Received: 26 July 2019 | Revised: 3 January 2020 | Accepted: 5 March 2020
DOI: 10.1002/isaf.1471

RESEARCH ARTICLE

WILEY

Using clustering ensemble to identify banking business models

Bernardo P. Marques[1,2] | Carlos F. Alves[2]

[1]Universidade Católica Portuguesa, Católica Porto Business School, Porto, Portugal.
[2]Universidade do Porto, Faculdade de Economia, CEF.UP, Porto, Portugal

Correspondence
Bernardo P. Marques, Universidade Católica Portuguesa, Católica Porto Business School, Porto, Portugal.
Email: bmarques@porto.ucp.pt

Funding information
Fundação para a Ciência e a Tecnologia, Grant/Award Numbers: SFRH/BD/135939/2018 (Phd scholarship), UIDB/00731/2020 (CPBS), UIDB/04105/2020 (CEF.UP)

Summary

The business models of banks are often seen as the result of a variety of simultaneously determined managerial choices, such as those regarding the types of activities, funding sources, level of diversification, and size. Moreover, owing to the fuzziness of data and the possibility that some banks may combine features of different business models, the use of hard clustering methods has often led to poorly identified business models. In this paper we propose a framework to deal with these challenges based on an ensemble of three unsupervised clustering methods to identify banking business models: fuzzy c-means (which allows us to handle fuzzy clustering), self-organizing maps (which yield intuitive visual representations of the clusters), and partitioning around medoids (which circumvents the presence of data outliers). We set up our analysis in the context of the European banking sector, which has seen its regulators increasingly focused on examining the business models of supervised entities in the aftermath of the twin financial crises. In our empirical application, we find evidence of four distinct banking business models and further distinguish between banks with a clearly defined business model (core banks) and others (non-core banks), as well as banks with a stable business model over time (persistent banks) and others (non-persistent banks). Our proposed framework performs well under several robustness checks related with the sample, clustering methods, and variables used.

KEYWORDS
banking, business models, clustering ensemble, fuzzy clustering, self-organizing maps

1 | INTRODUCTION

This paper deals with the special methodological requirements that emerge from the task of business model identification—a task that has gained particular relevance in the context of recent efforts to reform the regulation and supervision of banks in Europe (EBA, 2014; ECB, 2018). In particular, policymakers and researchers have become increasingly focused on grouping banks based on the similarity of their business model choices (such as size, types of activities, funding, and diversification). However, in doing so, they have faced significant challenges in finding clearly separated and homogeneous clusters. This occurs chiefly because business choices are likely to follow a fuzzy, rather than a crisp, logic; for example, some banks may choose to combine features of different business models following a merger or acquisition (DeSarbo and Grewal, 2008).

In general, by applying clustering analysis to the business choices of banks one may hope to achieve two main goals. First, to obtain an objective and stable taxonomy of business model classifications, which in turn may be used by supervisors to monitor the performance of banks in each business model (e.g. by identifying outliers)—in line with the guidelines for the supervisory review and evaluation process (EBA, 2014). Second, to obtain a better insight into the competitive structure of the banking sector, as banks with similar business choices may be expected to compete more intensely among themselves (Porter, 1979). The former goal (i.e. attaining an objective and stable taxonomy of business models) seems particularly timely given that the method currently used by supervisors to identify business models (expert judgement) may lead to inconsistencies and potential conflict of interests. In particular, under the principle of proportionality, different banking business models may be expected to entail different degrees of monitoring effort for the supervisor. Hence, if the allocation of banks to business models is based on the subjective assessment of supervisors, under first principles these may have the incentive to allocate banks into business models that are easier to monitor or subject to stricter regulatory requirements (e.g. higher capital requirements). The same rationale may be applied to business model self-reporting by banks. In this context, we argue that finding an objective and reliable method to allocate banks into business models is paramount for the implementation of business model specific regulation and supervision.

The use of clustering analysis to identify banking business models, however, bears significant challenges of its own, including those related with the choice of method. For instance, a recent strand of research has relied exclusively on hard clustering methods (e.g. hierarchical clustering (HC)) to identify the business models of banks, failing to apply methods that enable banks to have some affinity with more than one business model, such as fuzzy c-means (FCM). For instance, Mergaerts and Vander Vennet (2016) applied HC on seven business model variables for a sample of European banks (1998–2013) and reported an average silhouette width of 0.20 for a partition in three clusters, a value that is below the threshold of 0.25 for minimum quality of clustering as proposed by Kaufman and Rousseeuw (1990). Similarly, Martín-Oliver *et al*. (2017) applied HC on six variables for a sample of Spanish banks and reported persistency levels of business model classification across consecutive periods (1999–2002 versus 2003–2007) that range from 10.4% (lowest) to 85.7% (highest). In our view, both studies raise some concerns regarding the usefulness and reliability of results that are obtained by applying hard classification methods to the identification of banking business models.

Conversely, by using fuzzy clustering to identify banking business models one may be able to measure the similarity that each bank holds with the prototypical models (i.e. percentage of cluster membership). In turn, such a measure may be used in several empirical contexts in business model analysis, such as: (i) the identification of whether a bank combines features of more than one business model, and which models those are; (ii) the use of the measure in its original format (i.e. continuous value from 0 to 1) as an explanatory variable in performance and riskiness-related fixed effects regression (not possible when the business model assignment is stable and discrete); and (iii) its conversion into a discrete measure, 0 or 1, based on the business model with which the bank has the highest percentage of membership, enabling, for instance, a supervisor to identify peer groups of banks based on their business model.

This paper contributes to the literature in several ways. First, we provide a formal definition of 'banking business model' grounded on strategic management literature, namely the configurational approach (Miller, 1986) and strategic groups theory (DeSarbo and Grewal, 2008; Reger and Huff, 1993). Second, by applying principal components analysis to an array of banking variables, we identify five strategic dimensions along which banks assume a long-term position relative to their peers (Galbraith and Schendel, 1983). Third, based on the notion of consensus-based classification (Kuncheva, 2004), we identify the business models of European banks using an ensemble of three unsupervised clustering methods: FCM (Bezdek *et al*., 1984), which allows us to handle fuzzy clustering; self-organizing maps (SOMs; Kohonen, 1997), which yield intuitive visual representations of the clusters; and partitioning around medoids (PAM; Kaufman and Rousseeuw, 1990), which circumvents the presence of data outliers. Fourth, we examine the level of similarity of banks operating with the same long-term business model (core versus non-core banks). Finally, we provide some evidence regarding the level of persistency of banks in terms of their business model, as well as examine the factors that influence the likelihood of non-persistency per business model.

Briefly put, our approach begins with the implementation of principal component analysis with the aim of identifying a set of business model components. This step allows us to perform clustering on a space with orthogonal dimensions, as well as to focus on the most relevant relationships between business model choices and, thus, hopefully mitigate the problem of data noisiness. The second step is to run three clustering methods (PAM, FCM, and SOMs), combine their classification output and assign each bank to the business model (cluster) with the majority of the 'votes' (clustering ensemble). Next, we label a bank as 'core' in a given business model if (i) the ensemble is unanimous (i.e. if the three methods assign the bank to the same business model) and (ii) the silhouette width using the clustering ensemble classification is above a threshold identified in literature. Finally, we look for persistent banks by dividing the full sample period (2005–2016) into four trienniums (2005–2007, 2008–2010, 2011–2013 and 2014–2016), identifying the business model of banks for each triennium separately (using triennium average values) and looking for banks for which the business model is the same in all the trienniums in which the bank is present in the sample.

By applying our method to the context of the European banking industry (2005–2016), we find evidence of four banking business models: retail focused, retail diversified funding, retail diversified assets and large diversified. Importantly, we test the stability of classification using alternative sub-sampling methods and find that the stability of classification is significantly higher when testing the samples of core banks, and core and persistent banks when compared to tests with the full sample. Also, we find that the mean values of key dimensions of each banking business model change significantly when using the sample of core and persistent banks when compared to other banks. These results (stability and mean difference) may be seen as evidence of the suitability of our approach to identify banking business models.

This paper is structured in the following way. In Sections 2 and 3 we respectively survey applications of the 'business model' concept in banking regulation and recent literature on methods used to identify banking business models. A conceptual framework for banking business models is established in Section 4 . Section 5 provides an overview of key concepts in the 'clustering ensemble' approach, as well as a brief description of the classification methods used in the paper. The data set is presented and described in Section 6. In Section 7 we identify the procedures used in our methodology. Section 8 deals with the results and discussion, and Section 9 presents robustness checks. In the final section, we conclude and identify opportunities for future research. For brevity reasons, we include the description of the clustering algorithms and valuation criteria in the Appendix.

2 | BUSINESS MODELS' RELEVANCE IN RECENT BANKING REGULATION AND SUPERVISION

The importance of monitoring the different aspects of banking business models has been stressed in recent efforts to reform the regulation and supervision of banks. For instance, in the 'High-Level Expert Group's report on reforming the structure of the EU banking sector', also known as the Liikanen Report (Liikanen, 2012), an entire section is dedicated to the analysis of banking business models in the EU (Liikanen, 2012: 32–66), concluding that: 'while all types of bank business model have been affected in the crisis, some characteristics have proven less resilient than others. The main bank failures have been attributed to overreliance on short-term wholesale funding, excessive leverage, excessive trading/derivative/market activity (…)' (p. 32).

Additionally, business model analysis has become a key procedure in the Supervisory Review and Evaluation Process (SREP)[1] since its implementation on 1 January 2016. In particular, according to the European Banking Authority's (EBA's) guidelines for the SREP, supervisors are required to monitor, assess, and challenge the business models of supervised entities (EBA, 2014). On the other hand, the business model is depicted as a key element for the implementation of the principle of proportionality by supervisors in CRD IV,[2] which, among other things, mandates the EBA to assess the impact of new liquidity and leverage requirements on different banking business models (EBA, 2013, 2015, 2016).

Lastly, two instances further illustrate the recent attention attributed by regulators and supervisors to banking business models: the identification of business model and profitability risk as a top priority for the 'single supervisory mechanism' between 2016 and 2018 (ECB, 2018); and the approval by the US Congress of the 'Financial CHOICE Act' regulatory package in August 2017, which, among other things, reduces the reporting burden of traditional banking organizations (i.e. banks following a traditional business model).

[1]The SREP is an annual process carried out by supervisors (national central bank or European Central Bank, according to the systemic relevance of the entity) with the goal of reviewing whether the arrangements, strategies, processes, and mechanisms implemented by the supervised entities are in compliance with recent legislation (e.g. CRD IV).

[2]The term 'CRD IV' is used in banking literature and regulation to refer to the Capital Requirements Directive (Directive 2013/35/EU) and the Capital Requirements Regulation (Regulation no. 575/2013), jointly approved by the EU Parliament in April 2013 and with effect since 1 January 2014, which constitute the transposition to EU law of the post-crisis global regulatory reform, 'Basel III Agreement' (Basel Committee on Banking Supervision, 2011).

3 | LITERATURE REVIEW ON METHODS USED TO IDENTIFY BANKING BUSINESS MODELS

Recent literature reports four methods to identify the business models of banks. The first method is to apply discretionary rules on a set of business model proxies (Chiorazzo *et al*., 2018; Curi *et al*., 2015). Curi *et al*. (2015), for instance, use three measures of Herfindhal–Hirshman diversification (asset, funding, and income) and, based on the graphical observation of the estimated distributions of each variable, apply threshold values (0.35, 0.35, and 0.30 respectively) below which the banks are labelled as following a focused business model. For a sample of foreign banks located in Luxembourg, the authors found that banks that are cumulatively classified as focused in each of the three dimensions are more efficient than their peers. Methodologically, such an approach is simple to replicate and intuitive. However, the methods used may be seen as excessively discretionary, and often authors do not provide any evidence of the quality of the groupings (or clusters), fuelling some doubts regarding the similarity of the banks in each group.

The second method is to apply dimensionality reduction techniques to account for the multivariate and simultaneous nature of business model choices (de Haan and Kakes, 2019; Mergaerts and Vander Vennet, 2016; van Ewijk and Arnold, 2014). After applying hierarchical clustering and finding clusters with low quality (silhouette width of 0.2 for a partition in three clusters), Mergaerts and Vander Vennet (2016) applied factor analysis on seven business model variables and retained two factors: retail-orientation and diversification. Moreover, the authors accounted for the long-term nature of business models by using an econometric approach (Mundlak estimator) that separates the between and within effects of each factor on profitability and riskiness, wherein the between effect is attributed to the business model. In general, this method seems to allow an adequate grasp of the interconnections between business model choices and is able to separate long and short-term effects of business models. However, it fails to produce a business model classification, which may impair the ability for supervisors to analyse the performance and riskiness of peer groups.

The third method is to use expert judgement to identify business models (Cernov and Urbano, 2018; Köhler, 2015). In this strand, Cernov and Urbano (2018) combined a qualitative step (expert) with a quantitative step (unsupervised clustering) to define business models. Namely, in the first step, using the EBA's business model classification, which is comprised of 16 models, supervisors are asked to assign each supervised entity to a business model; in the second step, using 10 business model variables, the authors mapped the main activities and funding sources of each model and removed redundant business models, narrowing the initial set of 16 business models to 11. The authors argued that this approach allowed them to challenge the initial business model classification of some banks. For instance, the approach identifies large diversified banks, such as Crédit Agricole Group (CAG), as outliers in the initial model of cooperative banks/saving and loans associations and recommends it to be reclassified under the cross-border universal banks model. In our view, this approach is able to expand the typical scope and depth of the information used to classify business models. However, the collection of such granular data is often costly, inaccessible, and/or self-reported, which may impede cross-study comparisons.

The fourth and final method consists of employing hard clustering algorithms (Ayadi *et al*., 2011; Ayadi and De Groen, 2015; Roengpitya *et al*., 2014, 2017; Hryckiewicz and Kozłowski, 2017; Martín-Oliver *et al*., 2017). For instance, Martín-Oliver *et al*. (2017) employed hierarchical clustering with Ward's method on a sample of Spanish banks using different sub-periods and found four business models (retail-deposits, retail-balanced, retail-diversified, and retail-market). The authors documented that small cooperative banks that migrated to more risky business models before the crisis suffered higher losses during the crisis than shareholder banks with the same business model. However, the level of persistency of classifications in different periods is very low. For instance, 56.3% of the bank's assets that followed a 'retail-balanced model' in the 1999–2002 sub-period migrated to the 'retail diversified model' in the 2003–2007 sub-period, whereas only 10.4% remained in the same cluster. Similarly, in the 'retail-deposits' and 'retail-diversified' models, only 28.6% and 15.4% respectively of the bank's assets remained in the same cluster in consecutive sub-periods. These results provide a strong indication of the presence of issues regarding cluster quality and suggest that the use of fuzzy clustering methods is perhaps more appropriate to capture the business models of banks.

In summation, we retain two main issues that have been mishandled in recent banking literature: the general lack of a conceptual framework that clearly guides the methodology used to identify business models; and the inadequate use of hard clustering techniques to model data, which is likely to have a positive association with more than one cluster. In the next sections, we aim at contributing to mitigate these issues by providing a definition of 'banking business model' and by proposing an approach to identify business models that combines the results of alternative clustering methods.

4 | DEFINITION OF 'BANKING BUSINESS MODEL'

The banking literature has provided several definitions of business model. For example:

- 'We (…) define a business model as a simplified representation of the activities that a bank performs to make money' (Cavelaars and Passenier, 2012: 404).
- 'The concept of business models originates from the literature concerning strategic groups, i.e. sets of firms that are active in a single sector and use similar strategies. (…) that reflect the long-term choices of bank management with respect to assets, funding, capitalization and diversification' (Mergaerts and Vander Vennet, 2016: 58).
- 'A banking business model consists in a pattern of assets and liabilities adopted by one or several banks that differs from the pattern adopted by other banks, each with different combinations of expected return and risk' (Martín-Oliver *et al*., 2017: 248).

These citations illustrate that there is no unique definition of business model (Zott *et al*., 2011). Moreover, they showcase that the banking literature has not accounted for the fuzzy nature of business models. For that reason, we offer an alternative definition of 'banking business model' that attempts to overcome this and other misconceptions identified in the literature review (e.g. lack of stability of business models over time). Namely, we define a 'banking business model' as a predominantly stable and long-term oriented organizational configuration that is adopted, with different levels of association, by a significant share of banks, resulting from a set of observable and interconnected managerial choices.

