Using clustering ensemble to identify banking business models

Bernardo P. Marques (presenter)
Carlos F. Alves

Univ. Porto, Faculty of Economics
Católica Porto Business School

October 16, 2020

DaSSWeb – Data Science and Statistics Webinar

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 program" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA).

Disclaimer: This presentation should not be reported as representing the views of the European Central Bank (ECB) or the Single Supervisory Mechanism (SSM). The views expressed are those of the authors and do not necessarily reflect those of the ECB or the SSM. Particularly, the methods presented do not reflect the current or past methods used in the Supervisory Review and Evaluation Process (SREP) to identify the business models of the supervised credit institutions.
Agenda

➢ General conditions for the application of the method
➢ Specific context of the paper
➢ Clustering methods and valuation criteria
➢ Step-by-step overview of the method
➢ Some results
➢ Key takeways & what’s next
General applications of the method

➢ The method will be particularly useful under the following pre-conditions:

1. The researcher suspects that **groups exist** in the data, based on literature/observation; **and**
2. The researcher suspects that datapoints may have **affinity with more than one group**; **and**
3. The researcher suspects that datapoints may suffer **changes over time**; **and**
4. The researcher has **no access** to the **true assignment** of each datapoint to a group.

➢ Example of applications in management studies

- The researcher is analyzing a dataset of heterogenous firms and suspects that the baseline results may change depending on the firms’ **business orientation** w.r.t.:
  - Activities (going beyond ‘assets per NACE classification’)
  - Funding sources (going beyond ‘debt/equity ratio’)
  - Products (going beyond ‘revenues per product’)
  - Geographies (going beyond ‘revenues per country’)

Checking pre-conditions: activities’ orientation
1. Groups: Group A. Real Estate + Construction + Transp. Group B. Retail + Food/Accom. (synergies)
2. Fuzziness: firm may operate in other groups of NACE codes (but less significantly).
3. Changes: merger may change of NACE mix.
4. True group: no prior info on group assignment.
What is a banking business model?

- Banks make money by capturing funds from savers and investing in financial assets (issued by borrowers) as well as providing financial services, such as payment services (financial intermediation), to different types of customers.

- Literature has looked at ways in which to summarize the variety of strategic choices made by bank managers (e.g. type of activities, funding sources, size, diversification, equity) into a single proxy (discrete or continuous) which may categorize the way that the bank operates, i.e. its business model.

  - On one extreme, there is retail banking (focused on customer deposits, loans to customers, standard services) → credit risk as main exposure.
  - On the other, there is diversified banking (significantly funded via wholesale lenders, invested in all types of financial assets) → credit and market risk, as well as systemic risk (TBTF, externality for the entire system).
Why does the identification of banking business models matter?

➢ Academia
   • most explanations that banking literature has put forward are likely to change depending on whether we are referring to small retail banks or large diversified players;
   • each business model is expected to be subject to specific risks and vulnerabilities.

➢ Regulators/Supervisors
   • viability and sustainability of the business model as an element of the Supervisory Review and Evaluation Process (SREP), with implications in terms of Pillar 2 capital requirements (P2R).
Why aren’t the methods used so far to identify business models satisfactory?

- Expert judgement (Kohler, 2015; Cernov & Urbano, 2018) → inability to replicate results and potential conflict of interests;

- Factor analysis (Ewijk & Arnold, 2014; Mergaerts & Vennet, 2016) → does not allow for benchmark analysis among peer banks;

- Hard clustering (Martín-Oliver et al., 2017) → poor clustering given that some banks operate with mixed business models (e.g. following M&A).
Clustering methods (1/3)

- **Fuzzy clustering**
  - fuzzy logic: membership of a data point to a cluster may be nuanced, and hence a binomial membership function is likely to be oversimplistic (Zadeh, 1965).
  - assignment of data points to clusters: continuous function, between 0 and 1 (percentage of cluster membership, wherein the nearer the membership value is to 1, the higher in the similarity between the observation and the cluster).

- **Fuzzy C-Means (Bezdek et al., 1984)**
  - 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 is the weighting scheme:

  \[
  \min F = \sum_{i=1}^{n} \mu_{ij}^{m} d^2 \left( \overline{x}_i, \overline{v}_j \right)
  \]

  - $\overline{x}_i$ is the data vector for each bank (of size $1 \times k$, where $k$ are the number of input features)
  - $\overline{v}_j$ is the vector of cluster centres ($1 \times k$)
  - $d$ is the dissimilarity measure.
Clustering methods (2/3)

➢ Self-Organizing Maps (Kohonen, 1997)
  • 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 2), which briefly put consists of:
    1. Identifying, sequentially and for each data point, the neuron that is closest to a given point (winning neuron), based on the weights vector.
    2. Each assignment leads the neuron to update its codebook vector, as well as the vector of neighbor neurons.
  • The information contained in the vector of each neuron can be used to build effective visualizing tools.
Clustering methods (3/3)

- Partitioning Around Medoids or K-medoids (Kaufman & Rousseeuw, 1990)
  - 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.
  - 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).
  - Adequate to handle datasets with outliers.

