Forecasting the migration of deposit volume towards term deposits for Nationale-Nederlanden Bank

by

Pieter Voogt

Master Thesis

Faculty of Behavioral, Management, and Social Sciences

Financial Engineering and Management

University supervisors

Prof. Dr. L. Spierdijk

Dr. B. Roorda

Company supervisors

R. Hagen (Zanders)

T. van Rijn (Nationale-Nederlanden Bank)

May 13, 2024

Preface

Dear reader,

Submitting this thesis marks the end of my studies at the University of Twente and the start of a new chapter in my career, where I am happy to say that I will be starting as a consultant at Zanders. I would like to thank all the employees at Zanders who gave their input and support for this research, but I would especially like to thank Richard Hagen for the helpful weekly meetings, support and feedback. I also want to express my gratitude to Tina van Rijn for supervising me and providing the opportunity to conduct my research at Nationale-Nederlanden Bank.

I want to thank Laura Spierdijk for her guidance and helpful feedback, which have been crucial in bringing this thesis up to academic standards. Her broad knowledge of the financial sector provided me with new insights. I would also like to thank Berend Roorda for serving as my second supervisor and for his useful feedback. Their combined expertise and support were greatly appreciated in bringing this thesis to a successful conclusion.

Finally, I want to thank my friends and family for all the support they have given me over the past few months. A special thanks to Floor Martens for being there through all the milestones and challenges of my studies. Your involvement and encouragement have been really helpful.

I hope you enjoy reading this thesis!

Kind regards,

Pieter Voogt
Utrecht, May 13, 2024

Abstract

In this research, we develop a model framework that can predict the migration of deposit volume from savings accounts towards term deposits. The research makes use of a dataset provided by Nationale-Nederlanden Bank (NN Bank), which consists of monthly customer account data. A model framework was developed that consisted of different interest rate and deposit volume models. The models were trained and tested on the dataset of the NN Bank and historical interest rates. During this process, we encountered issues with the non-stationarity of our time series due to the sudden increase in interest rates in 2023. In the research, we divided the different maturities of term deposits into three distinct maturity groups: the 1-year, 2-year, and 3+ year. We researched whether the spread between the savings rate and the term deposit rate, or the level of the term deposit rate itself, gave customers more incentive to open a term deposit. The model developed to predict the migration was unable to make accurate predictions on our test set due to the limited observations of term deposits opened in our dataset before 2023. Due to the data issues, it is not possible to draw a reliable conclusion from this research. Despite this, our findings indicate that a reduction in the spread between the two rates is associated with a decline in the number of term deposits opened. When the term deposit rate is used as an input for our model, we observe a more consistent outflow.

Contents

List of Tables

List of Figures

List of Abbreviations

ALM Asset Liability Management

DNB De Nederlandsche Bank

DNS Dynamic Nelson-Siegel

EBA European Banking Authority

ECB European Central Bank

Euribor European Interbank Offer Rate

IRS Interest Rate Swap

MAE Mean Absolute Error

NMD Non-Maturing Deposit

NN Bank Nationale-Nederlanden Bank

OLS Ordinary Least Squares

RMSE Root Mean Squared Error

VAR Vector Autoregressive

1 Introduction

1.1 Problem context

In the area of financial institutions and banking, the management of deposit volumes plays an important role in ensuring stability and profitability. One area of particular concern is the migration of deposit volume from savings accounts to term deposits within a bank. Term deposits, also known as time deposits or fixed-term deposits, are financial products offered by banks and financial institutions that allow customers to invest a specific sum of money for a predetermined period, typically ranging from a few months to several years. During this period, customers cannot withdraw their money or can only do so with a penalty, but they receive higher interest rates compared to regular savings accounts. Savings accounts, however, have the characteristic that money can be withdrawn at any time and are often referred to as a Non-Maturing Deposit (NMD).

In the last year, the European Central Bank (ECB) has been drastically increasing the interest rate to reduce the inflation rate in Europe (ECB, 2023). There are several rates set by the ECB, the main refinancing rate, the deposit facility rate, and the marginal lending facility rate. These are the rates at which commercial banks can borrow, deposit, and borrow money overnight from the ECB respectively. The ECB changing its rates often triggers corresponding shifts in the market rates set by banks. Market rates, as discussed by banks and financial institutions, typically refer to the rates at which they can lend to or borrow from other banks or financial entities. Examples of such rates are the European Interbank Offer Rate (Euribor) and Interest Rate Swap (IRS), which will be discussed later on in this research. As the ECB raises interest rates, the market rate changes as well, typically increasing in response to the central bank's monetary policy. The banks adjust their deposit rates accordingly.

The increasing interest rate environment of 2023 causes customers' saving behavior to change within the banking sector. One notable trend is the increased popularity of term deposit contracts. In the Netherlands, term deposits were far less popular than regular savings accounts for years. These deposits are preferred over regular deposits by both individual and institutional investors because of their reliability and higher fixed interest rates. In banking, customers and banks have different benefits from savings and term deposit accounts. With a savings account, customers can access their money anytime and add more whenever they want. Banks like this because they can change the interest rate they pay on these accounts every day, depending on the market. Term deposits, on the other hand, give customers a higher interest rate, so they can earn more money over time. Banks benefit because they know exactly how much money they will have and what they will pay in interest, making it easier for them to plan their finances.

As market rates climb, customers are drawn to higher yields offered by these fixed-term contracts that are being given out by banks, and therefore we see a rise in these contracts being concluded as of the beginning of 2023. Banks advertise their deposit rates to get customers' attention by offering competitive rates, which is a marketing tactic to attract customers in the financial market. This method affects their position and determines their market share relative to their competitors. While banks can invest customers' term deposits relatively safely, they have no recent data or experience on customer behavior when they migrate to other savings products within the bank. Savings volume will migrate out of a bank if customers need money or if other banks provide better yields.

1.2 Company description

This master thesis will be written at Zanders, established in 1994. The company specializes in treasury management, risk management, and corporate finance. From its eight offices, a team of over 450 employees offers global services to corporations, financial institutions, and public sector entities. Conducted within the risk advisory department of Zanders, I will be tackling a problem of their client, Nationale-Nederlanden Bank (NN Bank). NN Bank is part of NN Group, an international financial services provider active in 11 countries and leading in several European countries and Japan. Their roots stretch back over 175 years. NN Bank offers its customers in the Netherlands financial services in the areas of savings, bank savings, investing, mortgages, and lending. This research will follow up on a previous study done at NN Bank by Van der Ploeg (2023), which investigated the average maturity of NMD at NN Bank. This thesis will cover one of the subjects that was addressed for further research.

1.3 Research motivation

Understanding the dynamics of deposit migration is important for financial institutions. Research has been extensively conducted on the outflow of savings. However, given the current circumstances, migration is becoming increasingly important and remains a relatively unexplored topic. It directly impacts a bank's asset-liability management, financial risk management, and profitability. Incorrectly modeling the deposit migration could cause increased exposure to liquidity and interest rate risk (Hull, 2018). Liquidity risk refers to the potential that an institution may not have sufficient liquid assets on hand to meet its short-term financial obligations when they come due. In the specific context of deposit migration, liquidity risk management plays a critical role in matching inflows and outflows with available funds. To manage interest rate risk, it is crucial to match the maturities of assets and liabilities, ensuring that interest rate adjustments are aligned. This strategy helps mitigate the impact of market rate fluctuations on a bank's income.

Banks undertake a critical responsibility when individuals trust them with their money through term deposits. To effectively manage these funds, banks invest them for a specified period, where the duration is determined by forecasting models. These models help anticipate how much of the deposit volume will remain within the bank, either in the same or different accounts, and how much might migrate out. The accuracy of these predictions is important because any miscalculation can lead to liquidity problems. Overestimating funds staying within the bank may lead to missed profit opportunities, while underestimation can result in insufficient reserves to meet withdrawals. Finding the right balance between maximizing returns and maintaining liquidity is crucial for a bank's financial stability and success.

This research is relevant due to the uncertainty surrounding interest rates determined by the ECB. Banks rely on various interest rate models to forecast these rates but face the challenge of correctly setting deposit interest rates to retain customers. A misjudgment in deposit rates exposes banks to interest rate risk within their banking books (BCBS, 2016). If rates are set too high, it can decrease profitability, while rates that are too low may result in customers leaving for competitors. This emphasizes the need for precise research to strike the right balance between customer retention and risk management in an uncertain interest rate environment.

Another consequence of the growing popularity of term deposits is deposit cannibalization. This term describes the transfer of funds from the bank's older, lower-yielding products to their new, higher-yielding ones. This potentially creates a negative outcome for the bank, as the cost of funds increases without any new money being brought in. Carefully selecting the customer rate to attract new clients while minimizing the impact of deposit cannibalization is essential.

Understanding the connection between deposit volume, deposit rates, and market rates is crucial in finance. Deposit volume is how much money people save in banks, and the deposit rate is the interest rate the bank pays on those savings. The market rate affects the deposit rate and is influenced by the overall economy and central bank decisions. The aim is to understand the factors, triggers, and motivations behind customers transitioning from one savings product to another.

When the market rate changes, banks adjust their deposit rates accordingly. This affects people's desire to save with banks. If another bank or product in the same bank offers a better rate, people shift their money there. One of the areas that this research focuses on is looking at the history of these rates and deposit volumes. By studying past data, we aim to understand how these factors have affected each other over time. This can help us forecast the market rate, deposit rate, and deposit volume by understanding the interrelationship between the three in the financial sector. Figure 1.1 visualizes the relationship between the different variables.

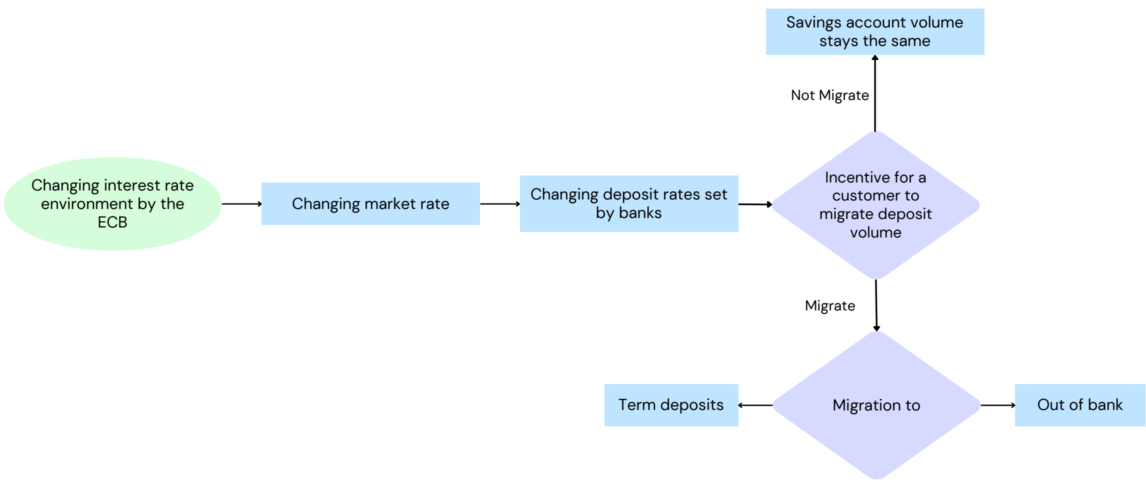

Figure 1.1: Visualization of the migration effect

Another relationship that is of importance, is the concept of supply and demand in the banking sector.

It covers the important topics of interest rate sensitivity and customer savings behavior. Due to the increasing interest rate environment, banks experience a change in the supply of loans and demand for deposits. The topic of supply and demand of deposits and its relation to interest rate changes is extensively researched in the existing literature and goes a long way back by articles of Friedman (1966) or Laidler (1966). A famous British economist Keynes (1936) proposed the so-called Keynesian Theory, which is one of the principles of supply and demand in modern-day economy. Banks are working hard to attract customers by offering good interest rates on loans and deposits, but they also have to be careful about how this affects their capital and profits (Ben-David et al., 2017). This situation creates a lot of competition among banks. Each one wants to be the best choice for customers looking for better interest returns, but they also need to keep their business healthy. This research also looks into how NN Bank positions itself in the Dutch banking sector and how deposit rates influence the demand for deposits.

Previous research on forecasting the migration of deposit volume is limited to volume models and outflow models of a savings portfolio. Well-known papers about deposit volume modeling of, for example, Jarrow and Van Deventer (1998), Nyström (2008), Kalkbrener and Willing (2004) cover various market rate, deposit rate, and volume models. The inclusion of the migration effect in these models has not been done extensively, hence the relevance for the academic literature. In Chapter 2, we will conduct a literature review to find out how the volumes of savings products are predicted.

1.4 Research design

The goal of this research is to forecast the migration of deposit volumes within the savings products of NN Bank. To do this, we want to clarify which variables influence the migration effect and how. This led to the following main research question:

"How can the migration of deposit volumes for Nationale-Nederlanden Bank's deposit products be accurately forecasted?"

To answer our research question, we need to explore different methods for predicting how much money is kept in bank accounts, which is done through volume models. These models help us predict the future amounts of money in bank accounts by looking at past trends and changes. A key part of our study is to see how customer behavior changes, especially when banks offer different interest rates on their savings accounts. In addition to volume models, we also use outflow models based on these volume predictions. These outflow models are important because they give us specific insights into how much of the money leaving savings accounts goes into term deposits instead. This helps us understand not just how much money is moving, but what portion of the total money leaving savings accounts is being transferred to term deposits. By combining volume and outflow models, we get a clearer picture of the movement of money in and out of bank accounts. This research aims to address the following sub-questions to acquire the necessary knowledge:

1. How are non-maturing deposits modeled?
   (a) Which market rate models are there in literature?
   (b) Which deposit rate models are there in literature?

(c) Which deposit volume models are there in literature?

By answering these sub-questions, we obtain knowledge on different market rate models, deposit rate models, and deposit volume models. Understanding deposit rate and market rate models is essential for modeling migration of deposit volume because it helps predict how changes in rates influence customer behavior and deposit flows. Furthermore, by looking at NMDs, we will find several aspects that are of interest when modeling NMDs. These aspects include behavioral assumptions, rate elasticity, and product characteristics. Previous research at NN Bank by Van der Ploeg (2023), estimating the maturity of NMDs, will be discussed in Chapter 2.

2. Which ways of capturing product migration are used in literature?

Recently, the migration effect has become a new trend in savings deposits, something we have not seen in the last ten years. In our research, we will explore the literature to identify methods for understanding and capturing this migration effect. Furthermore, we need to research which ways are suitable for the savings portfolio of NN Bank.

3. How do the characteristics and trends in the collected data enhance our understanding of deposit volume migration forecasting?
   (a) What are the long-term and short-term trends observed in the market rates, deposit rates, and volumes over the study period?

NN Bank will provide data on their clients with their corresponding accounts and deposit volume over time. We will introduce the datasets, describe the data cleaning and preparation processes, and present initial findings or patterns observed in the data. Furthermore, we will collect data on historic market rates and deposit rates.

4. How can we create a framework that incorporates the migration of deposit volume in volume models?
   (a) How can we incorporate the interest rate incentive of migration due to the difference in deposit rates between various savings products?

Once we answered sub-question four and obtained knowledge about the various ways of capturing the migration effect, we need to research how to include this into the volume model of the saving products of NN Bank.

5. How can we validate and accurately interpret the outcomes of our proposed models that include the migration effect?
   (a) What methodologies are most effective for assessing the impact of selected variables in our model?
   (b) What are the factors, triggers, and motivations behind customers transitioning from one savings product to another?
   (c) How do different interest rate scenarios impact the results of our model?

To conclude, we want to validate whether we included the deposit migration correctly to forecast the migrating deposit volumes within NN Bank around two years in advance. If we succeed in this, we help NN Bank with its Asset Liability Management (ALM) and financial risk management. Accurately predicting how much money customers will deposit and for how long is crucial for effective ALM. If NN Bank can anticipate future deposit volumes, it can better plan and balance its assets and liabilities. For instance, understanding deposit trends helps the bank decide how much money it can safely lend without risking a liquidity shortfall. Additionally, exploring the impact of different interest rate scenarios on migration is valuable. Predicting future interest rates with certainty is challenging. Investigating the consequences of NN Bank's competitive position in the market is also important. Understanding the implications of fluctuating market rates and the effects of NN Bank's competitiveness can provide valuable insights.

1.5 Problem approach

1.5.1 Methodology

To solve these research questions, we need to perform two types of research: qualitative and quantitative research.

Qualitative research

The qualitative research will mostly consist of finding a theoretical framework in the existing literature. Although the modeling of the migration effect has not been extensively reported in the literature, there is much to be found on deposit rate models, market rate models, and deposit volume models.

Quantitative research

The first step in quantitative research will be data processing and visualization. The data will mostly consist of time series of deposit volumes, deposit rates, and market rates. Once this is done, we can start looking at how best to use the data to build our models. NN Bank uses SAP software for their databases and the data will be extracted from there.

1.6 Scope, limitations and assumptions

This research focuses on investigating customer migration patterns between different saving products within NN Bank. The aim is to understand the factors, triggers, and motivations behind customers transitioning from one savings product to another. The research intends to provide insights into customer behavior and preferences, which can inform strategic decisions related to product offerings and customer retention strategies.

Many macroeconomic factors influence customers' saving behavior, such as inflation, house prices, or income growth (Fund, 1990). As mentioned in Section 1.3, we are researching deposit rates, market rates, and deposit volumes for specific reasons. First, by focusing on these variables, we can examine them in-depth to understand their unique coherence and impact. Second, we selected these variables for their reliable historical data, which is important for accurate analysis. Also, we limited our scope to make the results of the research more straightforward, avoiding confusion that could happen if too many different variables were included.

It is important to acknowledge that this focused approach has limitations. By not including other macroeconomic variables, such as inflation rates, economic growth, and demographic changes, the study may not capture the full picture of factors influencing deposit behavior. However, we understand that these limitations are trade-offs to achieve a more detailed analysis of the variables that are included. Another limitation is excluding customer-specific data such as age or gender. Choosing not to make assumptions about a customer's financial situation is a sound decision, as it can make the research more robust. However, this approach also acknowledges that customers likely pay more attention to their financial circumstances than to market conditions.

The recent trend of customers migrating from traditional savings products to term deposits is a relatively new development. This shift has resulted in limited historical data being available. Term deposits were available from the beginning of NN Bank's operations, but not to the extent that they are now. In addition, the data provided by NN Bank from its early years has some data quality issues, which we try to address in Chapter 3. The research operates on the premise that the forthcoming market conditions will be marked by a degree of uncertainty and volatility, which may differ from historical trends and patterns. This assumption guides the analysis and modeling efforts toward preparing for potential fluctuations in customer behavior.

1.7 Thesis outline

This research is organized into several chapters. Chapter 2 outlines our literature review, which aims to answer sub-questions 1 and 2. In Chapter 3, we will describe the data that is available and how it will be used and hereby answer sub-question 3. Chapter 4 provides a detailed description of our model framework and also includes the answer to sub-question 4. Chapter 5 will cover all the modeling and their results and hereby answer sub-question 5. Finally, Chapter 6 includes a conclusion and discussion of our research findings, limitations, and recommendations for future research.

2 Literature review

This chapter aims to answer sub-question 1 and 2. By answering these questions, we will know how NMDs are modeled and how we can account for product migration. Furthermore, it should become clear how the market rate and deposit rate are related and what their influence is on deposit volume models as visualized in Figure 1.1. During our study, a traditional literature review was conducted to create a solid foundation of the existing academic and professional knowledge regarding modeling NMDs and deposit volume migrations between financial products. The approach taken was comprehensive, allowing for a broad exploration of the topic. The review began with identifying key terms and concepts related to deposit volume migration, such as "market rate model," "deposit rate model," "financial product migration," "deposit outflows AND banks," "deposit volume model AND banks," and "fixed-term deposit migration." These terms guided the search through several databases, including JSTOR, EBSCO, and Google Scholar, ensuring a detailed scan of available literature. This review provided knowledge that we can use to make better models that predict deposit movement more accurately.

As previously mentioned, the thesis of Van der Ploeg (2023) looked at the average maturity of NMDs of NN Bank. The thesis and its results are interesting for our research because of the overlap in the three models needed to model NMDs (market rate, deposit rate, and deposit volume) at the same bank. However, the literature review of our research aims to find a theoretical framework for a better understanding of the existing models and how to model the migration effect explained in Figure 1.1. The results and insights found by Van der Ploeg's used models will be mentioned in Sections 2.1.1, 2.1.2, and 2.1.3.

2.1 Modeling of non-maturing deposits

NMD models are primarily designed to deal with deposits that do not have a contractual end date. The core of these models is to predict behaviors of such deposits, especially regarding their duration and stability. While NMD models are primarily designed for deposits without a fixed maturity, their core principles are relevant in understanding the behavioral dynamics of deposit migrations. When developing models for NMDs, several aspects are critical, including behavioral assumptions, economic conditions, and the competitive environment. Furthermore, the sensitivity of deposits to deposit rate changes is a crucial aspect of NMD models (Nyström, 2008). The following subsections examine the existing literature to determine the extent of this sensitivity, how the market rate influences the deposit rate, and how changes in the interest rate affect the supply and demand for deposits.

2.1.1 Modeling market rates

Market rates have a big influence on deposit rates. The different models for determining the deposit rate all include in some way the market rate (Jarrow & Van Deventer, 1998). When banks and financial institutions talk about the "market rate," they are often referring to the rate at which they can lend to or borrow from other banks or financial institutions. An example of such a rate is the Euribor, which is a rate comprising the average rates of a panel of major European banks that lend to each other in euros (Hayes, 2022). The short rate, often referred to as the short-term interest rate, is the rate of interest on short-term financial instruments. The short rate is influenced by central banks such as the ECB to control the interaction of supply and demand for funds in the money market. To increase or decrease inflation in the economy, the ECB sets out the rate at which banks can borrow money from them. A higher rate given out by the ECB will force banks to increase their deposit rate to attract cheaper funds than at the ECB. However, to cover for the expenses of paying out more interest on deposits, banks need to increase the deposit rate on their loans. The demand for loans gets lower and the supply of deposits increases, slowing the economy down and decreasing the inflation rate. The market rate, on the other hand, encompasses a broader range of interest rates available in the financial markets. This includes rates for various maturities and credit qualities, such as the Euribor. In the existing literature, future market rates are often modeled by short rate models.

Lundgren (2022) covers multiple models, such as the Vasicek model (Vasicek, 1977) and the Nelson-Siegel model (Nelson & Siegel, 1987). Furthermore, Nyström (2008) follows the model for the market rate of Jarrow and Van Deventer (1998) and the no-arbitrage approach of Heath et al. (1992). However, they argue that any other market rate model could be used as input for their model. We elaborate on several models below.

The one-factor Vasicek model is widely used in finance to model the evolution of the short rate using a stochastic differential equation. The model is commonly used in economics to model future short rates. It assumes that the short rate is mean-reverting towards a long-term mean level at a certain speed (Vasicek, 1977). The model uses the following equation:

$$r_t = r_{t-1} + a(b - r_{t-1})\Delta t + \sigma\sqrt{\Delta t}\epsilon_{t-1} \tag{2.1}$$

Where r_t is the short rate at time t, a is the speed of mean reversion, b is the long-term mean of the short rate, σ is the volatility of the short rate, and ϵ_t is the standard normal random variable which accounts for the randomness in the model. While the Vasicek model is elegant and has the advantage of mean reversion and its allowance for negative short rates, it has limitations because of its simplistic dynamics. As a result, numerous extensions and alternatives, like the Cox-Intersoll-Ross (CIR) model (Cox et al., 2005) or two-factor Vasicek model by Kalkbrener and Willing (2004), have been developed to address these issues.

The CIR model includes $\sqrt{r_t}$ in the volatility term to account for the non-negativity, which was initially seen as an advantage but lost relevance in the context of negative interest rate policies adopted by some central banks. The model was preferred in practical applications, especially in risk management and derivative pricing, where negative interest rates used to be unrealistic. This is because negative interest rates can pose challenges for financial institutions, such as affecting profitability and the functioning of money markets.

$$r_t = r_{t-1} + a(b - r_{t-1})\Delta t + \sigma\sqrt{r_{t-1}\Delta t}\epsilon_{t-1} \tag{2.2}$$

The two-factor Vasicek model incorporates two sources of stochastic variability. The intention behind extending the model to two factors is typically to better capture the observed dynamics and complexities in interest rate movements. The second factor could represent a long-term interest rate factor or a macroeconomic factor like inflation.

$$r_t = r_{t-1} + a_1(b_1 - r_{t-1})\Delta t + \sigma_1\sqrt{\Delta t}\epsilon_{1t-1}$$

$$y_t = y_{t-1} + a_2(b_2 - r_{t-1})\Delta t + \sigma_2\sqrt{\Delta t}\epsilon_{2t-1} \tag{2.3}$$

The short rate is then modelled as a combination of these factors which can be seen as the level and the slope of the term structure. By using two factors, the model can more accurately reflect the reality of how interest rates change over time, accounting for different market conditions that impact the yield curve. The error terms ϵ_{1t} and ϵ_{2t} are drawn from a bivariate normal distribution with a specified correlation coefficient ρ. This correlation coefficient measures the linear relationship between the two factors' movements and is typically between -1 and 1. The correlation is important for modeling the dynamics of the interest rates, as it allows the model to capture how different factors may influence each other.

Lundgren (2022) uses the Nelson-Siegel model (Nelson & Siegel, 1987) to simulate the market rate. The Nelson-Siegel model is a parametric exponential three-factor model. The model calculates the shape of the yield curve. The curve reflects the future short-term market rates of different maturities, if the market expects market rates to rise, the yield curve will typically be upward-sloping and vice versa. Furthermore, this model is capable of fitting a variety of shapes of yield curves, making it flexible for different market conditions which is relevant in the current interest rate environment. Both the one-factor Vasicek and CIR are single-factor models that may not capture the full range of yield curve shapes as effectively as the Nelson-Siegel model. The yield of maturity τ is calculated as follows:

$$y_\tau = \beta_1 + \beta_2(\frac{1-e^{-\lambda\tau}}{\lambda\tau}) + \beta_3(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}) \tag{2.4}$$

where y_τ is the yield of a given maturity τ and λ is a decay parameter that determines the rate at which the influence of the slope and curvature factors declines as maturity increases. This model can capture various yield curve shapes (e.g. upward-sloping, downward-sloping, humped) using the three β parameters. β_1, represents the long run levels of market rates, β_2 the short-term component and β_3 the medium-term component. However, Diebold and Li (2006) argued that the original model could be extended by a more dynamic framework. The extended model allows for forecasting of market rates by introducing a time-varying component:

$$y_{t,\tau} = \beta_{1t} + \beta_{2t}(\frac{1-e^{-\lambda\tau}}{\lambda\tau}) + \beta_{3t}(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}) \tag{2.5}$$

This allows us to analyze how the yield curve changes from one period to the next, which is useful for understanding the evolution of interest rates and the term structure through time. In the Dynamic Nelson-Siegel (DNS) model, we estimate the yield curve at multiple points in time. The regression model in equation (2.6) aims to estimate the latent factors β such that the factor loading matrix B times β approximates the observed yields y. The error term ϵ_t captures the discrepancy between the observed and estimated yields. The goal of the regression is to minimize the error term ϵ.

$$y_t = B(\tau_i)\beta_t + \epsilon_t \tag{2.6}$$

The factor loading matrix B in the DNS model is a matrix of functions that relate the latent factors (level, slope, and curvature) to the yields of bonds with different maturities. This matrix is defined based on the decay factor λ and the maturities τ. Each row corresponds to a different maturity and contains the loadings for the level, slope, and curvature factors.

$$B(\tau_i) = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} \\ \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \end{bmatrix}$$

Diebold and Li (2006) found that keeping the decay factor λ constant at 0.0609 gave the best results when using this model. The yield matrix y contains actual observed market data. This matrix is constructed from our dataset and consists of the yields for different maturities at different points in time.

$$y = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,N-1} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{M,1} & y_{M,2} & \cdots & y_{M,N-1} \end{bmatrix}$$

The DNS model is often used for predicting future yields and yield curves. The time-varying β parameters are modeled as autoregressive processes, enabling the model to capture the dynamic evolution of the yield curve. As a result, it can provide more accurate yield predictions. The study by Van der Ploeg (2023) used the DNS model and found a good fit for the interest rate data used. The effectiveness of the model in fitting the interest rate data as found by the study suggests that it can accurately capture the complex behaviors of interest rates across different maturities. This good fit indicates not only the model's capability to reflect the current interest rate environment but also its potential utility in forecasting future interest rate trends.

The goal of this part of the literature review was to provide an answer on how market rates are simulated for the use of NMDs modeling. The models discussed above provide a solid basis for modeling the market rate for several years, but with the occurrence of negative interest rates in the past year, the model characteristics of no negative interest rates no longer hold. Therefore, the CIR model is excluded. In Chapter 4, we will conclude which model for the market rate will be used in this research and why. Now we know how we can model the market rate, we can move on to how the market rate is used as a variable in deposit rate models.

2.1.2 Modeling deposit rates

Modeling the deposit rates that a bank should offer on its savings products is very important. It plays a vital role in the management of a bank's balance sheet, profitability, and overall risk profile. By offering competitive deposit rates, banks can attract or retain deposits, ensuring adequate liquidity to meet operational and regulatory requirements. When a bank offers competitive deposit rates, it increases the demand for its deposit products. If the bank lowers the rates, the demand for its deposits might decrease as customers look for better returns (Hubbard & O'Brien, 2018). Furthermore, banks lend out deposits at longer terms and different rates than they pay depositors. By accurately modeling deposit rates, banks can better manage the interest rate mismatch between assets and liabilities.

In the case of our research, changes in deposit rates could cause customers to migrate to term deposits or withdraw their money and shift to another bank. The deposit rates on a savings account could be changed daily, whereas, on a term deposit, the deposit rate is fixed for an agreed maturity (NN Bank, 2023). Therefore, offering the correct deposit rate for a term deposit is very important for a bank's risk management and profitability. Determining the deposit rate on a savings product is essentially the price the bank is willing to pay to attract and hold onto customers' money.

There are various ways described in literature on how to model deposit rates. Previous research has proven that deposit rates follow market rates (Jarrow & Van Deventer, 1998). The market rate often indicates the general direction of deposit rates in the economy. Furthermore, it is a key determinant of a bank's cost of funds. When market rates are high, the cost for a bank to obtain their funds increases. As a result, banks may raise deposit rates to attract more funds at a cheaper price.

