

IJACT 24-9-19

Identifying Cluster Patterns in Relationship Between Municipal Revenue Configuration and Fiscal Surplus: Application of Machine Learning Methodologies

¹Im Chunghyeok, ²Ryou Jaemin, ³Han JunHyun, ⁴Bae Jayon

¹ Lecturer. Dept. of Business Administration, Inha Univ., Korea
E-mail imchunghyeok@gmail.com

² Lecturer. Dept. of Global Public Administration, Yonsei Univ., Korea
E-mail min8483@yonsei.ac.kr

³ Senior Researcher. Fiscal Performance Management Institute, Korea
E-mail jh2-duran@hanmail.net

⁴ Researcher. AI-RPA Center, Fiscal Performance Management Institute, Korea
E-mail spy_neo@naver.com

Abstract

Net surplus serves as a crucial indicator of how efficiently local governments utilize their resources. This study aims to analyze and categorize the patterns of net surplus across 75 local governments in Korea. By employing machine learning techniques such as K-means clustering and silhouette analysis, this research delves into surplus patterns, revealing insights that differ from those provided by traditional analytical methods. Machine learning enables a broader spectrum of discoveries, leading us to identify three distinct clusters in the net surplus of Korean local finances. The characteristics of these three clusters show that the wealthiest cities have the highest surplus ratios. In contrast, mid-sized municipalities, constrained by limited central government support and scarce local resources, exhibit the lowest surplus ratios. Interestingly, a significant number of cities maintain a median surplus ratio even under challenging fiscal conditions. Additionally, we identify critical thresholds that differentiate the three clusters: a grant-in-aid ratio of 19.31%, a debt ratio of 3.52%, and a local tax ratio of 25.58%. This identification of thresholds is a key contribution of our study, as these specific thresholds have not been previously addressed in the literature.

Keywords: Government Fiscal Surplus, Machine Learning Methodologies, Revenue Configuration, Cluster by Financial Sources

1. INTRODUCTION

As for the fiscal management of municipal governance, it is not uncommon that the revenue streams exceed or fall short of initial projections. This inherent variability necessitates a recognition of the potential for a certain degree of unutilized funds, or the net surplus. This net surplus embodies a duality of implications: on

Manuscript received: June 5, 2024 / revised: July 2, 2024 / accepted: September 3, 2024

Corresponding Author: spy_neo@naver.com

Researcher, AI-RPA Center, Fiscal Performance Management Institute, Korea

Copyright©2024 by The International Promotion Agency of Culture Technology. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>)

the one hand, it signifies a positive aspect of having discretionary funds; on the other, it suggests a certain inefficiency in fiscal stewardship, reflecting a suboptimal allocation of financial resources.

In Korea, there are a total of 243 local government entities, each with distinct characteristics in generating fiscal surpluses. Traditional studies typically examine surplus generation by grouping these entities based on the size of their local populations. This study, however, takes a different approach by clustering them using a new technology—machine learning. The motivation behind this methodological shift is to gain fresh insights into the dynamics of surplus generation.

In this context, the study places particular emphasis on the configuration of the revenue sector, in addition to the expenditure side. By doing so, it aims to analyze the patterns of surplus generation, with a focus on how their revenue structures contribute to these surpluses. This approach offers a more nuanced understanding of the fiscal dynamics within these urban entities.

Chapter 2 reviews the existing literature, highlighting the need for applying new technologies to uncover novel findings in local finance. Chapter 3 outlines the adaptation of typical machine learning processes to the subject of this study. It details the characteristics of the dataset, preprocessing steps, analytic algorithms, and post-processing methods. Advanced analytics such as K-means clustering, Silhouette scores, and decision tree analysis are employed in this research.

2. LITERATURE REVIEWS

2.1 Financial Status Analysis

Surplus-related research generally falls into two main categories. The first type analyzes the current status of surpluses to identify their underlying causes. The second focuses on behavioral analysis, exploring the sources of surpluses in relation to financial, economic, and political factors.

A study by [1] investigates which revenue and expenditure items contribute most significantly to surpluses. It finds that non-tax revenues have the largest impact, while local taxes have the smallest. On the expenditure side, loans and capital expenditures in science and technology, as well as land and regional development, are the primary contributors to the surplus. [2] analyzes data from the 25 districts of Seoul Metropolitan between 2016 and 2021, revealing that a higher proportion of non-tax revenue and unspent budget balances lead to an increase in the proportion of net surplus funds. Finally, [3] highlights that the revenue components influencing surplus vary depending on the type of local government.

