Segmentation of Sharia Rural Banks based on Growth of the Business Performance

B Sartono¹, S Rahardiantoro¹, C Suhaeni¹, E A Irianto² and U Maulidya²

¹Department of Statistics, Bogor Agricultural University, Bogor, Indonesia
²Otoritas Jasa Keuangan

Email: bagusco@gmail.com, bagusco@ipb.ac.id

Abstract. The Otoritas Jasa Keuangan (OJK) is an institution that is responsible to supervise bank operational in Indonesia, including those which are BPRS’s. As a part of their tasks, they do a monitoring procedure to recognize the performance progress of BPRS. A time series analytics could help in studying time-to-time change of certain performance indicators. In term of the performance indicators, not all BPRS has a good growth since they were established. Indeed many of them are growing positively, but some of them may have up-and-down pattern, even tend to be negatively diminishing. Recognizing the time series pattern of the performance indicators of BPRS is useful for OJK to identify working strategy. This recent paper focuses on segmentation process to identify groups of BPRS based on their growth pattern in multivariate time series cluster analysis cases. The result is there are 3 groups of 156 rural banks, 128 banks in the first group, 18 banks in the second, and 10 banks in the third. The general patterns of each group follow: cluster 1 is a group of banks whith positive growth in most of the variable; cluster 2 is a group that the banks that belong to this cluster have negative growth along the time, especially in the last two years; whereas Cluster 3 consist of rural banks with negative growth at the beginning of the observation period and then got rebound at the rest of the time.

1. Introduction
Micro and small enterprises are believed to be one of the most important entities that support the economy of Indonesia. In 2016, the Ministry of KUKM estimates that at least 58 million enterprises exist in Indonesia and absorb 110 million human resources. However, most of them are not accessible by financial institutions, especially banks, by the reason of poor quality on the record of their business activities. One of the breakthroughs for this problem is the establishment of rural bank which in Indonesia called as Bank Perkreditan Rakyat (BPR) since 1990s. This kind of bank need more relax requirement to the micro and small enterprise to have financing support. As the development of the economics, recently it also exist Sharia/Islamic Rural Bank (Bank Pembiayaan Rakyat Syariah, BPRS).

The Otoritas Jasa Keuangan (OJK) is an institution that is responsible to supervise bank operational in Indonesia, including those which are BPRS’s. As a part of their tasks, they do a monitoring procedure to recognize the performance progress of BPRS. A time series analytics could help in studying time-to-time change of certain performance indicators. In term of the performance indicators, not all BPRS has a good growth since they were established. Indeed many of them are growing positively, but some of them may have up-and-down pattern, even tend to be negatively diminishing. Recognizing the time series pattern of the performance indicators of BPRS is useful for...
OJK to identify working strategies. For example, it may help in determining which bank should be supervised more and to be helped to win-back the business. It is also useful to learn more about the factor that influence the variation of performance among BPRS.

This recent paper focuses on segmentation process to identify groups of BPRS based on their growth pattern. The complication of the problem arises by the number of the series which is not one. So we would consider in handling a multivariate time series cluster analysis. Recall that what we would like to group is not the level of the indicators but the growth. Therefore the final segments would be uncorrelated with the business scale of the banks.

2. Time Series Cluster Analysis
In this section, would be described the main ideas of time series clustering which applied in this study. Based on Liao in 2005 [1], there are three time series clustering approaches, that are raw-data based, feature based, and the model based. The raw-based approach worked by calculating the distance among series in the raw-data, then the specific clustering methods implemented using this distance. The second approach worked based on the feature cores of the raw-data. The feature extractions would be the first step in this approach. The model based approach implemented using the model based on raw data. The necessary thing from this model is usually the coefficients of model, or the residuals of models.

This study applied the most popular approach above to find the best suggestion of cluster, that is the raw-data based approach. In this case, we use clustering based on univariate correlation matrices in multivariate time series. This method developed based on the raw-data. The main idea of this approach is get the transformation of correlation matrices become distance matrices. That is the correlation matrices mention above are the correlation among BPRS in every variables. Because there are more than one variable (in context multivariate data) the correlation matrices built from the average of the correlation matrices in every variable.