More specifically, a banking business model is expected to be 'predominantly stable and long-term oriented' in the sense that key strategic decisions are likely to bear significant investments that impose mobility barriers on banks, impeding them to freely shift across business models—an insight that is borrowed from strategic groups theory (Porter, 1979: 214). On the other hand, when we state that a business model may be seen as an 'organizational configuration (…) adopted by a significant share of banks' we allude to the configurational approach (Miller, 1986), in particular to the notion that '(…) elements of strategy, structure and environment often coalesce or configure into a manageable number of common, predictively useful types that describe a large proportion of (…) organizations' (Miller, 1986: 235–236). The fuzzy nature of business models is explicitly incorporated in our proposed definition of banking business model when we state that banks may exhibit 'different levels of association' with one or more business models. This idea is borrowed from the fuzzy approach to strategic groups theory (DeSarbo and Grewal, 2008; Reger and Huff, 1993). For instance, Reger and Huff (1993: 117) suggested that, based on the level of agreement among industry participants regarding the allocation of a firm to a strategic group, it is possible to segment the firms into core ('that are tightly associated and define the basic 'recipe' of a strategic group'), secondary ('that implement the strategic group recipe less consistently than core firms'), and transient ('whose strategies are changing from one strategic position to another, but along dimensions common to other firms in the industry'). In order to be able to replicate the clustering results, our definition requires business model choices to be 'observable'; by this we mean that the variables used to proxy for business model choices should, preferably, be publicly available. Finally, we require business model choices to be 'interconnected' in order to incorporate Miller's (1986) concerns regarding the need for configurational studies to provide an explanation of 'why and how' the elements of the configuration relate and complement with each other. In this study, we explore the 'why and how' in two ways: (i) by surveying banking theory and recent empirical studies for predictions regarding the way that some business model choices are expected to be interconnected; and (ii) by following Galbraith and Schendel (1983) and applying principal component analysis to extract the main business model components (note that the rationale underlying the use of principal component analysis is discussed in the introduction to Section 7).

Importantly, this definition of business model embodies two notions that are at the heart of our proposed methodology. First, the possibility that banks may vary in terms of the similarity of association relative to the most representative banks in the assigned business model, hence allowing for a distinction between core and non-core banks. Second, although we define business model as a 'predominantly stable and long-term oriented' concept, the choice of wording ('predominantly') deliberately makes way for the possibility of banks changing business models over time and, thus, the distinction between persistent and non-persistent banks. Both concepts, in our view, would make it ideal for us to select a clustering method that may, cumulatively, (i) capture the fuzzy logic of business models, (ii) yield an intuitive visualization of the clusters, and (iii) circumvent the potential presence of data outliers. In Section 5, we describe three methods that have been used in literature to address each of these requirements, and we provide an overview of literature on the method that combines the outputs of different clustering methods (i.e. clustering ensemble).

5 | CLUSTERING METHODS AND CLUSTERING ENSEMBLES

5.1 | Fuzzy Clustering

Fuzzy logic is founded on the idea that, in some real-world clustering problems, the membership of a data point to a given configuration or object may be nuanced, and hence a binomial membership function is likely to be oversimplistic (Zadeh, 1965). According to this logic, in some situations the assignment of data points to clusters may be better depicted as a continuous function, truncated between 0 and 1, whereby the nearer the membership value is to unity, the higher is the similarity between the observation and the cluster.

The application of fuzzy logic to clustering was popularized with the FCM algorithm, initially formulated by Dunn (1974) and improved by Bezdek *et al*. (1984). In FCM, data are clustered into a predetermined total number of clusters by iteratively minimizing the weighted within-group sum of squared errors, where the fuzzified membership of a data point to each cluster j ($\mu_{i,j}^m$) is the weighting scheme. The FCM objective function is

$$\min F = \sum_{i=1}^{n} \mu_{i,j}^{m} d^2\left(\vec{x}_i, \vec{v}_j\right)$$

wherein m is the fuzzifier ($m > 1$), $\vec{x}_i$ is the data vector for each bank (of size $1 \times k$, where k are the number of input features), $\vec{v}_j$ is the vector of cluster centres ($1 \times k$), and d is the dissimilarity measure. The total number of clusters J and m are predetermined. Regarding J, researchers may test the quality of different partitions using alternative valuation criteria, such as the silhouette width (Rousseeuw, 1987) or the Caliński–Harabaz index (Caliński and Harabasz, 1974). Concerning the choice of m, Bezdek *et al*. (1984) state that algorithms tend to perform well for fuzzifiers between 1.5 and 2.5.

Despite being subject to various revisions, Bezdek et al.'s original FCM algorithm is still widely used in a variety of fields, including in banking, where it has been applied to bankruptcy forecasting (Alam *et al*., 2000; de Andrés *et al*., 2011; Martin *et al*., 2011), branch efficiency (Azadeh *et al*., 2010), credit card issuance (Hsu, 2000), retail churn prediction (Popović and Bašić, 2009), credit scoring (Michalopoulos *et al*., 2002), and currency crisis prediction (Marghescu *et al*., 2010).

5.2 | Self-Organizing Maps

Another method that has been used in the literature to deal with high-dimensional fuzzy data is SOMs. Introduced by Kohonen (1997), a SOM is a form of artificial neural network that reduces dimensionality by projecting high-dimensional data (input layer) onto a two-dimensional space (output layer or lattice), using the concept of neurons (i.e. clusters). Each neuron is differentiated from the remaining neurons by a vector of weights attributed to the input variables (codebook vector). This vector is the result of the algorithm's training process (see Appendix A1). Briefly put, the training process consists of identifying, sequentially and for each data point, the neuron that is closest to a given point (winning neuron), based on the weights vector. Each assignment leads the neuron to update its codebook vector, as well as the vector of neighbour neurons (although to a lesser degree).

Importantly, the information contained in the vector of each neuron can be used to build effective visualizing tools. In SOMs, each data point is assigned to a single neuron, which may be seen as contrary to fuzzy logic. In order to circumvent this issue, we use the silhouette width (SW) to compare the distance between each data point, the remaining data points in the assigned neuron, and the data points assigned to the closest neighbour neuron. In our paper, the SW is one of the measures used to account for the fuzziness of the clustering output, in the sense that we expect banks that are close to the data points assigned to other business models (i.e. with low SW) to record some affinity with more than one business model (i.e. high fuzziness). Finally, in addition to the number of 'neurons' (or clusters), the SOM algorithm also requires the predefinition of a number of other parameters—such as the shape of the lattice, the distance function, the number of times the algorithm is rerun, the radius and the learning rate—which are often defined after experimentation with alternative specifications (Budayan *et al*., 2011; Curry *et al*., 2001).

The seminal applications of SOMs in finance-related literature cover a variety of topics, including bankruptcy prediction (Back *et al*., 1995) and financial diagnosis (Deboeck, 1998; Kiviluoto and Bergius, 1998; Serrano-Cinca, 1996). More recent applications include, for instance, the use of SOMs as a tool for macroprudential supervision (Sarlin, 2016) and the prediction of currency crises (Sarlin and Marghescu, 2011).

5.3 | Partitioning around Medoids

The PAM, or *k*-medoids, method is an iterative, partitional algorithm that groups data into a predetermined number of clusters *k* by finding a representative data point or medoid and assigning data points to the nearest (or least dissimilar) medoid (Kaufman and Rousseeuw, 1990). Compared with the *k*-means algorithm (MacQueen, 1967), PAM uses an actual data point (medoid) as the cluster centre, rather than the cluster mean (centroid).

This method has been deemed more adequate for some data structures, such as fuzzy data. Also, it allows us to document the representative (medoid) banks of each business model. The study of banking-related topics with the aid of PAM has been scarce and mostly oriented towards client segmentation (Aryuni *et al*., 2018; Liu *et al*., 2010). An exception to this has been the study by Hryckiewicz and Kozłowski (2017), who applied PAM on a sample of large banks operating in 65 countries. Their findings suggested that, during the financial crisis, banks that adopted an investment model contributed to the accumulation of systemic risk due to their reliance on wholesale funding.

5.4 | Clustering Ensembles

A recent stream of research in classification literature 'combines the information provided by the partitions' of different clustering methods (Jain, 2010: 660). As an illustration, if we compare a clustering result with a medical diagnosis, the clustering ensemble approach equates to combining the diagnoses performed by a variety of experts (clustering methods), based on a given consensus scheme, into a single medical diagnosis (ensemble), under the expectation that the robustness/accuracy of the ensemble diagnosis is improved as a result. Kuncheva (2004) referred to three consensus schemes: unanimity, simple majority, and plurality.[3] When using the unanimity scheme, a given observation is only assigned to a cluster if all the methods in the ensemble produce the same, unanimous classification; in the simple majority scheme, the observation is only classified in a cluster if the majority of methods assign the same classification; and, finally, the plurality scheme states that an observation is assigned to the cluster that receives the largest share of classifications among the methods in the ensemble. A number of methods have been used to handle ties (e.g. Ravikumar and Ravi, 2006). An established, and intuitive, result in the literature is that the performance of clustering ensembles is expected to improve with the diversity of clustering methods (Kuncheva, 2004).

The application of the clustering ensemble approach to banking has gained increasing attention over recent years, with particular focus on bankruptcy prediction (Alam *et al*., 2000; Ravikumar and Ravi, 2006; De Andrés *et al*., 2011; Davalos *et al*., 2014) and credit scoring (Al'araj and Abbod, 2016; Abellán and Castellano, 2017). For instance, Alam *et al*. (2000) combined FCM, SOMs, and a competitive neural network to form an ordinal ranking regarding the likelihood of bank failure. Budayan *et al*. (2011), on the other hand, studied the presence of hybrid strategic groups in Turkish construction firms by performing FCM, using cluster membership values as inputs to *k*-means, and visually representing the results using SOMs. The authors found three pure strategic groups and two hybrid groups. Our study relates to this strand of research, in the sense that we cumulatively apply PAM, FCM, and SOMs to a single setting. However, our method to identify banking business models combines the actual classification of the three methods, which relates more closely to the 'ensemble' approach than the paper by Budayan *et al*. (2011). Moreover, we also provide two additional layers of classification: the identification of core (and non-core) banks, and persistent (and non-persistent) banks. Finally, we provide evidence of the robustness of our approach across a variety of checks.

6 | DATA

6.1 | Sample Selection

Our sample includes 524 European banks, both listed and non-listed, from 2005 to 2016. We collect year-end consolidated data from Bankscope and Orbis Bank Focus. The following criteria are applied:

- headquarters in EU-28 countries;
- total assets greater than €5 billion in at least one year during the period 2005–2016;
- specialization—commercial, savings, cooperative, real estate and mortgage, investment, specialized governmental credit institution or bank holdings and holding companies;
- International Financial Reporting Standards or local generally accepted accounting principles accounting standards;
- both customer deposits and gross loans to customers greater than 5% of total assets;
- data available for at least three consecutive years.

Although we apply the correct consolidation code filter according to the Bankscope/Orbis standards, in some cases it is still possible to find both the bank entity and the holding company entity of the same group (e.g. HSBC Holdings Plc and HSBC Bank Plc). In such cases, we keep the bank entity and remove the holding company because we are mainly interested in studying the banking business. Regarding cooperative networks, we opt to remove cooperative entities operating in the lowest tier because, in most cases, the autonomy of these entities to set long-term strategic choices is reduced. This means that we only exclude the local cooperatives whenever groups have more

[3]Onan (2019) provides an overview of recently developed consensus schemes, such as homogeneous and heterogeneous consensus clustering-based undersampling schemes.

than one tier. For instance, the CAG consists of a three-tier cooperative network that includes the top tier cooperative, regional banks, and local banks; for this study we only keep CAG's entities belonging to the top-tier and regional banks. We winsorize the variables at the 1% and 99% percentiles.

6.2 | Business Model Variables

In this section, we identify 10 variables that have been used in extant studies to identify banking business models. All variables are taken from the financial statements, as these are well covered in the data set. The definition of each variable is presented in Table 1.

6.2.1 | Balance Sheet Structure

The ratio of *gross loans to customers* to total assets measures the bank's level of engagement in traditional 'originate to hold' lending activities, in line with the notion of banks as delegated monitors (Diamond, 1984). The ratio of *trading assets* to total assets captures the allocation of resources to financial assets at fair value. Banks engaged in such activities are typically investment banks; however, such activities may also be evidence of portfolio diversification strategies or search for yield. The ratio of *interbank lending* to total assets, on the other hand, reflects the involvement of banks in the creation of short-term liquidity. Though such operations constitute a key component of market liquidity for banking institutions, evidence shows that they may be a significant source of counterparty and guarantee risks (Gorton and Metrick, 2012). The ratio of *customer deposits* to total assets reflects the dependence of banks on the most traditional source of funding, also typically considered as the most stable source of funding due to the presence of deposit guarantee schemes (Diamond and Dybvig, 1983; Rajan, 1992). The ratio of *interbank borrowing* to total assets includes mainly bank deposits and other money market funds that have been documented as more fragile to negative shocks via refunding risk (Taylor and Williams, 2009). On the other hand, such funds may reflect the presence of internal capital markets; that is, the borrower–lender relations of firms belonging to the same group. Under this notion, subsidiary banks are likely to face different incentives than those faced by standalone banks (de Haas and Van Lelyveld, 2010). The ratio of *wholesale funding* to total assets reflects the dependence of banks on market funding. This type of funding has become increasingly used by banks; for instance, due to Basel rules on bail-in-able debt. However, a significant share of this type of funding is expected to be marked-to-market (e.g. trading liabilities), which may induce balance sheet volatility and riskiness.

6.2.2 | Diversification

The ratio of *derivative instruments* to total assets includes both trading and standard interest-rate hedging derivatives. Given the level of expertise required to deal with certain complex derivative instruments, these are expected to absorb a significant share of human and technological resources (Blundell-Wignall *et al*., 2014). The

TABLE 1 Variables description

Variable	Description
Balance sheet structure	
Gross loans to customers	Gross loans and advances to customers
Trading assets	Financial assets trading and at fair value through profit or loss
Interbank lending	Sum of (i) net loans and advances to banks, (ii) reverse repos, securities borrowed and cash collateral
Customer deposits	Customer deposits
Interbank borrowing	Sum of (i) bank deposits, (ii) repurchase agreements, securities loaned and cash collateral
Wholesale funding	Sum of (i) other deposits, (ii) short-term funding and debt securities (maturity <1 year), (iii) long-term borrowings and debt securities at historical cost, (iv) subordinated liabilities, (iv) other long-term borrowing
Diversification	
Total derivatives	Derivative financial instruments, asset and liability side
Income diversification	Following Elsas *et al*. (2010), income diversification is computed as an HHI. In banking, total operating income (TOR) includes net interest income (NII), net fees and commissions (NFC), net trading income (NTI), and other income (OTH). We use the absolute values of each component: $[1 - [(NII/TOR)^2 + (NFC/TOR)^2 + (NTI/TOR)^2 + (OTH/TOR)^2]]$
Size	
Total assets	Log of total assets in thousand euros
Leverage	
Total equity	Total equity

Notes: All variables computed as percentage of total assets, except income diversification (HHI) and total assets (log). Data are obtained from the Bankscope and Orbis databases.

Herfindhal–Hirshman *income diversification* reflects the bank's ability to diversify into fee-based financial services, such as bancassurance, investment advice, and credit card services (Elsas *et al*., 2010), which enable it to improve the screening and monitoring of customers due to access to additional information, as well as to diversify risks (Diamond, 1984).

6.2.3 | Size

The value of *total assets* may be an important indicator of banking business models, in the sense that different banking activities seem to bear different potentials for economies of scale (DeYoung, 2010). In particular, the main intuition is that hard-information-based activities, such as trading, wholesale funding, and wholesale lending, are more prone to economies of scale than are soft-information-based activities, such as relationship lending, because hard-information activities are standardizable and require investments in specialized technologies and human resources and hence tend to be performed by larger banks (Hunter and Timme, 1986). Soft-information activities, on the other hand, tend to be performed less effectively in large organizations; for instance, owing to the presence of multiple layers of hierarchy that impede the effective communication of soft information from subordinates to superiors (Liberti and Mian, 2008).

6.2.4 | Leverage

The ratio of accounting *equity* to total assets is also expected to vary with the choice of other business model variables, for a variety of reasons. For instance, large banks seem to benefit from too-big-to-fail state subsidies, which are likely to offset the excess risk premium of operating with lower than optimal equity (O'Hara and Shaw, 1990). Also, small regional banks are likely to face constraints in terms of asset growth and access to new sources of equity, which may yield a suboptimal level of leverage. Finally, large diversified banks may be tempted to offset agency issues by offering relatively generous buybacks and dividends to shareholders (Easterbrook, 1984), hence resulting in higher bank leverage.

