Journal of Industrial Policy and Technology Management, 7(2), 2024, 49-60

Examining The Impact of Income - Expense Ratios on Bank Success Using Machine Learning Methods

İsmail Dere[1] Münevver Turanlı[2] Selçuk Alp[3]
[1,2]*İstanbul Ticaret University, Türkiye*
[3]*Yıldız Technical University, Statistics Department, Türkiye*
[1]*dereismail7@gmail.com*, [2]*mturanli@ticaret.edu.tr*, [3]*alp@yildiz.edu.tr*

Received: Jul 20, 2024 Accepted: Aug 9, 2024 Published: Dec 30, 2024

Abstract: The study aims to examine the success of banks in the banking sector in terms of financial ratios. In this context, income-expense structure ratios for 24 selected banks operating in Türkiye for the period 2012-2022 were used. In the study, Net Operating Profit (Loss) / Total Assets was considered as the dependent variable. For each year of the 2012-2022 period, the average of the relevant variable was taken, and banks above the average were coded as successful, while those below were coded as unsuccessful. The examination of the period, especially considering the average taken for each year, allowed for the consideration of any emerging vulnerabilities or opportunities given the different cyclical conditions. Classification algorithms are used to determine which of the predefined classes a data point belongs to. The study predicted the 264 data points for the 24 banks for the period 2012-2022 using machine learning methods (Logistic Regression, k-Nearest Neighbors, Support Vector Machines, Artificial Neural Networks, CART, Random Forest, and Light Gradient Boosting Machine). According to the findings, the highest accuracy, precision, recall, and F1 score values among the machine learning methods used were shown by XGBoost, while the lowest values were shown by the KNN method. When examining the findings in terms of income-expense structure ratios, the most effective criteria for determining the success of a bank's operations were found to be total income / total expenses, net interest income after specific provisions / total assets, and other operating expenses / gross operating profit. The study found that banks maintaining a balance between income and expenses were successful. Among the machine learning methods used to determine the financial success of banks, XGBoost showed the highest performance. Additionally, the most effective criteria for assessing banks' operational success were identified, highlighting that these criteria are significant indicators for understanding bank performance. This study demonstrated the effectiveness of machine learning methods in evaluating the financial structures of banks and provided an important foundation for future research.

Keywords: Machine Learning, Banking Sector, Financial Performance

JEL Classification: C10, G20

1. Introduction

In the dynamic and ever-evolving financial landscape, the success of banks has become increasingly complex to analyze, with a myriad of factors contributing to their performance. One crucial aspect that has garnered significant attention is the impact of income-expense ratios on bank success, particularly in the context of leveraging machine learning methods to enhance our understanding of these relationships. This research paper aims to delve into the intricate interplay between income-expense ratios and bank success, exploring the potential insights that can be derived through the application of advanced analytical techniques.

The primary objective of this study is to examine the influence of income-expense ratios on the overall success of banks, with a specific focus on the ability of machine learning models to capture and elucidate these relationships. In this context, income-expense structure ratios for 24 selected banks operating in Türkiye for the period 2012-2022 were used. In the study, Net Operating Profit (Loss) / Total Assets was considered as the dependent variable. For each year of the 2012-2022 period, the average of the relevant variable was taken, and banks above the average were coded as successful, while those below were coded as unsuccessful. The examination of the period, especially considering the average taken for each year, allowed for the consideration of any emerging vulnerabilities or opportunities given the different cyclical conditions.

Through the integration of machine learning algorithms, this research endeavor will transcend the limitations of traditional statistical approaches, enabling a more comprehensive and dynamic analysis of the complex relationships at play. By training and validating a suite of machine learning models, including supervised and unsupervised techniques, we aim to identify the most influential income-expense ratio variables and their relative impact on bank success, while also exploring the potential for predictive capabilities.

Furthermore, this study will delve into the implications of these findings, discussing the practical applications for bank executives, regulators, and policymakers. The insights gleaned from this research can inform strategic decision-making, risk management practices, and the development of data-driven policies that foster the long-term sustainability and resilience of the banking sector.

