A Clustering Approach for Business Models of Iranian Banks; Analysis of Risks and Migrations

Sepideh Aghajani[a*]; Reza Samizadeh[b]

[a] Department of industrial engineering, Alzahra University, Tehran, Iran; Address: Tehran-Vanak- Alzahra University- Kharazmi Building- Postal code: 1993893973, Email: aghajani68@ut.ac.ir; Phone number: 09127608350

[b] Department of industrial engineering, Alzahra University, Tehran, Iran; Address: Tehran-Vanak- Alzahra University- Kharazmi Building- Postal code: 1993893973, Email: rsamizadeh@alzahra.ac.ir; Phone number: 09121116695

Abstract

Various studies have shown that different banking business models are related to several variables that will change banks' strategies and impose various risks. This paper examines the data from 2006 to 2021 for 33 Iranian banks while identifying different variables affecting business models. K-Means, FCM, and PAM clustering approaches are used to cluster different Iranian banks. Also, by analyzing the liquidity risk, credit risk, and insolvency risk, the impact of the business model on various risks is examined. In the following, the changes in banks' business models are examined by carefully analyzing the state of different business models from 2006-2021. We found that banking business models shifted from SME-invested banks to SME-operating during shock periods, while the change is reversed during stable periods. Furthermore, large public banks have a small tendency to become large-funding banks in a period of economic stability.

Keywords: Bank; Business Model; Clustering; Financial Risk; Migration

1. Introduction

In order to survive in the financial markets, banks always seek to have sufficient knowledge about their risk. For this purpose, they try to have a suitable dynamic structure by choosing a suitable business model (according to their goals. Historically, it has been observed that banks look for a change in their business model as a final solution when faced with high risks [1]. The importance of this issue is motivating topic for research in Iranian banks. Business model determination in banks is a relatively new subject that is considered in financial research. Different banks try to achieve their goals by choosing different strategies for their activities and balance sheet structure. These differences will lead to changes in their business models. In today's volatile business environment, choosing appropriate business models that cover banks' profitability and risk objectives is important. Also, analyzing the pattern of changes in banks' business models can be a beacon for the future. This issue has been studied in many countries. However, the geographical dimensions and governing the economies of countries have caused great diversity in banking business models [2]. The selection of influential variables in determining business models has also been very diverse. In general, the issues in this area are based on clustering, and it is executed with different approaches to grouping banks in different business models. A general structure of these issues is presented in Figure 1.

Figure 1. Here

According to this figure, the problem includes identifying and analyzing effective variables, selecting a clustering model, identifying business model groups, analyzing each model's risks, and analyzing conversion patterns from one model to another. In order to implement this mechanism, in the first phase, by examining the literature, different clustering approaches, the studied variables, and different business models are analysed. In general, these variables include structural variables like ownership structure [3,4], policy variables (Normal, Islamic) [5]; variables of activity type (participatory, commercial, investment, mortgage, fintech) [6-8]; and balance sheet or behavioral variables [9,10]. It can be seen that there is much diversity and lack of integration in the variables, so it is necessary to have a careful study of the variables and their structural features.

Choosing an accurate method for performing clustering to determine banking business models has always been challenging. The choice of hard approaches (allocating the bank to a single cluster only) or soft (allocating the bank to different clusters with different weights) such as hierarchical, fuzzy, self-organizing map (SOM), and partitioning around medoids (PAM) is the important points in these issues. The different types of financial risks that have been considered in banking business models are credit, liquidity, operational, and systemic risks. Recent studies in this field have shown that different business models can have various risks [11,12]; however, the analysis of the relationship between risk and business models is a very important issue that has not a rich literature. This issue has also been examined with variables such as the financial stability of banks, the risk of insolvency, and the probability of default.

Examining the trend of changes in business models and analyzing the cause of these changes concerning other financial risks is a topic that has been studied sparsely in some studies. In this paper, according to the specific geographical features of 33 Iranian banks, using the available data between 2006 and 2021, the following objectives have been examined:

- Identifying different banking business models according to specific environmental characteristics;
- Accurate selection of variables affecting the grouping of banks;
- Using a precise approach to clustering;
- Analysis of credit risks and liquidity of banks’ business models;
- Examine the trend of business model changes during the last 15 years.

The implementation of this research with data from Iran along with the integrated study of clustering criteria and simultaneous analysis of risk and migration of business models and the use of different clustering models are among the most contribution of this paper compared to previous studies.

2. Literature review

The issue of business models has come to the attention of researchers in the last two decades. In a complete definition, Chesbrough [13] believes that a business model is an approach that an organization implements to gain value from the market and that customers pay for it. Business models may vary in detail depending on the approach of the organizations or industries in which they operate, but their core components are usually common. These components are value proposition, customers or target group, resources, cost, and revenue structure [14]. Resource-based, economic, networked, activity-oriented, strategic, and knowledge-based business models are among the most popular models. Banking business models have also received special attention after the 2008 economic crisis, international rules of banking supervisory committees, and the change in the banks' services. According to Ayadi et al [15], the main difference between banking business models and other industries affected is the value they strive to create and the regulations imposed on them. Studies in the field of business models of banks are usually common in three areas: influential variables in clustering, clustering approach, and cluster analysis. By collecting data on variables with a data-driven approach, these studies classify them into different clusters and examine their commonalities, assigning different names to each cluster and analyzing them. The most important studies in this field will be reviewed in table 1.

A review of the literature shows that different studies in this field often use various variables for clustering, apply different clustering methods, and analyze clusters in terms of profitability, efficiency, and risk. Detailed analysis of these papers shows four challenges to the literature: first, dissimilarities in different clustering variables; second, dissimilarities in different clustering approaches; third, diversity in cluster analysis and business models; and, fourth, the importance of the studied countries in the results. In fact, it can be seen that different perspectives on different clusters of business models, influential variables in clustering, and analysis of cluster status have been completely dependent on the available data, the author's view, and the environment governing the region's economy. From this perspective, it is necessary to implement a precise structure for selecting variables, clustering, and groups of business models and carefully analyze their risk and profitability, as well as changes in business models in any economic environment. In short, the most important gaps are the use of a high-confidence method, the inability to generalize the results of the analysis of the banking business model in one country to other countries (especially countries with Islamic financial markets), and not paying proper attention to changes in the bank's business model based on risk. According to this point, in this paper, by identifying clustering variables and different groups of business models, business models for 33 Iranian banks will be examined. In order to select the variables, first, all the variables are summarized, and according to the common groups for business models according to the country of Iran, the variables of clustering implementation will be selected. Also, according to the clustering approaches in this paper, the ensemble approach will be used. In the third section, credit risks, liquidity, insolvency, and equity returns will be examined, and changes in business models and the reasons for these changes will be inspected.

Table 1. Here

3. Methodology

In this section, the presented approach of the study is presented. For this purpose, the variables used for clustering will be presented first. In the following, the methods of performing clustering will be mentioned. The next section will discuss the approach to calculating credit, liquidity, and insolvency risks. The next section mentions the necessary approach to creating clusters, and the change in cluster status will be examined.

3.1. Defining the variables of clustering

Among the different variables identified in the literature, we tried to select the complete set of variables given the nature of Iranian banks and the available data. The variables used to perform the clustering are presented in Table 2.

Table 2. Here

In selecting these variables, an attempt has been made to use a set of different factors that show the bank's structure in terms of assets and liabilities and its profitability. Each variable will be used to calculate the principal variables in the clustering model. Since one of the important parts of this paper is the risk analysis of different banks, these risks and related variables are presented in the next section.

3.2. Analyzing the risks of banks

In order to analyze the financial risks, credit, liquidity, and insolvency risks are considered. This study uses the Loan loss provision to calculate the credit risk to net interest income ratio (*LLPNII*). This ratio is used by Elahi and Poswal [26] to analyze the relationship between profitability and the credit risk of banks. Also, the ratio of liquid assets to demand deposits (*LADD*) is used to analyze the liquidity risk. Sharma [27] used this ratio to analyze the performance of Indian banks. Finally, in order to calculate insolvency risk, the Z-risk index is applied.

$$LADD = -\left(\frac{Liquidity\ Asset}{Demapnd\ Deposit}\right) \tag{1}$$

$$LLPNII = \frac{Loan\ Loss\ Provision}{Net\ Interset\ Income} \tag{2}$$

$$Z-risk = \frac{ROA+CAP}{\sigma(ROA_t)} \tag{3}$$

In this equation, *ROA* is the return on assets, *CAP* is the equity capital to total assets ratio (CAP is a measure to show the capital which is bank provided during large income decreasing), and *σ(ROA)* is the standard deviation of *ROA* [28]. Notably, a lower level of the *Z-risk* index implies a high-risk bank. In order to calculate the risks, the required data are required reserves, non-performing loans, demand deposits, liquid assets (including reserve deposits, investments, bonds, and cash), asset return records, and equity returns.

3.3. The clustering methods

In this section, the clustering method is presented. Clustering algorithms place data with similar properties close to each other into separate categories called clusters. Although most clustering algorithms have the same basis, there are differences in how similarities or distances are measured and the choice of labels for the objects in each cluster.

Despite the different methods, however, there is no more optimal approach to this issue, so usually, different studies try to use several methods or sometimes a combination of them for clustering. In this paper, four approaches of K-Means, FCM, PAM, and Ensemble approach will be used for clustering. We try to use different methods which consider the four characteristics mentioned above. K-means is the simplest method for clustering with low sophistication and one object optimization model, and its computational complexity and time complexity are low. However, the robustness and results of the algorithm are not excellent. FCM is a relatively simple algorithm with low complexity in time. Given that this method is a soft clustering approach and assigns each data to several clusters with one point, it is clear that it needs more calculations and provides a more accurate answer. This is because, in this algorithm, membership values are calculated based on the relative distance (not absolute distance) to the centers of the clusters. This causes the clustering results to be affected by erroneous and fragmented data. Also, depending on the approach type, this method's robustness will be more than the previous approach.

The PAM clustering algorithm is one of the simplest types of the k-medoids algorithm, which converges to the local optimal due to the random selection of representative and non-representative objects at the beginning of the algorithm and does not necessarily produce the optimal answer in clustering. This method has more complex calculations than the previous methods due to the constant change in medoids values, but naturally, the results will be more accurate and stable. Finally, combining several approaches, the ensemble approach seeks to achieve more robust and accurate results [29].

3.3.1. K-Means

The K-means method is a partitional clustering method in which each cluster is connected to its center (Centroid), and each point in the cluster is assigned to the nearest center. In this method, the number of clusters, k, must be specified. Its main algorithm is very simple and as follows [30].

a) For each cluster, compute the centroid, which is a p-dimensional mean vector of the observations in that cluster.

b) Based on the Euclidean distance, assign each observation to the cluster whose centroid is closest

We aim to find a good set of centroids C_k, which minimizes the entropy in each cluster of the partition induced by C_k. The optimization problem is then as follows:

$$\min_{c_k} d(X, C_k) = \sum_{x_i \in X} d(x_i, c(x_i)) \qquad (4)$$

To compute the objective function, an appropriate distance measure d has to be created. This paper focuses on the Mahalanobis metric since it accounts for different variances and the covariance structure within X. The metric is as equation 5:

$$d_M(x_i, c(x_i)) = \sqrt{(x_i - c(x_i))^T \times \sum\nolimits_X^{-1} (x_i - c(x_i))} \qquad (5)$$

In this method, proximity is measured by Euclidean distance. Many convergences occur in the first few iterations, and the pattern continues "until the points change". In order to evaluate the quality of clustering, the silhouette coefficient is applied. It is a criterion measuring the average distance of each observation within a cluster with its neighboring cluster. This criterion is between -1 and 1. The larger silhouette coefficient means the more distant the observation is from the neighboring cluster, and negative values show wrong clustering.