6.3 | Descriptive Statistics and Principal Component Analysis

The descriptive statistics presented in Table 2 are based on full-period average values, which allows us to account for the long-term nature of business model choices. The indicators show that asset allocation is mostly directed towards gross loans to customers (56.6%), and funding is mainly obtained via customer deposits (52.0%). This suggests that, on average, European banks are oriented towards traditional retail intermediation. Also, some variables display a substantially larger mean than median values (trading assets: mean 3.5% and median 0.8%; total derivatives: mean 5.1% and median 1.3%). This indicates that the distributions have significant right-sided skewness. In other words, a small share of observations has large values and a large share of observations has low values. In general, this seems to support the notion that though traditional retail intermediation prevails in European banking, significant heterogeneity may be observed across banks.

Table 3 presents the principal component analysis results. Literature has pointed out several advantages regarding the use of the retained principal components in clustering vis-à-vis using the original variables. First, using the retained components allows us to perform clustering on a space with orthogonal dimensions, given that they are uncorrelated, which is a desirable feature when using the Euclidean distance to compute dissimilarities (Sharma, 1996). Second, such an approach narrows the focus of the analysis on the most relevant relationships between business model choices and, thus, mitigates the problem of data noisiness in strategy variables (Galbraith and Schendel, 1983). Related to this, we interpret each retained component as a strategic dimension along which banks assume a long-term position relative to their peers. On the other hand, the main disadvantage of using the retained components is loss of information. However, we argue that the aforementioned benefits of using the components outweigh the issue of loss of information, particularly if the retained components explain a significant share of the variation of the original variables. For this reason, we retain a number of components that ensures that the total variation explained is greater than 80%. We also perform a varimax rotation in order to increase the interpretability of the components. Table 3 is divided into five groups of columns, each indicating a different sample period. As discussed Section 7, the full-period sample (2005–2016) is used to classify the long-term business model of banks, whereas the triennium samples (T1: 2005–2007; T2: 2008–2010; T3: 2011–2013; T4: 2014–2016) are used to assess the persistency of business models. In order to ensure comparability, we retain the same number of components for all sample periods, which allows us to cover close to 80% of the variation explained in all trienniums (T1: 81.2%; T2: 80.9%; T3: 79.3%; 77.3%).

In general, the retained components are the same across sample periods; however, the relevance of some components (and hence their order) has shifted over time. Taking the components derived from the full-period mean values as the reference (i.e. column (1)), the first component is loaded positively by trading assets, derivatives, and income diversification—and, thus, may be interpreted as a business orientation towards 'diversification' (div). The second component is loaded positively by gross loans to customers and negatively by interbank lending. Hence, we interpret the second component as an orientation towards 'traditional lending' (tlen). The third component is loaded positively by customer deposits and negatively by wholesale funding, which indicates an orientation that is focused on 'stable funding' (sfun). The fourth component is loaded positively by equity and negatively by total assets, suggesting a 'solvency'-oriented policy (sol). Finally, the fifth component is loaded positively by interbank borrowing and negatively by customer deposits, which indicates an orientation towards 'interbank funding'. In comparison with extant studies,

TABLE 2 Descriptive statistics

	Mean	SD	Min.	Median	Max.
Gross loans to customers	56.6	20.9	7.4	60.9	95.9
Trading assets	3.5	6.1	0.0	0.8	39.6
Interbank lending	15.9	16.0	0.2	10.5	79.7
Customer deposits	52.0	22.8	6.2	55.0	92.0
Interbank borrowing	17.9	14.8	0.0	14.1	72.9
Wholesale funding	13.2	14.6	0.0	8.8	65.9
Total derivatives	5.1	9.5	0.0	1.3	56.4
Income diversification	47.5	12.1	10.5	50.4	68.5
Total assets	7.3	0.6	6.1	7.1	9.0
Total equity	7.1	4.2	0.9	6.4	28.6

Sample based on full-period average for each bank (524 observations). Variables winsorized at 1st and 99th percentiles. All variables computed as percentage of total assets, except income diversification (HHI) and total assets (log).

the top three components are in line with those found in literature. For instance, van Ewijk and Arnold (2014) retained factors related with traditional funding and traditional lending; Mergaerts and Vander Vennet (2016) retained factors associated with retail orientation and diversification; and de Haan and Kakes (2019) labelled the retained factors as 'big investment banks' and 'retail banks'. Regarding the bottom two components (solvency and interbank funding), a possible explanation for their novelty (in relation to comparable studies) is likely related with the higher cut-off value of cumulative proportion of variation explained used in our study. For instance, de Haan and Kakes (2019) reported a cumulative proportion of variation explained of 69.7%, which compares with 80.5% for our study.

When we analyse each sub-period separately, it is possible to identify a significant shift in the order of two components that takes place in the 2011–2016 period. Namely, the 'diversification' component significantly loses relevance (i.e. the variation explained drops from 28% in 2005–2007 to 10% in 2014–2016), whereas as the 'stable funding' component gains relevance (i.e. the variation explained increases from 8% in 2005–2007 to 25% in 2014–2016). In our view, such a result is informative and in line with the recent events that took place in the banking sector. In particular, we interpret this shift as indicating that funding-related choices have become more important sources of strategic variation among European banks in recent years. This result has been documented in extant literature (e.g. Roengpitya *et al*., 2017) and is likely related to the implementation of the funding-related Basel III requirements, specifically the net stable funding ratio and the liquidity coverage ratio (Basel Committee on Banking Supervision, 2011).

7 | METHODOLOGY

The method used to identify banking business models may, in general, be summarized in the following way: first, we perform clustering analysis with alternative algorithms based on the retained components to identify the optimal number of business models, and combine the classification outputs of each algorithm into one single assignment, using a majority consensus rule (clustering ensemble); then, we apply a set of criteria to identify core banks, using a stricter consensus rule (unanimity) and information regarding the quality of clustering (silhouette width); finally, we identify as persistent banks those that hold the same business model classification for all the sample periods. In the following sub-sections we detail the methodological decisions made in each step.

7.1 | Classification of Banking Business Models (Clustering Ensemble)

In order to classify and describe the business models we apply the following procedure:

1 Using the retained components as inputs (full period mean values), run the three clustering methods: PAM, FCM, and SOMs (each algorithm is presented in Appendix A1). For parsimony, we run the algorithms for a range of three to nine clusters. Several decisions are made at this stage:

- distance measure (PAM, FCM, SOM)—Euclidean distance;
- fuzzifier (FCM)—2 (following the default value in ppclust R package[4]);
- grid size (SOM)—we adapt the grid configuration to enable us to represent the range of three to nine clusters J. Namely, for $J = 3$: 3×1; for $J = 4$: 2×2; for $J = 5$: 5×1; for $J = 6$: 3×2; for $J = 7$: 7×1; for $J = 8$: 4×2; for $J = 9$: 3×3;
- neighbourhood radius/type and topology (SOM)—radius 0.5 (minimum) and 1.0 (maximum), Gaussian function type and rectangular topology (defined after experimentation with alternative specifications);
- type of learning algorithm and initialization (SOM)—batch with linear initialization; that is, the algorithm defines the initial data point weights matrix by using 'the linear grids upon the first

[4]The list of all R packages used in the paper is given in Appendix A2.

TABLE 3 Principal component analysis

	(1) Full period, 2005–2016					(2) T1, 2005–2007					(3) T2, 2008–2010					(4) T3, 2011–2013					(5) T4, 2014–2016				
	PC1	PC2	PC3	PC4	PC5	PC1	PC2	PC3	PC4	PC5	PC1	PC2	PC3	PC4	PC5	PC1	PC2	PC3	PC4	PC5	PC1	PC2	PC3	PC4	PC5
Variable	div	tlen	sfun	sol	ifun	div	tlen	ifun	sol	sfun	div	tlen	sfun	ifun	sol	sfun	tlen	div	sol	ifun	sfun	tlen	ifun	sol	div
Rotated factor loadings																									
Gross loans to customers	−0.34	**0.88**				−0.32	**0.87**				−0.33	**0.88**					**0.91**	−0.30				**0.90**			−0.36
Trading assets	**0.78**	−0.15				**0.84**					**0.80**					−0.12	−0.20	**0.75**			−0.15	−0.22	−0.21		**0.73**
Interbank lending	−0.12	**−0.89**	0.11		0.17	−0.17	**−0.86**	0.21		0.16	−0.14	**−0.90**	0.12	0.17		0.11	**−0.92**	−0.13		0.12		**−0.92**	0.19		−0.14
Customer deposits	−0.28		**0.74**		**−0.58**	−0.26		**−0.54**		**0.77**	−0.26		**0.73**	**−0.60**		**0.69**		−0.26		**−0.64**	**0.81**		**−0.53**		−0.24
Interbank borrowing		−0.14			**0.97**		−0.14	**0.98**	−0.11			−0.11		**0.98**						**0.98**		−0.12	**0.95**		
Wholesale funding		0.13	**−0.87**	−0.17			0.14		−0.13	**−0.90**			**−0.92**	−0.11	−0.13	**−0.92**	0.11		−0.14		**−0.90**	0.10		−0.14	−0.13
Total derivatives	**0.71**	−0.14	−0.35	−0.24		**0.71**			−0.18	−0.23	**0.74**		−0.25		−0.28	−0.32	−0.16	**0.67**	−0.38		−0.39	−0.15		−0.46	**0.59**
Income diversification	**0.69**		0.38		0.15	**0.69**		0.23	0.15	0.17	**0.67**		0.22			0.22	0.12	**0.71**			0.23		−0.16		**0.72**
Total assets	0.47	0.11	−0.38	**−0.55**		**0.55**			−0.41	−0.45	0.45		−0.45		**−0.52**	−0.34		0.40	**−0.64**		−0.27	0.11		**−0.70**	0.35
Total equity				**0.94**					**0.99**						**0.95**		0.11	0.14	**0.88**					**0.83**	0.21
Variation explained																									
Sum of squared loadings	1.64	1.36	1.23	1.05	0.93	1.68	1.41	1.19	1.02	0.90	1.66	1.37	1.23	1.07	0.90	1.65	1.29	1.23	1.05	0.96	1.59	1.27	1.20	1.08	1.00
Variation explained (VE)	27.0	18.4	15.2	11.1	8.7	28.4	20.0	14.3	10.4	8.2	27.4	18.8	15.2	11.5	8.0	27.3	16.8	15.2	11.0	9.1	25.2	16.3	14.5	11.7	9.9
Cumulative VE	27.0	45.4	60.7	71.8	80.5	28.4	48.4	62.6	73.1	81.2	27.4	46.2	61.4	72.9	80.9	27.3	44.1	59.2	70.2	79.3	25.2	41.5	56.0	67.6	77.5
N	524					376					441					495					483				

The results in (1) are obtained using the full-period average of each input variable for all banks (n = 524). For the remaining results (2–5), we compute the triennium average value of the input variables observed by the banks present in the sample in at least one year of each triennium. Hence, the last line of the table represents the number of banks present in the sample in each triennium. Component loadings rotated using varimax rotation. Variables with higher than 0.5 loadings per component (absolute value) are in bold. Input data standardized. Labels of the principal components: div, diversification; tlen, traditional lending; sfun, stable funding; sol, solvency; ifun, interbank funding.

two principal components direction' (Chair and Charrad, 2017: 2), which allows for a deterministic, and hence reproducible, clustering output.

In FCM, each observation is assigned to the cluster for which it has the highest coefficient of membership. For each bank, collect the silhouette width SW using PAM, FCM, and SOMs and the percentage of cluster membership (PCM) using FCM.

2 Select the optimal number of clusters for each method by examining four criteria: average SW, Caliński–Harabasz index, Davies–Bouldin index, and Dunn index (see the description of each criterion in Appendix A3). In particular, for each method we rank the results obtained for each partition (J = 3 to J = 9) and count the number of times each partition is ranked as the best (#1) or second best (#2) value in each criterion. The partition with the highest count of #1 and #2 is labelled as the optimal number of clusters.
3 Combine the clustering assignment of each method into one single assignment (clustering ensemble step). To do so, compute the 'voting results' for each bank; that is, the count of classifications (1, 2, or 3) of a given bank in each business model and apply a majority consensus rule. That is, a bank is assigned to the business model for which the count of classifications is higher. For example, a bank may be classified as operating with business model BM1 by two methods (BM1 count = 2) and BM3 by the other method (BM3 count = 1). In this case, the bank is assigned to business model BM1 (as 2 > 1). When there is a tie (i.e. each method assigns the bank to a different business model), we follow the assignment made by the method with the highest silhouette width for that specific bank.
4 Assess the similarity of business model classifications between the ensemble classification and each of the alternative clustering methods (PAM, FCM, SOM) by computing the cross-tabulation of clustering results. In order to match the classes of each clustering method, we analyse the clusterwise mean values of each clustering output. The similarity of classifications is assessed using Simple Matching (SM), the Rand index (RI), the adjusted Rand index (ARI), and the Jaccard index (JI). The first two measures allow us to have an intuitive description of classification similarity; the third measure corrects the original RI for randomness; and the JI only considers as similar classifications those that are 'true positives', whereas the RI also considers true negatives. The SM method is the only method that directly uses the elements of the cross-tabulation and is computed as the sum of similar business model classification (elements in the diagonal) divided by the total number of elements. The remaining indices (RI, ARI, and JI) are computed based on pairs of elements. For brevity reasons, we refer the specification of these indices to the work by Milligan and Cooper (1986)—which is also used as the main reference paper for the R package used in this analysis ('clues'). We also run Pearson's chi-square independence test.
5 Finally, describe the composition of each business model by computing the mean and standard deviation of business model variables for each cluster (business model). Test for differences in the mean values between pairs of business models using the Tukey honestly significant difference (HSD) test. For each variable, identify the cluster with the highest and lowest values. We label as a 'cluster distinguishing feature' those variables for which a cluster records the highest or lowest value and its mean is statistically significant from one or more extant clusters.

7.2 | Identification of Core Banks

The identification of 'core' (and 'non-core') banks is made in the following way:

1 Use the clustering ensemble obtained for the cross-section sample (see Section 7.1, step 5) and compute the silhouette width for each bank.
2 Label as 'core' the banks that cumulatively meet the following criteria:

- **Criterion 1.** Unanimity scheme; that is, the bank is classified in the same business model using the three clustering methods (three out of three). The choice of the unanimity scheme, which is more restrictive than the alternative schemes, allows us to obtain a lower propensity for type I error, although at the expense of a higher likelihood of type II error. We take this decision given the focus of this step in the identification of those banks that clearly belong to a given business model. In other words, we decide to adopt such voting scheme because we place special importance in increasing the level of confidence regarding the accuracy of the classification attributed to banks labelled as core.
- **Criterion 2.** The bank records a silhouette width (clustering ensemble) above 0.2. This limit is based on the threshold values of average silhouette width as proposed by Kaufman and Rousseeuw (1990).

3 Check the fuzziness of core and non-core banks by comparing the following metrics:

- average values of the first and second-best cluster memberships (PCM1 and PCM2), wherein PCM1 and PCM2 correspond to the top two membership scores obtained via FCM for each bank;
- difference between PCM1 and PCM2;
- Herfindahl–Hirschman index (HHI) of cluster memberships, computed as the sum of squared PCMs (i.e. $PCM1^2+\cdots+PCMJ^2$) for each bank;
- average silhouette width based on the ensemble classification for each bank.
  For each metric, we compute the Tukey HSD test for comparison of means between the two subsamples: core and non-core banks.

