BANKING STABILITY MODELLING FOR ARMENIA: DECISION TREE APPROACH

Davit MINASYAN
PhD applicant at ASUE Department of EMM

Key words: banking stability, decision tree, modelling, Z-score.

Introduction

In Armenia, the banking sector is a critical component of the financial system, accounting for more than 85% of the entire financial structure. Therefore, the overall health of the financial system is largely determined by the stability of its banks. Banking stability is essential for a country's economic well-being: any deviation from stability can lead to significant disruptions across various economic sectors, potentially causing financial crises, loss of trust in the banking system, capital outflows, and even insolvency.

To ensure the stability of the banking system, it is important to identify and analyze the factors that influence banking behavior and predict future trends. However, the complexity and interconnectedness of economic systems make it difficult to detect hidden effects and interactions among different factors that could pose risks to banks. For example, systemic risks arising from regulatory changes can have widespread effects across the industry. Advanced machine learning and statistical models, such as decision trees and neural network models, can help capture these hidden risks. This article will explore the use of a decision tree model to better understand and enhance the stability of Armenia's banking sector.

Methodology

In this paper, the Decision tree model is implemented for estimating the Armenian banking stability and deciding which factors are playing crucial roles in changes in the stability score. For doing so, the data from 2013 January till 2023 September was used. The data is derived from the publicly available data of the Central Bank of Armenia, in monthly granularity. To measure banking stability of Armenia, is calculated the banking Z-score, which measures the distance from the default. The Z-score is calculated with the following formula:

Z-score=(ROA+(equity/assets))/sd(ROA), where: ROA indicates the return on assets

The independent variables have been chosen as macroeconomic indicators from different sectors of the economy: financial, real and foreign. The total number of factors is 47. For better performance of the model, all factors have been scaled using the StandardScaler function from the SKlearn library in Python. For each observation, it subtracts the mean and divides it to the standard deviation of the series.

Besides evaluating the current impact on the dependent variable, the corresponding lags of the independent variables have also been included in the model. The optimal Lag length has been selected with the help of the estimation of the VAR model. The lag length is selected by relying on Akaike information criteria. For testing decision tree model performance, the data is split into train and test subsets, with the proportion of 95:5. The model was trained on a training dataset, by predicting values matching the test dataset, and then comparing mean squared error among those 2 sets. The model was evaluated using the Python sklearn library.

Literature review

The application of tree-based models in banking stability have been mentioned in [Carmine, 2019], where the author describes the usage of regression tree technology with statistical algorithms like CRAGGING. It provides endogenously determined critical thresholds for a set of useful indicators, presented in the intuitive form of a decision tree structure. The framework takes into account the conditional relations between various indicators when setting early warning thresholds [Carmine, 2019]. Classification models and Decision trees started to be applied in economic modelling during the last decade, and have been borrowed from machine learning. One of the app-lications stands in early warning models, as described in the above-mentioned article.

Apart from decision tree models, random forest algorithms are also implemented in economic and financial stability modelling, aggregation of the results of a large number of single decision trees. Such an approach is mentioned in Alessi and Detken (2014), and has been used in identifying excessive credit and the associated build-up of systemic risk in the banking system [Alessi & Carsten, 2014]. About the application of decision trees in machine learning, there is a valuable work: "The Macroeconomy as a Random Forest" by Philippe Goulet Coulombe on a macroeconomic Random Forest (MRF), an algorithm adapting the canonical Machine Learning (ML) tool to flexibly model evolving parameters in a linear macro equation. Decision trees are widely used in financial modelling analyses, decision-making: choosing the best portfolio for investors [Maozhu, et.al., 2018].

In this article, the decision tree model will be used to predict banking stability modelling. The key indicator to determine whether the banking system is solvent or insolvent will be country-level banking Z-score, described by the World Bank group. The use of Z-score as a banking stability indicator has been mentioned in works of Edward Altman [Altman, 2000]. Also, Z-score is described in Steve Mercieca, et.al (2006) where the majority of institutions do not have publicly traded securities to examine the impact of a diversification index (Herfindahl Hirschman Index) on risk-adjusted performance, accounting measures of bank profits such as Return on Assets (ROA) and Return on Equity (ROE) can be used, to gauge bank performance [Mercieca et al., 2007].