The methodological approach employed in this study will encompass a comprehensive data collection and preprocessing phase, followed by the implementation of a diverse range of machine learning algorithms, including Logistic Regression, k-Nearest

Neighbors, Support Vector Machines, Artificial Neural Networks, CART, Random Forest, and Light Gradient Boosting Machine. The performance of these models will be rigorously evaluated using appropriate metrics, such as accuracy, recall, F1 scores, to ensure the reliability and generalizability of the findings.

By combining the analytical power of machine learning with the depth of understanding gained from a thorough examination of income-expense ratios, this research paper will contribute to the existing body of knowledge in the field of banking and finance, serving as a valuable resource for academics, industry practitioners, and policymakers alike.

2. Literature

The banking sector plays a critical role in economic stability and growth. In this context, evaluating and predicting the financial performance of banks is of great importance. Income-expense ratios are significant financial indicators that reflect the profitability, financial structure, and overall performance of banks. This study aims to examine the impact of income-expense ratios on bank success using machine learning methods.

Machine learning has become a prominent tool for analyzing large datasets, making predictions, and identifying patterns that are not easily noticeable with traditional statistical methods. In the banking sector, machine learning techniques are widely used for credit risk assessment, customer behavior analysis, and financial performance forecasting. Fethi and Pasiouras (2010) explored the application of operational research and artificial intelligence techniques, including machine learning, in evaluating bank efficiency and performance and concluded that these methods are highly effective. Similarly, Kou, et al. (2019) noted that machine learning techniques provide high accuracy in financial analysis and forecasting.

Machine learning methods offer powerful tools for assessing the impact of income-expense ratios on bank success. These methods reveal hidden patterns in complex datasets, providing more accurate predictions. Techniques such as artificial neural networks (ANN) and support vector machines (SVM) model the relationships between income-expense ratios and performance metrics, achieving high accuracy. Machado and Karray (2022) investigated the use of machine learning techniques in credit risk assessment in the banking sector, demonstrating their high accuracy.

Income-expense ratios are important financial indicators used to evaluate the financial health and performance of banks. These ratios, which include key indicators such as net interest margin, non-interest income/expense ratio, operating expense ratio, and loan loss provisions, provide insights into the bank's operational efficiency and profitability. Dietrich and Wanzenried (2011) highlighted the importance of these ratios in understanding bank performance, emphasizing that high income-expense ratios can negatively impact profitability. Similarly, Demirgüç-Kunt and Huizinga (1999) found that high income-expense ratios are associated with lower bank profitability.

Numerous studies have examined the impact of financial ratios comprising the income-expense ratio on bank performance. Athanasoglou, Brissimis, and Delis (2008) used panel data analysis (GMM) to investigate the determinants of return on equity and return on assets as dependent variables in Greek banks during the period 1985-2001. The findings showed that capital adequacy, personnel efficiency, and inflation positively affected bank profitability, while loan loss provisions, operating expenses, and concentration ratio negatively impacted profitability.

In a study by Jeon and Miller (2004), factors determining the performance of Korean banks during the period 1991-1999 were examined. Using balance sheet and income statement data from 16 domestic and foreign banks, the effects of these data on return on assets and return on equity as performance measures were investigated. The findings indicated that the ratio of equity to total assets, non-interest income, and the number of employees positively affected both return on assets and return on equity, while loan loss provisions and non-interest expenses had negative effects on both metrics.

Chantapong (2005) studied the impact of profitability in 23 domestic and foreign commercial banks operating in Thailand following the Asian crisis. In the study covering the period 1995-2000, return on assets was used as the dependent variable, while general administrative expenses, the ratio of loans to total assets, the ratio of non-performing loans to total assets, and the ratio of non-interest income to total assets were used as independent variables. The findings showed that the ratio of non-interest income to total assets positively affected return on assets, while the ratio of non-performing loans to total loans negatively impacted return on assets.

Sufian and Habibullah (2009) examined factors affecting the performance of 37 banks in Bangladesh during the period 1997-2004. Using average return on assets, average return on equity, and interest income ratio as measures of bank performance, the study found that liquidity risk, credit risk, and the ratio of non-interest expenses to total assets positively affected bank performance, while non-interest income had a negative impact.

Sufian (2009) studied the determinants of return on assets in domestic and foreign commercial banks operating in Malaysia during the period 2004-2004. The findings showed that the ratio of non-interest income to total assets, the ratio of general administrative expenses to total assets, and the ratio of equity to total assets positively affected banks' return on assets, while the ratio of non-performing loans to total loans negatively impacted return on assets.