3.3.2. PAM

Partitioning around Medoids (PAM), or k-medoids, is an iterative 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 [31]. Compared with the k-means algorithm, PAM uses an actual data point (medoid) as the cluster center rather than the cluster mean (centroid). In this method, instead of using the center of a cluster instead of a reference, medoid is used. Therefore, its implementation method can be formed like the previous approach on the principle of minimizing the sum of discrepancies between each object and its corresponding reference object. Next, each remaining object is clustered with the medoid, which is most similar. In the next step, this strategy is repeated by replacing objects to improve the quality of the clustering result (the cost function of the average dissimilarity between an object and the median). The execution algorithm is as follows [32]:

1. Selecting J prototype as medoids (C_j) randomly. where j are the clusters ($j = 1,...,J$. with pre-defined J).
2. Assigning data points to the nearest medoid based on the dissimilarity matrix;
3. Computing the sum of all distances to their relevant medoids in the same cluster;
4. Finding a new set of medoids. New medoids are the points that are closest to the data that is in the same cluster;
5. Updating the datapoint assignment to new medoids and computing the sum of distance function;
6. Comparing the new distance function new prototype with the previous and computing total swapping cost;
7. Repeating steps 4 to 6 until the total swapping cost becomes zero or negative.

3.3.3. FCM

The FCM algorithm is a fuzzy clustering method in which each data is assigned to each cluster with a membership value between 0 and 1. In this algorithm, membership values are calculated based on the relative distance (not absolute distance) to the centers of the clusters. This causes the clustering results to be affected by erroneous and fragmented data. Fuzzy logic was used in clustering methods in 1984 by Bezdek et al, [33]. In this method, the membership of each point (i) to each cluster (j) is considered with a fuzzy weight ($\mu^m{}_{i,j}$) and minimization of the sum of squared errors within the group is used to cluster the numbers. The main algorithm is as follows:

1. Initializing the membership matrix ($\mu_{i,j}$) randomly, where i are the observations and j are the clusters ($j=1, \ldots , J,$ with predefined J). The following constraints must be satisfied:

$$\mu_{i,j} = [0,1];\ 1 \le i \le n, 1 \le j \le J \tag{6}$$

$$0 \le \sum_{i=1}^{n} \mu_{i,j} \le n \quad ; \quad 1 \le j \le J \tag{7}$$

$$\sum_{i=1}^{n} \mu_{i,j} = 1 \quad ; \quad 1 \le i \le n \tag{8}$$

2. Calculating the prototype cluster centers (v_j , $1 \le j \le J$) using a predetermined measure of fuzziness ($1 \le m \le \infty$):

$$\overrightarrow{v_j} = \frac{\sum_{i=1}^{n} \mu_{i,j}^{m} \times \overrightarrow{x_i}}{\sum_{i=1}^{n} \mu_{i,j}^{m}} \tag{9}$$

3. Computing the dissimilarity matrix (d^2) (squared Euclidean distance), between the datapoints (x_i) and each cluster center (v_j).

$$d^2(\overrightarrow{x_i}, \overrightarrow{v_i}) = \left\| \overrightarrow{x_i} - \overrightarrow{v_i} \right\|^2 \quad ; \quad 1 \le i \le n \tag{10}$$

4. Updating the previous version of $\mu_{i,j}$; where the denominator is the sum of all weights and is used to normalize the membership scores.

$$\mu_{i,j} = \frac{\left(1 \big/ d^2(\overrightarrow{x_i}, \overrightarrow{v_i}) \right)^{\frac{1}{(m-1)}}}{\sum_{l=1}^{k} \left(1 \big/ d^2(\overrightarrow{x_i}, \overrightarrow{v_l}) \right)^{\frac{1}{(m-1)}}} \tag{11}$$

5. Repeating steps 2 to 4 until that objective function (equation 12) can not be improved:

$$\min J = \sum_{i=1}^{n} \sum_{i=1}^{n} \mu_{i,j}^{m} d^2(\vec{x}_i, \vec{v}_i) \tag{12}$$

3.3.4. Ensemble method

In recent years, the use of combining results in different clustering methods has been developed. This method was developed based on the fact that each method has its abilities, and combining the results will increase the reliability of the final results [34,35]. The final combination model will be very strong and stable because each model creates an attitude of the feature space, and combination creates the ability to use all different attitudes in the model. Three rules are emphasized in ensemble clustering: unanimity, simple majority, and plurality [29]. In the case of unanimity, data belongs to a specific cluster if placed in the same cluster in all clustering methods. In the simple majority, the data belongs to a cluster only if the majority of methods assign a cluster to it. Also, plurality is used when data is assigned to a cluster that receives the most involvement among different clustering methods. Indeed, if data in all clustering methods are placed in a specific cluster, the final output will be the same output. On the other way, when the data is placed in different clusters, the cluster that has selected more methods for the data is selected as the optimal cluster. In order to consider the capabilities of different models, it is necessary to give a different degree of importance (as quality index) to each method [36]. Here, the Dunn index as clustering quality index is applied. This method investigates the clustering condition according to the compactness and separation criteria. Dunn uses two criteria of the distance between clusters and diameter to calculate compactness and separation [37]. The larger value of this index leads to better separation and more effective clustering. Accordingly, if the ratio of the degree of separation to the diameter of the clusters is large, the clustering is well done. Finally, in order to combine the results of the weight of each applied method, the data will be assigned to the final cluster based on the weight driven by the Dunn index.

3.4. Business Model Migration

Given that we are interested in business model changes, we track each bank over its lifetime in the sample to assess whether it switches business model. To ensure we do not identify anomalous migrations (i.e., driven by one-off, extraordinary balance sheet operations), we consider a bank as having changed its business model only if it does not return to the previous business model in the following year. More specifically, we are interested in stable migrations that is when: (a) the bank maintains the same business model for at least two years after migration; or (b) the yearly change in business model refers to a continuous evolution of business models from focused to diversified or vice versa. Next, we analyze migrations by bank business model to account for the possible risk changes in business models.

3.5. The paper's approach

In this section, we propose our model for the study in a flowchart. Figure 2 shows the different steps of the model and details about it.
Figure 2. Here

4. Data and results

In order to implement the proposed model, 33 Iranian banks are chosen, and relevant data are gathered for them. These banks have a total of 21,446 branches in the country in 2021. Most of them do domestic activities and a very limited number are focused on international banking. The activities of these banks include traditional banking, financing, investment services, and specialized fields. To analyze the main problem of the study, almost 28000 data from 2006 to 2021 are gathered from the central bank of Iran. The three clustering methods are coded in Python 3.11 software. In this section, the results of clustering are presented.
First, the descriptive statistics based on full-period data are calculated. The results are presented in Table 3. According to Table 3, the value of the standard deviation in each variable is high. Also, the difference between the mean and the median of the data in each variable is relatively large. This suggests that there are different big and small data in the problem. It can be concluded that the values of the variables are fundamentally different for different banks, so in a preliminary analysis, the existence of several diverse clusters of banks can be predicted. Then, the proposed approach in the previous section is implemented to calculate the clusters.

Table 3. Here

4.1. Identification of the number of business models

In order to check the number of clusters in the problem, the Davies-Bouldin index is used. It measures the clustering status according to the number of selected clusters, the size of scattering within the cluster, and the distance between the clusters. The steps of this method are as follows [38]:

1. Calculate the scatter within the cluster by equation 13:

$$S_i = \left[\frac{1}{|C_i|} \sum_{x \in c_i} d^r(x, c_i) \right]^{\left(\frac{1}{r}\right)}, \quad r > 0 \tag{13}$$

In this equation, *Si* is the scattering rate of clusters *Ci* and *d* are also a function of Minkowski distance.

2. Calculate the amount of dissimilarity between clusters by Equation 14 based on the distance between the center points. In this equation, V_i and V_j are the centers of the two clusters *i* and *j*, and *d* is a function of the Minkowski distance.

$$D_{ij} = \left[\sum d(x, c_i)^t \right]^{\left(\frac{1}{t}\right)}, \quad r > 0 \tag{14}$$

3. Calculate the distance between two clusters based on Equation 15:

$$R_{ij} = \frac{S_i + S_j}{D_{ij}} \tag{15}$$

4. Determine the maximum distance for each cluster relative to the other clusters according to Equation 16:

$$R_i = \max_{i \neq j} R_{ij} \tag{16}$$

5. Calculate the average of maximum distances for all clusters (V*)* based on Equation 17:

$$V_{DB} = \frac{\sum_{i=1}^{k} R_i}{k} \tag{17}$$

6. Select the appropriate number of clusters according to the value of *V*. Notably, lower values of *V* have better clustering.

4.2. Clustering the data

In order to analyze the banks' business model three methods of K-Means, PAM, and FCM mentioned in the previous section were applied. Then using the ensemble clustering approach the results are combined. In the continuation of this research, different analyzes will be performed based on the ensemble clustering method. Previously, the similarity of the results of the ensemble method with other approaches is first examined. To analysis the results,we evaluated the efficiency of each clustering method using Davies-Bouldin criterion. According to this method, selecting four clusters provides an appropriate level of the v-index. In addition, the ensemble method has a very good ability to reduce the v index. Moreover, we examined how much data in different methods were paired in the same cluster. The results are presented in Figure 3. According to this figure, the ensemble method is more suitable than the other three methods. Furthermore, the situation of cluster similarity between this method and other methods is relatively high. In order to accurately evaluate this case, the chi-square index was calculated to analyze the independence of the Ensemble clustering method from other methods. The chi-square values between Ensemble and K-Means are 703.3, Ensemble and PAM are 823.3, and Ensemble and FCM are 913.4. Based on this, it was observed that the assumption of independence of clustering methods is rejected.

Figure 3. Here

The following results of descriptive statistics of different clusters based on the ensemble method are presented. According to table 4, the problem arises in the large difference in all data, including the maximum and minimum, and the high amount of standard deviation, after the implementation of clustering, has been greatly reduced within each cluster.

In order to perform a better analysis, we consider the status of each cluster's variables by considering the bank's size. The number of interbank loan receipts, commission income, operating income, investment amount, and foreign exchange positions are analyzed by modifying the bank size. For this purpose, the ratio of the amounts of deposits to the total bank deposits and loans to the total loans is used. From this perspective, the smallest size in banks belongs to cluster one. The largest cluster also belongs to clusters 3 and 4. By modifying this coefficient, the following results can be examined.

