Financial Health of United States Banks in 2016

Huafeng Zhang

Montana, the USA

Abstract

This study reports a k-means clustering analysis investigating patterns of financial health of all 5,870 banks in the United States in 2016 by using five key variables: net interest margin, return on assets, net charge-offs to loans, tier 1 capital ratio and total risk-based capital ratio. The results suggest that there are nine groups of banks in terms of their financial health. 1,224 banks are found be in the cluster that represented the most unhealthy institutions, whereas only 19 banks are in the cluster described as most healthy. The banks that failed in the first quarter of 2017 are all found in the first cluster.

Keywords: Banks, financial health, k-means cluster, PCP, geographical distribution, United States

1. Introduction

Banks play a vital role in keeping the economy going by attracting savings and granting loans. It is essential for economists as well as politicians to understand what banks do, and how healthy their financial systems are. A bank's financial health can be measured by its investment effectiveness, profitability, loan and debt health, capital stability, solvency.[1] This study aims to understand financial health by performing a clustering analysis of all the banks in the United States based on their financial health in 2016.

With help from Erika Refsland, an MSU graduate student who worked as a bank examiner in office of the Comptroller of the Currency for two years, I decided to use five variables as indicators in measuring the financial health of banks in the contiguous U.S.: net interest margin, return on assets, net charge-offs to loans, core capital ratio and total risk-based capital ratio. The definitions of the five variables from the Federal Deposit Insurance Corporation (FDIC) are stated in Table 1.[2] Following Table 1 are short discussions of the significance of the five variables.

Definitions of the Variables from FDIC	
Variables	**Definition**
Net interest margin	Total interest income less total interest expense (annualized) as a percent of the average of all loans and other investments that earn interest or dividends
Return on assets	Net income after taxes and extraordinary items (annualized) as a percent of the average total assets
Net charge-offs to loans	Gross loan and lease financing receivable charge-offs, less gross recoveries, (annualized) as a percent of average total loans and lease financing receivables
Tier 1 capital ratio	A percent of average total assets minus ineligible intangibles
Total risk-based capital ratio	Total risk based capital as a percent of risk-weighted assets as defined by the appropriate federal regulator for prompt corrective action during that time period

Net interest margin (NIM) measures a bank's investment effectiveness. A negative NIM suggests that the bank has not invested its funds efficiently. But often banks that have a higher NIM are actually riskier, as they could be loosening their underwriting standards and originating more risky loans. A moderate NIM may therefore be a better indicator of good health in a bank than a high NIM.

Return on assets (ROA) measures a bank's annual profitability. It usually gives investors an idea of how effectively a bank uses its assets to make a profit. The higher ROA is, the more profitable the bank is.

When **Net charge-offs to loans** (NCOL) is high, the bank has "bad debt" or "poor quality loans" that are unlikely to be recovered by the bank. A negative NCOL usually indicates that in the corresponding year, the bank has more recoveries than charge-offs.[3]

Tier 1 capital ratio (CCLR) measures a bank's capital stability,which is expressed as a ratio of a bank's average total consolidated assets and certain off-balance sheet exposures. The higher the CCLR is, the higher is the likelihood of the bank withstanding negative shocks to its balance sheet.

Total risk-based capital ratio (TRBCR) measures how well a bank prepares to protect its investors and depositors from riskiness. The Dodd-Frank Wall Street Reform and Consumer Protection Act (2010) requires that every bank's TRBCR is at least 8% as a response of the 2008 Financial Crisis.[4]

In this study, I assume that each of the five variables is equally important in measuring financial health, and banks with a healthy financial system have a moderately effective investment strategy, high profitability, a low net charge-offs to loans ratio, stable capital, and a high risk-based capital ratio.

2. Data

I collected three data sets for this study, and all of them are collected from the FDIC website.[2] The first raw data set (*Banks16.csv*), consists of 5,922 banks in the United States, which are identified and ordered by their FDIC certificate number (cert), and sixteen variables, five to measure financial health, and 11 for additional information about the banks. The second data set (*Assets.csv*) only includes a bank's cert and its total assets in 2016. The third data set (*Failed.csv*) includes the information for the banks that failed from 2000 to the first quarter of 2017 in the U.S. The banks in all of the data sets are ordered by its unique FDIC certificate number. Most of the information of interest is included in the first data set. The whole analysis was conducted using R, version 3.3.2. [5]

There were 85 missing values in the first data set, all from the columns of the five variables of interest. As indicated in Table 1, the greatest number of NAs was 48 for the variable NCOL. I removed the banks with missing values because the missing values are relatively few (85), only 0.2% ($\frac{85}{5{,}922*5}$ = 0.2%) of the total number of the five variables values of interest in the data set. This process removed 48 banks.