In a study by Alper and Anbar (2011) covering the period 2002-2010, indicators affecting the profitability of banks operating in Türkiye were examined using panel data analysis. The study concluded that asset size and non-interest income positively affected bank profitability, while non-performing loans and the size of the loan portfolio negatively impacted profitability.

Onuonga (2014) investigated the internal factors determining the return on assets of six commercial banks operating in Kenya during the period 2008-2013. The findings showed that the ratio of non-interest income to total operating income, the ratio of total loans to total assets, and the ratio of equity to assets positively affected return on assets, while the ratio of total operating expenses to total assets had a negative effect.

In a study by Shah and Jan (2014), factors determining the performance of the top 10 private banks operating in Pakistan were examined. Using return on assets as a measure of bank performance, the study found that operational efficiency and bank size negatively impacted bank performance, while operational efficiency had a negative relationship with interest income, and bank size had a positive relationship with interest income.

Yesmine and Bhuiyah (2015) examined the performance of public and private banks operating in Bangladesh using multiple regression analysis. In the study covering the period 2008-2014, return on assets was used as a performance indicator. The study found that the ratio of total interest income to total costs and the ratio of operating profit to total assets positively affected banks' return on assets, while the ratio of loan loss provisions to total loans negatively impacted return on assets. Additionally, the most influential factor in bank performance was the ratio of operating profit to total assets for public banks and the ratio of loan loss provisions to total loans for private banks, while the ratio of loans to deposits was not effective for either bank group.

Özgür and Görüş (2016) examined the impact of bank-specific variables and macroeconomic variables on profitability in deposit banks operating in Türkiye. In the study covering the period 2006/1-2016/2, the ratio of equity to total assets and the ratio of net interest income to total assets positively affected bank profitability, while the ratio of non-performing loans to total cash loans negatively impacted profitability.

Sevim and Eyüboğlu (2016) examined factors determining the average return on assets of 22 commercial banks, including 3 public banks and 19 private banks, operating in Türkiye. In the study covering the period 2006-2014, banks with above-average return on assets were classified as "highly profitable," while banks with below-average return on assets were classified as "low profitable." The findings showed that non-interest income positively affected profitability in both bank groups, but this effect was higher in banks classified as "low profitable." On the other hand, non-interest expenses, considered an indicator of operational efficiency, negatively affected profitability in both bank groups, but this effect was much higher in banks classified as "low profitable."

Saldanlı and Aydın (2016) investigated factors affecting the profitability of 23 deposit banks operating in Türkiye during the period 2004-2014. In the study, the ratio of net profit (loss) to total assets was determined as the dependent variable, while the ratio of equity to total assets, the ratio of liquid assets to short-term liabilities, the ratio of net non-interest income to total assets, and the ratio of interest income to interest expenses were determined as independent variables. The findings showed that only the ratio of net non-interest income to total assets had positive and significant effects on the dependent variable.

In a study by Işık, Noyan-Yalman, and Koşaroğlu (2017), factors affecting the profitability of 20 deposit banks operating in Türkiye were investigated. In the study using data from the period 2006-2014, credit risk and liquidity management negatively affected bank profitability, while bank capital, interest income, and non-interest income positively affected profitability. Additionally, economic growth was found to have a positive impact on bank profitability.

As seen in many studies in the literature, income-expense ratios play a critical role in evaluating the financial performance of banks. Machine learning methods offer advanced artificial intelligence techniques for analyzing the impact of these ratios on bank success, providing more accurate and predictive forecasts. This study aims to contribute to the literature by using machine learning methods to comprehensively analyze the impact of income-expense ratios on bank performance in the banking sector.

3. Data and Methodology

The aim of the study is to examine whether the banks in the banking sector are successful in terms of their activities or not in terms of financial ratios. In this context, machine learning techniques (LR, kNN, SVM, ANN, CART, RF, and GB) are used to evaluate the financial performance of 24 selected banks operating in Türkiye for the 2012-2022 period. In the study, to measure the success of banks in the context of their activities, the ratio of net operating profit (loss) value, which expresses only the profits/losses obtained from activities to total assets was determined as the dependent variable. The average of the relevant variable was taken for each year in the period under consideration, and banks above the average were coded as successful, while banks below the average were coded as unsuccessful. The averages taken for each year, especially based on the period examined, allowed considering the vulnerabilities or opportunities that may arise when cyclical differences are considered. The independent variable of the study is the income-expense balance ratios.

