

Balance as One of the Attributes in the Customer Segmentation Analysis Method: Systematic Literature Review

Uus Firdaus*, Ditdit Nugeraha Utama

Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

ARTICLE INFO

Article history:

Received: 30 April, 2020

Accepted: 23 May, 2020

Online: 29 May, 2020

Keywords:

Customer Balance

Customer Segmentation

RFM

K-Means

ABSTRACT

The banking industry is very competitive. To utilize the information, they have in order to be a competitive advantage winner is reasonably very crucial for the company. At present, the company does not only focus on the company's strategy that prioritizes products (e.g. product or service oriented), however also necessitates to focus on the company's strategy in prioritizing customers. Customer segmentation, its attributes, and the appropriate analysis method are going to get accurate data segmentation results, so that it is able to be used as a reference by the company and as a basis for determining its products' marketing strategies. This systematic literature review discusses the types of attributes operated, including customer balance attributes, whether or not they can be included in segmentation. In addition, it also discusses what popular analytical methods are widely used in the customer segmentation process. Literature searching in the digital library resulted in a total are 592,363, 1,361, and 21 papers respectively in the first, second, and third stage. 10 papers found finally in the final stage that were considered capable of answering research questions. Based on 10 papers selected, it can be concluded that customer balances can be functioned scientifically as one of attributes for segmentation use. The popular analytic methods operated for customer segmentation are recency, frequency, monetary (RFM) model (4 times appeared), K-Means algorithm (6 times occurred), and C-Means (2 times emerged).

1. Introduction

In business competition, to use of existing capabilities as much as possible in order to compete with other companies is realistically required by the companies. They have to create and obtain the characteristics of their customers, the information is used as a strategy to develop and market their products. Thus, the companies are able to set target rightly and customers are going to be more interested and satisfied. Presently, the companies should not only focus on the companies' strategy prioritizing products, but also need to focus on the companies' strategy prioritizing customers (customer oriented) [1].

The banking industry is very aggressive. To optimizing its information for convincingly winning the competition is very imperative to be conducted by banking institution [2]. The marketing department has an important role to make strategy in the business competition of other companies. Service-oriented companies selling their products to customers will generally face marketing problems. The huge amount of data and parameters collected in the banking area can be one of the marketing problems.

In any company including financial institutions, data is a valuable asset that can be used for corporate strategy. Proper data processing from a collection of raw data is going to produce critical and beneficial information. The Company has a very large data set that can be used for commercial use [3]. Prominent banking data (i.e. transaction data, bank accounts, loan customer data, all payment data, etc.) stored in a database. In essence, this data is very useful and can be used to predict revenue and analyze sales that have been achieved [3].

Customer segmentation is demanded to classify customers who have similar characteristics to find out consumer behavior. This helps in controlling the right marketing strategy to expediently improve company revenue [1]. Numerous methods for segmenting customers are many. The most popular is RFM analysis. It is for behavioral-based data processing that extracts customer profiles using the recency (current), frequency, and monetary values [4].

This literature review aims to find out other customer segmentation attribute in the banking industry, the analytical methods frequently operated, and the parameters involved. The elements used as analysis in the segmentation process are customer transaction habits and customer nominal balances.

*Uus Firdaus, Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, +6281280118811, uus.firdaus@binus.ac.id

Customer segmentation and suitable analysis methods will get accurate data segmentation results so that it can be used as a reference by the company. The results of this customer segmentation can be used as a basis for determining or strategy in marketing its products. Systematic literature review (SLR) method is a mechanism adopted scientifically in conducting studies.

2. Research Methodology

There are three reasons in this study to conduct an SLR. First, gather knowledge from previous research on related topics. Second, identify the elements used from previous studies. And final one, support research in finding new methods or new topics with basic information needed. For this reason, SLRs are able to meet the needs of achieving various objectives while maintaining a strong evidence base [5]. Also apply five stages conducted in reviewing articles; topic selection, determination of the scope of the review, selection of online libraries, selection of literature, and complete review of much of the literature [6].

2.1. Research Question

The focus of this study is to show that the balance attribute as one of the attributes that can be included and an important part of the customer segmentation analysis process. This customer is part of a company / financial institution that has a balance component. This SLR is done also to answer two main research questions as a reference in conducting further research. The research question defined as a guidance for doing the SLR are:

1. RQ1: What can balance component be operated in customer segmentation?
2. RQ2: What are analytical methods functioned in customer segmentation?