Table 4. Here

According to table 4 clusters 1 and 2 have lower levels of loans (customer loans, industrial and commercial loans, mortgage loans, agricultural loans, Government funding, and Interbank loans) and liabilities (deposits and interbank debt) than clusters 3 and 4. Indeed, clusters 1 and 2 have small or medium sizes, and clusters 3 and 4 have large sizes. Furthermore, the object market of cluster 1 retails, cluster 2 is international, cluster 3 is public, and cluster 4 is corporate. The main income in cluster 1 is investing, cluster 2 is operational income, cluster 3 is commission and fees, and cluster 4 is interest income. The first cluster is banks with small or medium size that seeks to invest in markets, the second cluster are banks with small or medium size which conduct foreign exchange operations, the third cluster is large banks that perform traditional tasks such as lending and receiving deposits, and the fourth cluster is large banks focus on funding. Thus, in the continuation of this paper, cluster 1 belongs to banks with an SME-investing business model, cluster 2 belongs to banks with an SME-operating business model, cluster 3 belongs to banks with a large-public business model, and cluster 4 belongs to the large-funding business model.

4.3. Analyzing the risk of clusters

The results of liquidity, credit, and insolvency risk are presented in Tables 5 to 7. Also, the trend of risk changes during different years for each cluster is presented in Figures 4 to 6. These figures generally show that the highest amount of risk belongs to the first and second clusters. Regarding credit risk, a higher level of risk was observed in the first cluster and then in the second cluster. Regarding liquidity risk, the highest risk was observed in the second and then the first clusters. Also, regarding the insolvency risk, the lowest values of the Z-score were observed in the first and second clusters.

These results confirm the results obtained from the literature regarding the risk of different banking business models. The issue of high credit risk in the investment SMEs and operational SMEs has also been seen in the studies of Ayadi and Geroen [16], Galletta and Mazzu [11], and Lueg et al [20]. In addition, regarding the higher liquidity risk in the second cluster, the results of Sudrajad and Hubner [19] are similar to the presented paper. In addition, regarding the insolvency risk, which has also been examined by Ronegpitya et al [6] and Ayadi and Geroen [16], the insolvency risk in larger banks of the public or funding type is less than the other models.

In terms of the changing trend, credit risk was observed in the years of the global economic crisis of 2008 and inflation due to currency prices in 2017 and 2018. Regarding liquidity risk, a peak in risks has been observed during the currency crisis and the COVID-19 pandemic. This trend has also been observed in changes in the risk of insolvency for the years leading up to 2021.

Table 5. Here

Table 6. Here

Table 7. Here

Figure 4. Here

Figure 5. Here

Figure 6. Here

4.4. Business Model Migration

In order to analyze the changes in business models, clustering output was analyzed by the Ensemble method for each bank in different years. The results are presented in Figure 7. According to this chart, the most changes in the business model occurred in 2008 and 2021, during the economic crisis and the Corona pandemic. The results of this section are similar to the results of Cheng et al [23] and Seetharaman [1] which respectively mentioned the changes in the business model of banks before and after the economic crisis and before and after the COVID-19 pandemic.

Most business model changes have occurred from cluster large-funding to large-public (almost 2%), cluster SME-operating to SME- Investing (3.5%), and SME- Investing to SME- operating (4%). The trend of the data shows that in periods of environmental risk, a shift from more risky business models to lower risk models has been seen (conversion from business model 1 to 2 in 2008 and 2009, and conversion from business model 4 to 3 in 2009 and 2020). There has also been a shift to more risky and cost-effective business models during the stable years (conversion from business model 2 to 1 in 2015 and 2016, conversion from business model 3 to 4 in 2010, and change from business model 3 to 2 in 2017 and 2019). This issue has been seen in the study of Molin [17] and Lueg et al [20]. Also, we check the risk of each of the business models after changing the business model. The results show that in the shock period after changing the business model, banks' liquidity, credit, and insolvency risks decreased by 8.5%, 6%, and 4.3%, respectively.

Figure 7. Here

5. Discussion

This paper presents an analysis of the business models of Iranian banks from 2006 to 2021. Using data and implementing the ensemble clustering method, Iranian banks were assigned to 4 clusters: SME-Investment, SME-Operating, Large-Public, and Large-Funding. In order to discuss the designated clusters, each of the banks located in the four clusters was first examined in terms of size, target market, type of main activity, target income, expected return, and risk. The results of this section are presented in Table 8. Note that the size was considered based on each cluster's volume of receivables and deposits. The results of table 8 show that naming four different clusters in this paper seems logical. In the continuation of this research, the most important variables in clustering were investigated. The results of this section over different years are presented in Table 9. According to this table, the most important variables in this section are loans granted, deposits received, and operating income. The correlation between these three variables with different types of credit, liquidity, and insolvency risks was examined. Figures 8 to 16 show the relationship of these cases to different clusters.

Table 8. Here

Table 9. Here

Figure 8. Here

Figure 9. Here

Figure 10. Here

Figure 11. Here

Figure 12. Here

Figure 13. Here

Figure 14. Here

Figure 15. Here

Figure 16. Here

In the continuation of this section, hypothetical tests were performed regarding the significance of the relationship between different risks and clusters.

Hypothesis A: there is not a significant difference between the LADD of clusters 1 to 4

Hypothesis B: there is not a significant difference between the LLNPII of clusters 1 to 4

Hypothesis C: there is not a significant difference between the Z-Score of clusters 1 to 4

Hypothesis D: there is not a significant difference between the LADD of clusters 1 to 4 during shock periods (2008-2009 and 2019-2021) and stable years (2010-2018)

Hypothesis E: there is not a significant difference between the LLNPII of clusters 1 to 4 during shock periods (2008-2009 and 2019-2021) and stable years (2010-2018)

Hypothesis F: there is not a significant difference between the Z-Score of clusters 1 to 4 during shock periods (2008-2009 and 2019-2021) and stable years (2010-2018)

In order to analyze the hypothesis, Welch's t-test is applied. In this manner, it was checked whether there was a significant difference between these values among the clusters by testing the mean values of LADD, LLNPII, and Z-score. The results are provided in table 10. According to these results, the significant difference in LADD values among clusters 1-3, 1-4, 2-3, and 2-4 were proved; however, no reason was found to accept the assumption of similarity of LADDD values for clusters 1-2, and 3-4. The cases for which the significance of the sameness of the mean was not proved are highlighted and shown in red color in the table. Also, in the other panels of Table 10, the similarity of risk values in each cluster during the shock and stable periods was checked. For example, according to table 10, except for the credit risk in the third cluster, for other clusters, a significant difference in the risk status was discovered in shock and stable periods.

Table 10. Here

6. Conclusions

Analyzing the business models of banks is a challenging issue which can lead to different results in different geographical locations. Also, the high diversity of banks' activities makes this issue necessary to assign banks to suitable clusters based on the similarities of the structure of the balance sheet, income statement, and other bank information. Selecting these variables are very important which requires a comprehensive and targeted perspective to determine its effect on the selected cluster, and choosing a correct method for clustering is of particular importance. As a prominent point, compared to other research in this field, this paper has used three main methods (K-Means, PAM, and FCM) and a combined method to perform clustering (ensemble). In this paper, in order to investigate the Iranian banks' business model and their relevant risks, the data were collected from 2006 to 2021. It was observed that the variance of the data for different banks is high, there is a big difference between their median and average, and the maximum and minimum values are far from each other. Considering this issue, it was tried to allocate banks in different clusters based on their business model. The final results of the ensemble clustering showed that the high spreading of the data had been reduced to a great extent in the new mode within the cluster. Further, by examining the main characteristics of the banks of each cluster, the designated clusters were introduced with four titles: SME-Investing, SME-Operating, Large-Public, and Large-Funding. It seems to the title of the cluster is very close to the nature of the input information of the banks which are assigned to each cluster.

In the next step, the financial risks of the banks of each cluster, including credit risk, liquidity, and insolvency, were calculated and analyzed. The results of this phase showed that SME-

Investing and SME-Operating clusters have a higher level than Large-Public and Large-Funding in all three types of risk. Cluster risk analysis also seems logical based on the subject literature.

In the next part, an analysis of the migration of banks' business models in the studied years was done. Based on the obtained results, most changes in business models have been made between SME-Investing and SME-Operating. Also, the migration from large-public to large-funding and vice versa has also been observed. The results of this section showed that in different years, except for large-public and large-funding clusters, there was usually a risk difference between other clusters. Another point is that the risk values of each cluster significantly differed in the shock and stable period for almost all clusters. However, there are issues regarding the clusters that can be taken into consideration by policymakers. The difference between risk and return in the first and second clusters is fundamentally different from the third and fourth clusters. Considering this limitation that it is mostly not possible to move between clusters 1 or 2 to clusters 3 or 4 in a short time, the major changes was from clusters 1 to 2, 2 to 1, 3 to 4, and 4 to 3. Another important point is that changing the business model cannot be considered a substitute for implementing precautionary measures for risk management. Rather, changing the business model is a relatively lengthy. Therefore, policymakers should consciously follow the change of business model while emphasizing the strict implementation of the Basel Committee's precautionary regulation as another solution. This change can be the transformation of the business model of SME-investing to SME-operating during the shock period such as the pandemic, or the transformation from SME-operating to SME-investing due to inflation shock. Likewise, for larger banks, reducing funding operations and the strategy of changing from the funding business model to the public in the crisis period seems more prudent.

This paper used a set of variables based on the balance sheet and income statements to analyze banks. Future studies can develop another type of business model by using more comprehensive data, including the distribution of asset and debt resources, ownership structure, and the status of service implementation, and by using other variables regarding risk and return to analyze the status of each model. Nevertheless, as mentioned at the beginning of the paper, banking business models can be different with changes in the geographical environment. For example, international banking business models, banking based on new investment tools, and banking based on financial technology, can be considered in another geographical environment. Therefore, future studies in this regard can analyze the risk of these business models by focusing on fintech banking business models [8]. Especially, these new business models have been able to create great influence in traditional banks that support technology by reducing the risk of error, increasing the speed of transactions, and user-friendly services.

References

[1] Seetharaman, P. “Business models shifts: Impact of Covid-19.” *International Journal of Information Management* 54 (2020): 102173.

[2] Merck, M., Rixtel, A. V., & González Mota, E. “Business models of international banks in the wake of the 2007-2009 global financial crisis.” *Estabilidad financiera. Nº 22 (mayo 2012), p. 97-121* (2012).

[3] Ayadi, R. "Banking business models: Definition, analytical framework and financial stability assessment. " Springer, 2019.

[4] Ayadi, R. “Bank Business Models in Europe: Why Does it Matter for the Future of Regulation and Resolution?.” *International Research Centre on Cooperative Finance Policy Paper Jul* (2016).

[5] Beck, T., Demirgüç-Kunt, A., Merrouche, O. “Islamic vs. conventional banking: Business model, efficiency and stability.” *Journal of Banking & finance* 37.2 (2013): 433-447.

[6] Roengpitya, R., Tarashev, N. A., Tsatsaronis, K., et al. “Bank business models: popularity and performance.” (2017).

[7] Ferstl, R., Seres, D. “Clustering Austrian banks’ business models and peer groups in the European Banking Sector.” *Financial Stability Report* 24 (2012): 79-95.

[8] Zhao, J., Li, X., Yu, C. H., et al. “Riding the FinTech innovation wave: FinTech, patents and bank performance.” *Journal of International Money and Finance* 122 (2022): 102552.

[9] Köhler, M. “Which banks are more risky? The impact of business models on bank stability.” *Journal of Financial Stability* 16 (2015): 195-212.

[10] Portes, R. “Interconnectedness: mapping the shadow banking system.” (2018).

[11] Galletta, S., Mazzù, S. “Liquidity risk drivers and bank business models.” *Risks* 7.3 (2019): 89.