There are several types of deposit rate models in the existing literature that this literature review will examine. Our analysis begins with a linear model that establishes a direct relationship with the market rate without an intercept, as described by Nyström (2008).

$$d_t = \beta_1 r_t + \epsilon_t \tag{2.7}$$

In this equation, d_t represents the deposit rate at time t of a bank, r_t represents the market rate at time t and ϵ_t represents the error term at time t that captures the portion of the deposit rate that is not explained by the linear relationship with the market rate. The term follows a normal distribution. Because this is the simplest form of deposit rate modeling, the formula is not able to capture complex relationships between the deposit rate and the market rate. The reaction of deposit rates to market rates might be more pronounced in a high-interest-rate environment than in a low-interest-rate environment. Elkenbracht and Nauta (2006) extend equation (2.7) by adding the constant β_0:

$$d_t = \beta_0 + \beta_1 r_t + \epsilon_t \tag{2.8}$$

The introduction of β_0 can be seen as acknowledging a base level for the deposit rate, independent of the deposit rate. This base level could represent the minimum return depositors expect or a basic rate that is offered by financial institutions regardless of the market rate.

Jarrow and Van Deventer (1998) present a discrete-time model for predicting future deposit rates by using current information and ensuring that the predicted rates are consistent with the no-arbitrage principle as given in equation (2.9). The no-arbitrage principle states that it is impossible to make a profit without taking any risk by exploiting price differences on the market. When it comes to deposit rates, this means that the rates predicted by the model should eliminate arbitrage opportunities when compared to other investments with similar risks. By using the present market rate r_t and the historical market rate pattern in the model, the prediction of deposit rates is based on the market actualities and past observations.

$$d_t = d_0 + \beta_0 t + \beta_1 \sum_{j=0}^{t-1} r_{t-j} + \beta_2 (r_t - r_0) + \epsilon_t \tag{2.9}$$

The parameters β_0, β_1, and β_2 need to be estimated by using a method such as Ordinary Least Squares

(OLS). By incorporating a summation of past market rates and the difference between the current and initial rates, the model provides a complete overview of the rate's progression over time. The initial rates, represented by $t = 0$, refer to the values of the variables at the start of the period being analyzed or modeled. The base level deposit rate d_0 acts as a starting point upon which the effects of market rates are built. t captures general tendencies in deposit rates over time, separate from market rate fluctuations. The historical summation by $\sum_{j=0}^{t-1} r_{t-j}$ provides a memory of where rates have been, using the fact that past market rate conditions can have an impact on the current state of deposit rates (O'Brien et al., 1994). However, this feature could lead to overestimated deposit rates when used for long-term forecasting. $(r_t - r_0)$ represents the difference between the current market rate and the initial market rate. It captures how changes in the market rate relative to its initial level affect the deposit rate, indicating the model's responsiveness to recent market movements. The current rate changes incorporate the recent market conditions, allowing the model to be dynamic and responsive to new information. However, the effectiveness of the model depends on the accurate estimation of the β parameters. The error term ϵ_t captures the random fluctuations and all other unaccounted-for influences on the deposit rate that are not explained by the model's included variables.

O'Brien (2000) states that deposit rates exhibit stickiness and that rate adjustments tend to be asymmetric. This implies that the spreads of the deposit rates differ when market rates increase or decrease. Banks may be slow to lower deposit rates when market rates fall to maintain good customer relationships and loyalty. Furthermore, when market rates rise, banks may quickly increase deposit rates to avoid losing customers to competitors offering higher rates. The following formulas are proposed:

$$\Delta d_t = (\lambda^+ I_t + \lambda^-(1 - I_t))(d_t^e - d_{t-1}) + \epsilon_t \tag{2.10}$$

$$d_t^e = br_t - g$$

$$I_t \equiv I_{[d_t^e - d_{t-1} > 0]}$$

where Δd_t is the one-period change in the deposit rate and r_t is the market rate. d_t^e is a conditional equilibrium rate in that the actual deposit rate adjusts to d_t^e which gets corrected by the one-period lagged deposit rate d_{t-1}. It is determined by the current market rate r_t multiplied by a factor b, and adjusted by a constant g. The factor b reflects the sensitivity of the deposit rate to the market rate. The constant g is determined by empirical analysis by fitting the model to historical data, and g is estimated alongside other parameters like b, λ^+, and λ^-. The value of g adjusts the baseline level of deposit rates, accounting for factors that influence deposit rates but are not directly related to current market rates. Furthermore, I_t is an indicator function that equals 1 if the expected deposit rate d_t^e exceeds the previous period's deposit rate d_{t-1} (indicating an increase), and 0 otherwise. ϵ_t represents the zero-mean process. The asymmetric part of this function is modeled by the parameters λ^+ and λ^- as they represent the upward and downward adjustment, respectively. If $\lambda^+ < \lambda^-$, then the upward adjustment is slower than the downward adjustment.

Avsar and Ruimy (2021) came up with a model for modeling the average term deposit rate. The model introduces the London Interbank Offered Rate (Libor 6M), which is equivalent to the Euribor explained in Subsection 2.1.1. The idea behind using this rate is that term deposit rates would generally depend on not only the overnight market rate but also the term structure of the rates. The model is described as:

$$\Delta d_t = \beta_1 \Delta l_t^{6M} + \beta_2 X_t + \epsilon_t \tag{2.11}$$
$$X_t = d_{t-1} - r_{t-1}^*$$
$$r_t^* = \beta_3 l_t^{6M}$$

where r_t^* is the equilibrium term deposit rate, which is estimated using the long-run relationship between against the Libor 6M; l_t^{6M} is the Libor 6M. Furthermore, X_t is the deviation between the term deposit rate and the equilibrium rate, d_t^t and r_t^* respectively. ϵ_t accounts for the estimation errors. The use of Libor 6M acknowledges that term deposit rates are influenced not just by short-term market rates (as overnight rates), but also by the broader term structure of interest rates. Term deposits, being medium- to long-term products, are naturally influenced by longer-term market rates. The model suggests that the term deposit rate adjusts based on its deviation from the equilibrium rate. If the term deposit rate is below what is suggested by the Libor 6M, it might increase, and vice versa.

Van der Ploeg (2023) investigated ten different deposit rate models based on several methodologies in academic literature from November 2014 till May 2023. The models varied between linear models and first difference (FD) models. Within these models, various predictors were tested such as moving average, autoregressive terms, and asymmetry. The different models are tested on the Akaike information criterion (AIC) performance indicator. The deposit rate model that exhibited the best fit for NN Bank's savings rate is as follows:

$$\Delta d_t = \beta_0 + \beta_1 \Delta d_{t-1} + \beta_2 \Delta MA_t + \beta_3 t + \epsilon_t \tag{2.12}$$

where the model consists of a constant β_0, Δd_{t-1} is the lagged difference of the deposit rate, ΔMA_t is the first difference of the three-month moving average of the market rate, t accounts for a time factor in the model and the error term ϵ_t captures the randomness in the model by using a random standard normal distribution. In time series analysis, using lagged values can help in understanding how past values of a variable influence its current value. The difference part indicates that the model is focusing on the change in deposit rates rather than their absolute levels, which can be useful for highlighting trends or cycles and it is also done to ensure the data is statistically stationary, facilitating more reliable analysis and forecasting. The use of a moving average can smooth out short-term fluctuations and highlight longer-term trends in the market rate. However, the use of t as a variable is unusual for a deposit rate since it is not expected to keep rising continuously over a long period without any declines. It could be the case that adding the time trend in the formula was found to be significant because of the recent deposit rate increase.

The models suggested by Nyström (2008) and Elkenbracht and Nauta (2006) might not be able to capture complex relationships between the deposit rate and the market rate. The sudden increase in interest rates in the banking sector is an example of such a complex relationship. Therefore, relying on historical data for these model parameters could lead to incorrect predictions by assuming that past relationships between deposits and market interest rates will hold without considering how sudden market changes could change these dynamics. Therefore, we need to find a more complex model that does not only look at the current value of the market rate. Jarrow and Van Deventer (1998) incorporate a summation of past market rates and the difference between the current and initial rates. This model would already be an improvement in comparison to the two models of Nyström (2008) and Elkenbracht and Nauta (2006). However, this model does not include the lagged deposit rate, and thus it may lose the context of the previously forecasted rates. The model developed and validated by Van der Ploeg (2023) for the savings rate at NN Bank might be well-suited for this research, given the prior research conducted at the same institution. This previous work provides a strong foundation and context, enhancing the relevance and applicability of the model for our current analysis.

In the context of answering sub-question 1 and understanding the relationship between deposit rates, market rates, and deposit volumes, it is clear that the market rate is a crucial factor in modeling deposit rates. Now, to extend this analysis further, it is important to explore how both deposit rates and market rates influence deposit volume models.

2.1.3 Modeling of deposit volumes

The modeling of deposit volumes is often distinguished into two general types, a volume model or an outflow model. The models in this section use the market rate, deposit rate, or both, as variables. This is because these rates affect customers' saving behavior and act as an incentive for customers to save money. Furthermore, both market and deposit rates are indicators of broader economic conditions and influence the demand for deposits. Higher interest rates make saving more attractive because they offer greater returns on money deposited in savings accounts or other interest-bearing financial instruments. Simultaneously, higher interest rates also make borrowing more expensive, as the cost of loans and credit increases. This combination typically motivates consumers to save more, earn more interest, and spend less, especially on items or services that would require borrowing or financing. As a result, banks often see an increase in deposit volumes due to this shift in consumer behavior towards saving. A variety of relevant deposit volume models that are found in this literature review are described in the following.

Aupérin (2020), Kalkbrener and Willing (2004) and Marena et al. (2023) use an Ornstein-Uhlenbeck (OU) process to capture the dynamics of deposit volume where the volume fluctuates around a trend $\tilde{D}_t$. The dynamics of the deposit volume follow an OU process, but the drift term is a trend that varies over time.

$$D_t - D_{t-1} = k(\tilde{D}_{t-1} - D_{t-1})\Delta t + \sigma\sqrt{\Delta t}\epsilon_{t-1} \tag{2.13}$$

The primary feature of the OU process is its mean-reverting nature. However, the key is that the process does not only revert to a fixed mean but is attracted to a moving target. $\tilde{D}_t$ is the dynamic trend function and this approach acknowledges that the average level itself may shift over time. In the context of a bank's deposit, the trend could represent the expected growth or decline in deposit volumes due to the bank's varying customer base or due to inflation. Fluctuations in the number of customers can lead to variations in deposit amounts. An increasing customer base typically results in higher deposits, while a declining base might reduce them. Inflation impacts the value of money, influencing consumer decisions to save or spend. High inflation may discourage saving due to decreased purchasing power, leading to lower deposits, whereas lower inflation can encourage saving, increasing deposits. These factors contribute to the dynamic nature of average deposit levels, requiring a model like the OU process that accounts for such time-varying trends. Additionally, σ represents the volatility of the deposit volume changes and quantifies the degree to which the deposit volume can fluctuate around its dynamic trend. ϵ_t represents the random shock at time t drawn from a standard normal distribution which introduces randomness into the model subject to unpredictable changes.

Paraschiv and Schürle (2010) incorporate deposit rates in the deposit volume model. The model is designed to capture the dynamic behavior of deposit volumes over time and how they are influenced by the deposit rates offered on deposits by the bank.

$$D_t = \beta_1 + \beta_2 t + \beta_3 D_{t-1}[d_{t-3} - (\delta r_{t-3}^{short} + (1-\delta) r_{t-3}^{long})] + \epsilon_t \tag{2.14}$$

where t represents the time trend, D_{t-1} is the lagged value of the deposits volumes. The next explanatory variable is the spread between the deposit rate d_{t-3} and a weighted average between short and long maturities of market rates, r_{t-3}^{short} and r_{t-3}^{long} respectively. The coefficients $(\beta_1, \beta_2, \beta_3)$ quantify the impact of each factor on deposit volumes. β_1 sets the base level, β_2 indicates the growth or decline rate over time, and β_3 shows the impact of past deposit volumes and interest rate spread on current volumes. In the paper δ is fixed at 0.33, resulting in a higher weight for the longer maturity. The rates are taken with a lag of three periods, suggesting that there is a delay before changes in both the market and deposit rates affect deposit volumes.

Frauendorfer and Schürle (2007) observe that the volume of savings deposits is low when market rates are above average and vice versa. Therefore, a dependency on market rates must be taken into account in a volume model. Frauendorfer and Schürle model relative changes in the deposit volume D_t over time as:

$$\ln D_t - \ln D_{t-1} = \beta_0 + \beta_1 t + \beta_2 \eta_{1t} + \beta_3 \eta_{2t} + \epsilon_t, \tag{2.15}$$

where the constant β_0 and the time component $\beta_1 t$ cause the total deposit volume to follow a linear trend. η_{1t} and η_{2t} represent factors from a term structure model of the market rate that are used as explanatory variables. η_{1t} could represent a factor related to short-term interest rate movements or trends, capturing the immediate response of deposit volumes to changes in the market rates. η_{2t} might capture longer-term aspects of the interest rate environment, such as expectations of future interest rate movements or the economic outlook. These factors could exhibit a correlation between short-term and long-term interest rates. Interest rates across different maturities tend to move together in response to common economic forces, regulatory policies, or market sentiments. This correlation is due to the tendency of interest rates to be influenced by the same factors. The stochastic factor ϵ_t introduces some randomness into the model.

Wyle (2014) proposes a model for the average deposit balance to calculate estimations of future month average balances for various NMD accounts:

$$D_t = D_{t-1} * \frac{SFact_t}{SFact_{t-1}} * (1 + \beta_1 * \frac{\Delta D_{t-1}}{D_t} + \beta_2 * r_t$$
$$+ \beta_3 * r_{t-1} + \beta_4 * \frac{\Delta TD_{t-2}}{D_{t-1}} * \beta_5) + \epsilon_t \quad (2.16)$$

Here, D_t is the total average account balance in month t, ΔD_t is change in total average account balance from month $t-1$ to month t, r_t is the reference interest rate level at the end of month t, ΔTD_t is the change in total term deposits $<$ \$100k account balance from month $t-1$ to month t, $SFact_t$ is the seasonal factor that applies for month t, and finally β_i are the model parameter constants. Term deposit accounts smaller than 100k are chosen because these accounts are most popular due to the Deposit Guarantee System. This model takes multiple factors into account such as seasonality, interest rate levels, and changes in balances. Also adding lagged variables can help capture delayed effects in the system.

Additionally, besides a market and deposit rate model, Van der Ploeg (2023) also used a deposit volume model for modeling the total volume of the savings accounts for each month. The regression model has the following formula:

$$\Delta logD_t = \beta_0 + \beta_1 \Delta logD_{t-1} + \beta_2 \Delta d_t + \beta_3 (c_{St} - dt_t) + \beta_4 (d_t - c_{Bt}) + \beta_5 t + \beta_6 S_{12,t} + \epsilon_t \quad (2.17)$$

where $\Delta logD_{t-1}$ is the lagged change in the logarithm of deposit volumes. Including this term helps to account for autoregression, suggesting that past changes in deposit volumes can influence current changes. Δd_t is the first difference of the deposit rate, c_{St} and c_{Bt} are the deposit rates of small competitors and big competitors respectively. Taking the difference between the bank's deposit rate and that of its competitors shows the competitive effect and their position in the market. t captures any long-term trends in deposit volume growth as might be the case for NN Bank. $S_{12,t}$ is a dummy variable which is equal to 1 for every twelfth month and zero otherwise, and finally, ϵ_t captures factors that are uncorrelated with the factors included in the model that influence deposit volume growth or decline but are not included in the model. All these independent variables aim to predict changes in the deposit volume represented by the dependent variable $\Delta logD_t$. The use of the logarithm transformation is common in economic modeling as it can help stabilize variance, remove heteroskedasticity, and make relationships more linear. Van der Ploeg (2023) tested the significance of all the independent variables and used the AIC to find the best fit on the actual deposit volume of NN Bank. The model that best fits the data used all variables except the dummy variable, despite observable seasonality in the deposit volume. This decision can be understood in the context of an overall decreasing trend in deposit volume over the last year. The overarching trend may have had a more significant impact on the deposit volume than the dummy variable was intended to capture.

Since we are not only interested in the total deposit volume of NN Bank per month but also want to model the deposit outflow of NN Bank, we found the following outflow model in the literature. Gyllberg (2022) suggests using the following outflow model:

$$Y_t = \phi_1 Y_{t-1} + \Phi_1 Y_{t-52} - \phi_1 \Phi_1 Y_{t-53} + \epsilon_t \quad (2.18)$$

where Y_t is the log change at time t, ϕ_1 is the coefficient of the autoregressive variable of the change, Φ_1 is the coefficient for the seasonal autoregressive variable, and ϵ_t is the error term for white noise in the estimations.

In this section, we looked at various models that we could use to model the deposit volume for NN Bank. Each proposed model from the literature has its own view on how to do this and which variables to include. The models of Paraschiv and Schürle (2010), Frauendorfer and Schürle (2007) and Wyle (2014) allow for the expansion or combination of more explanatory variables to capture more complex relationships between market and deposit rates and the corresponding volume of savings accounts and term deposits. In Chapter 4, we will explain which model for deposit volume we will use for the remainder of this thesis.

2.2 Migration modeling

In this section, we will explore how it would be possible to incorporate the deposit rate incentive in a migration model and by doing so answer sub-question 2. The incentive for the customers that will migrate their deposit volume is when the additional interest income is high enough for them. Interest rate sensitivity is important in financial economics and behavioral finance, especially in the context of bank deposits and customer behavior. This includes studying how changes to savings rates affect a customer's choice to stay with their current bank or move to another competitive bank offering better rates. To retain and attract customers, banks try to offer a competitive deposit rate in comparison to other banks in the market. Higher deposit rates will attract more customers but will decrease the net interest margin. The previous section discussed how deposit rates are used in models that calculate the amount of deposits. However, not a lot of attention has been given to the underlying behavior of customers that causes them to move between different types of savings products or withdraw funds from the bank altogether. Therefore, this section describes these behavioral concepts and how they can be measured. We will describe a model that considers the rates other banks offer or the difference in rates between different savings products within the same bank. This will help us understand and predict how customers will respond to changes in deposit rates. In the illustration of Figure 1.1, this is the choice for customers whether to migrate their deposits or not.

2.2.1 Interest rate incentive

Gerritsen and Bikker (2020) researched the savings transfers between banks by retail depositors and how interest rates affect the transfers. The study focuses on the size of the transfer of savings to another account and considers both transfers to newly opened accounts and reallocation across existing accounts. The authors found that interest rate differences play a significant role in depositors' decision to switch banks. Furthermore, the researcher covers the topic of switching costs for a customer. These costs do not have to be monetary but are also psychological, effort-based, and time-based Kiser (2002). In the Dutch banking sector, it is free for a customer to switch from one bank to another. The paper also states that after a depositor has decided to open an additional savings account, these costs are lower. Switching costs are therefore less relevant for the decision on how much savings to transfer to the newly opened account. Other papers by Chakravarty et al. (2004) and Manrai and Manrai (2007) categorize bank customers as having either switched or stayed with their original bank. However, this oversimplifies the reality as many depositors hold savings accounts with multiple financial institutions simultaneously. Gerritsen and Bikker (2020) investigate both the complete transfer of deposits to recently opened accounts at different banks and the trend of partial switches, which is when depositors move a portion of their total savings to a different account.

The ideas expressed by Gerritsen and Bikker (2020) not only clarify movement between banks but also provide a basis for understanding deposit migrations more generally. Although their research focuses on inter-bank processes, the same underlying behaviors and motivations apply to movement between products within a bank. Just like how depositors can move money to other banks to find better deposit rates, customers at a single bank can also transfer their money between different savings accounts or products that have different benefits and returns. Understanding the similarities between these two types of moves can help banks learn from their behavior when transferring funds between different savings products within the same bank. The paper proposes a regression model to capture the proportion of savings that a depositor transfers between banks

$$TransferProportion_{it} = \beta_0 + \beta_1 Rate_{it} + \beta_2 H_{it} + \beta_3 C_t + \beta_4 \lambda_{it} + \epsilon_{it} \tag{2.19}$$

where $Rate_{it}$ represents the difference in deposit rates between the bank receiving the new deposit and the bank losing the deposit for depositor i in year t, λ_{it} is the inverse Mills ratio (Pinelis, 2019). It is a statistical concept used in regression analysis to address selection bias, which occurs when the sample of data analyzed is not representative of the population being studied. It is calculated from the distribution of the variable causing the selection bias and is used to adjust coefficients, ensuring they are more accurate and representative of the entire population. Selection bias, in this research, might arise because not all depositors have the same probability of being included in the study, especially those who actively manage their savings by opening new accounts or reallocating funds. This bias could distort the analysis, making findings less representative of the general depositor population. The instrument is a variable that predicts the likelihood of a depositor's decision to open a new account or change the distribution of savings without directly influencing the actual amount of savings transferred. H_{it} is a vector of control variables that account for other factors influencing the transfer proportion, C_t represents any time-specific effects and finally, ϵ_{it} is the error term. In addition to the article of Gerritsen and Bikker (2020), the article by Zhao et al. (2022) explores the key factors for bank-switching behavior in retail banking. The study revealed that besides effective advertising from competitors, inconvenience, and service failures, the price factor significantly influences customer's retail bank-switching behavior. The price factor includes the interest paid on deposits, which is relevant for this research.

2.2.2 Inertia

Deuflhard (2018) covers the concept of inertia in the Dutch retail deposit market. Inertia refers to the resistance of customers to change or move from their current financial position, even if they could get a better offer from another financial institution. This concept is an important behavioral factor in financial decision-making and can have a significant impact on the movement of deposits between savings products within the same bank. Inertia represents a form of switching cost that is behavioral and rooted in customers' resistance to change. The paper provides new evidence that monetary switching benefits and online banking usage are important determinants of the decision to switch. As the invested volume grows, the relative costs associated with inertia decrease, but in absolute monetary terms, these inertia costs increase. Additionally, the study finds several descriptive patterns in the data consistent with inertia. These patterns could indicate resistance to change, such as long periods where customers maintain their savings in the same account despite changes in interest rates or the introduction of new products. We could identify the frequency and triggers of switching between products. If customers rarely switch, or most of the time do so only when there is a significant change in the terms, this might indicate inertia.

2.3 Conclusion

This chapter aimed to address sub-questions 1 and 2 through a traditional literature review. We tackled sub-question 1 by identifying and explaining different models related to market rates, deposit rates, and deposit volumes. This exploration has given us the knowledge to effectively apply these models to the savings portfolio of NN Bank. Our review revealed that the market rate is often used as a key variable in predicting deposit rates. These market and deposit rates are then, in turn, used as predictor variables in models forecasting deposit volumes, thereby influencing the dynamics of deposit supply and demand.

The focus of sub-question 2 was to understand the motivations behind a customer's decision to move their deposit volume and how to integrate this understanding into our models. However, we found that the current literature does not thoroughly explore models where interest rates or pricing significantly influence customer behavior in terms of switching banks and altering deposit volumes. To bridge this gap, we propose adopting the method of Gerritsen and Bikker (2020). This approach will model the migration of deposit volumes, providing insights into how changes in interest rates and pricing strategies may influence customer decisions to switch savings products. By incorporating this method, we aim to develop a more comprehensive understanding of the factors driving deposit volume migration and increase our predictive capabilities in this area. Furthermore, in our analysis of the data and the interpretation of the results we will consider the concept of inertia. The specific models selected from our literature review for use in our methodology, along with the reasons for their selection, will be detailed in Chapter 4.

3 Data

In this chapter, we aim to answer sub-question 3. We gather all the necessary data for our study and discuss the sample statistics. The data includes historical market rates, NN Bank's deposit rates, its competitors' rates, and the deposit amounts for NN Bank's different savings products. This data is needed for our predictive models and to understand the relationship between the different variables.

Firstly, in Section 3.1, we focus on the market rates by looking at the Euribor rates. These rates help us understand the broader financial environment that NN Bank operates within. Next, in Section 3.2, we examine and compare NN Bank's deposit rates with those of other banks. This comparison shows how NN Bank's rates have changed over time and how they stack up against the competition. Lastly, Section 3.3 presents the changes in the total amount of money held in NN Bank's various savings products. We break down these amounts by product class to see where most customers are putting their money.

The data in these sections gives us a clear picture of how deposit volumes have shifted over time in response to changing interest rates and competition. The historical data is also used as the input for our models to forecast the market rate, deposit rate, and migration between deposit products. One of the datasets that we are utilizing in this study, is a dataset provided by NN Bank, which includes monthly deposit volumes for each account, spanning from February 2015 to December 2023. This is a different dataset than the one used in the research of Van der Ploeg (2023), which only contained aggregated monthly data. The dataset features key information such as the client ID, account number, account opening date, deposit rate, and for term deposits, both the maturity date and term length in years. NN Bank offers savings accounts and term deposits. To open a term deposit, customers must first have a savings account. This requirement helps us monitor deposit flow among customers. Our initial step was to address data quality issues, including missing values and implausible account records, such as those with starting dates as early as 2008, which is implausible because the bank did not exist in that year. To handle missing values, we attempted to impute them by identifying if absent client numbers appeared in later months on an account of the same client. This method was also applied to term deposits lacking maturity data. After removing data rows that still contained unresolved missing values or unusual outliers. The advantage of our dataset, compared to the dataset used in Van der Ploeg (2023), is that we can look at the deposit volume on a client-based level, providing us with additional insights into the monthly behavior of the clients as will be explained in Section 3.3. Table 3.1 shows an example of the monthly data of a client.

Table 3.1: Example rows of data of a client

date	account_nr	client_id	product	principal	coupon	start_dt	maturity_dt	term
30-11-2023	20009344	20582931	Savings account	3000	1.5	31-10-2020	-	
30-11-2023	13591039	20582931	Term deposit	1500	2.6	31-01-2021	31-01-2024	3

3.1 Historical market rates

As we have introduced in Subsection 2.1.1, the Euribor represents the average interest rates at which a large panel of European banks borrow from one another in euro. It is a key reference point in the banking sector, influencing the interest rates that banks offer on savings products, which tend to follow the trend of the Euribor to some extent. In Figure 3.1, we illustrate the trends of the 1-month (1M), 3-month (3M), and 6-month (6M) Euribor over a given period. These specific maturities have been selected due to their relevance in the banking sector and reflect a range of market conditions. Because they are benchmarks in financial products and have liquidity and stability, they are dependable indicators. They offer a varied perspective of interest rate trends in the short to medium term, which is critical for precise market rate forecasting. The historical data of the Euribor rates is gathered from the database of the ECB.

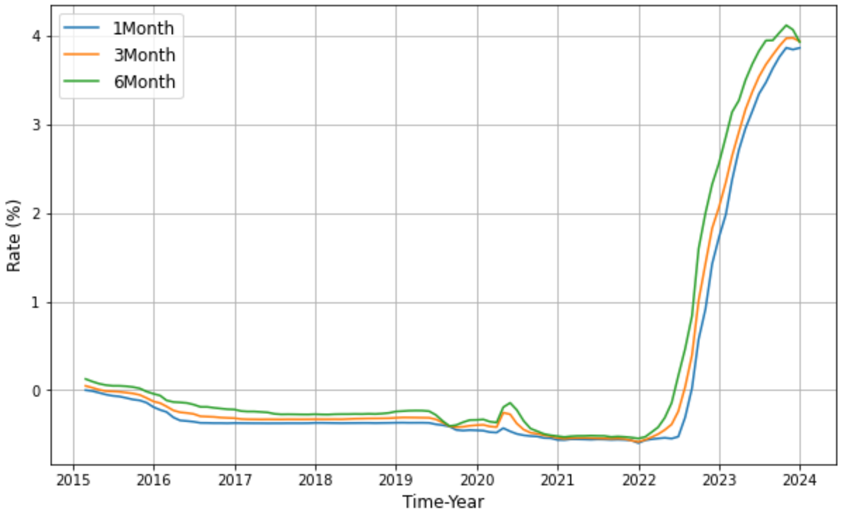

Figure 3.1: Historical Euribor rates

Additionally, Figure 3.2 displays the interest rates for Euro IRS with durations from one year to ten years. These swap rates represent the agreed-upon interest payments for exchanges over predefined time frames and show how banks hedge against and speculate on future changes in interest rates. We choose swaps with maturities of one to ten years because those are the lengths of the term deposits that NN Bank offers.

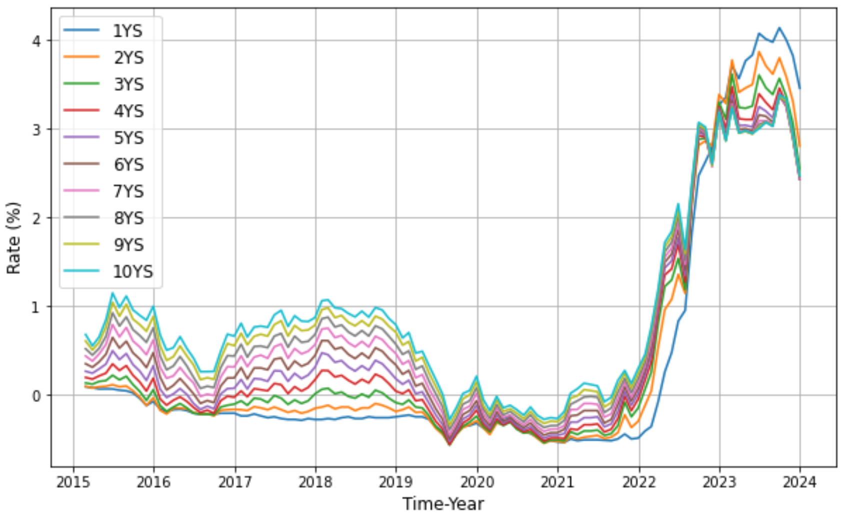

Figure 3.2: Historical swap rates

For both Euribor and swap rates, the upward trend towards the end of the observed period is the result of a shift in monetary policy, such as central banks, like the ECB, raising rates to combat inflation leading to higher demand for capital and hence higher interest rates. Swap rates across all maturities converged noticeably around 2020, suggesting a period of uncertainty or a significant market event. Currently, the shorter-term interest rates provide more interest than the longer maturities, resulting in an inverted yield curve. This phenomenon has historically preceded recessions and is uncommon in the economy (Fonseca et al., 2023). An inverted yield curve can result in a challenging period for banks, where it could impact their margin on deposit rates and they could be exposed to interest rate risk.

When examining term deposits, the IRS rates for corresponding maturities can be a valuable indicator of the economic outlook. For term deposits with fixed maturities, such as one year to ten years, the fixed leg of an IRS with the same maturity can provide insight into market expectations for future interest rates. This is because the fixed rate on a swap is essentially the market's consensus on the average level of the floating rate (like Euribor) over the swap's term. Banks often use these swap rates as a reference rate to price their term deposit products. By comparing the fixed rate of a swap to the fixed interest rate offered on a term deposit of the same maturity, banks can assess the relative attractiveness of the term deposit. We will use the Euribor to forecast the market rate. However, due to the large increase in the interest rate, the data characteristics of the historic Euribor data may lead to incorrect statistical properties of the time series. Therefore, we will test for stationarity of the time series in Section 3.4.

3.2 Historical deposit rates

In this section, we will show the deposit rates given out by NN Bank and its competitors since February 2015 on savings accounts and term deposits. The reason for choosing February 2015 is that NN Bank's dataset has its earliest records in that month.

