

Article

Not peer-reviewed version

Machine Learning for Predicting Bank Stability: The Role of Income Diversification in European Banking

Karim Farag , [Loubna Ali](#) , [Noah Cheruiyot Mutai](#) * , [Rabia Lugman](#) * , Ahmed Mahmoud , Nol Krasniqi

Posted Date: 21 April 2025

doi: [10.20944/preprints202504.1691.v1](https://doi.org/10.20944/preprints202504.1691.v1)

Keywords: income diversification; bank stability; Support Vector Machines (SVM); Random Forest (RF); financial supervision



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Machine Learning for Predicting Bank Stability: The Role of Income Diversification in European Banking

Karim Farag ¹, **Loubna Ali** ², **Noah Cheruiyot Mutai** ^{1*}, **Rabia Luqman** ¹, **Ahmed Mahmoud** ²
and **Nol Krasniqi** ¹

¹ Faculty of Economics and Business Administration, Berlin School of Business and Innovation (BSBI), Germany

² Faculty of Computer Science and Informatics, Berlin School of Business and Innovation (BSBI), Germany

* Correspondence: noah.mutai@berlinsbi.com

Abstract: This study investigates the impact of income diversification on bank stability in European financial institutions, leveraging machine learning algorithms to enhance predictive accuracy. Using a dataset of 315 commercial banks across 29 European countries from 2011 to 2021, we apply Support Vector Machines (SVM) and Random Forest (RF) classifiers to evaluate how different income structures influence the likelihood of bank distress. Our findings indicate that banks with higher levels of non-interest income diversification are more prone to instability, particularly in low-regulation environments. Among the models tested, SVM outperforms RF in classifying at-risk banks, offering a more reliable early warning signal. These results contribute to the growing body of FinTech research by demonstrating how advanced algorithms can support real-time financial supervision, risk assessment, and regulatory decision-making in the banking sector.

Keywords: income diversification; bank stability; Support Vector Machines (SVM); Random Forest (RF); financial supervision

JEL Classification: G21; G32; C45; G28; C55

1. Introduction

Banks are instrumental in advancing sustainable economic development by efficiently allocating capital to productive sectors, thereby enhancing output and long-term growth. However, the post-crisis financial environment has become increasingly dynamic, requiring banks to adopt more proactive and strategic approaches to portfolio and risk management [1]. In response to systemic vulnerabilities exposed during the global financial crisis, regulatory standards such as Basel III increased the minimum capital requirement to 10.5% of risk-weighted assets, up from 8% under Basel II [2]. This shift underscores the importance of accurate risk exposure measurement for ensuring financial system resilience and supporting economic stability.

Globalization and financial reforms have further driven the deregulation of banking activities and expanded the scope for income diversification. Financial institutions are increasingly adopting digital tools and FinTech-enabled services to improve operational efficiency, expand product offerings, and manage competition [3]. Consequently, banks have shifted away from a heavy reliance on interest income toward a broader mix of fee-based and off-balance-sheet activities—such as securitization, derivatives trading, trade finance, underwriting, and financial advisory services—which theoretically help mitigate concentration risk [4]. Despite its theoretical appeal, the impact of income diversification on bank stability remains contentious. Some scholars argue that it contributes to systemic risk, particularly when fee-based activities involve complex financial instruments like subprime mortgages and derivatives—key triggers of the 2007–2009 financial crisis [1,5]. In response, regulatory bodies such as the Bank of England enforced the separation of retail and investment

banking, while Asian regulators—including in Korea and Taiwan—imposed restrictions on off-balance-sheet activities to protect financial stability [6].

This ongoing debate calls for a reassessment of income diversification strategies through advanced analytical tools. This study contributes to the FinTech literature by evaluating the effect of income diversification on bank stability using both econometric and machine learning approaches. By combining Generalized Method of Moments (GMM) with predictive algorithms like Support Vector Machines (SVM) and Random Forests (RF), the study explores the causal linkages and predictive potential of diversification in the context of European commercial banks. Our findings aim to support evidence-based regulation and highlight how machine learning can be leveraged for real-time risk monitoring in a fast-evolving financial ecosystem.

According to Figure 1, the income diversification ratio of European commercial banks increased noticeably in 2015. This suggests that fee-based and off-balance sheet activities increase more frequently than traditional investments in loans. Basel III emphasized the importance of asset diversification to better meet capital requirements. Furthermore, [7] stated that following the 2007–2009 financial crisis, banks were increasingly motivated to diversify from conventional assets to protect themselves from credit and insolvency risks. Additionally, [1] contended that the decline in market interest rates caused European banks to reallocate their reserves to non-traditional assets. Furthermore, banks' use of technology to expand internationally and join global financial markets was greatly aided by technological advancements, which in turn increased non-interest-bearing investment. On the other hand, as illustrated in Figure 2, Bank Z's score has considerably improved, however, in 2021, the Z score decreased while income diversification increased, suggesting a hazy explanation that necessitates further study on the relationship between income diversification and bank stability to better advise regulators and bankers on how to manage income for a consistent level of profitability that supports the growth of the European economies.

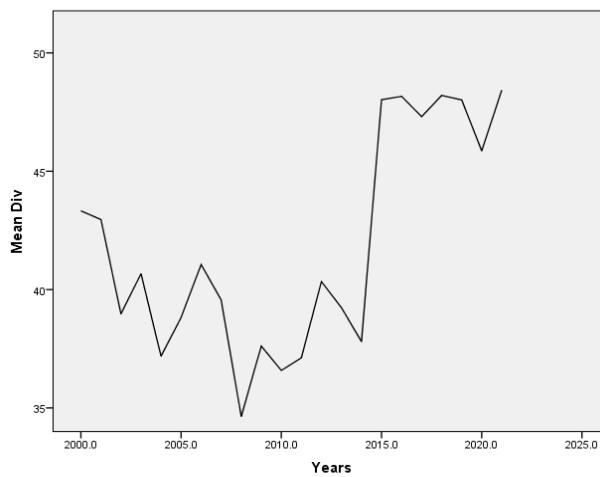


Figure 1. Income Diversification Ratio in Europe during 2000–2021.