2.2 Behavioral Analysis

The following studies of [4], [5], [6], [7], [8], and [9] examine the influence of fiscal, economic, and political factors on surplus generation. According to [4], an increase in the proportion of local taxes correlates with a rise in the ratio of surplus funds, while municipalities receiving a higher proportion of subsidies from the central government tend to have a lower surplus ratio. Additionally, a higher ratio of local government debt is linked to a reduced proportion of surplus funds.

[5] finds that the re-election of local government leaders tends to decrease net surplus funds, although there is typically an increase in surplus funds during election years. The Unreserved General Fund Balance is positively influenced by factors such as the volatility of local revenues, the labor cost ratio, the debt ratio, and the capital expenditure ratio [6].

In [7], it is observed that the overall ratio of the supplementary budget to the main budget significantly increases as the surplus grows, while political variables do not show significant correlation. In [8], revenue diversity and the unassigned fund balance as a percentage of total revenues are positively related, but only in spatial regressions. Revenue volatility shows a positive relationship in both fixed effects and spatial regressions. Finally, [9] finds that the political culture of voters and the financial management capacity of local governments influence the creation of fiscal stabilization funds.

2.3 Lesson for This Study

Numerous studies have explored the various factors contributing to the generation of financial surpluses. However, surpluses result from the simultaneous influence of multidimensional factors, which may combine to form discernible patterns. Segmenting groups of municipalities could further illuminate the relationships between fiscal structure and surplus generation.

The advancement of machine learning and artificial intelligence also holds promise for uncovering new insights in financial management. For instance, [10] attempts to cluster Korean cities based on their differing approaches to participatory budgeting. Similarly, [11] employs machine learning techniques to enhance the use of financial data. Building on this growing trend, the present study seeks to apply both supervised and unsupervised machine learning methodologies to the analysis of surplus patterns.

3. RESEARCH DESIGN FOR MACHINE LEARNING

3.1 Dataset and Preprocessing

Data on municipal surpluses were obtained from the government authority, encompassing 149 variables across 75 Korean municipal governments. However, this study focuses on a select number of variables directly related to financial surplus. Municipal fiscal resources are drawn from seven distinct sources: local tax revenue, non-tax revenue, grants-in-aid, subsidies, municipal bonds, compensations, and carryover. The variable names are labeled to reflect these revenue sources as follows: NetSurplus, LocalTax, NonTaxRev, Grant, Subsidy, Bond, Comp, and Carryover. These diverse streams collectively form the financial foundation of city governments, enabling them to fulfill their varied responsibilities.

A preliminary investigation of the data revealed that Gwacheon City is an outlier, leading to its exclusion from the analysis. Gwacheon City's outlier status is due to its unique financial structure, as the city generates significant revenue from its racetrack, making it one of the few cities that do not receive grants from the central government. Consequently, the final number of city governments included in the analysis is reduced to 74.

Following a standard machine learning approach, this study employs random selection to allocate 52 municipalities, representing 70% of the 74, to the training dataset. The remaining 22 municipalities, or 30%, form the testing dataset. This division is crucial for evaluating the model and analysis results, serving as a prerequisite for accurate forecasting in real-world scenarios.

3.2 Process of Analysis and Algorithms

The K-means technique is employed for the analysis, serving as a clustering algorithm that partitions a dataset into K distinct clusters by minimizing the variance within each cluster. The algorithm iteratively assigns data points to the nearest cluster center and updates the cluster centers until convergence. However, a notable limitation of K-means is its sensitivity to the initial placement of cluster centers and the choice of the number of clusters (K).

To address this issue, silhouette analysis is subsequently performed. By calculating the silhouette score for different values of K, we can identify the number of clusters that maximizes the silhouette score, thereby determining the most appropriate number of clusters. Once the optimal number of clusters is established for the Korean municipalities, a decision tree analysis is conducted to identify the criteria that define these clusters.

The final stage of the machine learning analysis is evaluation. After generating estimates by applying the testing data to the prediction model derived from the training data, these estimates are evaluated to determine how closely they match the actual values. Metrics such as AIC (Akaike Information Criterion) and CA (Classification Accuracy) are used in this evaluation process.

4. Results of Analysis

4.1 Current Status of Municipal Revenue Configuration and Surplus

The following [Table 1] demonstrates the current situation of the financial surplus and revenue configurations, as of 2021. In an examination of municipal fiscal metrics, it is noteworthy that the average global surplus ratio for city governments stands at 7.65%. However, Paju City presents an anomalously high figure of 19.13%. Additionally, when considering the proportion of local government debt, Gumi City exhibits a considerably elevated level at 10.55%. Overall, however, it can be posited that the general range of these financial indicators typically falls within conventional parameters.

