University of Glasgow

Adam Smith Business School

WORKING PAPER SERIES

Bayesian Nonparametric Inference in Bank Business Models with Transient and Persistent Cost Inefficiency.

Dimitris Korobilis, Emmanuel C. Mamatzakis and Vasileos Pappas

Paper No. 2025-02
March 2025

Bayesian Nonparametric Inference in Bank Business Models with Transient and Persistent Cost Inefficiency

Dimitris Korobilis [*1,2], Emmanuel C. Mamatzakis[3], and Vasileios Pappas[4]

[1]Adam Smith Business School
[2]BI Norwegian Business School
[3]Birkbeck Business School
[4]Surrey Business School

March 3, 2025

Abstract

This paper introduces a novel econometric framework for identifying and modeling bank business models (BBMs), which dynamically evolve in response to changing financial and economic conditions. Building on the stochastic frontier literature, we extend the traditional cost-efficiency models by decomposing inefficiency into persistent and transient components. We propose a Bayesian nonparametric approach that adapts to the data through an infinite mixture model with predictor-dependent clustering, enabling a flexible classification of banks into distinct business models. Our method, based on the Logit Stick-Breaking Process (LSBP), provides a data-driven way to capture the heterogeneity in bank strategies, allowing for dynamic transitions between business models over time. This model offers a significant advancement over existing parametric and kernel-based approaches by combining the scalability of nonparametric methods with efficient computational routines. We apply the model to a dataset of European banks and identify four distinct business model clusters, providing novel insights into the evolution of bank performance and efficiency. Our findings contribute to the growing literature on the identification and measurement of bank business models, offering valuable implications for policy and regulatory frameworks.

Keywords: Stochastic Frontier Analysis; Bayesian Nonparametrics; Logit Stick-Breaking Process; Dynamic Clustering; Cost Efficiency

JEL Classification: C11; C14; C23; D24; G21; G28

This paper is part of the research activities at the Centre for Applied Macroeconomics and commodity Prices (CAMP) at BI Norwegian Business School.

*Corresponding Author: Dimitris Korobilis, Professor of Econometrics, Adam Smith Business School, University of Glasgow, 2 Discovery Place, Glasgow, G11 6EY, United Kingdom, Dimitris.Korobilis@glasgow.ac.uk

1 Introduction

Understanding the nature and evolution of bank business models (BBMs) is a fundamental issue in banking research. The core challenge lies in identifying and modeling how banks' strategic decisions on asset composition, funding sources, and income generation drive their business models, especially in light of economic shocks and regulatory changes (DeYoung and Rice, 2004a,b; DeYoung et al., 2004). While much of the banking literature has focused on the classification of business models using cross-sectional data (Cerqueiro et al., 2011; Zott et al., 2011), the dynamic nature of these models — particularly in response to shifts in market conditions — remains underexplored. This paper contributes to the econometric literature by proposing a novel dynamic framework for identifying BBMs, which accounts for both persistent and transient inefficiencies within a performance measurement model (Tsionas and Kumbhakar, 2014; Badunenko et al., 2021).

We extend existing approaches by incorporating a flexible Bayesian nonparametric model, which allows for the endogenous identification of business models that evolve over time. The proposed model adapts the classical stochastic frontier framework to disentangle cost inefficiency into short-run and long-run components, with special attention to unobserved bank heterogeneity (Greene, 2005; Kumbhakar et al., 2007). By employing a mixture of distributions, we allow for the data-driven clustering of banks based on their evolving characteristics (Custodio João et al., 2024). In contrast to traditional clustering techniques, which often impose rigid assumptions about business model transitions, our framework enables the seamless capture of these transitions as banks adjust their strategies in response to both internal factors and external shocks, such as financial crises and regulatory changes (Fahlenbrach et al., 2012; van Oordt and Zhou, 2019).

Our starting point is the four-way error component stochastic frontier model of Tsionas and Kumbhakar (2014) that is able to decompose inefficiency into persistent and transient components. We extend this model into a Bayesian nonparametric framework, which leverages an infinite mixture representation with predictor-dependent clustering, enabling a flexible, data-driven classification of banks into distinct business models (BBMs). To capture the complexity of bank behavior, we employ a Logit Stick-Breaking Process (LSBP) (Ren et al., 2011), which introduces predictor-dependent mixture weights. This allows us to model bank-specific heterogeneity and dynamic transitions between BBMs over time. By extending existing nonparametric methods, our approach offers a more structured and scalable solution to identifying BBMs compared to traditional parametric and kernel-based methods (Dunson and Park, 2008; Dunson and

Rodríguez, 2011), while maintaining computational efficiency through Pólya-gamma data augmentation (Polson et al., 2013). We derive an efficient Markov chain Monte Carlo (MCMC) algorithm for this model that is able to handle the large dimensions of our empirical dataset. By means of synthetic data experiments, we establish the good numerical properties of our new algorithm and estimator.

Our approach is particularly relevant for empirical studies of bank performance and efficiency. First, we contribute to the literature on the econometric identification of BBMs by offering a method that allows for the flexible classification of banks over time, incorporating both the evolution of their business model and their efficiency (Tsionas and Kumbhakar, 2014). Second, by introducing a Bayesian nonparametric framework with predictor-dependent clustering, we overcome the limitations of standard clustering methods that assume banks transition smoothly between models (Custodio João et al., 2023; Kumbhakar et al., 2007). This innovation is especially important for understanding the impact of policy changes and market conditions on bank performance, providing richer insights into the long-term sustainability and short-term adjustments of banking strategies (Assaf et al., 2019; Tsionas et al., 2023a). Third, in our empirical application to 493 European banks, we identify four distinct business model clusters over the period 2008–2022. The model reveals heterogeneity across banks in terms of their strategic choices and operational efficiencies though one BBM, that focuses on traditional bank asset and funding management, appears dominant. Notably, our model reveals that banks respond to shocks, such as the global financial crisis, by switching between business models to recover part of their losses in efficiency. These findings emphasize the underlying dynamics of BBM and the role of strategic transitions in improving cost efficiency in response to shocks (Badunenko and Kumbhakar, 2017; Wheelock and Wilson, 2000).

The remainder of the paper is structured as follows. Section 2 outlines the econometric methodology of our proposed model, detailing the Bayesian nonparametric framework and its application to BBM identification. Section 3 presents the results from a series of Monte Carlo simulations, evaluating the performance of our model in terms of both parameter estimation accuracy and cluster identification. Section 4 applies the model to a dataset of European banks, providing an empirical analysis of the identified business models and their efficiency dynamics over time. Finally, Section 5 concludes with a summary of the main findings, policy implications, and suggestions for future research.

2 Econometric Methodology

We adopt the production technology framework from Badunenko et al. (2021) and Badunenko and Kumbhakar (2017), which uses cost functions to capture inefficiencies in input use, aligning with standard banking studies that emphasize cost minimization. To model the cost function in panel data, we use the four-way error component stochastic frontier model by Tsionas and Kumbhakar (2014), which decomposes inefficiency into persistent and transient components. The log-cost function for bank $i = 1, ..., n$ at time $t = 1, ..., T_i$ is

$$y_{it} = h(x_{it}, \beta) + a_i + \eta_i^+ + u_{it}^+ + v_{it}, \tag{1}$$

where $y_{it} = \log(c_{it})$ is the logarithm of total cost, $h(x_{it}, \beta)$ is a translog function with parameters β and covariates x_{it} that consist of bank input prices and bank outputs.[1] Here $\eta_i^+ \geq 0$ and $u_{it}^+ \geq 0$ denote, respectively, persistent inefficiency that is consistent over time and transient inefficiency that varies with time, while a_i reflects unobserved bank-specific differences, and v_{it} represents traditional random noise. Therefore, the model is completed by assuming $a_i \sim N(0, \sigma_a^2)$, $\eta_i^+ \sim N_+\left(0, \sigma_\eta^2\right)$, $u_{it}^+ \sim N_+(0, \sigma_u^2)$, $v_{it} \sim N(0, \sigma_v^2)$, where σ_i^2 are scalar variance parameters for $i = a, \eta, u, v$ and N_+ denotes the half-normal distribution.

Badunenko et al. (2021) tailor the model above to the problem of BBM identification, by allowing the variances σ_η^2 and σ_u^2 to be functions of bank-specific characteristics. In contrast, we propose an alternative method for BBM identification, following the literature that employs mixture models to classify banks into different BBMs.

2.1 A smoothly mixing nonparametric panel data stochastic frontier model

To model complex dependencies and heterogeneity in bank behavior, we extend the stochastic frontier model in equations (1)–(3) using a Bayesian nonparametric framework. Bayesian nonparametric methods allow model complexity to adapt to the data, using an infinite mixture representation of the general form

$$f(y) = \int K(y; \theta)\, p(d\theta) \equiv \sum_{k=1}^{\infty} \pi_k K(y; \theta),$$

[1] In the empirical section we approximate the function $h(\bullet)$ using linear, quadratic, and cross-product terms in x_{it}. The precise form does not influence our analysis in this section, provided that the function remains linear in the parameters β. This linearity assumption is not restrictive, as it encompasses the polynomial (quadratic) approximation we adopt in our empirical exercise, as well as highly nonlinear approximations such as splines and Gaussian processes.

where $K(y;\theta)$ is a parametric kernel and p is a random probability measure and π_k a discrete probability. We extend this framework by employing an infinite mixture representation with predictor-dependent weights, enabling a flexible clustering of banks based on their characteristics. To address dependencies on covariates, we adopt the Logit Stick-Breaking Process (LSBP) proposed by Ren et al. (2011), which introduces predictor-dependent mixture weights $\pi_k(z_{it})$, with z_{it} a $q \times 1$ vector, providing a structured and data-driven model for bank-specific behavior.

In place of the parametric kernel $K(y;\theta)$ we use the stochastic frontier model of Tsionas and Kumbhakar (2014) introduced previously. Therefore, the proposed mixture model is expressed as:

$$y_{it} = \sum_{k=1}^{\infty} \pi_k\left(z_{it}\right)\left[h\left(x_{it},\beta_k\right) + a_i + \eta_i^{+} + u_{it,k}^{+} + v_{it,k}\right], \tag{2}$$

where $k = 1, 2, 3, \ldots$ indexes the mixture components, $\pi_k(z_{it}) = \Pr(G_i = k \mid z_{it})$ are predictor-dependent probabilities, and $G_i \in \mathbb{N}$ denotes the cluster assignment. Variances of the noise components $u_{it,k}$ and $v_{it,k}$ are cluster-specific, while the random effects a_i and η_i^{+} vary only by unit i:

$$a_i \sim N\left(0, \sigma_a^2\right), \tag{3a}$$
$$\eta_i^{+} \sim N^{+}\left(0, \sigma_\eta^2\right), \tag{3b}$$
$$u_{it,k}^{+} \sim N^{+}\left(0, \sigma_{u,k}^2\right), \tag{3c}$$
$$v_{it,k} \sim N\left(0, \sigma_{v,k}^2\right). \tag{3d}$$

The translog function $h(x_{it},\beta_k)$ has cluster-specific parameters β_k.

The LSBP defines the weights $\pi_k(z_{it})$ using a stick-breaking process:

$$\pi_k\left(z_{it}\right) = \nu_k\left(z_{it}\right)\prod_{l=1}^{k-1}\left[1 - \nu_l\left(z_{it}\right)\right], \tag{4}$$

where the probabilities $\nu_k(z)$ depend on the predictors through a logit function:

$$\nu_k\left(z_{it}\right) = \frac{\exp(z_{it}'\alpha_k)}{1 + \exp(z_{it}'\alpha_k)}. \tag{5}$$

The parametrization in equations (4)-(5) is referred to as a continuation ratio model (Agresti, 2010). It also relates to the smoothly mixing regressions introduced by Geweke and Keane (2007) in the context of finite mixture models. The weights $\nu_k(z_{it})$ are interpreted as the probability of selecting component k given that the $k-1$ earlier

components were not selected. This can be seen by noting that ν_k can be written as $\nu_k(z_{it}) = \frac{\pi_k(z_{it})}{1-\sum_{l=1}^{k-1}\pi_l(z_{it})} = \frac{\Pr(G_i=k|z_{it})}{\Pr(G_i>k-1|z_{it})}$. As we show in the next subsection, the continuation-ratio approach simplifies Bayesian inference by transforming the mixture allocation problem into a series of *independent* binary logistic regressions, leveraging Pólya-gamma data augmentation for computational efficiency (Polson et al., 2013). Rigon and Durante (2021) illustrate how the logit stick-breaking process enables scalable posterior computations across multiple algorithms, such as Gibbs sampling, expectation-maximization, and variational Bayes, while preserving the flexibility of Bayesian nonparametric models; see also Schröder (2024) for an application of these ideas to inflation modeling.

Compared to existing approaches, such as the kernel stick-breaking process (Dunson and Park, 2008) and the probit stick-breaking process (Dunson and Rodríguez, 2011), the LSBP introduces predictor-dependent clustering while benefiting from efficient computational routines. Theoretical results in Ghosal et al. (1999) and Tokdar (2006) ensure that the Dirichlet process model in equation (2) can consistently approximate the true data density. Although the proposed model is grounded in nonparametric infinite mixtures, it can be practically estimated with a finite number of clusters, K. In fact, the truncation can involve as few as two clusters, offering remarkable flexibility.[2] This adaptability makes the model particularly suitable for capturing heterogeneity in panel data applications, such as those encountered in banking. The mixture representation allows to follow a large literature on identifying BBMs using mixtures of distributions; see Lucas et al. (2019) and Custodio João et al. (2023). At the same time, our model is nonparametric in nature, aligning with the recent state of the art in nonparametric BBM estimation (Custodio João et al., 2024) as well as nonparametric stochastic frontier analysis (Tsionas et al., 2023b; Kumbhakar et al., 2007).