7.3 | Identification of Persistent Banks

The following procedure is used to separate 'persistent' from 'non-persistent' banks:

1 Compute the clustering ensemble for each 'triennium sample'. Namely, we perform principal component analysis and clustering with alternative methods for each triennium sample: 2005–2007 (T1), 2008–2010 (T2), 2011–2013 (T3), and 2014–2016 (T4).
2 Assess the general persistency of business model classification for each method over consecutive trienniums. In order to do this, for each method, organize the classification results so that each row represents a bank-specific pair of classifications obtained in consecutive trienniums. To illustrate this, if a given bank (e.g. A) is present in our data set in four trienniums (T1 to T4), we compute the clustering method (e.g. FCM) for each triennium separately and transform the results into three pairs of consecutive business model: 'A.FCM.T1-T2', 'A.FCM.T2-T3', 'A.FCM.T3-T4'. We compute the SM, RI, ARI, JI (Milligan and Cooper, 1986), and chi-square independence test for each method.
3 Label as a 'persistent bank' those banks for which the business model classification (ensemble) is the same across all trienniums.
4 For the subset of 'non-persistent banks', investigate the number of changes per bank. This step allows us to understand whether non-persistency derives from business policy changes (if the majority of non-persistent banks change their business model once) or, on the contrary, from the inability of our approach to capture the business models in a consistent fashion over time. Such analysis may be seen as a time-varying adaptation of the concept of 'fustrated clusterings' proposed by Gates *et al.* (2019), which refers to observations that the method 'cannot consistently decide on a grouping' (p. 8).
5 Identify the distinctive features of non-persistent banks (relative to persistent ones) by comparing banks that changed their business model in a given triennium (t+1) with other banks that held the same business model in the triennium prior to the change (t) and did not change their business model in t+1, with respect to the features exhibited by both banks in triennium t. To undergo this analysis, we run Bayesian logistic regressions (J regressions; i.e. one per business model prior to change) using a sample of bank–triennium observations. The model specification is as follows:

$$\text{Prob}\left(Y^j_{i,t+1}=1\right)=1-\Lambda(\alpha+\beta\mathbf{X}_{i,t}+\gamma\text{BM}_{i,t+1}+\tau\text{Triennium}_t+\delta\text{Fuzziness}_{i,t}+\varepsilon_{i,t})$$

In which $Y^j_{i,t+1}$ is a dummy which takes on the value 1 if bank i changes its business model from j in triennium t to another business model ($-j$) in triennium t+1 and assumes the value 0 if the bank remains in business model j, wherein t = 2005–2007, 2008–2010, 2011–2013, and 2014–2016; α is the model constant; $\mathbf{X}_{i,t}$ is the vector of business model variables observed in triennium t for each bank i; $\text{BM}_{i,t+1}$ is the destination business model of bank i, which for banks that change business model will be $-j$ and for the persistent banks will be j. Notice that the aim is to capture forward looking fixed effects that are specific to the business model change direction. For instance, a bank that changes from a retail diversified funding model to a retail focused model is expected to have previously invested in technology to capture customers deposits (Berger *et al.*, 2005); Triennium_t is a dummy for the triennium and is used to capture period specific fixed effects (e.g. new regulation or a financial crisis); Fuzziness_t is 1 minus the difference between the top two percentages of business model membership in t for each bank i; that is, 1 − (PCM1 − PCM2); β, γ, τ, and δ are the regression coefficients; and $\varepsilon_{i,t}$ is the disturbance term.

8 | RESULTS AND DISCUSSION

8.1 | Banking Business Models

Table 4 displays the results of the selection criteria for each partition and method. In general, the results seem consistent across clustering methods. For instance, for a partition of four clusters (J = 4), the three methods record high mean values of similarity of banks within the assigned business model vis-à-vis those assigned to other business models, as given by the highest value of average silhouette width SW for PAM (0.23) and the second highest for FCM (0.18) and SOM (0.19). Similarly, the J = 4 partition records the highest ratio of between to within cluster dispersion (Calińksi–Harabazs index) for PAM (137.57) and the second highest for SOM and FCM (122.01 and 128.82 respectively). Moreover, we may observe J = 4 is identified as the partition with the highest count of #1 ranked criteria for PAM (3) and the highest count of #2 ranked criteria for FCM (3) and SOM (2). Based on these results, we conclude that the optimal number of clusters is J = 4. In other words, our findings suggest that the European banking sector may be characterized by the presence of four distinct banking business models. This result is in line with the number of banking business models identified by previous studies (Ayadi and De Groen, 2015; Martín-Oliver *et al.*, 2017; Roengpitya *et al.*, 2017). Also, we note that, in line with the results of Mergaerts and Vander Vennet (2016), the values of average silhouette width are below the threshold of 0.25 for minimum quality of clustering as proposed by Kaufman and Rousseeuw (1990). In our framework, such findings support the notion that some banks may be combining the 'recipes' of different business models (non-core banks) and/or have changed their business model in the sample period (non-persistent)—two analyses that are conducted in subsequent sections.

Next, Table 5 describes the composition and popularity of each banking business model, using the clustering ensemble approach.[5] The table shows the mean values (and standard deviation) of all business model variables, as well as the results of the test for mean differences across pairs of business model. In general, results show that, when a business model records the highest or lowest mean value in comparison with the other models, the number of significant pairwise

[5]A similar table with classifications obtained for each method (separately) may be found in Appendix A4.

TABLE 4 Selection criteria: number of business models

	ASW	SW > 0.5	SW < 0	CHI	DBI	DI	Count of #1 rank	Count of #2 rank
FCM								
J = 3	**0.23**	**8.21**	18.70	**133.89**	1.65	0.039	3	0
J = 4	**0.18**	0.00	**18.32**	**128.82**	1.59	**0.049**	1	3
J = 5	0.15	0.00	**17.94**	119.25	1.92	0.036	1	0
J = 6	0.14	0.00	20.99	108.96	1.63	0.021	0	0
J = 7	0.13	0.00	23.47	105.95	**1.40**	0.036	0	1
J = 8	0.10	0.00	27.29	96.20	**1.36**	0.025	1	0
J = 9	0.09	0.00	30.53	87.21	2.09	**0.040**	0	1
SOM								
J = 3	0.11	0.00	20.61	110.96	1.68	**0.041**	0	1
J = 4	**0.19**	**1.15**	19.27	**122.01**	1.51	0.032	1	2
J = 5	**0.21**	0.00	**13.36**	**127.57**	**1.14**	0.039	3	1
J = 6	0.11	0.00	27.67	86.66	1.46	0.019	0	0
J = 7	0.14	0.00	**17.56**	98.67	**1.05**	**0.045**	2	1
J = 8	0.11	0.00	26.34	97.20	1.18	0.024	0	0
J = 9	0.07	0.00	35.31	74.05	1.33	0.036	0	0
PAM								
J = 3	0.15	0.00	**9.54**	114.21	1.47	0.042	1	0
J = 4	**0.23**	**1.53**	10.50	**137.57**	1.44	0.036	3	0
J = 5	0.21	0.00	9.92	126.00	1.27	0.036	0	0
J = 6	0.19	**0.38**	12.60	123.25	1.10	**0.044**	0	2
J = 7	**0.21**	**0.38**	**9.54**	**129.94**	**1.09**	**0.049**	2	4
J = 8	0.20	0.00	14.31	123.44	**0.96**	0.042	1	0
J = 9	0.20	0.00	14.69	117.15	1.23	0.042	0	0

Notes: Results of running the PAM, SOM, and FCM algorithms on the full-period average sample, with inputs PC1 to PC5 for different number of clusters *J*. Selection criteria: average SW (ASW), percentage of observations with SW > 0.5 and SW < 0, Caliński–Harabasz index (CHI), Davies–Bouldin index (DBI), and Dunn index (DI). The partitions with the top values (#1 and #2) for each criterion are presented in bold. Note that the best partitions minimize SW < 0 and DBI and maximize the remaining criteria (see Appendix A3).

TABLE 5 Composition of business models: clustering ensemble

	BM1	BM2	BM3	BM4
Number of banks	203	124	109	88
Gross loans to customers	68.3 (12.6)$^{++}$	67.6 (14.1)$^{++}$	35.6 (16.0)$^{++}$	40.1 (18.2)$^{++}$
Trading assets	1.8 (3.4)$^{+}$	1.9 (2.5)$^{+}$	2.0 (4.9)$^{+}$	**11.4 (8.9)**$^{+++}$
Interbank lending	8.2 (5.4)$^{++}$	8.9 (6.3)$^{++}$	**37.6 (19.1)**$^{+++}$	16.7 (12.7)$^{+++}$
Customer deposits	**67.3 (13.4)**$^{+++}$	37.5 (16.2)$^{+++}$	58.5 (23.0)$^{+++}$	**29.3 (15.9)**$^{+++}$
Interbank borrowing	**11.5 (8.6)**$^{+++}$	21.5 (16.6)$^{+}$	24.5 (19.7)$^{+}$	19.4 (10.3)$^{+}$
Wholesale funding	7.2 (6.5)$^{++}$	**25.5 (17.0)**$^{+++}$	4.5 (8.8)$^{++}$	20.5 (15.4)$^{+++}$
Total derivatives	1.4 (2.1)$^{++}$	3.8 (4.0)$^{+++}$	1.0 (2.9)$^{++}$	**20.3 (14.5)**$^{+++}$
Income diversification	47.2 (11)$^{++}$	43.2 (13.0)$^{++}$	46.6 (12.5)$^{+}$	**55.1 (9.4)**$^{+++}$
Total assets	7.0 (0.3)$^{++}$	7.5 (0.5)$^{+++}$	7.0 (0.4)$^{++}$	**8.1 (0.7)**$^{+++}$
Total equity	**8.9 (4.5)**$^{+++}$	6.0 (3.3)$^{+}$	6.7 (4.3)$^{++}$	5.2 (2.9)$^{++}$

Notes: Mean values and standard deviation in parentheses, except number of banks (count). The classification is obtained using the clustering ensemble of PAM, SOM, and FCM classification output following a majority consensus rule (see Appendix A4 for detailed results per method). The input variables used in the clustering process are PC1 to PC5 for the full period, as identified in Table 3. For each variable, we compute the Tukey HSD test for comparison of means per pair of business models; that is, for a given variable, the mean value of each business model is potentially different from the mean of the remaining three business models (only two, only one, or none). The number of ($^{+}$) indicates the number of pairwise comparisons that are statistically different at the 5% level. Values in bold indicate the business models with the highest and lowest mean values for each variable, when the number of plus signs is ($^{+++}$). All variables computed as percentage of total assets, except income diversification (HHI) and total assets (log).

differences is consistently two ($^{++}$) or three ($^{+++}$). This finding indicates the ability of the ensemble to significantly differentiate between banking business models. Taking each business model separately, we observe that:

- **BM1** is followed by 203 banks (38.7% of the total number of banks), making it the most popular configuration. This model records the highest mean value of customer deposits (67.3%) and total equity (8.9%) and the lowest value of interbank borrowing (11.5%). Moreover, along with BM2, BM1 registers the highest mean value of gross loans to customers (68.3%). We label BM1 the *retail focused* model.
- **BM2** comprises 124 banks (23.7%). As a whole, banks included in BM2 exhibit the highest mean value of wholesale funding (25.5%) and second largest size (7.5, log). Additionally, *ex-aequo* with BM1, BM2 records the highest value of gross loans to customers (67.6%). We label BM2 the *retail diversified funding* model.
- **BM3** includes 109 banks (20.8%). This model records the second highest value of customer deposits (58.5%), the highest mean value of interbank lending (37.6%), and lowest loans to customers (35.6%). We label BM3 the *retail diversified assets* model.
- **BM4** is followed by 88 banks (16.8%), making it the least common model. BM4 shows the highest mean value of trading assets (11.4%), total derivatives (20.3%), income diversification (55.1, HHI), and size (8.1, log). Moreover, BM4 records the lowest mean value of customer deposits (29.3%), which makes it the least traditional and more diversified business model. We label BM4 the *large diversified* model.

In general, such distribution of banks across business models, as well as the description of each model, is in line with the results obtained by Ayadi and De Groen (2015). Next, we provide a novel visual representation of banking business models by using SOMs (Figure 1). In order to obtain a richer representation, we modify the SOMs used in the clustering ensemble by expanding the SOM grid from 2×2 to 4×4. Then, we cluster the SOM code vectors and map the four banking business models, which are identified in the top-left map of Figure 1 (labelled 'Business models'). Observing the maps of business model features, we may see that the results are consistent with those described in Table 5. In particular, the panel corresponding to BM1 (retail focused) shows high values (in yellow and green) for gross loans to customers, customer deposits, and equity, and also contains the largest number of banks. In the same line, BM2 (retail diversified funding) registers high values of gross loans to customers, wholesale funding, and size. BM3 (retail diversified assets) records low values in gross loans to customers and high values in interbank lending, interbank borrowing, and customer deposits. Lastly, BM4 (large diversified) records high values of trading assets, derivatives, income diversification, and size. Interestingly, the 'neurons' located in the four extreme corners of each map seem to adhere more closely to each 'typical' business model configuration. On the contrary, the 'neurons' located in the centre part of each map seem to be less well defined, exhibiting some features that adhere less closely to each archetypal business model. For instance, in the panel corresponding to BM1 (retail focused), the 'neuron' located closest to the centre exhibits relatively lower values than the corner 'neuron' in terms of gross loans to customer, customer deposits, and total equity, and higher values in terms of income diversification. Overall, we view these results as suggesting that banks may have different levels of adherence to their assigned banking business model, an analysis that we explore in more detail in the next section.

Finally, we perform the cross-tabulation of the business model classifications obtained in the ensemble and each method in order to assess the suitability of methods included in our ensemble. In the upper panel of Table 6, we test the classification similarity of the pair ensemble/FCM and find that 94.7% of the banks in our sample record the same business model classification in both methods (SM). Moreover, the RI of 0.943 indicates that 94.3% of the 132,026 unordered pairs of observations, $524(524 - 1)/2$, include elements that are either (i) both classified in the same business model by both methods (true positives) or (ii) both classified in different business models by both methods (true negatives)—a value that reduces slightly to 0.857 when adjusting for chance (ARI). The main sources of dissimilarity stem from 18 banks that are classified as BM2 by FCM and as BM1 by the ensemble. In the middle panel, the pair ensemble/SOM shows a similar SM score to that of FCM (93.5%), but presents higher values for other similarity measures, including ARI (0.866). In the lower panel, the pair ensemble/PAM also shows high values for the measures of similarity, despite displaying the lowest ARI (0.741). Finally, the null hypothesis of statistical independence of the classifications is rejected at a 1% level for all methods. In general, we interpret these results as supporting our approach, in the sense that (i) the similarity of classification between the ensemble and each method seems sufficiently high to give us some confidence that the classifications are not spurious; (ii) the sources of dissimilarity differ across pairs of methods, indicating the presence of diversity, which is seen as a positive attribute of ensemble compositions (Kuncheva, 2004).