[12] Van Oordt, M., Zhou, C. “Systemic risk and bank business models.” *Journal of Applied Econometrics* 34.3 (2019): 365-384.

[13] Chesbrough, H. “Business model innovation: opportunities and barriers.” *Long range planning* 43.2-3 (2010): 354-363.

[14] Richardson, J. E. “The business model: an integrative framework for strategy execution.” *Available at SSRN 932998* (2005).

[15] Ayadi, R., Bongini, P., Casu, B., et al., 2021. “Bank business model migrations in Europe: Determinants and effects.” British Journal of Management, 32(4), pp.1007-1026.

[16] Ayadi, R., De Groen, W. P. “Banking Business Models Monitor 2014: Europe. CEPS Paperbacks, 14 October 2014.” (2014).

[17] Molin, F. “Cluster analysis of European banking data.” (2017).

[18] Lucas, A., Schaumburg, J., Schwaab, B. “Bank business models at zero interest rates.” *Journal of Business & Economic Statistics* 37.3 (2019): 542-555.

[19] Sudrajad, O. Y., Hübner, G. “Empirical evidence on bank market power, business models, stability and performance in the emerging economies.” *Eurasian Business Review* 9 (2019): 213-245.

[20] Lueg, R., Schmaltz, C., Tomkus, M. "BUSINESS MODELS IN BANKING: A CLUSTER ANALYSIS USING ARCHIVAL DATA." *TRAMES: A Journal of the Humanities & Social Sciences* 23.1 (2019).

[21] Marques, B. P., Alves, C. F. "Using clustering ensemble to identify banking business models." *Intelligent Systems in Accounting, Finance and Management* 27.2 (2020): 66-94.

[22] Bonaccorsi di Patti, E., Palazzo, F. "Bank profitability and macroeconomic conditions: Are business models different?." *Economic Notes* 49.2 (2020): e12155.

[23] Cheng, M., Ma, C., Geng, H. "The effects of business models on bank risk before, during and after financial crisis: Evidence from China." *Applied Economics* 52.20 (2020): 2147-2164.

[24] Farnè, M., Vouldis, A. T. "Banks' business models in the euro area: a cluster analysis in high dimensions." *Annals of Operations Research* 305.1-2 (2021): 23-57.

[25] Tran, D. V., Hoang, K., Nguyen, C. "How does economic policy uncertainty affect bank business models?." *Finance Research Letters* 39 (2021): 101639.

[26] Elahi, M., Poswal, B. A. "Factors Influencing Cost Efficiency in Leading Banks "A Comparative Study of Banks Operating in UK and Germany Listed on London Stock Exchange"." *International Journal of Scientific & Engineering Research* 8.1 (2017): 620-639.

[27] Sharma, V. K. "PERFORMANCE EVALUATION OF STATE BANK OF INDIA AND ITS ASSOCIATE BANKS THROUGH CAMEL ANALYSIS." *CLEAR International Journal of Research in Commerce & Management* 8.3 (2017).

[28] Rahman, A. A. "Financing structure and insolvency risk exposure of Islamic banks." *Financial Markets and Portfolio Management* 24.4 (2010): 419.

[29] Kuncheva, L.I. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2014.

[30 Ikotun, A. M., Ezugwu, A. E. "Boosting k-means clustering with symbiotic organisms search for automatic clustering problems." *Plos one* 17.8 (2022): e0272861.

[31] Kaufman, L., Rousseeuw, P. J. "Finding groups in data: an introduction to cluster analysis." John Wiley & Sons, 2009.

[32] Maechler, M., Rousseeuw, P., Struyf, A., et al. 2012. Cluster: cluster analysis basics and extensions.

[33] Bezdek, J. C., Ehrlich, R., Full, W., et al. "FCM: The fuzzy c-means clustering algorithm." *Computers & geosciences* 10.2-3 (1984): 191-203.

[34] Ala'raj, M., Abbod, M. F. "A new hybrid ensemble credit scoring model based on classifiers consensus system approach." *Expert systems with applications* 64 (2016): 36-55.

[35] Abellán, J., Castellano, J. G. "A comparative study on base classifiers in ensemble methods for credit scoring." *Expert systems with applications* 73 (2017): 1-10.

[36] Badhani, S., Muttoo, S. K. “CENDroid—A cluster-ensemble classifier for detecting malicious Android applications.” *Computers & Security* 85 (2019): 25-40.
[37] Ncir, C. E. B., Hamza, A., Bouaguel, W. “Parallel and scalable Dunn Index for the validation of big data clusters.” *Parallel Computing* 102 (2021): 102751.

[38] Ramadhani, S., Azzahra, D., Tomi, Z. “Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student’s Thesis.” *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi* 13.1 (2022): 24-33.

Authors’ biography

Sepideh Aghajani

Sepideh Aghajani received her BSc degree in Industrial Engineering in 2012 from Alzahra University, Tehran, Iran and her MSc degree in Industrial Engineering in 2014 from Tehran University, Tehran, Iran. She is now PhD Candidate of Industrial Engineering in Alzahra University, Tehran, Iran. Her main research interests are supply chain management, financial risks, Data mining, Decision Science, and business analysis.

Reza Samizadeh

Dr Reza Samizadeh holds his BSc and MSc Degrees from Sharif University of Technology, Tehran, Iran and he also received his PhD degree in IT from Tarbiat Modarres University. He is now faculty member of industrial engineering department at Alzahra University, Tehran, Iran. His research interests are Data Mining, Decision Science, CRM, IT, and business analysis.

Table 1. The summary of the literature.
Table 2. Defined variables of the study
Table 3. Descriptive statistics results
Table 4. Descriptive statistics of different clusters in ensemble clustering
Table 5. The results of credit risk (LLPNII) for different clusters during 2006-2021
Table 6. The results of liquidity risk (LADD) for different clusters during 2006-2021
Table 7. The results of insolvency risk (Z-Score) for different clusters during 2006-2021
Table 8. The main characteristics of different clusters
Table 9. Analyzing the effect of each variable on the clustering
Table 10. The results of Welch’s t-test about the significant difference between clusters’ risks
Figure 1. The general structure of the banks’ business model problems
Figure 2. The flowchart of the proposed approach
Figure 3. The similarity of clustering results between different methods and ensemble model
Figure 4. Changes of LLPNII for each cluster during 2006-2021
Figure 5. Changes of LADD for each cluster during 2006-2021
Figure 6. Changes of Z-Score for each cluster during 2006-2021
Figure 7. The changes in the business model of banks from 2006 to 2021
Figure 8. Analysis of the relationship between LLNPII and Customer loans for different clusters
Figure 9. Analysis of the relationship between LADD and Customer loans for different clusters
Figure 10. Analysis of the relationship between Z-score and Customer loans for different clusters

Figure 11. Analysis of the relationship between LLNPII and Customer deposit for different clusters
Figure 12. Analysis of the relationship between LADD and Customer deposit for different clusters
Figure 13. Analysis of the relationship between Z-score and Customer deposit for different clusters
Figure 14. Analysis of the relationship between LLNPII and operating income for different clusters
Figure 15. Analysis of the relationship between LADD and operating income for different clusters
Figure 16. Analysis of the relationship between Z-score and operating income for different clusters

Table 1. The summary of the literature.

Paper	Clustering variables	Clustering groups	Clustering method	Area of analysis	Analysis of risk	Analysis of profitabilit	Analysis of migrating	Other main findings
Ferstl et al [7]	Net interest income, trading income, income from fees and commissions, operating income, customer deposits, total loans, total assets	Five clusters	K-centroid	European and Australian banks	-	-	-	The business models in European and Australian banks are different
Beck et al. [5]	fee income, non-deposit funding, operational income, deposits, overheads, total assets	Islamic and conventional	-	Islamic countries	✓	✓	-	Islamic banks are better capitalized, have higher asset quality, and are less risky
Roengpitya et al. [6]	Gross loans, trading book, Interbank lending, wholesale debt, stable funding, deposits	Wholesale funding, retail funding, trading	Statistical clustering	International	✓	✓	-	The banks' business model between emerging and advanced economics are different
Ayadi and Geroen [16]	Ownership and financial activity	Investment, wholesale, diversified retail, focused-retail	Ward	European countries	✓	✓	-	The financial performance and risk of different groups are not similar.
Molin [17]	Net fee and commission income, total deposits, total assets, operating income, loan to retail/ corporate/ banks	Universal banking with a wholesale focus and universal banking with a retail focus	Calinski-Harabasz criterion	European countries	-	-	✓	-
Lucas et al. [18]	Leverage, loan to assets, trading assets, derivative, net interest income, net fees and commissions, trading income	Large universal, diversified international lenders, fee-focused, diversified domestic lenders, domestic retail lenders, small	Time-varying component median	European countries	✓	-	-	-

Paper	Clustering variables	Clustering groups	Clustering method	Area of analysis	Analysis of risk	Analysis of profitabilit	Analysis of migrating	Other main findings
		international						
Galletta and Mazzu [11]	loans, investments, deposits	Commercial, investment, saving	Classification	European countries	✓	-	-	Saving banks and commercial banks have more liquidity risk than investment banks
Sudrajad and Hubner [19]	Fee and commissions, trading and derivatives, deposit from banks, cash collateral, loans, assets	Commercial, investment, mortgage, saving	Classification	Asian countries	✓	✓	-	Bank stability based on the Z-index, ROA and ROE volatility are different for different class.
Lueg et al. [20]	loans, deposits, trading assets, interbank liabilities, fees and commissions income, operating income, tangible common equity, total assets	Retail, investment, universal	Ward	European and American countries	-	-	✓	-
Marques and Awes [21]	Gross loan to customers, trading assets, interbank lending and borrowing, customer deposits, size, wholesale funding, total derivatives, income, leverage, diversification	Retail-focused, retail diversified-focused, retail diversified asset model, large diversified	FCM/ SOM/ PAM	European countries	-	-	-	-
Patti and Palazzo [22]	Net income pretax, operating income, net interest income, fee income, trading income, operating expenses, asset write-downs, loans and credit impairments, securities, cash holding, retail deposits, capital, assets	-	Ward	European countries	✓	✓	-	There is a significant relationship between the bank’s business model and macroeconomic conditions.

Paper	Clustering variables	Clustering groups	Clustering method	Area of analysis	Analysis of risk	Analysis of profitabilit	Analysis of migrating	Other main findings
Cheng et al. [23]	Total equity, total assets, impaired loans, loan provisions, liquid assets, non-interest income, non-deposit funding, foreign investment	-	-	China	✓	-	-	-
Ayadi et al. [15]	Bank loans, derivatives, trading assets, debt liabilities, customer loans, bank liabilities, customer deposits	Investment, wholesale, focused retail, diversified retail type 1, diversified retail type 2	Ward	European countries	✓	✓	✓	The migrating business model for banks using a logistic regression method in two groups (migrating and non-migrating) is analyzed.
Farne and Vouldis [24]	derivative (for hedge or trading, off-balance sheet items, liabilities, total assets, available for sale assets, loans, trading assets	Wholesale funded, securities holdings, traditional commercial complex commercial	The robustified version of factorial K-means	European countries	✓	✓	-	-
Tran et al. [25]	-	small, medium, large	-	American Banks	-	✓	-	Small and medium banks in a state of economic policy uncertainty increase their net interest income activities, while the large banks do not change their policies.