Table 1. Data Describe

Ratio		Sub ratio
Banking Activity	S	Net Operating Profit (Loss) / Total Assets
Income-Expense Structure, %	R1	Net Interest Income After Special Provisions / Total Assets
	R2	Net Interest Income / Operating Gross Profit After Special Provisions
	R3	Non-Interest Income (Net) / Total Assets
	R4	Non-Interest Income (Net) / Other Operating Expenses
	R5	Other Operating Expenses / Operating Gross Profit
	R6	Loan Provisions / Total Assets
	R7	Interest Income / Interest Expenses
	R8	Total Revenues / Total Expenses
	R9	Interest Income / Total Assets
	R10	Interest Expenses / Total Assets
	R11	Interest Income / Total Revenues
	R12	Interest Expenses / Total Expenses

Source: TBB.

In the rapidly evolving field of data analytics, the selection of the appropriate machine learning algorithm has become a crucial consideration for researchers and practitioners alike. This paper provides a comprehensive evaluation of seven prominent machine learning methods: Logistic Regression, k-Nearest Neighbors, Support Vector Machines, Artificial Neural Networks, CART, Random Forest, and Light Gradient Boosting Machine.

The study begins by examining the performance of these algorithms across a diverse set of meta-data, as highlighted in previous research (Rashid et al., 2022). The authors note that the accuracy of these tools is highly dependent on the characteristics of the datasets used, underscoring the importance of careful algorithm selection and tuning.

A common approach for comparing supervised machine learning algorithms is to perform statistical comparisons of the accuracies of trained classifiers on specific datasets. However, this methodology has limitations, as it may not fully capture the nuances of real-world deployment scenarios(Olson et al., 2017) The authors therefore propose a more holistic evaluation framework that considers additional performance metrics beyond just accuracy, such as precision, recall, and F1-score.

One key finding from the literature is the potential pitfall of over-reliance on deep neural networks, which may not always be the optimal solution, particularly for certain types of tabular data.

The results of this analysis indicate that while deep neural networks have garnered significant attention in recent years, they are not always the optimal solution for every problem. In fact, simpler algorithms like Logistic Regression and CART may be preferable in certain contexts, particularly when interpretability and computational efficiency are key priorities (Olson et al., 2017).

4. Findings

To determine the success of banks in terms of their activities, average return on assets, average return on equity, pre-tax profit / total assets and net period profit (loss) / paid-in capital ratios within the main group of income-expense structure were discussed. Descriptive statistical values of the rates are given in Table 2.

Table 2. Income-Expense Structure Ratios Descriptive Statistics

Ratio	N	Mean	std	min	25%	50%	75%	max
R1	264	2.93	2.03	-6.53	2.10	2.79	3.40	19.29
R2	264	52.85	66.84	-956.33	44.10	58.40	68.51	174.06
R3	264	1.32	1.08	-4.41	0.78	1.25	1.74	6.54
R4	264	81.57	74.06	-159.26	38.34	62.44	113.19	407.69
R5	264	43.31	32.18	10.77	24.25	40.66	53.70	448.34
R6	264	1.10	0.99	-2.90	0.66	0.99	1.42	9.25
R7	264	216.35	122.55	101.14	167.76	187.95	213.87	1510.66
R8	264	157.19	49.84	82.70	130.08	143.64	163.91	473.86
R9	264	8.46	2.44	1.21	7.01	8.16	9.56	23.24
R10	264	4.49	1.84	0.34	3.49	4.28	5.27	13.09
R11	264	86.54	9.39	57.37	81.25	86.92	91.62	137.36
R12	264	67.11	15.14	16.56	57.94	67.58	79.42	92.23

Using the grid search method, parameters were tested in different combinations, considering income-expense ratios, and analyzes were carried out according to the parameters that gave the best results. The findings of the parameters of the methods used are presented in Table 3.