Analysis

The modelling of banking stability is a crucial aspect for each country. Having a reliable and high-performance model will help a great deal to financial decision-makers and policy creators to make data-based decisions and be prepared for the upcoming shocks. One of the well-known data-based directions of banking stability modelling is early warning models, where upcoming crises are predicted using different factors. Machine learning is proven to have good prediction power and there are a lot of ML models capable of making good predictions. The further modelling is done on the bases of the decision tree model because of its 2 main advantages:

● simplicity of implication: it is quite intuitive to implement decision trees and requires only clear data and well-chosen factors;

● Feature importance: compared to other "black box" machine learning models, it gives feature importance, showing which factors are responsible for the prediction. Here, the term "Black box" means that the majority of ML models give predictions without possibility to explain the outcome, as complex algorithms and connections among variables are not easily interpretable or understandable by humans.

About decision trees According to IBM, a decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical tree structure, which consists of a root node, branches, internal nodes and leaf nodes. Decision tree analysis involves making a tree-shaped diagram to chart out a course of action or a statistical probability analysis. It is used to break down complex problems on branches. Each branch of the decision tree could be a possible outcome. The benefit of a decision tree is that it lists out all the possible outcomes and the gain or loss attached to each. Derived information can then be used by the stakeholders to make important decisions, such as the optimal investment level company is planning. The tree structure in the decision model helps in drawing a conclusion for any problem which is more complex in nature. The model is widely used in corporate finance, philosophy, economic forecasting, etc. Decision tree regression is a special category of decision tree modelling, designed to make predictions from continuous numerical values. Its clear and interpretable model allows for easy understanding of the underlying rules and patterns. It handles both categorical and numerical features, automatically handles missing values and outliers, and is robust to irrelevant features. However, it can be prone to overfitting and sensitivity to small variations [Viswa, 2023]. Considering all the above, decision tree regression has been chosen for evaluating banking stability.

In Armenian Banking stability Modelling, The Central Bank of Armenia mainly relies on financial risk maps, stress tests as well as implements complex Bayesian model averaging approach to model banking stability and find indicators affecting banking stability. The implication of decision tree models, as well as comparison of the result of the models with currently used approaches, can benefit by modelling non-linear relationships more precisely, making better predictions of upcoming crises as well as by applying the latest modelling literature and ML tools in the Armenian banking system. The decision tree cannot completely replace existing methodologies, as it cannot provide exact coefficients of impact for independent variables, which is possible with classic OLS, VAR models, but can complement with its successful implementation in early warning models and capturing non-linear relationships.

Armenian Banking Stability Risk directions In our study we address the main risk directions of the Armenian banking system. According to the Central Bank's 2022 report, high rates of average annual growth of both the mortgage loan portfolio and real estate prices were recorded, which is an indication of overheating of the real estate market. The high level of dollarization of households' credit portfolios is also risky. Starting from 2016 (except 2020), the share of loans in foreign currency has continuously increased and now exceeds the share of household loans in AMD by more than four times. Compared to 2021, the percentage weight of non-performing loans, classified loans, as well as non-performing and loan written-off in general loans has also increased. Moreover, the share of non-performing loans has especially increased among business loans. In 2022, the weight of loan losses also increased, crossing the threshold of 3.5%, which the Central Bank of Armenia explains not so much by the increase in the weight of non-performing loans, but by the banks' conservative policy of reserving risky loans on their balance sheets (recognition of possible losses). The other direction of risks, which is mentioned in the CBA Report, is related to the negative revaluation of government bonds. Reserve for revaluation of securities included in the capital in 2022 decreased by 29.8 billion drams and amounted to -34.4 billion drams, the realised loss was 1.0 billion drams. The total negative result is 2.9% of the balance sheet capital. All this was combined with the rise of the bond yield curve.[Central Bank of Armenia, 2023,11-41]

Selecting factors The banking Z-score will be used for measuring banking stability, as well as the dependent variable of the decision tree model. Z-score is efficient, as it is simple to interpret and easy to calculate. One main advantage of using it in modelling is the availability of the monthly data. ML model reliability heavily depends on the quantity of the data, which means the Z-score will help to have more accurate predictions.