Table 2. Defined variables of the study

Assets related variables	Debts related variables	Income related variables	Other variables
Customer loans (CL) Governmental funding (GF) Interbank loans (IL) Reserve deposits (RD) Bonds Investments in different markets (IM) Fixed assets (FA) Cash Trading assets (TrA) Liquid assets (LA) Total assets (ToA)	Debt to the central bank (DCB) Debt to banks (DB) Customer deposits (CD) Total equity (TE) Total debts (TD)	Operating expenses (OE) Net income pretax (NIP) Operating income (OI) Net interest income (NII) Trading income (TI) Fee and commission income (FCI) Non-interest income (NoII) Capital	Foreign investment (FI) Agriculture loans (AgL) Industrials loans (InL) Commercial loans (CoL) Mortgage loans (MoL)

Table 3. Descriptive statistics results

	Average	**Stdev**	**Min**	**Max**	**Median**
CL	251549/157	241539/364	1150/238	2672335	92864/5
ICL	190769/120	252655/919	0	5013468	36214/75
MoL	54382/227	132543/046	0	1536664	6459/75
AgL	20290/725	55422/036	0	924265	343/5
GF	44837/432	119668/450	0	3023317	0
IL	27515/109	44042/772	0	758977	6677/75
FI	297643/638	704716/152	0	8980889	3340/75
TrA	18658/521	25813/235	98/531	511562	4081
RD	34536/368	40488/282	59/863	380126	13761/75
ToA	465446/471	495240/882	1299	7560794	147564/5
DCB	45299/670	74003/061	0	615055/61	2846
DB	15691/154	21050/660	0	210522/39	4146/22125
CD	209730/005	238096/815	128/612	1848829	74923/5
TE	50778/660	127765/653	2159/31	4063154	11981/5
OI	29678/690	37225/588	365/251	636122	11611/75
NoII	3207/412	4564/089	0	41332	1430/25
NII	8949/775	17426/847	3/12697	277073	2219/75
LLP	41881/346	70418/810	111/0953	2072267	11682
LA	70523/280	77168/369	1981/2089	1041218	15336/5
ROA	0/5298	0/8644	-0/9088	3/0758	0/2753
LLPNII	15/7345	35/8021	0/3267	97/1636	4/0933
LADD	1/6110	4/9916	0/1044	37/2441	0/1852
Zit	1/9293	3/1630	0/0756	8/7784	1/2484

Table 4. Descriptive statistics of different clusters in ensemble clustering

	Cluster1		**Cluster 2**		**Cluster 3**		**Cluster 4**	
Number of banks	**196**		**108**		**53**		**41**	
	Average (SD)	**Max (Min)**	**Average (SD)**	**Max (Min)**	**Average (SD)**	**Max (Min)**	**Average (SD)**	**Max (Min)**
CL	136456(227047)	1081(1511995)	201549(241539)	1150(2172335)	406642(256032)	1219(3832675)	261549(241539)	1150(3172335)
ICL	116492(260236)	0(4563872)	218923(303187)	0(6616162)	285046(245076)	0(5863064)	352615(202125)	0(6010774)
MoL	45681(111336)	0(1290798)	47313(115312)	0(1336898)	58452(149774)	0(1736430)	66083(153750)	9083(1782530)
AgL	18262(49880)	0(831839)	18870(51542)	0(859566)	20711(59302)	0(988964)	23320(60964)	5012(1016692)
GF	17663(100521)	0(20395)	31563(137619)	0(74768)	72011(138815)	0(3507048)	58112(101718)	0(2569819)
IL	19166(46685)	0(504515)	22742(52411)	0(50318)	65864(41400)	0(713438)	42287(35675)	0(614771)
FI	239314(514314)	0(6238116)	439587(803376)	0(9238211)	255974(606056)	0(7723565)	355974(589621)	0(5523313)
TrA	22017(30460)	116(603643)	19300(21167)	81(419481)	24847(29685)	113(588296)	18860(21941)	84(434828)
RD	30737(36035)	53(338312)	28665(33605)	50(315505)	38335(44942)	66(421940)	40408(47371)	70(444747)
ToA	355351(651426)	609(7492700)	512256(837981)	610(10039640)	632170(750819)	666(11567840)	628062(767150)	636(9847635)
DCB	32581(69563)	0(578152)	21675(68083)	0(565851)	68017(78443)	0(651959)	58923(79923)	0(664260)
DB	13965(18735)	0(187365)	13181(17683)	0(176839)	17417(23366)	0(233680)	18202(24419)	0(244206)
CD	177146(223811)	121(1737899)	142952(219049)	118(1700923)	272314(252383)	136(1959759)	256508(257145)	139(1996735)
TE	43670(109878)	1857(3494312)	46716(117544)	1987(3738102)	54841(137987)	2332(4388206)	57888(145653)	2462(4631996)
OI	19237(27612)	119(330119)	33537(42065)	413(718818)	25820(32386)	318(553426)	29679(37226)	365(636122)
NoII	2290(3971)	0(35959)	2151(4199)	0(38025)	4624(5157)	0(46705)	3464(4929)	0(44639)
NII	9039(17601)	3(279844)	9308(18124)	3(288156)	8860(17253)	3(274302)	8592(16730)	3(265990)
LLP	48164(80982)	128(2383107)	42300(71123)	112(2092990)	35599(59856)	94(1761427)	41463(69715)	110(2051544)
LA	61355(67136)	1724(905860)	64881(70995)	1823(957921)	76165(83342)	2140(1124515)	79691(87200)	2239(1176576)
ROA	0/567(0/925)	0(3/291)	0/62(1/011)	0(3/599)	0/493(0/804)	0(2/86)	0/44(0/717)	0(2/553)
LLPNII	18/095(41/172)	0/376(111/738)	16/364(37/234)	0/34(101/05)	13/374(30/432)	0/278(82/589)	15/105(34/37)	0/314(93/277)
LADD	1/74(5/391)	0/113(40/224)	1/627(5/042)	0/105(37/617)	1/482(4/592)	0/096(34/265)	1/595(4/942)	0/103(36/872)
Zit	1/621(2/657)	0/064(7/374)	1/698(2/783)	0/067(7/725)	2/238(3/669)	0/088(10/183)	2/161(3/543)	0/085(9/832)

Table 5. The results of credit risk (LLPNII) for different clusters during 2006-2021

		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
C1	AVERAGE	4/8707	10/156	7/3029	5/9566	11/8324	5/2391	13/6923	34/9151	36/2414	25/6757	18/3524	44/9394	45/0871	6/7616	4/8615	2/1303
	STDEV	3/1565	28/937	7/6704	12/868	15/176	6/4059	14/8671	9/1757	6/9956	12/0930	11/109	118/8079	87/9514	4/2127	3/2102	4/2260
C2	AVERAGE	10/679	19/327	5/6239	16/996	12/487	30/0667	33/057	7/0324	8/0194	9/4801	8/0756	24/694	16/6456	3/7785	3/385	-0/2253
	STDEV	15/4837	30/433	4/8583	18/460	10/357	56/4464	58/9921	7/6499	3/2948	3/6769	3/7576	90/0761	29/8828	9/3100	6/641	7/8075
C3	AVERAGE	5/7973	6/8711	22/5162	19/3004	10/4357	5/6475	6/4063	2/2413	2/7033	1/9211	1/6765	29/847	26/868	2/3917	3/2425	4/9589
	STDEV	7/5993	6/2011	19/5714	7/0413	11/8596	2/1947	2/0664	0/5979	0/7710	0/5294	0/8985	44/0664	299/304	2/2562	2/3254	3/6501
C4	AVERAGE	3/3683	17/45	7/2329	10/2387	12/6919	5/3968	10/1041	5/1781	4/5551	7/1119	6/5512	23/7554	29/8602	1/3355	1/1261	1/9255
	STDEV	0/6037	6/0825	7/4579	5/9332	8/3906	0/0287	7/9593	136/9112	33/8015	604/5592	6/1967	165/6793	208/009	3/6921	1/0451	0/1432

Table 6. The results of liquidity risk (LADD) for different clusters during 2006-2021

		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
C1	AVERAGE	0/5896	0/3171	2/7069	1/4239	1/0883	0/5279	0/2375	0/3894	0/4603	0/1937	0/1834	0/2811	0/3846	1/8462	2/6036	2/9537
	STDEV	1/0209	0/4139	8/1478	3/0519	1/9464	0/9905	0/2888	0/7853	1/3589	0/2039	0/2174	0/3309	0/6959	1/4088	2/4643	47/6988
C2	AVERAGE	0/1611	0/1646	0/1953	0/2053	0/1948	1/336	0/1583	0/078	0/0865	0/0981	0/0847	0/1772	0/1827	1/3146	1/4551	1/4692
	STDEV	0/0157	0/0080	0/0453	0/0745	0/0573	0/1332	1/2219	0/0383	0/0153	0/0502	0/0529	0/0318	0/1231	0/1844	1/0598	0/9650
C3	AVERAGE	0/107	0/1287	0/1827	0/2456	0/1859	0/1725	0/1856	0/2053	0/2033	0/1939	0/1825	0/1542	0/1826	0/8365	1/0993	0/5012
	STDEV	0/0329	0/0736	0/0221	0/0847	0/0256	0/1083	0/0291	0/0190	0/0222	0/0233	0/0118	0/0472	0/0824	3/3945	2/1885	0/2762
C4	AVERAGE	0/1872	0/1722	0/1649	0/1548	0/1558	0/0849	0/6393	0/1715	0/144	0/1273	0/1387	0/2052	0/244	0/9261	0/7543	0/6366
	STDEV	0/0181	0/0271	0/0566	0/0358	0/0247	1/4694	0/1019	0/0158	0/0093	0/0033	0/0189	0/0572	0/0777	1/2317	0/4321	0/1275

Table 7. The results of insolvency risk (Z-Score) for different clusters during 2006-2021

		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
C1	AVERAGE	1/1416	0/4927	0/2155	0/8922	0/1519	0/0849	0/9252	0/1103	1/4538	2/0374	1/6447	0/485	1/9908	0/4704	0/4748	0/646
	STDEV	1/4340	1/2909	1/0812	1/1790	1/2553	0/9905	0/1507	1/1721	1/4028	0/5098	1/3168	4/0471	0/9990	6/6614	15/8740	2/8931
C2	AVERAGE	0/0425	0/4472	0/4343	0/601	0/1528	0/1725	0/8822	0/7408	1/2757	1/4919	2/0422	0/3317	1/3313	0/4317	0/4802	0/6306
	STDEV	0/1437	0/1702	0/3523	0/2590	0/0875	0/1332	1/6403	0/2333	0/1790	0/1328	0/5426	0/2520	0/7428	0/1661	0/1215	1/1447
C3	AVERAGE	0/174	1/3035	1/4987	1/8835	0/5905	0/5279	2/868	0/4079	1/9148	4/0353	2/5335	1/6941	1/5856	1/2581	1/2171	1/7334
	STDEV	0/1286	0/1462	0/0503	0/1687	0/0966	0/1083	0/0050	0/5393	0/2090	1/2205	0/1112	0/1446	0/0532	1/0356	0/6011	0/7583
C4	AVERAGE	0/1174	0/5104	0/06895	0/0838	0/1282	1/336	2/5869	1/9723	1/845	3/45	1/0986	0/2308	1/2162	0/4865	0/3771	2/0312
	STDEV	0/1015	0/0167	1/8585	1/5343	0/0557	1/4694	0/0455	0/1418	1/1684	2/7789	0/2212	0/1545	0/2300	1/2753	4/2926	1/0520