2.2 Likelihood and priors

The likelihood of the proposed model is given by:

$$
\begin{aligned}
&p\big(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{\eta}^{+}, \boldsymbol{u}^{+}, \boldsymbol{\sigma}_v^2, \boldsymbol{z}, \boldsymbol{\alpha}\big) \\
&\quad = \prod_{i=1}^{n}\prod_{t=1}^{T}\sum_{k=1}^{\infty} \pi_k(z_{it})\, N\big(y_{it}; h(x_{it}, \beta_k) + a_i + \eta_i^{+} + u_{it,k}^{+}, \sigma_{v,k}^2\big),
\end{aligned}
$$

[2] For a discussion of cluster truncation, see the discussion in subsection 2.3. The estimation algorithm we propose is capable of handling even a single cluster, effectively fully nesting the model of Tsionas and Kumbhakar (2014) as a special case.

where the bold symbols ($\boldsymbol{y}$, $\boldsymbol{x}$, $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{\eta}^{+}$, $\boldsymbol{u}^{+}$, $\boldsymbol{\sigma}_v^2$, $\boldsymbol{z}$, $\boldsymbol{\alpha}$) represent collections, into vectors or matrices of conformable dimensions, of the respective parameters across all dimensions i, t, and k.

While this likelihood appears similar to a Gaussian mixture model, it is fundamentally different. As demonstrated by Tsionas and Kumbhakar (2014) for their simpler model (a special case of the above equation with $k = 1$), the convolution of the four latent components—comprising truncated normal and normal distributions—results in a multivariate skew-normal likelihood. Thus, our proposed specification extends the Tsionas and Kumbhakar (2014) framework into a flexible mixture of skewed distributions. It offers a structured perspective on production efficiency, distinguishing itself from the more generic skewed mixture specification proposed by Custodio João et al. (2023).

We next specify parametric prior distributions for all unknown parameters of interest. The prior for $\pi_k(z_{it})$ is given by the LSBP of equations (4)-(5).[3] Since we do not know a priori which banking variables z_{it} are most suitable for identifying different BBMs, we adopt a Bayesian variable selection approach. Specifically, the logit coefficients α_k follow a continuous spike-and-slab prior (George and McCulloch, 1993) of the form

$$\alpha_{j,k} \mid \gamma_{j,k} \sim (1-\gamma_{j,k})N\left(0,\tau_0^2\right) + \gamma_{j,k}N\left(0,\tau_1^2\right), \tag{6}$$

$$\gamma_{j,k} \sim \text{Bernoulli}(0.5), \quad \forall k \quad , j = 1, ..., q, \tag{7}$$

where $\gamma_{j,k}$ are binary variables indicating whether $\alpha_{j,k}$ is restricted ($\gamma_{j,k} = 0$) or not ($\gamma_{j,k} = 1$). When $\alpha_{j,k}$ is restricted (unrestricted), it has a zero-mean Gaussian prior with small (large) variance τ_0^2 (τ_1^2). Narisetty and He (2014) argue that fixing τ_0^2 and τ_1^2 to overly small and large values, respectively, can result in variable selection inconsistency. Following their recommendations, we set $\tau_0^2 = \frac{1}{nT/K}$ and $\tau_1^2 = \frac{\tau_0^2}{t_5\left(\sqrt{2.1\log(q+1)}\right)}$, where $t_5(\bullet)$ is the p.d.f. of the Student-t distribution with five degrees of freedom. The prior probability for all γ_j, k is set to 0.5, which is a standard uniform choice in Bayesian inference, implying that a priori, half of the covariates z_{it} are expected to be included in each cluster's logit specification. The coefficients of the stochastic frontier model have fairly noninformative prior distributions,

[3]Note that the logit stick-breaking process can be interpreted either as a multi-level model for $\pi_k(z_{it})$, forming part of the likelihood, or equivalently as a hierarchical prior in a Bayesian framework. These two terminologies are interchangeable, and here we choose to refer to it as the "LSBP prior" for the parameter $\pi_k(z_{it})$.

following closely Tsionas and Kumbhakar (2014). These take the following form

$$\beta_k \sim N\left(0, A^{-1}\right), \qquad k = 1, 2, ..., \tag{8}$$

$$\frac{Q_u}{\sigma_{u,k}} \sim \chi^2\left(N_u\right), \qquad \frac{Q_v}{\sigma_{v,k}} \sim \chi^2\left(N_v\right), \tag{9}$$

$$\frac{Q_a}{\sigma_a} \sim \chi^2\left(N_a\right), \qquad \frac{Q_\eta}{\sigma_\eta} \sim \chi^2\left(N_\eta\right), \tag{10}$$

where $A^{-1} = 10^{-4}$, $Q_\kappa = 10^{-4}$ and $N_\kappa = 1$, for $\kappa = a, \eta, v, u$, for all mixture components k.

2.3 Posterior Sampling

Given the hierarchical structure of the proposed model, we develop an efficient Markov chain Monte Carlo (MCMC) algorithm that sequentially samples from the conditional posterior distributions. Specifically, conditional on the parameters of the stochastic frontier kernel, we sample cluster assignments and the stick-breaking coefficients α_k. To circumvent the challenges of sampling from an infinite-dimensional parameter space, we truncate k to a maximum of K clusters. According to Theorem 1 in Rigon and Durante (2021), the estimated density under such truncation converges exponentially to the marginal density estimated under the representation using infinite clusters, that is, the representation in (2). This result implies that K does not need to be excessively large to achieve accurate nonparametric estimation of the underlying data density. Empirically, we set K to a "large enough value" based on the dataset at hand, and the algorithm is able to visit in a sequential manner only the required number of clusters, leaving redundant clusters empty. The simulation exercise of the next Section illustrates numerically this point.

For a given number of components K we can sample cluster assignments G_{it} randomly using probabilities defined in (4). Conditional on the cluster assignment indicators, latent variables ω_{itk} are introduced, simplifying the conditional posterior distributions. The conditional posterior distribution of the latent variable ω_{itk} is

$$\omega_{itk} \mid \bullet \sim \text{PG}(1, \psi_{itk}) \tag{11}$$

where $\mid \bullet$ denotes conditioning on data and other parameters, and $\text{PG}(1, \psi_{itk})$ denotes a Pólya-Gamma distribution[4] with parameters 1 and ψ_{itk}, the linear predictor associated with

[4]The Pólya-Gamma distribution is a continuous probability distribution with the probability density function given by:

$$f(w; b, c) = \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{g_k}{\left(k - \frac{1}{2}\right)^2 + \left(\frac{c}{2\pi}\right)^2}$$

$\nu_k(z_{it})$. The logistic regression coefficients α_k have a conditional posterior distribution:

$$\alpha_k \mid \bullet \sim N(m_k^\alpha, V_k^\alpha) \tag{12}$$

where the posterior mean m_k^α and variance V_k^α are updated based on the observed data, the continuous spike and slab prior, and the augmented variables ω_{itk}. Conditional on α_k the variable selection indicator γ_k in prior (6)-(7) is of the form

$$\gamma_k \mid \bullet \sim Bernoulli(\pi_\gamma), \tag{13}$$

with certain posterior hyperparameter π_γ.

Next, conditional on the cluster assignments and cluster probabilities, we derive the conditional posteriors of the stochastic frontier kernel, adapting the efficient sampling scheme from Tsionas and Kumbhakar (2014), see also Kumbhakar and Tsionas (2005). Given the model's high latency with four unobserved components, the authors propose a Gibbs sampler that avoids naive sequential updates of parameters. Instead, they sample the joint parameter $\delta_i = a_i + \eta_i^+$, followed by sampling η_i^+ conditional on the joint component $\xi_{it} = a_i + v_{it}$. Subsequently, a_i is computed as the difference between δ_i and η_i^+, reducing correlations in the resulting parameter samples. The conditional posterior of δ_i is non-conventional, and the authors use an efficient accept/reject algorithm with Gaussian proposals. Other parameters have conditional posteriors that have fairly standard forms: truncated normal for η_+ and $v_{it,k}^+$, and χ^2 for variance parameters which are functionally equivalent to more conventional inverse gamma priors. Under the assumption that the translog function $h(x_{it}, \beta_k)$ is linear in the parameters (as noted in footnote 1), the conditional posterior distribution of β_k follows a normal distribution, with hyperparameters derived in standard analytical forms.

The proposed Gibbs sampler proceeds by sequentially sampling each parameter one at a time, conditional on all other parameters remaining fixed. The exact steps are provided in Algorithm 1 below and detailed formulas are given in the Appendix.

where $g_k \sim \text{Gamma}(b, 1)$ are independent random variables following a Gamma distribution with shape parameter b and scale parameter 1.

Algorithm 1 *Logit Stick Breaking Process Stochastic Frontier Analysis panel data model*

1: Initialize all parameters.
2: Update cluster assignments G_{it} from multinomial.
3: Update latent variables ω_{itk} using Pólya-Gamma augmentation.
4: Update stick-breaking coefficients α_k from normal.
5: Update variable selection indicators γ_k from Bernoulli.
6: Update regression coefficients β_k from normal.
7: Update bank-specific heterogeneity δ_i using accept/reject using a normal proposal.
8: Update persistent inefficiency η_i from truncated normal with positive support.
9: Update transient inefficiency $u_{it,k}$ from truncated normal with positive support.
10: Update variances σ_a^2, σ_η^2, $\sigma_{u,k}^2$ and $\sigma_{v,k}^2$ from chi-square.

These sampling steps are fairly standard and easy to implement, resulting in a numerically stable MCMC algorithm that is also reasonably fast. Sampling from the Pólya-Gamma distribution can be efficiently achieved using Algorithm 6 in Windle (2013), which builds upon the work of Devroye (2009). Sampling from truncated normal conditional posteriors can also be implemented with high efficiency using the algorithm proposed by Botev (2017). The final algorithm cycles efficiently through the nine steps, enabling fast estimation of a complex model, even when thousands of panel data observations are available, as is the case with our bank data.

3 Simulation Study

We examine the ability of our model and Bayesian estimation algorithm to produce numerically accurate estimates and correctly identify the true number of clusters. The data generating process (DGP) for generating synthetic data is specified as follows:

$$x_{it} \sim N(0,1), \quad h(x_{it},\beta_k) = x_{it}\beta_k, \quad z_{it} \sim N(0,1), \tag{14}$$

$$\alpha = [1, 0.5, 0.8], \quad \beta = [1.8, 1.2, 0.6, 2.5], \tag{15}$$

$$\sigma_u = [0.1, 0.08, 0.05, 0.14], \quad \sigma_v = [0.1, 0.15, 0.22, 0.12], \tag{16}$$

$$\sigma_a = 0.2, \quad \sigma_\eta = 0.5. \tag{17}$$

We simulate models where the true number of clusters varies from $K = 2$ to $K = 4$. Depending on the specified number of clusters, the model is generated using the first K elements of all cluster-specific parameters, except for α, which requires $K - 1$ elements. For instance, for $K = 3$, we set $\alpha = [1, 0.5]$ and $\beta = [1.8, 1.2, 0.6]$, and apply similar

adjustments for all other parameters based on the true number of clusters in the DGP.

We further explore models with sample sizes $n = [100, 200, 500]$ and time periods $T = [5, 10, 15]$, considering all possible combinations of these values for n and T. We generate $n_{MC} = 100$ datasets and estimate each one using the proposed Gibbs sampler algorithm. A total of $110,000$ posterior samples are drawn, with the first $10,000$ discarded as burn-in. Every 100th sample is retained, resulting in $1,000$ effective posterior samples used for inference.

3.1 Numerical accuracy

Before presenting numerical accuracy statistics across all values of n, T, and K, we first examine parameter estimates in the case with $n = 200$, $T = 10$, and $K = 4$. The chosen values for n and T are intermediate among the considered options, while selecting the largest number of clusters provides a more informative assessment of numerical accuracy.

Across the 100 simulated datasets, the average number of observations in clusters one, two, three, and four are 394, 170, 204, and 1232, respectively, making cluster four the dominant cluster. When estimating clusters using our algorithm, these clusters are unlabeled, so we first reorder the estimated clusters to correspond to the generated ones. We achieve this by specifying a "cost matrix," which measures the Euclidean distance (quadratic loss) between the conditional mean of the estimated model and the conditional mean of the true model for each cluster. The clusters are then reassigned to their correct labels using the Hungarian algorithm (Kuhn, 1955).

Figure 1 shows the confusion matrix comparing true and estimated cluster assignments, averaged across the 100 Monte Carlo iterations. The rows represent the true cluster labels in the data generating process (DGP), while the columns show the estimated labels. Each cell indicates the count of observations assigned to a given cluster. Blank white spaces denote cases with zero observations. Therefore, the diagonal cells (highlighted in blue) represent the number of correct assignments for each cluster. For example, for cluster four, 1197 observations were correctly assigned out of a total of 1232. Off-diagonal cells (shown in pink) represent misclassifications. We observe that the total number of misclassified observations is 196, meaning that 1804 observations, or 90.2% of the data, were classified correctly. In general, for our specified DGP, misclassification rates range from 9% to 10.5% of the observations, which is a numerically good result. As a comparison, when using a naive clustering algorithm, in particular, the hierarchical clustering method of Ward (1955), we find misclassifications in over 70% of the observations. Most misclassifications in our algorithm occur when observations that belong to clusters one, two, and three are

mistakenly assigned to cluster four, as evidenced by the larger counts in the last column of Figure 1. This is primarily due to cluster four being generated with the largest number of observations.