3.2.1 Deposit rate on savings accounts

The historic deposit rates on saving accounts are gathered via an external party called MoneyView. The data is provided monthly and consists of the savings rate of NN Bank and other Dutch banks.

We can look at the deposit rates given out by NN Bank and the deposit rates given out by its competitors. In this research, the competitors of NN Bank will consist of other banks in the Netherlands. Although deposit volumes are shifting towards banks in other countries, we continue to concentrate on the Dutch banking sector as the core competitive landscape for NN Bank (Financieel Dagblad, 2023). For comparing the rate to its competitors, we should make a distinction between the sizes of the banks. The larger banks typically do not offer competitive deposit rates, while smaller banks strive to do so. Therefore, it is interesting to see where NN Bank positions itself in the Dutch retail banking market. In the Dutch retail banking sector, often the comparison is made between the three biggest banks (ABN Amro, ING, and Rabobank) based on their market share in the Netherlands and the other, somewhat, smaller banks (Banken.nl, 2023). Appendix A includes a list of which bank belongs to which of these two categories. NN Bank is a relatively small bank when comparing the market share to bigger banks. Smaller banks aim to attract more customers and deposits by offering more competitive interest rates. Figure 3.3 shows the historical rates of NN Bank, the average of the three biggest banks, and the average of the other smaller banks for the deposit accounts with the range €0-100,000. The figure provides information on how NN Bank's savings rate compares to other banks. Their rate is consistently higher than big banks, except during the period when interest rates were around zero. This tells us that the bank has deliberately chosen to keep its savings rate competitive over the past eight years and probably will do so shortly. For NN Bank, this is a way for banks to retain and attract customers.

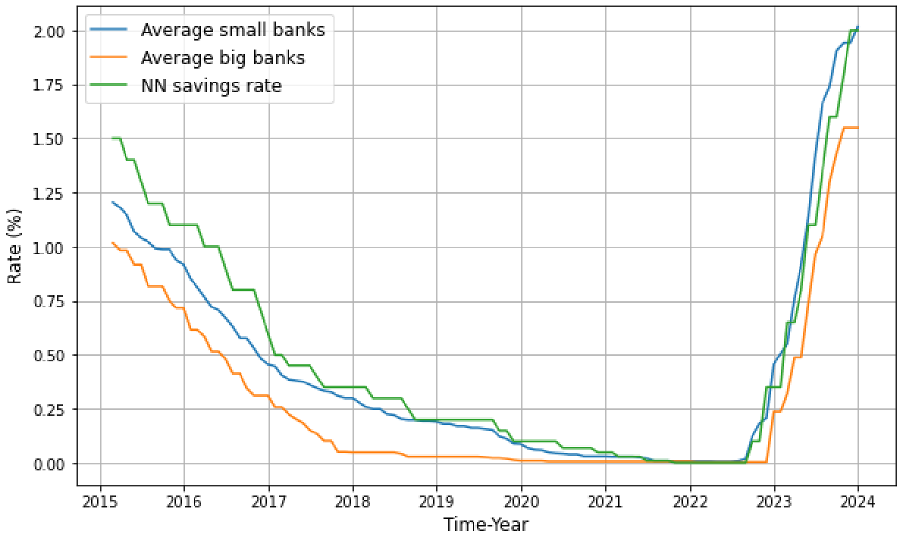

Figure 3.3: Savings rate of NN Bank vs Dutch small and big banks

Figure 3.3 illustrates that throughout the period, small banks offer higher rates than large banks. This is consistent with the competitive dynamics where smaller banks often offer more attractive rates to gain market share. There is a general downward trend in the savings rates for all banks from 2015 until around the end of 2021. This period corresponds to a low-interest-rate environment, often implemented by central banks to encourage borrowing and investment, especially during the COVID-19 period. In June 2022, there was a sharp increase in the savings rates, with NN Bank's rate and the rates of small banks increasing more significantly than those of large banks. Large banks often have a big customer base that is more loyal to the bank. Customers may prioritize the convenience, range of services, and belief in safety offered by larger institutions over slightly higher interest rates. This reduces the pressure on large banks to raise rates immediately to retain customers.

3.2.2 Deposit rate on term deposits

The past rates for term deposits were obtained from Excel files from NN Bank. These Excel files contained a monthly overview of the rates of NN Bank compared to its competitors. NN Bank now provides term deposits ranging from one to ten years. In previous years, these options were limited. Term deposits for periods shorter than six years were not available from December 2020 through March 2022, and the one-year deposit term was only available again from December 2022. Each maturity has its interest rate as can be seen in Figure 3.4. These rates are determined by experts of NN Bank by looking at the market rate, such as the market rates mentioned in Section 3.1, and the rates given out by competitors. By comparing Figure 3.2 and Figure 3.4, it is visible that the swap rates and deposit rates closely follow each other.

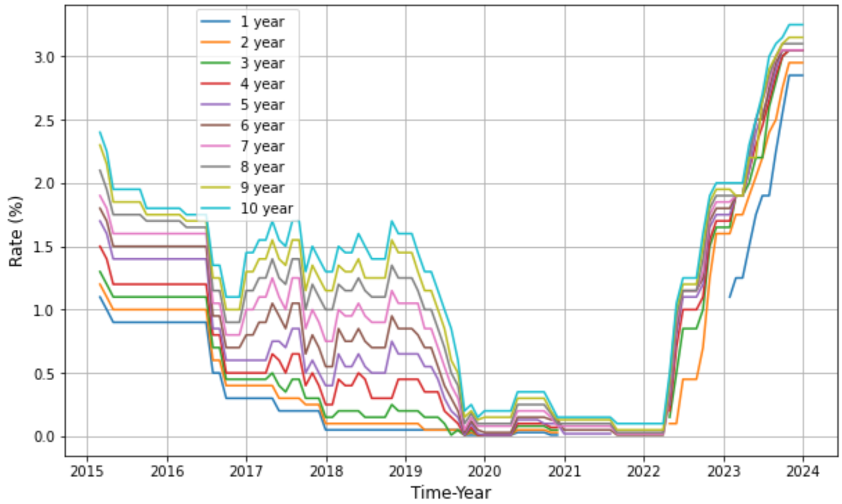

Figure 3.4: Term deposit rates of different maturities

Figure 3.5 shows the average term deposit rate of the Dutch big banks, smaller banks, and NN Bank. The average rate is calculated as the average for only the banks that provided the rate each month and each maturity. After that, each maturity is weighted by the number of banks that provided it. Similar to Figure 3.3, NN Bank has a competitive deposit rate in the Dutch banking sector. Determining the interest rates that NN Bank sets on its term deposits is a trade-off between maintaining and attracting deposit volume and paying too much interest on a term deposit instead of the interest paid on a savings account.

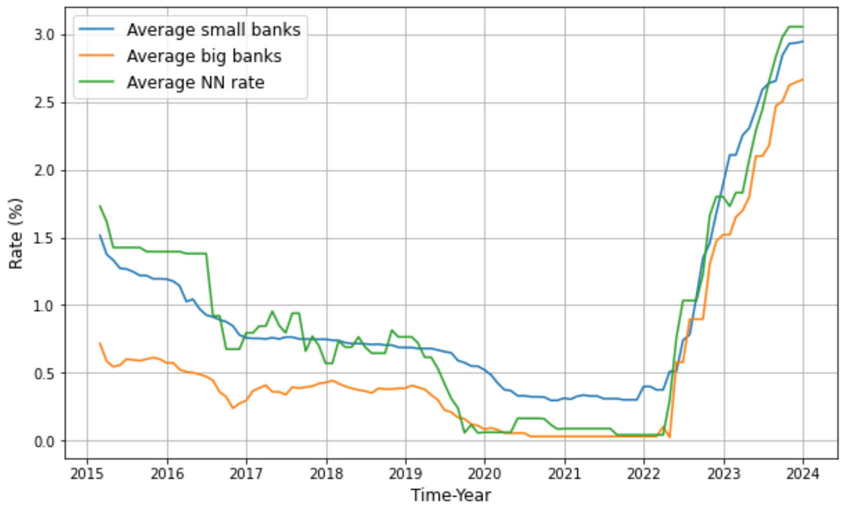

Figure 3.5: Average term deposit rates of NN Bank vs. Dutch small and big banks

3.2.3 Comparison of savings and term deposit rates

We will compare the spread between the average term deposit rates and the savings rate of NN Bank as this gives customers the incentive to migrate their deposit volume towards term deposits. As can be seen in Figure 3.6, the spread between the rates is not constant over time. It is interesting to note that the term deposit rates were raised about three months earlier than the savings rate. This leads to a rapidly increasing spread between the savings rate and the average term deposit rate. We can also see that the spread between the two rates has not been as large as in previous years. Furthermore, the savings rate seems to follow the Euribor rates more closely and the term deposit rates seem to follow the IRS more closely. The reason for this could be the forward-looking aspect of the IRS, just as the term deposit rates. Also, IRS rates include premiums for various risks, including credit risk and inflation risk. These premiums can change based on market sentiment and expectations, causing IRS rates to move differently than Euribor rates.

Another explanation for the earlier increase in term deposit rates is to attract more deposits and lock in funds for a fixed period. This can be beneficial for a bank in a rising interest rate environment, as term deposits provide more stability for managing interest rate risk and are a reliable source of funding. In addition, the deposit rates set for term deposits depend on the deposit rates set by other Dutch banks. To avoid losing customers seeking high returns in a rising deposit rate environment, a bank will try to keep pace with its competitors. If one of NN Bank's competitors decides to increase the deposit rates on term deposits, NN Bank will not be able to lag behind.

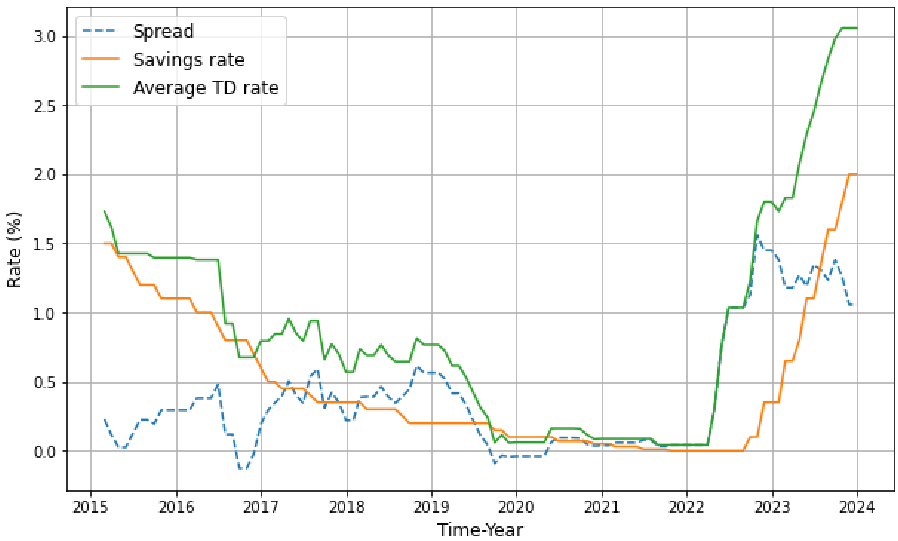

Figure 3.6: Comparison of savings rate vs. term deposit rates

3.3 Historical deposit volumes

As mentioned earlier in this chapter, we will use a dataset that includes NN Bank's savings accounts and term deposits on a client-based level. We can gain insight into the product volumes by aggregating them monthly. Figure 3.7 shows the total deposit volume per product of NN Bank. As can be seen in this figure, the deposit volume of term deposits has increased significantly in the last year due to higher term deposit rates. The higher rates serve as a stronger incentive for customers to choose term deposits over other saving options. On the contrary, the volume of savings accounts has decreased in 2023. We will map the flow to see if this money stayed in the bank to open a term deposit or if it flowed out of the bank.

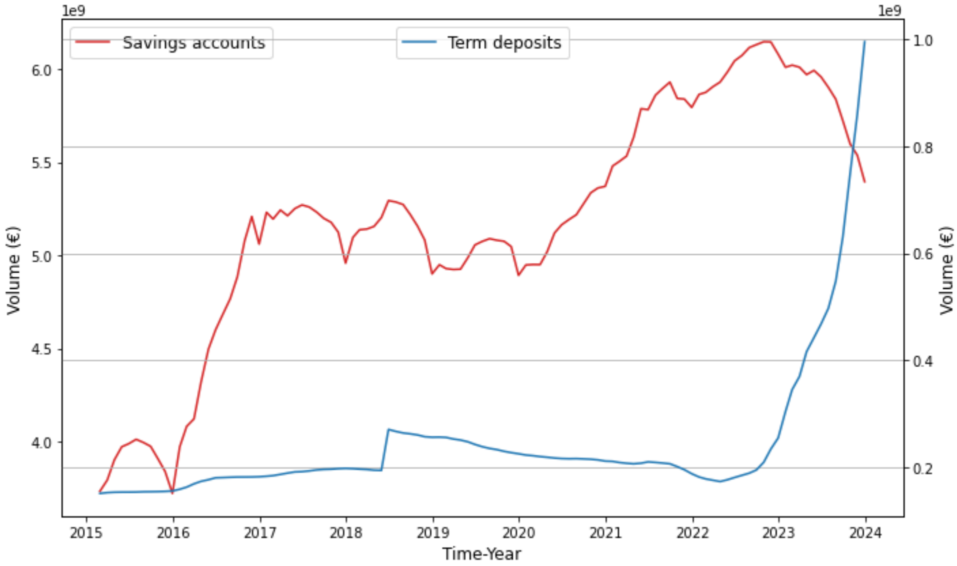

Figure 3.7: Historical deposit volume per savings product

Additionally, Figure 3.8 shows us how many term deposits started in which month of 2023 with a certain term. Of the 32098 term deposits that were opened at NN Bank from the beginning of the bank, 24431 were opened in 2023. This can be explained when looking at the deposit rates in Figure 3.4. Taking a term deposit with a relatively short maturity was not attractive because of the low interest rates at the time and the fact that a customer could not withdraw his money. The number of term deposits opened at NN Bank has increased significantly, but we notice that most of the term deposits have a maturity of one year, because people are willing to tie up their money for a relatively short period for a high-interest payment.

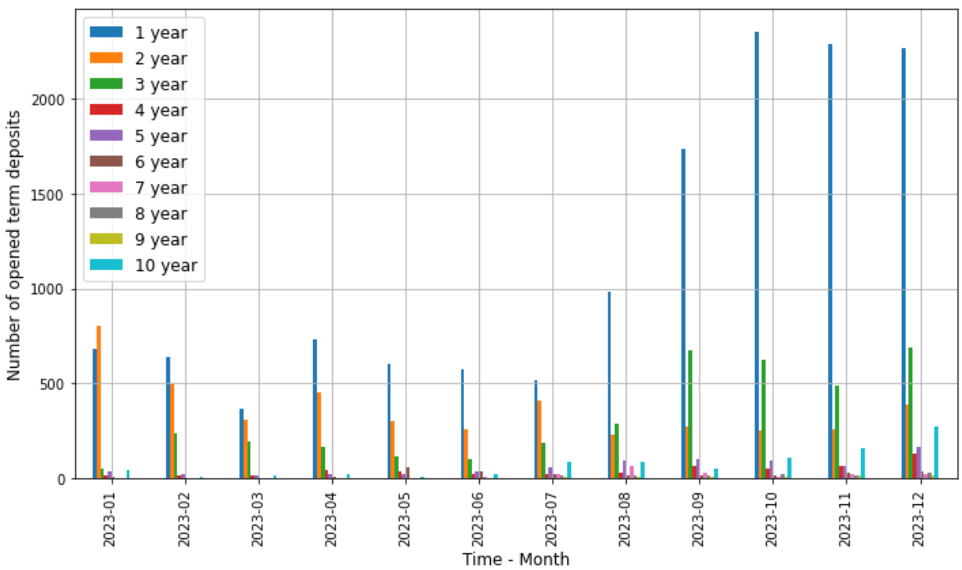

Figure 3.8: Amount of term deposits opened in 2023

Comparing Figure 3.6 with Figure 3.7 and 3.8, it is notable that people began migrating their money to term deposits when the term deposit rate reached a certain threshold, and because the spread between the savings rate and term deposit rate increased significantly. Interestingly, even though the spread between the savings and term deposit rates was greater at the end of 2022 than at the beginning of 2023, the migration of funds to term deposits started in early 2023, when this spread was smaller. This suggests that the decision to move money to term deposits was influenced by more than just the spread between the two rates. Specifically, it appears that the actual rate offered on term deposits was an incentive for customers. Customers are now willing to accept the downside of not being able to withdraw their savings for a higher deposit rate.

Besides looking at total deposit volume, we can use the client-based data to map certain flows of the savings portfolio of NN Bank. This involves examining specific flows of money, such as those exiting the bank, transfers from savings accounts to term deposits, and the proportion of new term deposits financed by previously held savings. Studying these deposit flows in detail helps us understand how clients make financial decisions, which will contribute to answering our main research question. In this research, we define the money that has been in a savings account in the previous month and then flows out of the bank or to a term deposit of NN Bank as 'migration of deposit volume'. We have programmed such a flow by looking at a negative savings volume difference of a client between two months. If the savings volume difference is negative and a term deposit has been opened, we classify the amount deducted from a customer's savings account as a migration within the bank. For term deposits, this migration may be the entire principal of a term deposit or a portion of it. This depends on whether a customer has deposited additional savings in their savings account to be used as funding for the term deposit. Table 3.2 shows a numerical example of how the migration is defined and extracted from our dataset. It demonstrates how savings volume can be used to open a term deposit — either fully, partially, or not at all.

Table 3.2: Example of deposit volume migration and new money

	Client x	Client y	Client z
Δ Savings account volume	-€ 1,000	-€ 500	€ 0
Term deposit opened	€ 1,000	€ 1,000	€ 1,000
Migration volume	€ 1,000	€ 500	-
New money volume	-	€ 500	€ 1,000

This research focuses on mapping and forecasting the migration of money that has been in a customer's savings account for some time before it migrates out of the savings account. The research questions of this research aim to create an understanding of how deposit volumes shift in response to various factors, leading to the development of a model that captures these dynamics.

Due to the fact that NN Bank itself does not have any current accounts, the money that flows out of a savings account either leaves the bank or goes into a term deposit of NN Bank. Customers must always have a savings account before they can open a term deposit. This makes it easier to map the inflows and outflows of NN Bank's savings products, but we only have month-end data for all customers. As a result, we have to make assumptions about the inflows and outflows of a savings account during the month:

- money is classified as a migration even as it have could be deposited a few days earlier

- outflow is only seen as a negative volume difference on an account from one period to the next
- inflow is seen as a positive volume difference on an account from one period to the next, this includes maturing term deposits and deposits made at the bank

Figure 3.9 gives us insights into the total flow of deposit volume. While the money flowing out of the bank used to be the same as the money flowing out of the savings accounts, we see a new trend in the last year. The money that would have flowed out of the bank is now flowing into term deposits. This can be seen by looking at the outflow from savings accounts and comparing it to the outflow from the bank. These lines were almost always the same until the end of 2022, where a portion of the money that would migrate out of the bank now migrates to term deposits as can be seen by the increasing line of 'migration to TD'. This flow could also be seen as internal cannibalization, i.e. deposit volumes that would normally flow out of the bank are now flowing into term deposits. The implications for NN Bank are twofold. On one hand, it is positive that the bank retains this volume; on the other hand, the bank has to pay additional interest because of the higher term deposit rates. This is related to the sub-question 4a, where we are interested in the interest rate incentive for customers to open a term deposit. In addition, we can see seasonality in the outflow of savings accounts, with a sharp increase in outflows at the end of each year. We are interested in the future outflow distribution of NN Bank and what proportion of outflow migrates towards term deposits. This gives NN Bank more insight into its future cash flows and will contribute to answering the main research question of this research.

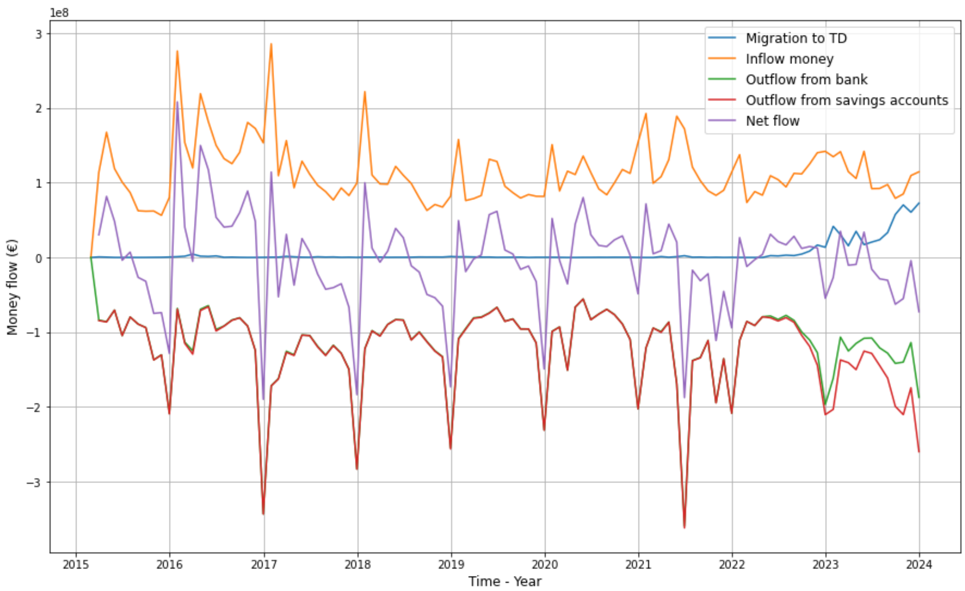

Figure 3.9: Inflow and outflow of deposit volume of NN Bank's savings portfolio

Furthermore, it is valuable to analyze the volume of the different maturities of the term deposits opened within a month. This analysis helps identify trends and preferences among customers, offering insights into the most and least favored term lengths during our dataset. This helps the bank to match the maturity of its interest-bearing products used to hedge the term deposits. We will divide the maturities into 1-year, and 2-year and group the rest of the maturities. This is because we see an increase in the volume of term deposits, especially in the 1-and-2-year term deposits. Term deposits with a longer maturity may only be attractive to customers who have enough savings to make such a decision. This can also be seen in Figure 3.10 where we see what proportions of the outflow of the savings accounts migrates to which maturity. The biggest proportion is towards the 1-year maturity, followed by the 2-year maturity, and finally the grouped maturities.

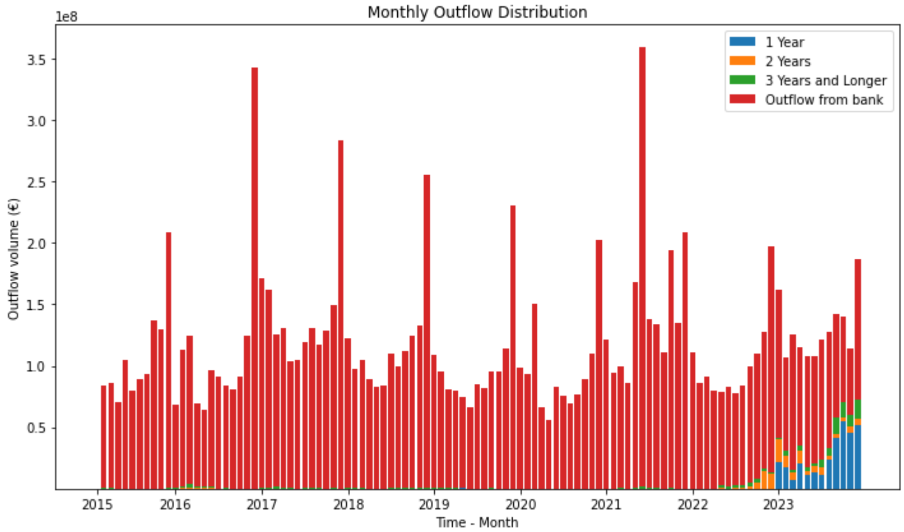

Figure 3.10: Proportional outflow from savings accounts

Additionally, the following analysis will examine the relationship between interest rates and deposit volumes to assess the correlation between the market rate, deposit rate, and deposit volume of both savings products. It will consider how deposit rates, whether market-driven or bank-specific, influence the volume of deposits consumers place in savings products. The correlation test will determine the interdependencies between these variables and their impact on savings behavior and savings product design as this is part of sub-question 3. The Pearson correlation coefficient will be used to determine the linear relationship.

Table 3.3: Correlation matrix savings account

	Deposit rate	Euribor 6M	Volume
Deposit rate	1.00	0.56	-0.46
Euribor 6M	0.56	1.00	0.37
Volume	-0.46	0.37	1.00

A correlation coefficient of 0.56 between the deposit rate and Euribor 6M suggests a moderate positive correlation. This means that as the Euribor 6M rate increases, the deposit rates offered by banks also tend to increase, and vice versa. This relationship is logical since deposit rates often adjust in response to changes in benchmark interest rates like the Euribor, reflecting broader economic conditions and central bank policies. A correlation coefficient of -0.46 between the deposit rate and volume indicates a moderate negative correlation. This suggests that an increase in the deposit rate is associated with a decrease in the volume of deposits, and vice versa. This may seem counterintuitive, as higher rates typically attract more deposits. However, the increase in savings volume at NN Bank despite low or decreasing deposit rates could explain this. Additionally, we have observed a decrease in deposit volume due to the migration towards term deposits in the current environment of increasing deposit rates. A correlation coefficient of 0.37 between Euribor 6M and volume indicates a weak to moderate positive correlation. This suggests that as the Euribor 6M rate increases, the volume of deposits also tends to increase. This relationship could reflect broader economic trends, where higher Euribor rates are associated with stronger economic activity, like inflation, or expectations, leading to increased savings and deposits.

To create the correlation matrix for the term deposit volume, we selected the deposit rate for the 1-year maturity and the 1YS rate. These rates were chosen because the 1-year maturity term deposits were the most popular at NN Bank and accounted for the biggest proportion of the volume. The 1YS rate was selected because it is expected to represent the relationship the best. Finding the relationship between these variables enhances our understanding of the characteristics and trends in migration modeling.

Table 3.4: Correlation matrix term deposits

	Deposit rate	1YS rate	Volume
Deposit rate	1.00	0.86	0.76
1YS rate	0.86	1.00	0.73
Volume	0.76	0.73	1.00

The correlation coefficient of 0.86 between the deposit rate and the 1YS rate indicates a strong positive correlation. This suggests that movements in the deposit rates are closely aligned with movements in the 1YS rates. When the 1YS rate increases, the deposit rates are likely to increase as well, and vice versa. This relationship is expected as the 1YS rate reflects the market's expectations for interest rates over the next year, which banks might use as a benchmark for setting their deposit rates. A correlation coefficient of 0.76 between the deposit rate and volume gives a strong positive correlation. Unlike the previous table, this indicates that as deposit rates increase, the volume of deposits also increases. This relationship is more in line with usual expectations, where higher interest rates attract more deposits by offering greater returns to savers. The correlation coefficient of 0.73 between the 1YS rate and volume also indicates a relatively strong positive correlation. This suggests that higher 1YS rates, which may reflect expectations of rising interest rates or stronger economic conditions, are associated with increased deposit volumes. This could be due to consumers seeking to lock in higher returns on their deposits in anticipation of rising rates or as a response to improved economic conditions.

3.4 Stationarity of the time series

3.4.1 Stationarity tests

We will use the historic data market rates, deposit rates, and deposit volumes or flows to forecast the migration of deposit volume. However, due to the large increase in the interest rate and its consequences, the characteristics of the historical data may lead to undesirable statistical properties of the time series. As mentioned in Chapter 2, the models to forecast the market rate, deposit rate, or deposit volume often use regression models. One of the properties of the data on which we want to test our times series is stationarity (Affek, 2019). Stationarity in a time series indicates that the process generating the series has constant statistical properties over time, specifically in terms of its mean, variance, and autocorrelation. A stationary process exhibits a constant mean, ensuring the time series does not trend in any particular direction. The variance remains constant over time, meaning the series does not exhibit periods of varying volatility. Finally, the autocorrelation or covariance of the series is dependent only on the time lag between observations, not on the time at which they are measured (Witt et al., 1998). These properties are important to effectively and precisely predict time series such as ours. The three types of stationarity include strict stationarity, seasonal stationarity, and trend stationarity. To test for stationarity, there are two commonly used statistical tests: the Augmented Dicky-Fuller test (ADF test) and the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS test) (Gimeno et al., 1999). The ADF test is a test that tests if a time series has a unit root. If a time series has a unit root, it shows a systematic pattern that is unpredictable. If the test statistic is less than the critical value and the p-value is less than 0.05, we reject the null hypothesis and conclude that the time series is stationary.

The KPSS test works the other way around. The KPSS test is also a unit root test, but it tests for the stationarity of a time series around a deterministic trend. If the p-value is less than 0.05, the series rejects stationary. The power of these tests, like all statistical tests, is affected by sample size. With 107 observations, there is a risk that these tests might not be as powerful in detecting stationarity or non-stationarity compared to a longer dataset. This means there is a higher chance of Type II errors (failing to reject a false null hypothesis). We will still perform these tests but we will use a graphical analysis, such as the plotting of the time series, to support the findings of the tests.

Table 3.5: Hypothesis tests for stationarity: ADF and KPSS

Test	Null Hypothesis (H_0)	Alternative Hypothesis (H_A)
ADF	The series has a unit root (is non-stationary).	The series does not have a unit root (is stationary).
KPSS	The series is stationary.	The series is non-stationary (has a unit root).

The most common approach to dealing with non-stationarity is to take the first difference of the series. Differencing involves subtracting the previous observation from the current observation. After taking the first difference or logarithm, we need to test for stationarity again. Besides the ADF and KPSS tests, we can also visualize the (non-)stationarity of a series using the autocorrelation function (ACF) and the partial autocorrelation function (PACF). ACF measures the correlation between a time series and lagged versions of itself. Peaks in the ACF plot indicate lags where the time series data points are significantly correlated. PACF measures the correlation between a time series and lagged versions of itself, but after removing the effects of shorter lags. Peaks in the PACF plot suggest a significant relationship at specific lags not accounted for by earlier lags. By looking at where the ACF/PACF plots cut off, we can identify the potential orders of AR and MA components for an ARIMA model (Ahmed, 2023). In Table 3.6 below, the results of the ADF and KPSS tests of the migration flow time series from the savings accounts of NN Bank are shown.

Table 3.6: ADF and KPSS test results for normal data, log-transformed data, and differentiated log-transformed data.