Clustering ensemble

- “combines the information provided by the partitions” of different clustering methods (Jain, 2010: p.660).
- Illustration: in medical diagnosis, we may wish to combine the diagnoses performed by a variety of experts, based on a given consensus scheme, into a single medical diagnosis, under the expectation that the accuracy of the diagnosis is improved as a result.
- Consensus schemes: unanimity, simple majority, and plurality.
- Accuracy: expected to increase with diversity of the methods (Kuncheva, 2004).
Valuation criteria

**Silhouette Width**
(Rousseeuw, 1987)

\[
SW = \frac{1}{n} \times \sum_{i=1}^{n} SW_i = \frac{1}{n} \times \sum_{i=1}^{n} \frac{b_i - a_i}{\max(a_i, b_i)}
\]

- \(a_i\): avg distance between obs. \(i\) and obs. in assigned cluster
- \(b_i\): avg distance between obs. \(i\) and obs. in nearest cluster

**Calinski–Harabasz index**
(Calinski & Harabasz, 1974)

\[
CHI = \frac{n - J}{J - 1} \times \frac{BGSS}{WGSS}
\]

- \(BGSS\): between group sum of squares
- \(WGSS\): within group sum of squares

**Davies–Bouldin index**
(Davies & Bouldin, 1979)

\[
DBI = \frac{1}{J} \times \sum_{j=1}^{J} M_j
\]

- \(M_j\): avg of largest within dispersion-to-between separation of cluster \(j\)

**Dunn index**
(Dunn, 1974)

\[
DI = \frac{d_{\min}}{d_{\max}}
\]

- \(d_{\min}\): min distance between obs. in different clusters
- \(d_{\max}\): max distance between obs. in same cluster

*Other standard nomenclature*
- \(n\): nbr of obs
- \(J\): nbr of clusters
Step-by-step overview of the method

1. Check conditions to use method
2. Select input variables
3. Perform PCA
4. Select clustering methods
5. Perform clustering with selected methods
6. Identify optimal number of clusters
7. Harmonize clustering output
8. Join clustering output (ensemble)
9. Analyze clustering ensemble output
10. Identify core banks
11. Identify persistent banks
12. Check robustness

- **Clustering methods**
  - FCM (fuzzy data)
  - SOM (visualization)
  - PAM (outliers)

- **Internal valuation**
  - Silhouette Width
  - Caliński-Harabasz
  - Davies-Bouldin
  - Dunn Index

- **Similarity between ensemble vs FCM, SOM, PAM**
  - Measures: SM, JI, ARI, CIT

- **Voting scheme**: unanimity
  - Silhouette Width > 0.20

- **Triennium values (not full period)**
  - Same classification over time

- **Random sampling**
  - Clustering methods (HC, MBC)
  - Types of variables (original)

- **Legend**:
  - Only researcher can do it!
  - Only software can do it!
  - There is software but depending on the level of programming skills researcher may also do it
  - R code in the Appendix
  - Formulae in the Appendix
Result 1: we found four business models for which, generically, the main types of risk exposures seem to be significantly different (1/2)

