Modeling Customer Behavior of Non-Maturity Deposits

MIKAELA JENNEROT
Modeling Customer Behavior of Non-Maturity Deposits

MIKAELA JENNEROT

Degree Projects in Financial Mathematics (30 ECTS credits)
Master's Programme in Applied and Computational Mathematics (120 credits)
KTH Royal Institute of Technology year 2019
Supervisors at Handelsbanken: Christian Alexandersson
Supervisor at KTH: Boualem Djehiche
Examiner at KTH: Boualem Djehiche
Abstract

The modeling of non-maturity deposits has become a highly relevant subject in the financial sector since these instruments constitute a significant portion of banks’ funding. A non-maturity deposit may look relatively simple, however, it has features that complicate the handling of these products. This thesis has the purpose of building a model based on the identification, integration and significance level of factors that influence customer behavior related to non-maturity deposits. Moreover, a mathematical approach based on a selection of these factors is made with the aim to analyze client behavior related to these products. The developed model uses simple linear regression and multiple linear regression with dummy variables to model long-term behavior. In contrast to the statistical methods that banks typically apply in this context, this thesis can contribute to the modeling of non-maturity deposits by highlighting customer behavior. Although, the evaluation of the mathematical approach indicates that the model might not be appropriate to use in real practice, it may arise ideas of alternative methods for the handling of non-maturity deposits.
**Sammanfattning**

**Modellering av kundbeteende för icke tidsbunden inlåning**

Acknowledgements

I would like to thank my supervisor at KTH, Boualem Djehiche for the support and advice throughout the thesis. Furthermore, I am grateful to my supervisors Christian Alexandersson and Bianca Dufour Partanen at Handelsbanken for the opportunity, guidance and contribution of valuable inputs. Finally, I would like to thank family and friends for all the support and encouragement during the project.

Stockholm, May 2019

Mikaela Jennerot
List of figures

1. Integration analysis .................................................................................................34
2. A framework for customer behavior of NMDs ..........................................................38
3. Deposit volume evolution .........................................................................................47
4. Trend estimation of female categories .....................................................................48
5. Trend estimation of male categories .........................................................................49
List of tables

1. Identification of internal factors ................................................................. 31
2. Identification of external factors ................................................................. 32
3. Integration analysis .................................................................................. 33
4. Significance level of internal factors ......................................................... 35
5. Significance level of external factors ......................................................... 36
6. Distribution of individuals in the customer categories .......................... 41
7. Multiple regression with dummy variables .............................................. 44
8. Trend coefficients .................................................................................... 50
9. Relative deposit volume changes ............................................................. 51
10. Regression outcomes ............................................................................. 52
1. Introduction

1.1 Background

A non-maturity deposit (NMD) is a deposit that is characterized by not having a specific contractual maturity. Common types of NMDs are saving accounts and current accounts. These financial products may seem relatively simple, however, there are two main features that complicate the modeling of NMDs. The customer has the opportunity to add or withdraw balances at any time and usually at no penalty. Simultaneously, the bank has the optionality to adjust the deposit rate.

After the financial crisis of 2008, NMDs gained a significant portion of banks’ liabilities. Dzmuranova and Teply (2016) describe that the attractiveness of these instruments mainly comes from that they allow for a relatively high return, even in a low interest rate environment. The structure of NMDs, i.e. having unknown future cash flows and uncertain pricing, implies mainly two types of risks: interest rate risk and liquidity risk. It appears that the risk management of NMDs is quite challenging for banks due to the difficulties in making reasonable modeling assumptions.

However, the complexity of NMDs goes beyond their structure. Behind every deposit is a unique customer whose decisions are influenced by a considerable amount of variables. Such factors can be related to depositors’ individual characteristics and their relationship with the bank, but also how they perceive the general macroeconomic environment may have an impact. Baumann et al. (2007) describe that apart from the difficulties of identifying such variables, what might be even more challenging is how to actually measure and translate clients’ intentions and attitudes into real actions.