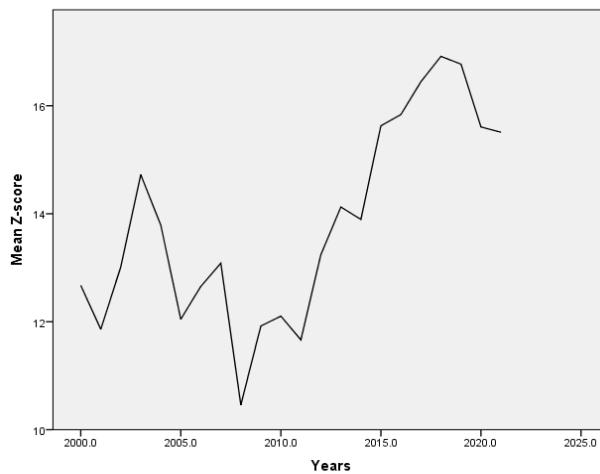


Figure 2. Bank Z score in Europe during 2000-2021.

Literature Review

Previous Studies in Europe

[8] investigated the relationship between income structure and bank profitability using panel data from 2002 to 2012 and a sample of European banks to observe how the banks' practices changed following the crisis. The results showed that while income diversification has a negative effect on profitability, this effect decreases during times of crisis arguing that income diversification is more favorable when banks are expecting any crisis but in normal events, it is favorable to have more diversification in traditional investments. Additionally, [9] employed regression to study the effect of income diversification on bank stability and profitability by using a sample of 1250 banks in the USA and Europe from 2008 to 2016. The findings found that income diversification is positively related to bank stability in the USA while in Europe, it has an insignificant effect on bank stability and profitability. This suggests that European banks would rather diversify their traditional investments in loans than non-interest income-bearing investments to increase profitability and bank stability.

Furthermore, [10] used the panel smooth transition regression model to explore the impact of income diversification on bank stability in 114 European commercial banks, utilizing panel data from 2010 to 2019. According to the findings, increasing income diversification through non-traditional banking activities has a detrimental impact on bank stability and financial performance. In addition, [11] examined the impact of non-interest income, loan, and geographic diversification during the COVID-19 pandemic using a sample of 56 European banks. The results showed that, in contrast to loan and geographic diversifications, the non-interest income ratio is the only variable that supports enhancing the stability of the European banks during the crisis and pandemics. Moreover, [12] employed the system GMM model to study the effect of income diversification on bank performance during the COVID-19 period, using a sample of 1,231 banks in 90 countries from 2018 to 2021. The findings revealed that income diversification has a positive effect on bank stability in developing and developed countries during the COVID-19 periods which confirms the importance of encouraging banks to engage in fee-based, trading, and FOREX activities to absorb the negative effect of the recessionary periods for more stability and growth in the credit markets.

Previous Studies in the Rest of the World

[13] argued that nonperforming loans (NPL), which are defined as loans that have been passed due for more than 90 days without payments, are used to gauge a bank's exposure to credit risk, which is a key factor in determining the stability of the bank. To prevent unforeseen bad debt

expenses that could jeopardize bank stability and degrade capital, banks must lower the non-performing loan ratio. In this regard, banks must diversify their loan portfolios by lending to businesses in various industries and offering a range of loan products, including credit cards, mortgages, auto loans, personal loans, commercial and industrial loans, and so forth. However, this diversification proves unsuccessful during recessions and crises, leading most banks to reallocate a portion of their reserves to non-interest-bearing investments to safeguard their capital against unforeseen losses during recessions and crises. In China,[14] employed the GMM to investigate the effect of income diversification on bank stability by using a sample of 101 Chinese banks with panel data from 2006 to 2016. The results indicated that income diversification has a negative effect on bank stability. They argued that the reason behind such a negative relationship is that the Chinese banks are still at the early stage of non-interest activities and have limited control. Additionally, increasing banks' engagement in non-interest activities could reduce banks' concern about their core business in loan investments raising the volatility of income and a lack of government supervision. In this respect, Chinese banks are more likely to diversify their traditional loan investments than non-traditional ones to improve their financial stability.

Additionally, [5] studied a sample of 200 commercial banks operating in South Asian countries discovering that income diversification has a positive influence on bank stability, except for fees and commission income activities, which have a negative impact on bank stability, implying that not all non-interest activities are beneficial to the financial health of the South Asian banking system. Additionally, [15] employed a fixed effect model using a sample of commercial banks from Malaysia and the results revealed that income diversification enhances the financial performance of the banks. Furthermore,

[3] used fixed-effect and GMM models to investigate the impact of income diversification on bank stability by taking a sample of 169 BRICS commercial banks from 2001 to 2015. The findings showed that the income diversification of large-sized banks positively affected bank performance.

In contrast, the small-sized ones had a negative effect, which provided better insights to regulators that income diversification is not favorable for all. Vidyarthi (2020) used a sample from Indian banks and confirmed the results of Sharma and Anand (2018). Furthermore, [16] discovered that Islamic banks in the Gulf Cooperation Council (GCC) nations prefer income diversification to increase bank stability. In contrast, the results of conventional banks showed that income diversification positively affected non-performing loans (NPL) and negatively impacted Z-score. This suggests that the conventional banks in the GCC would become less stable if they relied too much on income diversification from fee-based and off-balance sheet activities. Additionally, [6] used multivariate regression on a sample of commercial banks operating in 34 countries members of the Organization for Economic Co-operation and Development (OECD) using unbalanced panel data from 2002 to 2012. The results showed that while a moderate increase in income diversification can improve bank stability, excessive diversification, particularly during a crisis, can worsen stability.