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>Selection criteria: number of business models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASW</td>
</tr>
<tr>
<td>FCM</td>
<td></td>
</tr>
<tr>
<td>J = 3</td>
<td>0.23</td>
</tr>
<tr>
<td>J = 4</td>
<td>0.18</td>
</tr>
<tr>
<td>J = 5</td>
<td>0.15</td>
</tr>
<tr>
<td>J = 6</td>
<td>0.14</td>
</tr>
<tr>
<td>J = 7</td>
<td>0.13</td>
</tr>
<tr>
<td>J = 8</td>
<td>0.10</td>
</tr>
<tr>
<td>J = 9</td>
<td>0.09</td>
</tr>
<tr>
<td>SOM</td>
<td></td>
</tr>
<tr>
<td>J = 3</td>
<td>0.11</td>
</tr>
<tr>
<td>J = 4</td>
<td>0.19</td>
</tr>
<tr>
<td>J = 5</td>
<td>0.21</td>
</tr>
<tr>
<td>J = 6</td>
<td>0.11</td>
</tr>
<tr>
<td>J = 7</td>
<td>0.14</td>
</tr>
<tr>
<td>J = 8</td>
<td>0.11</td>
</tr>
<tr>
<td>J = 9</td>
<td>0.07</td>
</tr>
<tr>
<td>PAM</td>
<td></td>
</tr>
<tr>
<td>J = 3</td>
<td>0.15</td>
</tr>
<tr>
<td>J = 4</td>
<td>0.23</td>
</tr>
<tr>
<td>J = 5</td>
<td>0.21</td>
</tr>
<tr>
<td>J = 6</td>
<td>0.19</td>
</tr>
<tr>
<td>J = 7</td>
<td>0.21</td>
</tr>
<tr>
<td>J = 8</td>
<td>0.20</td>
</tr>
<tr>
<td>J = 9</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Result 1: we found four business models for which, generically, the main types of risk exposures seem to be significantly different (2/2)

**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)

*table appendix*
Result 2: level of affinity of banks with allocated model differs per model.

<table>
<thead>
<tr>
<th>TABLE 7</th>
<th>Core banks per business model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Criterion 1. Same BM classification across all methods ('unanimity')</td>
<td></td>
</tr>
<tr>
<td>FCM = PAM</td>
<td>427 (81.5%)</td>
</tr>
<tr>
<td>SOM = FCM</td>
<td>468 (89.3%)</td>
</tr>
<tr>
<td>PAM = SOM</td>
<td>422 (80.5%)</td>
</tr>
<tr>
<td>(a) PAM = SOM = FCM</td>
<td>400 (76.3%)</td>
</tr>
<tr>
<td>Criterion 2. SW &gt; 0.2 for the clustering ensemble</td>
<td></td>
</tr>
<tr>
<td>FCM</td>
<td>276 (52.7%)</td>
</tr>
<tr>
<td>SOM</td>
<td>266 (50.8%)</td>
</tr>
<tr>
<td>PAM</td>
<td>316 (60.3%)</td>
</tr>
<tr>
<td>(b) Ensemble</td>
<td>281 (53.6%)</td>
</tr>
<tr>
<td>Core banks (a, b)</td>
<td>273 (52.1%)</td>
</tr>
</tbody>
</table>

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.

- 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
- BM2 and BM3 banks represent a significantly lower share (12.1% vs 23.7%, 11.4% vs 20.8% respectively).

Level of affinity/fuzziness

➢ According to strategic groups literature, some firms may operate with mixed strategies, e.g. following a merger or acquisition (DeSarbo & Grewal, 2008).
Result 3: type and height of mobility barriers vary across business models.

Table 12: Likelihood of non-persistency: logistic regressions

<table>
<thead>
<tr>
<th></th>
<th>BM1</th>
<th>BM2</th>
<th>BM3</th>
<th>BM4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross loans to customers</td>
<td>-0.02</td>
<td>-0.04**</td>
<td>0.07***</td>
<td>0.01</td>
</tr>
<tr>
<td>Trading assets</td>
<td>0.03</td>
<td>-0.14*</td>
<td>0.02</td>
<td>-0.01</td>
</tr>
<tr>
<td>Interbank lending</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.06***</td>
<td>-0.04</td>
</tr>
<tr>
<td>Customer deposits</td>
<td>-0.12**</td>
<td>0.06**</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Interbank borrowing</td>
<td>0.01</td>
<td>-0.04*</td>
<td>0.00</td>
<td>-0.06</td>
</tr>
<tr>
<td>Wholesale funding</td>
<td>0.09</td>
<td>-0.05**</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Total derivatives</td>
<td>0.08</td>
<td>-0.10</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>Income diversification</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Total assets</td>
<td>1.55**</td>
<td>-1.48***</td>
<td>0.47</td>
<td>-2.30***</td>
</tr>
<tr>
<td>Total equity</td>
<td>-0.27***</td>
<td>0.05</td>
<td>0.11</td>
<td>-0.05</td>
</tr>
<tr>
<td>Fuzziness</td>
<td>1.73</td>
<td>4.54***</td>
<td>2.41*</td>
<td>0.60</td>
</tr>
<tr>
<td>Bank-tribennium obs. (non-persist.)</td>
<td>469 (63)</td>
<td>327 (94)</td>
<td>256 (47)</td>
<td>219 (26)</td>
</tr>
<tr>
<td>Akaike information criterion</td>
<td>147.86</td>
<td>220.40</td>
<td>139.86</td>
<td>52.25</td>
</tr>
<tr>
<td>McFadden's pseudo R²</td>
<td>0.687</td>
<td>0.520</td>
<td>0.558</td>
<td>0.873</td>
</tr>
</tbody>
</table>

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-tribennium 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 $R^2 = 1 - \frac{ln(LM)/ln(L0)}{ln(LM)/ln(L0)}$ 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.
Key take-aways

➢ The method has a variety of applications in management studies (particularly if: groups expected to exist, datapoints with affinity to more than one group, may change group over time, true classification unknown).