Armenia, being a developing, small country, heavily relies on imports, and has to bear the shocks happening in the global economy. The shocks are related to oil price changes in exchange rates, political changes, and different sanctions. To model the effects coming from the foreign economies, variables such as dollarisation rate, USD/AMD, EUR/AMD, RUB/AMD exchange rates, as well as net export have been included in the model. As recently, crypto is increasing its share in the world monetary transactions, the price of top cryptocurrencies also is included in the variables. If in 2013, the housing loans ratio to consumer loans was 25.6 %, in 2020 they reached 30 %, while in 2023 they reached a 95 % phenomenal rate. That's why it's important to also include in the modelling the amount of housing loans. The Armenian Central Bank also tracks 7 financial stability indicators published regularly by the IMF, which cover a wide range of risk aspects, like nonperforming loans, banking assets, capital, and liquidity. An important set of variables is investment companies' financial indicators, like assets, liabilities, profit, REPO transactions. Those will indicate how changes in the structure of investments can impact stability of the banking system. Money aggregates are also included in the model, which will include the impact of the liquidity of the banks, the impact of cash level, as well as the impact of deposits based on their expiration.

Armenian inflation has been increasing constantly in recent years (see Chart 1), which has been one of the biggest influences on banking stability. That is why CPI is also included in the list of the independent variables. Government bonds profitability also is a good indicator of the overall health of the financial system, as they directly and indirectly influence banks' investment returns, liquidity management, capital adequacy, economic confidence, and regulatory compliance. Last but not least, foreign and government reserves are included in the model, showing central banks' ability to lend to other banks, the level of systematic risks, as well as capability of banks to implement different policies. On the other hand, optimal levels of foreign reserves provide currency stability and become an emergency tool to enhance liquidity.

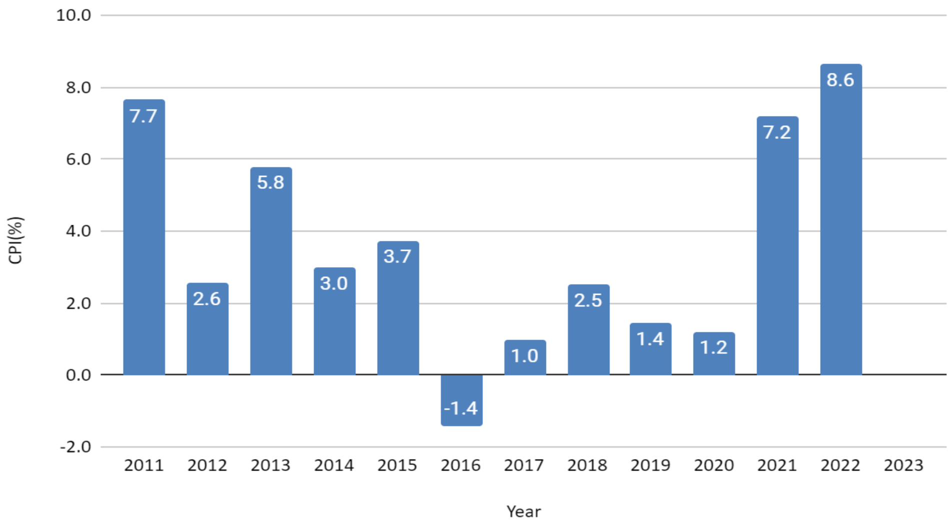

Figure 1. Inflation level in Armenia computed by CPI

Model evaluation The Model has been evaluated with the help of sklearn.tree sublibrary, using the Python programming language. The significant factors have been chosen after 3 levels of filtering:

First, lasso regression models have been implemented, to see which factors are significant. Lasso gives a 0 coefficient to the low-quality predictors. It does this by adding a penalty term to the residual sum of squares (RSS), which is then multiplied by the regularisation parameter (lambda or λ)[Kavlakoglu Eda, 2024].

The second step for filtering features is Ridge regression.

One of the most important things about ridge regression is that without wasting any information about predictions it tries to determine variables that have exactly zero effects. Ridge regression is popular because it uses regularisation for making predictions and regularisation is intended to resolve the problem of overfitting [Verma, 2024].

The 3rd step is estimating the VAR model with all the pool of factors to identify the optimal lags for the model, which is selected with AIC (Akaike information criterion). Based on that, optimal lag was selected as 2.