Metric	**Normal Data**		**Log Transf. Data**		**Diff. Log Transf. Data**	
	ADF	**KPSS**	**ADF**	**KPSS**	**ADF**	**KPSS**
Test Statistic	1.72	0.77	-2.03	0.53	-12.43	0.12
p-value	0.99	0.01	0.27	0.034	$3.86e^{-23}$	0.1
Lags Used	9	5	1	5	0	3
Number of Observations Used	97	102	105	101	104	101
Critical Value (10%)	-2.58	0.34	-2.58	0.34	-2.58	0.347
Critical Value (5%)	-2.89	0.46	-2.88	0.46	-2.89	0.463
Critical Value (2.5%)	-	0.57	-	0.57	-	0.574
Critical Value (1%)	-3.49	0.73	-3.49	0.73	-3.49	0.739

The original time series representing migration into term deposits is non-stationary, as indicated by both the ADF and KPSS tests. Initially, we attempted to achieve stationarity by applying a logarithmic transformation to the time series, but this approach was not successful. Specifically, we failed to reject the null hypothesis of the ADF test, which suggests non-stationarity, and we rejected the null hypothesis of the KPSS test, further indicating non-stationarity for the log-transformed series. However, when we applied the first difference to the logarithmically transformed series, the time series became stationary as per both the ADF and KPSS tests. This result implies that the time series is "Integrated of order 1" or $I(1)$. In other words, the time series showed non-stationary characteristics, but it obtained stationarity after being differenced once. This successful transformation confirms the $I(1)$ classification of the time series.

In our analysis, we also focused on two aspects: the total number of observations in the dataset and the number of lags used in the ADF and KPSS tests. The number of observations is important for the reliability of the test. For selecting the appropriate number of lags, we used the AIC to balance capturing the data's autocorrelation without making the model too complicated. This approach ensured a statistically sound analysis of the time series data's stationarity. However, with only 107 observations, we cannot fully rely on the test outcomes due to low statistical power where including more lags results in losing more observations. Therefore, we should be cautious when interpreting the results and use additional plots of the transformed time series to confirm that they are indeed stationary. Transforming this time series is useful to use the data for forecasting using regression models.

However, transforming the data to achieve stationarity has its downsides. By differencing the data to model interest rates, we potentially lose the predictive power inherent in the level of interest rates. This could result in inaccurate fits and predictions for our models. Therefore, in our research, we will try to account for stationarity but should focus on the predictive power of our models and whether valuable information is lost. This is a careful consideration cause it could have a big impact on the results of our research.

3.4.2 Vector Autoregressive model

To account for the problem of losing predictive power, we propose using a Vector Autoregressive (VAR) model to test if we can use this model to overcome this problem (Zivot & Wang, 2006). The VAR model is a statistical tool for analyzing multivariate time series data. It captures the linear interdependencies among multiple time series by modeling each variable as a linear function of its past values and the past values of all other variables in the system. Each equation describes one variable as a linear combination of lagged values of itself and lagged values of the other variables. The number of lags included (the lag order) is an important parameter that can be selected based on various criteria to best fit the data. A simple VAR model of two time series and lag order one is given in equation (3.1) below:

$$\begin{aligned} y_{1,t} &= c_1 + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + e_{1,t} \\ y_{2,t} &= c_2 + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + e_{2,t} \end{aligned} \tag{3.1}$$

The model is useful for understanding the relationships between variables and predicting future values (Zivot & Wang, 2006). A VAR model can be effective in ensuring the predictive power of time series data that are differenced twice. This approach addresses the challenges that come with over-differencing, which can lead to the loss of valuable information and the introduction of complexity into the model. The model could capture the underlying dynamics of multiple interrelated time series, even when individual series are differenced twice. In Section 4.3, we will explain how we will apply the VAR model for attempting to forecast the migration of deposit volume and which time series will be included in the model.

3.5 Conclusion

The purpose of this chapter was to answer sub-question 3 by analyzing and understanding the features and patterns in the data that we have gathered. This analysis aimed to improve our understanding of the key drivers behind the movement of deposit volumes. Additionally, we aimed to explore how market rates and deposit rates are interconnected and their impact on deposit volumes. We used graphs to visualize the time series and how to interpret them. It is visible from our observations that deposit rates tend to mirror market rates. Furthermore, an analysis of the graph illustrating term deposit volume reveals a recent trend in which a significant number of new term deposits are being opened, as a result of the attractive high deposit rates. We also introduced the statistical property of stationarity, which needs to be tested before we can use the time series for regression models. Following this, VAR models are introduced as a powerful tool for forecasting time series data, using the interdependencies among multiple variables. Now that we have analyzed our data, we can design our methodology for this research in Chapter 4.

4 Analytical framework

In this chapter, we aim to address sub-question 4 by developing a framework that models the market rate, deposit rate, and deposit volume to eventually model the migration of deposit volume. As we discovered in Chapter 2, there is limited existing literature on the integration of interest rate incentives into the modeling of deposit volume migration within a bank and on modeling the migration itself. However, our proposed framework includes a model that incorporates this incentive. Following the establishment of this framework, we apply the data gathered in Chapter 3 to calibrate our models and their coefficients, as detailed in Chapter 5. Subsequently, we use these calibrated models to predict future rate trends and attempt to forecast the migration of deposit volumes. Since we deal with non-stationarity issues in our data as described in Chapter 3, we also propose a separate framework for using the different time series of our data in a VAR model. In this chapter, we explain which time series we use and how we developed our VAR model.

4.1 Approach

To provide a clear overview, this section describes our methodology and how we use all the models together. We also describe the performance measures we will use and how we will fit our models and forecast the different time series.

4.1.1 Model overview

From our literature review on deposit rate modeling, we conclude that both our savings rate model and our term deposit rate model require input from the market rate as an explanatory variable which is shown in Figure 4.1. Therefore, we need a market rate model that can capture the historical market rate dynamics and calculate the yield, which is described in Section 4.2.1. We also need to model and forecast the deposit rate on savings accounts and term deposits as they are explanatory variables in our deposit volume models and migration model. Since these rates have different characteristics in terms of their maturity, we use different models to forecast these rates. The main difference is the possibility for the bank to change the deposit rate on savings accounts daily. The models for both rates are discussed in 4.2.2.1 and 4.2.2.2 respectively. Additionally, we construct a model for the monthly volume of savings accounts, which serves as an explanatory variable in our outflow analysis. The outflow model offers insights into the outflow distribution from NN Bank's savings accounts and determines what fraction of this outflow is likely to migrate towards term deposits each month. The interconnectedness of our models and their mutual influence are depicted in Figure 4.1.

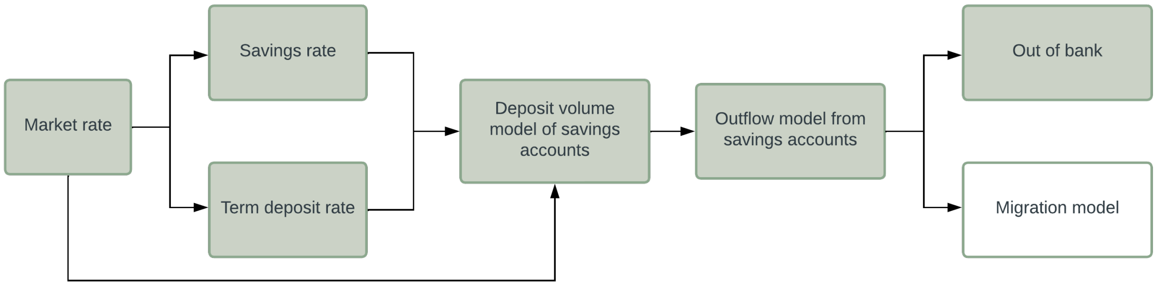

Figure 4.1: Model overview

The migration volume that we want to predict in this research is highlighted by the white cell in the figure above. Our proposed model that aims to predict this monthly migration volume requires the input of the savings rate, term deposit rate, average vintage of the customers, and the market position of NN Bank. The overview of this model is given in Figure 4.2 along with the explanatory variables representing equation (4.8). To calculate the average vintage of customers, we take the average number of months that customers who opened a term deposit in a given month have been clients at NN Bank. The market position of NN Bank is determined by comparing the term deposit rate given out by NN Bank with other Dutch banks and whether their rate is among the top 3 rates. The model is explained further in Section 4.2.4.

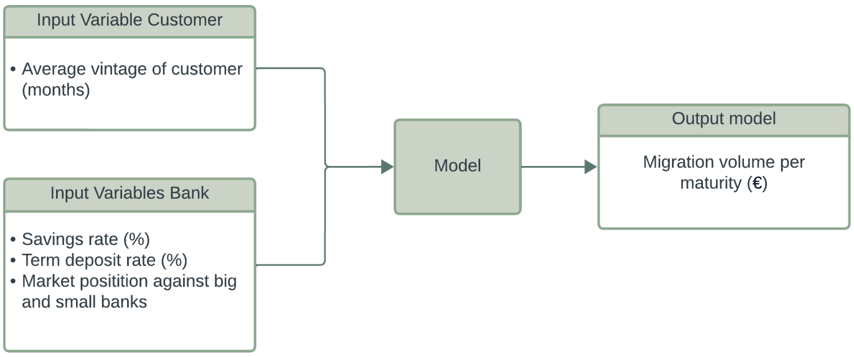

Figure 4.2: Migration model overview using individual migration volume

The goal of this methodology is to use historical data to predict the future migration of deposit volume and provide insight into what proportion of the total outflow of NN Bank comes from this migration. This gives the bank insight into which variables influence customer behavior and their decision to migrate their savings volume to term deposits. If accurately predicted, the bank can consider this information in its cash flow projections. This allows for a more precise estimation of whether funds leaving savings accounts will remain within the bank.

Besides these models we also use a VAR model for forecasting the market rates, deposit rates, deposit volumes, and ultimately, what we are researching in this thesis, the migration of deposit volume.

4.1.2 Performance measures

To test the performance of our models and to evaluate the relationship between our dependent and independent variables, we use various performance measures. To do so, we split our data up into a train and test set. The train set is the subset of the data used to train a model. The models learn to make predictions by fitting this data and understanding the relationship between the dependent variables and the independent variables. The test set is utilized to evaluate the performance of the trained model in generalizing its predictions to new, unseen data. There is no standard rule for determining the split point in a data set, but often a training set of 70/80% and a test set of 30/20% is used. In our research, we need to carefully determine this split point because of the increase in interest rates and the migration of deposit volume in the last part of our data. To evaluate the performance of the models, we categorize our measures into two types: goodness of fit models (in-sample) and forecast error measures (out-of-sample). For the goodness of fit, we consider the R^2, R^2_{Adj}, Root Mean Squared Error (RMSE), AIC, and BIC as performance measures. R^2 is a statistic that gives information about the goodness of fit of a model. It is a measure of how well the regression predictions approximate the real data points, as shown in equation (4.1) where $\hat{y}_i$ are the predicted values, y_i are the actual values and $\bar{y}$ is the mean of the observed data.

$$\mathrm{R}^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{4.1}$$

The R^2_{Adj} statistic extends the concept of R^2 to provide a measure that accounts for the number of predictors in a model. This adjustment is important when comparing models with different numbers of variables, as R^2 can artificially increase with the addition of more variables, regardless of their relevance to the prediction (Bhandari, 2024). The R^2_{Adj} compensates for this by penalizing the addition of irrelevant predictors, offering a more accurate indication of model performance across different models. An R^2_{Adj} close to 1 suggests that the model's predictions fit the data well while efficiently using a minimal number of predictors. The equation for R^2_{Adj} is shown in equation (4.2), where n is the number of observations and p is the number of predictors.

$$\mathrm{R}^2_{Adj} = 1 - \left(1 - R^2\right) \frac{n-1}{n-p-1} \tag{4.2}$$

This formulation adjusts the original R^2_{Adj} by taking into account the sample size and the number of predictors, making it a more reliable statistic for comparing the goodness of fit across different regression models. The RMSE is the square root of the average squared differences between the predicted values $\hat{y}_i$ and the actual values y_i. It measures the standard deviation of the residuals. RMSE is commonly used to evaluate the differences between values predicted by a model or an estimator and the values observed.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{4.3}$$

The AIC estimates the relative amount of information lost by a given model: the less information a model loses, the higher the quality of the model. It deals with the trade-off between the goodness of fit of the model and the complexity of the model. It is used for model selection, where a lower AIC suggests a better model.

$$\text{AIC} = 2k - 2\ln(L) \tag{4.4}$$

In equation (4.4), k is the number of estimated parameters and L is the maximized value of the likelihood function for the mode. Similar to AIC, the BIC measures the trade-off between model fit and complexity, but it penalizes models with more parameters more heavily than AIC. It is also used for model selection, with a preference for models with lower BIC values.

$$\text{BIC} = \ln(n)k - 2\ln(L) \tag{4.5}$$

On the other hand, for forecast error measures, we use the RMSE again and the Mean Absolute Error (MAE). The MAE is a measure used to quantify the accuracy of a model's predictions. It calculates the average magnitude of the errors between the predicted values and the actual values, without considering their direction. The MAE is particularly useful as it provides a straightforward interpretation of the average error magnitude. The formula for MAE is given by:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4.6}$$

These chosen performance measures offer insights into how well a model fits the observed data and facilitate comparing different models based on their goodness of fit and the complexity of the model, as well as their ability to minimize forecast errors.

Additionally, we examine the coefficients of the models to understand the impact of each variable. By calculating the p-values for these coefficients, we can determine the statistical significance of each variable and measure how well they explain the dependent variable. The p-values associated with each coefficient test the null hypothesis that the coefficient is equal to zero (no effect). A small p-value below 0.05 suggests that there is strong evidence against the null hypothesis, indicating that the variable has a statistically significant impact on the dependent variable. By comparing the p-values and the coefficients, we can determine which variables contribute the most to explaining the variation in the dependent variable.

4.1.3 Fitting and forecasting

To estimate the model's coefficients we use the OLS method. This is a common technique for estimating the coefficients of linear regression equations which is done by minimizing the the sum of square differences between the observed and predicted values. The models themselves are fit to actual historical data to avoid additional errors in our estimates. Furthermore, we check all our models for multicollinearity, which happens when the independent variables in a model are too similar to each other. This issue can make it hard to understand how each variable affects the outcome. If we find multicollinearity, we note it down and may adjust our models by removing variables to keep our findings accurate and reliable.

Our forecasting approach utilizes dynamic forecasting. This involves using predicted values as autoregressive inputs for subsequent predictions. We implement a one-step-ahead forecasting strategy without refitting the model as new data points become available. This method helps avoid overfitting, which is a concern when recent trends overly influence the model's parameters. To evaluate the validity and performance of our models, we apply the goodness of fit tests described in Section 4.1.2 to the in-sample data. Furthermore, we assess the forecast performance on out-of-sample data to guide our model selection process.

4.2 Model framework

This section describes the models of market rates, deposit rates, and deposit volumes that will be used in this research for the proposed framework. The relationship between the models will be mentioned in each subsection, as also shown in Figure 4.1.

4.2.1 Market rates

In this research, we use the Dynamic Nelson-Siegel model to model the market rate given in equation (2.5) because of its ability to capture market dynamics better by using multiple yield curves. The predicted yields help us understand how changing market rates affect the rates offered on savings and term deposits as they are used in our deposit rate models shown in Figure 4.1. Furthermore, it illustrates more accurately how interest rates for short, medium, and long terms move and relate to each other over time. The DNS model represents the yield curve at any point in time using a function of three factors, which correspond to the level, slope, and curvature of the yield curve. These factors are modeled as functions of time, allowing the DNS model to describe changes in the yield curve over time effectively. To use the model, we focus on determining the parameters, denoted as βs, at time t in equation 2.6). These parameters are important as they represent the aforementioned latent factors (level, slope, and curvature) at a specific point in time. The estimation of these parameters is performed through regression analysis, where the objective is to minimize the error term ϵ_t. The OLS method is used for this purpose.

Once we have predicted all β_i for each t using the OLS method, we need to predict the future. To do this, we use a VAR model since the βs are a vector of three values at each t. We look at the AIC and BIC scores to determine the lag order of the VAR model. Using the future βs we can use the equation (2.5) and calculate the yield curve for each time step t. The yield curves can be used to create a market rate path for each maturity τ. To achieve this, we extract the rate for each maturity from the yield curve at a specific time point t. Then, we sequentially append these rates. This process results in a continuous path for market rates that reflects the progression of yields over time for each maturity. The calculated yields are used as an explanatory variable for the savings rate and term deposit rate models in the following section.

4.2.2 Deposit rates

In this section, we discuss how we model and predict the savings rate and term deposit rate of NN Bank. Both rates are used as explanatory variables in our deposit volume model and migration model.

4.2.2.1 Savings accounts

In our literature review on how to predict deposit rates, we have found that these rates often change in line with market rates, which affect how much banks pay for funds. When market rates go up, banks might increase deposit rates to attract funds at a lower cost than borrowing funds from other financial institutions. A deposit rate model from a previous study conducted at NN Bank by Van der Ploeg (2023) is utilized. To determine if this model, as shown in equation (2.12) and model 'FD-t' in Table 4.1, performs well after the addition of six months of new data on deposit rates, we test and propose variations that may also have a good fit on the last six months of historical data. To accomplish this, we use historical data on deposit and market rates, including a three- and six-month average of market rates. We use the performance measures described in Section 4.1.2 to determine the performance of each model. To make predictions using these savings rate models, we use the predicted yields of the Euribor 6M from the DNS model represented by MA_t or r_{t-1} in Table 4.1. Besides the proposed model, we examine variations of the deposit rate model in Table 4.1. These variations are based on the expert opinion of employees within NN Bank and Zanders by using the knowledge of market practice at Dutch banks.

Table 4.1: Savings rate models

Description	Model
FD - t	$\Delta d_t = \beta_0 + \beta_1 \Delta d_{t-1} + \beta_2 \Delta MA_t + \beta_3 t + \epsilon_t$
FD	$\Delta d_t = \beta_0 + \beta_1 \Delta d_{t-1} + \beta_2 \Delta MA_t + \epsilon_t$
FD - Diff	$\Delta d_t = \beta_0 + \beta_1 \Delta d_{t-1} + \beta_2 \Delta MA_t + \beta_3 (d_{t-1} - r_{t-1}) + \epsilon_t$
3m MA	$d_t = \beta_0 + \beta_1 d_{t-1} + \beta_2 MA_t + \epsilon_t$
6m MA	$d_t = \beta_0 + \beta_1 d_{t-1} + \beta_2 MA_t + \epsilon_t$

To better capture the dynamics in the current interest rate environment, we exclude the time trend from the original model proposed by Van der Ploeg (2023) to test the influence of this variable. The inclusion of the term $(d_{t-1} - r_{t-1})$ in the model described in Table 4.1 serves to capture the lagged difference between deposit and market rates. This term is important for understanding how changes in market rates might influence deposit rates over time. By examining the difference between these rates from one period to the previous one, the model can determine the responsiveness or adjustment of deposit rates to fluctuations in market rates. This lagged difference helps in analyzing how quickly and to what extent the deposit rates adjust to changes in the market environment. The sign and magnitude of the variable provide insights into the competitive positioning of deposit rates relative to market rates. These dynamics are important to understanding the relationship between deposit and market rates and their impact on savings behavior and bank funding strategies. If deposit rates significantly exceed market rates, banks may lower deposit rates to reduce the cost of funds, especially if the market rates continue to remain low or decrease.

The models '3m MA' and '6m MA' step away from the first difference models and test the model on a moving average of the market rate for three and six months. We do this because first differences models are sensitive to short-term fluctuations and these models might be better at capturing the long-term relationship. Moving averages smooth out short-term volatility and random noise in the data, providing a clearer view of the underlying trend of the market rate. Furthermore, incorporating the moving average of the market rate helps in capturing the general trend of the financial market conditions. Savings rates do not always adjust immediately to changes in market rates due to factors like the policy of the financial intuition and its risk management causing stickiness of the savings rate (O'Brien, 2000).

In Chapter 5, we discuss which model variation has the best in-sample and out-of-sample performance. Based on both performance measures, a decision is made regarding which model to use for predicting the savings rate.

4.2.2.2 Term deposits

This section outlines the methodology for modeling and forecasting term deposit rates. The reason for additionally modeling the term deposit rate is that we are interested in the spread between savings rates and term deposit rates as we want to use this as an explanatory variable in our volume models. This is an incentive for customers to migrate their money from their savings accounts to term deposits. The model introduced by Avsar and Ruimy (2021) in equation (2.11) integrates the Libor 6M as a key variable. The choice of Libor 6M, similar to the Euribor, is based on the understanding that term deposit rates are influenced not only by the overnight market rate but also by the term structure of interest rates. This reflects the longer maturity periods associated with term deposits.

To effectively utilize this model for forecasting, we estimate its parameters with the OLS method using the historic average term deposit rates and the relevant market rates, as detailed in Section 3.2. Instead of using the Libor 6M, we use the Euribor 6M rate, which is used by NN Bank as a reference rate. As seen in Chapter 3, the deposit rates across different maturities closely align with the Euribor 6M rates. Therefore, we fit the parameters of the model for each term deposit rate based on the historical Euribor 6M rate. As a result, we end up with ten fitted models that we can independently use to forecast the deposit rate of each term deposit maturity. These rates are one of the explanatory variables for our forecast of deposit volumes.

4.2.3 Deposit volume

We propose to model the volume of savings accounts and term deposits separately. The volumes are based on different product characteristics. Customers can always withdraw their money from savings accounts, but not from term deposits, except with a penalty. The monthly volume of savings accounts described in Subsection 4.2.3.1 serves as an explanatory variable in our outflow model, as described in Subsection 4.2.3.2.

4.2.3.1 Savings accounts

For this research, we use the model by Paraschiv and Schürle (2010), given in equation (2.14) and also represented in Table 4.2 as 'Model 1', because the model incorporates market and deposit rates. Since we are interested in the effect of these rates, the model is used as a basis. To see the effect of different combinations of variables on the performance and fit of the model on the historical data, we propose to make modifications to the original model. We propose changes to the model based on new data and insights from industry experts. These changes aim to improve the model's fit and predictive accuracy by incorporating updated economic conditions and potentially relevant new variables. This ensures our model remains robust and reflective of the latest trends in financial markets. The use of the variable t is controversial because we see that the volume of NN Bank does not just grow. Additionally, we separate the lagged deposit volume variable from the interest rate distribution and include it as a separate term. We can deduce from Figure 3.9 that there are signs of seasonality in the outflow of savings accounts. It makes sense to include this as an independent variable in the volume model of savings accounts. For modeling the volume of savings accounts, we propose the following models:

Table 4.2: Volume models for total savings account volume

	Deposit Volume Model
1	$D_t = \beta_0 + \beta_1 t + \beta_2 D_{t-1}[d_{t-3} - (\delta r_{t-3}^{short} + (1-\delta) r_{t-3}^{long})] + \epsilon_t$
2	$D_t = \beta_0 + \beta_1 D_{t-1}[d_{t-3} - (\delta r_{t-3}^{short} + (1-\delta) r_{t-3}^{long})] + \epsilon_t$
3	$D_t = \beta_0 + \beta_1 D_{t-1}[d_{t-3} - (\delta r_{t-3}^{short} + (1-\delta) r_{t-3}^{long})] + \beta_2 S_{12,t} + \epsilon_t$
4	$D_t = \beta_0 + \beta_1 D_{t-1} + \beta_2 (d_{t-1}^s - r_{t-1}) + \beta_3 S_{12,t} + \epsilon_t$
5	$D_t = \beta_0 + \beta_1 D_{t-1} + \beta_2 (d_{t-1}^{td} - d_{t-1}^s) + \beta_3 S_{12,t} + \epsilon_t$
6	$D_t = \beta_0 + \beta_1 D_{t-1} + \beta_2 d_{t-1}^s + \epsilon_t$

We chose to remove the variable linear time trend t in the second model in Table 4.2 to see the effect of t on the performance of the model. The third model includes a binary dummy variable to account for seasonal patterns in savings account deposit volume. This variable is set to one in December to capture the typical decrease in deposits during this month and is zero for all other months. In the fourth and fifth models, we model the difference between the deposit rate and the market rate, and the difference between the term deposit rate and the savings rate, respectively. In this way, we want to include the interest rate incentive in the model separately from the lagged deposit volume. For r^{short} and r^{long} we use the Euribor 1M and 6M respectively. The sixth model includes the lagged deposit rate of savings accounts as an independent variable. Additionally, since we know when term deposits will mature in the future, we can add this volume to our predictions in the forecast.

In Chapter 5, we discuss which model variation has the best in-sample and out-of-sample performance. Based on both performance measures, a decision is made regarding which model to use for predicting the savings volume. The predictions of the savings volume are used as an explanatory variable in our outflow model.

4.2.3.2 Outflow model

To give a more comprehensive overview of how much the migration of deposit volume contributes to the total outflow from the savings accounts of NN Bank, we propose using the following outflow model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 D_{t-1} + \beta_3 S_{12,t} + \epsilon_t \tag{4.7}$$

where Y_t represents the monthly outflow from the savings accounts for a given month t. The variable D_{t-1} indicates the previous month's deposit volume of these savings accounts, serving as a measure of lagged deposit volume. The term $S_{12,t}$ is a binary variable that captures the effect of December's unique seasonal impact on savings behavior, being set to one in December and zero in other months. By incorporating the forecasted deposit volume of savings accounts as an input, our model not only reflects the inflows but also assesses how these inflows affect subsequent outflows. The model thereby considers the dynamic relationship between the savings account's total volume and its outflow patterns. In addition, this approach enables us to account for seasonal variations that significantly affect deposit behaviors, as shown in Figure 3.9. By including the $S_{12,t}$ variable, our model can account for the increased outflow usually seen in December. This ensures that our forecasts accurately reflect these yearly fluctuations.

4.2.4 Migration flows

At this point in our methodology, we have covered the models for interest rates and deposit volumes. Our predictions for savings and term deposit rates are used as input for our migration model, which we discuss in this section. The predictions from our outflow model, based on our savings account deposit volume, provide context for our migration volume predictions.

For modeling the migration of deposit volume, we use the model proposed by Gerritsen and Bikker (2020) which was originally used to capture the proportion of savings that a depositor transfers between banks. However, to capture the proportion of savings that transfer from savings accounts to term deposits we adjust the formula accordingly. By examining the proportions within the data, it is possible to determine which factors are driving customer behavior. This analysis provides significant insight into the allocation of their savings by identifying which variables are most indicative of how and where customers choose to allocate their savings. Furthermore, the original formula uses the inverse Mills ratio which is used to correct for selection bias. The inverse Mills ratio is typically included in regression models as a way to correct selection bias, particularly in situations where the sample used in the analysis is not randomly selected from the population, leading to potential biases in the estimated coefficients. However, in our research, we are researching all the customers of NN Bank and not a sample. Together with the data needed to estimate the inverse Mills ratio, which is not available in our data set, we decided to remove this term from our formula.

We can identify three distinct cash flow streams for NN Bank to provide a comprehensive overview of their savings and term deposit volume dynamics:

- Term deposit opened by a new client
- Term deposit opened by an existing client with old money
- Term deposit opened by an existing client with new money

NN Bank is interested in the second cash flow, described in Chapter 3 as 'old money' because additional interest has to be paid on savings volume that was already present in the bank but migrated from savings accounts towards term deposits. However, understanding the factors that motivate customers to transfer their savings and indicating the percentage of their savings that migrate can be extremely valuable in NN Bank's risk management and profitability.

We model the transfer proportion as the proportion of migrating old money in the total outflow from savings accounts in a month for NN Bank for the three maturity groups described earlier: 1-year, 2-year, and 3+ year. The model's equation, as shown in Figure 4.2, is given equation (4.8).

$$\text{Migration}_{i,t} = \beta_0 + \beta_1 R_{i,t} + \beta_2 V_{i,t} + \beta_3 C_{i,t} + \epsilon_t \tag{4.8}$$

where for maturity group i and month t:

- $R_{i,t}$ is the spread between savings rate and term deposit rate / term deposit rate
- $V_{i,t}$ is the average vintage of customers that opened a term deposit
- $C_{i,t}$ is the competitive position of NN Bank compared to Dutch banks

The spread between the savings rate and the term deposit rate is included as this is the incentive for customers to transfer their savings. To test whether the incentive for the customers is the spread between the two rates or whether it is the level of the term deposit rates itself, we also test the term deposit rate itself as an independent variable. Another independent variable that we obtain from our dataset is the average account vintage of a set of customers. The average vintage could indicate customer loyalty or confidence in the bank. Customers with longer account histories might be more inclined to open term deposits, suggesting a relationship between account age and customer trust or satisfaction. If we find a significant relationship between account vintage and term deposit openings, this metric can serve as a predictive indicator. It could help in forecasting future term deposit volumes based on the current composition of account vintages. Finally, we include an independent binary variable representing whether the term deposit rate is among the top three rates offered by competing banks, with this term deposit rate corresponding to the transfer proportion of the specific maturity that is being analyzed.

Using this method, we get a clear overview of which variables drive customer behavior when migrating their savings to a term deposit and which portion of NN Bank's customers are most likely to migrate their money. We also segment the different proportions in terms of maturities, which gives NN Bank more insight into how to hedge against this migration.

4.3 Vector Autoregressive model

In the previous sections of this chapter, we presented various models that ultimately led to the migration model. However, due to non-stationarity issues in our time series, we also suggest testing whether a VAR model could be used to achieve stationarity and make accurate predictions. To make the best use of a VAR model, we must ensure that we are using the correct time series based on their causality and autocorrelation. We also need to determine the number of lags, which can be done by looking at the scores of the AIC and BIC on the fitted VAR model. The time series we are going to look at consists of the following list:

- Market rate (Euribor 6M)
- Savings rate of NN Bank
- Weighted average term deposit rate of NN Bank
- Total monthly savings volume

- Total monthly term deposit volume
- Total monthly migration volume

These time series are chosen because there is a correlation between them, as calculated in Section 3.3. The Granger Causality test is used to determine if one time series can predict another. This hypothesis test determines which time series have a significant impact on each other and can be used in our VAR model. A p-value less than 0.05 rejects the null hypothesis that there is no causality between two time series at a 95% confidence level. For a VAR model to work, we need to make sure that the time series are stationary, as this is one of the requirements for using a VAR model. We use the ADF and KPSS tests as described in Section 3.4 to test for stationarity. To test whether the different time series are not autocorrelated, we use the Durbin Watson statistic. Autocorrelation in time series data can weaken the reliability of statistical tests, reduce forecasting accuracy, and lead to incorrect model conclusions. The statistic tests for autocorrelation in the residuals of a regression analysis at lag one. A value of about two tells us that there is no autocorrelation between the time series. Once we have determined the optimal setup for our VAR model, we can test its performance on our test set based on the performance measures described in Section 4.1.2.

4.4 Scenario analysis

We aim to assess the outcomes of our model framework and VAR model under different interest rate scenarios, as proposed in sub-question 5c, through scenario analysis, focusing on understanding the impact of various market conditions and deposit rate changes on the results. This approach is important for effective risk management in banking and can be seen as a form of stress testing. Given the uncertainties in predicting future interest rate trends, our analysis explores different rate scenarios to determine potential impacts on NN Bank. Understanding how customers might change their saving behavior with different deposit rate scenarios is important. Looking at various interest rate situations can help NN Bank get ready for possible shifts in the migration of deposit volume towards term deposits. This knowledge is important for planning future changes to what banking products are offered and how interest rates are set to keep customers interested. It also affects how the bank decides on its pricing strategy. By doing this, NN Bank can better understand how changes in interest rates could affect the stability of its deposits, which helps the bank manage its risk related to interest rates more effectively.