Table 3. Parameters for Machine Learning Methods for Income Expense Structure Ratios

Machine Learning Methods	Parametreler
Logistic Regression	-
K-Nearest Neighbor (KNN)	K=39
Support Vector Machines (SVM)	C=9, Kernel = rbf
Artificial neural networks	solver = "lbfgs",activation='logistic', alpha = 0.005, hidden_layer_sizes = (3, 5))
CART (Classification and Regression Tree)	max_depth = 5, min_samples_split = 20
Random Forests	max_features = 5, min_samples_split = 5, n_estimators = 100
Gradient Boosting Machines	learning_rate = 0.01, max_depth = 2, n_estimators = 300
XGBoost	learning_rate= 0.01, max_depth= 3, n_estimators= 1000, subsample= 0.6
Light GBM	learning_rate= 0.1, max_depth= 2, n_estimators= 100
CatBoost	depth= 5, iterations= 100, learning_rate= 0.01

Accuracy, precision, sensitivity and F1 criterion values considered in the machine learning methods used in the study are given. Accuracy metrics of the artificial intelligence algorithms used are included in Table 4 for income-expense structure ratios. XGBoost gave the highest value and KNN method gave the lowest value.

Table 4. Accuracy Metrics for Income-Expense Structure Ratios

Machine learning methods	Accuracy	Precision	Sensitivity/Recall	F1- criterion
Logistic Regression	0.81	0.81	0.81	0.81
KNN	0.75	0.77	0.76	0.75
Support Vector Machines	0.80	0.81	0.80	0.80
Artificial neural networks	0.78	0.80	0.78	0.77
CART	0.78	0.77	0.77	0.77
Random Forests	0.86	0.87	0.87	0.86
Gradient Boosting	0.84	0.84	0.84	0.84
XGBoost	0.86	0.87	0.87	0.86
Light GBM	0.84	0.84	0.84	0.84
CatBoost	0.80	081	0.80	0.80

When the findings are examined in terms of income-expense structure ratios, it is seen that the most effective criteria in determining whether the bank is successful in terms of its operations are total income / total expenses, net interest income after special provisions / total assets and other operating expenses / operating gross profit, respectively.

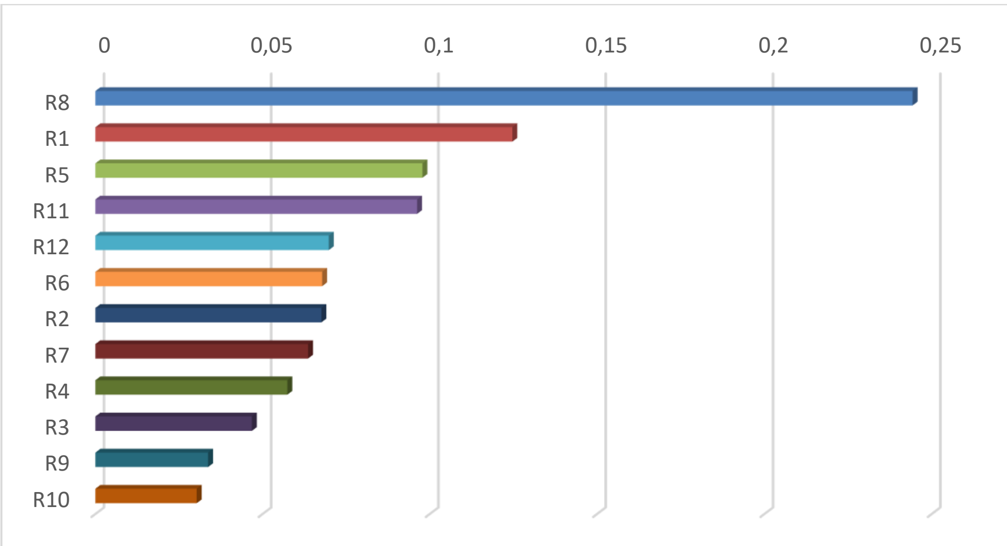

Figure 1. Importance Scores of Variables for Income Expense Structure Ratios

5. Conclusion

Based on the examination of the findings through the lens of income-expense structure ratios, it is evident that three principal financial metrics significantly determine the operational success of a bank. The most effective criterion is the total income to total expenses ratio, which indicates a bank's efficiency in managing its resources, with a higher ratio signifying strong financial health. The second critical metric is the net interest income after special provisions to total assets ratio, reflecting the bank's ability to generate income from its core lending activities after accounting for potential loan losses. A higher ratio in this context demonstrates effective risk management and operational efficacy. Lastly, the ratio of other operating expenses to operating gross profit measures the proportion of operating costs relative to the bank's gross profit from operations. A lower ratio here indicates efficient management of operating expenses, thereby maximizing profitability. Collectively, these ratios provide a comprehensive assessment of a bank's operational performance, emphasizing the importance of income generation, risk management, and cost control in achieving sustainable profitability and financial stability.