RQ1 focuses on discovering previous research to get answer of how important and how many elements of customer balances are included in the customer segmentation process at financial institutions. This will require analysis and find in any research related to customer segmentation in companies or financial institutions that include nominal balances in the process, directly the nominal balance will affect the information and data on the final results of the study so that it can be useful for researchers. The right elements and methods of analysis are needed, especially in relation to the quality of this segmentation data. While RQ2 is for finding and showing most of the analytical methods used in customer segmentation, this research question strongly states that the appropriate analysis methods and parameters used in segmentation will produce accurate and useful information for researchers.

2.2. The Search Process

Several standard indexes are used in the process of finding literature related in online databases. The papers included is in January 2015 until currently. Three online digital libraries used in this process; IEEE Xplore, Scimedirect, and Springer link. The combination of words in the search criteria are:

- (customer segmentation OR bank OR balance)
- (models AND customer segmentation)

In search strings use the "find at least one in context or title" feature of an advanced digital library search. This is to find all about nominal balance paper. Whereas in the second search

string use "find exactly the same" in the advanced search. It's used to filter research papers that contain analytical methods for customer segmentation.

2.3. Study Selection

Inclusion and exclusion criteria were used to minimize the focus of study selection [5]. By including the article; research, journals, conferences, reviews, pattern grouping, bank channel filters, engineering, business management, IT in Business or banking & finance articles as much as possible to compare each method and element used and by avoiding research papers that are not related to customer segmentation and method of analysis. Inclusion criteria are: Papers related to customer segmentation, bank, balance, and analysis methods from 2015 to the present, case studies based on and / or research reports and experience. A year back is an overview that is operated to see trends that are usually operated through several studies.

Exclusion criteria are:

1. The paper focuses on segmentation but the object is not the customer.
2. Paper that discusses customers but does not have a balance component
3. Studies that do not discuss customer segmentation related to customer behavior.
4. Non-English papers or complete book papers.
5. Papers that contain methods without explaining certain parameters.

Table 1. Questions for Quality Assessment

Number	Questions
1	Does this paper fit the topic?
2	Is this paper about research?
3	Are examples discussed in this paper / implementations of methods used?
4	Are there research methods in the paper?
5	Are the findings clearly defined?

Start a search on a digital library by only using the search string without making changes to the search criteria. Next is to conduct the application of the inclusion criteria and exclusion criteria to get a smaller number of papers discovered. The next step is to filter papers from titles and abstracts to get papers that can answer research questions. The final step, the remaining papers are papers that are in accordance with quality assessment and with full text analysis. Questions for quality assessment as seen in Table 1, and complete text analysis, approach descriptively according to topic and read full text of papers.

3. Result of study

Overall selection results as shown in Fig. 1. Start by searching in a digital library by using search strings. Obtained a total of 42,122 papers from IEEEExplore, 549,074 papers from Scimedirect, and 1,167 papers from SpringerLink. The total number at the first search stage was 592,363 papers. Implementation of the inclusion and exclusion criteria resulted in 821 papers from IEEEExplore, 516 papers from Scimedirect, and 24 papers from SpringerLink. This stage provides 1,361 papers. The third stage, search by title and abstract, produced 21 papers. The final stage, filtering our search with quality assessment and full text analysis, found 10 papers that can answer research questions.

Each paper is reviewed to determine the fulfillment of the questions in the research mentioned earlier. Table 2 shows 10 papers obtained from the previous selection process. In the following paper the method used to segment the customer and the components included. Some of the topics found in this paper are about customer segmentation, models, and some of them about the methods used in the segmentation.