NN Bank gains insights from scenario analyses of interest rates. However, due to stationarity issues and estimation errors in our prediction models, our interest rate forecasts can become inaccurate. To address this, we propose comparing our forecasts with different interest rate scenarios based on historical data.

Our scenario analysis consists of six supervisory shock scenarios described by the European Banking Authority (EBA) that are used to assess the resilience of banks under adverse and extreme conditions. The six scenario shocks proposed by the EBA cover various economic and financial conditions, including interest rate changes and market stress scenarios. The scenarios are (EBA, 2022):

(a) Parallel shock up, where there is a parallel upward shift of the yield curve with the same positive interest rate shock for all maturities;

(b) Parallel shock down, where there is a parallel downward shift of the yield curve with the same negative interest rate shock for all maturities;

(c) Steepener shock, where there is a steepening shift of the yield curve, with negative interest rate shocks for shorter maturities and positive interest rate shocks for longer maturities;

(d) Flattener shock, where there is a flattening shift of the yield curve, with positive interest rate shocks for shorter maturities and negative interest rate shocks for longer maturities;

(e) Short rates shock up, with larger positive interest rate shocks for shorter maturities to converge with the baseline for longer maturities; and

(f) Short rates shock down, with larger negative interest rate shocks for shorter maturities to converge with the baseline for longer maturities.

In our study, we employ the same shock scenarios utilized by NN Bank in their scenario analysis for De Nederlandsche Bank (DNB). Our approach involves utilizing forecasts from our market rate model, which predicts future yield curves at each point in time t. For each scenario, we apply the corresponding shocks to the appropriate tenor of the yield curve at every t. As a result, we generate six different scenarios showcasing the various outcomes of our market rates under different conditions. The rates for term deposits are adjusted to reflect changes in the yield curve on a one-to-one basis. This means that any fluctuation in the yield curve is directly reflected in the rates offered for term deposits, with the rates adjusting based on the shocks applied to the matching tenors of the yield curve. We chose this method because there is a strong correlation between the market rate and the term deposit rate, as shown in the correlation matrix in Section 3.3.

We suggest another strategy for setting the savings rate of NN Bank in these interest rate scenarios. We recommend a 70% pass-through rate. This means that if market rates change, the savings rates will not match these changes exactly. Instead, if market rates go up or down by 1%, the savings rates only change by 0.7%. This approach shows that banks are careful not to adjust savings rates too quickly in response to market changes which can be deducted from the correlation matrix in Section 3.3. By doing this, banks aim to balance making a profit, keeping customers satisfied, and staying competitive. This 70% rule helps manage how savings rates react to the shocks of the market rate. The interest rates from these scenarios are used as inputs to the proposed migration forecasting methodology. The market and deposit rates of both products are used in the deposit volume models. In addition, both deposit rates are used in the transfer rate model.

4.5 Assumptions

When evaluating the results from our models, it is critical to be cautious because the statistical characteristics of our time series change over time, making our analysis more challenging. This non-stationarity means we must carefully assess our findings to ensure they accurately reflect potential trends. Additionally, our use of forecasts to predict how deposits might move introduces the possibility of errors in our predictions. Relying on forecasts means we are working with assumptions that might not always match up with what happens in the future. For our analysis of the total volume of term deposits, we are assuming that there will not be any early withdrawals. This assumption simplifies our calculations by not having to consider variations in depositor behavior, such as withdrawing funds sooner than expected. While this makes our model less complex, it is worth noting that this simplification might overlook some real-life behaviors.

4.6 Conclusion

In this chapter, we aimed to find an answer to sub-question 4, where we wanted to find a method for modeling and forecasting the migration of deposit volume, and where we could also include the interest rate incentive. Two methods were proposed to predict migration: a model framework where individual interest rates are modeled and used as input for the models of deposit volume of savings accounts, outflow, and migration, and a VAR model where different time series are used to forecast each other. The first method uses a DNS model for the market rate, a savings rate model from our model selection, and the term deposit rate model by Avsar and Ruimy (2021). Furthermore, the choice of model for determining the deposit volume of savings accounts is based on model selection. We also proposed an outflow model and discussed our model for predicting the migration volume per maturity group. The second method for predicting the migration of deposit volume is a VAR model based on relevant time series that can be used to predict the migration volume. The selection of the time series is discussed in Chapter 5. Lastly, it is important to consider that using predictions with uncertainty as input for other predictions can increase the overall prediction uncertainty, potentially leading to inaccuracies in the final models.

5 Results

This chapter aims to address sub-question 5, which involves reflecting on the factors, triggers, and motivations that lead customers to migrate their deposit volume. The chapter contains the results of our proposed model framework and VAR model from our methodology and we reflect on all the results. We validate whether our models have a good performance and try to understand what variables drive the motivation of customers to migrate their deposit volume towards term deposits. We decided to split our data into 80% for training and 20% for testing because this is when interest rates started to rise. In this way, our models can learn from this important change, otherwise we run the risk of underperforming.

For our model framework proposed in Section 4.1.1, we faced problems with model performance after transforming data for stationarity during our model fitting and analysis. We attempted various methods to tackle the non-stationarity issues, including second differencing, first differences of logarithmically transformed data, and removing a quadratic trend, but none of these methods led to reasonable results. Our models performed poorly on the performance measures using the transformed data. We could see the poor performance in a low R^2_{Adj} on our training set and high prediction errors for the RMSE and MAE on our test set. Looking at our data, we can see a structural break. This break was marked by a sudden increase in interest rates and an unexpected shift in deposit volumes, which our models had difficulty accommodating. This break can affect the level, trend, or variance of the series. To account for the structural break, we tried using a dummy variable. This variable was set to zero before the break and one after. However, we faced challenges since the break occurred at the end of our time series. Our models could not effectively learn the new dynamics of the data due to insufficient post-break data. All these factors make it difficult for our models to find meaningful patterns in the data and lead to the poor performance of our performance measures. Therefore, we have chosen not to transform the data to obtain stationarity. This choice implies that the results of our models may not be fully reliable. Our decision was influenced by the performance of our models with transformed data, which produced inaccurate results. However, to achieve some degree of accuracy in our model results and their forecasts, we decided to work with non-transformed data. This approach is expected to yield more reasonable results, allowing us to effectively address the sub-question 5 and ultimately our main research question.

5.1 Model framework

This section discusses the results of the different models of our analytical framework proposed in the previous chapter.

5.1.1 Market rates

We use the DNS model to fit and forecast market rates of different maturities which will be used as input for our savings rate and term deposit rate models. We used a train and test set of 86 and 21 observations, respectively. Our model provides βs at each t, which can be found in Table B.2 in Appendix B.1. We calculate the R^2_{Adj} and RMSE to check whether the estimated βs provide a good fit to the actual interest rate paths of each maturity, as shown in Table B.1 in Appendix B.1. The overall high R^2_{Adj} and low RMSE values indicate that the model fits well with the actual rates. The calculation of the AIC and BIC of the DNS model was not included, as this would require calculating the values for each time step

t of the train set. These calculations do not provide much information about the model's goodness of fit. Additionally, it is important to note that we are not comparing different models. Figure 5.1 below shows the out-of-sample predictions of our DNS model on our test set where we highlighted market rates for different maturities for clarity.

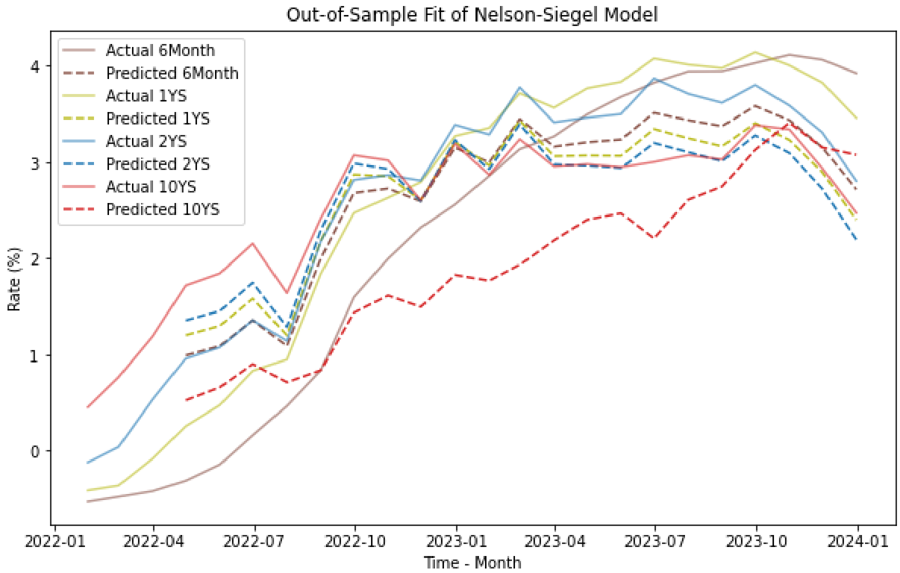

Figure 5.1: Actual market rates vs. fitted market rates

Now, to forecast the yield curves, we need to forecast the βs at each t. We use a VAR model fitted on our historical βs to forecast the βs and determine the best number of lags to use. The model has the best AIC score with nine lags, while the BIC has the lowest score with one lag. The AIC criterion is often used for determining the number of lags of a VAR model (Cromwell, 1994). Therefore, we use 9 lags to fit our VAR model. Our forecast covers the next 24 months. The VAR model's forecasted βs are also shown in Appendix B.1. We can use these βs as input for equation (2.5) and calculate the yield curve at time t. The yield curves can then be used to create the graph as shown in Figure 5.2.

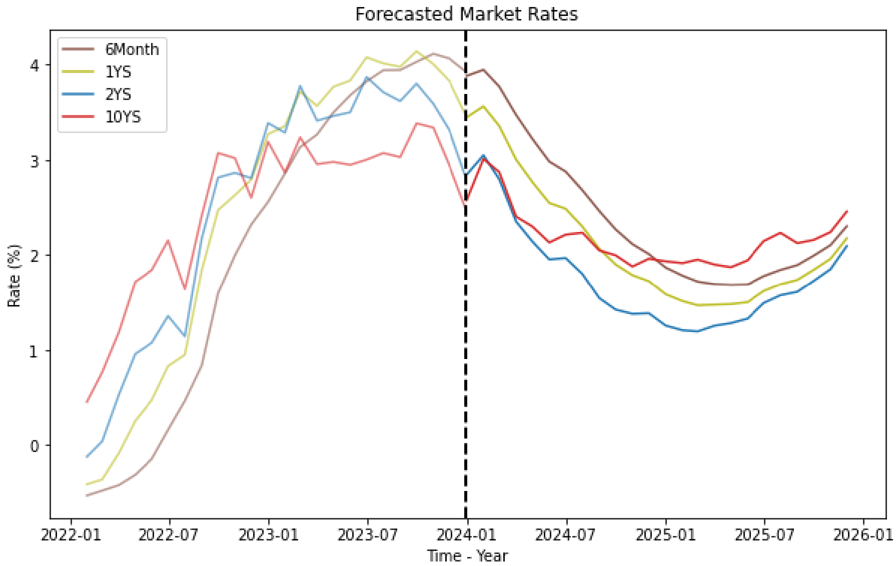

Figure 5.2: Out-of-sample forecast of market rates for a 2-year period

The figure shows that the model predicts a decrease in all market rates, followed by an increase in the second half of 2025. Additionally, an inverted yield curve is present at the beginning of 2024. An inverted yield curve is an unusual situation because it goes against the typical market condition where longer maturities offer higher yields to compensate investors for the increased risk over time. Our DNS model predicts that the yield curve will turn around again over the next 24 months which can be seen by higher interest rates for longer maturities. These forecasted rates are used as input for the remainder of our models.

5.1.2 Deposit rates

In this section, we compare and evaluate the performance of our proposed deposit rate models both in-sample and out-of-sample. Our train and test set consisted of 86 and 21 observations, respectively. We use the performance measures explained in Section 4.1.2 and conduct visual inspections. We use the performance measures for the model selection of our savings rate model. Additionally, we reflect on the predictions of the models 24 months in the future and their reliability based on past behavior of the deposit rates. For the prediction beyond our dataset, we use predictions of our market rates as input.

5.1.2.1 Savings accounts

We examine the results of our various deposit rate models for the savings rate of NN Bank that we proposed and select the best-performing model. We have fitted our models to the train set and calculated their predictions in the out-of-sample period. Table B.4 in Appendix B.2 contains the coefficients estimations and the results of the goodness of fit tests. The predictions and their performance of prediction errors can be found in Figure B.1 and Table B.5 in Appendix B.2, respectively. Based on the performance of the models both in-sample and out-of-sample, we conclude that 'Model 3m MA' best reflects NN Bank's savings rate given in equation (5.1).

$$d_t = \beta_0 + \beta_1 d_{t-1} + \beta_2 MA_{3,t} + \epsilon_t \quad (5.1)$$

The estimation of the coefficients alongside the in-sample performance measures is given in Table 5.1.

Table 5.1: In-sample coefficients and performance measures of Model 3m MA

	Constant	d_{t-1}	MA_t
Model 3m MA	0.0217 (0.255)	0.9415 (0.000)*	0.0561 (0.201)
R^2_{Adj}		0.995	
AIC		-348.9	
BIC		-341.6	

Note: * next to p-values indicate significance.

The table shows a high R^2_{Adj} for the model, which may indicate non-stationarity in the data. Therefore, it is important to handle the model's predictions with care and keep the prediction uncertainty in mind. Furthermore, the out-of-sample predictions of the model in Figure 5.3 are compared to the actual savings rate of NN Bank. We can see that the model can reproduce the actual savings rate of NN Bank for the most part.

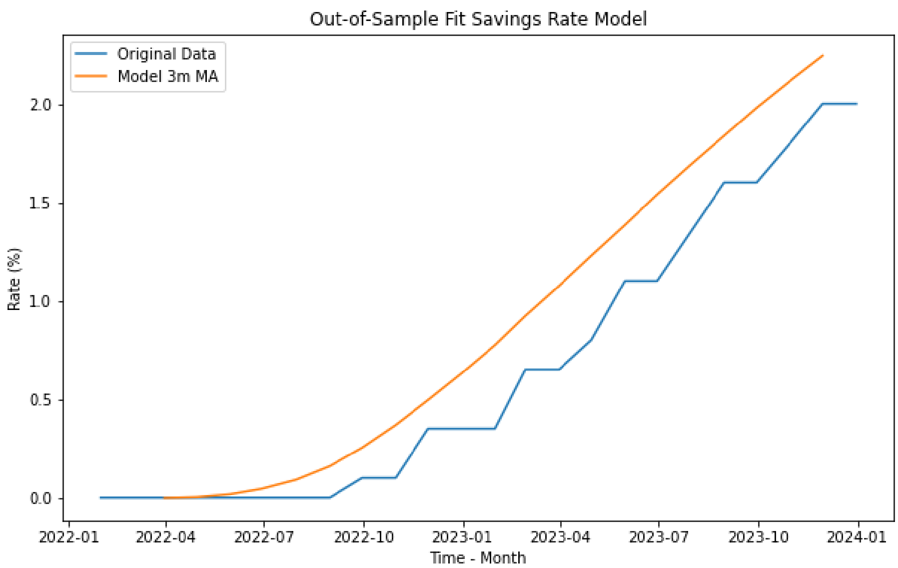

Figure 5.3: Actual savings rate vs. fitted rate

Using the coefficients of our fitted model on our data and the forecasted market rates of Section 5.1.1 as input, we forecast the savings rate using our selected model. We use the predicted Euribor 6M as input for our savings rate model since this is used by NN Bank as the reference rate for determining their savings rate. Figure 5.4 below shows the prediction of our savings rate model. We can see that the model predicts that the savings rate will continue to rise until the end of 2024. This is not very intuitive in the absence of an increase in market interest rates, since there is little incentive for banks to increase their savings rates. However, the predictions can be used as a reasonable input for our other models, but with a degree of uncertainty.

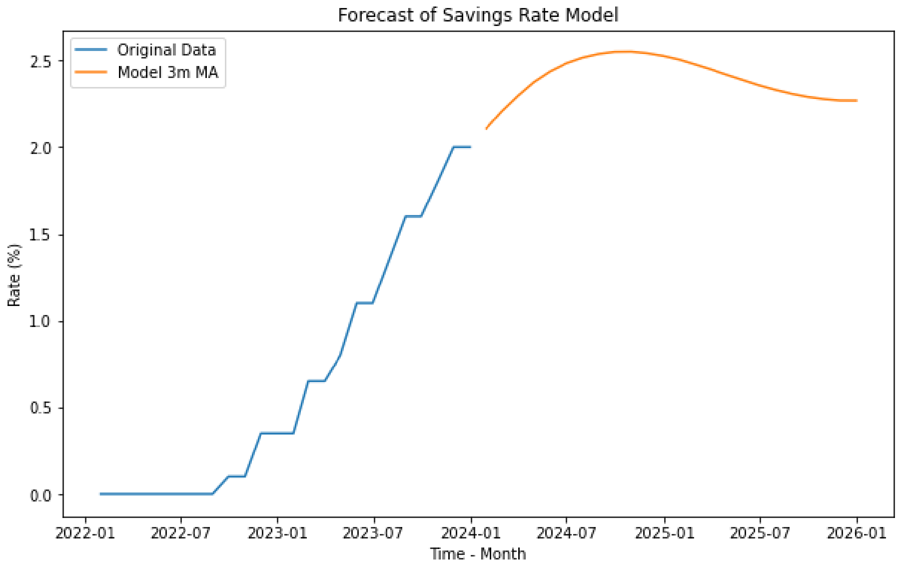

Figure 5.4: Out-of-Sample forecast of savings rate over a 2-year period

The predictions of this model are used as an independent variable for our deposit volume model and migration volume models.

5.1.2.2 Term deposits

As explained in Chapter 4, we use a term deposit rate model that incorporates the Euribor 6M for each term deposit rate given in equation (2.11). From December 2020 until 2022 it was not possible to take a term deposit with a maturity of one to five years. That causes the problem that there are no term deposit rates available in this period to fit our model on. We have fitted the model for the different maturities on the train set and predicted on the test set as can be seen in Figure 5.5 where we highlighted three different maturities for clarity. The coefficients of each maturity were estimated using the OLS method. The out-of-sample predictions for all maturities can be found in Figure B.2 in Appendix B.3.

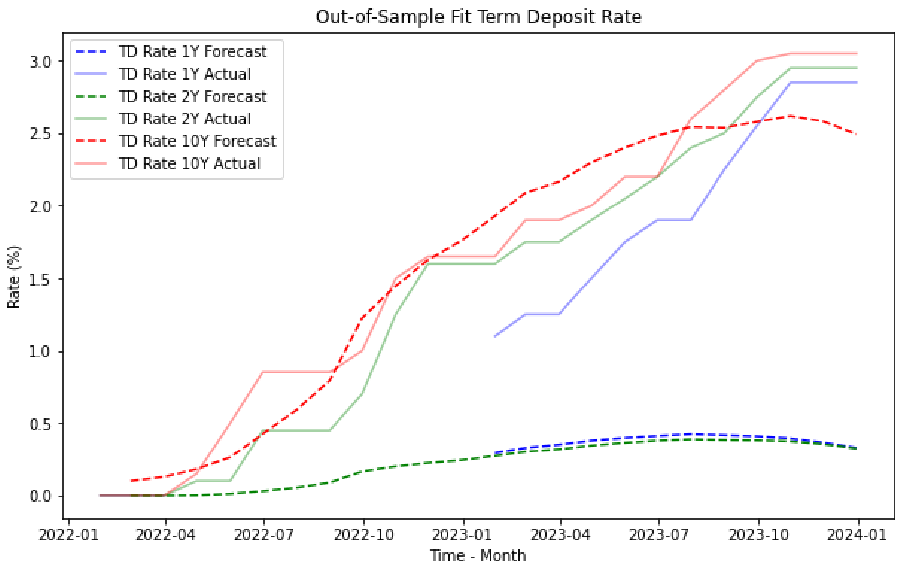

Figure 5.5: Actual term deposit rates vs. fitted rates

The model demonstrates a more accurate fit for longer maturity periods. This improved fit is caused by the historically higher rates associated with longer maturities, resulting in an increased β_3 coefficient. Additionally, a comparison of coefficients, detailed in Appendix B.3, between longer and shorter maturity rates reveals that β_1 and β_2 are lower for shorter maturities. This difference explains why predictions of term deposit rates are lower under rising Euribor 6M rates for shorter maturities.

Longer maturities tend to be more sensitive to changes in the economic outlook, as reflected by a higher β_3 coefficient. This is because long-term rates are more influenced by investors' future market expectations, which affect the term premium. The term premium is the extra return that investors demand to lock in their money for longer periods. Because long-term rates have historically been higher to compensate for increased risk over time, the model fits these rates better. The deposit rate for maturities of one, two, and ten years was forecasted using the fitted models, as shown in Figure 5.6 for clarity. The predictions for all maturities can be found in Figure B.3 in Appendix B.3. The predicted Euribor 6M rate was used as input for the models. All maturities show a decreasing trend in term deposit rates over the forecasted period. This is caused by the decreasing Euribor 6M. The figure also shows that longer-term maturities have higher rates than shorter-term ones. This difference in rates reflects the term premium. Investors usually demand higher rates for longer commitments because of the increased risk and opportunity cost of locking in funds for an extended period.

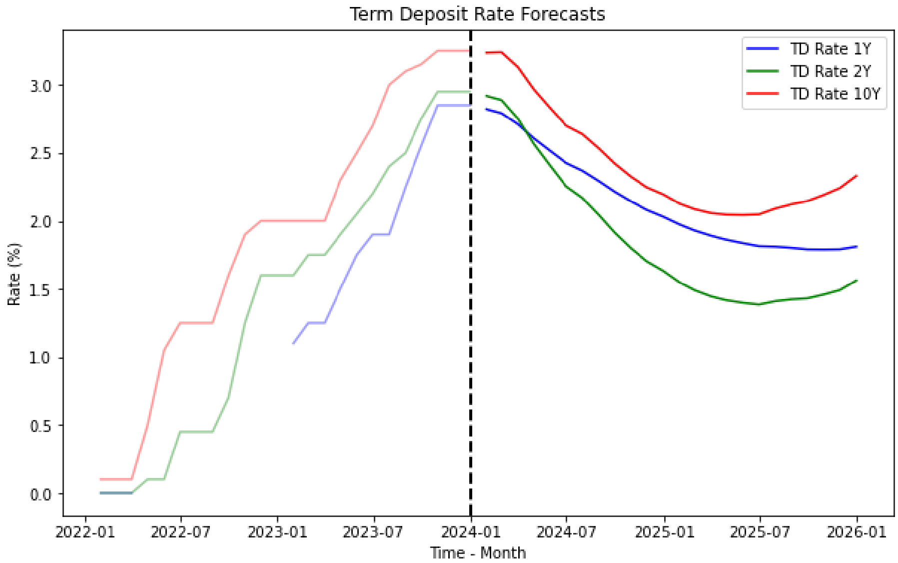

Figure 5.6: Out-of-sample forecast of term deposit rates

The predictions of the term deposit rates are used as an independent variable for our deposit volume model and migration volume models. However, the predicted rates include uncertainty because of the non-stationary data of the term deposit rates.

5.1.3 Deposit volume models

This section discusses the results of the deposit volume model for savings accounts and the outflow model. Various methods were researched to determine how well each model predicts the amount of money deposited. The models' predictions were compared to actual figures using performance measures. The best-performing model for the deposit volume of savings accounts is selected and used to predict the volume.

5.1.3.1 Savings accounts

The different deposit volume models for the savings accounts proposed in Table 4.2 require the estimation of the different coefficients. The OLS method was used to estimate the coefficients on our train set that consisted of 86 observations. The outcome of the calibration of the deposit volume models is given in Table B.8 in Appendix B.4. Furthermore, we used the estimated coefficients of the models to make predictions on our test set which consisted of 21 observations. The results of these predictions can be found in Figure B.4 in Appendix B.4. Based on the goodness of fit on our train set and the performance on our test set, we conclude that 'Model 5' can reproduce the monthly deposit volume of NN Bank's savings accounts the best given in equation (5.2). The coefficients and performance measures of the model can be found in Table 5.2 and the out-of-sample forecast is illustrated in Figure 5.7.

$$D_t = \beta_0 + \beta_1 D_{t-1} + \beta_2 (d^{td}_{t-1} - d^{s}_{t-1}) + \beta_3 S_{12,t} + \epsilon_t \tag{5.2}$$

The model is based on the lagged deposit volume, the spread between the savings rate and the weighted average term deposit rate, and a seasonality variable. Once again, we stumble across the non-stationarity issues of our data which can be deducted from the high R^2_{Adj}.

Table 5.2: In-sample coefficients and performance measures of Model 5

	Constant	D_{t-1}	$(d^{td}_{t-1} - d^{s}_{t-1})$	$S_{12,t}$
Model 5	1.989e+08 (0.004)*	0.9717 (0.000)*	-9.354e+07 (0.015)*	-1.542e+08 (0.000)*
R^2_{Adj}	0.985			
AIC	3189.			
BIC	3199.			

Note: * next to p-values indicate significance.

The model's out-of-sample predictions indicate a significant decrease in deposit volume. This decrease is likely due to a wider spread between the term deposit rate and the savings rate in our test set, resulting in a lower volume due to the negative coefficient of the spread variable.

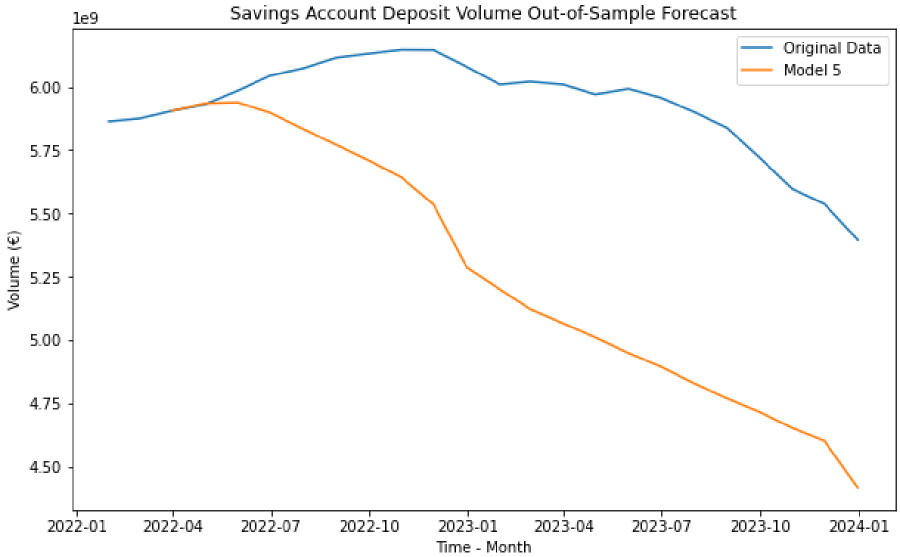

Figure 5.7: Actual deposit volume of savings accounts vs fitted volume

Figure 5.8 displays the out-of-sample predictions of our selected model for a 2-year period. The model utilizes dynamic forecasting, incorporating the previously predicted volume as a lagged variable. The model uses the predicted savings rate and term deposit rate as input for the independent variable representing the spread. It is important to note that both rates already include forecast uncertainty, and the predictions of the deposit volume also contain this uncertainty. As described in the methodology chapter, we adjust the predicted monthly volume by adding the maturing principal of the term deposits in the upcoming months, which we have obtained from our dataset, as this is a significant amount. There is a predicted increase in deposit volume over the next two years. This prediction is based on the historical increase in NN Bank's deposit volume. It is important to note that the maturing term deposit volume contributes significantly to the inflow.

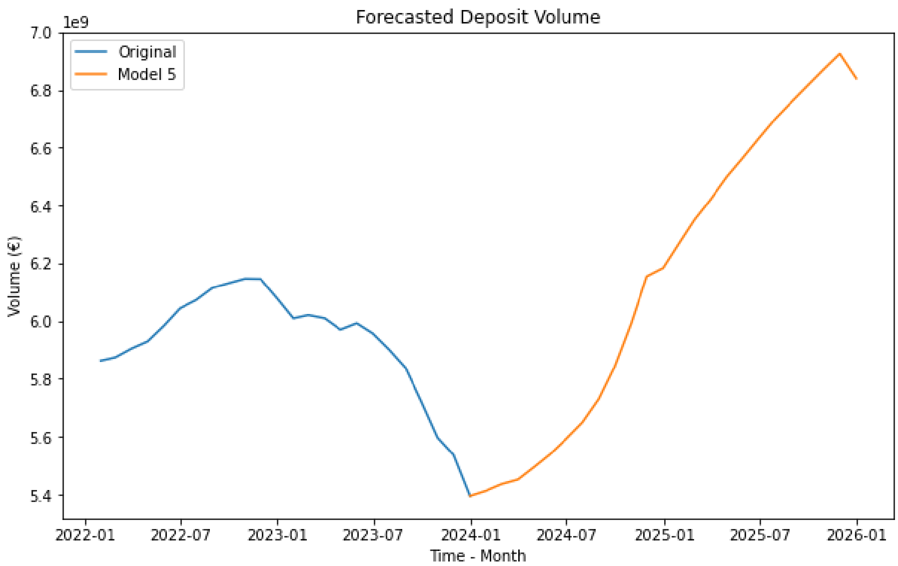

Figure 5.8: Out-of-sample forecast of deposit volume of savings account for a 2-year period

We now have the predictions of the deposit volume of the savings accounts where we used our predicted savings rate and term deposit rate as input. The predictions of the deposit volume will serve as an independent variable for our outflow model.

5.1.3.2 Outflow model

In this section, we discuss the results of our proposed outflow volume model framework given in equation (4.7) and its out-of-sample fit as well as the predictions for the future. The model's coefficients were estimated using the OLS method on the train set consisting of 86 observations. Table 5.3 contains the values of our coefficients and performance measures.

Table 5.3: In-sample coefficients and performance measures of outflow model

	Constant	Y_{t-1}	D_{t-1}	$S_{12,t}$
Outflow model	4.497e+07 (0.457)	0.2455 (0.003)*	-0.0229 (0.045)*	-1.415e+08 (0.000)*
R^2_{Adj}		0.549		
AIC		2838.		
BIC		2848.		

Note: * next to p-values indicate significance.

