeezed net interest margins for new loans and investments. In addition, and trivially, some bank assets are worth more at lower rates.

Second, bank leverage is predicted to increase as yields decline. This correlation needs to be interpreted with caution. Leverage declined most strongly between 2010 and 2012, when euro area yields were increasing owing to the sovereign debt crisis. In addition, leverage is influenced by changes in financial regulation which we do not control for.

Third, the composition of bank assets is sensitive to changes in the yield curve factors. The loans-to-assets ratio decreases by approximately 2% on average across business models in response to a 100 bps drop in long-term rates. By contrast, the sizes of banks' trading and derivative books increase to some extent. This change in balance sheet composition is driven mostly by the larger banks (components A to C; Table 5), and could reflect a decreased demand for new loans from the private sector in an environment of strongly declining rates; see Abbassi et al. (2016). In this environment, large banks may invest in tradable securities such as government bonds instead of expanding their respective loan books; see Acharya and

Steffen (2015).

Fourth, we observe little variation in the shares of income sources in response to falling yields. In particular, the share of net interest income is not significantly (at 5%) associated with contemporaneous changes in yields. Two opposing effects could be at work. On the one hand, banks funding cost also decrease, and may even do so at a faster rate than long-duration loan rates. In addition, banks' long-term loans and bond holdings are worth more at lower rates, leading to mark-to-market gains. On the other hand, low long term interest rates squeeze net interest margins for newly acquired loans and bonds. The former effects could approximately balance the latter in our sample.

Finally, some banks appear to decrease their deposits-to-loans ratio in response to falling short-term rates. We refer to Web Appendix H, Table H.4, for the respective coefficient estimates. As the slope factor declines by 100 bps, banks in components A, B, and D decrease their deposits-to-loans ratio by approximately 2-5%.

Changes in term structure factors can also be added to the econometric specification as discussed in Section 2.4.3. Given the limited number of $T = 32$ time series observations, however, we need to pool the coefficients B_j across mixture components $B_j \equiv B$ to reduce the number of parameters. Web Appendix I discusses the parameter estimates. The increased log-likelihood values suggest a slightly better fit than the baseline specification. Coefficients in B are, however, rarely statistically significant according to their t-values.

We conclude that bank business model characteristics appear to adjust to changes in the yield curve. Given their direction for falling rates — increased size, increased leverage, increased complexity through larger trading and derivatives books, and possibly less stable funding sources — the effects are potentially problematic and need to be assessed from a financial stability perspective.

6 Conclusion

We proposed a novel score-driven finite mixture model for the study of banking data, accommodating time-varying component means and covariance matrices, normal and Student's t

distributed mixtures, and term structure factors as economic determinants of time-varying parameters. In an empirical study of European banks, we classified more than 200 financial institutions into six distinct business model components. Our results suggest that the global financial crisis and the euro area sovereign debt crisis had a substantial yet different impact on banks with different business models. In addition, banks' business models adapt over time to changes in long-term interest rates.

Acknowledgements

Lucas and Schaumburg thank the European Union Seventh Framework Programme (FP7-SSH/2007–2013, grant agreement 320270 - SYRTO) for financial support. Schaumburg also thanks the Dutch Science Foundation (NWO, grant VENI451-15-022) for financial support. Parts of this paper were written while Schwaab was on secondment to the ECB's Single Supervisory Mechanism (SSM). We are particularly grateful to Klaus Düllmann, Heinrich Kick, and Federico Pierobon from the SSM. We also thank the three Referees and the Editor whose many insightful suggestions have helped us to reshape and improve the paper.

References

Abbassi, P., R. Iyer, J.-L. Peydro, and F. Tous (2016). Securities trading by banks and credit supply: Micro-evidence. *Journal of Financial Economics 121*(3), 569{594.

Acharya, V. and S. Ste en (2015). The \greatest" carry trade ever? Understanding eurozone bank risks. *Journal of Financial Economics 115*(2), 215 { 236.

Altunbas, Y., S. Manganelli, and D. Marques-Ibanez (2011). Bank risk during the nancial 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.

Bankscope (2014). Bankscope user guide. Bureau van Dijk, Amsterdam, January 2014. Available to subscribers.

Calinski, T. and J. Harabasz (1974). A dendrite method for cluster analysis. *Communications in Statistics 3*(1), 1{27.

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 (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*(4), 552{563.

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.

Creal, D. D., R. B. Gramacy, and R. S. Tsay (2014). Market-based credit ratings. *Journal of Business & Economic Statistics 32*, 430{444.

Davies, D. L. and D. W. Bouldin (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

de Amorim, R. C. and C. Hennig (2015). Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences 324*(10), 126145.

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.

Draghi, M. (2016). Addressing the causes of low interest rates. *Speech in Frankfurt am Main, 2 May 2016*.

Fruehwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Fruehwirth-Schnatter, S. and S. Kaufmann (2008). Model-based clustering of multiple time series. *Journal of Business and Economic Statistics 26*, 78{89.

Hamilton, J. D. and M. T. Owyang (2012). The propagation of regional recessions. *The Review of Economics and Statistics 94*, 935{947.

Hannoun, H. (2015). Ultra-low or negative interest rates: what they mean for nancial stability and growth. *Speech given at the BIS Eurofi high-level seminar in Riga on 22 April 2015*.

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

Heider, F., F. Saidi, and G. Schepens (2017). Life below zero: Bank lending under negative policy rates. *Working paper, available at SSRN: https://ssrn.com/abstract=2788204*.

Liu, C. and D. B. Rubin (1994). The ecme algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika*, 633{648.

Lucas, A. and X. Zhang (2016). Score driven exponentially weighted moving average and value-at-risk forecasting. *International Journal of Forecasting 32(2)*, 293302.

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

Meng, X.-L. and D. B. Rubin (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika 80*(2), 267{278.

Milligan, G. W. and M. C. Cooper (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika 50*(2), 159{179.

Nouy, D. (2016). Adjusting to new realities banking regulation and supervision in Europe. Speech by Daniele Nouy, Chair of the ECB's Supervisory Board, at the European Banking Federation's SSM Forum, Frankfurt, 6 April 2016.

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

Smyth, P. (1996). Clustering sequences with hidden markov models. *Advances in Neural Information Processing Systems 9*, 1{7.

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

Svensson, L. E. O. (1995). Estimating Forward Interest Rates with the Extended Nelson & Siegel Method. *Quarterly Review, Sveriges Riksbank* (3), 13{26.

Wang, Y., R. S. Tsay, J. Ledolter, and K. M. Shrestha (2013). Forecasting simultaneously high-dimensional time series: A robust model-based clustering approach. *Journal of Forecasting 32*(8), 673{684.

Yellen, J. L. (2014). Monetary Policy and Financial Stability. *Michel Camdessus Central Banking Lecture, International Monetary Fund, Washington, D.C., 2 July 2014*.

Web Appendix to

"Bank business models at zero interest rates"

Andre Lucas, Julia Schaumburg, Bernd Schwaab

Web Appendix A: Details on model selection when the number of components is unknown

In our empirical study of banking data, the number of clusters, i.e. bank business models, is unknown a priori. A number of model selection criteria and so-called cluster validation indices have been proposed in the literature, and comparative studies have not found a dominant criterion that performs best in all settings; see e.g. Milligan and Cooper (1985) and de Amorim and Hennig (2015). We therefore run an additional simulation study to see which model selection criteria are suitable to choose the optimal number of components in our multivariate panel setting.

All criteria and de nitions are listed in Web Appendix A.1. The simulation outcomes are reported in Web Appendix A.2. Overall, we nd that the standard likelihood-based information criteria such as AIC and BIC tend to systematically overestimate the number of clusters in a multivariate panel setting. Distance-based criteria corresponding to the within-cluster sum of squared errors perform better. The Davies-Bouldin index (DBI; see Davies and Bouldin (1979)) and the Calinski-Harabasz index (CHI; see Calinski and Harabasz (1974)) do well. In an empirical setting, therefore, we recommend looking at a variety of model selection criteria for a full picture, but to pay particular attention to DBI and CHI whenever the selection criteria disagree.