Table 2. Selected Papers

Paper	Year	Paper Type	Topics	Method(s)
[7]	2019	Conference (ICITBS)	Bank customers analysis for strategies and contributions to increase profits	PCC Model, Account Profit Analysis
[8]	2018	Conference (ICIMTech)	Customer Segmentation in Banking	RFM Score, K-Medoids and K-Means Clustering
[9]	2018	Conference (SIET)	Two-Step Mining Method for Customer Segmentation	RFM model, K-Means, Silhouette, Connectivity
[10]	2018	Journal	Effective approach to customer segmentation	RFM, Fuzzy C-Means and K-Means
[11]	2016	Conference (ICIC)	Customer potential segmentation	Customer C4.5 & K-Means Algorithm
[12]	2020	Conference (ICIC)	Marketing strategies and improving customer relationships with customer segmentation	RFM values and K-Means Algorithm, SSE, Elbow
[13]	2015	Journal	Profiling banking user Clustering	KDDM processes
[14]	2018	Conference (INNS)	segmentation helps marketing to achieve goals	Dataset, K-Means, SOM
[15]	2018	Conference (UV)	Bank Client Clustering	Fuzzy C-Means Clustering Bayesian
[16]	2018	Conference (ICAITI)	Financial performance of BPR Syariah	Information Criterion (BIC), CF Tree

In [7] the author used bank data as data objects, and had analyzed intensively on the bank customer profit contribution model. Focusing on income from the customer's loan business. The calculation of the main part of bank loans based on customer balances, loan rates, in the loan period. Furthermore, based on the profit contribution from assets, income and liability contributions from the middle business class, the basic model evaluates the contribution from the customer benefits built. The things proposed in this paper are verified to be in line with company needs. That's to increase customer loyalty to the bank; otherwise it can expand the company segment with effective customer support [7].

[8] discussed customer segmentation that has been applied to bank customer data in internet banking users by clustering. It is an unsupervised data mining technique that is used by segmentation of customers. Attribute names: Balance, Transaction Date, Posting Date, Account Number, CustomerID, Debit Code, AmountOf Transaction Code. This research was conducted to build a model of customer profile data based on the use of their IBs in banks and grouping methods using the K-Means and K-Medoids methods based on RFM scores from customer transactions. The results of his research show that the K-Means method outperforms the K-Medoids method, on the Davies-Bouldin index, K-Means performed slightly better than K-Medoids [8].

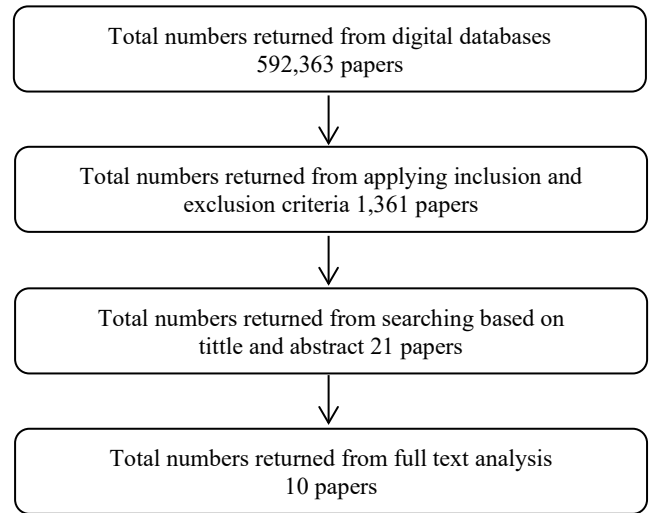


Fig. 1. Paper selection process

[9] explained customer relationship management (CRM) that has been researched to gain insight into customer requests and needs. The combination of user data such as the average customer balance was used for mining and computing methods, although this approach methodically still had several limitations. Also, several studies related to the RFM model in understanding customer habits. The model was functioned to segment customers in terms of customer transactions, frequencies, and monetary conditions. The first step, the results of the RFM model data and grouped with the k-Means algorithm. Then, data from each cluster was analyzed by association to make customer characteristics represented by IF-THEN rules. Clustering results were analyzed by silhouette size and connectivity [9].

In [10] the author examined the customer segmentation that has been run and efficiently in a company. Customers were categorized into the same behavior groups based on the customer's RFM values. Segmentation has provided a good understanding of customer needs and helped identify potential company customers. Dividing customers into several segments of the min-max balance component in an effort to increase company revenue. Retaining customers is considered more important than recruiting new customers. This can be in the form of implementing specific marketing strategies for individual segments to retain customers. This research begins with RFM analysis on transactional data and is developed into the same cluster using traditional K-means and Fuzzy C-Means algorithms. This paper also provided ideas for selecting initial centroids in the proposed K-Means algorithm [10].