3. Data
The data that are involved in the analysis consist of eight time series of the indicators. Those indicators are including:

- Return on Equity (ROE). It is a measure of profitability that calculates how many dollars of profit a company generates with each dollar of shareholders' equity. The formula for ROE is: 
  \[ \text{ROE} = \frac{\text{Net Income}}{\text{Shareholders' Equity}} \]
  ROE is sometimes called "return on net worth."
- Return on Asset (ROA). It is obtained by dividing the BPRS’s annual income by its total assets. This indicator measures how profitable a bank is relative to its total assets and how efficient management is at utilizing the assets to generate earnings.
- Third party funds (Dana Pihak Ketiga, DPK). It consists of demand deposit, time deposit, savings, certificates of deposit, and other short-term deposits at bank.
- Total Financing (Pembiayaan yang Dilakukan, PYD). It is the total amount of loan provide by the rural bank.
- Year-to-date profit and loss statement (Laba Rugi Berjalan). It is a document that shows the incoming and outgoing money from January 1 of the current year to the date the document is being prepared. This statement is a tool used by business owners to determine whether or not the business is lucrative.
- Total Assets. It refers to the total amount of assets owned by a person or entity. Assets are items of economic value, which are expended over time to yield a benefit for the owner. Total Assets is sum of all current and noncurrent assets and must equal the sum of total liabilities and stockholders’ equity combined.
- Earning Assets (Aset produktif). Earning Assets are anything that directly generates income, such as stocks that pay dividends, bonds that pay interest, real estate properties that generate rents, copyright and patents that bring in licensing fees, and machinery that allows you to produce goods for sale at a profit.
• Core Capital (Modal Inti). It is the minimum amount of capital that a thrift bank, such as a savings bank or savings and loan company, must have on hand in order to comply with Federal Home Loan Bank regulations. Core capital consists of equity capital and declared reserves.

There are as many as 156 BPRS’s that are of interest and the series of the data were on the period of Jan 2011 – Dec 2015.

4. Method
To be able to have a good segmentation of BPRS, we did the following steps:
1. General checking to the data quality which included logical violation to the values (mostly extremes) and sudden changes. A series of graphical exploration such as time series plots and histograms were generated to identify the anomaly within the dataset.
2. Constructing among-bank correlation matrix for each performance indicator, so that we had eight matrices. The correlation
3. Combining eight correlation matrices into a single matrix by averaging the correspondent elements
4. Calculating between-bank distance matrix. The between-bank distance is obtained by the following formula:

\[ D(i, i') = \left( \frac{1 - C(i, i')} {1 + C(i, i')} \right)^\beta ; \beta > 0 \]  

where

\( C(i, i') \) is the correlation coefficients between \( i \) – th BPRS and \( i' \) – th BPRS

\( D(i, i') \) is the distance between \( i \) – th BPRS and \( i' \) – th BPRS

5. Implementing an agglomerative clustering algorithm using complete linkage strategy [2]
6. Evaluating the result of the clustering, using some measurement of cohesion and separation statistics, that is the numerical measurement to evaluate the cluster result [3]. This step determines the number of optimal clusters/groups. The formula of cohesion and separation by Tan et al 2005 [4] describe below

\[ \text{cohesion(Cluster}_i\text{)} = \sum_{x \in \text{Cluster}_i} \sum_{y \in \text{Cluster}_i} \text{proximity}(x, y) \]

\[ \text{separation(Cluster}_i, \text{Cluster}_j\text{)} = \sum_{x \in \text{Cluster}_i} \sum_{y \in \text{Cluster}_j} \text{proximity}(x, y) \]

7. Profiling the growth pattern of each group of banks

5. Results and Discussion
5.1. Data Distribution
Figure 1 presents histograms of univariate distribution of the indicators we are interested in. The histograms are summarized from 156 observations, one for each bank. The condition of each bank was summarized using the average value of the last six months.

We could see from the histograms that most of the indicators which are related to business scale or the size such as DPK, PYD, asset and capital are skewed to the right. It happens because there are a small number of banks whose very large scale of business compared to other banks. Whereas, the ROA and ROE tend to have a left-tailed distribution. A bank with a negative and large value of ROA and ROE is a bank which is almost bankrupt or has very bad business. From the graphs we could see that there are two-four banks of that bad.
5.2. Clustering Results
The agglomerative clustering algorithm that was implemented to the distance matrix using complete linkage resulted a dendrogram that records the clustering process as shown in Figure 2. The dendrogram informs us which banks are grouped at the first iteration and what are at the next, until all banks are grouped together into a single cluster.

