

# Clustering cooperative banks: pre/post crisis business model dynamics (2005–2015)

## Data cleaning and feature engineering

We applied a light but necessary preprocessing step: we removed an irrelevant export column, converted the main variables to numeric format, and handled missing values (especially the RWA ratio) using median imputation within each period. We also engineered two key variables: trading intensity (trading assets / total assets) to capture business model orientation, and log(total assets) to summarize bank size in a robust way (used for interpretation and later regressions, not for clustering).

Importantly, size was not used as a clustering variable: it mainly captures scale (highly asymmetric) rather than business model structure. Instead, size was used later to interpret heterogeneity in transitions.

## Clustering setup and choice of k

Because features are on different scales (returns, ratios, risk measures), all clustering variables were standardized (z-scores) using StandardScaler. We selected the number of clusters using two internal criteria:

- Elbow method: examines inertia (within-cluster SSE) as k increases; the “elbow” indicates diminishing returns from adding clusters.
- Silhouette score: measures separation/compactness; higher values indicate a better partition.

Both diagnostics suggested  $k = 4$  as a good trade-off between interpretability and quality.

## K-means results: four business model types

Cluster profiles supports a simple economic interpretation:

- Cluster 0 — “Market-exposed but stable”: high trading assets intensity (~36%) but trading income near zero; medium risk and good profitability  
→ trading assets likely used for balance-sheet management rather than speculative profit.

- Cluster 1 — “Trading-oriented”: very high trading income ratio (~16%) and relatively high trading assets; lower profitability → revenue depends more strongly on trading.
- Cluster 2 — “Traditional but high-risk”: low trading intensity (~17%) but highest RWA ratio (~0.69), with decent profitability → risk comes mainly from balance-sheet exposures (e.g., credit).
- Cluster 3 — “Distressed”: strongly negative ROA/ROE and negative trading income on average → a small group of underperforming banks.

The silhouette score (~0.27) indicates moderate but meaningful separation, which is reasonable in real financial systems where business models can overlap.

### Pre- vs post-crisis comparison

After identifying four business model types on the pooled sample, we compared the pre and post-crisis periods in two ways. First, we looked at how banks redistribute across clusters, which reveals a clear post-crisis reallocation: the system shifts away from the more traditional/high-risk profile and toward a more trading-oriented profile, while the “stable” model remains dominant overall and the distressed group stays small. Second, we compared cluster profiles within each period to see how the same business model behaves after the crisis. The main pattern is that banks adapt through a mix of composition effects (moving to different models) and within-model adjustments (changes in trading exposure, risk, and profitability), consistent with a post-crisis environment of tighter constraints and lower profitability.

### Transition analysis

To study business model dynamics, we built a transition dataset by keeping only banks observed in both periods and linking each bank’s pre-crisis cluster to its post-crisis cluster. We then computed a transition count matrix and a row-normalized transition probability matrix, which makes it easy to see which clusters are stable and which ones tend to change.

Conclusion: Overall stability is high (about 74% of banks remain in the same cluster), suggesting strong persistence in business models. However, the transition matrix also highlights a clear reallocation across types: Cluster 2

shows a non-negligible shift toward Cluster 1, while Cluster 3 appears much less stable, with banks frequently moving to other clusters.

### Transitions by size and logistic modelling of the 2→1 shift

To test whether transitions depend on bank size, we merged the transition dataset with pre-crisis total assets and grouped banks into small/medium/large classes using quantiles (so size is measured before the outcome). The size-stratified transition matrices show that the key shift from the traditional/high-risk profile toward the trading-oriented one is much more frequent among larger institutions. We then formalized this pattern with a logistic regression focusing on banks starting in cluster 2, where the dependent variable is whether they switch to cluster 1 and the predictor is  $\log(\text{pre-crisis assets})$ .

The estimated coefficient is positive and the predicted probability increases with size; performance ( $\text{ROC-AUC} \approx 0.73$ ) indicates meaningful discriminatory power despite class imbalance. Overall, this supports the interpretation that larger banks were more likely to move toward a trading-oriented model after the crisis.

### Robustness of the model

We tested whether the K-means solution is stable with respect to random initialization by rerunning the algorithm with different seeds and comparing partitions using the Adjusted Rand Index (ARI). ARI values were consistently very high (min/median  $\approx 0.98$ ), meaning the partition is highly reproducible and not driven by initialization randomness.

### Extension with GMM

To complement K-means, we also fitted a Gaussian Mixture Model (GMM), which performs probabilistic (soft) clustering and can capture elliptical, overlapping groups—an appealing feature when bank business models are not cleanly separable. We selected the number of components using BIC and, for comparability with the main analysis, focused on  $k = 4$ . While the overall agreement with K-means (ARI) is low, the comparison is still informative: the most distinctive regimes are robust across methods, with the distressed group and the trading-oriented group mapping largely to single GMM components, whereas differences mainly arise in how the two large “non-distressed” clusters are split, suggesting a more continuous spectrum of profiles with partial overlap.

## Conclusion

The key findings are: four interpretable business model types, a strong post-crisis increase in trading-oriented banks, a major transition channel from traditional/high-risk to trading-oriented models, and a clear size effect, where larger institutions are more likely to switch toward trading-oriented profiles. Robustness checks (ARI) confirm the stability of the K-means solution, while GMM complements the analysis by highlighting overlap and confirming the most distinct regimes.