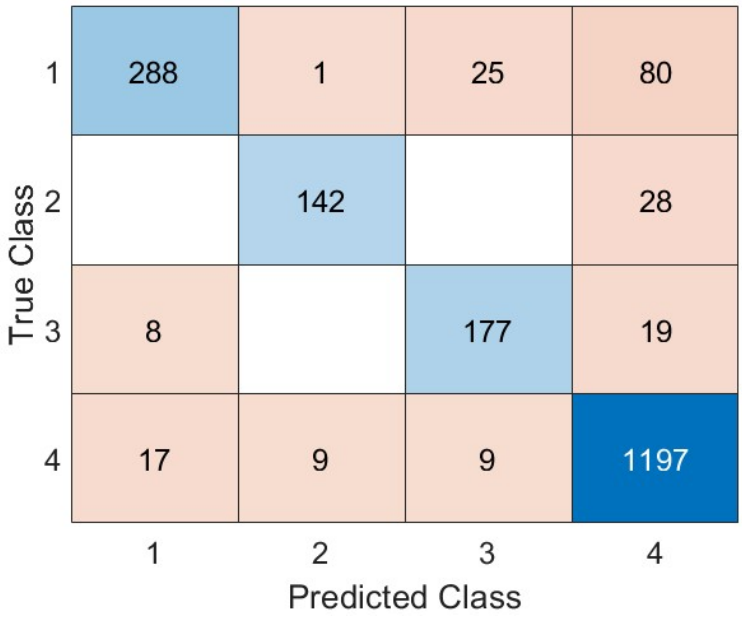

Figure 1: **Confusion matrix comparing true and estimated cluster assignments:** The rows represent the true cluster labels, while the columns correspond to the estimated cluster labels after realignment using the Hungarian algorithm. The values indicate the number of observations assigned to each cluster, with darker shades representing higher counts. Perfect alignment would result in a diagonal structure, where most observations are concentrated along the main diagonal. Results are for DGP with $n = 200$, $T = 10$, $K = 4$, averaged over 100 simulated datasets.

Table 1 presents the posterior means of the parameters along with their posterior standard errors (in parentheses) for each of the four clusters. These are averaged over all Monte Carlo iterations to ensure that the results reflect the general performance of the model across different simulated datasets. The estimated parameters for each cluster are generally very close to the true values specified in the DGP, especially for the β values and the σ_a and σ_η parameters. The small discrepancies between the true and estimated values can be attributed to sampling variability, especially given the relatively small sample size of n=200 and T=10 time periods in the simulation. However, the precision of the estimates, as indicated by the posterior standard errors, suggests that the model is working effectively to recover the true parameter values. The somewhat higher variability in the estimates for σ_u and σ_v across clusters could be explained by the inherent complexity of modeling heterogeneity across multiple clusters and the corresponding

uncertainty in estimating cluster-specific variances.

Overall, these results indicate that the Bayesian estimation method is robust and performs well in accurately estimating the parameters in the DGP, even when the true number of clusters is correctly identified as K=4. The estimates are both numerically accurate and precise, validating the effectiveness of the model and the Bayesian estimation algorithm for the given Monte Carlo simulation setup.

Table 1: Estimated posterior means and posterior standard errors (in parentheses) for each parameter for each cluster, averaged over all Monte Carlo iterations

Parameter	Cluster 1	Cluster 2	Cluster 3	Cluster 4
β	1.7990	1.2018	0.6121	2.4858
	(0.0038)	(0.0176)	(0.0134)	(0.0776)
σ_a		0.1770		
		(0.0359)		
σ_η		0.4889		
		(0.0264)		
σ_u	0.0921	0.0735	0.0530	0.1244
	(0.0425)	(0.0552)	(0.0344)	(0.0543)
σ_v	0.1067	0.0998	0.1530	0.1222
	(0.0099)	(0.0270)	(0.0387)	(0.0505)

We next move to compare numerical accuracy of the algorithm across different configurations of n, T and K. Table 2 shows the Mean Squared Error (MSE) for the estimation of key parameters in a model with varying sample sizes (n), time periods (T), and the number of clusters (K). The MSE values in the table represent the average error over these 100 datasets and it is calculated as follows:

$$\text{MSE}(\hat{\theta}) = \frac{1}{n_{MC}} \sum_{i=1}^{n_{MC}} (\hat{\theta}_i - \theta)^2$$

where $\hat{\theta}_i$ is the estimate of a given parameter θ from the i-th Monte Carlo dataset, and θ is the true value of the parameter. The MSE values in the table are averaged over the key parameters β, σ_a, σ_η, σ_u, and σ_v. Specifically, the MSE for each configuration of n, T, and K is computed as:

$$\text{Average MSE} = \frac{1}{5}\left(\text{MSE}(\hat{\beta}) + \text{MSE}(\hat{\sigma}_a) + \text{MSE}(\hat{\sigma}_\eta) + \text{MSE}(\hat{\sigma}_u) + \text{MSE}(\hat{\sigma}_v)\right)$$

This averaging procedure provides an overall measure of estimation accuracy across all the key parameters. The table displays the average MSE for different values of n, T, and K, with each entry corresponding to the computed MSE for a particular configuration of these factors. The results reflect how well the model parameters are estimated under varying conditions of sample size, time period length, and cluster number.

Table 2: Mean square error (MSE) for different values of n, T, and K.

n	T	K	Cluster 1	Cluster 2	Cluster 3	Cluster 4
100	5	2	0.0218	0.0182		
100	10	2	0.0044	0.011		
100	15	2	0.004	0.0107		
200	5	2	0.0411	0.0385		
200	10	2	0.0118	0.0176		
200	15	2	0.0052	0.0096		
500	5	2	0.0389	0.0374		
500	10	2	0.0127	0.021		
500	15	2	0.0072	0.0129		
100	5	3	0.0436	0.0548	0.0588	
100	10	3	0.0035	0.0345	0.0313	
100	15	3	0.0051	0.0108	0.0155	
200	5	3	0.037	0.0417	0.0229	
200	10	3	0.0048	0.0047	0.0053	
200	15	3	0.0021	0.0054	0.0067	
500	5	3	0.0453	0.0902	0.0614	
500	10	3	0.0071	0.0401	0.027	
500	15	3	0.0054	0.0101	0.0122	
100	5	4	0.0215	0.0304	0.0404	0.038
100	10	4	0.0061	0.0129	0.0181	0.0089
100	15	4	0.0065	0.0152	0.0166	0.0125
200	5	4	0.0609	0.1821	0.0197	0.0971
200	10	4	0.006	0.0052	0.0056	0.0152
200	15	4	0.002	0.0062	0.0074	0.0039
500	5	4	0.0464	0.0599	0.0373	0.0457
500	10	4	0.0094	0.0415	0.0353	0.0121
500	15	4	0.0041	0.0346	0.0228	0.0104

The results in Table 2 suggest that as the sample size increases, the MSE generally decreases across all values of K and T. Specifically, for smaller sample sizes ($n = 100$), the MSE tends to be higher, particularly for models with more clusters. For instance, when $K = 4$, the MSE is noticeably larger compared to when $K = 2$ or $K = 3$. As the number of time periods increases, the MSE decreases for most configurations, indicating that having more data over time helps improve estimation accuracy.

For the medium sample size ($n = 200$), the MSE decreases with increasing T, with the most noticeable improvement occurring when T moves from 5 to 15. For larger sample sizes ($n = 500$), the MSE continues to decrease with increasing T, though some configurations with more clusters (e.g., $K = 3$ and $K = 4$) still exhibit higher variability, especially at smaller values of T.

Overall, increasing the sample size and the number of time periods improves the accuracy of the parameter estimates, with more clusters leading to higher MSE at smaller sample sizes and fewer time periods.

3.2 Cluster allocation

We investigate our algorithm's ability to assign the correct number of clusters to the data. Since the cluster allocation in the LSBP is sequential, we can impose a large number of potential clusters and allow the algorithm to only visit the necessary ones. This approach—overfitting the number of clusters while restricting the algorithm to the relevant ones—was also used by Bauwens et al. (2015) in the context of Markov Switching and structural break models, where they overfit the number of breaks and let the algorithm select the appropriate ones.

Table 3 below shows the average estimated number of clusters across all Monte Carlo iterations for various sample sizes (n) and time periods (T). In this experiment, the true number of clusters in the data generation process (DGP) is $K = 2$, but the model was estimated with $K_{max} = 20$ clusters.[5] Although we overfit the number of clusters, the algorithm does not always visit all clusters, as it follows a sequential process and may not reach the maximum number of clusters.

The results indicate that as both the sample size n and the number of time periods T increase, the algorithm tends to overfit the number of clusters. However, given the maximum number of clusters imposed ($K = 20$), the performance of the algorithm is still satisfactory as it avoids computing multiple models with different numbers of clusters.

[5]For computational reasons, we do not repeat this exercise for true number of clusters being $K = 3$ and $K = 4$, as in the previous part of this section.

Table 3: Average Estimated Number of Clusters across Monte Carlo Iterations, when true number is $K = 2$ and maximum possible number of clusters is 20

n	T	# of Clusters
100	5	2.2
100	10	2.5
100	15	2.5
200	5	2.3
200	10	2.7
200	15	3
500	5	3.5
500	10	3.8
500	15	4.3

In practice, this method allows for the estimation of a model with many clusters, without the need to calculate computationally expensive marginal likelihoods for models with different numbers of clusters. This strategy is particularly beneficial when one does not know the exact number of clusters beforehand. The results suggest that this approach is valid as long as the number of clusters is not excessively large. For the empirical analysis, we impose a conservative number of clusters ($K = 4$) as our benchmark model, based on our experience with the dataset. Additionally, in the online supplement, we will estimate models with $K = 10$ clusters and compare the results.

4 Empirical Results

4.1 Data and technology specification

Our sample consists of $n = 493$ European commercial, cooperative, and savings banks observed annually over the period 2008–2022, implying $T = 15$. The resulting balanced panel consists of 7,395 bank-year observations. BankFocus is the source of balance sheet data of annual frequency.

Identification of BBMs is via the variables we denote as z_{it} in our model specification. We follow Badunenko et al. (2021), Curi et al. (2015), and Elsas et al. (2010) and opt for three key dimensions: (i) asset composition (or lending and investment in Badunenko et al., 2021), (ii) sources of funding and their quality, and (iii) sources of income. However, rather than using some aggregate measures of those dimensions, we employ all their available determinants. To this end, we opt for loans, loans to banks, derivatives, securities, non-earning assets, and fixed assets that determines the asset composition dimension of BBM. The funding quality dimension is based on deposits, bank deposits, short-term funding,

derivatives, long-term funding, reserves, and equity. Lastly, the income mix dimension focuses on interest income and non-interest income. In total, we assemble a comprehensive thirteen variables that would provide underlying information to identify BBM.

So far, we have left specification of the translog function $h(x_{it}, \beta)$ unspecified. We do that here, so we also specify the input and other variables denoted as x_{it}. The technology of bank i $(i = 1, \ldots, n)$ observed over time period t $(t = 1, \ldots, T)$ is specified by the following translog cost function

$$\begin{aligned} h(x_{it}, \beta) = \beta_0 &+ \sum_{j=1}^{2} \beta_j^Y \log Y_{j,it} + \sum_{m=1}^{3} \beta_m^W \log W_{m,it} \\ &+ \frac{1}{2} \sum_{j=1}^{2} \sum_{l=1}^{2} \beta_{jl}^{YY} \log Y_{j,it} \log Y_{l,it} + \frac{1}{2} \sum_{m=1}^{3} \sum_{n=1}^{3} \beta_{mn}^{WW} \log W_{m,it} \log W_{n,it} \\ &+ \sum_{j=1}^{2} \sum_{m=1}^{3} \beta_{jm}^{YW} \log Y_{j,it} \log W_{m,it} \\ &+ \sum_{j=1}^{2} \beta_j^{\delta} \log Y_{j,it} \times t + \sum_{m=1}^{3} \beta_m^{\theta} \log W_{m,it} \times t \\ &+ \beta_{\tau} t + \beta_{\tau^2} t^2. \end{aligned} \tag{18}$$

The outputs Y_1 and Y_2 represent loans and other earning assets, respectively, while the inputs W_1, W_2, and W_3 correspond to the prices of labor, physical capital, and financial capital. The first-order terms β_j^Y and β_m^W capture the direct effects of outputs and input prices on costs. The quadratic terms β_{jl}^{YY} and β_{mn}^{WW} account for nonlinear effects and possible economies of scale, while the interaction terms β_{jm}^{YW} capture how input prices and outputs jointly affect costs. The terms β_j^{δ} and β_m^{θ} allow for time-varying effects of outputs and input prices, while the linear and quadratic time trend terms β_{τ} and β_{τ^2} capture technological changes in cost efficiency. This translog cost function specification allows for a flexible representation of banking cost structures, accommodating nonlinear relationships, interactions, and technological progress over time.

