Using clustering ensemble to identify banking business models

Bernardo P. Marques (presenter)
Carlos F. Alves

University of Porto, Faculty of Economics

May 19, 2020

Presentation made at the
FinTech Workshop on AI, Financial Automation and Market Risk

- What is a banking business model?

  - Banks make money by capturing funds from savers and investing them in financial assets and providing financial services.

  - However, such financial intermediation may vary significantly.

  - On one extreme, there is **retail banking** (focused on customer deposits, loans to customers, standard services).

  - On the other, there is **diversified banking** (significantly funded via wholesale lenders, invested in all types of financial assets).

- 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.

  - Regulators: following the twin banking crises (2007-09 and 2010-12) regulators have been pushed to approve rulebooks that account for the differences in complexity of banking business models (EBA, 2014; US Congress, 2017).

- Why aren't the methods used so far entirely satisfactory?

  - Expert judgement (Kohler, 2015; Cernov & Urbano, 2018) → excessive discretion potentiates conflict of interests for bank supervisors;

  - 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).

- Our proposed method to identify banking business models

1. Principal component analysis based on business model variables identified in literature
2. Unsupervised clustering based on retained components (>80% of variation explained): Fuzzy C-Means, Self-Organizing Maps, Partitioning Around Medoids.
3. Optimal number of clusters using various internal valuation criteria: Silhouette Width, Calinksi-Harabasz, Davies-Bouldin and Dunn Index;
4. Ensemble using a majority rule
5. Core banks (‘votes’ are unanimous and silhouette width above 0.2)
6. Persistent banks (banks with the same business model in all trienniums)

Result 1: we find four significantly different business models.

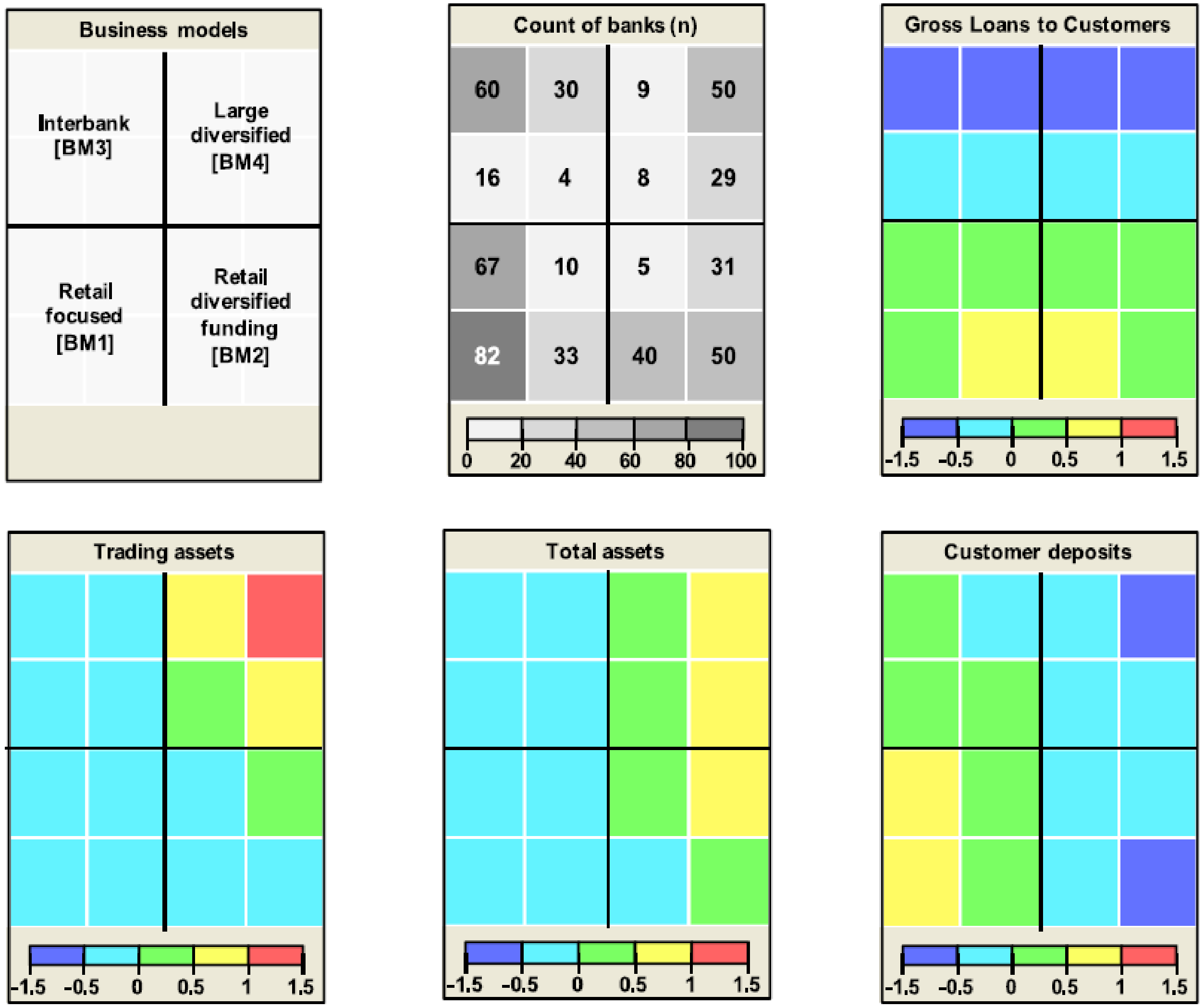

FIGURE 1 SOM of business model features. Notes: The frontier between business models was obtained by performing the clustering ensemble approach on the codebook vectors. The values presented for each variable, ranging from −1.5 to +1.5, correspond to the codebook vectors obtained by performing a batch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

Result 2: fuzziness is significantly different in core vs non-core banks.

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.

Result 3: business models tend to be persistent, and non-persistency occurs mostly due to one-off changes (not ‘frustrated clusterings’).

TABLE 11 Number of business model changes

	Total	Number of trienniums bank is present in the sample			
		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 10*n* (= 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.

- Our method responds well to several robustness checks

  - Different sub-samples (random sampling)
  - Different clustering methods (hierarchical and model-based clustering)
  - Different types of variables (original instead of the retained components)
  - Additional banks (US and Japan)

Using clustering ensemble to identify banking business models

Bernardo P. Marques (presenter)
Carlos F. Alves

University of Porto, Faculty of Economics

May 19, 2020

Presentation made at the
1st FinTech Workshop on AI, Financial Automation and Market Risk