The R^2_{Adj} value in Table 5.3 indicates that there is still some unexplained variability in the data that is not accounted for by our variables. We attempted to improve the model by testing additional independent variables, such as the deposit rate and a Y_{t-12} value, but this resulted in worse performance. The randomness in the data may be partially attributed to the sudden increase in outflow during the first half of 2021, which makes it difficult for the model to fit the data accurately. Figure 5.9 displays the predicted outflow of savings accounts on our test set, which consists of 26 observations.

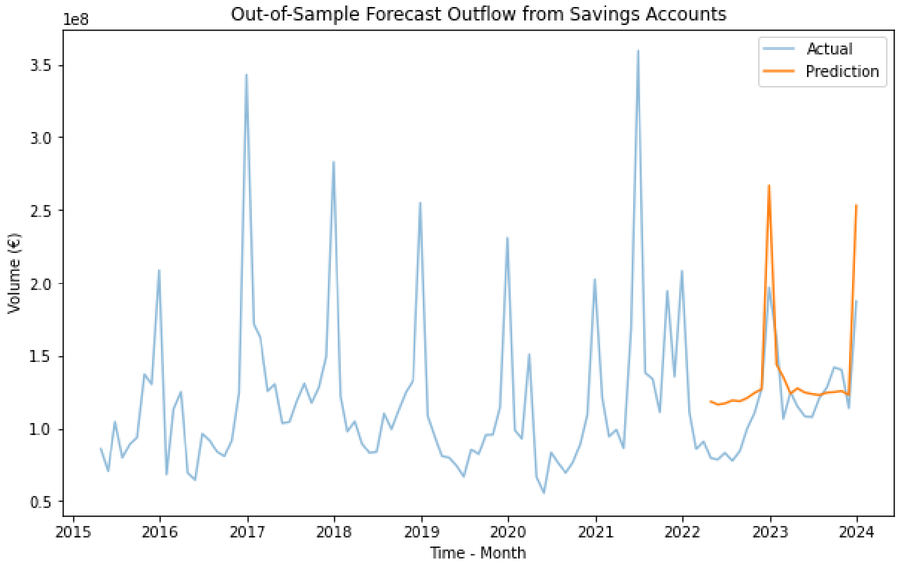

Figure 5.9: Actual outflow volume vs. fitted volume

The outflow model predicts an excessive outflow in the first few months of our test set. Additionally, the outflow in December is slightly overshot. During December 2023 and December 2024, the outflow remains relatively constant until the larger outflow of December 2024. The table shows that all coefficients have a p-value below 0.05, except for the constant. Figure 5.10 displays NN Bank's projected outflow for 2025 and 2026. To create this forecast, we used lagged outflow predictions, a binary variable, and the predicted deposit volume from Section 5.1.3.1 as an explanatory variable. The model predicts a consistent outflow throughout the year, except for December. While this pattern has not been observed historically, the constant outflow line represents the average outflow during all the months except December. Upon visual inspection of our dataset, these predictions appear reasonable.

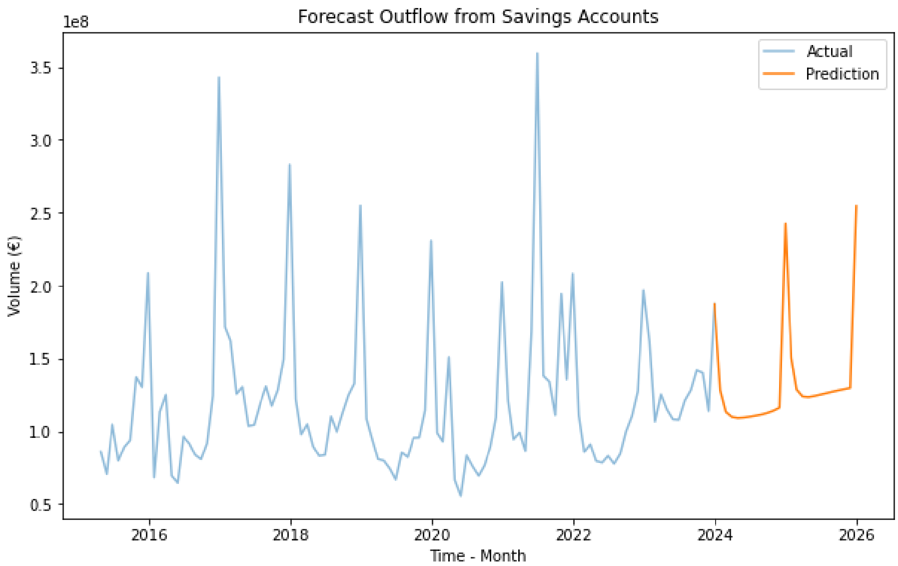

Figure 5.10: Out-of-sample forecast of outflow model for a 2-year period

The predictions of our outflow model are used to provide more insight into the proportion of the migration volume, which is calculated in the following section, compared to the total outflow.

5.1.4 Migration volume

In this section, we reflect on the results of our migration volume model given in equation (4.8) and hereby aim to answer sub-question 5b. It is very insightful to see what the incentive is for customers to migrate their savings. As explained in Section 4.1.1, the independent variables consist of either the spread between NN Bank's savings rate or the term deposit rate individually, the average vintage of the customer's savings accounts, and whether the NN Bank's term deposit rate was competitive with large and small banks in the Netherlands. We highlight the results of the largest migration volumes in our data, grouped into maturities of 1-year, 2-year, and 3+ year. We compared the competitiveness of NN Bank's term deposit rates with those of the small banks in the Netherlands, as this is where NN Bank focuses when setting its rates. In addition, the term deposit rate has always been in the top 3 compared to the big banks, so this does not add any valuable information to the model. We decided to increase our train set to 90% of our data to include some observations of months where migration was present. Furthermore, it is common to take the logarithm of data that consists of large numbers like the migration volumes because it often enhances the interpretability, performance, or validity of statistical assumptions (Curran-Everett, 2018). However, fitting the model to transformed data and using it to predict the migration volume gave worse results than using the original data. This could be explained by the structural break in the data, which causes a significant change in the underlying process generating the data. As a result, the relationship between variables that the model was trained on no longer holds, leading to inaccurate predictions when applied to data affected by the break.

We have created two subsections. Section 5.1.4.1 uses the spread between the savings rate and term deposit rate as an explanatory variable. Section 5.1.4.2 uses the term deposit rate as an explanatory variable. We analyze this difference to determine whether the incentive for customers to move their money is the spread between the two rates or the level of the term deposit rate. It is assumed that the average vintage of customers who open a term deposit will increase uniformly over the next two years. Additionally, it is assumed that NN Bank will maintain a competitive position in the market compared to other small Dutch banks.

5.1.4.1 Spread between savings and term deposit rate

The coefficients of the model that utilizes the spread as an explanatory variable are calibrated using the OLS method on our train set. Table 5.4 shows these coefficients and the performance measures of the migration model. The performance measures suggest a reasonable level of accuracy in our model's predictions. However, this performance is due to the model's good fit to the training data and the non-stationarity of our data. The training set contains few observations of migrating volume which is less challenging to predict, making it difficult to assess the true performance of the model. It is also difficult to fully trust the significance of the coefficients and their p-values due to the non-stationarity issues of the data.

Table 5.4: In-sample coefficients and performance measures of migration volume model (spread)

	Maturity 1-year	Maturity 2-year	Maturity 3+ year
Constant	2.831e+05 (0.521)	7.171e+05 (0.000)*	-4.783e+05 (0.143)
$d^{td}_{t-1} - d^{s}_{t-1}$	1.141e+06 (0.434)	8.168e+06 (0.000)*	1.376e+06 (0.000)*
Vintage	28.3579 (0.997)	-2.061e+04 (0.044)*	-1811.0633 (0.696)
Top 3	1.429e+07 (0.000)*	7.171e+05 (0.000)*	5.657e+05 (0.205)
R^2_{Adj}	0.861	0.749	0.593
AIC	2257.	2541.	2778.
BIC	2267.	2549.	2788.

Note: * next to p-values indicate significance.

The estimated coefficients of the different maturity groups are used to predict the migration volume per group on our test set. As shown in Figure 5.11, there is a significant difference between our model's migration volume predictions and the actual migration volume for each maturity group. This difference is most likely due to the minimal migration in our dataset until 2022, which may not have provided our model with enough variability to learn from and make accurate predictions for future periods. Our model's predictions for the 1-year and 3+ year maturities are consistently below the actual observed values, suggesting that the model may be underestimating the factors that contribute to migration for these maturities. On the other hand, the model overestimates migration for the 2-year maturity, suggesting that it may be overweighting some coefficients.

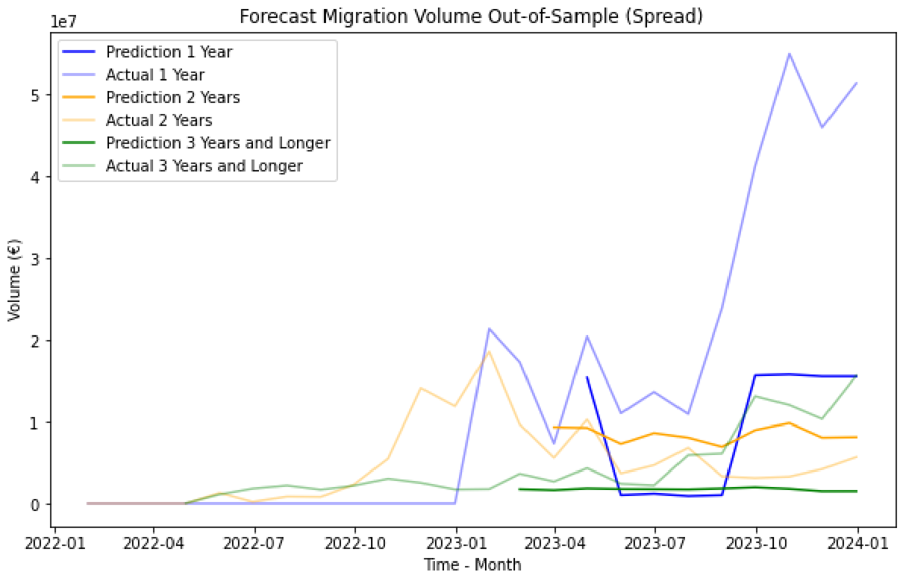

Figure 5.11: Actual migration volume per maturity group vs. fitted volumes (spread)

The values of the coefficients and p-vales as well as the performance measures in Table 5.4 make it difficult to make solid conclusions about the performance and inclusion of certain variables in our model. Additionally, the model was fitted to the full dataset, given in Table B.11 in Appendix B.5, and used to predict three different maturity groups, as shown in Figure 5.12. We used the predicted savings rate and term deposit rate as input for the model. The model predicts that deposit volume will continue to migrate towards the 1-year term deposit but will decrease until the end of 2025. After that point, the volume appears to stabilize just below €5.000.000. While there is some migration of deposit volume for the 2-year maturity group, it eventually ends up at zero. The migration volume of the 3+ year maturity group is predicted to remain constant over the coming two years. This decrease in migration volume is due to the negative constant of both groups on the full dataset, which eventually leads to a decrease in migration volume once the spread between both rates becomes too small.

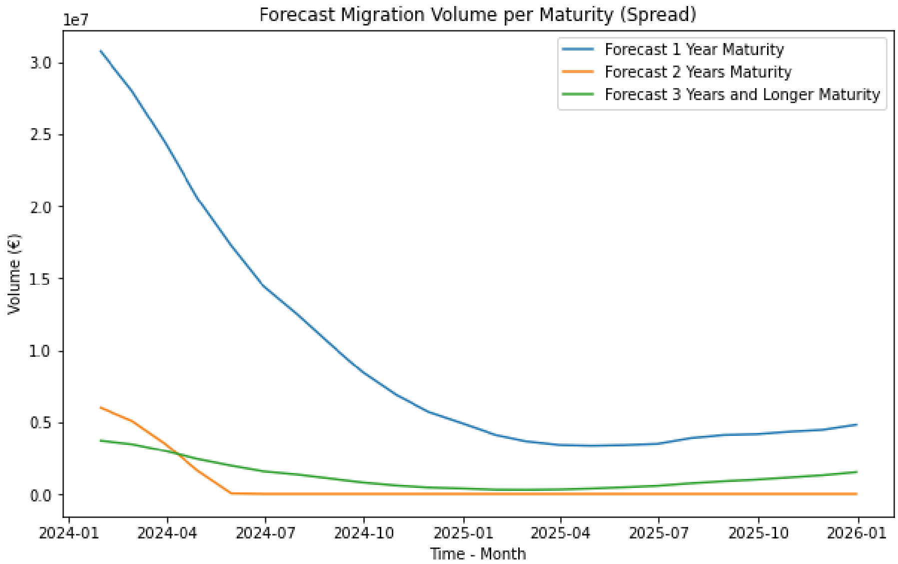

Figure 5.12: Out-of-sample forecast of migration volume (spread)

5.1.4.2 Term deposit rate

Using the term deposit rate as an explanatory variable, we see the same type of results. Table 5.5 displays the results of the estimation of the coefficients of the migration model per maturity group. The three values of the R^2_{Adj} suggest a decent in-sample fit on the train set, however, the results can not fully be relied on for the same data issues discussed in the previous subsection.

Table 5.5: In-sample coefficients and performance measures of migration volume (TD rate)

	Maturity 1-year	Maturity 2-year	Maturity 3+ year
Constant	3.41e+04 (0.923)	-9.726e+05 (0.000)*	2.763e+05 (0.533)
d_{t-1}^{td}	-1.063e+05 (0.844)	3.752e+06 (0.000)*	7.357e+05 (0.000)*
Vintage	2617.8605 (0.779)	5.771e+04 (0.000)*	2.482e+04 (0.000)*
Top 3	1.524e+07 (0.000)*	-9.726e+05 (0.000)*	-9.31e+05 (0.177)
R^2_{Adj}	0.860	0.517	0.515
AIC	2258.	2594.	2795.
BIC	2267.	2602.	2805.

Note: * next to p-values indicate significance.

The predictions on our test set of the models are off when compared to the actual migration volumes per maturity group shown in Figure 5.13. The same type of results can be seen when looking at Figure 5.12, where the migration volumes of the 1-year and 3+ year groups are underestimated and the 2-year group is overestimated. This is again the result of the limited observations of the migration in our train set.

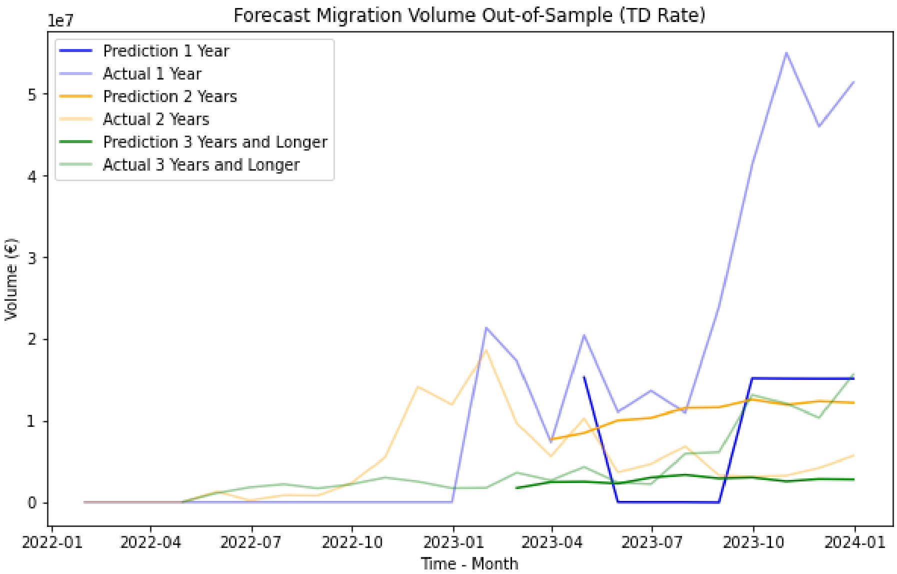

Figure 5.13: Actual migration volume per maturity group vs. fitted volumes (TD Rate)

When we fit our models to our full dataset, given in Table B.12 in Appendix B.5, we see some differences in predicted migration volume in Figure 5.14 compared to Section 5.1.4.1. The input for the model is the same, only the term deposit rate itself is used instead of the spread. For all three maturity groups, we see a migration volume that has approximately the same initial size compared to using the spread as an explanatory variable. However, for all groups, we see a flatter line, and for the 2-year and 3+ year groups, the model predicts that NN Bank will continue to have a migrating volume for the next two years. This can be explained by the historically high predicted rates for term deposits in the different maturity groups and the assumed competitive market position of NN Bank. These factors will continue to encourage customers to shift their deposit volumes to term deposits over the next two years. However, there is a contradiction in these forecasts as our predicted savings rate is higher than the predicted term deposit rates per maturity group. Customers would normally prefer not to tie up their money for a certain period and keep it in their savings account, but if a customer expects deposit rates to fall, they might choose to open a term deposit and take advantage of the fixed deposit rate. Finally, we should keep in mind that our predictions contain a fair amount of uncertainty because of the use of forecast uncertainty in the predicted term deposit rate, as well as the non-stationarity issues of our data to which the model is fitted.

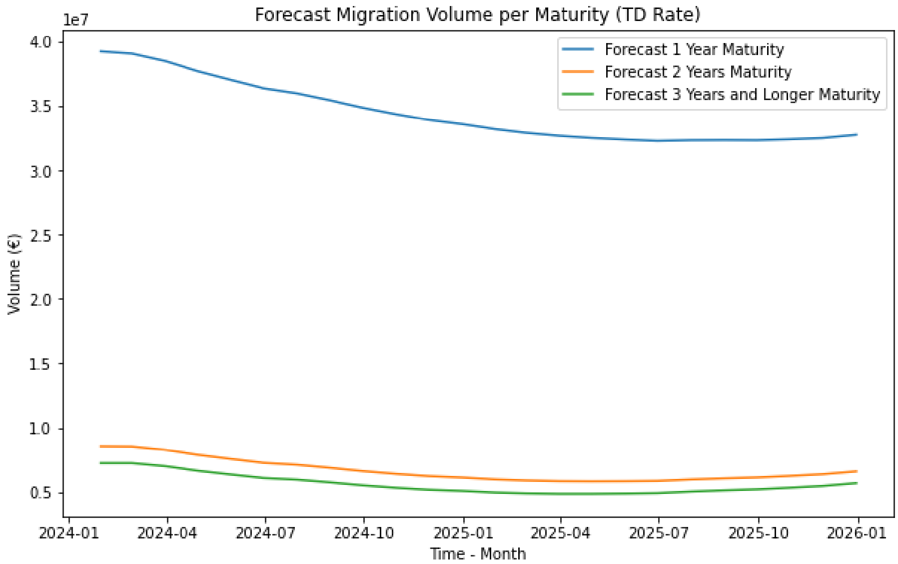

Figure 5.14: Out-of-sample forecast of migration volume (TD rate)

5.1.4.3 Outflow distribution

In our base scenario, we use the predicted outflow volume from Section 5.1.3.2 and combine it with the predicted migration volume per maturity group for both the spread input and the term deposit rate input. The outflow from the bank is calculated by subtracting the total migration volume per month from the outflow of the savings accounts. This outflow distribution indicates what proportion of the outflow will migrate to each maturity group and can help a bank in its liquidity management and hedging strategies.

Figure 5.15 shows the historical outflow distribution of 2023 and the predicted distribution over a 2-year period using the spread as input. To the left of the dashed line is the historical outflow distribution and to the right is the predicted outflow distribution including the predicted outflow and migration volume to each maturity group from Section 5.1.4. The proportion of deposit volume migrating towards term deposits gradually decreases over time. This is due to the reduced incentive for customers to open term deposits, as the spread between the savings rate and the term deposit rate also decreases over time.

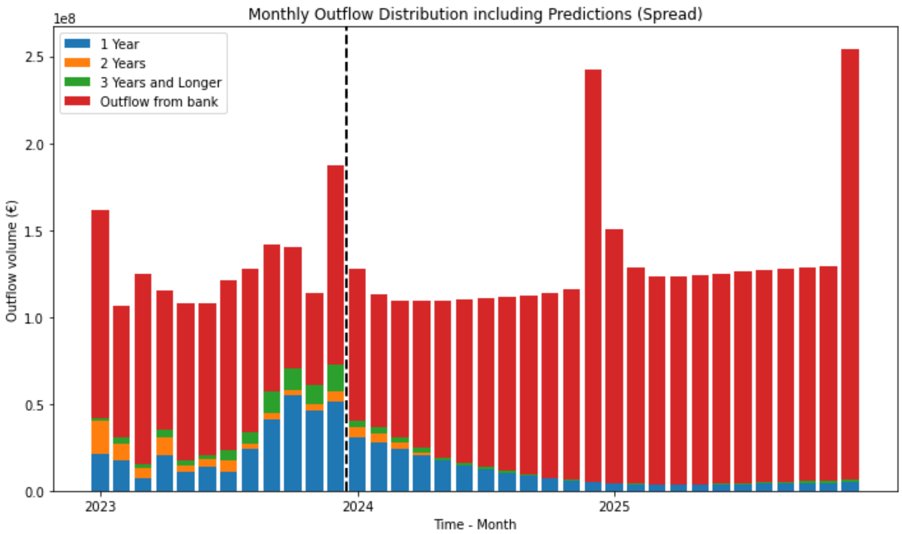

Figure 5.15: Out-of-sample forecast of the base scenario outflow distribution (spread)

Figure 5.16 displays the same information as Figure 5.15, but with the term deposit rate as input. According to the base scenario, a significant proportion of the outflow is expected to migrate towards term deposits. From the money that flows out of the savings accounts, a significant proportion will remain in the bank. This results in additional interest having to be paid on this proportion, as less interest is paid on savings accounts than on term deposits. This is an example of deposit cannibalization.

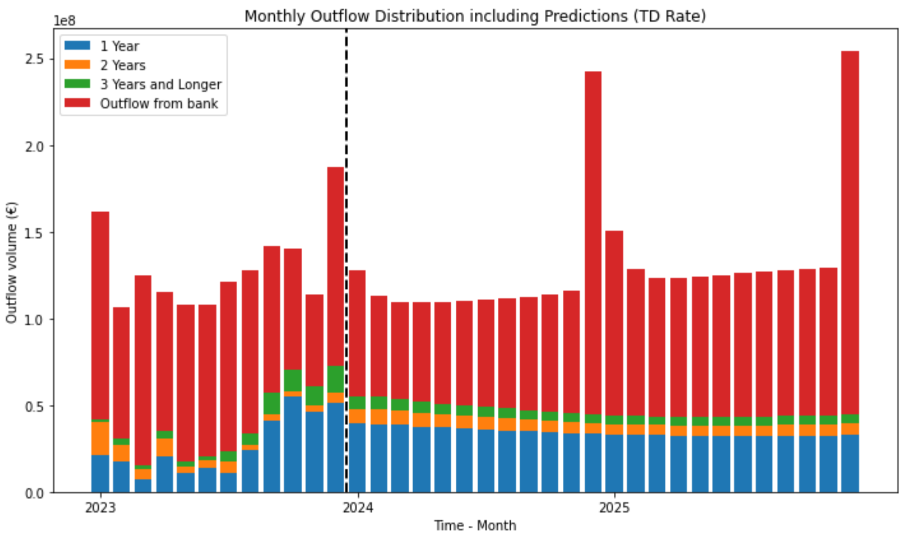

Figure 5.16: Out-of-sample forecast of the base scenario outflow distribution (TD rate)

Finally, it is difficult to conclude which variable, the spread or the term deposit rate, provides a better model since both variables are significant in both models. However, it is important to note that the significance and goodness of fit of our data may be limited due to non-stationarity issues and the limited amount of observations.

5.2 Vector Autoregressive model

This section discusses the results of our VAR model where we aim to predict the migration volume over a 2-year period. We begin by identifying the appropriate time series using the Granger Causality test. This test helps us find the time series that have causality between them, which is necessary for forecasting all the time series. Figure C.1 in Appendix C shows the results of this test in a matrix. Time series with a p-value lower than 0.05 indicate causality between them. We chose this set of time series because we believe they influence each other and can be used to forecast the migration of deposit volume towards term deposits. The figure shows that this is true for almost all the time series, except for the relationship between the savings volume of term deposits and migration of deposit volume with the savings volume of savings accounts. In a Granger Causality Matrix, the presence of causality in one direction does not necessarily imply its presence in the opposite direction. The lack of causality between the series and the volume of term deposits, as well as the migration of deposit volume, does make sense. This can be explained by the increase in savings account volume, while there was almost no increase in term deposit volume and no migration towards it. However, towards the end of the time series, there was a sudden increase in both, which may be difficult to explain for the model. The savings volume is a crucial time series that significantly affects other time series in the opposite direction. To ensure a proper VAR model, we must select the correct time series and confirm that they are stationary. To make all the time series stationary, we transformed the data by taking the second difference. Then, we applied the ADF test to test for stationarity. All time series are now $I(2)$ as all the time series have a p-value lower than 0.05 for the ADF test. Furthermore, the Durbin Watson (DW) test results in Table 5.6 indicate various levels of autocorrelation for the different time series.

Table 5.6: In-sample Durbin Watson test results

Variable	Durbin Watson Statistic
Market rate	1.96
Term deposit rate	2.14
Savings rate	2.1
Savings account volume	2.41
Term deposit volume	2.35
Migration volume	2.87

Overall, the DW test results suggest that most rates exhibit low to no autocorrelation, which is promising for the reliability of the forecasts derived from these series. However, we observe negative autocorrelation for the old money volume which lies outside the acceptable range of 1.50 - 2.50. Hence, we could experience problems with the predictive accuracy and reliability of this time series. We now determine the optimal lag order for the model. Since we have multiple time series, we evaluate the performance of the BIC, which accounts for the number of variables. The BIC indicates the lowest score with a lag of one, so we fit our model with a lag order of one. To test the performance of our VAR model on our data, we use a test set of all the time series. Figure C.2 in Appendix C shows the model's performance on the test set after fitting it on the train set. The figure shows that the VAR model consistently predicts an extension of the last observed trend for each time series. In other words, if the last known data point was increasing, the model forecasts a continuation of that increase, and vice versa for a decrease. To model shows a limitation in capturing a turning point or changes in trend. This is a common issue where the prediction essentially becomes a naive forecast. We also tested whether fitting the VAR model with a higher lag order could improve the predictions, however, this gave similar results. Figure C.3 in Appendix C displays the predictions of the VAR model when fitted on the entire data set. The same issue arises as in the test set, where the predictions align with the direction of the last observations. The issue persists even after fitting the VAR model to all available data. This suggests that the model may not be capturing the underlying processes that drive changes in the time series. This can occur if the time series is subject to complex dynamics that are not reflected solely by past values or if there are external factors or structural breaks in the series that the model does not account for. However, we attempted to address the structural break by introducing an exogenous dummy variable. Unfortunately, this approach did not resolve the issue, which was previously described in this chapter. Insufficient data is available after the structural break to accurately model this characteristic.

The prediction of the VAR model for the migration volume over the 2-year period depicted in Figure C.3 is not feasible as it overestimates the migration volume and continues to predict the current trend of the line. Therefore, we have decided to continue using the predictions made with our model framework of Section 4.2 in our scenario analysis.

5.3 Scenario analysis

In this section, we aim to answer sub-question 5c by using our model framework given in Section 4.2. Although it is difficult to assess the true performance of our migration model due to data issues, we have decided to perform a scenario analysis. This scenario analysis provides insight into the impact of different interest rate scenarios on our model's predictions. However, it is important to acknowledge the considerable uncertainty present in the predictions when interpreting them.

All banks in Europe have to follow the guidelines set by the EBA to keep a bank financially stable and liquid in times of unpredictable interest rate scenarios. The six shocks given by the EBA need to be applied by every bank to check if they still comply with all the guidelines that are set. These scenarios are also applied in this research as described in Section 4.4. In Appendix D, all the different interest rates with the EBA shock scenarios are shown. Next, we can examine the results when we include these shocks in our models. We use these rates as input to our deposit volume model and with the new deposit volume forecast we use this in our outflow model as a lagged variable. We also use both shocked deposit rates in our migration model.

Table 5.7 and 5.8 shows the percentage change in migration volume per maturity group for each scenario compared to our base scenario in Section 5.1.4, using the spread and term deposit rate as input, respectively. Based on the current interest rate environment, it is unlikely that short-term interest rates will increase further. The economy will eventually return to a normal yield curve. Our market rate model already predicts this, but by applying the Steepener scenario, we can return to a historically 'regular' scenario. This scenario will result in lower short-term rates and higher long-term rates, which is typical for a normal yield curve. Therefore, we highlight the predictions of the outflow distribution in this scenario as shown in Figure 5.17 and 5.18.

Table 5.7: Percentage change in prediction volume compared to base scenario (spread)

Scenario	1-Year (%)	2-year (%)	3+ year (%)
Up200	+98.82%	+154.07%	+257.37%
Steepener	+5.70%	+112.50%	+773.27%
ShortUp	+38.39%	−32.44%	−100.00%
Down200	−73.48%	−80.28%	−74.52%
Flattener	+10.28%	−57.41%	−100.00%
ShortDown	−38.39%	+39.83%	+595.87%

Using the spread as input, the Steepener scenario resulted in a large increase of 773.27% in the migration volume to the 3+ year maturity group. The significant increase can be explained by the higher term deposit rates for the longer maturities from the shock. This increases the incentive for customers to open a term deposit with a maturity of 3+ year as this is attractive because of the additional amount of interest. Additionally, we see an increase in migration volume towards the 2-year maturity group because of the bigger spread between the savings rate and the term deposit rate that this scenario imposes. Lastly, the small increase in migration volume towards the 1-year maturity group can also be explained as an increase in spread because the savings rate got a larger shock than the 1-year maturity term deposit rate. This significant shift indicates that depositors are interested in taking advantage of the higher yields offered by longer-term deposits.

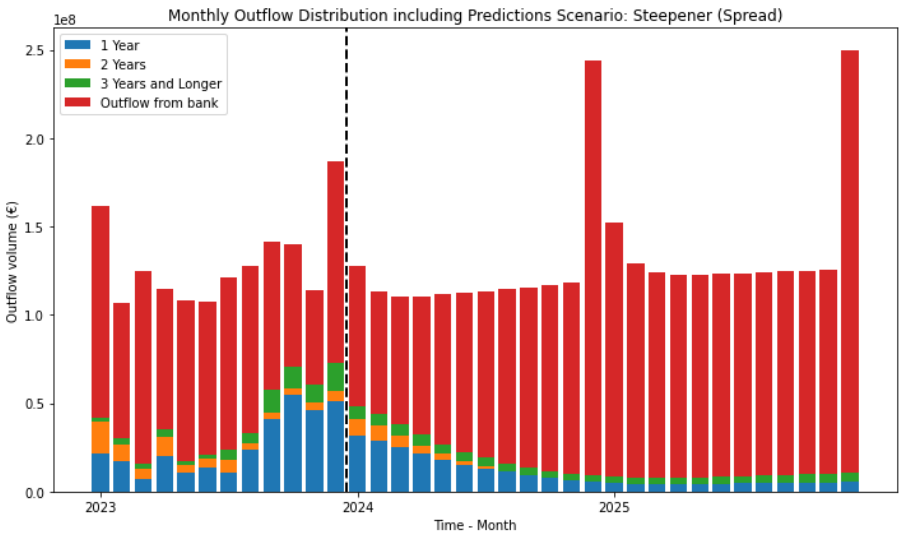

Figure 5.17: Outflow distribution in the Steepener scenario (spread)

We will now also highlight the Steepener scenario, but with the input of the term deposit rate.