4.2 Cluster selection

In the previous section, we established that our model accommodates a large maximum number of clusters. However, the algorithm only sequentially visits the most probable clusters, leaving the remaining ones empty. While this approach is computationally efficient, it comes at the cost of increased parameter uncertainty due to the estimation of an unnecessarily large number of clusters.Based on prior experience with the data, we

expect only a small number of clusters. For instance, Custodio João et al. (2023) identify six clusters using quarterly data and a different clustering methodology, but this number may be too high for our annual dataset.

To assess the appropriate number of clusters, we estimate models with three, four, and ten clusters. Table 4 presents the average number of banks per cluster over the $T = 15$ years in our sample, under the cases $K = 3$, $K = 4$ and $K = 10$. A clear pattern emerges, as all models indicate a dominant cluster with over 300 banks. In the ten-cluster model, three clusters contain zero or one bank, and five clusters have fewer than nine banks. This suggests that, depending on how clusters are defined in terms of membership size, the ten-cluster model effectively reduces to five or fewer meaningful clusters.

Both the three- and four-cluster models result in clusters with a sufficient number of banks. However, in the four-cluster model ($K = 4$), Cluster 2 is relatively small, averaging only 12 banks. Given these findings, we proceed with presenting results for the four-cluster model in the main text while providing results for the three-cluster model in the online supplement. Importantly, key estimates of permanent and transient inefficiencies remain largely consistent across the three- and four-cluster models.

Table 4: Cluster membership for models with $K = 3$, 4, and 10

	$K = 3$	$K = 4$	$K = 10$
Cluster 1	80	64	3
Cluster 2	339	12	9
Cluster 3	74	316	43
Cluster 4		101	1
Cluster 5			302
Cluster 6			1
Cluster 7			22
Cluster 8			81
Cluster 9			0
Cluster 10			31
Total	493	493	493

Notes: Entries are the average number of banks per cluster over $T = 15$ years. Clusters are unlabelled, meaning that, for example, Cluster 2 in the $K = 3$ model does not necessarily correspond to Cluster 2 in the $K = 4$ or $K = 10$ models.

Adopting the four cluster model as the benchmark, Figure 2 shows the number of banks in each of the four clusters at each point in time. The figure shows the number of banks assigned to each of the four clusters over time, based on the four-cluster model. The x-axis

represents the years from 2008 to 2022, while the y-axis indicates the number of banks in each cluster. Cluster 1 experiences a decline in membership from over 100 banks in 2008 to lower levels in the mid-2010s, followed by a rebound in 2018 and 2022, suggesting banks may transition into this cluster under specific economic conditions. Cluster 2 is marked by a sharp spike in 2011, with membership exceeding 35 banks, followed by a period of lower and fluctuating levels until another increase in 2022, potentially reflecting changes in banking characteristics during financial stress or regulatory shifts. Cluster 3 is the largest, consistently containing over 300 banks for most of the period, growing from 2008 to 2013–2014 before stabilizing and experiencing a modest decline in 2022. Cluster 4 remains relatively stable, with membership ranging between 100 and 120 banks, though it dips in the mid-2010s before rising sharply in 2021 and 2022, suggesting that banks may transition into this cluster in response to economic or regulatory changes.

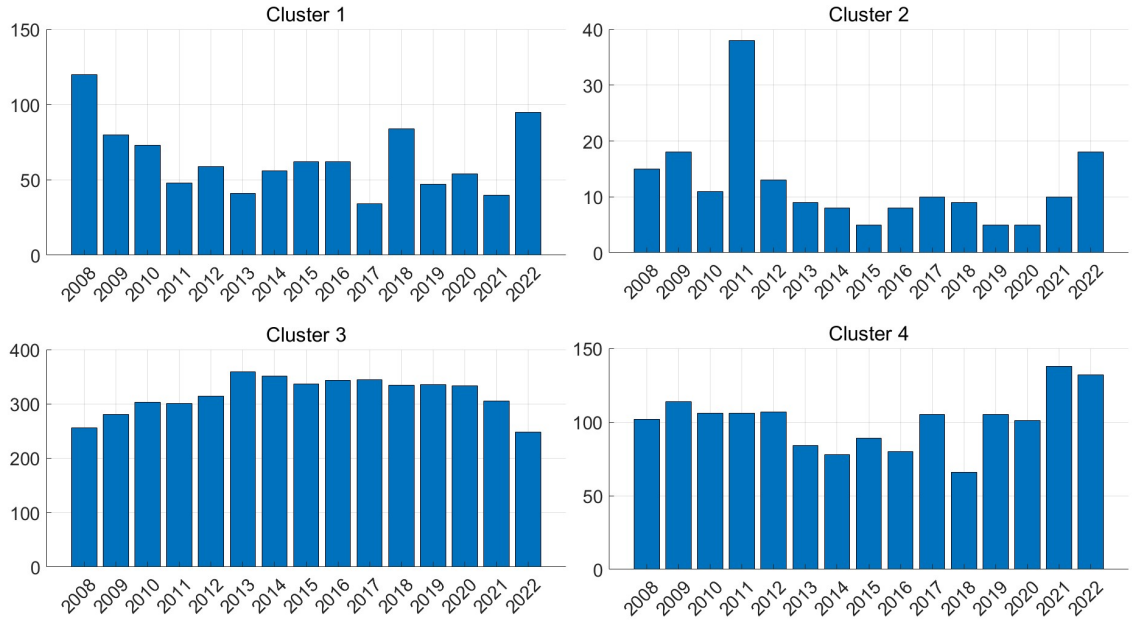

Figure 2: **Cluster transition and popularity**
This figure shows the number of banks allocated to each of the four clusters at each point in time. Bank i is assigned to cluster k at time t for which the cluster probability is maximal.

4.3 Efficiency measures

We report estimates of efficiency measures in Figure 3. Efficiencies are defined as the transformations $\exp\left(-u_{it}^{+}\right)$ and $\exp\left(-\eta_{it}^{+}\right)$ of transient and persistent inefficiency components, respectively. Transient efficiency takes higher values than the persistent efficiency, as expected, since banks tend to become more efficient in the long run. The

distribution of transient efficiency has a more complex shape as it is a mixture of four components. This is a desirable feature of our analysis, as it provides crucial insights for identifying the underlying BBM at the bank level. On average, banks exhibit a high level of efficiency, with a mean persistent cost efficiency of approximately 0.9189. The lowest estimate, around 0.876, suggests minimal long-term cost efficiency among European banks, reflecting a uniform and well-functioning banking environment in European countries. These results are consistent with Badunenko et al. (2021), Tsionas and Kumbhakar (2014), and Tsionas et al. (2023b). Figure 3 also illustrates the distribution of transient (short-run) cost efficiency, which is lower than persistent efficiency, with a mean of approximately 0.8778. However, the noticeable left tail suggests that some banks exhibit significantly lower transient efficiency, highlighting opportunities for substantial short-term efficiency gains if they switch BBM. This transient inefficiency warrants further investigation; panel (c) reports its values over time. In the aftermath of the global financial crisis, there is a clear peak in transient inefficiency (low efficiency). Since then, efficiency has exhibited a rollercoaster-like pattern, emphasizing the need for continued monitoring.

As we show in Figure S.2. of the online supplement, these estimates of permanent and transitory efficiencies are fairly similar to the ones obtained from the model with three clusters. However, as shown in Figure S.1. there are noticeable differences with the estimates of transient efficiency in the model with $K = 1$ clusters, which is the exact model in Tsionas and Kumbhakar (2014). This is to be expected, as in this model transient efficiency is not a mixture of cluster-specific distributions.

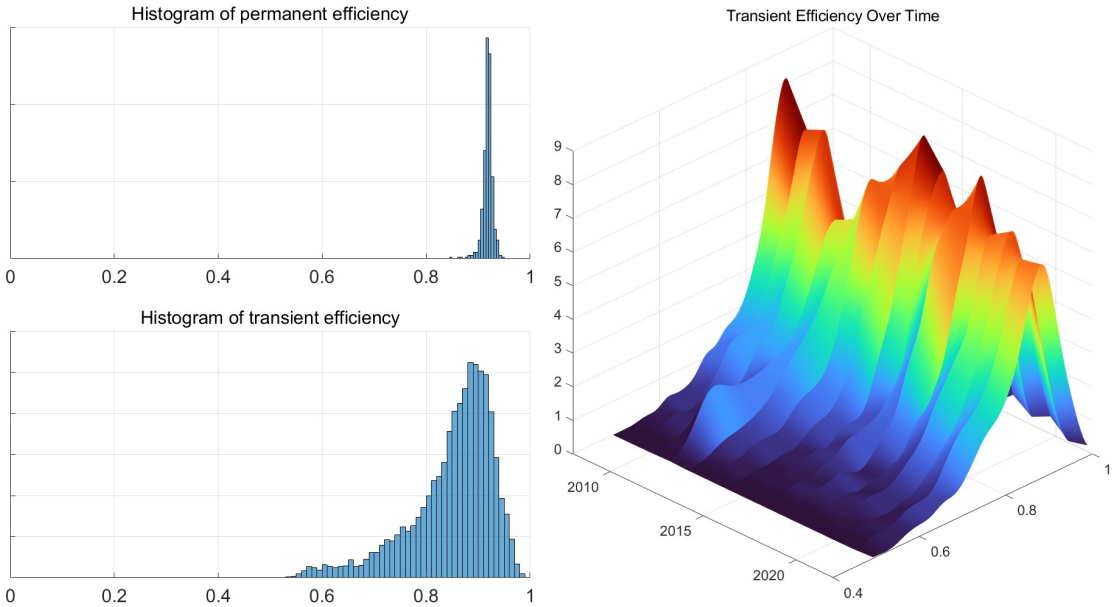

Figure 3: **Efficiency Decomposition in the LSBP Stochastic Frontier Model** This figure presents the decomposition of efficiency into permanent and transitory components. The top left panel displays the histogram of permanent efficiency, illustrating the distribution of long-term efficiency levels across banks. The bottom panel shows the histogram of transient efficiency, capturing the short-term fluctuations in efficiency across all banks and time periods. The right panel, spanning the right half of the figure, presents a 3D histogram of transient efficiencies, visualizing how these vary across banks and years.

To further understand how aggregate transient efficiency is a mixture of transient efficiency in the four clusters, Figure 4 displays the evolution of transitory efficiency estimates for each of the four clusters over time. The x-axis represents time from 2008 onwards, while the y-axis captures the efficiency scale, with the third dimension illustrating the distribution of efficiency levels across banks in each cluster. Cluster 1 exhibits a significant drop in efficiency during the period surrounding the global financial crisis, with fluctuations in subsequent years. Cluster 2 experiences the most pronounced decline in efficiency around the 2011 eurozone debt crisis, indicating that banks in this group were particularly affected by the sovereign debt turmoil. Cluster 3, which contains the largest number of banks, shows a more stable evolution of efficiency, though some periods of deterioration are visible. Cluster 4 demonstrates variability over time, with efficiency levels experiencing noticeable peaks and troughs, potentially reflecting shifts in market conditions or regulatory environments. The decomposition into clusters reveals that overall transitory efficiency is shaped by diverse patterns across bank groups, rather than a uniform trend.

Figure S.1 in the online supplement shows similar estimates of permanent and

transient efficiencies in the model with $K = 1$ clusters, which is the exact model in Tsionas and Kumbhakar (2014). As expected, there are noticeable differences in the estimates of transient efficiency in the $K = 1$ model, where this model

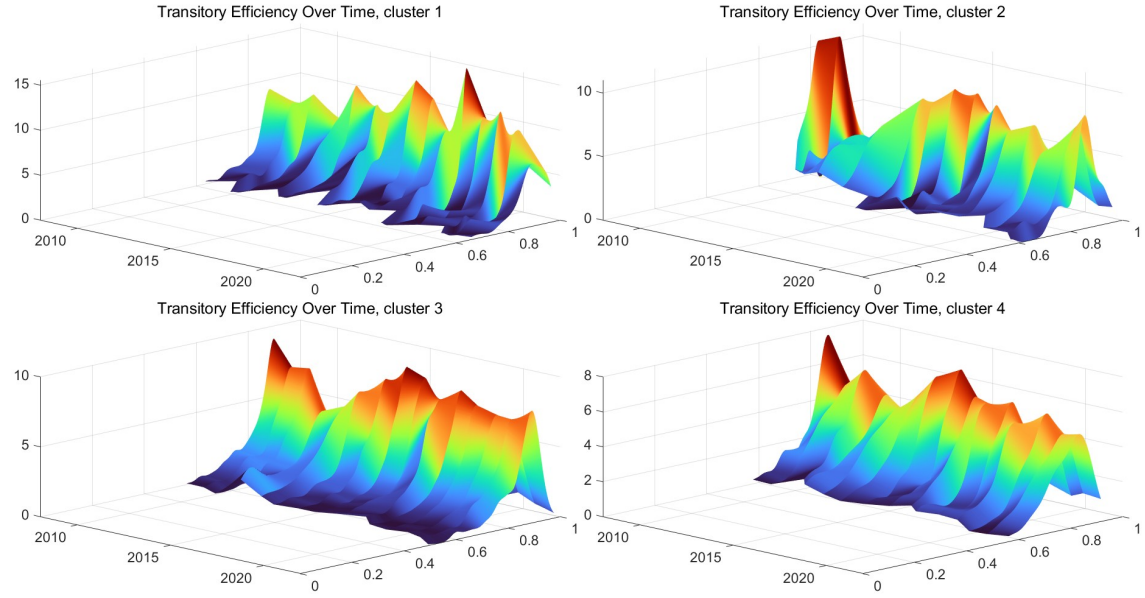

Figure 4: **Transient efficiency per period and per cluster**
The figure shows 3D histogram of transient efficiencies for each of the four clusters over the $T = 15$ years in the sample.