8.2 | Core Banks and Fuzziness Analysis

Next, we are interested in discriminating between core and non-core banks. Table 7 identifies the number of core banks per business model. In the upper panel (criterion 1), the results indicate that 400 banks are classified in the same business model by the three methods. In the lower panel (criterion 2), the assessment of partition quality is similar across the different methods. In particular, the percentage of banks with a silhouette width above 0.2 ranges from 50.8% (SOM) to 60.3% (PAM). For the clustering ensemble, the number of banks that meet this criterion is 281 (53.6%). Also, we find that in each criterion the methods that yield the lowest number of similar classifications per business model vary considerably. For instance, for criterion 2, PAM is more

FIGURE 1 SOM of business model features. Notes: The frontier between business models was obtained by performing the clustering ensemble approach on the codebook vectors. The values presented for each variable, ranging from −1.5 to +1.5, correspond to the codebook vectors obtained by performing a batch SOM on the full list of business model variables (standardized)

restrictive for BM4 (i.e. identifies a lower number of BM4 banks), SOM is more restrictive for BM1 and BM3, and FCM is more restrictive for BM2. We interpret these results as an indication of the suitability of the choice of clustering methods due to their diversity (Kuncheva, 2004). Importantly, when imposing criteria 1 and 2 we identify a total of 273 core banks (52.1% of total banks). Also, when comparing the distribution of core banks per business model relative to the full sample, we observe significant differences across business models. Namely, BM1 banks represent a significantly higher share of the sample of core banks (62.3%) when compared with the full sample (38.7%); BM4 banks represent a similar share; and BM2 and BM3 banks represent a significantly lower share (12.1% versus 23.7% and 11.4% versus 20.8% respectively). This finding seems to suggest that a bank following a retail-

TABLE 6 Similarity of business model classifications per pair of methods

	Ensemble				
	BM1	BM2	BM3	BM4	Total
FCM					
BM1	**184 (99.5%)**	0 (0.0%)	1 (0.5%)	0 (0.0%)	185 (100%)
BM2	18 (13.4%)	**116 (86.6%)**	0 (0.0%)	0 (0.0%)	134 (100%)
BM3	1 (0.9%)	6 (5.2%)	**108 (93.9%)**	0 (0.0%)	115 (100%)
BM4	0 (0.0%)	2 (2.2%)	0 (0.0%)	**88 (97.8%)**	90 (100%)
Total	203 (38.7%)	124 (23.7%)	109 (20.8%)	88 (16.8%)	524 (100%)
SM	0.947				
RI (ARI)	0.943 (0.857)				
JI	0.811				
χ^2	1154.8***				
SOM					
BM1	**198 (99.5%)**	0 (0.0%)	1 (0.5%)	0 (0.0%)	199 (100%)
BM2	0 (0.0%)	**103 (98.1%)**	2 (1.9%)	0 (0.0%)	105 (100%)
BM3	5 (4.7%)	0 (0.0%)	**101 (95.3%)**	0 (0.0%)	106 (100%)
BM4	0 (0.0%)	21 (18.4%)	5 (4.4%)	**88 (77.2%)**	114 (100%)
Total	203 (38.7%)	124 (23.7%)	109 (20.8%)	88 (16.8%)	524 (100%)
SM	0.935				
RI (ARI)	0.947 (0.866)				
JI	0.823				
χ^2	948.9***				
PAM					
BM1	**203 (88.3%)**	23 (10.0%)	4 (1.7%)	0 (0.0%)	230 (100%)
BM2	0 (0.0%)	**101 (70.6%)**	2 (1.4%)	40 (28%)	143 (100%)
BM3	0 (0.0%)	0 (0.0%)	**103 (100%)**	0 (0.0%)	103 (100%)
BM4	0 (0.0%)	0 (0.0%)	0 (0.0%)	**48 (100%)**	48 (100%)
Total	203 (38.7%)	124 (23.7%)	109 (20.8%)	88 (16.8%)	524 (100%)
SM	0.868				
RI (ARI)	0.892 (0.741)				
JI	0.691				
χ^2	972.1***				

Notes: Count of banks per business model classification over pairs of clustering methods. The values in parentheses are the percentage of each line total. Simple matching (SM) is computed as the number of observations with the same classification divided by the total number of observations. Remaining similarity measures are described in detail in Milligan and Cooper (1986); namely, the RI is measured as the proportion of pairs of observations labelled as true positive and true negative divided by the total number of pairs of observations (Rand, 1971); the ARI corrects the original RI for randomness (Hubert & Arabie, 1985); the multi-class mean JI is defined as the proportion of pairs labelled as true positives divided by the total number of pairs of observations excluding true negatives (Jaccard, 1901). Finally, we compute Pearson's χ^2 independence test. *, **, and *** indicate the statistical significance at the 10%, 5%, and 1% levels respectively. Classifications are obtained using PC1 to PC5 as input variables.

focused (BM1) or trading (BM4) model is more likely to follow the ‘standard’ group strategy, whereas banks allocated to the retail diversified funding (BM2) or retail diversified assets (BM3) models may be more prone to operate less tightly under the peer group's typical business strategy.

Following this line of inquiry, we next compare the core and non-core banks in terms of their level of adherence (or fuzziness) to the assigned business model (Table 8). First, we compare the average values of the first and second-best percentage of cluster membership (PCM1 and PCM2, obtained via FCM) for core and non-core banks. As expected, the results show that core banks record a significantly higher value of PCM1 (0.52 versus 0.43) and a lower value of PCM2 (0.23 versus 0.26) than non-core banks do. In line with this, the difference between PCM1 and PCM2 (third line) is significantly higher for core banks (0.29) than for non-core banks (0.17). Similarly, core banks record a higher concentration of cluster memberships (0.38) and a higher silhouette width (0.34) than non-core banks do (0.32 and 0.05 respectively). Interestingly, in comparison with the only other study

TABLE 7 Core banks per business model

	Total	BM1	BM2	BM3	BM4
Criterion 1. Same BM classification across all methods ('unanimity')					
FCM = PAM	427 (81.5%)	184	93	102	**48**
SOM = FCM	468 (89.3%)	**180**	100	100	88
PAM = SOM	422 (80.5%)	198	**80**	**96**	**48**
(a) PAM = SOM = FCM	400 (76.3%)	180	77	95	48
Criterion 2. SW > 0.2 for the clustering ensemble					
FCM	276 (52.7%)	169	**36**	32	39
SOM	266 (50.8%)	**155**	40	**30**	41
PAM	316 (60.3%)	200	50	33	**33**
(b) Ensemble	281 (53.6%)	172	37	31	41
Core banks (a, b)	273 (52.1%)	170	33	31	39

Notes: Results for each criterion are computed separately from other criteria, except in the last line (a, b). For each criterion (1 and 2) and business model (BM1, BM2, BM3, BM4) we identify the most restrictive method (PAM, SOM, or FCM) in bold; that is, the method that identifies the lowest number of banks that meet the criterion.

TABLE 8 Fuzziness analysis: core versus non-core banks

	Core	Non-core	Diff.
First best cluster membership (PCM1)	0.52 (0.12)	0.43 (0.12)	0.09^{***}
Second best cluster membership (PCM2)	0.23 (0.06)	0.26 (0.05)	-0.04^{***}
PCM1 − PCM2	0.29 (0.17)	0.17 (0.16)	0.13^{***}
HHI	0.38 (0.1)	0.32 (0.08)	0.05^{***}
ASW	0.34 (0.08)	0.05 (0.11)	0.29^{***}
Number of banks	273	251	

Notes: Mean values and standard deviation in brackets, except number of banks (count). The first and second-best cluster memberships (PCM1 and PCM2) correspond to the top two membership scores obtained via FCM for each bank. In other words, for each bank we identify the business models with which the bank has the two highest membership scores and label them as PCM1 and PCM2 respectively. Note that the sum of all membership scores per bank is 1. A core bank is expected to record a higher PCM1 and a lower PCM2 than non-core banks. The 'PCM1 − PCM2' is computed as the difference between the top two membership scores for each bank. A core bank is expected to record a higher PCM1 − PCM2 than non-core banks. The HHI is computed as the sum of squared PCMs (i.e. $PCM1^2 + PCM2^2 + PCM3^2 + PCM4^2$) for each bank. A core bank is expected to record a higher HHI than non-core banks. The ASW is based on the ensemble classification for each bank and is calculated as the difference between the average distance to banks in the closest neighbour business model minus the average distance to banks in the assigned business model, divided by the maximum of the two distances. A core bank is expected to record a higher ASW than non-core banks do. For each metric, we compute the Tukey HSD test for comparison of means between the two subsamples: core and non-core banks. Results are reported in the final column. *, **, and *** indicate the statistical significance of the difference at the 10%, 5%, and 1% levels respectively.

that reports the silhouette width in the identification of banking business models, the average silhouette width of core banks is significantly higher than the one reported by Mergaerts and Vander Vennet (2016), which ranges between 0.14 to 0.20 for a three to six clusters' partition. We interpret these results as an indication that, for banks labelled as core, the assignment to a single, discrete business model is, on average, appropriate, whereas it may be more suitable to depict non-core banks as following a combination of 'recipes' of different business models. Moreover, untabulated results show that the majority of non-core banks with below mean PCM1 − PCM2 (<0.17) seem to be better depicted as combining the retail-focused model (BM1) with either the retail diversified funding (BM2: 48) or the retail diversified assets (BM3: 24) models.

8.3 | Persistent Banks

The first step in the persistency analysis is to assess the general level of persistency of business model classifications. Using the RI, Table 9 shows that 83.2% of pairs of consecutive triennium observations are classified in the same business model. Furthermore, the null hypothesis of independence of classification over consecutive trienniums is rejected for all methods. As expected, these values are relatively lower when adjusting for chance (ARI). In other words, the proportion of pairs of bank–triennium observations that cumulatively change business model may be non-negligible. In sum, whereas the general results indicate that our approach is able to capture the long-term stable nature of business models, evidence also suggests that a significant

TABLE 9 Persistency of business models in consecutive trienniums

Ensemble (*t*)	Ensemble (*t* + 1) BM1	BM2	BM3	BM4	Total
BM1	406 (**82.7%**)	53 (10.8%)	27 (5.5%)	5 (1.0%)	491 (100%)
BM2	24 (8.6%)	233 (**83.5%**)	8 (2.9%)	14 (5.0%)	279 (100%)
BM3	35 (13.6%)	7 (2.7%)	209 (**81.0%**)	7 (2.7%)	258 (100%)
BM4	4 (1.6%)	34 (14%)	12 (4.9%)	193 (**79.4%**)	243 (100%)
Total	469 (36.9%)	327 (25.7%)	256 (20.1%)	219 (17.2%)	1,271 (100%)
SM	0.819				
RI (ARI)	0.832 (0.578)				
JI	0.530				
χ^2	2,202.4***				

Notes: Count of bank–triennium observations per business model classification over pairs of consecutive trienniums. The total number of observations (1,271) corresponds to the sum of banks in each of the first three trienniums (T1: 376; T2: 441; T3: 495) minus the banks that exited the sample in the last three trienniums (T2: 6; T3: 13; T4: 22). Values presented in parentheses are a percentage of the row total. Simple matching (SM) is computed as the number of observations with the same classification divided by the total number of observations. Other similarity measures follow Milligan and Cooper (1986): the RI is measured as the proportion of pairs labelled as true positive or true negative divided by the total number of pairs of observations (Rand, 1971); the ARI corrects the original RI for randomness (Hubert & Arabie, 1985); the multi-class mean JI is defined as the proportion of pairs labelled as true positives divided by the total number of pairs of observations excluding true negatives (Jaccard, 1901). Finally, we perform Pearson's χ^2 independence test. *, **, and *** indicate the statistical significance at the 10%, 5%, and 1% levels respectively.

share of banks in our sample changed business model in the period between 2005 and 2016. Though such a finding is in line with extant literature (e.g. Martín-Oliver *et al*., 2017; Roengpitya *et al*., 2017), it fuels our quest to understand more precisely what share of banks is non-persistent, how many times non-persistent banks changed business model during the 2005–2016 period, and which features of banks potentiate such changes.

The identification of persistent banks per pair of consecutive trienniums is detailed in Table 10. In particular, in the first three lines we identify the number of banks with the same business model classification in each pair of consecutive trienniums (T1 and T2, T2 and T3, T3 and T4). This step unveils the exact evolution of persistency across pairs of trienniums. The results indicate that the persistency of business model classifications remains stable throughout the sample period (T1 and T2: 82.4%; T2 and T3: 84.6%; T3 and T4: 79.1%). In general, our findings compare favourably to the persistency levels of business model classifications reported by Martín-Oliver *et al*. (2017), which range from 10.4% (lowest) to 85.7% (highest) for a sample of Spanish banks, using a single clustering method (hierarchical clustering) and comparing the periods 1999–2002 and 2003–2007. We view this result as supporting the notion that the ensemble approach is less pliable to clustering stochasticity than using each method separately. Importantly, the results for the full sample period show that 63.6% of banks (*n* = 321) are classified in the same business model in all trienniums. Finally, we observe that the distribution of persistent banks per business model is similar to the full sample (BM1: 41.7%; BM2: 19.3%; BM3: 19.3%; BM4: 19.6%).

A possible concern regarding these results is whether the changes in business model made by the non-persistent banks reflect actual changes in business policy or, rather, reveal fragilities in our clustering approach to adequately capture the business models of banks. To address this concern, we analyse the number of business model changes per bank. The intuition behind this analysis is that one-off changes are more likely to reflect clear policy changes, whereas

TABLE 10 Persistent banks per business model and period

	n	Persistent	BM1	BM2	BM3	BM4
T1 and T2	370	305 (82.4%)	108	83	55	59
T2 and T3	428	362 (84.6%)	137	82	77	66
T3 and T4	473	374 (79.1%)	161	68	77	68
All trienniums (**persistent banks**)	505	321 (63.6%)	134	62	62	63
Core and persistent banks	505	211 (41.8%)	130	22	25	34

Notes: Number of banks with the same business model classification across consecutive trienniums. A persistent bank records the same business model in every triennium it is present in the sample. The full sample period is divided in four trienniums: 2005–2007 (T1), 2008–2010 (T2), 2011–2013 (T3) and 2014–2016 (T4). The clustering is obtained using the ensemble approach for each triennium separately. The sample size *n* for each pair of consecutive trienniums is presented in the first column. Note that the number of banks considered in the fourth line ('All trienniums') is 505, rather than 524 (original full sample), because 19 banks are present in only one triennium in our sample and hence were excluded from the analysis of business model persistency. The percentage value presented in the column 'Persistent' for each period is calculated as the number of banks with the same business model classification divided by the sample size *n* for the period (e.g. T1 and T2 = (305/370) × 100 = 82.4%).

multiple changes indicate the presence of clustering stochasticity. Table 11 shows that 79.3% of banks record one-off changes in the sample period (146 of 184 non-persistent banks), even though they are mostly present in three or four trienniums in our sample. Alternatively, when considering all the banks in the sample, the number of banks that change more than once is only 7.3%. In our view, these findings attest that the approach followed is able to capture actual business policy changes. Regarding the banks with multiple business model changes (n = 38), these banks record high values of fuzziness (mean PCM1 = 0.41, PCM2 = 0.28, PCM1 − PCM2 = 0.13, HHI = 0.32, SW = 0.06), which suggests that they are likely combining the 'recipes' of more than one business model. Another possible explanation for multiple business model changes may lie in the structural shifts that have occurred in the banking sector during our sample period (2005–2016), including the global financial crisis, the sovereign debt crisis, and the implementation of new regulatory requirements under the Basel III Accord with significant impacts on the business model choices of banks (e.g. net stable funding ratio, liquidity coverage ratio, leverage ratio).

Finally, we may be interested in investigating the distinctive features of non-persistent (relative to persistent) banks. In other words, which features significantly impact the likelihood of a bank changing its business model? To study this, we perform a logistic regression wherein the explained variable is a dummy that takes on the value 1 if the bank changes its business model in a given triennium. The explanatory variables are all the business model variables. In brief, the results presented in Table 12 seem to suggest that the impact of some variables on the likelihood of a business policy change is significantly different across business models.

Namely, for retail-focused banks (BM1) the likelihood of a business policy change increases significantly with size. In a way, such a result may seem counterintuitive, in the sense that large banks could be expected to be less mobile given the amount of resources invested in their activity. In other words, one could suspect that larger banks are subject to higher mobility barriers than their smaller peers are (Caves and Porter, 1977). However, recall that, on average, BM1 banks are relatively small (mean total assets: 7.0, log) and, as such, an increase in size may represent an opportunity to have access to alternative sources of diversification and market funding with the goal, for instance, of reducing funding costs (Huang and Ratnovski, 2011) or signalling creditworthiness (Demirgüç-Kunt and Huizinga, 2010). The signal and significance of the coefficient of size for BM1 banks provides some support for the 'size as opportunity' rather than 'size as barrier' narrative. On the other hand, larger values of deposits and equity seem to deter retail banks away from changing their business model. Such barriers to mobility seem to reflect the fact that equity and customer deposits are often labelled as the most stable sources of bank funding; for the latter, this has become particularly true since the introduction of deposit guarantee schemes in the 1930s (Diamond and Dybvig, 1983; Rajan, 1992).

For retail diversified funding banks (BM2), which contrary to BM1 are, on average, among the largest in our sample, increases in size seem to reduce the likelihood of mobility—which is consistent with the 'size as opportunity' explanation presented early. Also, with the remaining business model variables held constant, both loans to customers and wholesale funding seem to impede mobility. On the other hand, BM2 banks with higher customer deposits are more likely to change their business model. In line with this, untabulated results indicate that the main destination of non-persistent BM2 banks is BM1 (53 in 94 BM shifts). This finding suggests that shifts from BM2 to BM1 do not occur in a short period of time, and require a smooth transition that is visible in the high values of customer deposits observed in the triennium that precedes the actual business model change—an observation that is consistent with the notion that retail banking requires investments in customer proximity that are likely to take some time to become productive (Berger *et al*., 2005). Importantly, note that this result is not driven by banks that operate with mixed business models, given that we control for the business model fuzziness (which, as expected, is positively correlated with non-persistency). In a similar vein, the majority of non-persistent banks following the retail diversified assets model (BM3) transit to BM1 (27 out of 47). By looking at the regression coefficients in Table 12, we observe that the propensity to change business model in BM3 banks increases for banks with larger lending portfolios. Again, in line with the

TABLE 11 Number of business model changes

		Number of trienniums bank is present in the sample			
	Total	**1**	**2**	**3**	**4**
Total banks	524	19	80	84	341
Banks with no changes*	340	19	64	53	204
Banks with changes	184 (100%)		16	31	137
one change	146 (79.3%)		16	26	104
two changes	30 (16.3%)			5	25
three changes	8 (4.3%)				8

Notes: The classification is obtained using the ensemble classification output following a majority consensus rule for each triennium. To illustrate how we compute the number of business model changes, consider a bank that is present in four trienniums in our sample (last column) and we obtain the following clustering (ensemble) results: T1 = BM1, T2 = BM2, T3 = BM1, T4 = BM1. In this case, we record two business model changes. *Note that the 'Total' obtained for the row 'zero changes' (n = 340) corresponds to the number of 'persistent banks' identified in Table 10n (= 321) plus 19 banks that are present in only one triennium in our sample, and hence were excluded from the persistency analysis in Table 10.