➢ Paper defines the method step-by-step.

➢ In the context of our empirical context, we found that:

  ▪ The European banking sector seems to be populated by four business models;

  ▪ Banks following the retail diversified model tend to be fuzzier than others – for these banks it may be advisable to monitor more closely the evolution of the main vulnerabilities;

  ▪ Height and type of mobility barriers differ across business models, wherein size, reliance on customer deposits and exposure interbank lending seem to be the main barriers to change.
What’s next?

➢ Implementation

➢ Improvement
  • Incorporate time dimension (time-series clustering).
  • Use interval data instead of means.
Using clustering ensemble to identify banking business models

Bernardo P. Marques (presenter)
Carlos F. Alves

University of Porto, Faculty of Economics

October 16, 2020

Data Science and Statistics Webinar

Thank You!

I am happy to answer any questions!

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 program” under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA).

Disclaimer: This presentation should not be reported as representing the views of the European Central Bank (ECB) or the Single Supervisory Mechanism (SSM). The views expressed are those of the authors and do not necessarily reflect those of the ECB or the SSM. Particularly, the methods presented do not reflect the current or past methods used in the Supervisory Review and Evaluation Process (SREP) to identify the business models of the supervised credit institutions.
References


Appendix 1. Some lines of code in R (1/2)

**PCA**

```
1 library(stats)
2 library(xlsx)
3 data <- read_excel("C:/data.xlsx")
4 pca <- prcomp(data, center = TRUE, scale. = TRUE)
5 tve <- summary(pca)[[6]]  # tve: total variation explained
6 ncomp <- 5  # ncomp: number of retained components
7 rawloadings <- pca$rotation[,1:ncomp] %>% diag(pca$sdev, ncomp, ncomp)
8 varimax(rawloadings)$loadings
```

**Clustering**

```
1 library(opclust)
2 fcm_nclust <- fcm(comp, centers=nclust)$cluster  # nclust: number of clusters
3 library(multisom)
4 somgrid_nclust <- somgrid(xdim = nx, ydim = ny, topo = "rectangular")  # nx\times ny=nclust
5 som_nclust <- as.integer(BatchSOM(as.matrix(comp), grid = somgrid_nclust,
6     mtn.radius=0.5, max.radius=1, maxit=1000,
7     init="linear", radius.type="gaussian")$classif)
8 library(cluster)
9 dist <- dist(comp, method ="euclidean")  # dist: retained pca components
10 pam_nclust <- pam(dist, k=nclust, cluster.only=TRUE)  # nclust: number of clusters
```

**Valuation criteria**

```
1 library(cluster)
2 sw_method_nclust <- silhouette(method_nclust, dist)[,3]  # method_nclust: method(fcm,som,fcm) & nbr of clusters
3 library(clusterCrit)
4 chi_method_nclust <- as.numeric(intCriteria(as.matrix(comp),method_nclust,"Calinski_Harabasz"))
5 db_method_nclust <- as.numeric(intCriteria(as.matrix(comp),method_nclust,"Davies_Bouldin"))
6 du_method_nclust <- as.numeric(intCriteria(as.matrix(comp),method_nclust,"Dunn"))
```
Appendix 1. Some lines of code in R (2/2)

Clustering ensemble step

```r
ensemble <- ifelse(clusters$fcm == clusters$som, clusters$fcm,
                   ifelse(clusters$fcm == clusters$pam, clusters$fcm,
                          ifelse(clusters$som == clusters$pam, clusters$som,
                                 ifelse(clusters$som == clusters$fcm_sw & clusters$som_sw > clusters$pam_sw, clusters$fcm,
                                           ifelse(clusters$som_sw > clusters$fcm_sw & clusters$som_sw > clusters$pam_sw, clusters$som,
                                                  ifelse(clusters$pam_sw > clusters$fcm_sw & clusters$pam_sw > clusters$som_sw, clusters$som, clusters$pam, 0))))))))
```