4.4 Cluster allocation and transitions

We use all available determinants of the three main bank dimensions (asset composition, funding quality, and income mix) to model the cluster allocation probabilities. These determinants, collected in the vector z_{it}, impact upon the likelihood of belonging to a specific cluster. The inclusion of a spike-and-slab prior on α allows the model to select only the most relevant variables for clustering, reducing dimensionality and improving interpretability. We do so by calculating posterior inclusion probabilities (PIPs) which are calculated as the posterior mean of the variable indicators $\gamma_{j,k}$, for $j = 1, ..., q$ and $k = 1, ..., K$, that is, $PIP_{j,k} = \mathbb{E}\left(p\left(\gamma_{j,k}|\bullet\right)\right)$. Following Barbieri and Berger (2004) we label as important those variables that have $PIP > 0.5$, also referred to as the median probability model. Table 5 reports z_{it} variables selected in the Logit Stick-Breaking Process for each cluster, revealing the underlying BBM. There are four identified BBMs, which we can label as follows:

1. Cluster 1 consists mostly of banks with an emphasis on the asset dimension and

non-interest income.

2. Cluster 2 consists mostly of banks with an emphasis on the funding dimension and interest income.

3. Cluster 3 consists mostly of banks with an emphasis on both the asset and funding dimensions but without the income mix dimension.

4. Cluster 4 consists mostly of banks with an emphasis on the asset dimension and also derivatives from the funding quality dimension.

The four clusters identified in this study exhibit distinct characteristics aligned with their determinants. Cluster 1 is predominantly defined by banks that focus on asset composition, with their main income derived from non-interest sources. The reliance on non-interest income provides a diversified source of revenue beyond traditional interest-bearing assets like loans and securities. This strategy aligns with the need for banks to reduce their reliance on volatile interest rate environments and develop alternative sources of revenue. The inclusion of loans, securities, and derivatives in the asset composition reflects a risk-management strategy, enabling these banks to hedge against adverse market conditions, while the focus on non-interest income may indicate a shift towards fee-based services such as investment banking or wealth management.

In contrast, Cluster 2 places a greater emphasis on the funding dimension, with banks primarily generating interest income. Banks operating under this model prioritize stability and security in their funding sources, as indicated by the heavy reliance on deposits, short-term funding, and equity. This model is indicative of a more conservative banking approach where interest income from loans and other interest-bearing assets remains a primary revenue source. Thus, Cluster 2 reflects banks that focus on sustainable and low-risk funding sources while still capturing the profitability from interest-bearing activities. The preference for equity and deposits as key funding sources suggests a low-risk appetite, which is often associated with financial stability. As shown in Figure 2, there was a spike in bank transitions to Cluster 2 in 2011, following the global financial crisis. This shift reflected a move toward traditional interest-bearing activities over securitization. However, banks in Cluster 2 experienced a decline in efficiency in 2012, indicating that their strategic decision to switch BBM came at the cost of reduced efficiency. This triggered a dynamic process of further re-switching between BBMs.

Cluster 3 combines aspects of Cluster 1 and Cluster 2, focusing on both asset and funding dimensions but excluding the income mix, thereby suggesting a mixture type of BBM that relies on balance sheet strength and funding structures, like loans and deposits.

This cluster is the dominant one, as depicted by Figure 3, panel C. This mixture bank model suggests an underlying adaptive bank managerial strategy, where banks may be adjusting their business models in response to evolving market conditions and shocks. This could indicate that banks in the aftermath of a shock, like the global financial crisis, were in a phase of rebalancing and restructuring, focusing on the structural elements of their operations, such as asset and funding composition, without diversifying their income sources. This model may reflect a more cautious, traditional banking approach, therefore, as banks adjusted primarily their risk profiles and operational strategies through their assets and funding, while alternative income streams were not a priority.

Lastly, Cluster 4 is characterized by a focus on asset composition, with derivatives from the funding quality dimension playing a notable role. Derivatives, in this case, are versatile instruments, serving both as an asset management tool and a mechanism for managing the bank's funding and risk profile. The inclusion of derivatives in both asset and funding dimensions indicates that these banks are using them not only to manage their investments but also to stabilize their funding sources, particularly in volatile market conditions. This may reflect banks' proactive strategy in managing liquidity, interest rate risks, and market exposure.

These identified clusters suggest distinct managerial approaches among banks. Although it is worth noting that the engagement with derivatives appears similar across all clusters, indicating a more uniform approach in terms of funding and risk management. That said, banks within a specific cluster can still display notable heterogeneity in certain characteristics. Banks in Cluster 2 exhibit an income mix strategy more geared towards interest revenues, representing deposit-oriented banks. Banks in Cluster 1 specialize in non-interest income but also in asset composition such as loans, indicating an efficient loan-making process. Banks in Cluster 3 follow a more traditional type of BBM that relies on the loan-making process, which is inherently labor-intensive, requiring skilled personnel to assess creditworthiness, manage client relationships, and tailor financial solutions. This is reflected in the high labor costs associated with these banks. Banks in Cluster 4 consist of securities- and derivative-oriented banks, with their business model likely to rely heavily on computerized securities trading, including high-frequency and algorithmic trading. This model places a significant emphasis on physical capital, such as trading infrastructure, servers, and advanced IT systems, which is evident in the high marginal effect of physical capital on total costs. This suggests that technological infrastructure is a primary driver of costs for these banks. Additionally, these institutions show minimal sensitivity to labor costs, pointing to a lower reliance on human-intensive processes compared to

loan-oriented banks. However, there are fewer EU banks in Cluster 4 compared to the other clusters, suggesting that the EU banking sector is dominated by more traditional banking models that rely on loans and deposits.

Table 5: Z variables selected in the Logit Stick-Breaking Process for each cluster

Predictor	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Intercept	✓	✓	✓	✓
Panel A. Asset composition				
Loans	✓	✓	✓	✓
Loans to banks	✓	✓	✓	
Derivatives	✓	✓	✓	
Securities	✓		✓	✓
Non-earning assets	✓		✓	
Fixed assets		✓		
Panel B. Funding quality				
Deposits		✓	✓	
Bank deposits		✓	✓	
Short-term funding		✓	✓	
Derivatives	✓	✓	✓	✓
Long-term funding	✓			
Reserves	✓			
Equity	✓	✓	✓	
Panel C. Income mix				
Non-interest income	✓			
Interest income		✓		

Figure 5 presents key insights into the four identified BBM during the sample period, segmented across four panels. Panel A depicts the frequency of business model switches per year, Panel B illustrates the number of business model switches made by each bank, Panel C shows the number of transitions into each cluster across all time periods and banks, and Panel D provides the total number of switches across all banks. Panel A shows that transitions picked in the aftermath of the global financial crisis in 2011 at a year that transitory bank efficiency dropped at it lowest point. Based on our findings it appears that the transition between BBMs was strategically opted with some lags to support recovery in efficiency that dropped due to the crisis. Banks, therefore, switch BBM with lags in response to a major shock. Interestingly transitions between BBMs picked again in 2021

due to the COVID-19 pandemic at a year the transitory (short-run) efficiency dropped. Panel B of Figure 5 highlights that a low portion of banks of around 28.0 percent do not switch between two different business models over the sample period, suggesting that the vast majority of banks will switch BBM at least once during the sample period as response to changing market conditions. The average number of transitions between BBMs per bank i, where $i = 1, ..., 493$, is around three times and ranges from zero to eight (but this is relevant just for one bank). It is worth noting that for around 100 banks out of 493 transition takes place more than four times within the sample period, implying that switching BBMs might not be as rigid as previous literature suggested (Custodio João et al., 2023). One of the main advantages of our model is that allows for multiple transitions for a given bank and this provides flexibility compared with previous models. Panel D of Figure 5 demonstrates this flexibility using a histogram.

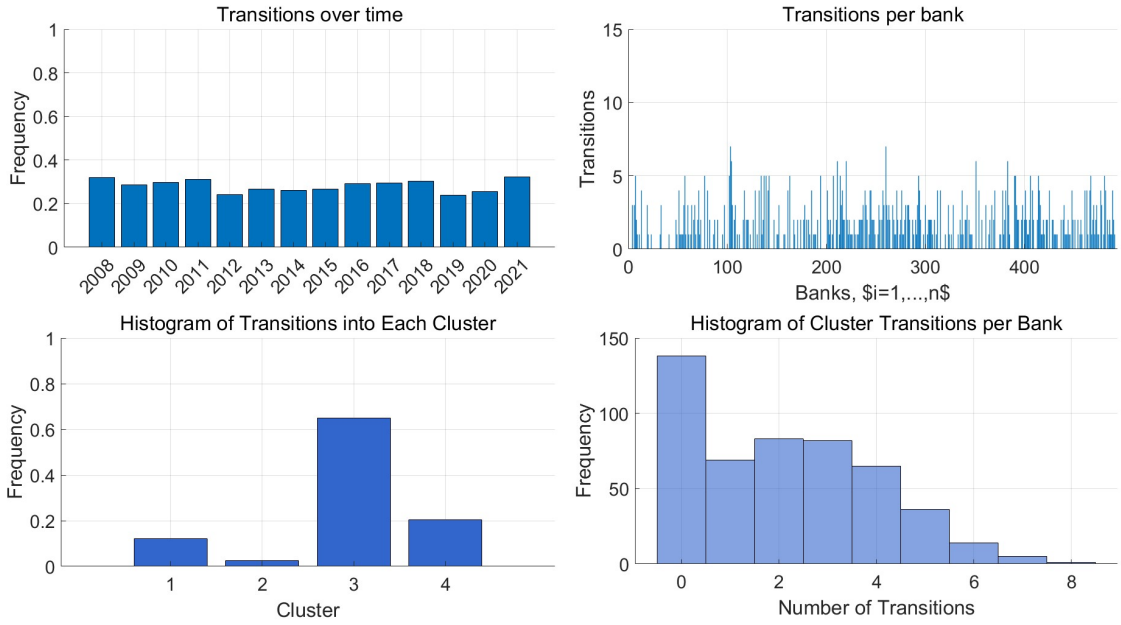

Figure 5: **Analysis of Cluster Transitions for Banks Over Time**
This figure presents a detailed analysis of the cluster transitions of banks over time, divided into four panels: the top left panel shows the frequency of transitions for each bank per year, illustrating how often banks change clusters over time; the top right panel displays the distribution of transitions per bank, indicating how many times each bank transitions clusters throughout the observation period; the bottom left panel presents the histogram of transitions to each cluster, showing how frequently banks move into each of the possible clusters across all time periods; and the bottom right panel provides the histogram of cluster transitions per bank, highlighting the frequency with which banks transition from one cluster to another over time.

4.5 Robustness of our results

We perform a series of robustness tests. First, we control for time-varying country-level characteristics in the efficiency estimation to capture differences in macroeconomic conditions, market structure and financial development. Second, to address potential endogeneity concerns, we perform a robustness check by lagging the BBM variables by one period. Third, we adjust the model to use both a smaller and a larger number of clusters to assess how sensitive the clustering results are to the choice of (maximum) cluster count. Finally, we consider a restricted version of the translog cost function where certain coefficients are constrained to follow theoretically motivated sign restrictions. While our cluster model is heavily parametrized resulting in quantitative discrepancies between different specifications (especially when we consider theory-based sign restrictions), overall our results remain qualitatively consistent across these tests. For brevity, we provide these additional results in the online supplement.

5 Conclusions

We propose a novel Bayesian approach to identifying BBMs, departing from conventional clustering methods. Our four-error component translog cost function model is based on the underlying economic fundamentals that imply banks are optimising in terms of cost minimization. This modelling approach is dynamic, allowing for transitions between different BBMs over time. Over a long period of time that major shocks took place our modelling approach is of particular importance. In our empirical investigation of EU banks from 2008 to 2022, we uncover BBM clusters and their dynamics, highlighting the heterogeneous nature of banks operating across identified clusters. We also find compelling evidence of banks switching between BBMs following a major shock like the global financial crisis and the COVID-19. Our analysis shows that switches between BBMs are linked to significant changes in transitory bank efficiency, as bank management seeks to improve efficiency.