TABLE 12 Likelihood of non-persistency: logistic regressions

	BM1	BM2	BM3	BM4
Gross loans to customers	−0.02	−0.04**	0.07***	0.01
Trading assets	0.03	−0.14*	0.02	−0.01
Interbank lending	−0.02	−0.03	−0.06***	−0.04
Customer deposits	−0.12**	0.06**	0.03	0.05
Interbank borrowing	0.01	−0.04*	0.00	−0.06
Wholesale funding	0.09	−0.05**	0.04	0.00
Total derivatives	0.08	−0.10	0.08	0.02
Income diversification	0.02	0.02	0.00	0.04
Total assets	1.55**	−1.48***	0.47	−2.30***
Total equity	−0.27***	0.05	0.11	−0.05
Fuzziness	1.73	4.54***	2.41*	0.60
Bank–triennium obs. (non-persist.)	469 (63)	327 (94)	256 (47)	219 (26)
Akaike information criterion	147.86	220.40	139.86	52.25
McFadden's pseudo R^2	0.687	0.520	0.558	0.873

Notes: Values presented are the coefficient estimates of a pooled Bayesian logistic regression with fixed effects for the trienniums and post-change business model. Explained variable: for each bank–triennium observation we label as non-persistent (dummy = 1) if a change occurs in the business model in the next triennium and label as persistent (dummy = 0) if the business model remains the same. Explanatory variables: business model variables, and 'business model fuzziness' given by 1 minus the difference of the top two percentage of cluster membership in t, i.e. 1 − (PCM1 − PCM2). Fixed effects included: triennium and business model recorded in t + 1. Predictors with statistically significant positive values are positively correlated with the likelihood of a bank being non-persistent, whereas predictors with a statistically significant negative value are inversely related with such likelihood. McFadden's pseudo $R^2 = 1 - [\ln(LM)/\ln(L0)]$ wherein ln(LM) is the log-likelihood of the fitted model and ln(L0) is the log-likelihood of the model with the intercept as the only predictor. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level respectively.

explanation put forward for BM2 non-persistent banks, we view this finding as supporting the idea that, prior to shifting to the retail-focused model, banks are required to invest in retail-specific technology that takes some time to yield results. Finally, the only significant feature of non-persistent banks with a large diversified business model (BM4) is their smaller size, which seems to support the notion that the possibility of attaining significant economies of scale in transactional banking is likely driving the long-term persistency of large diversified banks (van Ewijk and Arnold, 2014).

9 | ROBUSTNESS CHECKS

9.1 | Different Subsamples

A possible concern regarding our clustering approach may be whether the business model classifications (ensemble, core, persistent) hold after imposing disturbances to the baseline sample. In order to test this, we start by drawing 100 random subsamples without replacement, imposing a 1% disturbance on the baseline sample (d_1 = 1%). In other words, each subsample comprises 99% of the total observations in the baseline sample (i.e. n = 519). Then, we run our business model classification approach on each subsample and test the similarity of classifications (ensemble, core, persistent) relative to the classifications obtained for the baseline sample. We use four similarity measures: SM, RI, ARI, and JI. Results are reported as the mean similarity across subsamples. Finally, we repeat this procedure using different subsample sizes, namely comprising 95% and 90% of total observations in the baseline sample, d_2 and d_3 respectively (with the same number of random subsamples each, i.e. 100).

Table 13 shows that, when imposing a 1% disturbance on the baseline sample (upper panel), the ensemble classification remains virtually unchanged (SM = 0.988, RI = 0.987, ARI = 0.967, JI = 0.956). Similar results are found when focusing on the identification of core and persistent banks, although at relatively lower levels (e.g. RI = 0.955 and 0.897 respectively). In the same line, the vast majority of business model classifications perform well when imposing 5% and 10% disturbances on the baseline sample (middle and lower panels). In particular, the RI ranges from 0.807 to 0.966 for all types of classification (ensemble, core, persistent). As expected, as the disturbances become larger, the similarity of classifications with the baseline sample reduce for all measures (e.g. Gates *et al.*, 2019). This result suggests that, although the approach handles small disturbances well, practitioners should strive to use a stable sample in a scenario where business model analysis is performed in a time-varying setting (e.g. monitoring the evolution of performance and riskiness of banks per business model over time).

9.2 | Different Clustering Methods

Another potential source of concern may lie in the choice of methods included in the ensemble. In order to test whether the combination of methods significantly impacts the classification results, we remove one of the original methods at a time (PAM, SOM, FCM), replacing it with one of two alternative methods discussed in literature related to

TABLE 13 Stability of business model classification for different subsamples

	SM	RI	ARI	JI
***d_1* = 1% (*n* = 519)**				
Ensemble classification	0.988	0.987	0.967	0.956
Core versus non-core	0.975	0.955	0.909	0.922
Persistent versus non-persistent	0.945	0.897	0.794	0.823
***d_2* = 5% (*n* = 498)**				
Ensemble classification	0.967	0.966	0.915	0.888
Core versus non-core	0.938	0.892	0.783	0.821
Persistent versus non-persistent	0.912	0.840	0.679	0.735
***d_3* = 10% (*n* = 472)**				
Ensemble classification	0.914	0.954	0.886	0.853
Core versus non-core	0.954	0.860	0.720	0.774
Persistent versus non-persistent	0.891	0.807	0.614	0.689

Notes: We test the stability of our classification approach using three sets of 100 random subsamples drawn from the baseline sample without replacement. Each set has a different subsample size: 99% of the baseline sample (d_1, n = 519), 95% (d_2, n = 498), and 90% (d_3, n = 472). We report the mean value of each measure obtained for the three sets of samples. Simple matching (SM) is computed as the number of observations with the same classification divided by the total number of observations. Remaining similarity measures are described in detail in Milligan and Cooper (1986); namely, the RI is measured as the proportion of pairs labelled as true positive or true negative divided by the total number of pairs of observations (Rand, 1971); the ARI corrects the original RI for randomness (Hubert and Arabie, 1985); the multi-class mean JI is defined as the proportion of pairs labelled as true positives divided by the total number of pairs of observations excluding true negatives (Jaccard, 1901).

banking business models and strategic groups (Martín-Oliver *et al*., 2017; Mergaerts and Vander Vennet, 2016; Zúñiga-Vicente and Vicente-Lorente, 2006): HC and model-based clustering (MBC). This produces the following six alternative combinations of methods: HC/SOM/FCM, PAM/HC/FCM, PAM/SOM/HC, MBC/SOM/FCM, PAM/MBC/FCM, and PAM/SOM/MBC.

The (untabulated) results show that, on average, 91.4% of the observations retain the same classification using the alternative and the original combinations of methods (maximum similarity: 98.3%, MBC/SOM/FCM; minimum similarity: 87.4%, PAM/SOM/MBC). This finding suggests that the clustering results are not significantly dependent on the choice of methods. Importantly, there are fundamental reasons for why we use the original combination of methods (PAM/SOM/FCM). First, the combination has been used in extant literature (Alam *et al*., 2000; Budayan *et al*., 2011). Second, the classification obtained with PAM is not prone to outliers and yields a deterministic solution. Third, the FCM method yields information on the best and second-best cluster assignments per bank, a result that is a central input in the fuzziness analysis. And, finally, SOMs yield particularly informative graphical representations for banking business model analysis, as presented in Figure 1.

On the other hand, the choice of methods may also impact the identification of core banks. To test whether this is the case, we run a similar experiment to the one presented earlier; that is, we remove one of the original methods at a time (PAM, SOM, FCM) and replace it with HC and MBC, iteratively. The results show that 92.3% of the banks are labelled in the same category (core or non-core) using the alternative and the original combinations of methods (maximum similarity: 95.6%, MBC/SOM/FCM; minimum similarity: 89.1%, PAM/SOM/HC). Similarly, we test the impact of using the different combinations of methods on the identification of persistent banks. On average, 82.2% of banks are labelled in the same category (persistent or non-persistent) when compared with our original ensemble (maximum: 89.7%, MBC/SOM/FCM; minimum: 76.0%, PAM/SOM/MBC). Both findings may be interpreted as an indication that the identification of core and persistent banks is not significantly impacted by the choice of methods.

9.3 | Clustering with the Original Variables

Next, we assess whether the approach is insensitive to the decision regarding the use of original variables versus retained principal components. To do this, we run our clustering ensemble approach using the original variables, rather than the retained components, as inputs, and we check (i) the similarity of classifications and (ii) the business model composition. Regarding (i), the untabulated results indicate that 85.5% of the observations are labelled in the same business model using the original variables and the retained components. Regarding (ii), the composition of business models using the original variables seems relatively similar to the one that results from using the retained components as inputs (Table 5). Namely:

- **BM1** (retail focused): loans to customers (65.4%), customer deposits (68.9%);
- **BM2** (retail diversified funding): loans to customers (71.5%), wholesale funding (24.0%);
- **BM3** (retail diversified assets): interbank lending (35.8%), customer deposits (58.6%);

- **BM4** (large diversified): trading assets (8.7%), derivatives (15.6%), income diversification (51.6, HHI) and total assets (7.9, log).

Both results (classification similarity and composition) seem to suggest that our approach is, in fact, not entirely insensitive to the choice of inputs (retained components versus original variables). For this reason, we have explained with particular care the decision to use the retained components as inputs in the clustering process (see Section 6.3).

9.4 | Business Model Interpretation after Core and Persistency Treatments

An additional concern we may face is whether using the treated samples leads to a loss of interpretation of the business models or, rather, contributes to an increased clarity of interpretation. To answer this concern, we examine whether the mean values of core and persistent banks (C&P) are significantly different from the mean values of other banks (i.e. non-core, non-persistent, or both) per business model, and assess whether the sign of the changes increases (or reduces) model interpretability. The results are reported in Table 14. When we compare the mean value of C&P versus other banks, we find statistically significant differences across the main variables of each business model. Namely:

- **BM1** (retail focused): loans to customers (+4.5 pp), customer deposits (+12.3 pp);
- **BM2** (retail diversified funding): loans to customers (+4.2 pp), wholesale funding (+19.8 pp);
- **BM3** (retail diversified assets): interbank lending (+24.3 pp), customer deposits (+11.5 pp);
- **BM4** (large diversified): trading assets (+5.1 pp), derivatives (+14.4 pp), income diversification (+6.5, HHI), total assets (+0.4, log).

Importantly, the difference between the mean value of C&P and other banks is positive for the variables in which the business model records the highest value compared to other models (see Table 6). This implies that the differences between business models become accentuated in the C&P sample, apparently conforming more closely to distinctive organizational configurations. Figure 2 shows the graphical representation of the data points labelled as C&P using three of the five retained principal components (PC1, PC2, and PC3), representing a total variation explained of 61.0%. In general, the graphical inspection provides further evidence of a clear separation between models and homogeneity within each model. These results, in our view, sustain the usefulness of using core and persistent banks as a method to increase the robustness of business models analysis.

9.5 | Out-of-Sample Examples of Banks per Business Model

In our final check, we assess whether our approach allows the identification of business models for out-of-sample banks. To perform this analysis, we take exemplary banks and rerun our procedure to identify banking business models. More precisely, we begin by adding 25 US and Japanese global banks included in the study by Roengpitya *et al*. (2017) to our original sample (n = 549 = 524 + 25) and selecting

TABLE 14 Composition of business models per sub-sample: core and persistent (C&P) vs other banks

	BM1			BM2			BM3			BM4		
	C&P	Other	Diff.	C&P	Other	Diff.	C&P	Other	Diff.	C&P	Other	Diff.
Number of banks	130	73		22	102		25	84		34	54	
Gross loans to customers	70.0	65.4	4.5**	71.1	66.9	4.2	26.8	38.2	−11.4***	32.8	44.7	−11.9***
Trading assets	1.3	2.7	−1.3***	1.3	2.1	−0.8	0.5	2.4	−1.9*	**14.5**	**9.5**	**5.1*** **
Interbank lending	8.3	7.9	0.4	6.4	9.5	−3.1**	**56.2**	**32.0**	**24.3*** **	18.0	15.9	2.1
Customer deposits	**71.7**	**59.4**	**12.3*** **	27.6	39.7	−12.1***	67.3	55.8	11.5**	**24.5**	**32.3**	**−7.8** **
Interbank borrowing	**9.7**	**14.8**	**−5.1*** **	17.8	22.3	−4.5	22.5	25.1	−2.7	21.7	18.1	3.6
Wholesale funding	5.8	9.7	−3.8***	**41.8**	**22.0**	**19.8*** **	2.6	5.1	−2.5	16.5	23.1	−6.5*
Total derivatives	1.2	1.8	−0.6**	3.5	3.9	−0.3	0.5	1.1	−0.6	**29.2**	**14.7**	**14.4*** **
Income diversification	46.8	47.9	−1.1	35.4	44.9	−9.5***	49.4	45.7	3.7	**59.1**	**52.6**	**6.5*** **
Total assets	6.9	7.1	−0.1***	7.4	7.5	−0.1	6.9	7.0	−0.1	**8.4**	**7.9**	**0.4*** **
Total equity	8.7	9.2	−0.5	4.2	6.4	−2.2***	5.5	7.1	−1.6*	4.0	6.0	−2***

Notes: Mean values, except number of banks (count). Classification obtained via clustering ensemble using PC1 to PC5 as input variables. For each variable and business model, we compute the Tukey HSD test for comparison of means between the two subsamples: core and persistent banks (C&P) and other banks (Other). *, **, and *** indicate the statistical significance of the difference of each pairwise comparison at the 10%, 5%, and 1% levels respectively. This table is similar to Table 5 (full sample) but is decomposed into two subsamples; thus, both tables are comparable. Values in bold indicate the variables, per business model, identified both as (i) statistically different across subsamples (P < 5%) and (ii) the highest or lowest across business models. All variables computed as percentage of total assets, except income diversification (HHI) and total assets (log).