Table 5.8: Percentage change in prediction volume compared to base scenario (TD rate)

Scenario	1-year (%)	2-year (%)	3+ year (%)
Up200	+102.35%	+196.81%	+129.97%
Steepener	−5.81%	−0.36%	+24.17%
ShortUp	+100.48%	+160.57%	+40.51%
Down200	−38.73%	−87.13%	−96.29%
Flattener	+82.06%	+120.16%	+4.23%
ShortDown	−36.86%	−66.66%	−32.05%

There is a slight decrease of -5.81% and -0.36% for the 1-year and 2-year maturity groups, respectively. However, there is a significant increase of 24.17% for 3-year and longer maturities. This suggests a preference shift towards longer-term deposits due to their higher yields, reflecting strategic investor behavior to maximize returns.

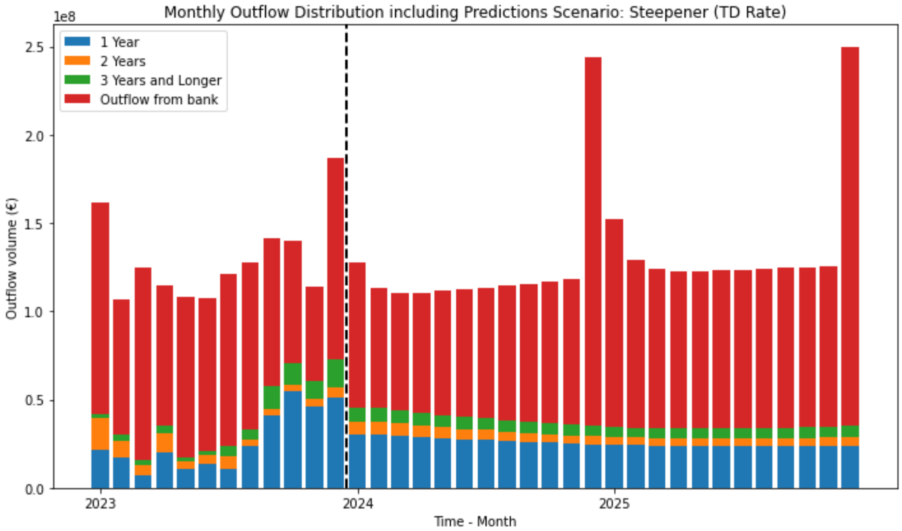

Figure 5.18: Outflow distribution in the Steepener scenario (TD rate)

In conclusion, our scenario analysis provides valuable insights into how deposit migration responds to interest rate shocks in the market. NN Bank can use these insights to improve its risk management frameworks and align its deposit strategies with anticipated market trends. However, the predictions of the different scenarios contain forecast uncertainty which should be kept in mind when interpreting the results.

5.4 Conclusion

This chapter aims to answer sub-question 5, which concerns the validation and interpretation of our models' results. We also aim to understand the incentives for customers to migrate their funds under different interest rate scenarios. The use of a VAR model to deal with the data issues did not provide reasonable predictions of the migration of deposit volume. The validation of our models of the model framework and their significance was challenging due to data issues, but the models made reasonable predictions for the 2-year forecast period. These predictions were used to eventually predict our migration volume and see how it contributes to the future outflow distribution of NN Bank. We have seen that there is a significant difference in the use of taking the spread as an incentive or the term deposit rate as an incentive. A decrease in the spread leads to less motivation for customers to migrate their savings towards term deposits. On the other hand, the use of the predicted term deposit rate itself led to a more constant migration volume to all the maturity groups. We presented our findings and the performance of our models on the test set. Additionally, we utilized our fitted models to forecast the migration volume in various scenarios, providing insights into why customers choose to migrate their savings volume towards term deposits and why they do not. We also examined six different interest rate scenarios and their impact compared to our base scenario.

Our results validated our expectations regarding the incentives for customers to migrate their deposit volume. We observed that a smaller spread diminishes the incentive to open a term deposit. Similarly, a decrease in the term deposit rate also reduces the motivation to open a term deposit. This insight is crucial for banks as it highlights the direct impact of rate adjustments on customer behavior. Finally, we did not include confidence intervals for our predictions as it is beyond the scope of this research.

6 Conclusions & Discussion

In Chapter 1, we discussed the relevance of researching a model framework for predicting the migration of deposit volume from savings accounts towards term deposits. This focus is driven by the recent trend of increased deposit migrations due to the high interest rate environment. In this chapter, we conclude the research by discussing our results, acknowledging the limitations, and proposing recommendations for future research.

6.1 Conclusions

The goal of this research was to forecast the migration volume from savings accounts towards term deposits. We focused on designing a model framework that incorporates interest rates as an incentive for customers to open a term deposit. Therefore, the following research question was formulated:

"How can the migration of deposit volumes for Nationale-Nederlanden Bank's deposit products be accurately forecasted?"

To address our research question, we began with a literature review focusing on the modeling of non-maturing deposits and the modeling of deposit volume migration. We explored how market rates, deposit rates, and deposit volumes are related and the best methods for modeling these elements. First, we examined several market rate models and chose the Dynamic Nelson-Siegel model because it dynamically calculates a yield curve at each time step, making it well-suited for our research. Next, we investigated various models for deposit rates described in the literature. We selected models that best represent the savings and term deposit rates at NN Bank. For term deposits, we found a model in the literature that uses the Euribor 6M as a reference rate. We researched various deposit volume models to estimate the monthly deposit volume of NN Bank's savings accounts. Additionally, we investigated how to model the outflow from these accounts. We discovered that there is limited existing literature that covers the migration of deposit volume towards term deposits. This is likely due to the historically limited popularity of term deposits. We found a model for modeling the switching behavior of bank customers that closely aligns with the literature we were seeking on the migration of deposit volume. The proposed adjustment to the model, created for modeling bank-switching customers, includes an interest rate incentive, along with customer-specific and market data. Two versions of the model were estimated; one version with the spread between the savings rate and the term deposit rate, and the other covering the term deposit rate itself. This was done to determine if there was a difference in the significance and size of the coefficient. We applied our proposed model framework to our test set to evaluate its performance and predict various interest rates and deposit volumes. We also tested the impact of different interest rate shock scenarios by the EBA on our results. In addition to this model framework, we also tested the use of a VAR model to address data issues.

6.1.1 General conclusions on the results

The results of our research consisted of various predictions of market rates, deposit rates, and deposit volumes. The relationships of these predictions have been discussed extensively throughout the research. Our selected models generally had difficulty fitting the time series with the sudden increase in interest rates and the corresponding shifts in deposit volumes. The objective was to forecast the migration volume towards term deposits. However, the migration model could not accurately predict the monthly migration volume in the test set for different maturity groups. It was observed that the model had an overall good in-sample fit, which suggests overfitting on the train set or non-stationarity issues with the data. The model's predictions of migration volume based on the spread and the term deposit rate appear reasonable when compared to the historical monthly migration volume. We found that the decreasing spread between the term deposit rate and savings rate results in less migration volume, while the slowly decreasing predicted term deposit rate leads to a constant flow of migration volume. Our model forecasts that, in our base scenario, the migrating deposit volume for nearly every maturity group will decline gradually over the next two years, based on both the spread and the term deposit rate. It is difficult to determine whether the use of the spread or the term deposit rate provides more accurate predictions due to the unreliability of our models, which is further explained by their inability to accurately predict the migration volume in the test set for different maturity groups and the non-stationarity of our data.

The scenario analysis, particularly focusing on the Steepener scenario, demonstrates that applying the shock across various yield curve tenors results in the most significant increase in migration volume for the 3+ year maturity group, for both the spread and the term deposit rate. This analysis highlights the sensitivity of our model's predictions to different interest rate scenarios, illustrating that our model is particularly responsive to changes in deposit rates.

Furthermore, the utilization of a VAR model to address our non-stationarity issues did not yield reasonable predictions. The predictions on our test set and the forecast over a 2-year period exhibited signs of overfitting, where the recent trend of the time series was extrapolated.

6.1.2 Conclusion on the research question

Our research proposed a model framework for modeling the migration volume from savings accounts towards term deposits where different models for the market rate, deposit rate, and deposit volume can be used. Our data shows a significant increase in deposit volume migration towards term deposits in NN Bank's savings portfolio. We tested the correlation between the market rate, deposit rates, and deposit volume. We found that the correlation between the market rate, term deposit rate, and total term deposit volume was strongly positive. This led to the conclusion that the recent increase in deposit rates provided an incentive for customers to migrate their deposit volume, and therefore this incentive is included in our migration model. However, we discovered an interesting dynamic: a large spread between the savings rate and term deposit rate does not necessarily lead to the greatest incentive for customers to migrate. The level of the term deposit rate also plays a role. Despite this, the migration model, including both incentives, was not able to accurately predict migration volumes. Nonetheless, the research provides valuable insights into investigating and modeling the recent trend of opening term deposits and their volume for a bank under different interest rate scenarios.

6.2 Discussion

In this section, we analyze the results of our research and address the limitations and assumptions made. We conclude with recommendations for future research.

6.2.1 Data limitations

The results of this research may not be entirely reliable due to data issues. These issues include limited observations and end-of-sample bias, which contribute to the level of prediction uncertainty. This uncertainty is caused by the estimation of model parameters and the error term in the regression analysis, as well as the fact that our predictions are based on prior predictions that already include uncertainties.

We encountered challenges in our analysis due to data non-stationarity caused by a structural break in the time series. Despite attempts to address this issue through various methods, such as second difference, logarithmic transformations, and a binary dummy variable for the structural break, we did not observe significant improvements in the results. As a result, models that did not correct for non-stationarity often provided better predictions. Therefore, it is important to approach any predictions from this research cautiously. It should be recognized that non-stationarity and structural breaks could affect the reliability and accuracy of the outcomes.

Furthermore, confidence intervals were not calculated due to the computational effort and time required. Managing the error term for multiple models while estimating parameters requires considerable time and computational resources. Given the complexity and scope of the research, calculating confidence intervals for all models was deemed impractical within the constraints of this study. Moreover, the uncertainty originating from the first model impacts the subsequent models, and so forth. Consequently, the 'true' confidence interval for the final output is likely extremely wide. Given the likely wide confidence interval, it may not be worth spending a lot of time and resources on precisely quantifying this uncertainty.

Finally, we did not apply the common practice of log-transforming large numerical data in regression to normalize distributions and reduce the influence of outliers on our deposit volume data. Typically, log-transformations help in stabilizing the variance and improving model fit by making the data more symmetrical. However, when we applied log transformations to the deposit volume data, especially in our migration volume model, the results significantly deteriorated. This negative effect could be caused by the characteristics of our data. The log transformation might have hidden or changed important patterns and relationships needed for accurate predictions. Hence, we chose to keep the original scale of the data to better understand the actual dynamics of deposit volume movements.

6.2.2 Assumptions

During this research, we made several key assumptions. For the savings rate of NN Bank, we focused on accounts with balances between €0-100,000, assuming that rates for higher account classes are typically 10 basis points lower. This assumption was based on the fact that the average account deposit volume fell within this class. Additionally, we defined migration volume as funds that have been in the bank for at least a month. This helped us standardize the measurement of migration volume across different accounts. Moreover, while modeling the migration volume, we considered deposit rates as an incentive. However, it is important to note that such decisions are also influenced by market sentiment and macroeconomic variables like inflation and GDP. Finally, in our migration model, we assumed that NN Bank's customer base would remain stable over the next two years and that the bank would continue to hold a competitive position in the market relative to other small Dutch banks. It is important to note that changes in these two variables over time could potentially lead to different results.

6.2.3 Contribution to literature

This research is relevant because of the increased popularity of term deposits and the lack of existing literature on the migration of deposit volume between savings accounts and term deposits. Additionally, there is currently no market practice at Dutch banks on how to model this. Despite the uncertainty in our results, we have developed a framework that could provide a stepping stone for future research.

6.3 Future research

After conducting this research, we found several recommendations that could enhance and expand the study. It is recommended to differentiate between parameter uncertainty and process uncertainty. This differentiation can be computationally intensive, often requiring simulations within simulations and specialized programming techniques such as parallel processing. By using this approach, we can calculate confidence intervals efficiently, providing a clearer understanding of the range and reliability of our model predictions.

Furthermore, the model framework should be re-calibrated as more data becomes available to improve the accuracy of the predictions. These predictions could assist banks in mitigating deposit cannibalization and managing interest rate risk. Future research could also explore how much banks might save by implementing these strategies. The current model framework is limited to savings accounts and term deposits, but most banks also have current accounts. It would be interesting to develop a framework that includes these as well. Transactional deposit data could be particularly valuable in this context. Moreover, banks could further refine the model to focus on specific customer groups. This would allow for a deeper analysis to determine if different incentives, such as spreads or interest rates, might be more effective for certain segments. Additionally, the model framework could be extended by including macroeconomic variables as explanatory variables, such as the inflation rate or GDP. This would be useful as shifts in economic indicators like inflation and GDP often influence customer behavior and investment preferences.

Finally, with the rising popularity of term deposits, a significant volume of deposits will flow back to savings accounts. For NN Bank, understanding customer behavior when a term deposit matures is important. Customers may opt to keep their money in their savings account, withdraw it, or reinvest in a new term deposit. Accurately predicting these actions will enable the bank to incorporate this information into their cash flow forecasts and liquidity calculations.

References

Affek, S. P. (2019). Detecting stationarity in time series data. https://towardsdatascience.com/detecting-stationarity-in-time-series-data-d29e0a21e638

Ahmed, I. (2023). What are acf and pacf plots in time series analysis? https://ilyasbinsalih.medium.com/what-are-acf-and-pacf-plots-in-time-series-analysis-cb586b119c5d#

Aupérin, A. (2020). Demand deposits modeling.

Avsar, E., & Ruimy, B. (2021). Modeling nonmaturing deposits: A framework for interest and liquidity risk management. *The Journal of Risk*. https://doi.org/10.21314/jor.2021.017

Banken.nl. (2023). Marktaandeel. https://www.banken.nl/bankensector/marktaandeel

BCBS. (2016). Interest rate risk in the banking book. https://www.bis.org/bcbs/publ/d368.pdf

Ben-David, I., Palvia, A., & Spatt, C. (2017). Banks' internal capital markets and deposit rates. *Journal of Financial and Quantitative Analysis*, *52*(5), 1797–1826. https://doi.org/10.1017/S0022109017000606

Bhandari, A. (2024). Key difference between r-squared and adjusted r-squared for regression analysis. https://www.analyticsvidhya.com/blog/2020/07/difference-between-r-squared-and-adjusted-r-squared/

Chakravarty, S., Feinberg, R., & Rhee, E.-Y. (2004). Relationships and individuals' bank switching behavior. *Journal of Economic Psychology*, *25*(4), 507–527.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (2005). A theory of the term structure of interest rates. In *Theory of valuation* (pp. 129–164). World Scientific.

Cromwell, J. B. (1994). *Multivariate tests for time series models*. Sage.

Curran-Everett, D. (2018). Explorations in statistics: The log transformation. *Advances in physiology education*, *42*(2), 343–347.

Deuflhard, F. (2018). Quantifying inertia in retail deposit markets.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of econometrics*, *130*(2), 337–364.

EBA. (2022). Regulatory technical standards specifying supervisory shock scenarios.

ECB. (2023). Key ecb interest rates. https://www.ecb.europa.eu/stats/policy_and_exchange_rates/key_ecb_interest_rates/html/index.en.html

Elkenbracht, M., & Nauta, B. (2006). Managing interest rate risk for non-maturity deposits. *RISK-LONDON-RISK MAGAZINE LIMITED-*, *19*(11), 82.

Financieel Dagblad. (2023). Spaarders zetten hun geld bij de buren in europa. https://fd.nl/financiele-markten/1493467/spaarders-zetten-hun-geld-bij-de-buren-in-europa

Fonseca, L., McQuade, P., Van Robays, I., & Vladu, A. L. (2023). The inversion of the yield curve and its information content in the euro area and the united states. https://www.ecb.europa.eu/pub/economic-bulletin/focus/2023/html/ecb.ebbox202307_02~78906aa989.en.html

Frauendorfer, K., & Schürle, M. (2007). Dynamic modeling and optimization of non-maturing accounts. *Liquidity risk: Measurement and management. Singapore: John Wiley and Sons*, 327–359.

Friedman, M. (1966). Interest rates and the demand for money. *The Journal of Law and Economics*, *9*, 71–85.

Fund, I. M. (1990). *The role of national saving in the world economy: Recent trends and prospects: Occa paper no.67*. https://doi.org/10.5089/9781557751348.084

Gerritsen, D. F., & Bikker, J. A. (2020). Bank switching and interest rates: Examining annual transfers between savings accounts. *Journal of Financial Services Research*, *57*, 29–49.

Gimeno, R., Manchado, B., & Minguez, R. (1999). Stationarity tests for financial time series. *Physica A: Statistical Mechanics and its Applications*, *269*(1), 72–78.

Gyllberg, C. (2022). Forecasting the outflow from non-maturity deposits using astressed seasonal autoregressive monte carlo simulation.

Hayes, A. (2022). Euro interbank offer rate (euribor) definition, uses, vs. eonia.

Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, 77–105.

Hubbard, R. G., & O'Brien, A. P. (2018). *Money, banking, and the financial system*. Pearson.

Hull, J. (2018). *Risk management and financial institutions*. John Wiley.

Jarrow, R. A., & Van Deventer, D. R. (1998). The arbitrage-free valuation and hedging of demand deposits and credit card loans. *Journal of Banking & Finance*, *22*(3), 249–272.

Kalkbrener, M., & Willing, J. (2004). Risk management of non-maturing liabilities. *Journal of Banking & Finance*, *28*(7), 1547–1568.

Keynes, J. M. (1936). *The general theory of employment, interest and money* [14th edition, 1973]. Macmillan.

Kiser, E. K. (2002). Predicting household switching behavior and switching costs at depository institutions. *Review of Industrial Organization*, *20*, 349–365.

Laidler, D. (1966). The rate of interest and the demand for money–some empirical evidence. *Journal of Political Economy*, *74*(6), 543–555.

Lundgren, F. (2022). Modelling of non-maturity deposits.

Manrai, L. A., & Manrai, A. K. (2007). A field study of customers' switching behavior for bank services. *Journal of retailing and consumer services*, *14*(3), 208–215.

Marena, M., Romeo, A., & Semeraro, P. (2023). Non-maturing deposits modelling in a ornstein-uhlenbeck framework. *Applied Stochastic Models in Business and Industry.*

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, 473–489.

NN Bank. (2023). Uitleg over spaarrentes. https://www.nn.nl/Particulier/Sparen/Uitleg-over-de-spaarrente.htm

Nyström, K. (2008). On deposit volumes and the valuation of non-maturing liabilities. *Journal of Economic Dynamics and Control*, *32*(3), 709–756.

O'Brien, J. M. (2000). Estimating the value and interest rate risk of interest-bearing transactions deposits. *Available at SSRN 256712.*

O'Brien, J. M., Orphanides, A., & Small, D. H. (1994). Estimating the interest rate sensitivity of liquid retail deposit values. *Available at SSRN 5602.*

Paraschiv, F., & Schürle, M. (2010). Modeling client rate and volumes of non-maturing accounts. *Institute for Operations Research and Computational Finance.*

Pinelis, I. (2019). Exact bounds on the inverse mills ratio and its derivatives. *Complex Analysis and Operator Theory*, *13*(4), 1643–1651.

Van der Ploeg, S. (2023). Estimating the maturity of non maturing deposits: A dynamic replicating portfolio approach. *Erasmus University.*

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, *5*(2), 177–188.

Witt, A., Kurths, J., & Pikovsky, A. (1998). Testing stationarity in time series. *physical Review E*, *58*(2), 1800.

Wyle, R. (2014). Asset and liability management: Applications for the management and modeling of non-maturing deposits. *Moody's Analytics*. https://www.moodysanalytics.com/-/media/whitepaper/2014/2014-22-05-Asset-Liability-Management-Applications-Modeling-Non-Maturing-Deposits.pdf

Zhao, C., Noman, A. H. M., & Asiaei, K. (2022). Exploring the reasons for bank-switching behavior in retail banking. *International Journal of Bank Marketing*, *40*(2), 242–262.

Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS*, 385–429.

Appendix A: List of banks

Table A.1: List of competitor banks in the Netherlands

Big banks	Small banks
ABN Amro	Anadolubank
ING	ASN
Rabobank	DHB Bank
	Garantibank
	Knab
	LeasePlan Bank
	Lloyds Bank
	Regiobank
	SNS
	Triodos Bank
	Yapi Kredi Bank

Appendix B: Model prediction results

B.1 Market rates

Table B.1: Performance measures for different maturities of DNS model

Maturity	R^2_{Adj}	RMSE
1Month	0.4679	0.0988
3Month	−0.2538	0.1387
6Month	−0.6434	0.1813
1YS	0.4626	0.2174
2YS	0.8579	0.2661
3YS	0.9141	0.2876
4YS	0.9480	0.2972
5YS	0.9580	0.2911
6YS	0.9666	0.2880
7YS	0.9429	0.2772
8YS	0.8740	0.2614
9YS	0.7363	0.2528
10YS	0.5024	0.2443

Table B.2: In-sample regression coefficients of the DNS model

β_1	β_2	β_3
-3.956	3.933	4.321
-0.2169	0.1899	0.8356
3.926	-3.943	-3.770
2.664	-2.690	-1.617
-0.4426	0.4489	-0.3249
0.5387	-0.5574	0.075
1.631	-1.631	-2.782
2.908	-2.926	-3.742
4.366	-4.397	-5.472
-6.564	6.518	8.917
4.027	-4.034	-6.503
-1.642	1.613	1.227
-4.45	4.413	5.659
0.4348	-0.4593	0.1884
-1.038	1.037	0.6218
1.059	-1.065	-1.802
-2.594	2.583	2.347
-1.608	1.599	2.053
1.709	-1.709	-2.101
-1.881	1.857	3.56
3.597	-3.604	-3.421
-1.134	1.132	1.281
-0.2225	0.1928	1.14
0.1484	-0.1355	-0.9989
-3.42	3.407	4.736
1.075	-1.082	-1.181
1.29	-1.292	-1.639
-2.048	2.03	3.241
2.808	-2.814	-3.111
0.5565	-0.5422	-1.635
-0.9804	0.9668	1.869
1.578	-1.573	-2.24
-2.917	2.914	3.553
-4.53	4.514	5.81
-1.341	1.317	2.693
0.6271	-0.6218	-0.7098
2.105	-2.094	-3.114
-1.572	1.571	1.947
2.22	-2.202	-3.117
1.691	-1.688	-2.255
-1.726	1.719	2.424
0.7931	-0.7795	-1.351
-3.471	3.453	4.841
1.658	-1.65	-2.192
1.971	-1.959	-2.877
3.094	-3.075	-4.181
-2.3	2.315	1.954
-1.219	1.219	1.714
0.6583	-0.6441	-2.011
0.0223	-0.02784	0.1771
2.559	-2.552	-4.001
2.684	-2.709	-3.883
-0.4064	0.3725	-0.1904

β_1	β_2	β_3
0.6008	-0.6278	-2.055
-4.852	4.84	6.55
-2.599	2.605	3.944
-1.547	1.555	2.02
-0.7996	0.7939	1.782
1.293	-1.283	-2.936
0.3225	-0.3451	-0.8754
-4.612	4.615	6.363
5.237	-5.106	-7.468
-2.4	2.399	3.043
1.026	-1.104	-1.182
-1.12	1.053	1.305
-0.1262	0.07998	0.7715
1.213	-1.227	-1.854
2.234	-2.256	-2.847
-2.05	2.04	2.612
-1.029	1.01	1.285
2.346	-2.354	-2.427
-0.7001	0.7012	1.86
2.84	-2.841	-3.203
-1.164	1.16	1.737
-0.7433	0.7416	0.8573
-1.251	1.25	1.437
0.2276	-0.2153	-1.231
-0.5292	0.5193	0.9326
-1.853	1.841	3.25
-14.2	14.18	17.83
6.355	-6.357	-8.424
-3.49	3.46	5.173
-9.366	9.378	12.08
-2.951	2.963	5.119
-24.28	24.31	31.18
-11.5	11.57	16.12
0.9546	-0.8534	-0.9727
-3.265	3.446	4.735
21.5	-21.19	-29.65
-30.51	30.92	38.99
-2.868	3.489	4.447
14.28	-13.9	-18.94
20.46	-20.02	-28.4
-13.52	13.79	18.19
17.63	-17.32	-24.06
-9.113	9.434	11.61
27.37	-27.15	-35.19
10.66	-10.4	-13.72
6.471	-6.281	-8.677
-12.65	12.84	14.81
14.8	-14.7	-17.84
8.539	-8.457	-10.76
-0.7417	0.8409	2.276
15.07	-15.01	-18.61
9.284	-9.304	-13.09
18.72	-18.8	-24.44

Table B.3: Out-of-sample regression coefficients for the DNS model

β_1	β_2	β_3
-6.531	6.456	8.779
-7.72	7.728	11.53
10.6	-10.74	-13.13
13.89	-14.12	-18.42
-1.486	1.219	2.287
-4.649	4.39	5.82
-10.36	10.2	13.65
2.331	-2.537	-2.003
5.453	-5.654	-6.852
-5.548	5.326	7.352
-13.36	13.16	16.52
-9.161	9.012	12.17
1.414	-1.56	-1.329
-2.899	2.798	3.823
-3.865	3.78	5.22
-11.96	11.9	14.63
-5.085	5.061	6.176
-3.694	3.68	4.93
-6.402	6.46	8.611
-1.462	1.52	2.003
-2.487	2.542	2.317
-3.854	3.945	4.551
-2.337	2.442	2.88
-5.241	5.419	6.819

B.2 Savings rate models

Table B.4: In-sample coefficients and performance measures of savings rate models

	FD -t	FD	FD - Diff	3m MA	6m MA
constant	-0.0390 (0.000)*	-0.0129 (0.003)*	0.0204 (0.026)*	0.0217 (0.255)	0.0257 (0.206)
Δd_{t-1}	0.0556 (0.614)	0.2048 (0.056)	0.0391 (0.710)	-	-
ΔMA_t	-0.0236 (0.873)	0.0450 (0.772)	-0.1036 (0.480)	-	-
t	0.0005 (0.002)*	-	-	-	-
$d_{t-1} - r_{t-1}$	-	-	-0.0548 (0.000)*		
d_{t-1}	-	-	-	0.9415 (0.000)*	0.9389 (0.000)*
MA_t	-	-	-	0.0561 (0.201)	0.0680 (0.153)
R^2_{Adj}	0.160	0.047	0.213	0.995	0.994
AIC	-338.9	-330.4	-344.3	-348.9	-341.6
BIC	-329.2	-323.1	-334.6	-341.6	-334.4

Note: * next to p-values indicate significance.

Table B.5: Performance measures for savings rate models

Description	R^2_{Adj}	MAE	RMSE
Model FD - t	-1.086	0.6935	0.9387
Model FD	-0.6636	0.6357	0.8626
Model FD - Diff	0.9723	0.0821	0.1080
Model 3m MA	0.9355	0.1408	0.1698
Model 6m MA	0.9166	0.1455	0.1930

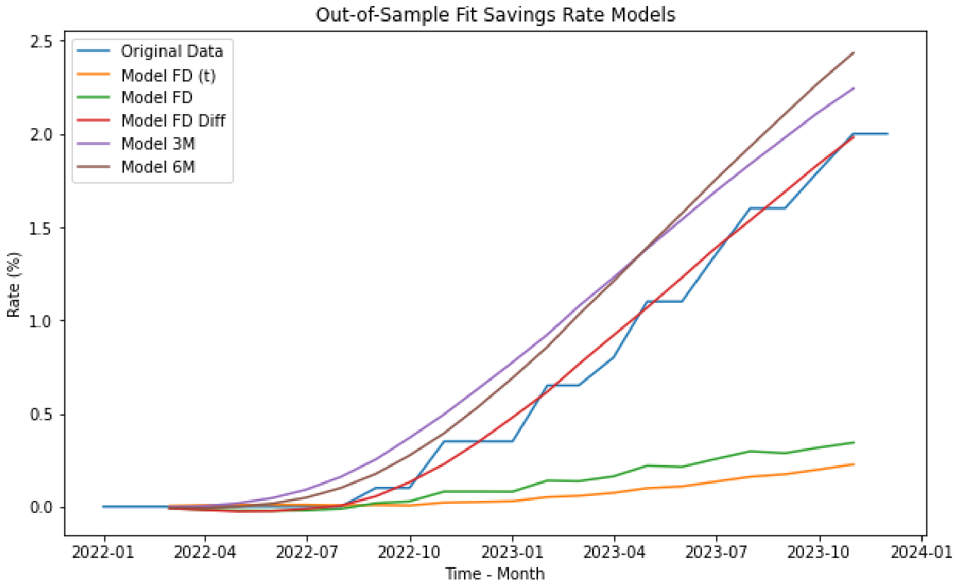

Figure B.1: Out-of-sample forecast of savings rate

B.3 Term deposit rates

Table B.6: Model coefficients and corresponding p-values for term deposit maturities

Model	β_1	β_2	β_3
1 Year	0.0888 (0.6388)	-0.0288 (0.0296)*	0.50 (0.00)*
2 Years	0.0925 (0.6279)	-0.0273 (0.0187)*	0.59 (0.00)*
3 Years	0.1962 (0.3812)	-0.0234 (0.0527)	0.63 (0.00)*
4 Years	0.3243 (0.2791)	-0.0199 (0.1535)	0.63 (0.00)*
5 Years	0.4076 (0.2546)	-0.0169 (0.2308)	0.64 (0.00)*
6 Years	0.4518 (0.2484)	-0.0139 (0.3049)	0.63 (0.00)*
7 Years	0.5061 (0.2170)	-0.0121 (0.3407)	0.63 (0.00)*
8 Years	0.5432 (0.1806)	-0.0130 (0.2490)	0.63 (0.00)*
9 Years	0.5801 (0.1760)	-0.0135 (0.2217)	0.62 (0.00)*
10 Years	0.5765 (0.1903)	-0.0127 (0.2249)	0.64 (0.00)*

Note: * next to p-values indicate significance.