This highlights the significance of traditional investment concentration in loan and deposit investments during a crisis rather than engaging in non-interest-bearing investments to stabilize bank operations. Moreover, [17] used the GMM to study the effect of income diversification on bank stability, using a sample of Vietnamese commercial banks from 2006 to 2015. The findings revealed that relying on fee-based activities rather than traditional loan investments can reduce bank stability. In addition, [18] employed the GMM to explore how income diversification affects bank performance in Sub-Saharan banks by conducting a comparative study among emerging, regional, and global banks. The findings showed that income diversification enhanced bank performance in global and emerging markets than the regional African and domestic banks which demonstrates the importance of adopting diversification in bank investments to stabilize their financials. Further, [18] used the GMM on a sample of 48 banks operating in India to investigate how bank diversification influences bank stability, with geographic, loan portfolio, and functional diversifications as dependent variables in the study. The findings demonstrated that all levels of bank diversification have a positive impact

on stabilizing bank performance, arguing that more engagement in different alternative investments would reduce the overall risk and increase the stability of the banks.

In addition, [19] used a regression model to investigate the impact of income diversification on bank stability from 2008 to 2017 using a sample of Tunisian commercial banks. The results demonstrated that income diversification significantly and favorably affects bank stability, suggesting that greater income diversification will raise the Z-score indicating improvement in bank stability. Furthermore, [20] examined a sample of 45 African commercial banks from 2000 to 2020 and discovered that income diversification improves bank stability whereas excessive diversity diminishes it. He also discovered that larger liquidity and interest margins, as well as increased operational inefficiencies, had a negative influence on bank stability. In contrast, GDP and inflation have a significant impact on banks' financial health. Further, [21] used panel data from 2002 to 2019 using GMM to examine the effects of income and asset diversification on bank stability in the United States commercial banks. Bank stability is positively impacted by assets and funding diversification, which in turn encourages banks to increase their traditional lending investments. On the other hand, they discovered that revenue diversification adversely affects bank stability, which makes banks' financial issues worse. [21] contends that income diversification is favorable to bank stability, however, excessive diversification could negatively affect bank stability in the African markets.

Furthermore, [21] used panel data from 2012 to 2021 and a sample from MENA countries to examine the effect of income diversification on bank stability using the fixed effect regression model. The results showed that income diversification significantly and positively affects bank stability. Additionally, [13] and [22] used a sample of Egyptian commercial banks with panel data from 2011 to 2020 to examine the impact of macroeconomic and bank-specific factors on bank stability. They found that bank-specific factors had a greater impact on corporate credit risk than on retail credit risk, additionally, the findings of the income diversification found an insignificant effect on retail and corporate NPL. According to [4] Zimbabwean commercial banks have a low-income diversification ratio because they rely heavily on loans and neglect investments in off-balance sheet and fee-based income activities. This makes these banks susceptible to high systemic risk. To investigate the impact of income diversification on bank performance, they used the modified OLS and difference GMM. The results showed that income diversification has a positive effect on banks' ROE, highlighting the necessity for Zimbabwean banks to alter their revenue strategy by increasing their level of diversification to improve stability and growth. Additionally, a sample of 271 commercial banks operating in the MENA countries from 2009 to 2020 was used by [23] to examine the impact of asset and income diversification on bank stability using the two-step GMM. They additionally investigated how political stability influences the relationship between diversification and bank stability. The results showed that while income and asset diversification contribute to bank stability, a greater proportion of non-interest income compared to interest-income activities has a negative impact on the benefits of asset diversification, and political stability undermines bank stability, which in turn reduces the benefits of investment diversification. Furthermore, the advantages of diversification differ depending on the size and market power of banks. This indicates that larger banks may use diversification to lower systemic risk more effectively than smaller ones, which are more vulnerable to systemic risks.

Literature Gap

After reviewing the literature, particularly in the European countries, the paper concluded that there is no crystal-cut evidence that offers a clear relationship between income diversification and bank stability after the periods of COVID-19. Additionally, to the best of researchers' knowledge, the research found limited studies applied in Europe after COVID-19. Further, the researchers noticed that bank stability measured by Z-score has declined recently as shown in Figure 2 along with some struggles in the income diversification ratio as shown in Figure 1. As a result, the research aims to investigate the impact of income diversification on bank stability in the banking sector of Europe by formulating the following hypotheses.

H₁: Income diversification significantly affects bank stability in Europe.
 H_{1a}: Income diversification significantly affects bank Z-Score in Europe.
 H_{1b}: Income diversification significantly affects bank Credit risk in Europe.
 H_{1c}: Income diversification significantly affects bank insolvency risk in Europe.

2. Materials and Methods

This study investigates the effect of income diversification on bank stability using a balanced panel dataset comprising 572 observations from 26 European countries over the period 2000 to 2021. The countries included in the analysis are Germany, the United Kingdom, France, Italy, Austria, Belgium, Croatia, Cyprus, the Czech Republic, Denmark, Sweden, Estonia, Finland, Greece, Hungary, Ireland, Bulgaria, Latvia, the Netherlands, Poland, Malta, Portugal, Slovakia, Slovenia, Lithuania, and Luxembourg. Data were obtained from the World Bank's Global Financial Development Database (GFDD) and the World Development Indicators (WDI).

To capture the causal relationship between income diversification and bank stability, we employed the System Generalized Method of Moments (System GMM) estimator. This approach was selected over Difference GMM due to its superior efficiency in exploiting additional moment conditions and minimizing potential bias from weak instruments. Additionally, Fixed and Random Effects models were estimated for robustness checks. Bank stability was proxied using the Z-score (a measure of insolvency risk), the non-performing loan (NPL) ratio, and the capital adequacy ratio (CAR), consistent with prior studies [10,13,16].

In parallel, this study implemented two machine learning algorithms—Random Forest and Support Vector Machine (SVM)—to complement the econometric analysis with predictive modeling. These models were trained on the same macro-financial dataset to classify bank distress, defined using a binarized Z-score threshold. Random Forest combines the outputs of multiple decision trees to capture non-linear relationships and reduce overfitting through ensemble averaging. SVM was selected for its ability to operate effectively in high-dimensional feature spaces and to identify optimal hyperplanes for class separation. Both models were evaluated based on accuracy, precision, recall, and F1 score, and cross-validation was used to enhance generalization.