Table 1. Percentage share of various revenue sources relative to total revenue

	Surplus	Local Tax	Non-tax Revenue	Grant-In-Aid	Subsidy	Debt	Compensation
Mean	7.65	15.85	5.82	26.03	36.83	1.67	1.55
Max	19.13	38.54	17.89	54.45	51.17	10.55	2.34
Min	2.56	4.49	2.82	1.04	13.16	0	0.69
Std.Error	0.40	0.96	0.29	1.65	0.81	0.3	0.04

Source: Created by the Authors

4.2 Identifying Clusters by K-means

The K-means technique identified three clusters of NetSurplus, with silhouette analysis used to validate the classification. Silhouette scores, which range from -1 to 1, help assess the quality of clustering; the closer the score is to 1, the better the clustering. In Figure 1, the silhouette scores for the 74 municipalities range from 0.55 to 0.72. These scores suggest that the clustering is well defined, though there may be some overlap between clusters.



Figure 1. Silhouette scores by individual municipalities

Figure 2 illustrates the characteristics of NetSurplus among the municipalities. Cluster 2 has the highest amount of NetSurplus, with six cities—such as Seongnam, Yongin, Cheongju, Jinju, and three other cities—falling into this cluster. Cluster 1 includes 46 of the 74 cities, typically small in population size and financial capacity. Medium-sized cities, such as Anyang, Andong, Gimcheon, and 20 others, belong to Cluster 3. The degree of NetSurplus compared to the total revenue is also medium.

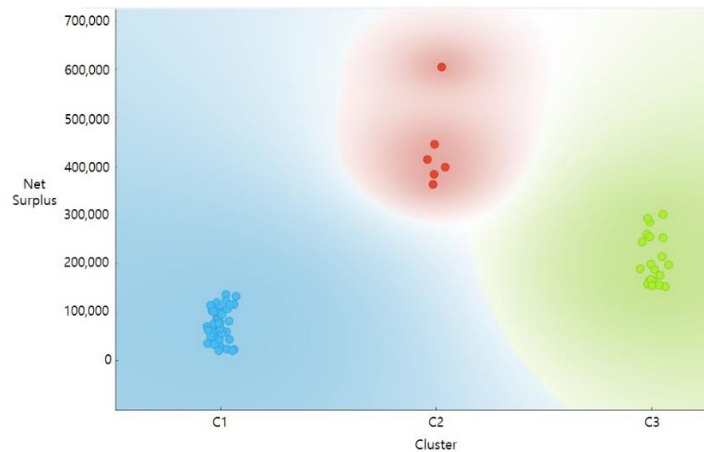


Figure 2. Three clusters found by K-means

4.3 Financial Threshold for Classification

A separate decision tree analysis was conducted to identify the financial characteristics of the clusters. The results revealed that a grant-in-aid ratio of 19.31%, a debt ratio of 3.52%, and a local tax ratio of 25.58% are critical thresholds for classification.

Group C1 consists of city governments where the grant-in-aid ratio exceeds 19.31% while maintaining a debt ratio below 3.52%. This suggests that these cities, likely due to financial difficulties, receive substantial grants from the central government but engage in limited urban development, resulting in a lower debt proportion. They can be characterized as municipalities with relatively inactive fiscal activities. Out of the total 74 cities, 43 belong to this cluster, with an average surplus fund ratio of 6.69%.

Group C2 includes six cities with a grant-in-aid ratio of less than 19.31%, indicating that their budget sizes are sufficiently large to avoid substantial grants from the central government. However, these cities have a debt ratio greater than 3.52% and a local tax proportion of less than 25.58%, suggesting that while they are financially stable now, they may face financial stress in near future. The average surplus ratio for this group is 5.47%. The city of Gwacheon was excluded from the analysis as an outlier, but it might have belonged to this group if included.

Group C3 comprises city governments with a grant-in-aid ratio of less than 19.31%, a debt ratio greater than 3.52%, and a local tax income exceeding 25.58%. These cities are relatively financially stable and development-oriented, actively engaging in fiscal projects. Of the 74 city governments, 21 fall into this category, with an average surplus ratio of 10.78%.