References

Anbar, A., & Alper, D. (2011). Bank specific and macroeconomic determinants of commercial bank profitability: Empirical evidence from Turkey. *Business and Economics Research Journal*, *2*(2), 139-152.

Athanasoglou, P. P., Brissimis, S. N., & Delis, M. D. (2008). Bank-specific, industry-specific and macroeconomic determinants of bank profitability. *Journal of International Financial Markets, Institutions and Money*, *18*(2), 121-136.

Chantapong, S. (2005). Comparative study of domestic and foreign bank performance in Thailand: The regression analysis. *Economic Change and Restructuring*, *38*(1), 63-83.

Demirgüç-Kunt, A., & Huizinga, H. (1999). Determinants of commercial bank interest margins and profitability: some international evidence. *The World Bank Economic Review*, 13(2), 379-408.

Dietrich, A., & Wanzenried, G. (2011). Determinants of bank profitability before and during the crisis: Evidence from Switzerland. *Journal of International Financial Markets, Institutions and Money*, *21*(3), 307-327.

Fethi, M. D., & Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research*, *204*(2), 189-198.

Işık, Ö., Yalman, İ. N., & Koşaroğlu, Ş. M. (2017). Türkiye'de mevduat bankalarının karlılığını etkileyen faktörler. *İşletme Araştırmaları Dergisi*, *9*(1), 362-380.

Jeon, Y., & Miller, S. M. (2004). The effect of the Asian financial crisis on the performance of Korean nationwide banks. *Applied Financial Economics*, 14(5), 351-360.

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., & Herrera Viedma, E. (2019). Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, 25(5), 716-742.

Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications*, *200*, 116889.

Olson, R S., Cava, W L., Mustahsan, Z., Varik, A., & Moore, J H. (2017, January 1). Data-driven Advice for Applying Machine Learning to Bioinformatics Problems. Cornell University. https://doi.org/10.48550/arxiv.1708.05070

Onuonga, S. M. (2014). The analysis of profitability of Kenyas top six commercial banks: Internal factor analysis. *American International Journal of Social Science*, *3*(5), 94-103.

Ozgur, O., & Gorus, M. S. (2016). Determinants of deposit bank profitability: evidence from Turkey. *Journal of Applied Economics and Business Research*, 6(3), 218-231.

Rashid, F., Gargaare, S M A., Aden, A H., & Abdi, A. (2022, January 1). Machine Learning Algorithms for Document Classification: Comparative Analysis. *Science and Information Organization*, 13(4). https://doi.org/10.14569/ijacsa.2022.0130430

Saldanli, A., & Aydin, M. (2016). Bankacılık sektöründe karlılığı etkileyen faktörlerin panel veri analizi ile incelenmesi: Türkiye Örneği. *Ekonometri ve İstatistik Dergisi*, (24), 1-9.

Sevim, U., & Eyüboğlu, K. (2016). Ticari banka performansının içsel belirleyicileri: Borsa İstanbul Örneği. *Doğuş Üniversitesi Dergisi*, *17*(2), 211-223.

Shah, S. Q., & Jan, R. (2014). Analysis of financial performance of private banks in Pakistan. *Procedia-Social and Behavioral Sciences*, 109, 1021-1025.

Sufian, F. (2009). Factors influencing bank profitability in a developing economy: Empirical evidence from Malaysia. *Global Business Review*, *10*(2), 225-241.

Sufian, F., & Habibullah, M. S. (2009). Determinants of bank profitability in a developing economy: Empirical evidence from Bangladesh. *Journal of Business Economics and Management*, *10*(3), 207-217.

Yesmine, S., & Bhuiyah, M. S. U. (2015). Determinants of banks' financial performance: A comparative study between nationalized and local private commercial banks of Bangladesh. *International Journal of Business and Management Invention*, *4*(9), 33-39.