Table 8. The main characteristics of different clusters

	C1	**C2**	**C3**	**C4**
Size	small	small	large	Large
Market	retail	international	public	Corporate
Type	investment	operating	loan/deposit	Funding
Income	interest	operational	fee and commission	Interest
ROA	high expected	high expected	Less expected	Less expected
Risk	riskiest	risky	low risk	low risk

Table 9. Analyzing the effect of each variable on the clustering

	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**Average**
CL	0/95	0/80	0/88	0/90	0/97	0/90	0/95	0/94	0/94	0/92	0/79	0/84	0/90	0/68	0/86	0/86	0/88
GF	0/75	0/57	0/76	0/79	0/89	0/70	0/75	0/66	0/65	0/68	0/78	0/68	0/59	0/67	0/64	0/64	0/70
IL	0/66	0/66	0/44	0/46	0/42	0/66	0/66	0/78	0/65	0/68	0/72	0/58	0/74	0/58	0/56	0/56	0/61
RD	0/73	0/54	0/46	0/45	0/42	0/72	0/73	0/69	0/78	0/73	0/67	0/70	0/79	0/61	0/67	0/67	0/65
ToA	0/90	0/48	0/69	0/48	0/49	0/82	0/90	0/88	0/89	0/88	0/88	0/85	0/94	0/79	0/83	0/83	0/78
TrA	0/64	0/64	0/64	0/63	0/82	0/50	0/64	0/59	0/65	0/64	0/57	0/64	0/77	0/28	0/27	0/27	0/58
DCB	0/77	-0/14	-0/16	-0/15	-0/14	0/73	0/77	0/71	0/66	0/75	0/84	0/72	0/69	0/81	0/80	0/80	0/53
DB	0/68	0/57	0/83	0/84	0/81	0/82	0/68	0/60	0/50	0/58	0/48	0/72	0/69	0/81	0/80	0/80	0/70
CD	0/79	0/75	0/88	0/86	0/95	0/76	0/79	0/78	0/84	0/79	0/78	0/70	0/88	0/62	0/60	0/60	0/77
TE	0/80	0/77	0/96	0/71	0/91	0/80	0/80	0/70	0/66	0/75	0/77	0/02	0/39	0/45	0/43	0/43	0/65
ICL	0/63	0/79	0/45	0/57	0/63	0/61	0/63	0/60	0/56	0/60	0/62	0/73	0/90	0/63	0/67	0/67	0/64
MoL	0/74	0/54	0/76	0/89	0/74	0/70	0/74	0/69	0/66	0/65	0/74	0/53	0/03	0/51	0/49	0/49	0/62
AgL	0/45	0/81	0/90	0/90	0/98	0/53	0/45	0/36	0/40	0/23	0/25	0/31	0/07	0/37	0/36	0/36	0/48
FI	0/58	0/76	0/63	0/64	0/71	0/65	0/58	0/52	0/52	0/41	0/24	0/70	0/72	0/48	0/66	0/66	0/59
OI	0/90	0/87	0/76	0/79	0/91	0/78	0/90	0/91	0/93	0/85	0/79	0/85	0/75	0/71	0/57	0/57	0/80
NoII	0/69	0/58	0/92	0/90	0/84	0/63	0/69	0/61	0/56	0/62	0/60	0/50	0/59	0/24	0/23	0/23	0/59

Table 10. The results of Welch's t-test about the significant difference between clusters' risks

Welch's t-test		LADD				Welch's t-test		LLNPII				Welch's t-test		Z-Score			
		C1	C2	C3	C4			C1	C2	C3	C4			C1	C2	C3	C4
LADD	C1		2.67	**4.98**	**5.53**	LLNPII	C1		1.67	3.11	**4.23**	Z-Score	C1		**3.66**	**4.43**	**5.66**
	C2			**4.06**	**4.78**		C2			**3.66**	**3.59**		C2			**5.02**	**6.29**
	C3				**3.42**		C3				2.05		C3				**3.53**
	C4						C4						C4				
Welch's t-test		LADD- Stable period				Welch's t-test		LLNPII- Stable period				Welch's t-test		Z-Score- Stable period			
		C1	C2	C3	C4			C1	C2	C3	C4			C1	C2	C3	C4
LADD- Shock period	C1	**7.66**				LLNPII- Shock period	C1	**9.12**				Z-Score- Shock period	C1	**9.97**			
	C2		**8.42**				C2		**5.65**				C2		**6.59**		
	C3			**3.87**			C3			3.27			C3			**4.43**	
	C4				**4.09**		C4				**4.29**		C4				**5.12**

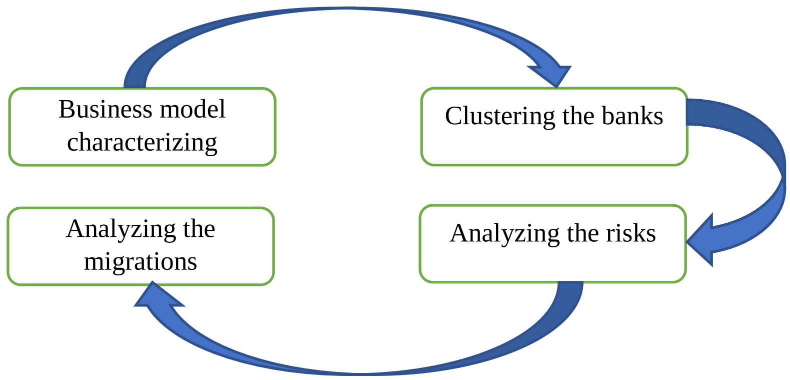

Figure 1. The general structure of the banks' business model problems

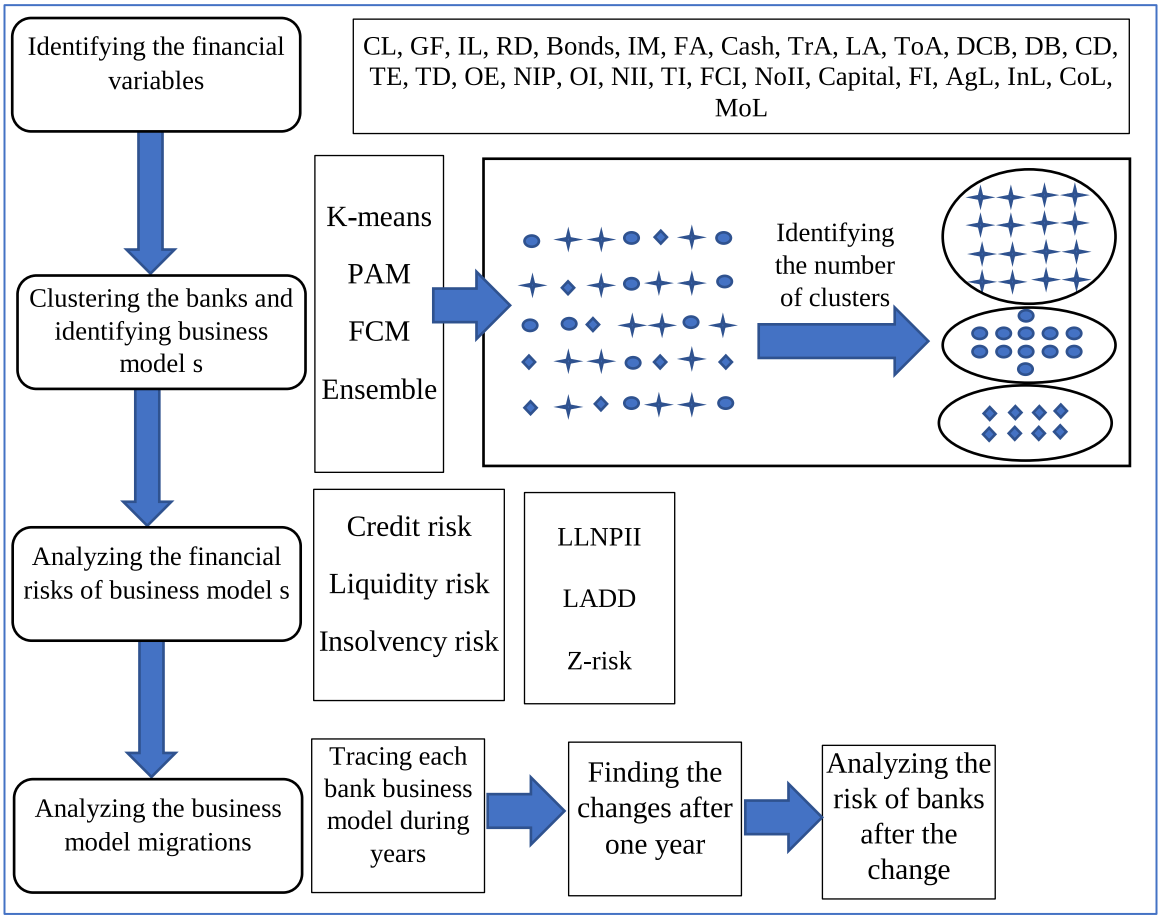

Figure 2. The flowchart of the proposed approach

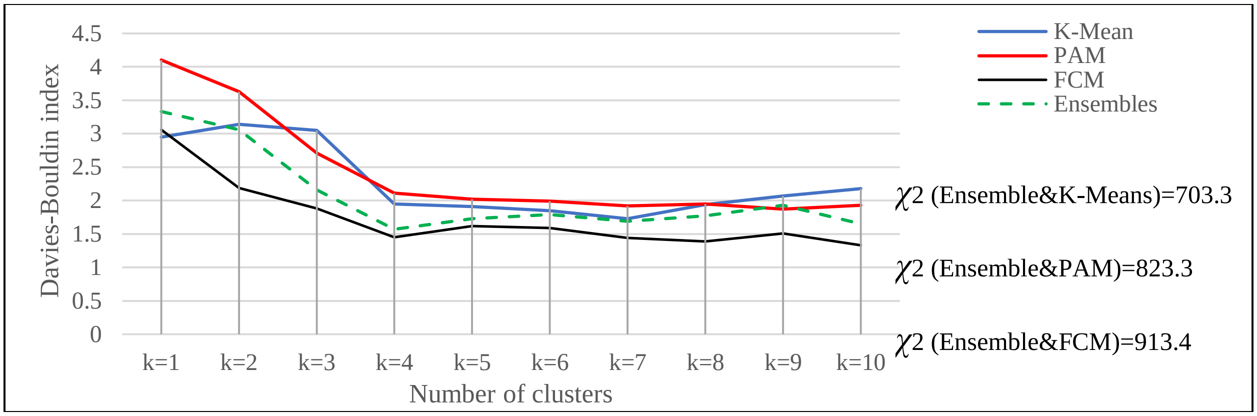

Figure 3. The similarity of clustering results between different methods and ensemble model