5. CONCLUSION

In summary, the most prosperous cities exhibit the highest ratios of surplus funds, while mid-sized municipalities, which receive limited support from the central government and have constrained local resources, show the lowest surplus ratios. This trend signals potential fiscal challenges for these municipalities. Interestingly, a significant number of cities manage to maintain a median surplus ratio despite facing suboptimal financial conditions.

These insights are challenging to uncover using traditional analytics, as such methods often rely on a minimal number of variables based on the "principle of parsimony." This approach can exclude valuable

information from the analysis. For instance, during the COVID-19 pandemic from 2020 to 2022, local finances were heavily impacted. Simultaneously, the years 2020 and 2021 saw a sharp increase in real estate values in Korea. Such fluctuations leave municipalities with little room to adjust, leading to higher surpluses. Consequently, wealthier municipalities experienced significant discrepancies between anticipated and actual revenue from property taxes, resulting in greater financial surpluses at the end of the year[12].

The advantage of machine learning analytics lies in its ability to uncover new insights. For example, if a city in Cluster 1 aims to move to Cluster 3, the key strategy would be to attract more residents. An increasing population drives up real estate prices and, consequently, property tax revenues. This raises the subsequent question: how can a city attract more people?

This study suggests that net surplus is not just a figure on the balance sheet but also an indicator of financial resilience. Entrepreneurial leadership is essential to improving the current financial situation and preparing for future challenges[13].

REFERENCES

- [1] S. I. Im and H. Lee, 'Budget Factors Influencing the Net Budget Surplus of Local Finance', *Korean Governmental Accounting Review*, Vol 12 No. 2, pp. 1-25, 2014.
- [2] C. H. Im. 'Effect of Local Revenue Structure on the Budget Surplus and Unused Budgets', Ph.D. Thesis, University of Inha, 2024.
- [3] M. S. Song and I. S. Kang, 'A Study of Factors Affecting Local Governments' Budget Surplus : Focusing on the 25 boroughs of Seoul'. *Korean Public Management Review*, Vol 37 No. 1, pp. 119-137, 2023. DOI : 10.24210/kapm.2023.37.1.006.
- [4] I. M. Pai, 'Analysis of the Effects of Local Government Revenue Structure on Net Surplus', *The Korea Local Administration Review*, Vol 34 No. 4, pp. 103-130, 2020.
- [5] J. Kim, 'Determinants of Local Government Fiscal Slack: Organizational, Fiscal, and Political Perspectives'. *The Korean Association for Policy Studies*, Vol 30, No. 4, pp. 195-232, 2021.
- [6] Su. M, 'Understanding the accumulation of local government savings: A dynamic analysis', *International Journal of Public Administration*, Vol 42 No. 11, pp.893-903. 2019. DOI : <https://doi.org/10.1080/01900692.2018.1522506>
- [7] S. H. Lee, 'Examining the effect of net surplus budget on the local governments' supplementary budget expansion: Focused on the theory of Niskanen' discretionary budget-maximization'. *Korean Governmental Accounting Review*. Vol. 18 No.2, pp. 77-112. 2020.
- [8] J. Shon and S. Kwak, 'Managing Fiscal Volatility: An Empirical Analysis of California County Governments' Saving Behavior'. *The American review of public administration*, Vol. 50 No.3, pp. 328-345. 2020. DOI: <https://doi.org/10.1177/0275074019893809>
- [9] D. Snow, G.A. Gianakis and J. Haughton, 'The politics of local government stabilization funds'. *Public Administration Review*, Vol 75 No. 2, pp. 304-314. 2015. DOI: <https://doi.org/10.1111/puar.12317>
- [10] J. H. Han, J.M. Ryou, J.Y. Bae and C. H. Im, 'A Study on Korean Local Governments' Operation of Participatory Budgeting System: Classification by Support Vector Machine Technique', *The Journal of the Convergence on Culture Technology (JCCT)*. Vol. 10 No.3, pp. 461-466. 2024.
- [11] M. K. Kim, 'Utilization of Machine Learning to Promote Use of Fiscal Data'. *Research Paper 23-10. Korea Fiscal Information Service*, pp. 1-63. 2023.
- [12] J. S. Yang and H.J. Oh, 'An Effect of Local Financial Management System on the Budget Execution in Local Governments: Focusing on Unused and Carried-forward Budget', *Korean Public Management Review*, Vol 34 No 3, pp. 25-48, 2020. DOI : 10.24210/kapm.2020.34.3.002
- [13] D. Y. Yoon, 'Three Factors Supporting Sustainability and Development of Local Industries', *Juo Economic Press*. Japan, 2014.