![Dendrogram of agnes(x = dist1, diss = T, method = "complete")](image)

**Figure 2.** The dendrogram of the clustering process with complete linkage approach.

The chart of a dendrogram might be useful to determine the number of optimal clusters. Recall that a horizontal line in the dendrogram represents a merge step of two nearest clusters. Using the rule of largest jump between two horizontal lines in the dendrogram, we can decide how many clusters should be produced. It is obvious that the jump between the most above line and the second above are the largest jump, so that it indicates that 2 (two) clusters would be good. However we prefer to use other assessment called cohesion and separation statistics to determine the number of clusters. To do that, we assessed using statistics if the number of clusters is 2, 3, 4 and 5. Then, we decide the cluster number with the number whose largest difference of cohesion and separation. Table 1 provides the statistics and 3 (three) clusters are the optimal using this criterion.
Table 1. The assessment of clustering in several numbers of clusters.

| Number of Clusters | Cohesion | Separation | Difference |
|--------------------|----------|------------|------------|
| 2                  | 0.4421   | -0.1195    | 0.5616     |
| 3                  | 0.4528   | -0.1164    | 0.5692     |
| 4                  | 0.4907   | 0.1421     | 0.3486     |
| 5                  | 0.4927   | 0.1420     | 0.3508     |

Based on the last evaluation, we decided to group the BPRS into three groups of banks. The numbers of banks in each group are as follows: 128 banks in the first group, 18 banks in the second, and 10 banks in the third. Table 2 provides us information about the distribution of banks in each group/cluster based on their location/zone and their age group. In the management of OJK, the zones of BPRS are distinguished into four zones: (1) Greater Jakarta, (2) Java and Bali Island, including Bogor, Depok, Tangerang, and Bekasi, (3) capital of Province in other island than Java and Bali, (4) the other locations. It can be seen from Table 2 that the members of each cluster distribute almost uniformly in every zone and every age group.

Table 2. The distribution of group membership based on location zones and age groups.

| Zone | Cluster 1 | Cluster 2 | Cluster 3 | Total |
|------|-----------|-----------|-----------|-------|
| 1    | 16        | 2         | 0         | 18    |
| 2    | 72        | 6         | 5         | 83    |
| 3    | 14        | 2         | 3         | 19    |
| 4    | 26        | 8         | 2         | 36    |
| Total| 128       | 18        | 10        | 156   |

| Age Group | Cluster 1 | Cluster 2 | Cluster 3 | Total |
|-----------|-----------|-----------|-----------|-------|
| more than 15 years | 49        | 11        | 8         | 68    |
| 8 – 15 year     | 38        | 6         | 1         | 45    |
| less than 8 year | 41        | 1         | 1         | 43    |
| Total           | 128       | 18        | 10        | 156   |

5.3. Growth Profiling

It is now interesting to explore the growth pattern of each group resulted by the cluster analysis. Figure 3 depicts line charts, one for each variable, to examine the pattern of monthly growth of BPRS. To be able to have the general pattern, first we standardized the original values across the time for each bank, and then calculated an average within the group so that we have three series for every variable.

The charts on Figure 3 suggest the following general pattern. Cluster 1 is a group of banks with positive growth in most of the variable. The banks in this cluster also tend to have low changes in terms of ROA and ROE. By this profile the authors agree to say that it is a group of banks with good business.

A different pattern was found in Cluster 2. The banks that belong to this cluster have negative growth along the time, especially in the last two years. Whereas Cluster 3 consist of rural banks with negative growth at the beginning of the observation period and then got rebound at the rest of the time. The rural banks in this group tend to have higher monthly fluctuation in their businesses.
a. ROA  
b. ROE  
c. third party funding  
d. total financing  
e. current profit/loss  
f. total assets  
g. total productive assets  
h. core capital

Figure 3. General growth pattern of groups of banks for eight variables under interest.
6. Conclusion
This paper discovered that there are three time series growth patterns of the Sharia Rural Banks in Indonesia. The result was generated by implementing a time series cluster analysis by utilizing an agglomerative algorithm based on correlational distance between banks. Profiling to the three groups shows that there are 128 banks with good growth, 18 banks with negative progress, and 10 banks with worsened at the beginning and tent to be recovered at the recent period.

References
[1] Liao TW 2005 Pattern Recognition 38(11) 1857-1874
[2] Kaufman L and Rousseeuw PJ 1990 Finding Groups in Data : An Introduction to Cluster Analysis Wiley New York
[3] Johnson RA, Wichern DW 2007 Applied Multivariate Statistical Analysis 6th ed Pearson
[4] Tan PN, Steinbach M, Kumar V 2005 Introduction to Data Mining Addison-Wesley Longman Publishing Co. Inc Boston MA USA