A.1 Information criteria and selection indices

Standard likelihood-based information criteria As our framework is likelihood-based, we can apply standard information criteria such as AICc and BIC for model selection. AICc is the likelihood-based Akaike criterion, corrected by a nite sample adjustment; see Hurvich and Tsai (1989). However, these criteria might not give a large enough penalty when too many clusters are assumed. Therefore, we also consider a range of other criteria from the literature, that have been applied to select the number of components in di erent settings.

Distance-based information criteria The average within-cluster sum of squared errors simply measures the average deviation of an observation from its respective cluster mean. It does not take into account how well clusters are separated from each other. Therefore, we expect this

criterion to over t in some cases.

$$wSSE(J) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{i \in C_j} k y_{i;t} \quad {}_{j;t} k^2 \tag{A.1}$$

where kxk denotes the Euclidean norm of vector x.

The Akaike information criterion has been adapted to select the number of clusters when using the k-means algorithm (see Peel and McLachlan (2000) and Aggarwal and Reddy (2014)). It adds a penalty function to the within-cluster sum of squared errors:

$$AICk(J) = wSSE(J) + 2JD$$

with $wSSE(J)$ as de ned in (A.1). Bai and Ng (2002) provide criteria to optimally choose the number of factors in factor models for large panels. Their setting is similar to ours, so we include the criterion they suggest in their study:

$$IC_2(J) = \frac{1}{ND} wSSE(J) + \frac{J(N+T)}{NT} \ln(c_{NT}^2)$$

with $wSSE(J)$ from (A.1) and sequence $c_{NT} = \min f \sqrt{p}_{\overline{N;}} \sqrt{p}_{\overline{T}} g$.

Calinski-Harabasz Index The Calinski-Harabasz Index (CHI), which was introduced by Calinski and Harabasz (1974), has been successful in various comparative simulation studies, see, e.g., Milligan and Cooper (1985). We de ne a weighted version of the within-cluster sum of squared errors as

$$wSSE_M(J) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{i \in C_j} d(y_{i;t}; \quad {}_{j;t})$$

with $d(x;y)$ denoting the Mahalanobis distance between two vectors x and y, i.e. $d(y_{i;t}; \quad {}_{j;t}) = (y_{i;t} \quad {}_{j;t})' \quad {}^{1}(y_{i;t} \quad {}_{j;t})$ where is the unconditional covariance matrix of the data. Similarly, we de ne the weighted between-cluster sum of squared errors as

$$bSSE_M(J) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{J} N_j d(y_{i;t}; y_t)$$

with $y_t = \frac{1}{N} \sum_{i=1}^{N} y_{i;t}$ and N_j denoting the number of units contained in cluster j. Then, the CHI corresponding to J clusters is

$$CHI(J) = \frac{N - J}{J - 1} \frac{bSSE_M(J)}{wSSE_M(J)};$$

J is chosen such that $CHI(J)$ is as large as possible.

Davies-Bouldin index The Davies-Bouldin index (DBI, see Davies and Bouldin (1979)) also utilizes the Mahalanobis distance. It relies on a similarity measure for each pair of clusters l and m, which is often de ned as

$$R_{l;m} = \frac{s_l + s_m}{d_{l;m}}$$

where $s_j = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_j} d(y_{i;t}; \ _{j;t})$, measures the distance of a typical observation in cluster j to its mean, and $d_{l;m} = \frac{1}{T} \sum_{t=1}^{T} d(\ _{l;t}; \ _{m;t})$ measures the distance between the centroids of clusters l and m. J is then chosen such that the maximum ratio of (weighted) within- and between cluster sums of squares is as small as possible:

$$DBI(J) = \frac{1}{J} \sum_{j=1}^{J} R_j$$

with

$$R_j = \max\{R_{jk}\}; \quad j \neq k;$$

Average silhouette index The silhouette index measures how well each observation ts into its respective cluster by comparing its within-cluster cohesion with its dissimilarity to the observations in the closest neighboring cluster. The silhouette value lies between -1 and 1, and the larger, the better is the t. It was rst introduced in Rousseeuw (1987) and its average over all observations has been shown to work well as a tool to determine the number of clusters in simulation studies, see, for instance, de Amorim and Hennig (2015)

$$SI = \frac{1}{N} \sum_{i=1}^{N} \frac{b(Y_i) - a(Y_i)}{\max\{a(Y_i); b(Y_i)\}}$$

where $a(Y_i)$ is the average distance over time of $Y_i \in S_j$ to all other $Y_m \in S_j$ and $b(Y_i)$ is the average distance to all $Y_m \in S_l$, where S_l is the closest cluster to S_j.

A.2 Simulation tables for an unknown number of clusters

We rely on the simulation settings as described in Section 3.1. This means we continue to simulate from bivariate mixtures consisting of two components, distinguishing between high and low signal-to-noise ratios, and overlapping and non-overlapping mean trajectories. As before, we consider the impact of two types of mis-speci cation. By contrast, however, we do not assume that the number of clusters is equal to the true number (two) when estimating the model, but also estimate model parameters assuming one and three mixture components, respectively.

Tables A.1 and A.2 list the number of times each information criterion chooses 1, 2, or 3 as the number of clusters. Overall, we nd that the standard likelihood-based information criteria AICc and BIC tend to systematically overestimate the number of clusters. As a result, these criteria do not work well. Distance-based criteria corresponding to the within-cluster sum of squared errors perform better, but almost all of them break down in one or more of the considered settings. The Davies-Bouldin index (DBI; see Davies and Bouldin (1979)) chooses the correct number of clusters in all cases reported here. Similarly, the Calinski-Harabasz index (CHI; see Calinski and Harabasz (1974)) often gives good indications as well. In our empirical application, we therefore report a variety of model selection criteria to get a full picture, but pay particular attention to these two validation indices whenever the criteria disagree.

Table A.1: Number of clusters

Chosen number of clusters by ten information criteria over 100 simulation runs. Correct speci cation refers to the case of simulating from a normal mixture and estimating assuming a mixture of normal distributions. In the case of mis-speci cation 1, data are simulated from a t-mixture with 5 degrees of freedom, but the model is estimated assuming a Gaussian mixture distribution. In the case of mis-speci cation 2, a $t(3)$-mixture is used to simulate the data, and a xed value = 5 is assumed in the estimation.

radius=4, distance=8									
	correct spec.			misspec. 1			misspec. 2		
no. clusters	1	2	3	1	2	3	1	2	3
AICc	0	65	35	0	66	34	0	54	46
BIC	0	70	30	0	71	29	0	57	43
SSE	0	69	31	0	75	25	0	56	44
AICk	0	100	0	0	100	0	0	94	6
BNG1	0	100	0	0	100	0	0	85	15
BNG2	0	100	0	0	100	0	0	86	14
BNG3	0	100	0	0	99	1	0	84	16
CHI	0	100	0	0	100	0	0	100	0
SI	0	85	15	0	89	11	0	75	25
DBI	0	100	0	0	100	0	0	100	0

radius=4, distance=0									
	correct spec.			misspec. 1			misspec. 2		
no. clusters	1	2	3	1	2	3	1	2	3
AICc	0	53	47	0	55	45	0	54	46
BIC	0	55	45	0	58	42	0	58	42
SSE	0	70	30	0	71	29	0	56	44
AICk	0	100	0	0	100	0	0	98	2
BNG1	0	100	0	0	100	0	0	90	10
BNG2	0	100	0	0	100	0	0	91	9
BNG3	0	100	0	0	100	0	0	86	14
CHK	0	100	0	0	100	0	0	100	0
SI	0	96	4	0	99	1	0	79	21
DBI	0	100	0	0	100	0	0	100	0

Table A.2: Number of clusters, ctd.