This dual-method approach—econometric modeling for inference and machine learning for prediction—provides a comprehensive framework to assess the implications of income diversification for bank stability, while accounting for potential endogeneity, heterogeneity, and non-linear dependencies.

$$\begin{aligned}\Delta ZScore_{it} = & \alpha_i + \sum \beta_1 DIV + \sum \beta_2 CON \\ & + \sum \beta_3 EFF + \sum \beta_4 SMR + \sum \beta_5 SPV \\ & + \sum \beta_6 PROF + \sum \beta_7 GDP + \sum \beta_8 INF + \sum \beta_9 UNEMP + e_{it} \\ \Delta NPL_{it} = & \alpha_i + \sum \beta_1 DIV + \sum \beta_2 CON \\ & + \sum \beta_3 EFF + \sum \beta_4 SMR + \sum \beta_5 SPV \\ & + \sum \beta_6 PROF + \sum \beta_7 GDP + \sum \beta_8 INF + \sum \beta_9 UNEMP + e_{it} \\ \Delta CAR_{it} = & \alpha_i + \sum \beta_1 DIV + \sum \beta_2 CON \\ & + \sum \beta_3 EFF + \sum \beta_4 SMR + \sum \beta_5 SPV \\ & + \sum \beta_6 PROF + \sum \beta_7 GDP + \sum \beta_8 INF + \sum \beta_9 UNEMP + e_{it}\end{aligned}$$

Additionally, the formulas used in the machine learning algorithms are as follows:

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP}+\text{FN})}$$

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{FP}+\text{FN}+\text{TN})}$$

$$\text{F1 Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

Table 1. Definition of Variables and Measurements.

Variables	Measurements
<u>Dependent variables:</u>	
Z-Score Non-performing loan (NPL) ratio Capital Adequacy Ratio (CAR)	(ROA + equity to assets ratio) / STDEV of ROA NPLs / total gross loans Total equity / total assets
<u>Independent variables:</u>	
Income Diversification (DIV) Concentration risk (CON) Operating efficiency (EFF) Stock market return (SMR)	Non-interest income / total income Assets of the three largest banks / total assets of all banks Total expenses / total income End of Period Market Capitalization–Beginning of Period Market Capitalization/ End of Period Market Capitalization
Stock price volatility (SPV)	Square Root of (Sum of (Daily Return minus Average Daily Return) squared divided by Number of Days)
Profitability (PROF) Economic growth (GDP) Inflation (INF) Unemployment (UNEMP)	ROA, ROE, and NIM Real GDP growth rate CPI in Current Year – CPI in Previous Year / CPI in Previous Year Number of Unemployed People / Total Labor Force

3. Results

Descriptive Analysis

The paper conducted a descriptive analysis by describing the collected data in terms of mean, standard deviation (STDEV), minimum (MIN), and maximum (MAX) as shown in Table 2. The average score of Z-Score is 13.78 which indicates that most of the banks in Europe are well capitalized and have stable levels of earnings, which shows that the banks are stable. Furthermore, with an STDEV of 6.97% indicating a moderate level of volatility in the NPL ratio, the mean of the NPL ratio is 5.72%, indicating that the commercial banks of Europe are confronting a moderate level of credit risk exposure, reaching almost 6%. Furthermore, banks in Europe hold more capital than is necessary to be prepared to absorb any unforeseen losses in their portfolios, as evidenced by the CAR's means of 16.19%, which is significantly higher than the minimum regulatory capital requirements. Additionally, the CAR's STDEV of 4.51 indicates some stability in the CAR level. Moreover, the DIV's mean of 41.77% demonstrates that non-interest income accounts for 41.77% of the total income generated by European banks.

Furthermore, the average EFF is 58.90%, signifying that, on average, all expenses in European banks equal 58.90% of the total income. In addition, the SMR has the highest STDEV which shows that the European stock market indices have high volatility from its mean of 6.27% followed by the CON, ROE, and EFF having a STDEV of 16.75%, 13.87, and 12.09 respectively. Additionally, Europe's average ROA is 0.62%, whereas the USA and UAE have averages of 1.6% and 1.6%, respectively. Additionally, the European GDP growth rate is 2.43%, the inflation rate is 2.24%, and the

unemployment rate is 4.70%. All these figures appear to be normal, and the dataset shows no anomalies. Nonetheless, the data shows that DIV, NPL, CAR, and EFF are high in addition to high STDEV in Z-Score, highlighting the importance of researching how income diversification affects bank stability to give regulators and bankers better insights for improved financial and economic outcomes.

Table 2. Descriptive Analysis.

Variables	Mean	STDEV	MIN	MAX
Z-Score	13.78	9.25	-.33	57.44
NPL	5.72	6.97	.10	47.75
CAR	16.19	4.51	7.00	35.65
DIV	41.77	12.96	7.39	82.49
CON	71.63	16.75	28.56	100.00
EFF	58.90	12.09	14.75	97.17
SMR	6.72	24.84	-74.56	124.98
SPV	20.34	8.85	6.33	61.52
ROA	.62	1.06	-9.53	4.36
ROE	7.41	13.87	-117.67	37.46
NIM	2.11	1.18	.18	5.79
GDP	2.43	3.93	-14.84	24.48
INF	2.24	2.09	4.45	15.40
UNEMP	4.70	2.66	.996	20.86

Regression Results and Discussion

As shown in Tables 3 and 4, the study tested the hypotheses of the gathered data using the GMM and Fixed-effect models. The P-value of the Sargan test for all the GMM models exceeds 0.05 as shown in Table 3 indicating that the instruments are valid, and they are not correlated with the error term. Additionally, the p-values of the autocorrelation are above 0.05 which means that there is no significant autocorrelation in the residuals. Moreover, the p-values of the Wald test are less than 0.05 which shows that the tested coefficients are significant which signifies that the chosen independent variables have a significant impact on the NPL, CAR, and Z-Score. Accordingly, the models are robust and reliable. The findings of the GMM model illustrated that DIV, CON, and ROA had a negative impact on NPL and a positive effect on Z-Score suggesting that higher profitability and more income diversification and concentration of assets in the largest banks can lower credit risk and improve bank stability in Europe. Therefore, the results support H_{1a} and H_{1b} and are consistent with the findings of [3,11,15,23,24]. This suggests that the advantages of income diversification and superior risk management techniques enable large banks to manage their portfolios better. Moreover, the fixed effect results support the GMM findings that increased income diversification, and ROA can raise the stability levels of European banks, highlighting the significance of diversifying income through non-traditional investments.