A parallel coordinates plot (Figure 1) was created using lattice (Deepayan, 2008)[6]. Figure 1 suggests that there are three banks with extremely high TRBCR, and another bank's ROA is also a lot bigger than other banks. Three of the four banks are smaller than over half of the 5,874 banks (their assets: $0.017, $0.138, $0.06 billion; the median asset of the 5,874 banks: $0.205 billion). Although one of the four banks is larger (assets: $0.217 billion), all of their assets are much smaller than the mean assets ($2.855 billion), which indicates that these four banks have small shares of the financial markets. So it is reasonable to remove these four banks in an effort to avoid the clustering of 5,870 banks distorted by the four small banks. This process returned the final version of the data set that was used to do clustering analysis.

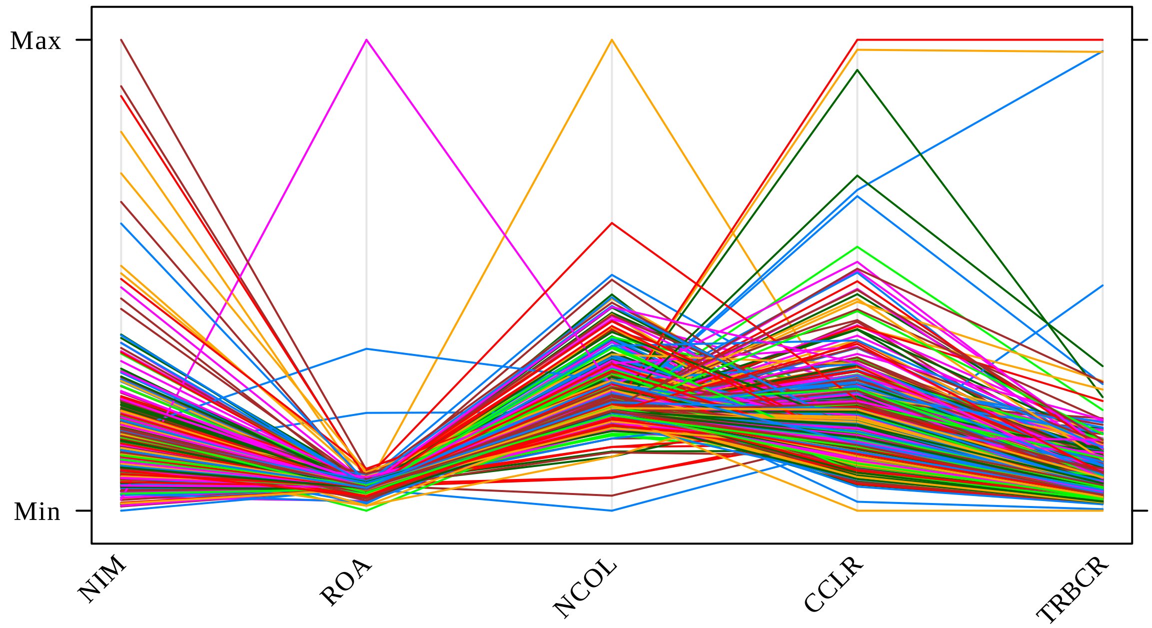

Figure 1: PCP for the Raw Data

Table 1 below, provides the summaries of the five variables in the cluster analysis.

	Variable	Size_withoutNAs	n_NAs	Mean_withNAs	Mean_withoutNAs	SD_withoutNAs
1	NIM	5860	10	3.68	3.70	1.09
2	ROA	5861	9	1.03	0.94	1.28
3	NCOL	5822	48	0.17	0.17	0.54
4	CCLR	5861	9	11.68	11.27	3.89
5	TRBCR	5861	9	26.31	18.84	10.02

Table 1: Summaries of the Five Variables

The correlation plot created using corrplot (Wei Simko, 2016)[7] as well as the table of the correlation coefficients (Table 2) shows that CCLR and TRBCR are highly correlated with each other (the correlation coefficient betweeen CCLR and TRBCR is 0.75). This is reasonable because TRBCR is defined as the the sum of CCLR and other capital ratios. The difference between CCLR and TRBCR is that TRBCR measures a bank's level of protection, whereas CCLR focuses on its capital stability.