Chosen number of clusters by ten information criteria over 100 simulation runs. Correct specification refers to the case of simulating from a normal mixture and estimating assuming a mixture of normal distributions. In the case of misspecification 1, data are simulated from a t-mixture with 5 degrees of freedom, but the model is estimated assuming normal mixtures. In the case of misspecification 2, a $t(3)$-mixture is used to simulate the data while in the estimation, a fixed value = 5 is assumed.

radius=1, distance=8									
	correct spec.			misspec. 1			misspec. 2		
no. clusters	1	2	3	1	2	3	1	2	3
AICc	0	55	45	0	64	36	0	60	40
BIC	0	63	37	0	67	33	0	60	40
SSE	0	68	32	0	68	32	0	55	45
AICk	0	100	0	0	100	0	0	96	4
BNG1	0	100	0	0	100	0	0	80	20
BNG2	0	100	0	0	100	0	0	81	19
BNG3	0	100	0	0	99	1	0	78	22
CHK	0	100	0	0	100	0	0	100	0
SI	0	99	1	0	98	2	0	93	7
DBI	0	100	0	0	100	0	0	100	0

radius=1, distance=0									
	correct spec.			misspec. 1			misspec. 2		
no. clusters	1	2	3	1	2	3	1	2	3
AICc	0	53	47	1	55	44	0	59	41
BIC	0	55	45	2	56	42	0	64	36
SSE	0	64	36	0	67	33	14	46	40
AICk	100	0	0	100	0	0	94	6	0
BNG1	1	99	0	61	39	0	67	28	5
BNG2	4	96	0	66	34	0	70	25	5
BNG3	0	100	0	45	55	0	58	35	7
CHI	0	100	0	0	98	2	0	100	0
SI	0	100	0	0	100	0	0	99	1
DBI	0	100	0	0	100	0	0	100	0

Web Appendix B: Data details

Our sample under study consists of $N = 208$ European banks, for which we consider quarterly bank-level accounting data from SNL Financial between 2008Q1 { 2015Q4. Banks that underwent distressed mergers, were acquired, or ceased to operate for other reasons during that time, are excluded from the analysis. We assume that di erences in the remaining banks' business models can be characterized along six dimensions: size, complexity, activities, geographical reach, funding strategies, and ownership structure. We select a parsimonious set of $D = 13$ indicators from these six categories. Table B.1 lists the respective indicators.

We augment the proprietary data from SNL Financial with con dential supervisory data from the European Central Bank. While the publicly-available data is generally of good quality, it occasionally has insu cient coverage for some bank-level variables given the purpose of this study. Speci cally, we incorporate information from bank-level Finrep reports for euro area signi cant institutions when the respective information is missing from SNL Financial. This is necessary for the bank-level breakdown between retail loans and commercial loans, and the share of domestic loans to total loans.

Our augmented multivariate panel data is unbalanced in the time dimension. Missing values occur routinely 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 (back lling). If variables are missing in the beginning of the sample, we use the most adjacent future value. In the cross-section, we require at least one entry for each variable and bank.

We consider banks at their highest level of consolidation. In addition, however, we also 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 (54%) and the European Union (E.U., 73%). European non-E.U. banks are located in Norway (12%), Switzerland (4%), and other countries (11%).

Table B.1 also reports the data transformation used in the applied modeling. For example, some ratios lie strictly within the unit interval. We transform such ratios into unbounded continuous variables by mapping them through an inverse Probit transform. We take natural logarithms of large numbers, such as total assets, CET1 capital, derivatives held for trading, etc. Finally, the

Table B.1: Indicator variables

Bank-level panel data variables for the empirical analysis. We consider J=13 indicator variables covering six different categories. The third column explains which transformation is applied to each indicator before the statistical analysis. $^{1}(\,)$ denotes the inverse Probit transform.

Category	Variable	Transformation
Size	1. Total assets	ln (Total Assets)
	2. Leverage w.r.t. CET1 capital	$\ln \frac{\text{Total Assets}}{\text{CET1 capital}}$
Complexity	3. Net loans to assets	$^{1}\,\frac{\text{Loans}}{\text{Assets}}$
	4. Risk mix	$\ln \frac{\text{Market Risk}+\text{Operational Risk}}{\text{Credit Risk}}$
	5. Assets held for trading	$\frac{\text{Assets in trading portfolios}}{\text{Total Assets}}$
	6. Derivatives held for trading	$\frac{\text{Derivatives held for trading}}{\text{Total Assets}}$
Activities	7. Share of net interest income	$\frac{\text{Net interest income}}{\text{Operating revenue}}$
	8. Share of net fees & commission income	$\frac{\text{Net fees and commissions}}{\text{Operating income}}$
	9. Share of trading income	$\frac{\text{Trading income}}{\text{Operating income}}$
	10. Retail loans	$\frac{\text{Retail loans}}{\text{Retail and corporate loans}}$
Geography	11. Domestic loans ratio	$^{1}\,\frac{\text{Domestic loans}}{\text{Total loans}}$
Funding	12. Loan-to-deposits ratio	$\frac{\text{Total loans}}{\text{Total deposits}}$
Ownership	13. Ownership index	-

Note: Total Assets are all assets owned by the company (SNL key field 131929). CET1 capital is Tier 1 regulatory capital as defined by the latest supervisory guidelines (220292). 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, and 248880, respectively), which are as reported by the company. Trading portfolio assets are assets acquired principally for the purpose of selling in the near term (224997). 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). Retail loans are expressed as a percent of retail and corporate loans (226957). 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.

`ownership' variable combines information on ownership and organizational structure with information on whether a bank is listed at a stock exchange. It distinguishes listed stock corporations, non-listed stock corporations, other limited liability companies, mutual/cooperative banks, and government-owned/state banks. We standardize the resulting categorical variable to zero mean and unit cross-sectional variance, and add standard-normally distributed error terms to make the variable `ownership index' continuous. Doing so allows us to also consider information on organizational structure within our framework of normal and Student's *t*-distributed mixtures. We checked that omission of the ownership variable only leads to small di erences in cluster allocation among small banks. The choice of number of clusters is robust to the inclusion of the ownership variable.

Web Appendix C: Diagnostic statistics

C.1 Point-in-time probability plots

Figures C.1 { C.3 plot the point-in-time component membership probabilities $_{ijjt}$ as given in (23). Most rms are fairly unequivocally allocated to one component. Persistent switches of component membership appear to be rare. In a few cases, however, component membership may have changed. If so, the new cluster is almost always a `neighboring' cluster for which the component means are relatively similar. Such rms may be `in-between'-cases, and harder to classify.