The GMM results found that DIV has an insignificant effect on CAR which illustrates that income diversification does not affect the amount of capital held by the banks arguing that the amount of capital varies based on the changes in the level of credit risk and insolvency risk expectations. In this regard, the results do not support H_{1c}. In addition, the ROA was found to be significant and negatively associated with NPL stating that low-profit margin banks are more incentive to engage in risky investments leading to a higher level of NPL. In this regard, the low-profit margin European banks should be more careful while they are managing their portfolios to avoid any unexpected insolvency risk that might threaten their survival and growth in the credit markets. Moreover, the EFF had a negative significant impact on NPL, confirming the need to keep expenses under control in relation to income to better minimize credit risk exposure and increase operational stability. The SMR had a negative impact on CAR, claiming that lower stock market index

levels indicate larger economic issues that put further strain on bank capital as corporate borrowers experience financial difficulties, compelling banks to boost their capital buffer to protect themselves from insolvency risk threats. On the other side, the findings of the macroeconomic variables were found significant showing that GDP has a negative association with NPL and CAR indicating that at times of booming the borrowers have better repayment capacity to fulfill their obligations to banks reducing the level of NPL while lower level of inflation enhanced bank stability and increases in unemployment rate raises the level of NPL which threatens bank stability. In this regard, European banks should carefully monitor the macroeconomic indicators to act accordingly to protect their bank solvency from any negative threats.

Table 3. The Results of the GMM Model.

Variables	NPL		CAR		Z-Score	
	Estimate	Pr(> t)	Estimate	Pr(> t)	Estimate	Pr(> t)
Lags in dep. Var.	0.6617440	0.0000***	0.8041697	0.0000 ***	-0.057710	0.3598736
DIV	-0.0600427	0.054423.	-0.0062308	0.775403	0.176086	0.000 ***
CON	-0.0142226	0.503962	0.0141811	0.086588.	0.091744	0.00017 ***
EFF	-0.0602007	0.020025 *	0.0096383	0.543389	0.039840	0.1938525
SMR	0.0075872	0.235614	-0.0140217	0.003924 **	-0.014962	0.2316336
SPV	0.0348037	0.231596	-0.0283258	0.198989	-0.090804	0.1100486
ROA	-0.3435750	0.358737	0.1989617	0.094636	1.603595	0.0542526.
ROE	-0.0330053	0.02612 *	-0.0062834	0.722972	-0.118989	0.1230773
NIM	-0.4996805	0.271358	-0.1693435	0.430917	1.211837	0.0225761 *
GDP	-0.1139556	0.02793 *	-0.1085273	0.0000 ***	-0.117773	0.3139549
INF	0.0124079	0.91447	-0.2250472	0.006123 **	-0.392149	0.000541 ***
UNEMP	07747466	0.00010 ***	0.0937027	0.195391	0.113528	0.6346114
Sargan Test	26 (p-value = 1)		26 (p-value = 1)		26 (p-value = 1)	
Autocorrelation Test (1)	normal = -2.059052 (p-value = 0.339489)		normal = -2.291137 (p-value = 0.521955)		normal = -1.408896 (p-value = 0.15887)	
Autocorrelation Test (2)	normal = 0.5974548 (p-value = 0.5502)		normal = 1.140559 (p-value = 0.25405)		normal = 0.7279215 (p-value = 0.46666)	
Wald Test	chisq(12) = 1668.669 (p-value = < 0.00000)		chisq(12) = 1831.75 (p-value = < 0.00000)		chisq(12) = 2318.332 (p-value = < 0.00000)	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 4. The Results of the Fixed Effect Model.

Variables	NPL		CAR		Z-Score	
	Estimate	Pr(> t)	Estimate	Pr(> t)	Estimate	Pr(> t)
DIV	0.042937	0.033569 *	0.09224	0.0000***	0.09224	0.0000***
CON	0.045269	0.054029.	0.036232	0.06349.	0.03623	0.06349.
EFF	0.000566	0.979999	0.01904	0.30787	0.01904	0.30787
SMR	-0.02635	0.006476 **	-0.00969	0.22320	-0.00969	0.22320
SPV	-0.05842	0.033568 *	-0.117733	0.0000***	-0.11773	0.0000***
ROA	-1.02345	0.010345 *	1.337261	0.0000***	1.33726	0.0000***
ROE	0.023166	0.488981	-0.11047	0.0000***	-0.11047	0.0000***
NIM	-0.21657	0.576269	-0.54226	0.09209.	-0.54226	0.09209.
GDP	0.080210	0.228705	-0.22664	0.0000***	-0.22664	0.0000***
INF	-0.21054	0.044879 *	-0.59337	0.0000***	-0.59337	0.0000***
UNEMP	1.893341	0.000000 ***	-0.02799	0.78968	-0.0279898	0.78968

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Machine Learning Results and Discussion

In this section, the paper uses the machine learning algorithm to study the relationship between income diversification (DIV) and bank stability (Z-Score binary) and to predict the financial distress in banks in Europe. The paper utilized a binary classification approach to classify the Z-score above 3 as a non-distressed bank while the Z-score below 3 is a distressed bank. To do so, the paper used a threshold of 3 which is used in the financial distress prediction models. The threshold confirms that banks with strong financial health are categorized as non-distressed banks, while those at higher risk of bankruptcy are classified as distressed banks. In this regard, this binary classification approach simplifies decision-making, providing a clear-cut distinction between stable banks and troubled ones. Therefore, the paper used two types of machine learning algorithms: Random Forest and Support Vector Machine (SVM). Random Forest was adopted for its ability to handle large, complex datasets, manage non-linear relationships, and provide robust predictions by aggregating the results of multiple decision trees. As a result, it reduces the risk of overfitting and enhances model generalization. On the other side, SVM is selected for its ability to work effectively in high-dimensional spaces and its ability to find the optimal hyperplane that separates the data points into distinct classes. These algorithms were trained on the financial metrics DIV and evaluated based on their ability to predict the Z-score Binary classification. Metrics such as accuracy, F1 score, and confusion matrix components (True Positives, True Negatives, False Positives, and False Negatives) were used to assess the models' performance. Furthermore, the use of a Z-score threshold of 3 is consistent with industry practices for financial distress prediction, making the binary classification approach both practical and interpretable by comparing the performance of Random Forest and SVM, we aimed to determine which model best captures data patterns and provides the most reliable predictions for financial distress detection.