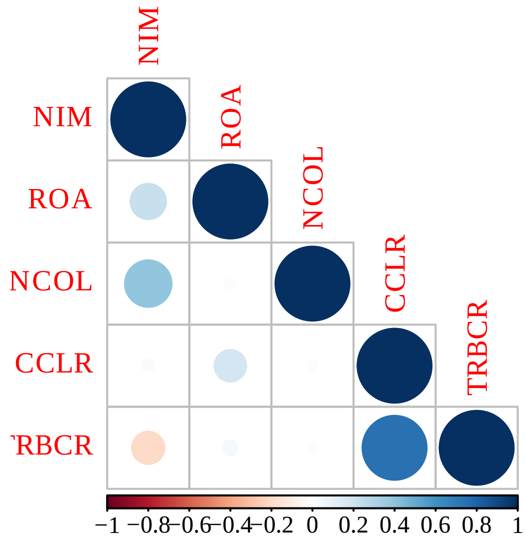

Figure 2: Correlation Plot

	NIM	ROA	NCOL	CCLR	TRBCR
NIM	1.00	0.23	0.40	0.03	-0.19
ROA	0.23	1.00	-0.02	0.18	0.04
NCOL	0.40	-0.02	1.00	0.02	-0.01
CCLR	0.03	0.18	0.02	1.00	0.75
TRBCR	-0.19	0.04	-0.01	0.75	1.00

Table 2: the Correlation Coefficients

3. Methods

This study employs k-means clustering, which is widely used in the financial field.[8] This method partitions observations into k clusters defined by their centroids. Given 5,870 banks in the data set, a k-means clustering method is a good starting point for exploratory analysis because of its simplicity. We can change to use other clustering techniques depending on the results of the k-means clustering. Before doing clustering, I scaled all five variables of interest using function ***scale*** in R to avoid the bias that is caused by the difference in magnitudes of the variables.

After getting the scaled data, I used the Elbow method to find the optimal number of clusters, since a small sum of squared errors within each cluster is needed for an optimized cluster. As the number of clusters increases, the sum of squared errors (SSE) within each cluster decreases. But it is not appropriate to use a big number of clusters (almost close to the total number of observations) since my goal is to find and compare clusters instead of comparing individual banks to one another. The Elbow method can help me find an optimal number of clusters, that is, a small value of k (the number of clusters) that still has a low SSE.

The figure below suggests that k=9 is an optimal choice for clustering. Because when there are less than nine, the variance explained by the clusters dropped sharply, and when there are more than nine, the sum of the variance explained decreases slowly, which indicates that nine should be my optimal choice for the number of clusters. k=6 also appears to be a reasonable choice, however, the resulting PCP plot showed no clear pattern in the clustering, hence this choice would not be as useful as k=9 for the economic analysis.