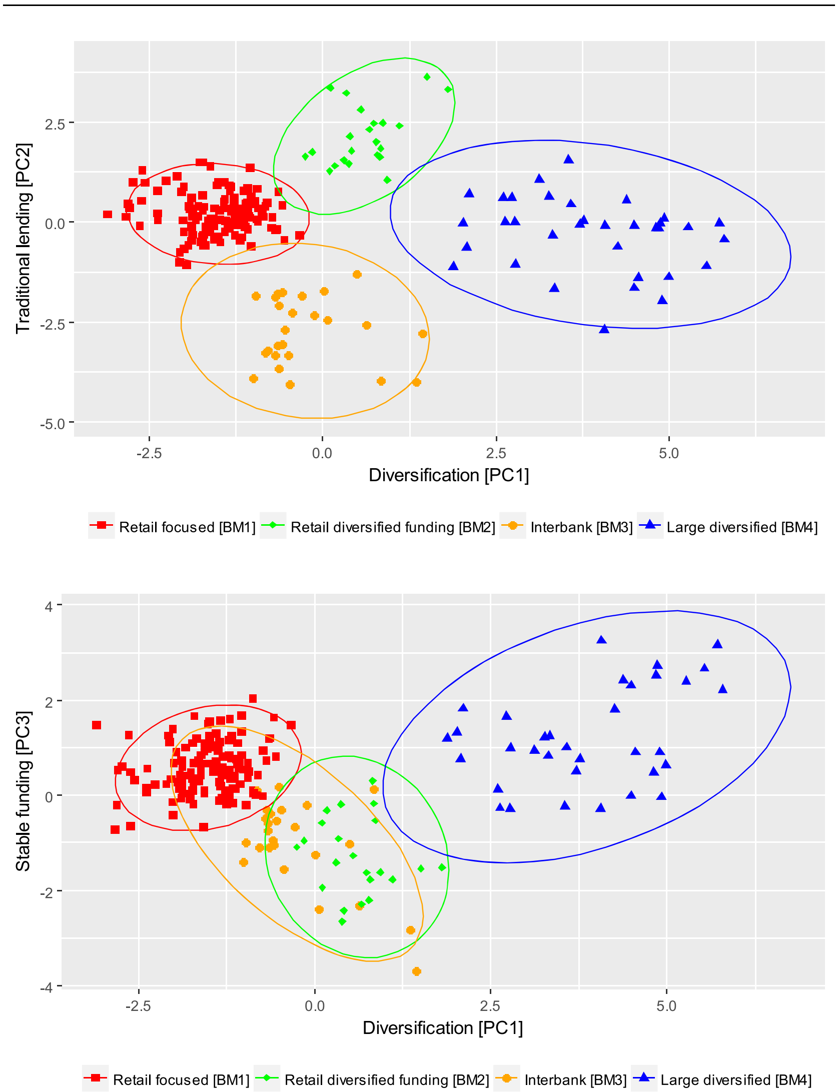

FIGURE 2 Business model representation. Notes: Sample, 'core and persistent banks' (*n* = 211: BM1 = 130, BM2 = 22, BM3 = 25, BM4 = 34). BM classification obtained from clustering ensemble. For parsimony, we plot the top three principal components, which explain 61.0% of total variation, as described in Table 3

the optimal number of clusters based on the valuation criteria of each clustering method (FCM, SOM, PAM). Untabulated results indicate that the original findings reported in Table 4 (that lead us to identify four business models) remain virtually unchanged. Next, we check whether the composition of business models suffered any significant changes. Using Tukey HSD tests to compare the mean values for each business model variable between the original sample (*n* = 524) and the new sample (*n* = 549), we were unable to reject the null hypothesis of equality of means—which, again, lends support to the stability of our initial composition. Finally, we report the business model classification of the 'new' banks (for brevity reasons, we only report the classification of the US banks):

BM1 (retail focused)

- Bank of America: gross loans to customers (52.0%), customer deposits (71.1%);
- BB&T: gross loans to customers (66.0%), customer deposits (69.1%);
- Capital One Financial: gross loans to customers (66.0%), customer deposits (63.3%);
- Comerica: gross loans to customers (73.2%), customer deposits (76.6%);
- PNC Financial Services: gross loans to customers (58.2%), customer deposits (67.5%);
- SunTrust Banks: gross loans to customers (71.4%), customer deposits (71.3%).

BM2 (retail diversified funding)

- Wells Fargo & Company: gross loans to customers (62.4%), wholesale funding (15.0%).

BM3 (retail diversified assets)

- Bank of New York Mellon: interbank lending (36.9%), customer deposits (66.2%);

- Deutsche Bank Trust Corp.: interbank lending (76.9%), customer deposits (54.2%);
- Northern Trust Corp.: interbank lending (31.2%), customer deposits (78.2%);
- State Street Corp.: interbank lending (29.7%), customer deposits (68.7%).

BM4 (large diversified)

- Goldman Sachs: trading assets (30.7%), derivatives (12.7%), income diversification (59.1, HHI), total assets (8.8, log);
- JPMorgan Chase Bank: trading assets (13.9%), derivatives (8.5%), income diversification (68.3, HHI), total assets (9.1, log);
- Morgan Stanley: trading assets (27.5%), derivatives (9.6%), income diversification (53.1, HHI), total assets (8.8, log).

We note that despite the out-of-sample banks only comprising large global banks and size being one of the business model variables used in the clustering process, we are still able to identify at least one bank per banking business model. In our view, such a result seems to suggest that our proposed approach is extensible to geographies beyond Europe and, hence, may be actionable for policy and managerial purposes.

10 | CONCLUSIONS AND OPPORTUNITIES FOR FUTURE RESEARCH

In this paper, we propose a clustering ensemble approach to distinguish banks in terms of (i) their long-term similarity with other banks operating with the same business model (core versus non-core) and (ii) their persistency over time (persistent versus non-persistent). In order to frame this approach, we provide a definition of business model that follows the fuzzy approach to strategic groups theory (DeSarbo and Grewal, 2008; Reger and Huff, 1993). The methods included in our ensemble are chosen based on their different strengths, including their ability to perform fuzzy clustering (FCM), to yield intuitive visualizations of the clusters (SOMs), and to circumvent the presence of data outliers (PAM). The consensus reached among the methods that form the ensemble is based on a majority rule that allows us to have an increased level of confidence regarding the accuracy of the classification attributed to the banks that meet the criterion.

We find four business models in the European banking industry (2005–2016): retail focused, retail diversified funding, retail diversified assets, and large diversified. The banks labelled as core exhibit a significantly higher adherence to their assigned business model (lower fuzziness) than non-core banks do. Moreover, the tests for stability of classification show that the ensemble classification is practically immune to disturbances to the sample, the choice of methods, and the choice of variables. Also, we detect significant differences in the mean values of the key dimensions of each business model when using the sample of core and persistent banks when compared with the remaining banks. Importantly, in the sample of core and persistent banks the differences between business models become clearer (i.e. more accentuated) when compared with the differences observed among the remaining banks. Finally, we provide evidence that expanding our approach to include banks in the USA and Japan yields meaningful results.

Based on our findings, we envision two areas of interest for future research. The first is the use of the concepts of core and persistent banks to study topics such as bank performance, riskiness. and systemic risk. Namely, by narrowing the analysis of bank performance to a sample of banks that are consistently classified under the same business model, researchers may expect to have a better handle on undesired sources of heterogeneity or endogeneity (e.g. shift in strategic orientation), and hence enjoy a cleaner testing environment to grasp the long-term effects of business models. The second area of interest is related to the study of the effects of business model changes by non-persistent banks on bank performance and riskiness.

ACKNOWLEDGEMENTS

This research is financed by European Union and Portuguese public funds through the European Social Fund (Portugal 2020) and FCT (Fundação para a Ciência e a Tecnologia, I.P.) under projects number UID/ECO/04105/2019 (CEF.UP) and UID/GES/00731/2019 (CPBS), as well as the Phd scholarship SFRH/BD/135939/2018. This research has also received funding from the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA). We are grateful for the helpful comments from Paula Brito (FEP) and three anonymous referees. We would also like to thank the participants of the 6th Young Finance Scholars Conference 2019 at University of Sussex, IFABS 2018 Porto conference, 1st PhD Student Workshop in Economics and Business Administration at Universidade do Minho, and 13th Católica Porto Business School Internal Conference for their valuable comments

ORCID

Bernardo P. Marques https://orcid.org/0000-0001-6289-4116
Carlos F. Alves https://orcid.org/0000-0003-3861-8554

REFERENCES

Abellán, J., & Castellano, J. G. (2017). A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications*, *73*(1), 1–10.

Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications*, *18*(3), 185–199.

Ala'raj, M., & Abbod, M. F. (2016). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Systems with Applications*, *64*(1), 36–55.

Aryuni, M., Madyatmadja, E. D., & Miranda, E. (2018). Customer segmentation in XYZ bank using *K*-means and *K*-medoids clustering. Proceedings of 2018 International Conference on Information Management and Technology (ICIMTech), 412–416.

Ayadi, R., Arbak, E., & De Groen, W. P. (2011). *Business Models in European Banking: A Pre and Post-Crisis Screening*. Brussels, Belgium: Center for European Policy Studies.

Ayadi, R., & De Groen, W. P. (2015). *Banking Business Models Monitor 2014: Europe*. Brussels, Belgium/Montreal, QC, Canada: Center for European Policy Studies/IOFSC.

Azadeh, A., Javanmardi, L., & Saberi, M. (2010). The impact of decision-making units features on efficiency by integration of data envelopment analysis, artificial neural network, fuzzy c-means and analysis of variance. *International Journal of Operational Research*, *7*(3), 387–411.

Back, B., Oosterom, G., Sere, K., & Van Wezel, M. (1995). *Intelligent information systems within business bankruptcy predictions using neural networks*. In 3rd European Conference on Information Systems, 1–3.

Basel Committee on Banking Supervision. (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Basel Committee on Banking Supervision. June 2011, Basel.

Berger, A., Clarke, G., Cull, R., Klapper, L., & Udell, G. (2005). Corporate governance and bank performance: a joint analysis of the static, selection, and dynamic effects of domestic, foreign, and state ownership. *Journal of Banking & Finance*, *29*(8), 2179–2221.

Bezdek, J., Ehrlich, R., & Full, W. (1984). FCM: the fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2–3), 191–203.

Blundell-Wignall, A., Atkinson, P., & Roulet, C. (2014). Bank business models and the Basel system: complexity and interconnectedness. *OECD Journal: Financial Market Trends*, *2013*(2), 43–68.

Budayan, C., Dikmen, I., & Birgonul, M. T. (2011). Hybrid strategic groups in construction. *Engineering Project Organization Journal*, *1*(3), 183–196.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics: Theory and Methods*, *3*(1), 1–27.

Cavelaars, P., & Passenier, J. (2012). Follow the money: what does the literature on banking tell prudential supervisors about bank business models? *Journal of Financial Regulation and Compliance*, *20*(4), 402–416.

Caves, R. E., & Porter, M. E. (1977). From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition. *The Quarterly Journal of Economics*, *91*, 241–261.

Cebeci, Z., Yildiz, F., Kavlak, A., Cebeci, C., & Onder, H. (2019). ppclust: probabilistic and possibilistic cluster analysis. R package version 0.1.2.

Cernov, M., & Urbano, T. (2018). Identification of EU bank business models: a novel approach to classifying banks in the EU regulatory framework. EBA Staff Paper Series 2/2018. EBA: London.

Chair, S., & Charrad, M. (2017). multisom: clustering a data set using multi-SOM algorithm. R package version 1.3.

Chang, F., Qiu, W., Carey, V., Zamar, R., Lazarus, R., & Wang, X. (2019). clues: clustering method based on local. R package version 0.6.2.2.

Chiorazzo, V., D'Apice, V., DeYoung, R., & Morelli, P. (2018). Is the traditional banking model a survivor? *Journal of Banking & Finance*, *97*, 238–256.

Curi, C., Lozano-Vivas, A., & 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.

Curry, B., Davies, F., Phillips, P., Evans, M., & Moutinho, L. (2001). The Kohonen self-organizing map: an application to the study of strategic groups in the UK hotel industry. *Expert Systems*, *18*(1), 19–31.

Davalos, S., Leng, F., Feroz, E. H., & Cao, Z. (2014). Designing an if–then rules-based ensemble of heterogeneous bankruptcy classifiers: a genetic algorithm approach. *Intelligent Systems in Accounting, Finance and Management*, *21*(3), 129–153.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(2), 224–227.

de Andrés, J., Lorca, P., de Cos Juez, F. J., & Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: a hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS). *Expert Systems with Applications*, *38*(3), 1866–1875.

de Haan, L., & Kakes, J. (2019). European banks after the global financial crisis: peak accumulated losses, twin crises and business models. *Journal of Banking Regulation*. Advance online publication. https://doi.org/10.1057/s41261-019-00107-y.

de Haas, R., & Van Lelyveld, I. (2010). Internal capital markets and lending by multinational bank subsidiaries. *Journal of Financial Intermediation*, *19*(1), 1–25.

Deboeck, G. J. (1998). Financial applications of self-organizing maps. *Neural Network World*, *8*(2), 213–241.

Demirgüç-Kunt, A., & Huizinga, H. (2010). Bank activity and funding strategies: the impact on risk and returns. *Journal of Financial Economics*, *98*, 626–650.

DeSarbo, W. S., & Grewal, R. (2008). Hybrid strategic groups. *Strategic Management Journal*, *29*(3), 293–317.

Desgraupes, B. (2016). clustercrit: clustering indices. R package version 1.2.7.

DeYoung, R. (2010). Scale economies are a distraction. *The Region, September 2010*, 14–16.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, *51*(3), 393–414.

Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, *91*(3), 401–419.

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, *4*(1), 95–104.

Easterbrook, F. H. (1984). Two agency-cost explanations of dividends. *American Economic Review*, *74*(4), 650–659.

EBA. (2013). EBA report on impact assessment for liquidity measures under Article 509(1) of the CRR. European Banking Authority: London, UK.

EBA. (2014). Guidelines on common procedures and methodologies for the supervisory review and evaluation process. EBA/GL/2014/13. European Banking Authority: London, UK.

EBA. (2015). EBA report on net stable funding requirements under Article 510 of the CRR. EBA/Op/2015/22. European Banking Authority: London, UK

EBA. (2016). EBA report on the leverage ratio requirements under Article 511 of the CRR. EBA/Op/2016/13. European Banking Authority. London, UK.

ECB. (2018). SSM thematic review on profitability and business models: report on the outcome of the assessment. European Central Bank: Frankfurt, Germany.

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

Galbraith, C., & Schendel, D. (1983). An empirical analysis of strategy types. *Strategic Management Journal*, *4*(2), 153–173.

Gates, A. J., Wood, I. B., Hetrick, W. P., & Ahn, Y. Y. (2019). Element-centric clustering comparison unifies overlaps and hierarchy. *Scientific Reports*, *9*(1), dash;1–13.

Gelman, A., & Su, Y. (2018). arm: data analysis using regression and multilevel/hierarchical models. R package version 1.1011.

Gorton, G., & Metrick, A. (2012). Securitized banking and the run on repo. *Journal of Financial Economics*, *104*(3), 425–451.

Hryckiewicz, A., & Kozłowski, Ł. (2017). Banking business models and the nature of financial crisis. *Journal of International Money and Finance*, *71*, 1–24.

Hsu, T.-H. (2000). An application of fuzzy clustering in group-positioning analysis. *Proceedings of the National Science Council, Republic of China*, *10*(2), 157–167.

Huang, R., & Ratnovski, L. (2011). The dark side of bank wholesale funding. *Journal of Financial Intermediation*, *20*(2), 248–263.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218.

Hunter, W. C., & Timme, S. G. (1986). Technical change, organizational form, and the structure of bank production. *Journal of Money, Credit and Banking*, *18*(2), 152–166.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, *37*, 547–579.

Jackman, S. (2017). pscl: Political Science Computational Laboratory. R package version 1.5.2.

Jain, A. K. (2010). Data clustering: 50 years beyond *K*-means. *Pattern Recognition Letters*, *31*(8), 651–666.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken; NJ, USA: John Wiley & Sons, Inc.

Kiviluoto, K., & Bergius, P. (1998). *Two-level self-organizing maps for analysis of financial statements*. In IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence, 1, 189-192.

Köhler, M. (2015). Which banks are more risky? The impact of business models on bank stability. *Journal of Financial Stability*, *16*, 195–212.

Kohonen, T. (1997). Exploration of very large databases by self-organizing maps. Proceedings of International Conference on Neural Networks, 1, PL1-PL6.

Kuncheva, L. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Liberti, J. M., & Mian, A. R. (2008). Estimating the effect of hierarchies on information use. *Review of Financial Studies*, *22*(10), 4057–4090.

Liikanen, E. (2012). Reforming the structure of the EU banking sector. Report of the High-Level Expert Group, chaired by Erkki Liikanen, European Commission: Brussels, Belgium

Liu, Y., Wang, X., & Zhai, D. (2010). Commercial banks exceptional client distinguish based on data mining. *Proceedings of the 2010 International Forum on Information Technology and Applications*, 1, 164-166.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(14), 281–297.

Maechler, M. (2018). cluster: 'finding groups in data': cluster analysis extended Rousseeuw et al. R package version 2.0.71.

Marghescu, D., Sarlin, P., & Liu, S. (2010). Early-warning analysis for currency crises in emerging markets: a revisit with fuzzy clustering. *Intelligent Systems in Accounting, Finance & Management*, *17*(3–4), 143–165.

Martin, A., Gayathri, V., Saranya, G., Gayathri, P., & Venkatesan, P. (2011). A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy c-means and MARS. *International Journal on Soft Computing*, *2*(1), 12–24.

Martín-Oliver, A., Ruano, S., & Salas-Fumás, V. (2017). The fall of Spanish cajas: lessons of ownership and governance for banks. *Journal of Financial Stability*, *33*, 244–260.

Mergaerts, F., & Vander Vennet, R. (2016). Business models and bank performance: a long-term perspective. *Journal of Financial Stability*, *22*, 57–75.

Michalopoulos, M., Dounias, G., Hatas, D., & Zopounidis, C. (2002). On the use of a combination approach to automated knowledge acquisition based on neural networks and fuzzy logic with regard to credit scoring problems. In C. Zapounidis (Ed.), *New Trends in Banking Management* (pp. 87–101). Heidelberg, Germany: Physica.