A Technical Appendix

A.1 Model Representation

The full model with response variable y_{it}, for firm i at time t, can be represented as follows

$$y_{it} \mid G_{it} = k, \beta_k, a_i, \eta_i^+, u_{it,k}^+, \sigma_{v,k}^2 \sim N\left(x_{it}'\beta_k + a_i + \eta_i^+ + u_{it,k}, \sigma_{v,k}^2\right), \tag{A.1}$$

where $a_i \sim N(0, \sigma_a^2)$ is the firm-specific heterogeneity for cluster k, $\eta_i^+ \sim N_+(0, \sigma_\eta^2)$ is the persistent (time-invariant) inefficiency, $u_{it,k} \sim N_+(0, \sigma_{u,k}^2)$ is the transient (time-varying) inefficiency for cluster k and $v_{it,k} \sim N(0, \sigma_{v,k}^2)$ is the noise for cluster k. The model is completed by specifying priors for the unknown parameters. First, the mixing weights $\pi_k(z_{it}) = prob(G_{it} = k|z_{it})$ depend on the exogenous variables z_{it} and are defined as

$$\pi_k = \nu_k(z_{it}) \prod_{l=1}^{k-1} [1 - \nu_l(z_{it})], \qquad \nu_k(z_{it}) = \text{logit}^{-1}(z_{it}'\alpha_k),$$

The regression coefficients α_k have a continuous spike and slab prior of the form

$$\alpha_{j,k} \mid \gamma_{j,k} \sim (1 - \gamma_{j,k})N(0, \tau_0^2) + \gamma_{j,k}N(0, \tau_1^2) \tag{A.2}$$

$$\gamma_{j,k} \sim Bernoulli(0.5), \tag{A.3}$$

for $j = 1, ..., q$ and $k = 1, ..., K$. Finally the remaining coefficients in (A.1) have noninformative priors, following Tsionas and Kumbhakar (2014)

$$\beta_k \sim N(0, A^{-1}), \qquad \frac{Q_{\kappa_1}}{\sigma_{\kappa_1,k}^2} \sim \chi^2(N_{\kappa_1}), \qquad \frac{Q_{\kappa_2}}{\sigma_{\kappa_2}^2} \sim \chi^2(N_{\kappa_2}), \tag{A.4}$$

where $A = 10^{-4}I$, $N_{\kappa_1} = N_{\kappa_2} = 1$ and $Q_{\kappa_1} = Q_{\kappa_2} = 10^{-4}$, for $\kappa_1 = u, v$ and $\kappa_2 = a, \eta$.

A.2 Conditional posteriors and sampling

The Gibbs sampler iteratively updates all parameters from the following conditional posterior distributions:

Step 1: Update Cluster Assignments G_{it}

The probability that $G_{it} = k$ is:

$$P(G_{it} = k \mid -) \propto \pi_k(z_{it}) \cdot N\left(y_{it} \mid x_{it}'\beta_k + a_i + \eta_i + u_{it,k}, \sigma_{v,k}^2\right),$$

where $\pi_k(z_{it})$ is the weight of component k. We can sample G_{it} from the multinomial distribution using the probabilities defined above.
Construct data matrices/vector x_k, z_k, y_k, and u_k by selecting only those observations (i,t) from x, z, y, and u for which $G_{it} = k$. these will be used in steps 2 and 3 below.

Step 2: Update Stick-Breaking Coefficients α_k
For each $k = 1, \ldots, K$:

1. Generate Pólya-Gamma latent variables:

$$\omega_{it,k} \sim PG(1, z'_{it}\alpha_k), \quad \forall i,t. \tag{A.5}$$

2. Update α_k from the Gaussian posterior:

$$\alpha_{j,k} \mid - \sim N\left(\Sigma_{\alpha_k}\left(z'_k(\zeta_k - 0.5) + \Sigma_\alpha^{-1}\mu_\alpha\right), \Sigma_{\alpha_k}\right), \tag{A.6}$$

   where $\Sigma_{\alpha_k} = (z'_k\Omega_k z_k + \Sigma_\alpha^{-1})^{-1}$, $\Omega_k = \text{diag}(\omega_{it,k})$, $\zeta_k = \{\zeta_{it,k}\}_{i,t}$, $\zeta_{it,k} = 1$ if $G_{it} = k$ and $\zeta_{it,k} = 0$ if $G_{it} > k$ and $\Sigma_\alpha = diag\left((1-\gamma_{j,k})\tau_0^2 + \gamma_{j,k}\tau_1^2\right)$

3. Update $\gamma_{j,k}$ from the Bernoulli posterior

$$\gamma_{j,k} \mid - \sim Bernoulli\left(\frac{\widetilde{\pi}_1}{\widetilde{\pi}_1 + \widetilde{\pi}_0}\right) \tag{A.7}$$

   where $\widetilde{\pi}_0 = (1-\pi_0) \times N(\alpha_{j,k}|0,\tau_0^2)$ and $\widetilde{\pi}_1 = (1-\pi_1) \times N(\alpha_{j,k}|0,\tau_1^2)$.

Step 3: Update Regression Coefficients β_k
For each cluster k, conditionally on $G_{it} = k$:

$$\beta_k \mid - \sim N\left(\Sigma_{\beta_k}\left(\frac{x'_k y_k}{\sigma_{v,k}^2}\right), \Sigma_{\beta_k}\right),$$

where $\Sigma_{\beta_k} = \left(\frac{x'_k x_k}{\sigma_{v,k}^2} + A\right)^{-1}$, $\tilde{y}_k = (y_k - \delta \otimes \iota_T - u_k)$.

Step 4: Update Bank-Specific Heterogeneity $\delta_i = a_i + \eta_i^+$
For each firm i sample from the posterior

$$p(\delta_i \mid -) \propto \exp\left(-\frac{(R_i - d_i\mathbf{1}_T)'(R_i - d_i\mathbf{1}_T)}{2\Sigma_v} - \frac{\delta_i^2}{\sigma_\delta^2}\right)\Phi\left(\frac{\lambda\delta_i}{\sigma_\delta}\right), \tag{A.8}$$

where $R_i = \sum_{t=1}^T y_{it} - x'_{it}\beta_{(G_{it}=k)} - u_{it}^+$ and $\sigma_\delta^2 = \sigma_a^2 + \sigma_\eta^2$ and Σ_v is an $nT \times 1$ vector with $\Sigma_{v,it} = \sum_{k=1}^K \mathbb{I}(G_{it} = k)\sigma_{v,k}^2$ for all i,t. As this is not a conditional posterior

we can directly sample from, we follow Tsionas and Kumbhakar (2014) and use a Metropolis-Hasting step, see Section 3.3.1 of their paper.

Step 5: Update Persistent Inefficiency η_i^+

For each firm i, sample η_i^+ from a truncated normal:

$$\eta_i^+ \sim N^+(\mu_i, \phi^2) \tag{A.9}$$

where $\phi^2 = \left(1/\sigma_\eta^2 + \mathbf{1}'_{nT}\Sigma^{-1}\mathbf{1}_{nT}\right)^{-1}$ and $\mu_i = \phi^2 \mathbf{1}_T^\top \Sigma^{-1} D_i$. Here $\Sigma = I_{nT} \otimes \sigma_a^2 (\mathbf{1}'_{nT}\mathbf{1}_{nT}) + \Sigma_v$ is an $nT \times nT$ covariance matrix, and $D_i = \sum_{t=1}^T y_{it} - x_{it}\beta_{(G_{it}=k)} - u_{it}^+$.

Step 6: Update Transient Inefficiency u_{it}^+

For each observation it, sample u_{it}^+ from a truncated normal:

$$u_{it}^+ \mid - \sim N^+\left(\tilde{u}_{it}, \frac{\Sigma_{v,it}\Sigma_{u,it}}{\Sigma_{v,it} + \Sigma_{u,it}}\right) \tag{A.10}$$

where $\tilde{u}_{it} = \frac{\Sigma_{u,it}\left(y_{it} - x_{it}\beta_{(G_{it}=k)} - \delta_i\right)}{\Sigma_{v,it}+\Sigma_{u,it}}$ where similar to Σ_v defined previously, Σ_u is also an $nT \times 1$ vector with $\Sigma_{u,it} = \sum_{k=1}^K \mathbb{I}(G_{it} = k)\, \sigma_{u,k}^2$.

Step 7: Update Variance Components

For each cluster k, update the variances as follows:

$$\begin{aligned}
\sigma_{v,k}^2 \mid - &\sim \frac{\sum_{i,t}\left(y_{it} - x'_{it}\beta_{(G_{it}=k)} - a_i - \eta_i^+ - u_{it}\right)^2}{\chi^2_{n_{G_k}}}, \\
\sigma_{u,k}^2 \mid - &\sim \frac{\sum_{i,t}\mathbb{I}(G_{it} = k)\, u_{it}^2}{\chi^2_{n_{G_k}}}, \quad n_{G_k} = \sum_{i,t}\mathbb{I}(G_{it} = k), \\
\sigma_\eta^2 \mid - &\sim \frac{\sum_{i:G_i=k}\left(\eta_i^+\right)^2}{\chi^2_n}, \\
\sigma_a^2 \mid - &\sim \frac{\sum_{i:G_i=k} a_i^2}{\chi^2_n}.
\end{aligned}$$

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, 2nd edition.

Assaf, A. G., Berger, A. N., Roman, R. A., and Tsionas, M. G. (2019). Does efficiency help banks survive and thrive during financial crises? *Journal of Banking & Finance*, 106:445–470.

Badunenko, O. and Kumbhakar, S. C. (2017). Economies of scale, technical change and persistent and time-varying cost efficiency in indian banking: Do ownership, regulation and heterogeneity matter? *European Journal of Operational Research*, 260(2):789–803.

Badunenko, O., Kumbhakar, S. C., and Lozano-Vivas, A. (2021). Achieving a sustainable cost-efficient business model in banking: The case of european commercial banks. *European Journal of Operational Research*, 293(2):773–785.

Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, 32(3):870–897.

Bauwens, L., Koop, D., Korobilis, D., and Rombouts, K. P. (2015). Markov switching and structural break models: An application to the u.s. economy. *Journal of Applied Econometrics*, 30(7):1130–1154.

Botev, Z. I. (2017). The normal law under linear restrictions: Simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):125–148.

Cerqueiro, G., Degryse, H., and Ongena, S. (2011). Rules versus discretion in loan rate setting. *Journal of Financial Intermediation*, 20(4):503–529.

Curi, C., Lozano-Vivas, A., and Zelenyuk, V. (2015). Foreign bank diversification and efficiency prior to and during the financial crisis: Does one business model fit all? *Journal of Banking & Finance*, 61:S22–S35.

Custodio João, I., Lucas, A., Schaumburg, J., and Schwaab, B. (2023). Dynamic clustering of multivariate panel data. *Journal of Econometrics*, 237(2, Part B):105281.

Custodio João, I., Schaumburg, J., Lucas, A., and Schwaab, B. (2024). Dynamic nonparametric clustering of multivariate panel data. *Journal of Financial Econometrics*, 22(2):335–374.

Devroye, L. (2009). On exact simulation algorithms for some distributions related to Jacobi theta functions. *Statistics & Probability Letters*, 79(21):2251–2259.

DeYoung, R., Hunter, W. C., and Udell, G. F. (2004). The past, present, and probable future for community banks. *Journal of Financial Services Research*, 25:85–133.

DeYoung, R. and Rice, T. (2004a). How do banks make money? a variety of business strategies. *Economic Perspectives-Federal Reserve Bank of Chicago*, 28(4):52.

DeYoung, R. and Rice, T. (2004b). Noninterest income and financial performance at us commercial banks. *Financial Review*, 39(1):101–127.

Dunson, D. B. and Park, J.-H. (2008). Kernel stick-breaking processes. *Biometrika*, 95(2):307–323.

Dunson, D. B. and Rodríguez, A. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis*, 6(1):145–177.

Elsas, R., Hackethal, A., and Holzhäuser, M. (2010). The anatomy of bank diversification. *Journal of Banking & Finance*, 34(6):1274–1287.

Fahlenbrach, R., Prilmeier, R., and Stulz, R. M. (2012). This time is the same: Using bank performance in 1998 to explain bank performance during the recent financial crisis. *The Journal of Finance*, 67(6):2139–2185.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Geweke, J. and Keane, M. (2007). Smoothly mixing regressions. *Journal of Econometrics*, 138(1):252–290.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 27(1):143–158.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2):269–303.

Kuhn, H. W. (1955). The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97.

Kumbhakar, S. C., Park, B. U., Simar, L., and Tsionas, E. G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics*, 137(1):1–27.

Kumbhakar, S. C. and Tsionas, E. G. (2005). Measuring technical and allocative inefficiency in the translog cost system: A Bayesian approach. *Journal of Econometrics*, 126(2):355–384.

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

Narisetty, N. N. and He, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *The Annals of Statistics*, 42(2):789–817.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349.

Ren, L., Du, L., Carin, L., and Dunson, D. (2011). Logistic stick-breaking process. *Journal of Machine Learning Research*, 12(7):203–239.

Rigon, T. and Durante, D. (2021). Tractable Bayesian density regression via logit stick-breaking priors. *Journal of Statistical Planning and Inference*, 211:131–142.

Schröder, M. (2024). Mixing it up: Inflation at risk. *arXiv:2405.17237v2 [econ.EM]*.

Tokdar, S. T. (2006). Posterior consistency of dirichlet location-scale mixture of normals in density estimation and regression. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 68(1):90–110.

Tsionas, E. G. and Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. *Journal of Applied Econometrics*, 29(1):110–132.

Tsionas, M., Parmeter, C. F., and Zelenyuk, V. (2023a). Bayesian artificial neural networks for frontier efficiency analysis. *Journal of Econometrics*, 236(2):105491.

Tsionas, M., Parmeter, C. F., and Zelenyuk, V. (2023b). Bayesian artificial neural networks for frontier efficiency analysis. *Journal of Econometrics*, 236(2):105491.

van Oordt, M. and Zhou, C. (2019). Systemic risk and bank business models. *Journal of Applied Econometrics*, 34(3):365–384.

Ward, J. H. (1955). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Wheelock, D. C. and Wilson, P. W. (2000). Why do banks disappear? the determinants of us bank failures and acquisitions. *Review of Economics and Statistics*, 82(1):127–138.

Windle, J. B. (2013). *Forecasting High-Dimensional, Time-Varying Variance-Covariance Matrices with High-Frequency Data and Sampling Pólya-Gamma Random Variates for Posterior Distributions Derived from Logistic Likelihoods*. PhD thesis, University of Texas at Austin.