Random sampling

```r
library(base)
for (rep in 1:100) {
  rm(.Random.seed)
  # random sample
  rand <- as.vector(sample(data$index, nobs, replace=FALSE)) ##nobs: count of subset observations
  rand_index <- bm[data$id %in% rand, c(2:12)] ##id: unique id per observation
  rand_index <- rand_index[order(rand_index$index),]
  data <- subset(rand_index, select=c(2:11)) ##to remove id column
}
Appendix 2. Clustering algorithms

**Partitioning Around Medoids.** The algorithm takes the following steps (Maechler, 2018):

1. Randomly select \( f \) prototype ‘representative data points’ or medoids \((\bar{x}_j)\), where \( f \) are the clusters \((j = 1, \ldots, f)\), with pre-defined \( f \).
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.

**Self-Organizing Maps.** Chair & Charrad (2017) implement a batch version of the following algorithm:

1. Initialize the ‘neurons’ weights matrix \((W_{j,p})\) based on the linear grids upon the first two principle components direction, where \( j \) are the ‘neurons’ \((j = 1, \ldots, f)\), with pre-defined \( f \) and \( p \) are the input variables.
2. Draw a sample training input vector \( \bar{x}_t \).
3. Find the winning neuron \( l(x_t) \) so that:
   \[
   \min d^2(\bar{x}_t, \bar{w}_j) = \|\bar{x}_t - \bar{w}_j\|^2 = d^2(\bar{x}_t, l(\bar{x}_t))
   \]
4. Compute weight update equation:
   \[
   \Delta w_{ji} = \rho(t) \, T_{ji}(l(\bar{x}_t)) (\bar{x}_t - \bar{w}_j)
   \]
   where \( T_{ji}(l(\bar{x}_t)) \) is the Gaussian neighbourhood and \( \rho(t) \) is the learning rate.
5. Repeat steps 2 to 4 until \( \rho(t) \) cannot be improved.

**Fuzzy C-Means.** The following algorithm was run (Cebeci et al., 2019):

1. Randomly initialize the membership matrix \((U_{i,j})\), where \( i \) are the data points and \( j \) are the clusters \((j = 1, \ldots, f)\), with pre-defined \( f \). The following constraints must be satisfied:
   \[
   \mu_{ij} = [0,1] ; \quad 1 \leq i \leq n, 1 \leq j \leq f
   \]
   \[
   0 \leq \sum_{j=1}^{f} \mu_{ij} \leq n ; \quad 1 \leq j \leq f
   \]
   \[
   \sum_{j=1}^{f} \mu_{ij} = 1 ; \quad 1 \leq i \leq n
   \]
   where \( i \) are the observations \((i = 1, \ldots, n)\), \( j \) are the clusters \((j = 1, \ldots, f)\) and \( f \) is pre-determined.
2. Calculate the prototype cluster centres \((\bar{v}_j, 1 \leq j \leq f)\) using a pre-determined measure of fuzziness \((1 \leq m < \infty)\):
   \[
   \bar{v}_j = \frac{\sum_{i=1}^{n} \mu_{ij}^m \bar{x}_i}{\sum_{i=1}^{n} \mu_{ij}^m}
   \]
3. Compute the dissimilarity matrix \((d^2)\), i.e. the squared Euclidean distance, between the data points \((\bar{x}_i)\) and each cluster centre \( \bar{v}_j \):
   \[
   d^2(\bar{x}_i, \bar{v}_j) = \|\bar{x}_i - \bar{v}_j\|^2
   \]
4. Update the previous version of \( \mu_{ij} \):
   \[
   \mu_{ij} = \left( \frac{1}{d^2(\bar{x}_i, \bar{v}_j)} \right)^{1/(m-1)}
   \]
   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:
   \[
   \min J = \sum_{i=1}^{n} \sum_{j=1}^{f} \mu_{ij}^m d^2(\bar{x}_i, \bar{v}_j)
   \]
Appendix 3. Valuation criteria

The Silhouette Width 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 of observation $i$ and observations in its assigned cluster ($a_i$). Hence, the average silhouette width ($SW$) is given by (Rousseeuw, 1987):

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

The value of $SW$ is positively related with 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 & Harabasz, 1974):

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

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 & Bouldin, 1974):

$$DBI = \frac{1}{J} \times \sum_{j=1}^{J} M_j$$

The value of $DBI$ is negatively related with 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):

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

A higher value of $DI$ values indicate better clustering output.
**Appendix 4. Composition of business models**

**TABLE 5** Composition of business models: clustering ensemble

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

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).