In [11] the author reviewed companies that know that regular customers have costs even though they are lower than

the costs of getting new customers. Customer loyalty is important to know so that companies can project revenue as a reference in corporate planning and strategy. The process begins by using certain attributes to perform the segmentation process with the K-means algorithm. Attributes such as exchange rates, balances, the number of days past payment or the number of months of debt, and the age of the customer. Dataset without segmentation, is intended for termination, product, service disconnection, balance, age of the customer, and always decides that the class of loyalty is classified. This model combines the K-means algorithm and the C4.5 classification algorithm [11].

In [12] the author deliberated the segmentation of customers for marketing strategies and to improve the relationship between customers and companies. The behavior of loyal customers from certain services or products provides benefits to the company because customers will continue to use the services or products. This research is to find the value and types of customers that can be used to determine which customers provide the most profit for the company. RFM and balance value as monetary are used as basic criteria to identify customers in forming clusters and are called clustering. In this study the clustering method or algorithm used is K-Means and also uses the results of the Elbow Method from sum square error (SSE) from several existing clusters [12].

In [13] the author portrayed the analysis that has been carried out by analyzing data sets to obtain and ascertain the full potential of customers from realized techniques, it is called knowledge discovery and data mining (KDDM). This technique is used to process survey data from internet banking users in Jamaica that include demographics, attitude and behavior variables. The results of summarizing internet banking user data can find out the services and patterns of internet banking usage that are most frequently used. Balance inquiry and bill payment are two of the most frequently used features of a customer's account [13].

In [14] the author stated that customer segmentation whose concept has been designed in marketing to increase business and increase revenue. Also discussed are various data analytic algorithms, specifically K-Means and self-organized maps (SOM). The K-Means algorithm has shown promising clustering results, and SOM is beneficial in speed, quality in grouping, and visualization. Two levels of grouping have been applied to large customer transaction data sets, transaction history, transaction times, balances, transaction amounts, etc. K-means algorithm and grouping step are applied in this customer segmentation. Recent research proposes a segmentation framework of a customer's lifetime value (LTV). Researchers have prepared a framework for segmenting customers, calculating the value of each segment for life, and estimating the future value of each segment [14].

In [15] the author explained the groupings that have been carried out in detail. The techniques discussed mostly depend on the characteristic features of the database. In analyzing the actual data, the algorithm should be efficient in large multidimensional data sets, noise, and outliers. The researchers clearly explained this phenomenon in this paper discussing PFCVI, which took a lot of results from a shared partition and gave the advantage of analyzing information on the nominal amount of savings / balance in a bank account. Finally, we created profiles of each user and their categories and postulated references for banks [15].

In [16] the author explained Islamic BPRs in Indonesia which carry out their business activities are based on sharia principles and there is no payment service. The researcher explains that the company must evaluate its capital (balance) to survive and compete with customers. The level of bank performance can measure the health of the bank. The variables used in measuring bank performance are capital (C) and others that are part of the CAMELS method [17].

The results of the analysis from the financial side of the BPRS such as capital, productive assets, third party funds, and increasing credit have a great opportunity to change for the better. The results of the cluster feature (CF) tree were analyzed by hierarchical groups using the agglomeration method, and the calculation of the number of groups in the Bayesian information criteria (BIC) for each group [16].

4. Discussion of Research Questions

RQ1: What can balance component be operated in customer segmentation?

Based on the paper reviewed in this study, various types of attribute usage that can be interpreted as balance attributes are obtained. They were obviously described in Table 3. The use of these balance attributes described by this paper has the same goal of supporting the process of segmenting or grouping customers. [7] discussed that customer balance was calculated and was an object of research analysis which considered as revenue and contributes to company profits. While [8] discussed internet banking user balances functioned in customer segmentation and clustering with data mining techniques, and [9] technically expressed the average customer balance operated for mining and computing methods for specified CRM.

In addition, [10] described the Min-Max balance component in effort to increase company revenue. Customer balances in the dataset officially exploited to classify customer loyalty were discussed in [11], while [12] discussed the balance as monetary benefited in the clustering method. In [13], balances and transactions were two features that widely utilized to discover potential customers in the KDDM process.