Zott, C., Amit, R., and Massa, L. (2011). The business model: recent developments and future research. *Journal of Management*, 37(4):1019–1042.

Online Supplement to "Bayesian Nonparametric Inference in Bank Business Models with Transient and Persistent Cost Inefficiency"

Dimitris Korobilis and Emmanuel Mamatzakis and Vasileios Pappas

S.1 Data Appendix

Our sample consists of commercial, cooperative, and savings banks from the European Union and the United Kingdom, covering the period 2008–2022. We obtain annual financial data from BankFocus, excluding banks without a complete data history. The data is cleaned according to standard empirical banking practices:

1. **Exclusion of incomplete data**: We remove banks with missing values for key financial variables, including assets, deposits, liabilities, equity, and income.
2. **Preference for consolidated reports**: If a bank provides both consolidated and unconsolidated financial statements, we use the consolidated version.
3. **Selection criteria for multiple reports**: When multiple financial statements are available, we prioritize restated over original, audited over unaudited, and International Financial Reporting Standards (IFRS) over Generally Accepted Accounting Principles (GAAP)-based reports.
4. **Outlier treatment**: To mitigate the influence of extreme values, we winsorize key financial variables at the 1st and 99th percentiles.

The final balanced panel consists of **7,395 bank-year observations**, corresponding to **15 annual observations** ($T = 15$) for **493 banks** ($n = 493$). The number of banks per country ranges from **one (Estonia, Greece)** to **258 (Germany)**. The sample includes all EU-27 countries except Cyprus, covering: Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom. All monetary values are converted to **real terms** using country-specific **GDP deflators** from the **IMF's International Financial Statistics database**, with **2010 as the base year**.

For cost efficiency estimation, consistent with the bank efficiency literature (Sealey and Lindley, 1977; Berger and Humphrey, 1997), we adopt the intermediary approach, which assumes that banks collect funds and use labor and physical capital to transform them into loans and other earning assets. The bank's technology is modeled using a translog cost function with three inputs—labor, physical capital, and financial capital—and two outputs: loans and other earning assets. The input prices are defined as follows: the price of labor is the ratio of personnel expenses to total assets, the price of physical capital is the ratio of administrative expenses to fixed assets, and the price of financial capital is total interest expenses divided by total interest-bearing borrowed funds. Total cost is defined as the sum of operating expenses, interest expenses, and non-interest expenses. Table S.1 presents key summary statistics.

Table S.1: Summary statistics: Outputs and inputs.

Variable	Mean	SD	P1	Median	P99	Obs
Loans	5,634.08	8,121.42	57.45	2,628.44	51,302.28	7,395
Other earning assets	1,586.20	3,027.31	0.01	842.94	25,604.47	7,395
Total cost	156.80	209.33	10.51	88.38	1,616.73	7,395
Price of labor	0.0091	0.0039	0.0001	0.0094	0.0215	7,395
Price of financial capital	1.3493	1.0809	0.0128	1.1108	5.0360	7,395
Price of physical capital	2.4035	8.1811	0.0100	0.6467	67.9978	7,395

Notes: The table presents summary statistics of the outputs, inputs, and total cost variables used in the cost efficiency estimation. The variables are winsorized at the 1/99 percentiles and deflated using the country-specific GDP deflator.

For identifying bank business models, we strategically select variables across three key dimensions: (i) asset composition, (ii) funding quality, and (iii) income mix. Our variable selection aligns with prior works of Badunenko et al. (2021), Curi et al. (2015), and Elsas et al. (2010). Specifically, asset composition includes loans, loans to banks, derivatives, securities, non-earning assets, and fixed assets. Funding quality comprises deposits, bank deposits, short-term funding, derivatives, long-term funding, reserves, and equity. Income mix focuses on interest income and non-interest income. Table S.2 presents key summary statistics.

Table S.2: Summary statistics: Bank business models

Variable	Mean	SD	P1	Median	P99	Obs
Panel A. Asset composition						
Loans	7,007.12	26,387.11	57.45	2,628.44	51,302.28	7,395
Loans to banks	1,398.54	6,103.98	1.72	304.97	18,595.35	7,395
Derivatives	1,191.47	28,055.88	0.00	0.00	1,936.91	7,395
Securities	4,127.35	53,105.80	0.00	842.94	25,604.47	7,395
Non-earning assets	711.71	6,856.80	0.14	98.29	6,571.46	7,395
Fixed assets	75.12	304.74	0.00	32.66	723.52	7,395
Panel B. Funding quality						
Deposits	6,369.59	24,671.78	57.91	2,904.58	46,887.70	7,395
Bank deposits	2,085.45	6,342.55	0.00	498.68	20,121.28	7,395
Short-term funding	8,732.18	42,136.56	0.00	3,425.51	65,277.09	7,395
Derivatives	1,182.34	27,555.03	0.00	0.00	2,570.07	7,395
Long-term funding	931.70	5,308.67	0.00	13.45	20,935.95	7,395
Reserves	511.36	8,727.03	4.48	63.30	7,353.87	7,395
Equity	929.28	3,879.91	46.05	353.04	6,217.61	7,395
Panel C. Income mix						
Non-interest income	136.03	1,004.11	-0.66	35.31	1,030.30	7,395
Interest income	341.23	2,906.40	13.62	114.21	2,666.30	7,395

Notes: The table presents summary statistics of the bank business models dimensions. The variables are deflated using the country-specific GDP deflator.

S.2 Additional empirical results

In this section we present results from additional runs of our model using different assumptions. In subsection S.2.1 we present efficiency estimates from our model with $K = 1$ and $K = 3$, where we remind that the special case $K = 1$ is the model in Tsionas and Kumbhakar (2014). In subsection S.2.2 and subsection S.2.3 we show results from various versions of our model. In one case we estimate the model with the LSBP prior being dependent on the first lag of z_{it}, i.e. we convert equation (4) in the main paper into

$$\pi_k\left(z_{it}\right) = \nu_k\left(z_{it-1}\right) \prod_{l=1}^{k-1} \left[1 - \nu_l\left(z_{it-1}\right)\right], \tag{S.1}$$

where nu_k is the same logit function for cluster $k = 1, ..., K$.

We also consider a restricted version of the translog cost function where certain coefficients are constrained to follow theoretically motivated sign restrictions. Specifically, we impose the following restrictions:

- **First-order coefficients**: The coefficients on the output terms (b_1, b_2) and input price terms (a_1, a_2, a_3) are constrained to be **positive**, ensuring that cost increases with output and input prices.
- **Second-order coefficients**:
  - The quadratic terms for input prices (a_{11}, a_{22}, a_{33}) are constrained to be **positive**, ensuring concavity of the cost function in input prices.
  - No sign restrictions are imposed on the quadratic terms for outputs (b_{11}, b_{12}, b_{22}) or on cross-terms between inputs (a_{12}, a_{13}, a_{23}).
- **Interaction terms between outputs and inputs** (g_{km}): No sign restrictions are imposed on these parameters.
- **Time trend terms** (τ_t, τ_t^2): The linear trend (τ_t) and quadratic trend (τ_t^2) are constrained to be **negative**, capturing potential cost-reducing technological progress over time.
- **Additional controls** $(d_1, d_2, \theta_1, \theta_2, \theta_3)$: No sign restrictions are imposed.

These restrictions are implemented to ensure consistency with economic theory while allowing sufficient flexibility in the estimation.

S.2.1 Efficiency measures

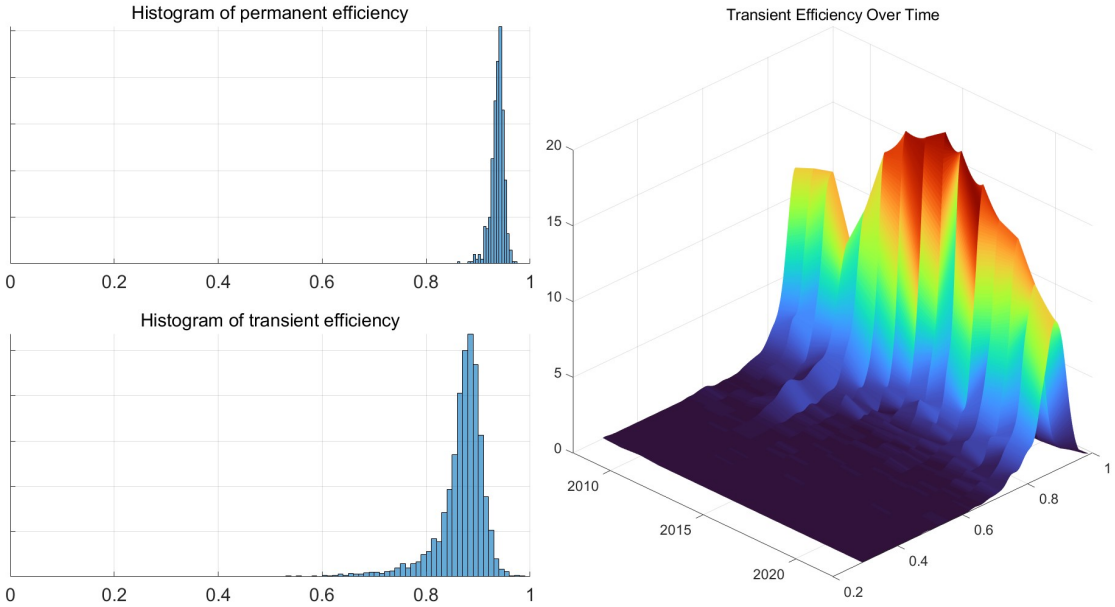

Figure S.1: **Efficiency Decomposition in the LSBP Stochastic Frontier Model ($K = 1$ clusters, that is, the model in Tsionas and Kumbhakar (2014)):**
This figure presents the decomposition of efficiency into permanent and transitory components. The top left panel displays the histogram of permanent efficiency, illustrating the distribution of long-term efficiency levels across banks. The bottom panel shows the histogram of transient efficiency, capturing the short-term fluctuations in efficiency across all banks and time periods. The right panel, spanning the right half of the figure, presents a 3D histogram of transient efficiencies, visualizing how these vary across banks and years.

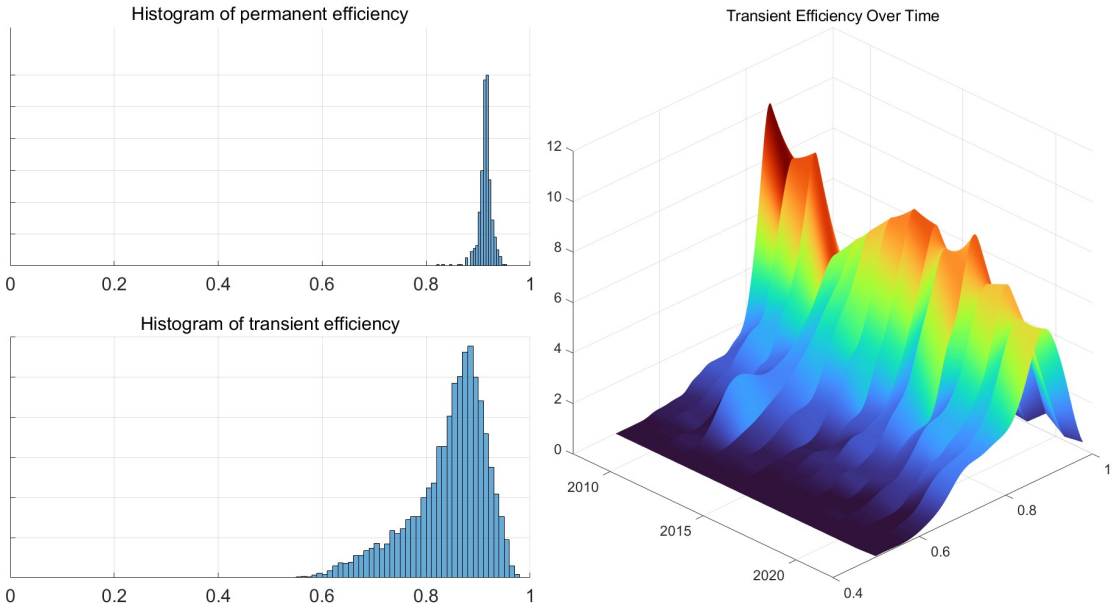

Figure S.2: **Efficiency Decomposition in the LSBP Stochastic Frontier Model ($K = 3$ clusters):**
This figure presents the decomposition of efficiency into permanent and transitory components. The top left panel displays the histogram of permanent efficiency, illustrating the distribution of long-term efficiency levels across banks. The bottom panel shows the histogram of transient efficiency, capturing the short-term fluctuations in efficiency across all banks and time periods. The right panel, spanning the right half of the figure, presents a 3D histogram of transient efficiencies, visualizing how these vary across banks and years.

S.2.2 BBM determinants under alternative assumptions

Table S.3: Z variables selected in the Logit Stick-Breaking Process for each cluster

Predictor	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Intercept	✓	✓	✓	✓
Panel A. Asset composition				
Loans	✓	✓	✓	✓
Loans to banks	✓	✓	✓	
Derivatives			✓	✓
Securities	✓			✓
Non-earning assets	✓		✓	
Fixed assets				
Panel B. Funding quality				
Deposits		✓	✓	✓
Bank deposits		✓	✓	
Short-term funding		✓	✓	
Derivatives	✓	✓	✓	✓
Long-term funding	✓			
Reserves	✓			
Equity	✓	✓		
Panel C. Income mix				
Non-interest income	✓	✓		
Interest income		✓		

Notes: This is the equivalent of Table 5 in the main text, but the variables z_{it} entering the LSBP specification with a lag, to address endogeneity concerns in the determination of the clusters/BBMs.