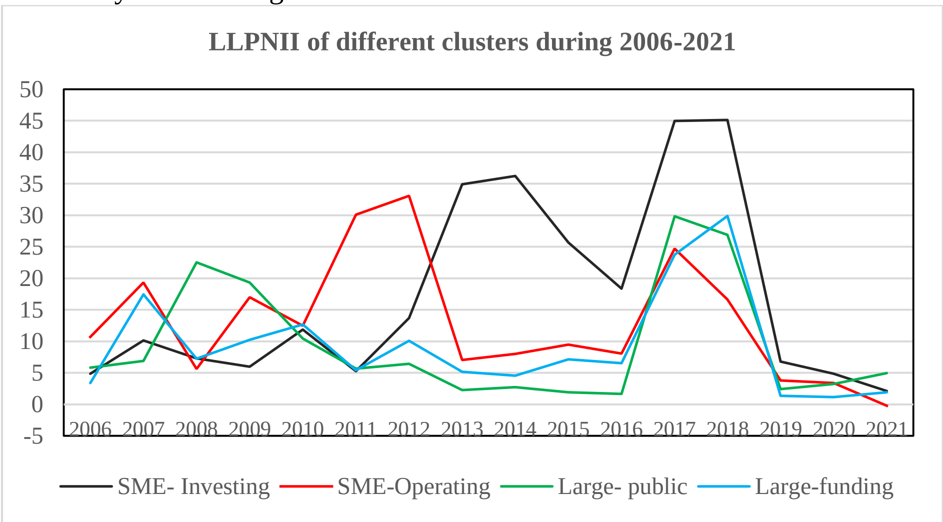

Figure 4. Changes of LLPNII for each cluster during 2006-2021

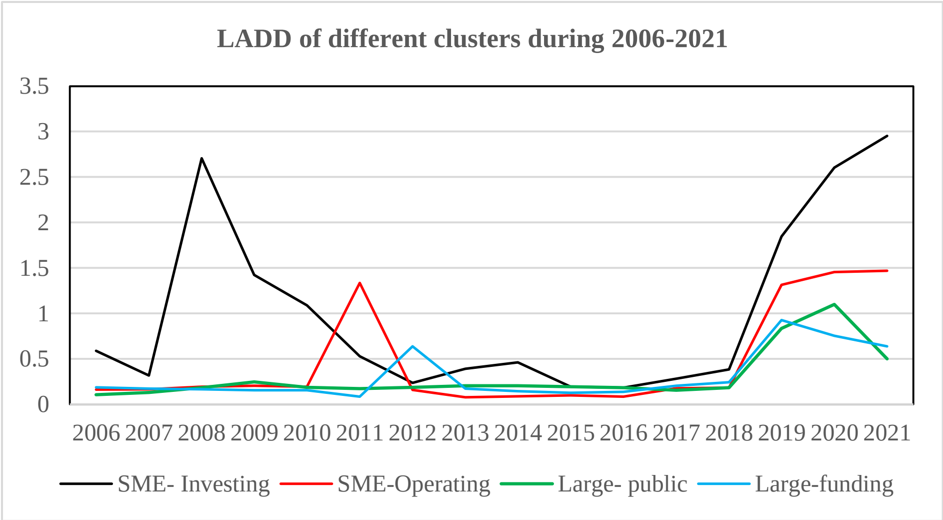

Figure 5. Changes of LADD for each cluster during 2006-2021

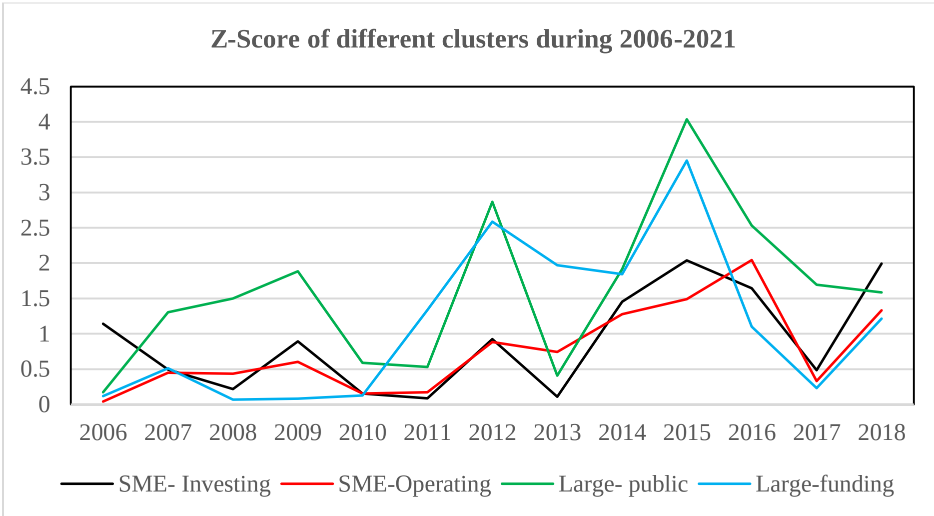

Figure 6. Changes of Z-Score for each cluster during 2006-2021

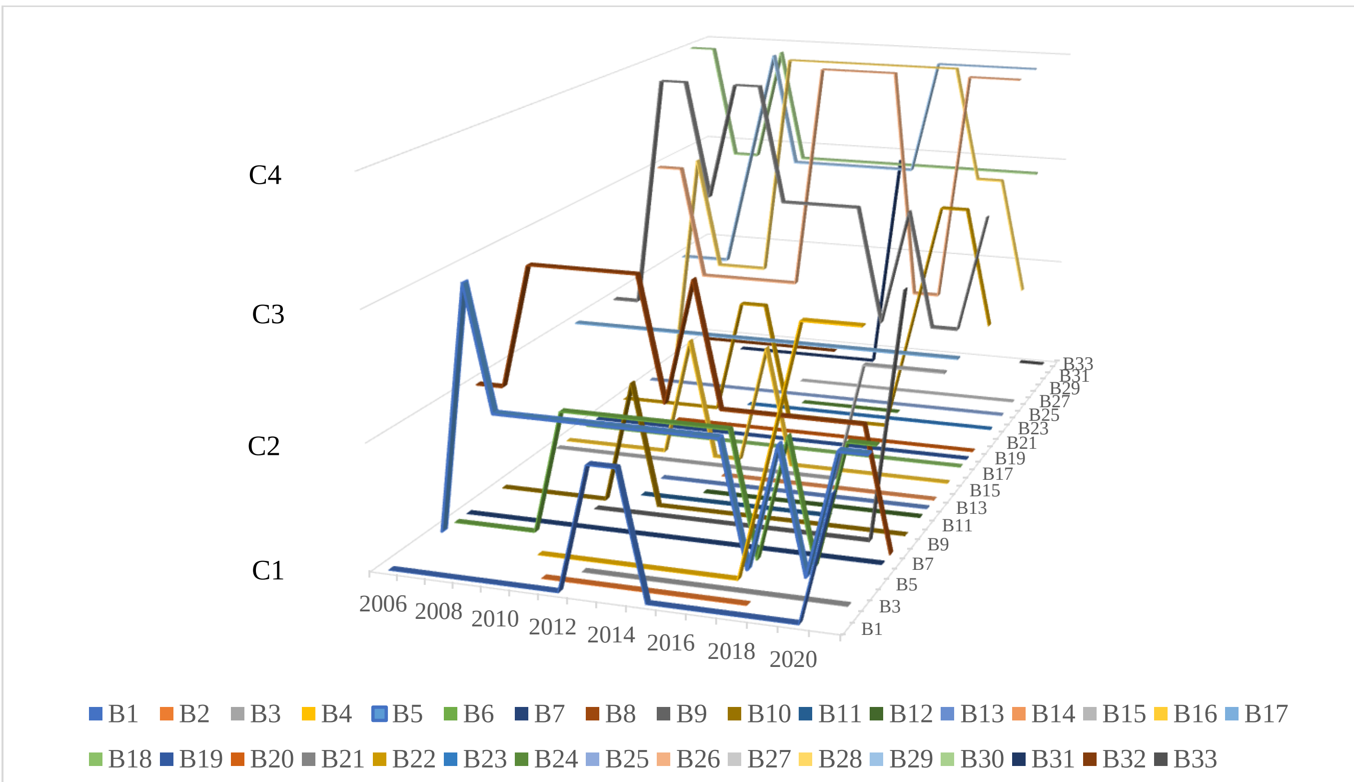

Figure 7. The changes in the business model of banks from 2006 to 2021

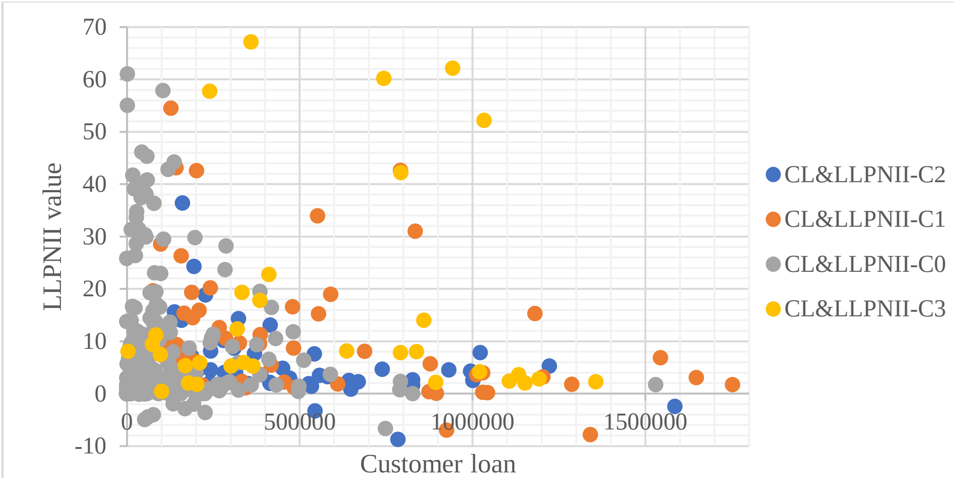

Figure 8. Analysis of the relationship between LLNPII and Customer loans for different clusters

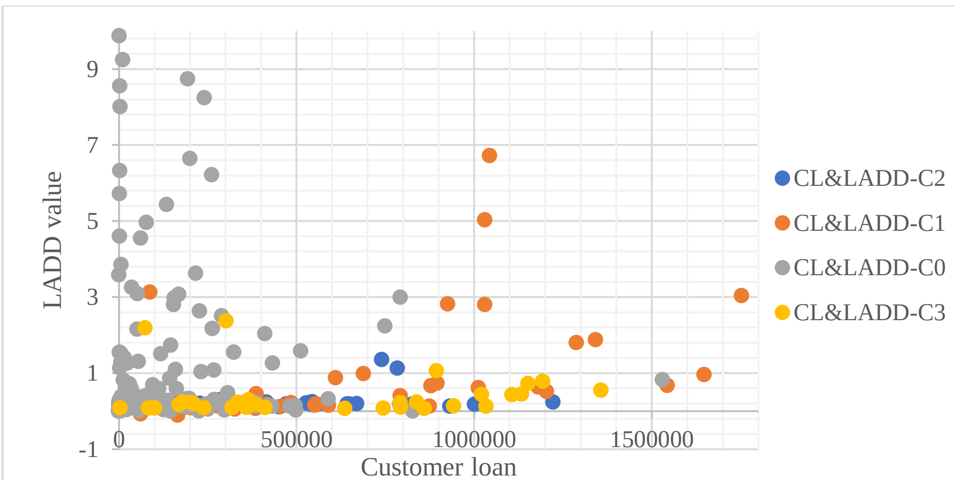

Figure 9. Analysis of the relationship between LADD and Customer loans for different clusters