In this section, the paper used the importance features of the Random Forest as shown in Figure 3 to illustrate how the income diversification and control variables are valuable or useful features and how much each feature contributes to the dependent variable of the Z-Score. The results illustrated that income diversification has a positive effect on bank stability which is consistent with the results of the regression models arguing that banks of Europe are recommended to increase their diversification in fee-based income activities to raise the bank stability levels. Additionally, it shows that income diversification has a more positive impact on Z-score than CAR, ROE, ROA, OPEFF, GDP, and stock market return. On the other hand, the findings also revealed that NIM, NPL, UEMP, INF, and CONS have a negative effect on Z-Score which is supported by the regression results.

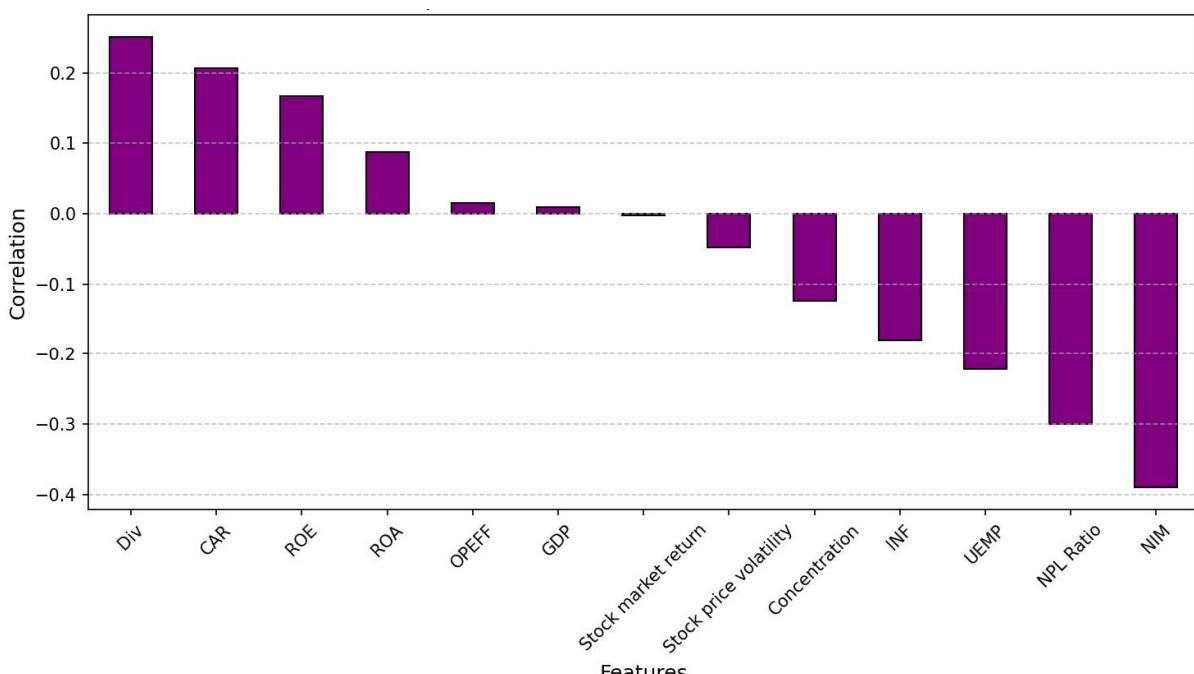
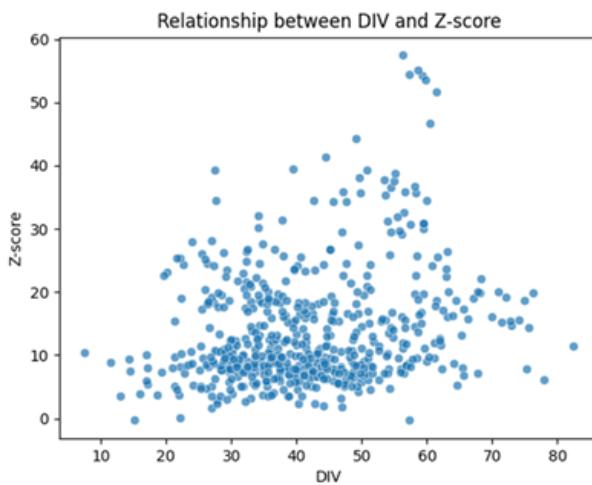
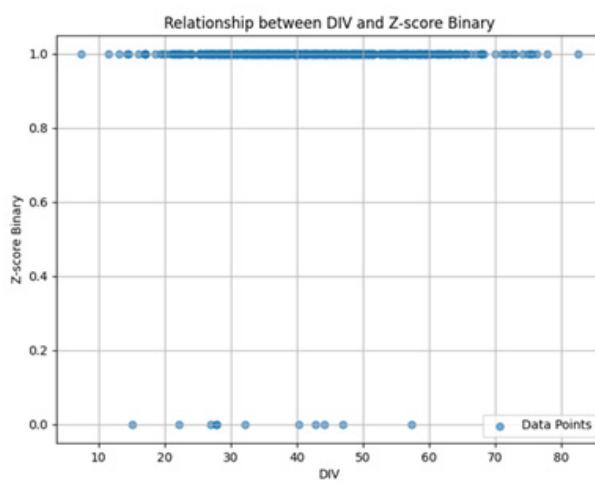


Figure 3. Importance Features.