Using 1st and 2nd lags of the independent variables, alongside level values, the decision tree model was estimated. Decision tree models have been used to predict the dependent variable: in our case, banking Z-score, with the help of the selected variables and 2 lags of each variable.

To estimate the efficiency of the model, the data was split into train and test datasets.

The length of the training dataset is 115 data points, and 7 data points non included in the model have been used for the prediction.

All the factors have been stationarised before feeding them to the model. In the result, we get quite impressive predictions, which are represented in the Chart 2.

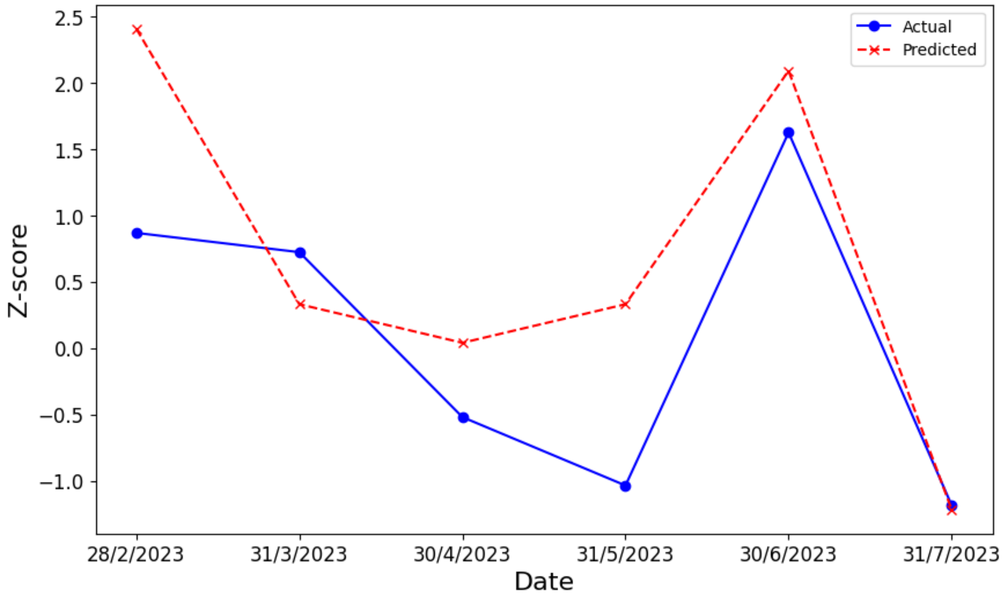

Figure 2. Comparison of predicted and test values of Z-score

The blue color line represents the actual value of the Z-score and the red line represents predictions using the decision tree for the period 02/2023-07/2023. While comparing test and forecasted data, we get a small mean squared error: a value close to 0.8.

The model is evaluated several times with different sets of variables: first including housing loans, internal and Central Bank's debt and then without those 3 factors.

The results show that the model without including housing loans, internal depth, and The Central Bank's Debt performs better.

Decision tree feature importance results are presented in the table below, which indicates that the net export: both commercial and noncommercial, inflation-related indicators, government bond profitability, investment indicators: REPO-transactions and accumulated profit of investment firms, previous lags of the Z-score, and the exchange rate of EUR/AMD, as well as the close price of cryptocurrency are the main indicators determining the banking Z-score

The top 10 factors based on their importance are presented in the Table 1:

Table 1. Feature importance score of top 10 factors affecting Z-score

Factor	Importance score
Net import (non commercial)	0.42
CPI (L1)	0.10
Z-score(L2)	0.06
Repo transactions of individuals(L2)	0.06
CPI Service sector (L2)	0.06
Zscore (L1)	0.05
EUR/AMD(L1)	0.04
The accumulated profit of the investment firms (L2)	0.03
Close crypto price	0.02
Net import (non commercial,L2)	0.02

Net import (non commercial)

Each value in importance score shows the relative prediction power of each measure in forecasting. The L1 and L2 correspond to the first and second lags of each variable.

It can be seen that net non-commercial import is the key factor determining the Z-score predicted value, and compared to other variables it is at least 4 times more important. This means that the gap between export and import is a key determinant for the banking crisis.