Table S.4: z_{it} variables selected in the Logit Stick-Breaking Process for each cluster

Predictor	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Intercept	✓	✓	✓	✓
Panel A. Asset composition				
Loans (aL)	✓	✓	✓	
Loans to banks (aLTB)	✓	✓	✓	
Derivatives (aDERV)	✓	✓	✓	
Securities (aSEC)	✓		✓	✓
Non-earning assets (aNEA)	✓		✓	
Fixed assets (aFA)	✓		✓	✓
Panel B. Funding quality				
Deposits (dDEP)		✓	✓	
Bank deposits (dDEPB)	✓			
Short-term funding (dSTB)		✓	✓	
Deposit variations (dDEVT)		✓	✓	
Long-term funding (dLTF)	✓			
Reserves (dREV)	✓			
Equity (dEQT)	✓	✓	✓	
Panel C. Income mix				
Non-interest income (iNII)	✓	✓		
Interest income (iII)		✓	✓	

Notes: This is the equivalent of Table 5 in the main text, but with coefficients of the translog function having sign restrictions.

S.2.3 Estimated translog function coefficients

Table S.5: Regression Coefficients by Cluster: Benchmark model, $K = 4$

Parameter	Cluster 1	Cluster 2	Cluster 3	Cluster 4
β_0	1.355 (0.530)	18.790 (5.289)	2.281 (0.430)	3.125 (0.393)
β_1^y	0.410 (0.081)	-1.403 (0.829)	0.472 (0.077)	0.029 (0.089)
β_2^y	0.170 (0.027)	0.044 (0.432)	0.146 (0.019)	0.018 (0.034)
β_1^w	0.475 (0.177)	3.353 (0.826)	0.912 (0.073)	0.391 (0.070)
β_2^w	0.325 (0.134)	0.822 (0.812)	0.189 (0.096)	0.050 (0.054)
β_3^w	0.123 (0.047)	0.132 (0.380)	0.132 (0.035)	0.032 (0.035)
β_{11}^{yy}	0.109 (0.011)	0.225 (0.090)	0.069 (0.009)	0.151 (0.008)
β_{12}^{yy}	-0.031 (0.008)	-0.044 (0.083)	-0.014 (0.005)	-0.030 (0.006)
β_{22}^{yy}	0.015 (0.002)	0.033 (0.014)	0.006 (0.001)	0.032 (0.002)
β_{11}^{ww}	0.157 (0.028)	0.384 (0.074)	0.255 (0.011)	0.133 (0.023)
β_{12}^{ww}	0.095 (0.031)	-0.124 (0.138)	0.120 (0.021)	-0.007 (0.019)
β_{13}^{ww}	-0.000 (0.010)	0.029 (0.041)	0.020 (0.006)	0.005 (0.008)
β_{22}^{ww}	-0.045 (0.011)	-0.037 (0.138)	-0.000 (0.010)	0.013 (0.007)
β_{23}^{ww}	0.004 (0.015)	0.020 (0.088)	-0.014 (0.010)	0.003 (0.014)
β_{33}^{ww}	0.002 (0.003)	-0.049 (0.030)	0.009 (0.003)	0.005 (0.004)
β_{11}^{yw}	0.109 (0.007)	-0.039 (0.075)	0.074 (0.006)	0.104 (0.015)
β_{12}^{yw}	-0.018 (0.010)	-0.135 (0.112)	0.010 (0.007)	-0.017 (0.008)
β_{13}^{yw}	-0.015 (0.004)	-0.009 (0.046)	-0.008 (0.003)	-0.003 (0.003)
β_{21}^{yw}	0.002 (0.005)	0.009 (0.029)	0.011 (0.003)	-0.013 (0.004)
β_{22}^{yw}	0.002 (0.007)	0.013 (0.037)	-0.000 (0.004)	0.007 (0.005)
β_{23}^{yw}	0.001 (0.002)	0.011 (0.018)	0.003 (0.002)	0.003 (0.001)
β_τ	0.030 (0.019)	-0.304 (0.247)	0.018 (0.016)	0.050 (0.018)
β_{τ^2}	-0.003 (0.001)	-0.008 (0.005)	-0.002 (0.000)	-0.000 (0.000)
β_1^δ	0.001 (0.002)	0.010 (0.024)	0.007 (0.002)	-0.001 (0.001)
β_2^δ	-0.003 (0.002)	0.015 (0.012)	-0.001 (0.001)	-0.002 (0.001)
β_1^θ	-0.005 (0.003)	-0.046 (0.020)	0.011 (0.002)	0.005 (0.003)
β_2^θ	-0.002 (0.003)	-0.020 (0.025)	-0.002 (0.002)	0.003 (0.002)
β_3^θ	-0.001 (0.001)	-0.001 (0.012)	-0.001 (0.001)	0.001 (0.001)

Table S.6: Regression Coefficients by Cluster: Model with sign restrictions, $K = 4$

Parameter	Cluster 1	Cluster 2	Cluster 3	Cluster 4
β_0	3.070 (0.170)	3.703 (0.362)	4.119 (0.194)	6.176 (0.948)
β_1^y	0.122 (0.020)	0.080 (0.038)	0.079 (0.021)	0.200 (0.117)
β_2^y	0.038 (0.016)	0.023 (0.007)	0.012 (0.007)	0.299 (0.137)
β_1^w	0.173 (0.035)	0.146 (0.045)	0.140 (0.031)	1.489 (0.169)
β_2^w	0.034 (0.033)	0.029 (0.025)	0.035 (0.044)	0.157 (0.115)
β_3^w	0.133 (0.050)	0.063 (0.043)	0.027 (0.028)	0.188 (0.086)
β_{11}^{yy}	0.044 (0.009)	0.040 (0.012)	0.030 (0.006)	0.052 (0.015)
β_{12}^{yy}	0.008 (0.003)	0.006 (0.004)	-0.001 (0.000)	-0.033 (0.029)
β_{22}^{yy}	0.005 (0.004)	0.001 (0.006)	0.003 (0.002)	0.017 (0.009)
β_{11}^{ww}	0.002 (0.003)	0.001 (0.001)	0.000 (0.000)	0.231 (0.025)
β_{12}^{ww}	-0.070 (0.047)	0.056 (0.012)	-0.007 (0.012)	-0.021 (0.045)
β_{13}^{ww}	0.023 (0.016)	0.020 (0.025)	-0.057 (0.007)	0.062 (0.017)
β_{22}^{ww}	0.006 (0.008)	0.002 (0.002)	0.002 (0.001)	0.011 (0.010)
β_{23}^{ww}	-0.051 (0.032)	-0.031 (0.019)	0.008 (0.004)	0.027 (0.027)
β_{33}^{ww}	0.005 (0.004)	0.001 (0.001)	0.001 (0.001)	0.015 (0.008)
β_{11}^{yw}	0.032 (0.006)	0.023 (0.009)	0.013 (0.004)	0.020 (0.010)
β_{12}^{yw}	-0.033 (0.023)	0.003 (0.004)	-0.004 (0.005)	-0.028 (0.018)
β_{13}^{yw}	0.002 (0.005)	0.005 (0.005)	-0.019 (0.001)	0.005 (0.008)
β_{21}^{yw}	0.002 (0.001)	0.003 (0.003)	0.002 (0.000)	0.036 (0.012)
β_{22}^{yw}	-0.004 (0.004)	0.004 (0.005)	-0.003 (0.003)	0.007 (0.015)
β_{23}^{yw}	-0.005 (0.004)	-0.004 (0.005)	0.001 (0.001)	0.001 (0.009)
β_τ	-0.014 (0.007)	-0.020 (0.007)	-0.016 (0.002)	-0.022 (0.019)
β_{τ^2}	-0.004 (0.002)	-0.002 (0.001)	-0.001 (0.000)	-0.003 (0.002)
β_1^δ	0.017 (0.003)	0.010 (0.002)	0.007 (0.002)	0.001 (0.006)
β_2^δ	-0.004 (0.001)	-0.002 (0.000)	-0.000 (0.000)	-0.003 (0.003)
β_1^θ	0.005 (0.002)	0.004 (0.002)	0.010 (0.001)	-0.015 (0.005)
β_2^θ	0.006 (0.003)	0.004 (0.003)	-0.001 (0.001)	0.003 (0.008)
β_3^θ	-0.011 (0.004)	-0.006 (0.002)	-0.000 (0.001)	-0.002 (0.004)

Table S.7: Regression Coefficients by Cluster: Model with country-specific predictors, $K = 4$

Parameter	Cluster 1	Cluster 2	Cluster 3	Cluster 4
β_0	0.433 (0.418)	21.709 (3.477)	3.614 (0.563)	4.967 (0.438)
β_1^y	0.408 (0.094)	-1.226 (0.610)	-0.089 (0.114)	-0.150 (0.091)
β_2^y	0.090 (0.038)	-0.296 (0.294)	0.139 (0.024)	0.084 (0.017)
β_1^w	-0.199 (0.078)	3.419 (0.693)	0.668 (0.107)	0.707 (0.078)
β_2^w	0.118 (0.068)	0.395 (0.951)	0.332 (0.045)	-0.036 (0.047)
β_3^w	0.013 (0.038)	0.260 (0.318)	0.033 (0.044)	0.141 (0.028)
β_{11}^{yy}	0.123 (0.012)	0.159 (0.100)	0.140 (0.014)	0.173 (0.010)
β_{12}^{yy}	-0.007 (0.010)	0.011 (0.084)	-0.025 (0.008)	-0.038 (0.006)
β_{22}^{yy}	0.016 (0.002)	0.036 (0.013)	0.017 (0.002)	0.028 (0.003)
β_{11}^{ww}	0.084 (0.011)	0.397 (0.086)	0.190 (0.011)	0.185 (0.010)
β_{12}^{ww}	0.051 (0.013)	-0.127 (0.170)	0.152 (0.014)	-0.007 (0.018)
β_{13}^{ww}	-0.039 (0.008)	0.094 (0.052)	0.015 (0.007)	0.024 (0.010)
β_{22}^{ww}	0.032 (0.010)	0.117 (0.132)	0.014 (0.007)	0.027 (0.006)
β_{23}^{ww}	-0.018 (0.008)	0.065 (0.129)	-0.037 (0.006)	-0.003 (0.005)
β_{33}^{ww}	-0.006 (0.004)	-0.044 (0.036)	0.007 (0.004)	-0.001 (0.002)
β_{11}^{yw}	0.146 (0.006)	-0.046 (0.058)	0.053 (0.008)	0.102 (0.006)
β_{12}^{yw}	-0.001 (0.005)	-0.113 (0.108)	-0.011 (0.004)	-0.003 (0.008)
β_{13}^{yw}	-0.016 (0.004)	0.006 (0.040)	0.010 (0.005)	-0.009 (0.002)
β_{21}^{yw}	0.009 (0.002)	0.012 (0.035)	0.010 (0.002)	-0.009 (0.003)
β_{22}^{yw}	-0.009 (0.002)	0.072 (0.050)	0.010 (0.005)	0.002 (0.004)
β_{23}^{yw}	0.002 (0.002)	0.019 (0.023)	-0.004 (0.001)	-0.000 (0.002)
β_τ	-0.039 (0.015)	-0.354 (0.215)	0.107 (0.020)	-0.039 (0.016)
β_{τ^2}	-0.001 (0.000)	-0.004 (0.005)	-0.000 (0.000)	-0.001 (0.000)
β_1^δ	0.005 (0.001)	0.014 (0.019)	-0.001 (0.002)	0.004 (0.002)
β_2^δ	-0.001 (0.001)	0.027 (0.012)	0.001 (0.001)	-0.005 (0.001)
β_1^θ	-0.002 (0.002)	-0.020 (0.019)	0.024 (0.002)	-0.009 (0.002)
β_2^θ	0.002 (0.001)	-0.008 (0.023)	0.005 (0.001)	0.004 (0.002)
β_3^θ	-0.001 (0.001)	-0.015 (0.012)	-0.004 (0.001)	0.000 (0.001)
Infl_CPI	-1.532 (0.300)	-2.809 (3.910)	0.101 (0.247)	0.749 (0.265)
GDP_real_gr	0.353 (0.167)	1.193 (1.386)	0.145 (0.137)	0.155 (0.118)
Credit_priv_sect	-0.003 (0.000)	-0.016 (0.003)	-0.003 (0.000)	-0.003 (0.000)

References

Badunenko, O., Kumbhakar, S. C., and Lozano-Vivas, A. (2021). Achieving a sustainable cost-efficient business model in banking: The case of european commercial banks. *European Journal of Operational Research*, 293(2):773–785.

Berger, A. N. and Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*, 98(2):175–212.

Curi, C., Lozano-Vivas, A., and Zelenyuk, V. (2015). Foreign bank diversification and efficiency prior to and during the financial crisis: Does one business model fit all? *Journal of Banking & Finance*, 61:S22–S35.

Elsas, R., Hackethal, A., and Holzhäuser, M. (2010). The anatomy of bank diversification. *Journal of Banking & Finance*, 34(6):1274–1287.

Sealey, C. W. and Lindley, J. T. (1977). Inputs, outputs, and the theory of production and cost of depository financial institutions. *Journal of Finance*, 32(4):1251–1266.

Tsionas, E. G. and Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. *Journal of Applied Econometrics*, 29(1):110–132.