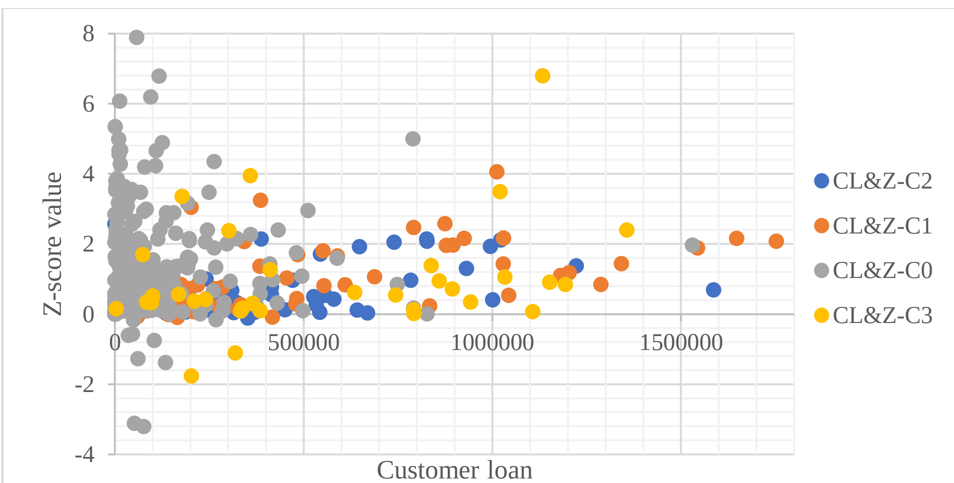

Figure 10. Analysis of the relationship between Z-score and Customer loans for different clusters

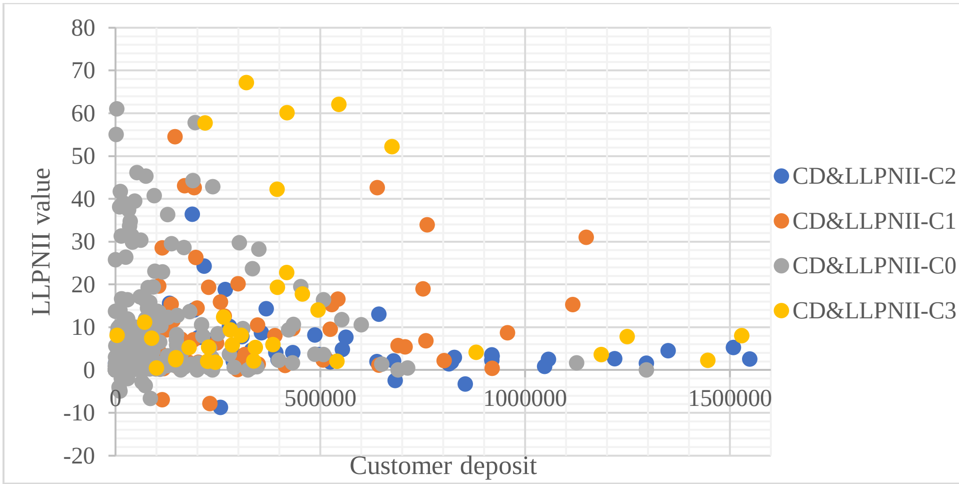

Figure 11. Analysis of the relationship between LLNPII and Customer deposit for different clusters

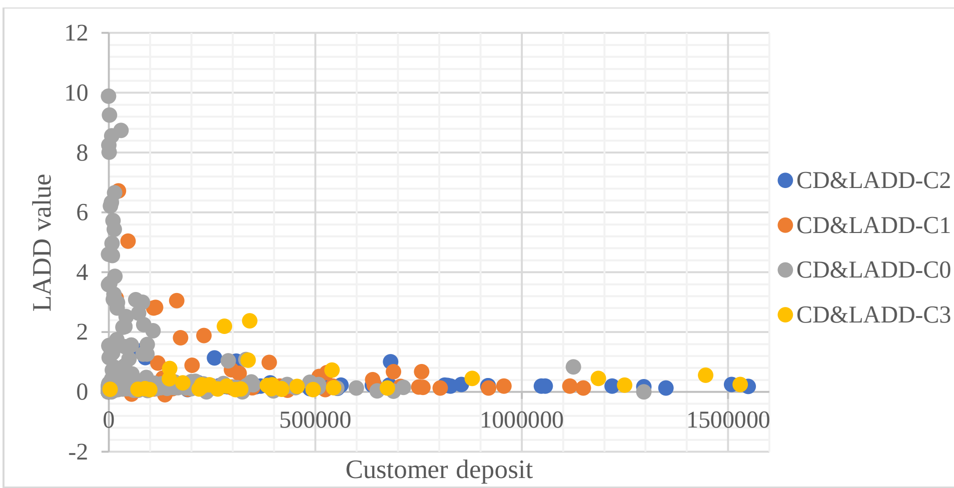

Figure 12. Analysis of the relationship between LADD and Customer deposit for different clusters

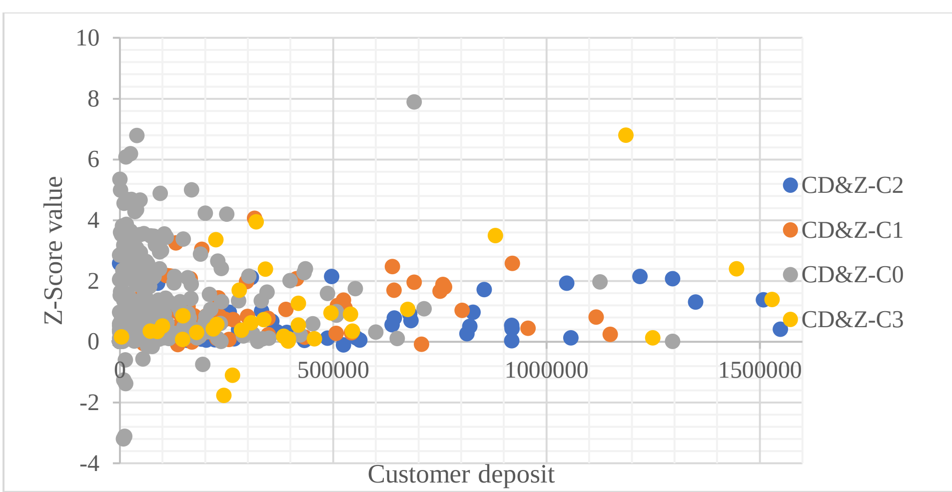

Figure 13. Analysis of the relationship between Z-score and Customer deposit for different clusters

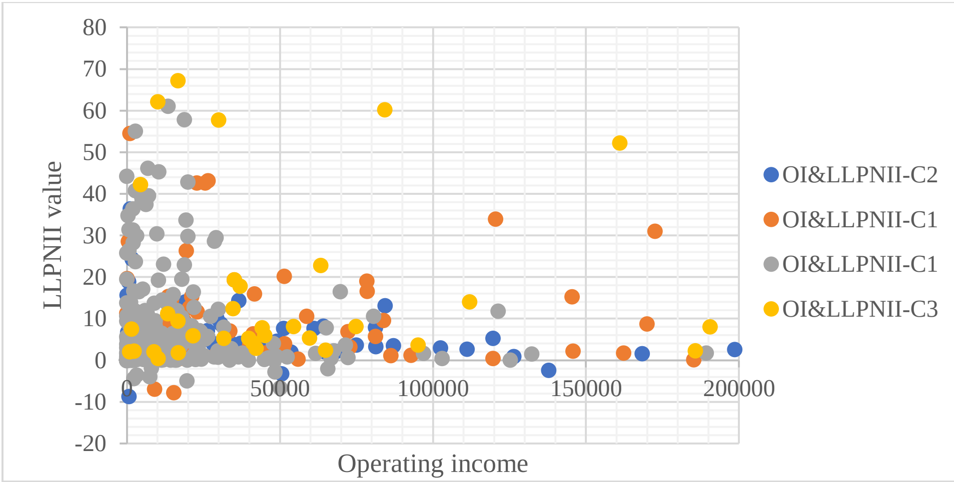

Figure 14. Analysis of the relationship between LLNPII and operating income for different clusters

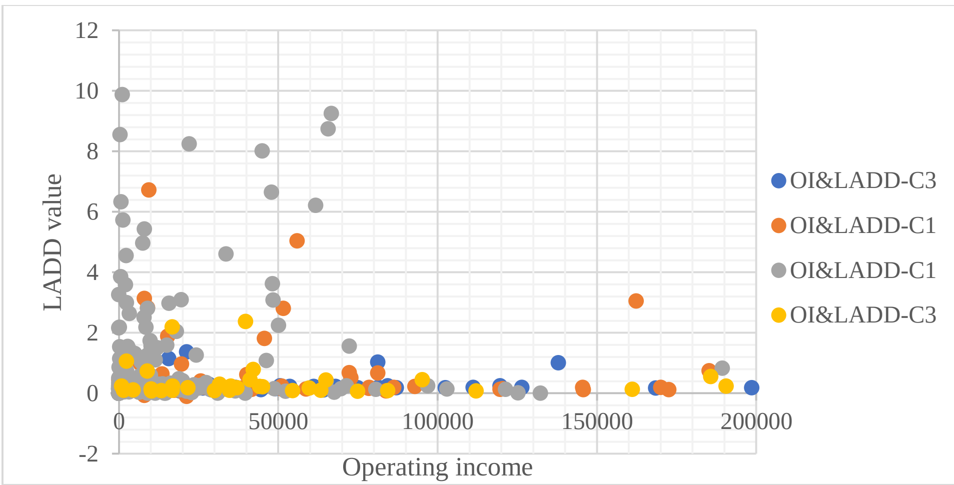

Figure 15. Analysis of the relationship between LADD and operating income for different clusters

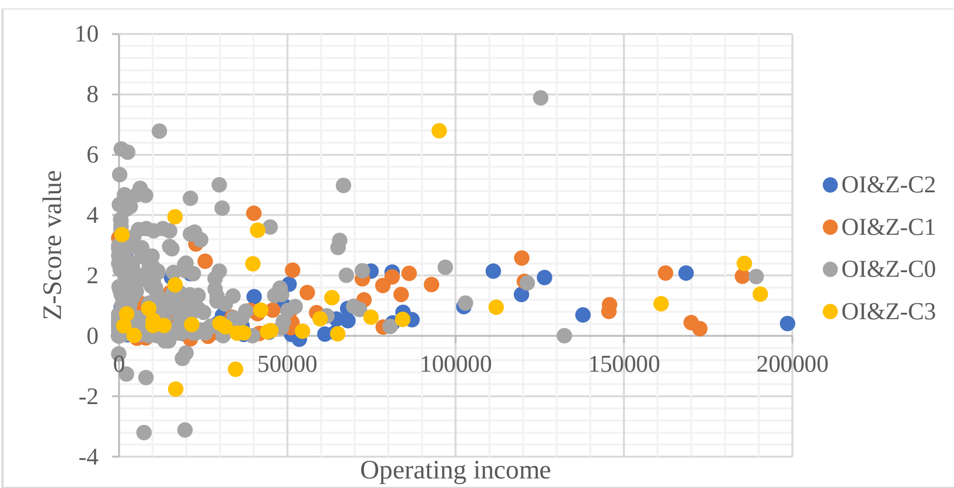

Figure 16. Analysis of the relationship between Z-score and operating income for different clusters