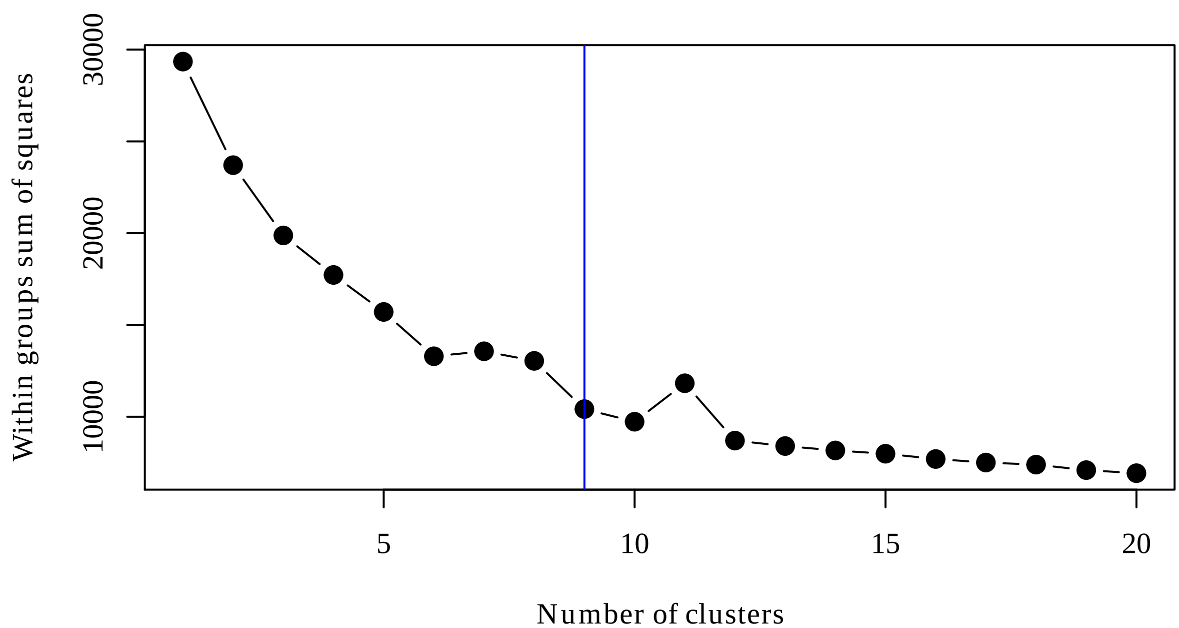

Figure 3: The Elbow Method

After determining the number of clusters, a k-means clustering analysis was conducted by using function ***kmeans***. Table 2 below displays the cluster summary, and a parallel coordinates plot (Figure 4) was created using lattice (Deepayan, 2008)[6] to visualize the characteristics of the banks in each cluster.

	Cluster	mean(NIM)	mean(ROA)	mean(NCOL)	mean(CCLR)	mean(TRBCR)	Size
1	1	3.87	0.93	0.12	9.48	13.70	2075
2	2	3.69	1.11	0.05	12.36	19.31	1143
3	3	18.28	10.29	4.35	18.62	26.59	12
4	4	2.96	0.52	0.11	9.34	17.02	1224
5	5	3.01	0.60	0.11	22.29	52.82	161
6	6	3.16	2.95	0.05	45.73	97.98	19
7	7	4.95	1.61	0.28	11.33	16.17	587
8	8	4.87	0.09	3.12	11.38	18.25	72
9	9	3.15	0.76	0.13	15.24	30.80	577

Table 3: Cluster Summary

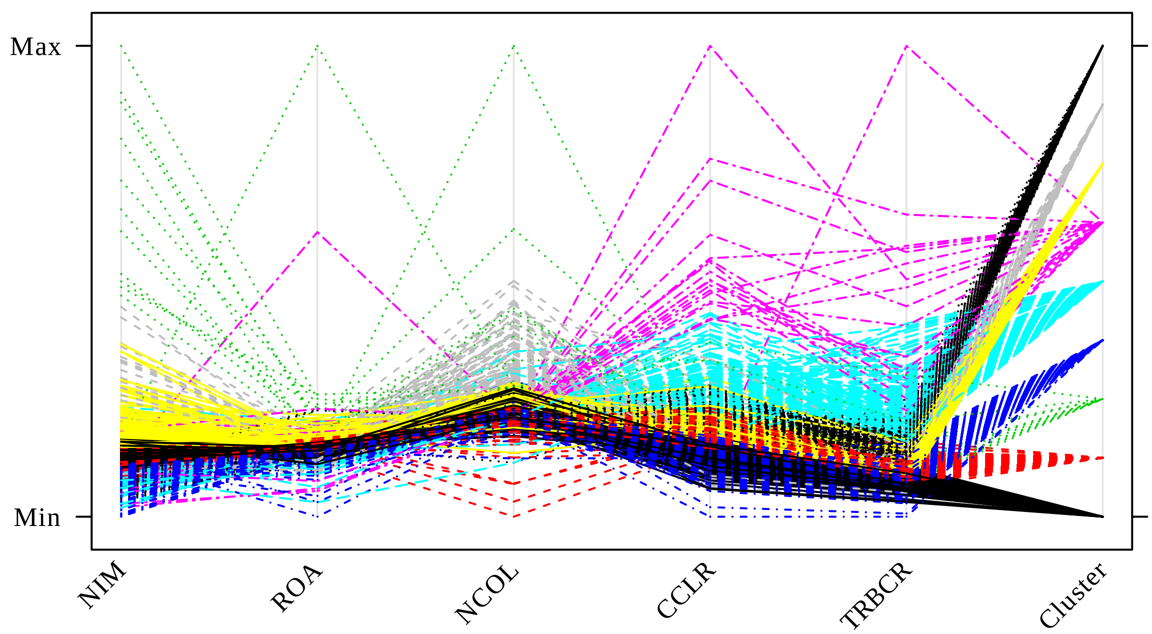

Figure 4: PCP for the 9 Clusters

4. Results

Cluster 1 (black solid line in Figure 4) has the most banks 2,075 banks. These banks' mean NIM (3.87) and ROA (0.93) are about average, but they have the lowest mean TRBCR (13.70), and the second lowest CCLR (9.48) and the third lowest mean NCOL (0.12) among the nine clusters, which suggests that these banks' investment effectiveness and profitability are at average level, even though they don't have much "bad debt" or "low quality loans" relative to their total loans or stable capital, and their protection for their investors and depositors is the worst amongst the nine clusters. I would describe these banks as “normal in making investments and profits, but not reliable or stable.”