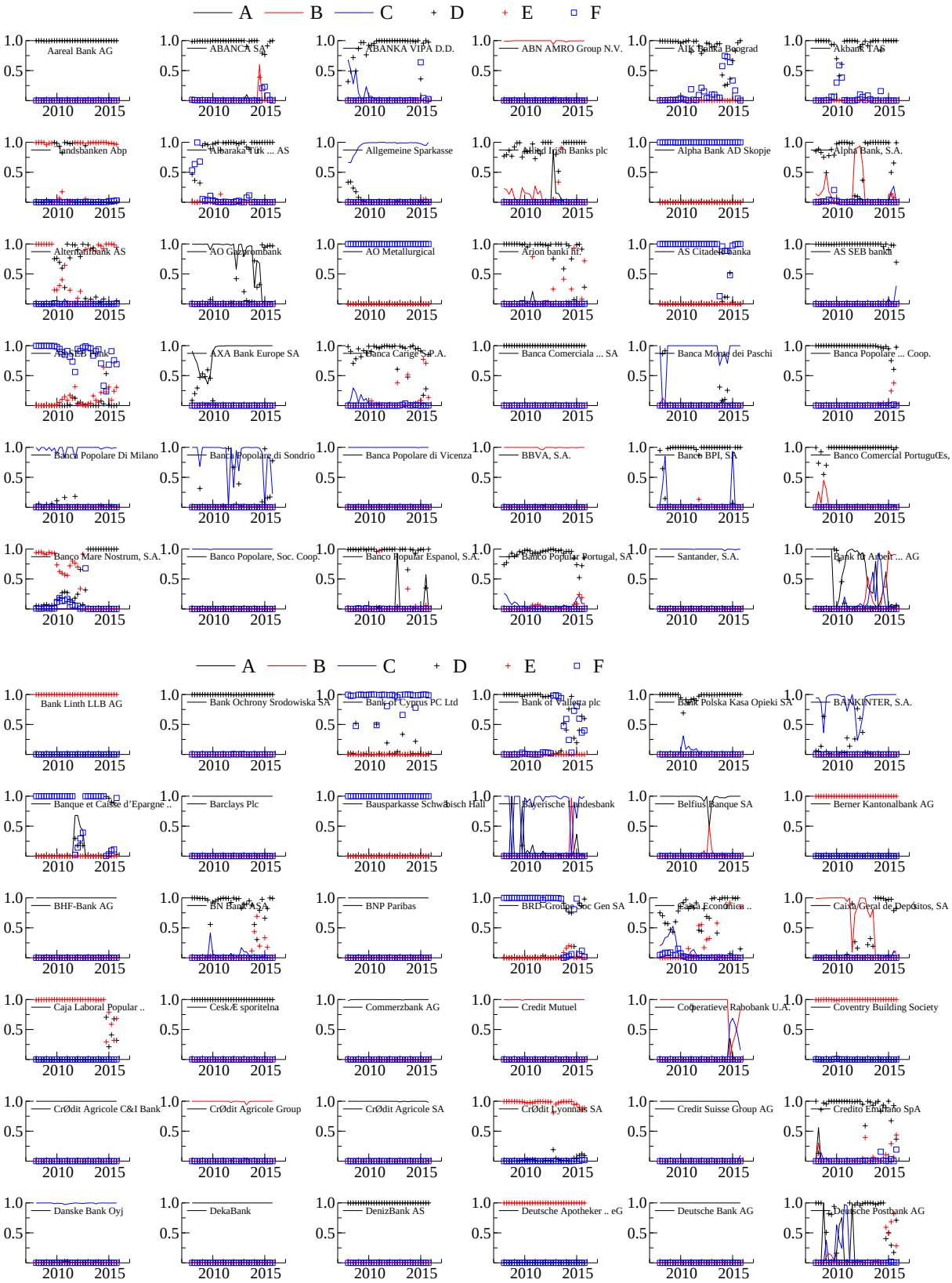

Figure C.1: Point-in-time component probabilities

Point-in-time posterior probabilities $_{ijjt}$ for rms 1-72.

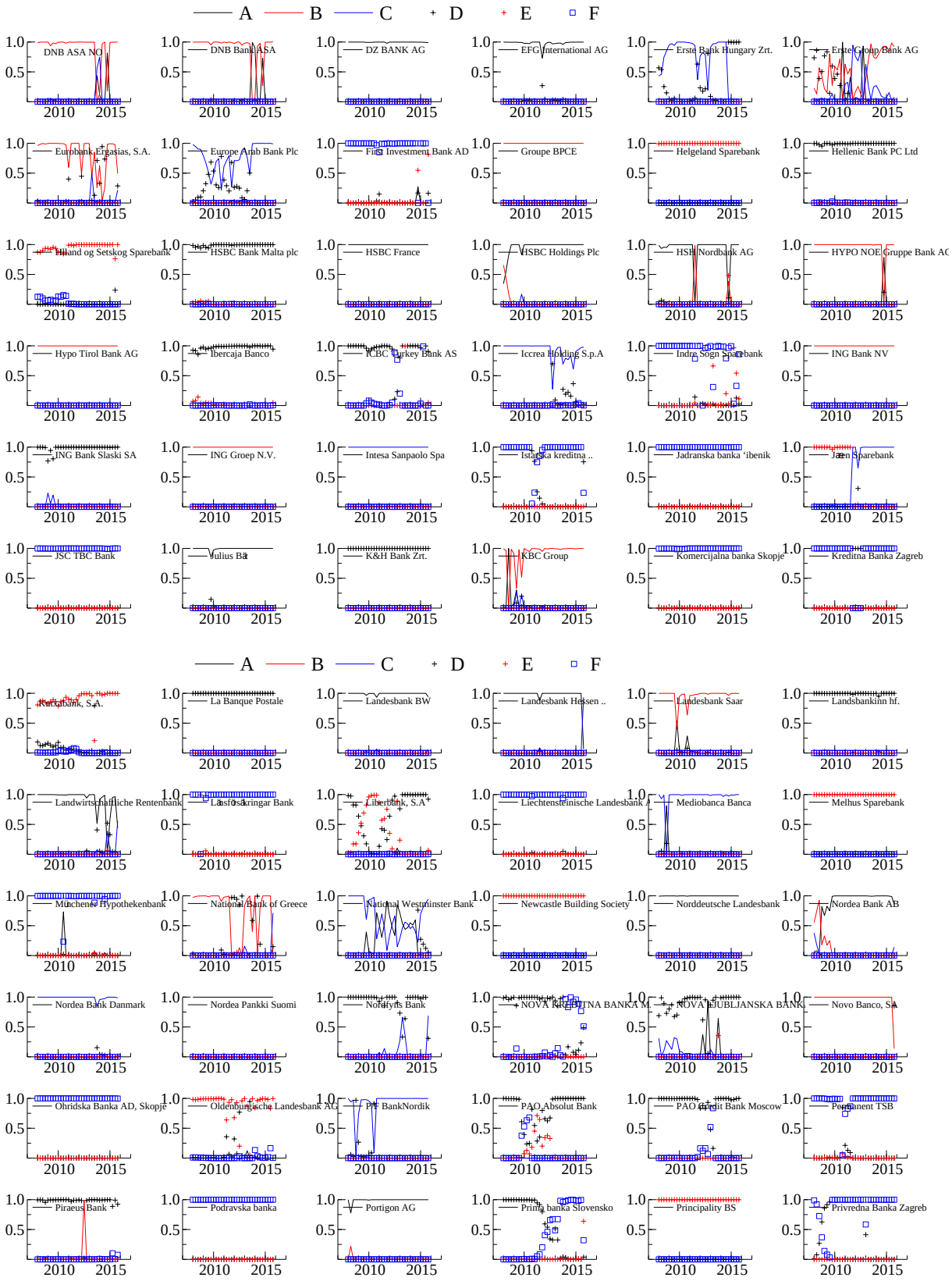

Figure C.2: Point-in-time component probabilities

Point-in-time component membership probabilities $_{ijjt}$ for rms 73-144.

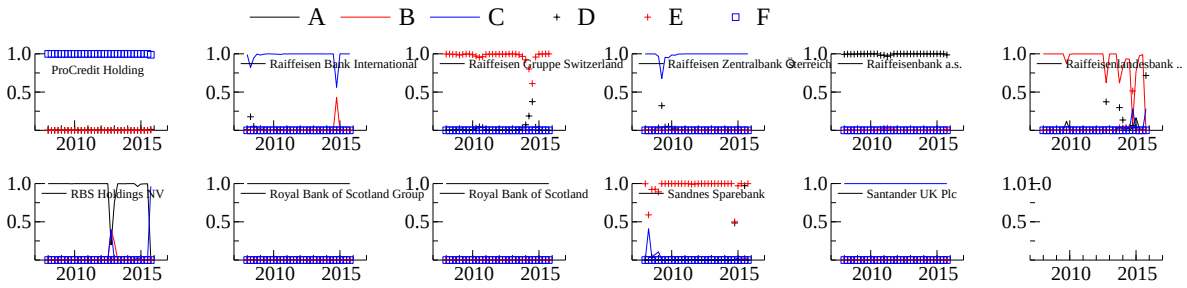

Figure C.3: Point-in-time component probabilities

Point-in-time component membership probabilities $_{ijjt}$ for rms 145-208.