Also, the balance before transaction discussed in [14]. It was used in the customer segmentation framework for customer lifetime value (LTV). [15] explained the analysis of information regarding the nominal amount of savings / account balance and becomes a reference for banks. Finally, [16] discoursed a capital (balance) as well as a variable or attribute exploited in practically measuring a company's financial performance.

Table 3. Approaches of customer segmentation methods

Number	Attribute	Description
1	Balance of Customer	Calculated, income from the customer's loan, for the bank profit contribution
2	Balance of IB users	Balance for customer segmentation and clustering in data mining technique
3	Average customer balance	Average customer balance used for mining and computing methods for CRM
4	Min-Max balance	Min-Max balance component in an effort to increase company revenue
5	Customer balance	Customer balance in the dataset to classify class of customer loyalty.

6	Balance as Monetary	Balance value as monetary for clustering method
7	Balances and Transactions	Two features are widely used to find potential customers in the KDDM
8	Balance before transaction	Used for the Bank Customer segmentation framework, on customer LTV
9	Value of savings	Create user profiles for each category and provide some advice for banks
10	Bank capital balance	To measure the health of the bank, and for the bank's financial performance

Discussing about the balance attributes used in the segmentation or clustering process will require analytic models that can process them and suitable methods so as to produce optimal information and results.

RQ2: What are analytical methods functioned in customer segmentation?

From some papers that have been reviewed, there are several models and methods used in segmentation. From some papers that have been reviewed, there are several models and methods used in segmentation as show in Table 4. Based on the paper that has been reviewed, there are several models and methods used in segmentation. The popular method used and appears in many discussions of this literature shows that the method is indeed suitable for use in customer segmentation. The RFM model appears 4 times from the papers selected for discussion, namely in the papers [8], [9], [10], [12], while the K-Means Algorithm appears most and is used in this literature as many as 6 times in papers [8], [9], [10], [11], [12], [14], and the most popular of which were used and appeared 2 times, namely C-Means in papers [10], [15].

Table 4. List of Popular Methods / Models

Number	Popular Methods	Paper
1	RFM models	[8], [9], [10], [12]
2	K-Means Algorithm	[8], [9], [10], [11], [12], [14]
3	C-Means	[10], [15]

Search results in the digital library showed 592,363 the first stage, filtered to 1,361 in the second stage, 21 papers in the third stage, and the final stage selected 10 papers that answered the research questions. In conclusion from the 10 papers, customer balances can be used as attributes for customer segmentation, and 3 analytical methods most often used; RFM models, K-Means algorithm, and algorithm and C-Means.

5. Limitations

The number of papers selected and reviewed may have a weak validity of the findings. One reason is in the number of digital libraries that discuss this topic is very much which will be able to show most of the other sophisticated methods or models in customer segmentation and / or the methods they use. However, this will also lead to a minimum number of relevant papers going through the screening process. Obviously, the search, selection, and quality assessment are very strict and reject some papers that do not meet some assessment questions.

6. Conclusion and Future Works

From the results of a systematic literature review that has been done, customer balances are included or used as attributes in the customer segmentation process. Responding to the RQ1

research question, all papers that have been reviewed show that balances are used as attributes in the customer segmentation process. This attribute is the total balance, the value of savings, the amount of capital, or the average value of the balance. There are 3 popular methods that are widely used in the customer segmentation process of the papers that have been reviewed. Future research can be carried out to improve the validity of the findings by including more digital libraries or expanding on inclusion criteria.

References

[1] B. E. Adiana, I. Soesanti and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering" JUTEI Edisi Volume 2, no. 2, 23-32, 2018. <https://doi.org/10.21460/jutei.2018.21.76>.

[2] S. M. Kostic, M. Duricic, M. I. Simic and M. V. Kostic, "Data Mining and Modeling Use Case in Banking Industry" in 2018 26th Telecommunications Forum (TELFOR), Belgrade, 1-4, 2018. <https://doi.org/10.1109/TELFOR.2018.8611897>.

[3] A. Subrahmanyam, "Big data in finance: Evidence and challenges" Borsa Istanbul Review, vol. 19, no. 4, pp. 283-287, 2019. <https://doi.org/10.1016/j.bir.2019.07.007>.