Cluster 2 (red line) with 1,143 banks, has the lowest mean NCOL (0.05). Its mean NIM (3.69), ROA (1.11), and TRBCR (19.31) are about average level. This indicates that the banks in Cluster 2 have similar profile to those in Cluster 1, but they have lower mean net charge-offs to loans.

Cluster 3 (green line) with the fewest banks (12), has the highest mean NIM (18.28), ROA (10.29) , and NCOL (4.35), the third highest mean CCLR (18.62) and fourth highest mean TRBCR (26.59). This shows that these 12 banks are the most profitable with the most effective investment strategy, but also potentially the most risky since their mean net charge-offs to loans is the highest, and they didn’t prepare as much risk based capital as they would need for their investors and depositors. I would call these 12 banks “most profitable but also most risky.”

Cluster 4 (blue line) with 1,224 banks has the lowest mean NIM (2.96), CCLR (9.34), the second lowest mean ROA (0.52), NCOL (0.11), and average level of mean TRBCR (26.59). These suggest that these banks have the least effective investment strategy and least stable capital, and they are nearly the lowest in profitability with only average level protections for their investors and depositors, even though they don’t have much net charge-offs relative to their loans. I would describe these banks as “most unhealthy financially.”

Cluster 5 (light blue line) with 161 banks, has the second lowest mean NCOL (0.11), the second largest mean CCLR (22.29) and TRBCR (52.82), even though they have medium to low mean NIM (3.01) and ROA (0.60). This suggests that these banks have more stable capital and better protection for their investors and depositors compared to banks in other clusters except Cluster 6. They also have the lowest mean net charges-offs to loans. However, their strategy of using assets and investments still needs to improve compared to banks in other clusters (except Cluster 4). I would call these banks “less effective, profitable, but stable, reliable.”

Cluster 6 (pink line) has 19 banks has the highest mean CCLR (45.73), TRBCR (97.89), the lowest mean NCOL (0.05), the second largest mean ROA (2.95), and the second lowest mean NIM (3.16). Despite the low NIM (which indicates ineffective investments), they still have the second highest profitability, the most stable capital, most risk-based capital to protect their investors and depositors, and the lowest percentages of bad debts to loans (net charge-offs). Therefore, we can call these banks the “most healthy" financially.

Cluster 7 (yellow line), with 587 banks, has the second highest mean NIM (4.95), the third highest mean ROA (1.61), NCOL (0.28), and the third smallest mean CCLR (11.33), TRBCR (16.17). This indicates that these banks have relative effective investments, "regular" profits, moderate net charge-offs to their loans. Their capital is relatively unstable, and is less reliable for their investors and depositors. I will use "on track, but potentially growing risky" to describe these 587 banks in terms of their financial health.

There are 72 banks in **Cluster 8** (grey line). It has the third highest mean NIM (4.87), the lowest mean ROA (0.09), the second highest mean NCOL (3.12), and the fourth smallest mean CCLR (11.38), TRBCR (18.25). This suggests that the banks in this cluster are the least profitable, but otherwise are not distinctive. I will describe this cluster as "least profitable."

Cluster 9 (black dash line), with 577 banks, has the third lowest effective investments (mean NIM=3.15), the fourth lowest profitability (mean ROA = 0.09), the third lowest proportion of bad debt out of their total loans (mean NCOL =3.12), the fourth most stable capital, and the fourth highest proportion of the risk-based capital to protect their investors and depositors out of their risk-weighted assets (mean CCLR = 15.24, TRBCR = 18.25). I would call these banks "on track."

FDIC data show that the three banks that have failed in the first quarter of 2017 are all from Cluster 4, the one that was found to be "most unhealthy financially." The map created using ggplot2 (Wickham, 2009)[9] shows the geographical distribution of banks in the contiguous U.S. (other regions have many fewer banks and no failures). The banks on the map are colored according to their clusters they belong to based on my k-means analysis. The three big black dots represent the three banks that failed in the first quarter of 2017. There is no obvious geographical association with financial health of banks, including the three failed banks.