In Figure 4 the paper used a scatter plot to visualize the relationship between the DIV and Z-score. The X-axis represents the variable of DIV with values ranging from 0 to 100 while the Y-axis represents the variables of Z-score with values ranging from 0 to 60. Additionally, the blue dots represent a specific data point demonstrating the combination of the DIV and Z-score values for that point. As shown in Table 4 below, we can notice that there are many points clustered tightly together in a certain area, it indicates that there is a moderately strong relationship between income diversification and bank stability because some of the points are spread out more widely across the scatter plot. Additionally, the paper has separated the classes of the classification in the scatter plot to provide a clear separation between the two classes effectively as shown in Figure 5 after the classification of the Z-Score there are few blue dots scattered at 0 while most of the dots are clustered at 1 demonstrating that most of the points belong to class 1. The clear separation illustrates that the DIV value can be used to distinguish between the two classes effectively, the clustering points of 1 signify that higher DIV values are associated with Z-Score Binary 1 and lower DIV values are associated with Z-Score binary 0.

**Figure 4.** Scatter Plot between DIV and Z-Score.**Figure 5.** Scatter Plot Before classification After classification between DIV and Z-Score.

The article employed the confusion matrix to evaluate the performance of the Random Forest and SPV classification models, as illustrated in Table 5. The findings reveal that random forest and SVM accurately predict 160 and 162 positive observations, respectively, whereas both models

correctly predict 0 negative data. Furthermore, the incorrectly anticipated positive observation values are 2 and 0, whereas the inaccurate negative values are 2 and 0. Accordingly, the findings show that there are high TP and TN values which confirm that the models correctly predicted most of the positive and negative classes. On the other hand, the low FC and FN show that there are few misclassifications, demonstrating a good model performance. Furthermore, the accuracy was used to evaluate the correctness of the two models among the total number of cases studied. In this respect, the accuracy score of both models is 96.39% and 97.59% respectively as shown below in Table 5 which indicates that there are a good proportion of correct predictions in both models. Moreover, the F1 score is also adopted in the examination to evaluate the harmonic mean of precision and recall, developing a single metric that balances both concerns. In this regard, the scores of the F1 score show 0.98 and 0.99 respectively which means that both models have good balances between precision and recall. Therefore, the results illustrate that income diversification (DIV) has a strong relationship with bank stability measured by the Z-Score and both models can be used to predict the future values of Z-Score to improve the prediction level of future bank stability in banks of Europe.

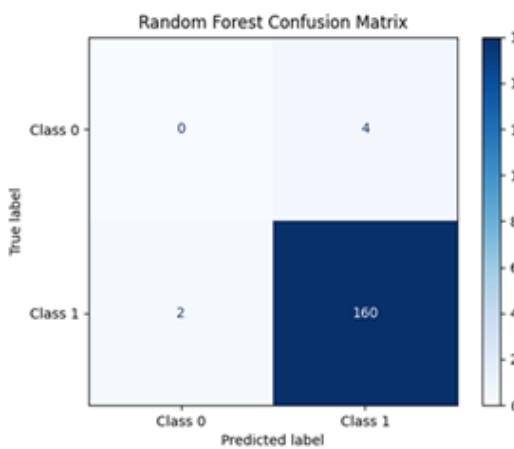


Figure 5. Confusion Matrix of Random Forest.

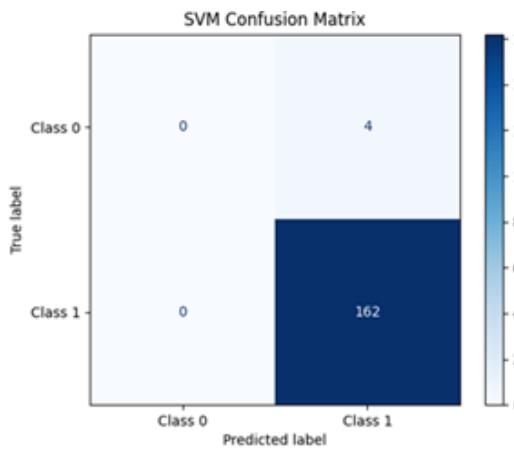


Figure 6. Confusion Matrix of SVM.

Table 5. The Results of The Confusion Matrix.

Confusion Matrix	Random Forest Scores	SVM Scores
TP (True Positive)	160	162
TN (True Negative)	0	0
FP (False Positive)	4	4
FN (False Negative)	2	0

Accuracy	96.39%	97.59%
F1 Score	0.98	0.99

5. Conclusions

The paper aimed to study the relationship between income diversification and bank stability in the banking sector of Europe by employing regression and machine learning approaches to provide a comprehensive analysis that helps to find answers to the ongoing debate about the pros and cons of income diversification in banks. Moreover, the paper aimed to enhance the prediction level by the bankers and to provide a better comprehension of the regulators which might affect positively making more effective regulations that can help in controlling the financial distress exposure by the European banks. The paper used the GMM and Fixed-effect models to examine the impact of income diversification along with some control variables on the bank stability measured by Z-score, NPL ratio, and CAR and the robust checks found such regression models are accurate and reliable econometric models that can be used for testing the hypotheses. The regression findings revealed that income diversification has a negative effect on NPL and a positive effect on Z-score, indicating that income diversification lowers bank credit risk exposure and enhances bank stability. On the other side, the research employed Random Forest and SVM as machine learning algorithms to create predictive models that can enhance the prediction performance of the bankers in the European banks. The results show that SVM has a higher accuracy score than the Random forest and both are above 95% which shows the importance of using the SVM model to predict the future movements of bank stability to be used as an early indicator to the bankers to make a proactive and precaution strategies that can help them in surviving and growing and to continue their effective intermediation role in supporting the growth of the European economies.

Funding: This research received no external funding

Data Availability Statement: Data can be requested from the corresponding author.