[4] M. Tavakoli, M. Molavi, V. Masoumi, M. Mobini and S. Etemad, "Customer Segmentation and Strategy Development based on User Behavior Analysis, RFM model and Data Mining Techniques: A Case Study" in 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), Xi'an, 119-126, 2018. <https://doi.org/10.1109/ICEBE.2018.00027>.

[5] R. P. Ghozali, H. Saputra, Suharjito, D. N. Utama and A. Nugroho, "Systematic Literature Review on Decision-Making of Requirement Engineering from Agile Software Development" in 4th International Conference on Computer Science and Computational Intelligence 2019 (IC2SCI), Jakarta, 274-281, 2019. <https://doi.org/10.1016/j.procs.2019.08.167>.

[6] R. Bria, A. Retnowardhani and D. N. Utama, "Five Stages of Database Forensic Analysis: A Systematic Literature Review" in 2018 International Conference on Information Management and Technology (ICIMTech), Jakarta, 246-250, 2018. <https://doi.org/10.1109/ICIMTech.2018.8528177>.

[7] Y. Jinping, "Analysis and Improvement Strategy for Profit Contribution of Bank Customer Under Big Data Background" in International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), China, 338-341, 2019. <https://doi.org/10.1109/ICITBS.2019.00089>.

[8] M. Aryuni, E. D. Madyatmadja and E. Miranda, "Customer Segmentation in XYZ Bank using K-Means and K-Medoids Clustering" in International Conference on Information Management and Technology (ICIMTech), Jakarta, 412-416, 2018. <https://doi.org/10.1109/ICIMTech.2018.8528086>.

[9] F. A. Bachtiar, "Customer Segmentation Using Two-Step Mining Method Based on RFM Model" in International Conference on Sustainable Information Engineering and Technology (SIET), Malang, 10-15, 2018. <https://doi.org/10.1109/SIET.2018.8693173>.

[10] A. J. Christy, A. Umamakeswari, L. Priyatharsini and A. Neyaa, "RFM Ranking – An Effective Approach to Customer Segmentation" Journal of King Saud University - Computer and Information, 2018. <https://doi.org/10.1016/j.jksuci.2018.09.004>.

[11] S. Moedjiono, Y. R. Isak and A. Kusdaryono, "Customer Loyalty Prediction In Multimedia Service Provider Company With K-Means Segmentation And C4.5 Algorithm" in International Conference on Informatics and Computing (ICIC), Mataram, 201-215, 2016. <https://doi.org/10.1109/IAC.2016.7905717>.

[12] Dedi, M. I. Dzulhaq, K. W. Sari, S. Ramdhan and R. Tullah, "Customer Segmentation Based on RFM Value Using K-Means Algorithm" in International Conference on Informatics and Computing (ICIC),

Semarang, 1-14,
2020.<https://doi.org/10.1109/ICIC47613.2019.8985726>.

- [13] G. Mansingh, L. Rao and K.-M. Osei-Bryson, "Profiling internet banking users: A knowledge discovery in data mining process model based approach" *Information Systems Frontiers* volume, vol. 17, no. 1, pp. 193-215, February 2015.<https://doi.org/10.1007/s10796-012-9397-2>.
- [14] W. Qadadeh and S. Abdallah, "Customers Segmentation in the Insurance Company (TIC) Dataset" in *INNS Conference on Big Data and Deep Learning* 2018, Edinburgh, 277-290, 2018.<https://doi.org/10.1016/j.procs.2018.10.529>.
- [15] J. Zheng, H. Cui, X. Li, L. Meng and T. Wang, "The Clustering for Clients in a Bank Based on Big Data" in *2018 4th International Conference on Universal Village (UV)*, Boston, 1-5, 2018.<https://doi.org/10.1109/UV.2018.8642136>.
- [16] M. F. Nazar, Maiyastri, D. Devianto and H. Yozza, "On the Clustering of Islamic Rural Banks Based on Financial Performance" in *International Conference on Applied Information Technology and Innovation (ICAITI)*, Padang, 108-113, 2018.<https://doi.org/10.1109/ICAITI.2018.8686755>.
- [17] A. Shaddady and T. Moore, "Investigation of the effects of financial regulation and supervision on bank stability: The application of CAMELS-DEA to quantile regressions" *Journal of International Financial Markets, Institutions and Money*, vol. 58, pp. 96-116, 2019.<https://doi.org/10.1016/j.intfin.2018.09.006>.