First lag of the inflation [CPI (L1)] expressed by CPI is the second most important feature. The rest of the factors have lower predictive power, but still determine the value of Z-score and lower the standard error of the model. We can see that investment indicators:

Repo transactions of individuals and accumulated profit of investment firms influence Z-score value after 2 periods, meaning that it takes longer for investments to impact the economy in general.

In the model were included exchange rates of EUR/AMD, USD/AMD, RUB/AMD, but only the exchange rate of Euro is the list of the top 10 important factors.

Last but not least, cryptocurrency prices also influence banking stability, which proves that increasing volume of the crypto market can determine the mood in the financial system of the country, and fluctuations and speculations in top coin prices can cause financial crises.

Conclusions

The article describes a complete methodology to model the banking stability in the Republic of Armenia. Banking Stability is measured with the help of the banking Z-score, which is calculated in monthly granularity based on country level financial data in Armenian. Then, using Z-score as a dependent variable, a decision tree regression model is implemented to predict the value of Z-score based on the set of selected variables. After model evaluation on the training data, it was tested on the test data and showed a low amount of mean squared error, proving the relevance of the selected factors and model efficiency.

In addition to good predictions, the decision tree model also provides feature importance scores, based on which are identified main factors determining changes in the Z-score for Armenia, which displays which variables have more impact on the stability or instability of the banking system. The key factors determining Armenian banking stability are as follows:

Net export, CPI first lag, previous month's value of the Z-score, second lag of REPO transactions, second lag of FDR reserves, the exchange rate of EUR/AMD, second lag of the accumulated profit of investment firms, the average profitability of government bonds and total capital. The rest of the factors have relatively low impact and importance in the modelling.

This modelling approach will be a complement to the existing early warning model approaches for the Central Bank of Armenia and will help in the detection of the weak spots in the banking system. First of all, the model will help to identify banking system collapse risks earlier, and then will indicate which directions to look for finding out key factors causing instability.

Decision tree models, on the other hand, will benefit with their ability to identify non-linear relationships and detect interrelations among different economic factors and banking stability.

References

1. Alessi Lucia, and Carsten Detken, "Identifying Excessive Credit Growth And Leverage." *European Central Bank*, 2014.
2. Altman Edward, "Predicting Financial Distress Of Companies: Revisiting The Z-Score And Zeta." Handbook of Research Methods and Applications in Empirical Finance, 2000.
3. Central Bank of the Republic of Armenia, Financial Stability Report 2022, 2023
4. Gabriele Carmine, "Learning from trees: A mixed approach to building early warning systems for systemic banking crises." *European Stability Mechanism*, 2019.
5. Kavlakoglu Eda, "Apply lasso regression to automate feature selection." *IBM* developer, 2024.
6. Steve Mercieca, Klaus Schaeck, Simon Wolfe, "Small European banks: Benefits from diversification?", 2007, Journal of Banking & Finance
7. Verma Yugesh, "A hands-on guide to ridge regression for feature selection." Analytics India Magazine, no. 2024.
8. Viswa, "Unveiling Decision Tree Regression: Exploring its Principles, Implementation." Medium, 2023.

Davit MINASYAN

Banking stability modelling for Armenia: Decision tree approach

Key words: banking stability, decision tree, modelling, Z-score

In the following article is presented methodology for modelling the banking stability of the Republic of Armenia, using decision tree models. As a measure of banking stability is suggested banking Z-score, which indicates how the banking system is close to default. For evaluating the model, 47 factors are included from real, financial and foreign sectors of the economy. For considering the lagged impact of factors on the dependent variable, apart from base values the first 2 lags of independent variables are included in the model. The proper lag is decided by estimating a VAR model. In the result, the model's feature importance scores show, that Armenian banking stability is mainly impacted by net import, inflation rate, previous lags of the dependent variable: which is state of the banking stability in previous period, investment indicators of individuals, exchange rate of Euro/AMD, as well as cryptocurrency price fluctuations. It is justified that the chosen methodology and variables are efficient for predicting the banking stability of Armenia, by giving small standard errors and accurate predictions. Also, including the previous lags of the variables improves the model predictive power. Apart from that, it is proved that the Z-score can be used as the measure of banking stability, and due to the availability of monthly data, act as a dependent variable to model the banking system's resilience to the shocks and key factors responsible for the stability.