Acknowledgments: We acknowledge Berlin School of Business and Innovation for providing a conducive environment for this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lahouel, B.B.; Taleb, L.; Kossai, M. Nonlinearities between Bank Stability and Income Diversification: A Dynamic Network Data Envelopment Analysis Approach. *Expert Systems with Applications* **2022**, *207*, 117776, doi:10.1016/j.eswa.2022.117776.
2. Velasco, P. Is Bank Diversification a Linking Channel between Regulatory Capital and Bank Value? *The British Accounting Review* **2022**, *54*, 101070, doi:10.1016/j.bar.2021.101070.
3. Sharma, S.; Anand, A. Income Diversification and Bank Performance: Evidence from BRICS Nations. *International Journal of Productivity and Performance Management* **2018**, *67*, 1625–1639, doi:10.1108/IJPPM-01-2018-0013.
4. Dzingirai, C.; Dzingirai, M. Threshold Effect of Non-Interest Income Disaggregates on Commercial Banks' Financial Performance in Zimbabwe. *Heliyon* **2024**, *10*, doi:10.1016/j.heliyon.2024.e31379.
5. Nisar, S.; Peng, K.; Wang, S.; Ashraf, B.N. The Impact of Revenue Diversification on Bank Profitability and Stability: Empirical Evidence from South Asian Countries. *International Journal of Financial Studies* **2018**, *6*, 40, doi:10.3390/ijfs6020040.
6. Kim, H.; Batten, J.A.; Ryu, D. Financial Crisis, Bank Diversification, and Financial Stability: OECD Countries. *International Review of Economics & Finance* **2020**, *65*, 94–104, doi:10.1016/j.iref.2019.08.009.
7. De Meo, E.; De Nicola, A.; Lusignani, G.; Orsini, F.; Zicchino, L. European Banks in the XXI Century: Are Their Business Models Sustainable? In Proceedings of the 5th EBA Policy Research Workshop

"Competition in banking: implications for financial regulation and supervision"; November 2016; pp. 28–29.

8. Maudos, J. Income Structure, Profitability and Risk in the European Banking Sector: The Impact of the Crisis. *Research in International Business and Finance* **2017**, *39*, 85–101, doi:10.1016/j.ribaf.2016.07.034.
9. Rossi, S.; Dreassi, A.; Borroni, M.; Paltrinieri, A. Does Revenue Diversification Still Matter in Banking? Evidence from A Cross-Country Analysis. *Journal of Financial Management, Markets and Institutions* **2020**, *8*, 2050003, doi:10.1142/S2282717X20500036.
10. Ben Lahouel, B.; Taleb, L.; Kočišová, K.; Ben Zaied, Y. The Threshold Effects of Income Diversification on Bank Stability: An Efficiency Perspective Based on a Dynamic Network Slacks-Based Measure Model. *Annals of Operations Research* **2023**, *330*, 267–304.
11. Simoens, M.; Vander Vennet, R. Does Diversification Protect European Banks' Market Valuations in a Pandemic? *Finance Research Letters* **2022**, *44*, 102093, doi:10.1016/j.frl.2021.102093.
12. Ho, T.H.; Nguyen, D.T.; Luu, T.B.; Le, T.D.; Ngo, T.D. Bank Performance during the COVID-19 Pandemic: Does Income Diversification Help? *Journal of Applied Economics* **2023**, *26*, 2222964, doi:10.1080/15140326.2023.2222964.
13. Farag, K.; Kassem, T.; Ramzy, Y. The Crucial Macroeconomic and Microeconomic Determinants of Retail and Corporate Credit Risks. *Journal of Accounting Finance and Auditing Studies (JAFAS)* **2023**, doi:10.32602/jafas.
14. Wang, C.-C.; Lin, Y.-C. The Influence of Income Diversification on Operating Stability of the Chinese Commercial Banking Industry. *Romanian Journal of Economic Forecasting* **2018**, *21*, 38.
15. Brahmana, R.; Kontesa, M.; Gilbert, R.E. Income Diversification and Bank Performance: Evidence from Malaysian Banks. *Economics Bulletin* **2018**, *38*, 799–809.
16. Abuzayed, B.; Al-Fayoumi, N.; Molyneux, P. Diversification and Bank Stability in the GCC. *Journal of International Financial Markets, Institutions and Money* **2018**, *57*, 17–43.
17. Le, T.D. Geographic Expansion, Income Diversification, and Bank Stability: Evidence from Vietnam. *Cogent Business & Management* **2021**, *8*, 1885149, doi:10.1080/23311975.2021.1885149.
18. Chandramohan, K.; Lunawat, C.D.; Lunawat, C.A. The Impact of Diversification on Bank Stability in India. *Cogent Business & Management* **2022**, *9*, 2094590.
19. Alouane, N.; Kahloul, I.; Grira, J. The Trilogy of Ownership, Income Diversification, and Performance Nexus: Empirical Evidence from Tunisian Banks. *Finance Research Letters* **2022**, *45*, 102180.
20. Adem, M. Impact of Diversification on Bank Stability: Evidence from Emerging and Developing Countries. *Discrete Dynamics in Nature and Society* **2022**, *2022*, 7200725.
21. Shahriar, A.; Mehzabin, S.; Azad, M.A.K. Diversification and Bank Stability in the MENA Region. *Social Sciences & Humanities Open* **2023**, *8*, 100520, doi:10.1016/j.ssho.2023.100520.
22. Farag, K.M. The Impact of Macroeconomic and Microeconomic Variables on Credit Risk: A Comparative Study between Retail and Corporate Credit Risks in the Banks of Egypt. *MSA-Management Sciences Journal* **2024**, *3*, 116–128.
23. Shabir, M.; Jiang, P.; Shahab, Y.; Wang, W.; İşik, O.; Mehroush, I. Diversification and Bank Stability: Role of Political Instability and Climate Risk. *International Review of Economics & Finance* **2024**, *89*, 63–92, doi:10.1016/j.iref.2023.10.009.
24. Vidyarthi, H. Dynamics of Income Diversification and Bank Performance in India. *Journal of Financial Economic Policy* **2020**, *12*, 383–407, doi:10.1108/JFEP-05-2019-0084.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.