C.2 Histograms of maximum average point-in-time and filtered component probabilities

Figure C.4 shows histograms of the maximum average point-in-time components probabilities ($\max_j T^{-1} \sum_{t=1}^{T} \hat{\gamma}_{i;jjt}$) for the 208 banks in the sample (left panel). The point-in-time component probabilities are defined as in equation (23)). The right panel shows a similar histogram for maximum *filtered* component probabilities as defined below (23). Both histograms show that the average probability of almost each bank to belong to its respective cluster is high, and many banks are classified to belong to their clusters with probability 1 across all time points.

Figure C.4

Web Appendix D: Supervisory survey on banks' peers

This section compares our classi cation results to a 2016 supervisory ECB/SSM bank survey (`thematic review').[1] The con dential survey asked SSM-supervised banks which *other* banks they considered to follow a similar business model. Of the banks that responded to the survey, 78 are in our sample (37.5%). Each bank could list up to ten peers.

The survey is not an ideal benchmark for our clustering outcomes, for several reasons. First, banks may have listed peers that are similar in only one aspect, such as a similar major business line, instead of peers at the consolidated level as the latter are considerably more di cult to determine (requiring, for example, a modeling framework as the one presented in the main text). Second, the survey responses may be biased towards larger and better-known banks. Third, country-e ects may play a role: banks frequently reported peers that are located in the same jurisdiction. Country location, however, played no role in our classi cation analysis. Fourth, strategic reporting may have been present in some cases. Finally, the survey was undertaken at one point in time (2016), while our empirical results are based on earlier multivariate panel data (between 2008{2015).

Table D.1: Contingency table

For banks in each business model (column 1) we report the fraction of business model membership of the reported peers. The two highest entries in a row are printed in bold. We distinguish A: large universal banks, including G-SIBs; B: corporate/wholesale-focused banks; C: fee-based banks/asset managers; D: small diversi ed lenders; E: domestic retail lenders; and F: mutual/cooperative-type banks.

	#1	#2	#3	#4	#5	#6	Total
A	**63%**	**25%**	0%	6%	6%	0%	16
B	0%	**30%**	**70%**	0%	0%	0%	10
C	19%	**31%**	**50%**	0%	0%	0%	16
D	0%	0%	**44%**	19%	4%	**33%**	27
E	0%	0%	0%	50%	50%	0%	2
F	0%	0%	**57%**	14%	0%	**29%**	7
Total							78

Despite these caveats, we may still expect to nd most of the reported peers to come from the same, or at least an adjacent[2], business model component as determined by our multivariate panel

[1] The clustering presented in the main text is not used for micro-prudential supervision. We thank the ECB MS I-IV team that assembled and conducted the thematic review.

[2] Components A{F are ordered in terms of declining average size (total assets), see Figure 2. In addition, the components are approximately ordered in terms of complexity, ranging from large universal banks, including G-SIBs (most complex), to smaller and more traditional banks (least complex).

data methodology. This is the case. Table D.1 reports the respective fractions. The two highest entries in a row are printed in bold. Rows A{C and E have more than 80% of responses in the same and adjacent columns. Smaller banks in rows D and F frequently listed larger (C) banks as their peers; these banks are probably comparable in only one business line, not at the consolidated level, see the caveats above. We conclude that our business model classi cation outcomes approximately corresponds to bank managements' own views.

Web Appendix E: Influence analysis

How sensitive are the clustering outcomes reported in the main paper to changes in the information set? For example, does the size of a bank, as captured by (log) total assets, play a dominant role? To explore these questions we de ne a divergence measure between clustering outcomes as

$$\text{div}(k) = \sqrt{\frac{2}{J(J-1)N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{J} \hat{\gamma}_{ij} \left(\tilde{\gamma}_{ij}^{\text{restr.}, k} \right)^2}; \tag{E.1}$$

where $\hat{\gamma}_{ij}$ are the full sample posterior probability estimates used in the main paper, and $\tilde{\gamma}_{ij}^{\text{restr.}, k}$ are the corresponding component membership probabilities estimated from a restricted sample which leaves out variable k. Table E.1 presents the deviation statistics.

The three highest entries of (E.1) are observed for *i*) the breakdown of total loans into retail and commercial loans, *ii*) the domestic loans to total loans ratio, and *iii*) the loans to assets ratio. These ratios play a major role in distinguishing the di erent business model groups. The rst variable, log total assets, is not particularly in uential for the clustering outcome. We note, however, that the information from the rst variable is to some extent also contained in the second variable, log leverage.

Table E.1: In uence of variables on clustering outcomes

The divergence statistics as de ned in (E.1). The three most in uential entries are printed in bold.

Variable	div(k)
1. Total assets	0.028
2. Leverage w.r.t. CET1 capital	0.120
3. Net loans to assets	**0.163**
4. Risk mix	0.098
5. Assets held for trading	0.152
6. Derivatives held for trading	0.133
7. Share of net interest income	0.139
8. Share of net fees & commission income	0.110
9. Share of trading income	0.147
10. Retail loans	**0.235**
11. Domestic loans ratio	**0.198**
12. Loan-to-deposits ratio	0.130
13. Ownership index	0.144

Web Appendix F: Additional figures

F.1 Data box plots across business models

Figure F.1 reports box plots for time-averages of the indicator variables; see Table B.1 for the variable descriptions.

Figure F.1: Box plots for indicator variables

We report box plots for twelve indicator variables; see Table B.1 for the variable descriptions. For each variable, we consider the time series average over $T = 32$ quarters. In each panel, we distinguish (from left to right): A: large universal banks, including G-SIBs; B: international diversi ed lenders; C: fee-based banks; D: domestic diversi ed lenders; E: domestic retail lenders; and F: small international banks.

F.2 Time-varying component standard deviations

Figure F.2 plots the filtered component-specific time-varying standard deviations $\text{std.dev}_{j;t}(d) = \sqrt{\Sigma_{j;t}(d;d)}$ for variables $d = 1;:::;D$ 1. Allowing for time-varying component covariance matrices results in significant increases in log-likelihood; see Table 3. The off-diagonal elements of $\Sigma_{j;t}$ are not reported. Overall, the standard deviations tend to decrease over time from the high levels observed during the financial crisis between 2008{2009. The variables (log) total assets and (the inverse Probit of) the share of domestic loans to total loans are particularly dispersed across units within a given component.

Figure F.2: Time-varying standard deviations

Filtered time-varying standard deviations $\text{std.dev}_{j,t}(d) = \sqrt{\Sigma_{j,t}(d;d)}$ for variables $d = 1;:::;D$ 1. Each panel refers to a business model component A { F, and contains twelve standard deviation estimates over time (one for each variable in Table B.1, except ownership, which is time-invariant). Mean and standard deviation estimates are based on a t-mixture model with $J = 6$ components and dynamic covariance matrices $\Sigma_{j,t}$.