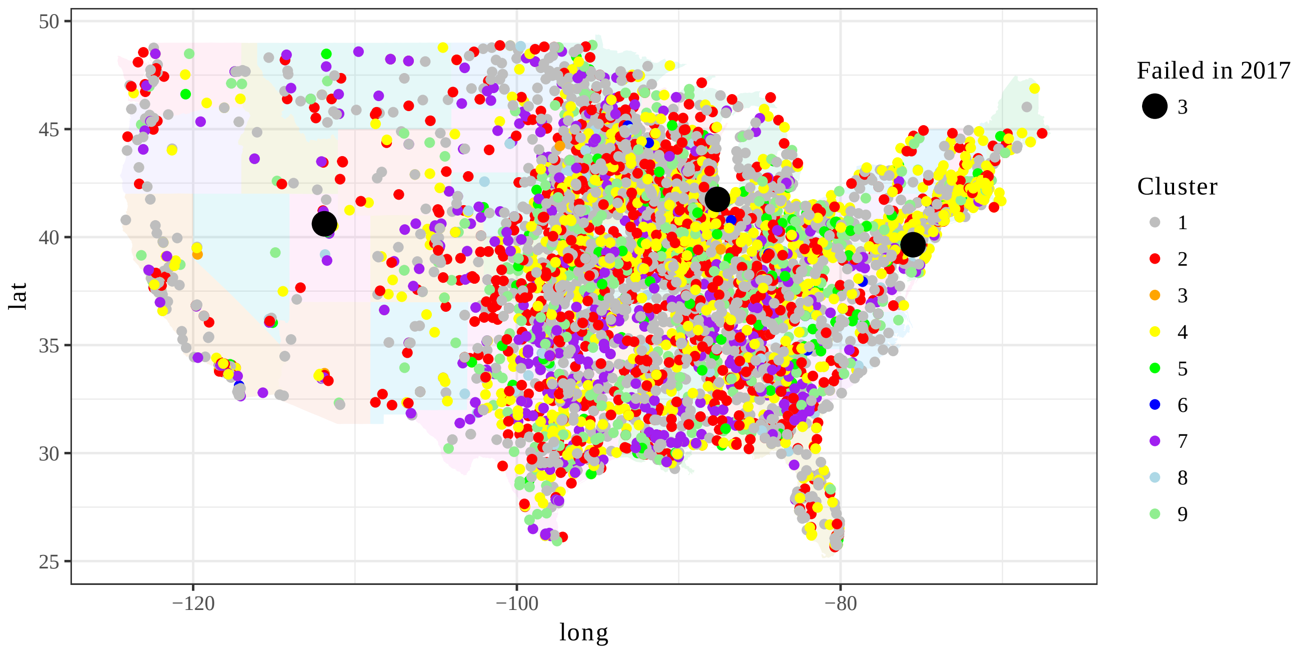

Figure 5: the U.S. Banks in 2016

5. Conclusion

For this study, I analyzed all 5,870 banks in the United States in 2016 using five indicators of financial health: moderate effective investment strategy, high profitability, a low net charge-offs to loans ratio, stable capital, and a high risk-based capital ratio. Using k-means clustering resulted in one cluster described as the most unhealthy banks. Indeed, all bank failures are from this cluster. It will be interesting to see if future bank failures follow the same pattern.

6. Reference

[1] Necas, S. (2016). Financial health of a commercial insurance company and its coherences. Trendy Ekonomiky a Managementu, 10(25), 22-32.doi:http://dx.doi.org.proxybz.lib.montana.edu/10.13164/trends.2016.25

[2] Each depositor insured to at least $250,000 per insured bank. (2017). Retrieved April 7, 2017, from https://www5.fdic.gov/sdi/main.asp?formname=customddownload

[3] Ashcraft, A. (2005). Are Banks Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks. The American Economic Review, 95(5), 1712-1730. Retrieved from http://www.jstor.org.proxybz.lil

[4] An Act (2010). Retrieved April 11, 2017, from https://www.gpo.gov/fdsys/pkg/PLAW-111publ203/html/PLA 111publ203.htm

[5] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[6] Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer,New York. ISBN 978-0-387-75968-5

[7] Taiyun Wei and Viliam Simko (2016). corrplot: Visualization of a Correlation Matrix. R package version 0.77. https://CRAN.R-project.org/package=corrplot

[8] Yu-Jie Wang, Hsuan-Shih Lee, A clustering method to identify representative financial ratios, Information Sciences, Volume 178, Issue 4, 15 February 2008, Pages 1087-1097, ISSN 0020-0255,http://doi.org/10.1016/j.

[9] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.