Miller, D. (1986). Configurations of strategy and structure: towards a synthesis. *Strategic Management Journal*, *7*(3), 233–249.

Milligan, G. W., & Cooper, M. C. (1986). A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research*, *21*(4), 441–458. https://doi.org/10.1207/s15327906mbr2104_5

O'Hara, M., & Shaw, W. (1990). Deposit insurance and wealth effects: the value of being 'too big to fail'. *Journal of Finance*, *45*(5), 1587–1600.

Onan, A. (2019). Consensus clustering-based undersampling approach to imbalanced learning. *Scientific Programming*, *2019*, 1–14.

Popović, D., & Bašić, B. (2009). Churn prediction model in retail banking using fuzzy c-means algorithm. *Informatica*, *33*(2), 243–248.

Porter, M. E. (1979). The structure within industries and companies' performance. *Review of Economics and Statistics*, *61*(2), 214–227.

Rajan, R. G. (1992). Insiders and outsiders: the choice between informed and arm's length debt. *The Journal of Finance*, *47*(4), 1367–1400.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850.

Ravikumar, P., & Ravi, V. (2006). Bankruptcy prediction in banks by an ensemble classifier. Proceedings of the 2006 IEEE International Conference on Industrial Technology, 2032–2036.

Reger, R. K., & Huff, A. S. (1993). Strategic groups: a cognitive perspective. *Strategic Management Journal*, *14*(2), 103–123.

Roengpitya, R., Tarashev, N., Tsatsaronis, K., & Villegas, A. (2017). Bank business models: popularity and performance. BIS Working Papers no. 682.

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

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

Sarlin, P. (2016). Visual macroprudential surveillance of banks. *Intelligent Systems in Accounting, Finance and Management*, *23*(4), 257–264.

Sarlin, P., & Marghescu, D. (2011). Visual predictions of currency crises using self-organizing maps. *Intelligent Systems in Accounting, Finance and Management*, *18*(1), 15–38.

Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, *17*(3), 227–238.

Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons.

Taylor, J. B., & Williams, J. C. (2009). A black swan in the money market. *American Economic Journal: Macroeconomics*, *1*(1), 58–83.

van Ewijk, S., & Arnold, I. (2014). How bank business models drive interest margins: evidence from US bank-level data. *The European Journal of Finance*, *20*(10), 850–873.

Warnes, G. R., Bolker, B., Lumley, T., & Johnson, R. C. (2018). gmodels: various R programming tools for model fitting. R package version 2.18.1.

Wickham, H., & Chang, W., (2019). ggplot2: create elegant data visualisations using the grammar of graphics. *R package version 2*. *2*(1).

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

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

Zúñiga-Vicente, J. Á., & Vicente-Lorente, J. D. (2006). Strategic moves and organizational survival in turbulent environments: the case of Spanish banks (1983–97). *Journal of Management Studies*, *43*(3), 485–519.

How to cite this article: Marques BP, Alves CF. Using clustering ensemble to identify banking business models. *Intell Sys Acc Fin Mgmt*. 2020;1–29. https://doi.org/10.1002/isaf.1471

APPENDIX A

A.1. | Clustering Algorithms

A.1.1. | Fuzzy C-Means

The following algorithm was run (Cebeci *et al*., 2019):

1 Randomly initialize the membership matrix ($\boldsymbol{U}_{i \times j}$), where i are the data points and j are the clusters (j = 1,…,J, with predefined J). The following constraints must be satisfied:

$$\mu_{i,j} = [0,1];\ 1 \le i \le n, 1 \le j \le J$$

$$0 \le \sum_{i=1}^{n} \mu_{i,j} \le n; 1 \le j \le J$$

$$\sum_{j=1}^{J} \mu_{i,j} = 1;\ 1 \le i \le n$$

where i are the observations (i = 1,…,n), j are the clusters (j = 1,…,J) and J is predetermined.

2 Calculate the prototype cluster centres ($\vec{v}_j$, $1 \le j \le J$) using a pre-determined measure of fuzziness ($1 \le m < \infty$):

$$\vec{v}_j = \frac{\sum_{i=1}^{n} \mu_{i,j}^{m} \vec{x}_i}{\sum_{i=1}^{n} \mu_{i,j}^{m}}$$

3 Compute the dissimilarity matrix d^2 (i.e. the squared Euclidean distance) between the data points $\vec{x}_i$ and each cluster centre $\vec{v}_j$:

$$d^2\left(\vec{x}_i, \vec{v}_j\right) = \left\|\vec{x}_i - \vec{v}_j\right\|^2$$

4 Update the previous version of $\mu_{i,j}$:

$$\mu_{i,j} = \frac{\left[1/d^2\left(\vec{x}_i, \vec{v}_j\right)\right]^{1/(m-1)}}{\sum_{l=1}^{k}\left[1/d^2\left(\vec{x}_i, \vec{v}_l\right)\right]^{1/(m-1)}}$$

where the denominator is the sum of all weights and is used to normalize the membership scores.

5 Repeat steps 2 to 4 until the objective function cannot be improved:

$$\min J = \sum_{i=1}^{n}\sum_{i=1}^{n} \mu_{i,j}^{m} d^2\left(\vec{x}_{i,}, \vec{v}_j\right)$$

A.1.2. | Self-Organizing Maps

Chair and Charrad (2017) implement a batch version of the following algorithm:

1 Initialize the ‘neurons’ weights matrix ($\boldsymbol{W}_{j \times p}$) based on the linear grids upon the first two principle components direction, where j are the ‘neurons’ (j = 1,…,J, with predefined J) and p are the input variables.
2 Draw a sample training input vector $\vec{x}_i$.
3 Find the winning neuron $I\left(\vec{x}_i\right)$ so that

$$\min d^2\left(\vec{x}_i, \vec{w}_j\right) = \left\|\vec{x}_i - \vec{w}_j\right\|^2 = d^2\left(\vec{x}_i, I\left(\vec{x}_i\right)\right)$$

4 Compute weight update equation:

$$\Delta w_{ji} = \rho(t)\, T_{j,I\left(\vec{x}_i\right)}(t)\left(\vec{x}_i - \vec{w}_j\right)$$

where $T_{j,I\left(\vec{x}_i\right)}$ is the Gaussian neighbourhood and $\rho(t)$ is the learning rate.

5 Repeat steps 2 to 4 until $\rho(t)$ cannot be improved.

A.1.3. | Partitioning around Medoids

The algorithm takes the following steps (Maechler, 2018):

1 Randomly select J prototype ‘representative data points’ or medoids $\vec{c}_j$, where j are the clusters (j = 1,…,J, with predefined J).
2 Based on the dissimilarity matrix, assign each data point to the nearest medoid and compute the sum of all distances to their medoids (‘cost’) and to other points in the same cluster.
3 Find a new prototype medoid, by taking the point with the lowest sum of distances to the other points in the same cluster.
4 Re-run step 2 (update of assignment and cost) with new prototype medoid.

5 Compute total swapping cost, by comparing the 'cost' of the new prototype with the previous.
6 Repeat steps 3 to 5 until total swapping cost becomes zero or negative.

A.2. | List of R Packages Used in the Paper

- 'arm: data analysis using regression and multilevel/hierarchical models', version 1.101 (Gelman and Su, 2018)—applied in the Bayesian logistic regression to capture differences in the observed variables between persistent and non-persistent banks.
- 'clues: clustering method based on local', version 0.6.2.2, (Chang et al., 2019)—used to compute the cluster similarities measures, namely the RI, the ARI, and the JI.
- 'cluster: "finding groups in data": cluster analysis extended Rousseeuw *et al.*', version 2.0.71 (Maechler, 2018)—used to employ the PAM algorithm.
- 'clustercrit: clustering indices', version 1.2.7 (Desgraupes, 2016)—employed to examine the quality of the clustering outputs; that is, to obtain the SW, the Caliński-Harabasz index, the Davies–Bouldin index, and the Dunn index.
- 'ggplot2: create elegant data visualisations using the grammar of graphics', version 2.2.1 (Wickham and Chang, 2019)—used to build 2D plots of business models.
- 'gmodels: various R programming tools for model fitting', version 2.18.1 (Warnes *et al.*, 2018)—implemented to perform the χ^2 independence test.
- 'multisom: clustering a data set using multi-SOM algorithm', version 1.3 (Chair and Charrad, 2017)—used to compute the SOMs.
- 'ppclust: probabilistic and possibilistic cluster analysis', version 0.1.2 (Cebeci *et al.*, 2019)—used to compute the FCM algorithm.
- 'pscl: Political Science Computational Laboratory', version 1.5.2 (Jackman, 2017)—applied to obtain the McFadden's pseudo R^2 in the logistic regression.
- 'stats: the R stats package', version 3.4.4 (this package is maintained by the R Core Team)—employed to obtain descriptive statistics as well as to compute the ex-post Tukey HSD test for comparison of means across clusters.

A.3. | Valuation Criteria

The SW for each observation SW_i is computed as the difference between the average distance of observation i to other observations in the nearest cluster b_i and the average distance between observation i and observations in its assigned cluster a_i. Hence, the average SW is given by (Rousseeuw, 1987)

$$\mathrm{SW} = \frac{1}{n} \times \sum_{i=1}^{n} \mathrm{SW}_i = \frac{1}{n} \times \sum_{i=1}^{n} \frac{b_i - a_i}{\max(a_i, b_i)}$$

The value of SW is positively related to cluster quality.

The Caliński–Harabasz index CHI is computed as the ratio of between-groups sum of squares (BGSS) to within-group sum of squares WGSS, for a given partition of J clusters (Caliński and Harabasz, 1974):

$$\mathrm{CHI} = \frac{n-J}{J-1} \times \frac{\mathrm{BGSS}}{\mathrm{WGSS}}$$

A higher value CHI is an indication of good cluster quality.

The Davies–Bouldin index DBI is the average value of the largest within dispersion-to-between separation of each cluster M_J (Davies and Bouldin, 1979):

$$\mathrm{DBI} = \frac{1}{J} \times \sum_{J=1}^{J} M_J$$

The value of DBI is negatively related to cluster quality.

The Dunn index DI measures the ratio between the minimum distance between observations in different clusters $d_{\min}$ and the maximum distance between observations in the same cluster $d_{\max}$ (Dunn, 1974):

$$\mathrm{DI} = \frac{d_{\min}}{d_{\max}}$$

A higher value of DI indicates better clustering output.

A.4. | Composition of Business Models per Clustering Method

	BM1	BM2	BM3	BM4
FCM				
Number of banks	185	134	115	90
Gross loans to customers	68.0 (12.8)$^{++}$	69.5 (11.9)$^{++}$	35.7 (16.1)$^{++}$	40.8 (18.6)$^{++}$
Trading assets	1.8 (3.5)$^{+}$	2.0 (2.5)$^{+}$	2.0 (4.8)$^{+}$	**11.2 (9)$^{+++}$**
Interbank lending	8.2 (5.5)$^{++}$	8.2 (5.8)$^{++}$	**36.6 (19.1)$^{+++}$**	16.5 (12.6)$^{+++}$
Customer deposits	**69.0 (12.5)$^{+++}$**	40.6 (15.7)$^{+++}$	56.3 (23.9)$^{+++}$	**28.8 (16.1)$^{+++}$**
Interbank borrowing	**10.2 (7.4)$^{+++}$**	21.1 (15.2)$^{++}$	**25.8 (19.8)$^{+++}$**	19.1 (10.5)$^{++}$
Wholesale funding	6.9 (6.2)$^{++}$	23.4 (16.3)$^{++}$	5.0 (9.2)$^{++}$	21.5 (16.6)$^{++}$
Total derivatives	1.4 (2.1)$^{+}$	3.2 (3.2)$^{+}$	1.2 (3.0)$^{+}$	**20.3 (14.3)$^{+++}$**
Income diversification	46.9 (11.3)$^{+}$	44.4 (12.4)$^{+}$	46.6 (12.4)$^{+}$	54.3 (10.6)$^{+++}$
Total assets	7.0 (0.3)$^{++}$	7.5 (0.5)$^{+++}$	7.0 (0.4)$^{++}$	**8.1 (0.7)$^{+++}$**
Total equity	**9.0 (4.4)$^{+++}$**	6.2 (3.2)$^{+}$	6.9 (4.7)$^{++}$	5.2 (3.0)$^{++}$
SOM				
Number of banks	199	105	106	114
Gross loans to customers	69.1 (11.4)$^{++}$	71.1 (11.3)$^{++}$	**35.2 (15.7)$^{+++}$**	41.4 (17.6)$^{+++}$
Trading assets	1.7 (3.2)$^{+}$	1.6 (2.3)$^{+}$	2.2 (4.9)$^{+}$	**9.6 (8.8)$^{+++}$**
Interbank lending	8.2 (5.5)$^{++}$	7.7 (5.3)$^{++}$	**36.4 (19.6)$^{+++}$**	17.8 (13.5)$^{+++}$
Customer deposits	**67.5 (13.2)$^{+++}$**	37.4 (16.6)$^{+++}$	61.2 (20.6)$^{+++}$	**30. (16.1)$^{+++}$**
Interbank borrowing	**11.6 (8.7)$^{+++}$**	21.4 (17.7)$^{+}$	21.9 (18)$^{+}$	22.0 (13.6)$^{+}$
Wholesale funding	7.3 (6.6)$^{++}$	**26.4 (18)$^{+++}$**	3.8 (6.7)$^{++}$	20.2 (14.8)$^{+++}$
Total derivatives	1.3 (1.7)$^{+}$	3.4 (3.9)$^{+}$	1.2 (3.4)$^{+}$	**16.7 (14.4)$^{+++}$**
Income diversification	47.2 (11)$^{++}$	41.8 (13.1)$^{+++}$	47.1 (12.3)$^{++}$	53.5 (10.4)$^{+++}$
Total assets	7.0 (0.3)$^{++}$	7.5 (0.5)$^{+++}$	7.0 (0.4)$^{++}$	**8.0 (0.7)$^{+++}$**
Total equity	**8.8 (4.5)$^{+++}$**	6.0 (3.4)$^{+}$	7.1 (4.6)$^{++}$	5.3 (2.9)$^{++}$
PAM				
Number of banks	230	143	103	48
Gross loans to customers	**69.1 (13)$^{+++}$**	61.3 (15.2)$^{+++}$	35.1 (15.9)$^{++}$	29.5 (14)$^{++}$
Trading assets	1.7 (3.2)$^{++}$	3.4 (4.3)$^{++}$	2.0 (5.1)$^{+}$	**15.5 (9.1)$^{+++}$**
Interbank lending	8.3 (5.6)$^{++}$	11.0 (9.8)$^{++}$	**38.0 (19.2)$^{+++}$**	19.3 (13.1)$^{+++}$
Customer deposits	**65.0 (14.7)$^{+++}$**	34.8 (16.2)$^{+++}$	59.3 (23)$^{+++}$	**25.9 (15.3)$^{+++}$**
Interbank borrowing	**13.2 (10.8)$^{+++}$**	19.7 (14.8)$^{++}$	24.9 (20.2)$^{++}$	20.3 (10.2)$^{+}$
Wholesale funding	7.9 (7.1)$^{+++}$	**28.1 (16.9)$^{+++}$**	**3.7 (7)$^{+++}$**	14.9 (12.4)$^{+++}$
Total derivatives	1.5 (2.2)$^{++}$	6.3 (5.7)$^{+++}$	0.8 (2.8)$^{++}$	**27.6 (15.6)$^{+++}$**
Income diversification	47.4 (10.8)$^{+}$	44.7 (12.7)$^{+}$	46.6 (12.7)$^{+}$	58.1 (9.5)$^{+++}$
Total assets	7.0 (0.4)$^{++}$	7.7 (0.6)$^{+++}$	7.0 (0.4)$^{++}$	**8.0 (0.8)$^{+++}$**
Total equity	**8.8 (4.6)$^{+++}$**	5.7 (3)$^{+}$	6.5 (3.9)$^{+}$	5.1 (3.5)$^{+}$

Notes: Mean values and standard deviation in parentheses, except number of banks (count). Classification obtained using PC1 to PC5 as input variables. For each variable, we compute the Tukey HSD test for comparison of means per pair of business models; that is, for a given variable, the mean value of each business model is potentially different from the mean of the remaining three business models (only two, only one, or none). The number of ($^{+}$) indicates the number of pairwise comparisons that are statistically different at the 5% level. Values in bold indicate the business models with the highest and lowest mean values for each variable, when the number of plus signs is ($^{+++}$). All variables computed as percentage of total assets, except income diversification (HHI) and total assets (log).