Web Appendix G: Additional discussion of time-varying component location parameters

G.1 Bank heterogeneity during crises

Banks may be more or less resilient to the occurrence of a financial crisis depending on their respective business model. For example, in a study of U.S. and European banks during the financial crisis between 2008–2009, Altunbas et al. (2011) argue that firms with higher leverage, larger size, greater reliance on market funding, and aggressive credit growth before the crisis suffered more when the crisis hit. Conversely, a strong customer deposit base and greater income diversification were mitigating factors. Beltratti and Stulz (2012) confirm that bank balance sheet and profitability characteristics played a role, and point to banks' risk profiles and risk governance as additional factors. Finally, in a study of U.S. banks, Chiorazzo et al. (2016) argue that smaller traditional banks, including mutual/co-operative-type banks, were the most resilient to stress.

In our study of European data, Figure 2 in the main text suggests that the global financial crisis between 2008–2009 had a differential impact on banks with different business models. Large universal banks, including G-SIBs, appear to be the most affected. Many such banks held substantial amounts of mortgage-related securities at the beginning of the 2008–2009 financial crisis. This exposure explains the high ratio of net interest income to total income in 2008 for these banks, as well as the negative contributions of trading income to total income in that year (panels 7 and 9 of Figure 2). By contrast, domestic retail lenders (E) and small international banks (F) experienced less variability in component means, particularly in the share of income sources.

We also observe differences in the variability of component means during the euro area sovereign debt crisis between 2010Q1–2012Q2; see ECB (2014). This time, large universal banks (A) and fee-based banks (C) appear to be the most affected. For both groups, net interest income decreases, to below 60%, and market risks increases relative to credit risk (risk mix). Conversely, and again, domestic retail lenders and small international banks were relatively less affected. This is in line with the findings by Chiorazzo et al. (2016), for a different sample of banks.

G.2 Post-crisis banking sector trends

This section continues our discussion of Figure 2 with a focus on banking sector trends. We also refer to ECB (2016) for a detailed discussion.

The financial crisis of 2008–2009 and subsequent new regulatory requirements have had a profound impact on banks' activities and business models. Pre-crisis profitability levels of many European banks were supported by high leverage ratios, reliance on relatively cheap wholesale (non-deposit) funding, as well as, in some cases, elevated risk-taking with concentration risks in securitization exposures and/or commercial real estate; see Ayadi et al. (2014) and ECB (2016). Changes in the post-crisis regulatory framework[3], however, have rendered some of these previously profitable business strategies much less viable. In addition, adverse macroeconomic outcomes and financial market conditions during the euro area sovereign debt crisis between 2010–2012 weakened most European banks further.

Figure 2 suggests that banks adapt their respective business models to a changing financial and regulatory environment. We focus on three main developments. First, we observe a stark reduction in leverage (second panel in Figure 2) and wholesale funding (bottom right panel). Before the crisis, euro area banks were more highly leveraged, on average, than their global peers, according to the IMF (2009) and ECB (2016).[4] After the crisis, banks' adjustment to higher capital requirements appear to have contributed to lower leverage ratios. The only exception is formed by mutual/co-op banks, whose leverage ratios have increased over time. Similarly, new regulatory requirements and the increased cost of wholesale funding seem to have pushed European banks to reduce their over-reliance on wholesale funding sources.

Second, changes in regulation made certain business lines more costly, in particular trading activities, thus providing banks with an incentive to scale down these activities.[5] Panels five and six of Figure 2 suggest that large universal banks substantially reduced their trading activities between 2012Q2–2015Q4. Derivatives held for trading declined from approximately 18% to approximately

[3] Examples include the Basel III rules, the Capital Requirements Directive (CRD) IV, the Bank Recovery and Resolution Directive, mandatory bail-in rules, and the move to single banking supervision (SSM) and a single resolution mechanism (SRM) in the euro area.

[4] Caveats apply. Relatively higher leverage ratios may have been related to different institutional practices, such as mortgage balance sheet retention. In addition, differences in accounting standards may have mattered, such as the different treatment of derivatives under IFRS and US GAAP.

[5] For E.U. banks, the so-called Liikanen report recommended increased capital requirements for trading business lines, and a mandatory separation of proprietary trading and other high-risk trading from core activities. The Liikanen report was published in October 2012.

14% of total assets on average for these banks. Assets held for trading declined from approximately 28% to 25% of total assets.

Third, there is some evidence of a shift towards retail business and away from investment banking and corporate/wholesale lending activities. Speci cally, the share of retail loans to total loans is increasing for most banks, particularly in the low interest rate environment between 2012{ 2015.

Web Appendix H: Term structure factor sensitivities

Tables H.1 – H.4 present factor sensitivity estimates obtained by fixed effects panel regression. Parameter estimates are either pooled across business models (column 2) and disaggregated across business models (columns 3 to 8). The pooled estimates are identical to the fixed effect panel regression estimates presented in the main text; see Table 4. Tables H.1 – H.4 report Driscoll-Kraay standard error estimates with three lags.

Table H.1: Disaggregated factor sensitivities

Yield factor sensitivities for bank-level accounting data. Factor sensitivity parameters are reported as pooled across business models (column 2) as well as disaggregated across business model components A { F (columns 3 to 8). Parameter estimates and standard errors are obtained by xed e ects panel regression. Standard errors are Driscoll-Kraay standard errors with three lags.

	Dependent variable: $_4$ ln(Total Assets)$_t$						
	All	A	B	C	D	E	F
$_4$ level$_t$	-0.0508***	-0.126***	-0.0612***	-0.0501***	-0.0286***	-0.0375	-0.0301
	(0.0118)	(0.0216)	(0.00568)	(0.0150)	(0.00809)	(0.0432)	(0.0253)
$_4$ slope$_t$	-0.0514***	-0.124***	-0.0662***	-0.0596***	-0.0256**	-0.0306	-0.0371
	(0.0123)	(0.0228)	(0.00804)	(0.0150)	(0.00921)	(0.0394)	(0.0253)
$_4$ level$_{t\ 4}$	-0.0169***	0.00622	-0.00888**	-0.0143***	-0.0162***	-0.0454***	-0.0266***
	(0.00422)	(0.00992)	(0.00341)	(0.00291)	(0.00261)	(0.0128)	(0.00924)
$_4$ slope$_{t\ 4}$	-0.0346***	-0.0287**	-0.0257***	-0.0442***	-0.0169***	-0.0700***	-0.0406***
	(0.00540)	(0.0132)	(0.00429)	(0.00590)	(0.00375)	(0.0173)	(0.00862)
Constant	0.0104*	-0.0411***	-0.00999*	0.00783	0.0174***	0.0494***	0.0283***
	(0.00578)	(0.0133)	(0.00490)	(0.00729)	(0.00502)	(0.0162)	(0.00915)
Observations	3,064	424	376	581	901	487	295
Number of groups	208	31	23	33	56	37	28
Within R^2	0.0508	0.215	0.152	0.145	0.0191	0.0501	0.0697
	Dependent variable: $_4$ ln(Leverage)$_t$						
$_4$ level$_t$	-0.0260**	-0.0678*	-0.0358	0.0107	0.00223	-0.0745***	-0.0474
	(0.0116)	(0.0354)	(0.0235)	(0.0177)	(0.0204)	(0.0255)	(0.0341)
$_4$ slope$_t$	-0.0267**	-0.0768**	-0.0435*	0.0116	0.00206	-0.0551*	-0.0666*
	(0.0112)	(0.0362)	(0.0252)	(0.0177)	(0.0222)	(0.0273)	(0.0388)
$_4$ level$_{t\ 4}$	0.00685	0.0128	0.0159	0.0117*	-0.00779	0.00930	0.000299
	(0.00490)	(0.0175)	(0.00987)	(0.00664)	(0.00898)	(0.0179)	(0.0243)
$_4$ slope$_{t\ 4}$	-0.00777	-0.0113	-0.0120	-4.47e-05	-0.0185*	0.00154	-0.00507
	(0.00567)	(0.0170)	(0.0110)	(0.00653)	(0.00996)	(0.0172)	(0.0188)
Constant	-0.0425***	-0.0675***	-0.0621***	-0.0345***	-0.0298***	-0.0463***	-0.0327
	(0.00344)	(0.0175)	(0.0103)	(0.00641)	(0.0102)	(0.0141)	(0.0207)
Observations	2,640	385	330	512	716	455	242
Number of groups	206	31	23	33	56	36	27
Within R^2	0.00758	0.0519	0.0475	0.0209	0.00447	0.0113	0.0273
	Dependent variable: $_4$ (Loan-to-Assets ratio)$_t$						
$_4$ level$_t$	1.824***	3.391***	2.236***	2.495***	1.204***	1.446**	0.0682
	(0.292)	(0.666)	(0.394)	(0.234)	(0.258)	(0.611)	(0.619)
$_4$ slope$_t$	1.470***	2.867***	2.102***	2.050***	0.672**	1.483**	-0.276
	(0.303)	(0.648)	(0.486)	(0.288)	(0.276)	(0.697)	(0.654)
$_4$ level$_{t\ 4}$	0.213**	0.186	-0.0928	0.714***	0.173	-0.245	-0.0133
	(0.0984)	(0.300)	(0.177)	(0.0943)	(0.151)	(0.293)	(0.259)
$_4$ slope$_{t\ 4}$	0.118	0.784**	-0.0369	0.702***	-0.197	-0.161	-0.658**
	(0.143)	(0.358)	(0.263)	(0.112)	(0.203)	(0.321)	(0.314)
Constant	0.236	0.914***	0.197	0.316**	-0.288***	0.582*	0.0182
	(0.138)	(0.276)	(0.212)	(0.128)	(0.0716)	(0.337)	(0.304)
Observations	2,902	368	341	530	885	487	291
Number of groups	208	31	23	33	56	37	28
Within R^2	0.0251	0.120	0.0654	0.0590	0.0268	0.0114	0.0244