Table B.7: Out-of-sample performance measures for term deposit rate model

Term Deposit Maturity	R^2_{Adj}	RMSE	MAE
1 Year	-1.0281	1.2913	1.0198
2 Years	-1.4574	1.5388	1.2713
3 Years	-1.0599	1.3762	1.1878
4 Years	-0.2010	1.0327	0.9073
5 Years	0.2265	0.8218	0.7267
6 Years	0.4227	0.7053	0.6225
7 Years	0.6315	0.5641	0.4851
8 Years	0.7149	0.4981	0.4193
9 Years	0.7697	0.4424	0.3724
10 Years	0.7560	0.4641	0.3829

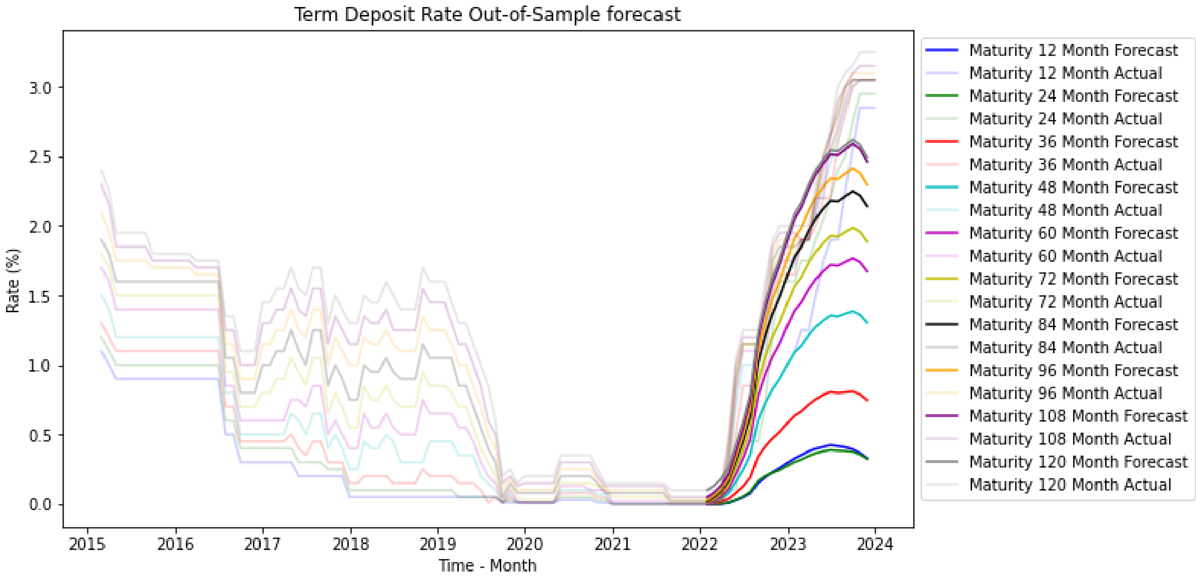

Figure B.2: Actual term deposit rates vs fitted rates

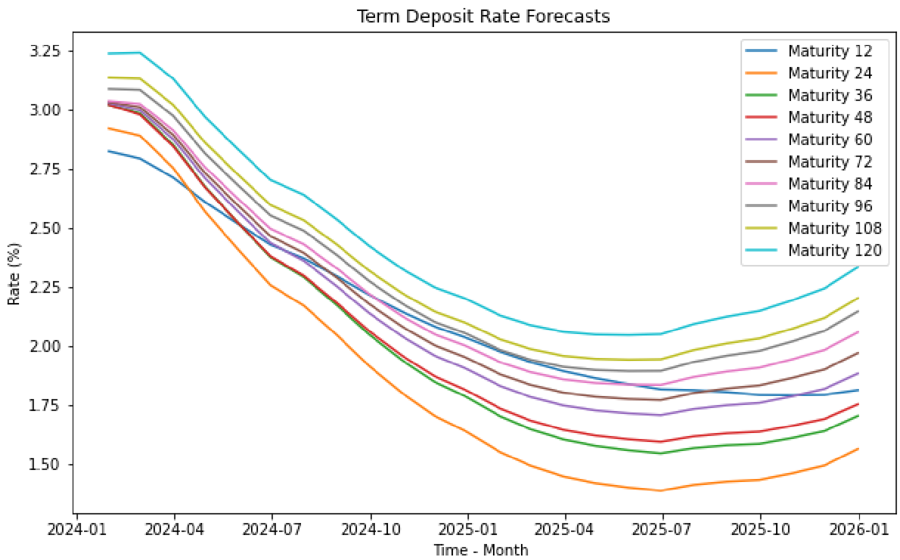

Figure B.3: Out-of-sample forecast of term deposit rates

B.4 Savings volume

Table B.8: In-sample coefficients and performance measures of savings accounts volume models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
constant	3.362e+09 (0.000)*	5.968e+09 (0.000)*	5.977e+09 (0.000)*	-5.902e+07 (0.660)	1.989e+08 (0.004)*	5.667e+07 (0.768)
$D_{t-1}\left[d_{t-3} - (\delta r_{t-3}^{short} + (1-\delta) r_{t-3}^{long})\right]$	0.16470 (0.002)*	-0.2497 (0.000)*	-0.2500 (0.000)*	-	-	-
t	2.441e+07 (0.000)*	-	-	-	-	-
$S_{12,t}$	-	-	-8.742e+07 (0.621)	-1.508e+08 (0.000)*	-1.542e+08 (0.000)*	-
D_{t-1}	-	-	-	1.0085 (0.000)*	0.9717 (0.000)*	0.9912 (0.000)*
$(d_{t-1}^{s} - r_{t-1})$	-	-	-	7.771e+07 (0.065)	-	-
$(d_{t-1}^{td} - d_{t-1}^{s})$	-	-	-	-	-9.354e+07 (0.015)*	-
d_{t-1}^{s}	-	-	-	-	-	2.822e+07 (0.530)
R^2_{Adj}	0.278	0.278	0.271	**0.985**	**0.985**	0.978
AIC	3415.	3415.	3416.	3192.	**3189.**	3220.
BIC	3419.	3419.	3423.	3202.	**3199.**	3228.

Note: * next to p-values indicate significance.

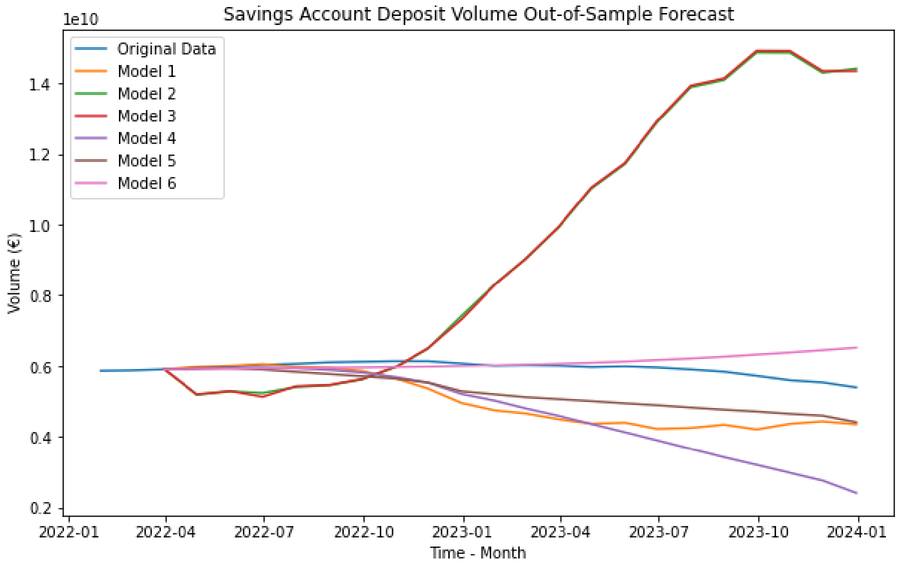

Figure B.4: Out-of-sample forecast of deposit volume of savings accounts

Table B.9: Out-of-sample comparison of savings account volume model performance

Model	R^2_{Adj}	RMSE	MAE
Model 1	-34.14	1.066×10^9	0.830×10^9
Model 2	-578.55	4.426×10^9	3.178×10^9
Model 3	-609.30	4.442×10^9	3.186×10^9
Model 4	-74.32	1.536×10^9	1.154×10^9
Model 5	-16.94	0.750×10^9	0.633×10^9
Model 6	-2.10	0.319×10^9	0.238×10^9

Table B.10: Out-of-sample performance measures of outflow model

Model	R^2_{Adj}	RMSE	MAE
Outflow model	0.4822	2.906×10^7	3.574×10^7

B.5 Migration volume

Table B.11: Coefficients and performance measures of migration volume (spread) - full set

	Maturity 1-year	Maturity 2-year	Maturity 3+ year
Constant	5.638e+06 (0.000)*	-1.981e+06 (0.033)*	-4.056e+06 (0.012)*
$d_{t-1}^{td} - d_{t-1}^{s}$	2.071e+07 (0.000)*	6.973e+06 (0.000)*	2.444e+06 (0.000)*
Vintage	-7.042e+04 (0.035)*	-1.551e+04 (0.165)	3.528e+04 (0.001)*
Top 3	1.47e+07 (0.000)*	3.173e+06 (0.001)*	3.436e+06 (0.071)
R^2_{Adj}	0.804	0.725	0.431
AIC	2767.	2875.	3408.
BIC	2777.	2885.	3419.

Note: * next to p-values indicate significance.

Table B.12: Coefficients and performance measures of migration volume (TD rate) - full set

	Maturity 1-year	Maturity 2-year	Maturity 3+ year
Constant	-4.302e+06 (0.000)*	-4.293e+06 (0.003)*	-2.893e+06 (0.007)*
d_{t-1}^{td}	8.293e+06 (0.000)*	2.353e+06 (0.000)*	2.397e+06 (0.000)*
Vintage	8.26e+04 (0.000)*	5.37e+04 (0.000)*	4.383e+04 (0.000)*
Top 3	1.496e+07 (0.000)*	2.949e+06 (0.024)*	3.222e+05 (0.803)
R^2_{Adj}	0.860	0.490	0.681
AIC	2740.	2931.	3346.
BIC	2749.	2941.	3357.

Note: * next to p-values indicate significance.

Appendix C: VAR model results

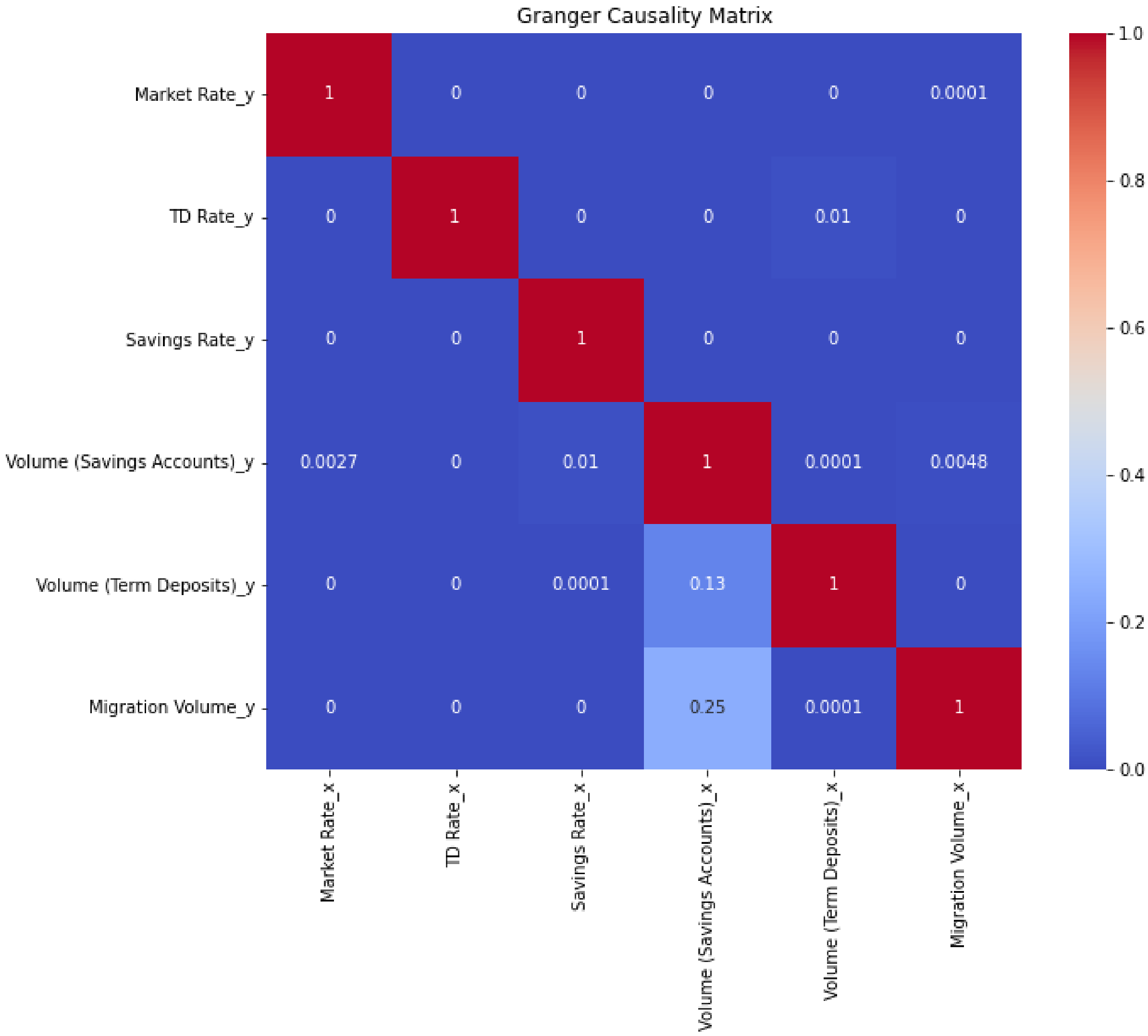

Figure C.1: Granger causality matrix of VAR model time series

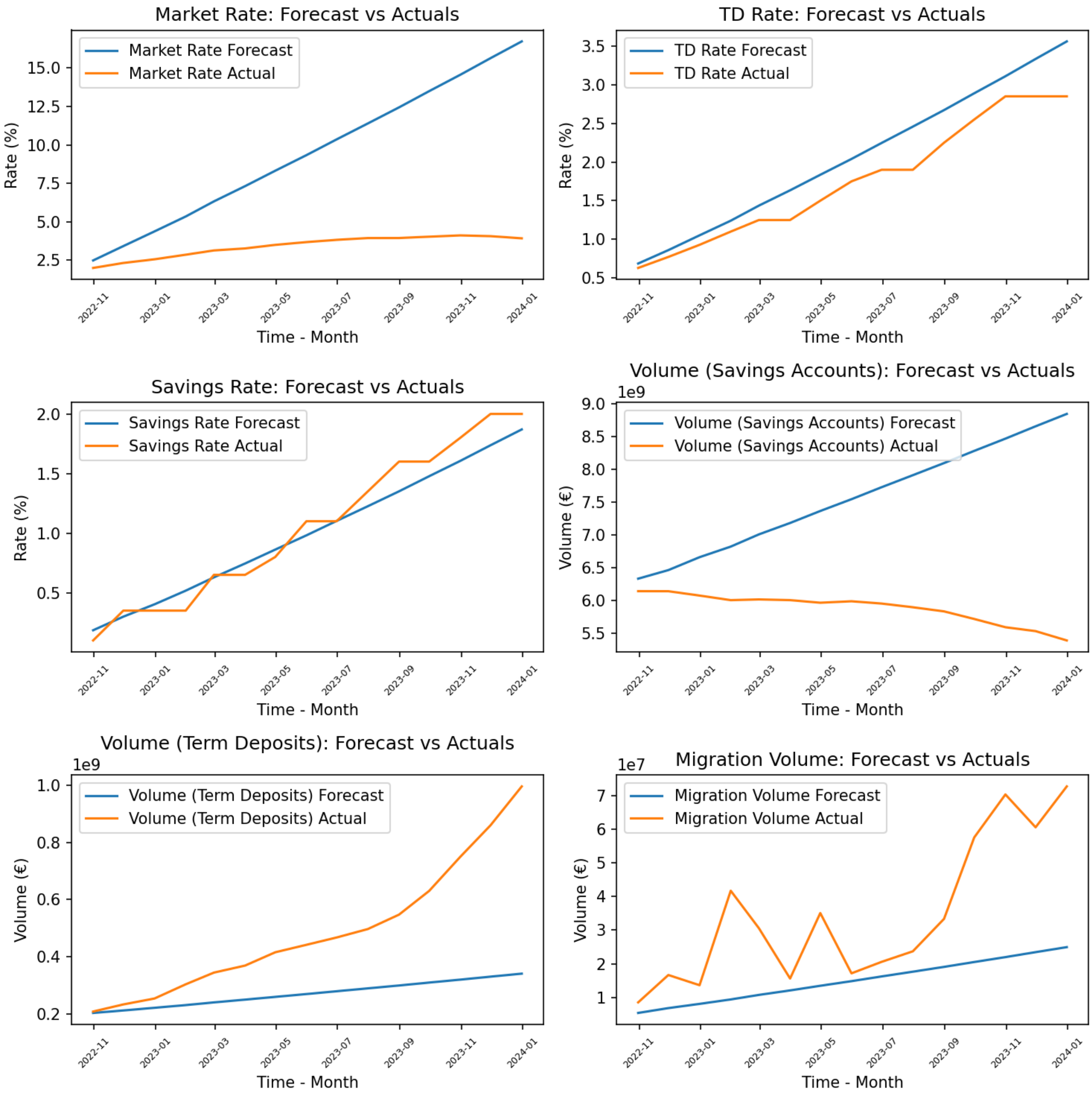

Figure C.2: Out-of-sample forecasts of VAR (1) model

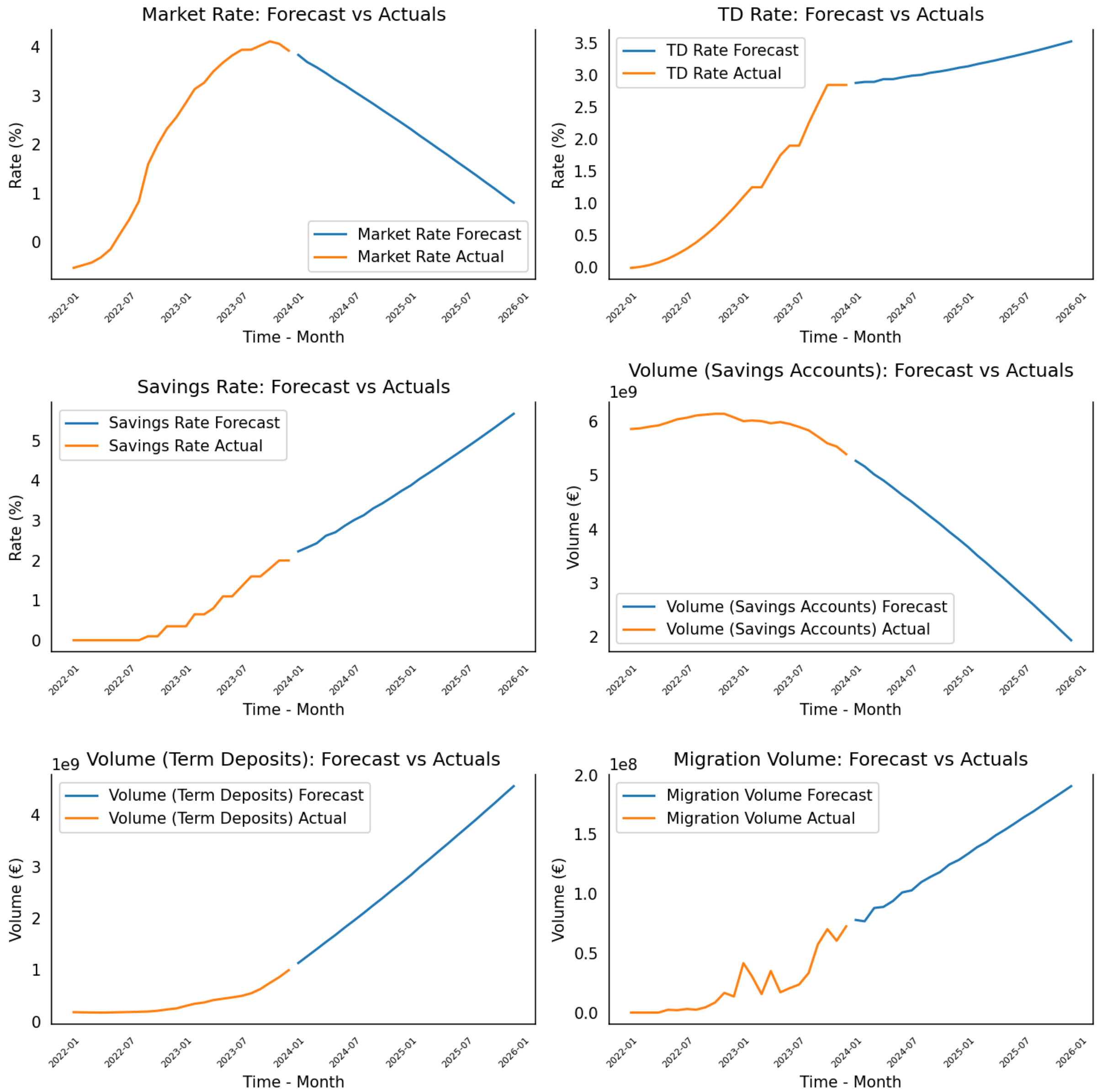

Figure C.3: Forecasts of VAR (1) model

Appendix D: Scenario analysis

D.1 Shocked market rates

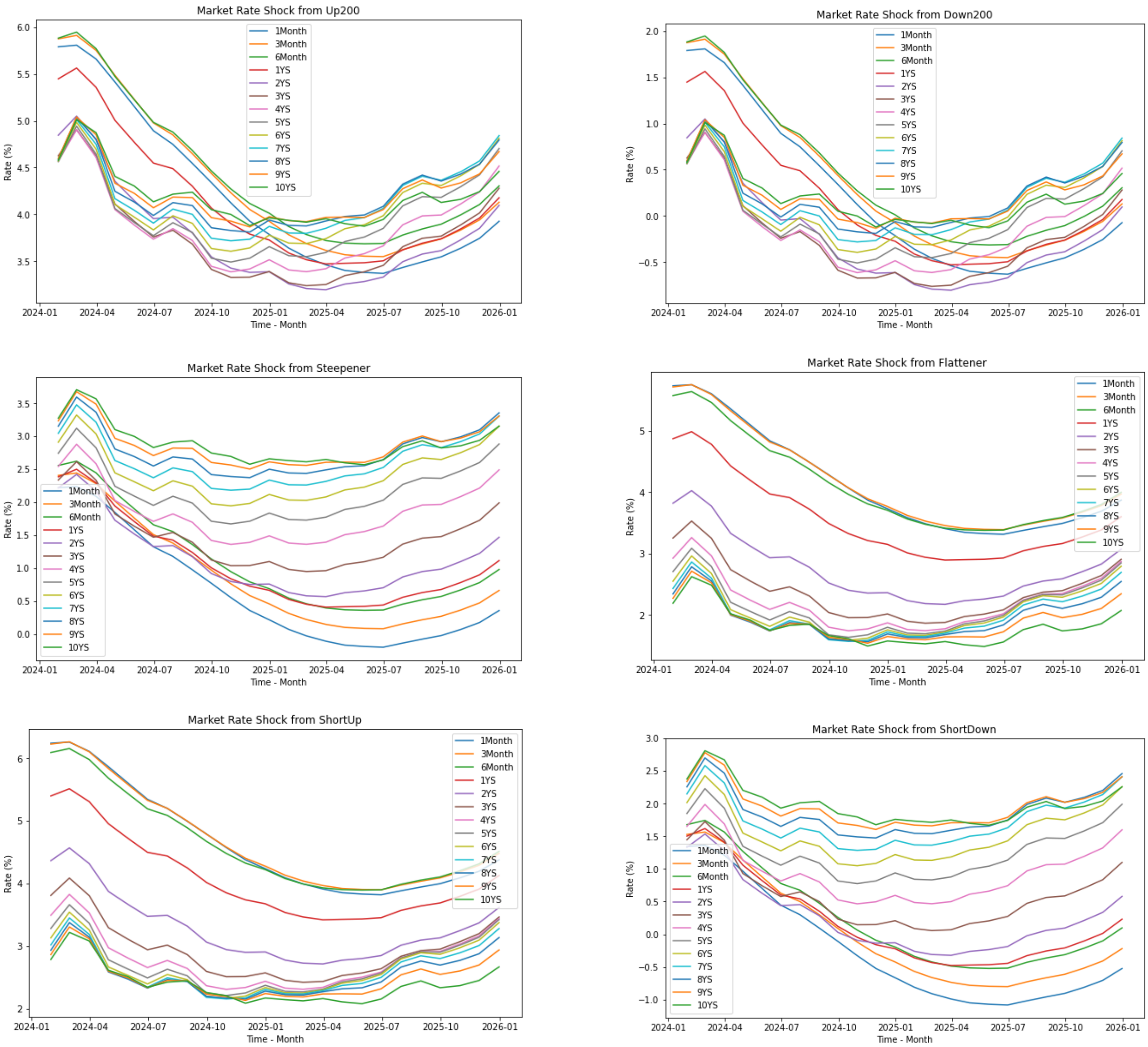

Figure D.1: EBA shock scenario's on market rates

D.2 Shocked term deposit rates

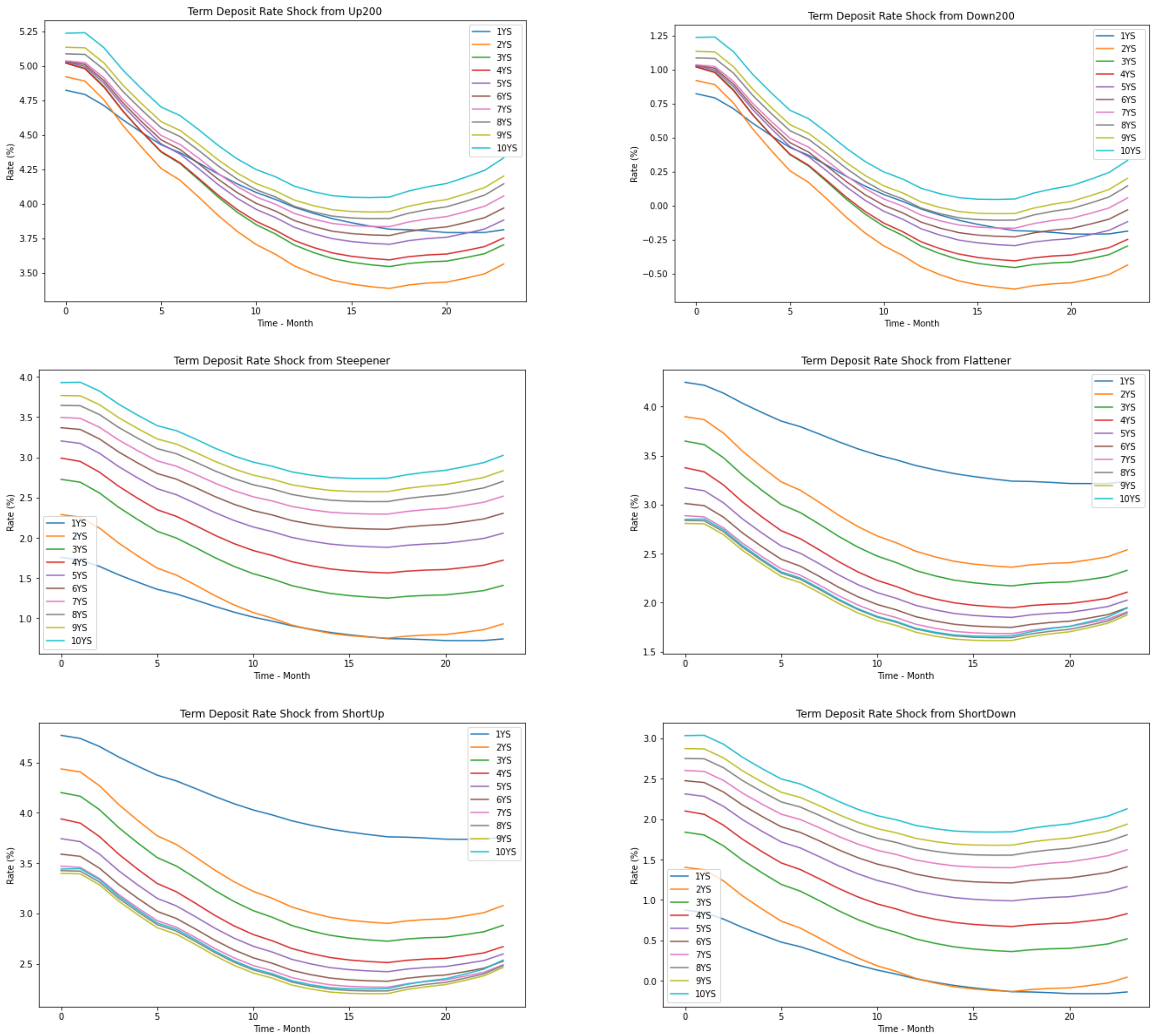

Figure D.2: EBA shock scenario's on term deposit rates

D.3 Shocked savings rates

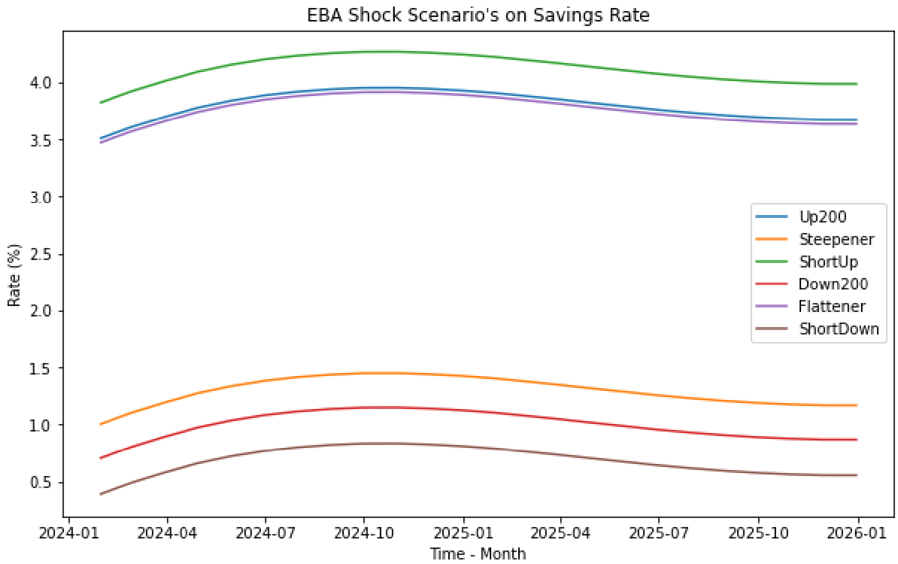

Figure D.3: EBA shock scenario's on savings rate