Table H.3: Disaggregated factor sensitivities, ctd.

Yield factor sensitivities for bank-level accounting data. Factor sensitivity parameters are reported as pooled across business models (column 2) as well as disaggregated across business model components A { F (columns 3 to 8). Parameter estimates and standard errors are obtained by xed e ects panel regression. Standard errors are Driscoll-Kraay standard errors with maximum lag 3.

Dependent variable: $_4$ (Net interest income/Operating revenue)$_t$							
Cluster	All	A	B	C	D	E	F
$_4$ level$_t$	1.931	-6.797	4.614	5.992**	2.985	0.184	-2.010
	(2.040)	(4.742)	(5.120)	(2.195)	(1.753)	(4.745)	(3.212)
$_4$ slope$_t$	2.712	-4.185	9.581	5.449**	1.983	2.432	-3.232
	(1.868)	(5.466)	(6.127)	(2.131)	(1.837)	(5.062)	(3.405)
$_4$ level$_{t\ 4}$	-1.160**	-2.262*	-3.095	-0.348	0.131	-1.947	-3.034***
	(0.536)	(1.174)	(2.170)	(0.617)	(0.863)	(2.644)	(0.971)
$_4$ slope$_{t\ 4}$	-2.191***	-5.764***	-5.356**	-1.426	0.482	-2.323	-4.436***
	(0.631)	(1.642)	(2.488)	(0.985)	(1.093)	(3.257)	(1.309)
Constant	0.118	-0.988	0.947	0.918	-0.250	0.315	-1.198
	(0.839)	(2.158)	(3.878)	(1.333)	(0.838)	(2.465)	(0.913)
Observations	2,836	363	337	511	847	488	290
Number of groups	208	31	23	33	56	37	28
Within R^2	0.00287	0.0286	0.0119	0.00784	0.00183	0.00547	0.0502
Dependent variable: $_4$ (Fee & commissions income/Operating income)$_t$							
$_4$ level$_t$	0.371	1.447	1.703	0.234	-0.986	0.184	2.591
	(0.664)	(0.954)	(1.767)	(0.728)	(0.987)	(1.631)	(2.216)
$_4$ slope$_t$	1.606**	2.757**	4.382*	0.497	0.0402	1.767	3.889*
	(0.740)	(1.147)	(2.121)	(0.638)	(1.107)	(1.798)	(2.207)
$_4$ level$_{t\ 4}$	-1.293***	-2.243***	-1.452*	-1.213***	-1.392	-1.290	0.844
	(0.414)	(0.586)	(0.737)	(0.206)	(0.897)	(0.945)	(0.905)
$_4$ slope$_{t\ 4}$	-1.335**	-2.091***	-1.630	-1.068***	-2.114	-0.895	1.834
	(0.572)	(0.479)	(1.005)	(0.322)	(1.314)	(1.249)	(1.429)
Constant	-0.0291	-0.184	0.197	0.356	-0.441	-0.270	0.980
	(0.313)	(0.867)	(1.241)	(0.257)	(0.856)	(0.827)	(0.814)
Observations	2,827	365	336	501	847	488	290
Number of groups	208	31	23	33	56	37	28
Within R^2	0.00391	0.0177	0.0165	0.0261	0.00205	0.0117	0.0270
Dependent variable: $_4$ (Trading income/Operating income)$_t$							
$_4$ level$_t$	-0.0509	12.40***	0.623	-1.792	-1.755	-4.171	1.448
	(1.871)	(4.188)	(2.611)	(2.585)	(1.725)	(3.010)	(1.518)
$_4$ slope$_t$	1.041	14.55***	0.587	-0.0568	-0.786	-3.110	1.636
	(1.610)	(4.831)	(2.344)	(2.168)	(1.532)	(2.844)	(1.509)
$_4$ level$_{t\ 4}$	0.686	0.575	1.988***	1.078	0.0964	0.189	1.125***
	(0.433)	(0.948)	(0.581)	(0.639)	(0.614)	(0.930)	(0.372)
$_4$ slope$_{t\ 4}$	1.243**	3.255**	2.931***	1.586*	0.0644	0.0130	1.710***
	(0.509)	(1.488)	(1.012)	(0.811)	(0.827)	(1.301)	(0.335)
Constant	-0.0375	1.492	0.923	-0.334	0.0888	-1.698	0.364
	(0.711)	(1.731)	(1.094)	(1.048)	(0.772)	(1.317)	(0.301)
Observations	2,737	329	328	499	835	477	269
Number of groups	208	31	23	33	56	37	28
Within R^2	0.00745	0.0448	0.0150	0.0234	0.00416	0.0166	0.0226

Table H.4: Disaggregated factor sensitivities, ctd.

Yield factor sensitivities for bank-level accounting data. Factor sensitivity parameters are reported as pooled across business models (column 2) as well as disaggregated across business model components A { F (columns 3 to 8). Parameter estimates and standard errors are obtained by xed e ects panel regression. Standard errors are Driscoll-Kraay standard errors with maximum lag 3.

Dependent variable: $_4$ (Retail loans/Retail and commercial loans)$_t$							
Cluster	All	A	B	C	D	E	F
$_4$ level$_t$	0.00562***	0.00129	0.00252	0.0108***	0.00141	0.0113***	-0.00195
	(0.00194)	(0.00316)	(0.00480)	(0.00351)	(0.00373)	(0.00398)	(0.00828)
$_4$ slope$_t$	0.00615***	0.00161	0.00905*	0.0125**	0.00598	0.00249	-0.00188
	(0.00174)	(0.00289)	(0.00521)	(0.00456)	(0.00391)	(0.00438)	(0.00911)
$_4$ level$_{t\ 4}$	-0.00210***	-0.00438	-0.00766***	-0.00175	-0.00252	0.00193	-0.00193
	(0.000714)	(0.00261)	(0.00191)	(0.00209)	(0.00191)	(0.00136)	(0.00498)
$_4$ slope$_{t\ 4}$	0.000269	-0.00122	0.00280	0.000726	-0.000720	-0.00326	0.00751
	(0.00102)	(0.00262)	(0.00266)	(0.00122)	(0.00243)	(0.00265)	(0.00616)
Constant	0.00728***	0.00404*	0.00165	0.00823***	0.0140***	0.00730***	-0.00500
	(0.00117)	(0.00231)	(0.00288)	(0.00172)	(0.00235)	(0.00230)	(0.00436)
Observations	1,895	232	222	301	564	359	217
Number of groups	181	22	23	24	50	35	27
Within R^2	0.00578	0.0149	0.122	0.0151	0.0176	0.0481	0.0184
Dependent variable: $_4$ (Domestic loans/Total loans)$_t$							
$_4$ level$_t$	1.225***	1.158*	3.736***	1.418***	0.384	0.0161	2.581*
	(0.291)	(0.563)	(1.009)	(0.374)	(0.447)	(0.205)	(1.466)
$_4$ slope$_t$	1.182***	0.975	3.281**	1.335***	0.843	-0.204	2.157
	(0.395)	(0.667)	(1.270)	(0.347)	(0.619)	(0.197)	(1.541)
$_4$ level$_{t\ 4}$	-0.200	-0.110	-0.972	-0.880***	-0.0727	-0.0505	0.812*
	(0.196)	(0.470)	(0.836)	(0.163)	(0.197)	(0.0907)	(0.404)
$_4$ slope$_{t\ 4}$	-0.111	-0.406	-0.753	-0.952***	0.0230	-0.105	1.529**
	(0.189)	(0.431)	(0.863)	(0.151)	(0.195)	(0.118)	(0.575)
Constant	0.362	0.394	0.0225	0.0795	0.359	-0.138*	1.669***
	(0.212)	(0.369)	(0.750)	(0.125)	(0.250)	(0.0742)	(0.514)
Observations	1,498	227	162	234	426	252	197
Number of groups	172	27	20	22	42	36	25
Within R^2	0.00633	0.00929	0.0423	0.0671	0.0140	0.0224	0.0188
Dependent variable: $_4$ (Loans-to-deposits ratio)$_t$							
$_4$ level$_t$	-0.230	-1.754	-3.275*	2.924	-1.534	0.347	0.620
	(0.498)	(1.837)	(1.705)	(2.080)	(1.510)	(1.642)	(1.462)
$_4$ slope$_t$	-1.047	-2.254	-4.992**	2.143	-2.711*	0.318	0.856
	(0.700)	(2.436)	(2.223)	(2.362)	(1.506)	(1.805)	(1.312)
$_4$ level$_{t\ 4}$	0.277	-0.381	-0.663	1.676**	1.335**	0.104	-3.743***
	(0.279)	(0.764)	(0.906)	(0.753)	(0.636)	(0.711)	(0.657)
$_4$ slope$_{t\ 4}$	0.0282	-0.828	-2.825**	2.533**	0.345	1.086	-2.350***
	(0.349)	(0.878)	(1.157)	(1.015)	(0.549)	(0.692)	(0.518)
Constant	-1.973***	-2.045*	-2.666**	-2.261*	-1.537***	-1.581**	-3.443***
	(0.362)	(1.152)	(1.220)	(1.138)	(0.424)	(0.597)	(0.610)
Observations	2,417	320	273	406	771	391	256
Number of groups	207	31	22	33	56	37	28
Within R^2	0.00426	0.00371	0.0395	0.0209	0.0168	0.0138	0.0466

Web Appendix I: Extended score-driven model

Term structure factors can also be added to the econometric speci cation as discussed in Section 2.4.3. Given the limited number of $T = 32$ time series observations, we pool the coe cients B_j across mixture components $B_j \quad B$ to reduce the number of parameters. Falling interest rates have been argued to push banks to do more business at squeezed rates, take more risk, and change their asset composition; see e.g. Hannoun (2015), Abbassi et al. (2016), and Heider et al. (2017). Consequently, yield curve factors could help predict future average business model characteristics.

Table I.1 reports the parameter estimates. The increased log-likelihood values suggest a slightly better t than the baseline speci cation. Most coe cients in B, however, are statistically insignificant according to their t-values. When signi cant, the predictive e ects appear to be small in economic terms. For example, a 100 bps drop in the level factor predicts a fall in the ratio of net interest income to operating revenue of 1.3 percentage points over the next quarter.

Table I.1: Factor sensitivity estimates: GAS-X model

We report parameter estimates associated with the extended GAS-X model as introduced in (22); see Section 2.4.3. The baseline model M0 contains no additional conditioning variables. Model M1 uses changes in the level factor l_t to predict one-quarter ahead changes in the mean of each variable. Model M2 uses changes in the short rate $(l_t + s_t)$ for prediction. Coe cients B_j are pooled across mixture components $B_j \quad B$, and scaled up by a factor of 100 for readability. Conditioning variables are in percentage points (ppt). Estimation sample is 2008Q1 { 2015Q4.

	M0		M1		M2	
X_t	{		l_t		$(l_t + s_t)$	
	par	t-val	par	t-val	par	t-val
A_1	0.95	23.6	0.93	21.4	0.91	21.2
A_2	0.62	67.7	0.62	62.9	0.62	67.6
	5.66	17.2	5.65	16.9	5.65	17.0
$\ln(TA)_t$			-0.12	0.1	-1.67	0.9
$\ln(Lev)_t$			0.10	0.3	1.40	2.2
$(TL/TA)_t$			0.07	0.7	0.12	0.7
$\ln(RM)_t$			0.26	0.6	-0.83	1.1
$(AHFT/TA)_t$			-0.01	1.3	0.01	0.4
$(DHFT/TA)_t$			-0.01	1.5	0.01	0.6
$(NII/OR)_t$			0.35	2.6	1.30	4.6
$(NFC/OI)_t$			0.00	0.0	0.33	2.1
$(TI/OI)_t$			-0.30	2.6	-1.62	5.8
$(RL/TL)_t$			-0.09	0.6	-0.32	1.3
$(DL/TL)_t$			-0.67	1.0	-1.16	1.0
$(TL/Dep)_t$			0.14	0.4	0.08	0.1
loglik	29,190.3		29,197.1		29,212.9	

References

Abbassi, P., R. Iyer, J.-L. Peydro, and F. Tous (2016). Securities trading by banks and credit supply: Micro-evidence. *Journal of Financial Economics 121* (3), 569{594.

Aggarwal, C. C. and C. K. Reddy (2014). *Data Clustering. Algorithms and Applications*. Chapman & Hall/CRC.

Altunbas, Y., S. Manganelli, and D. Marques-Ibanez (2011). Bank risk during the nancial 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.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 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.

Calinski, T. and J. Harabasz (1974). A dendrite method for cluster analysis. *Communications in Statistics 3* (1), 1{27.

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

Davies, D. L. and D. W. Bouldin (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

de Amorim, R. C. and C. Hennig (2015). Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences 324* (10), 126145.

ECB (2014). The determinants of euro area sovereign bond yield spreads during the crisis. ECB Monthly Bulletin article, May 2014.

ECB (2016). Recent trends in eirp area banks' business models and implications for banking sector stability. *ECB Financial Stability Review, May 2016, Special Feature C*.

Hannoun, H. (2015). Ultra-low or negative interest rates: what they mean for nancial stability and growth. *Speech given at the BIS Euro high-level seminar in Riga on 22 April 2015*.

Heider, F., F. Saidi, and G. Schepens (2017). Life below zero: Bank lending under negative policy rates. *Working paper, available at SSRN: https://ssrn.com/abstract=2788204*.

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

IMF (2009). Detecting systemic risk. *IMF Global Financial Stability Report, April 2009, Ch. 3*.

Milligan, G. W. and M. C. Cooper (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika 50*(2), 159{179.

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

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics 20*, 53{65.