The aim of modeling NMDs is to obtain realistic estimations of customers’ saving durations, i.e. the time period between deposit and withdrawal of funds. These approximations play a substantial role for determining a bank’s degree of maturity transformation as well as cash flow profile to a large extent. In that respect, banks typically apply statistical methods, for example static replicating portfolio. However, Maes and Timmermans (2005) explain that there are some shortcomings related to this approach, which has led to the development of new and more complex models within many banks such as dynamic replicating portfolio and net present value Monte Carlo simulation. The advantage of these models is that the scenarios are based on present market conditions, which means that they are continuously adjusted throughout the modeling. This enables the bank to adapt much quicker to changes in customer behavior as well as in the market environment. However, the new models appear to be very complex and thus many banks stay with the old and less complicated ones.
In recent years, authorities have increased their attention towards NMDs. The Basel Committee on Banking Supervision (2016) has developed a standardized framework regarding the treatment of NMDs. Furthermore, the European Banking Authority has released guidelines that further describe how banks can implement different procedures when managing NMDs. Interestingly, the 2018 publication of guidelines emphasizes institutions not to completely rely on quantitative results from statistical models, but to also incorporate alternative methods in the modeling of NMDs.

1.2 Problem

The difficulties that banks experience when modeling NMDs are mainly related to the complexity of customer behavior. Every client is unique and his or hers future actions are unpredictable, which complicates the making of proper modeling assumptions. Currently, the most frequently used techniques regarding NMDs are statistical models that use quantitative data, however, these fail to incorporate many of the possible scenarios related to customer behavior. This demands research for development of new models for NMDs, which also is in line with the latest issue of guidelines by the European Banking Authority. Hence, there are two aims for this thesis. Firstly, to identify and analyze factors affecting customer behavior related to NMDs from existing literature. Secondly, based on the findings and the author’s own assumptions, to construct a new model for client behavior of NMDs by the use of a qualitative and mathematical approach. Thus, in addition to current statistical methods, the proposed model could contribute to banks’ management of NMDs by highlighting aspects of customer behavior.

1.3 Research question

How should customer behavior of an NMD be modeled?

In order to answer this question, the following three questions must be answered first:

- What factors affect customer behavior?
- How can the factors integrate with each other?
- For each relevant factor, how significant is it?
1.4 Delimitations

Customer behavior in the banking sector has been widely studied over time and there exist a substantial number of studies related to this subject. Hence, the content and analysis of this thesis will be subject to the reviewed literature. Moreover, this thesis has been written at a large Swedish bank, which has provided the data that has been used. Consequently, the findings and conclusions of this study will depend on the circumstances in this particular setting and thus have limited generalizability.

1.5 Structure of the thesis

Chapter 2 gives an introduction of NMDs and a description of some of the regulations and modeling difficulties associated with these products. Moreover, a brief presentation of established saving theories is given and relevant factors associated with customer behavior regarding NMDs are identified based on existing literature. The last section in chapter 2 presents the mathematical theory applied in the study. In chapter 3, a new model for customer behavior related to NMDs is suggested and motivated. A description of the data and methodology is given in chapter 4. In chapter 5, the results are presented. Finally, chapter 6 discusses the findings of the study and chapter 7 presents the conclusions that have been made and suggestions for future research.
2. Literature and theory

This chapter presents the literature study of this thesis. In section 2.1 the fundamentals of NMDs are explained, followed by a description of regulations and guidelines related to NMDs in 2.2. Section 2.3 goes thorough basic saving theories and factors affecting bank customer behavior are described in 2.4. Lastly, section 2.5 presents the theory of the mathematical models used in the study.

2.1 Non-maturity deposits

Non-maturity deposits are complex financial products with stochastic cash flows and no contractually agreed maturity date. Customers can freely withdraw account balances at any time and banks possess the opportunity to adjust the deposit rate. According to Schlüter et al. (2015), these features turn them into instruments with embedded optionality, exposing banks to Asset/Liability Management (ALM) risk. The main types of financial risks associated with NMDs are interest rate risk and liquidity risk.

The interest rate risk arises from the uncertain level of future market rates. Maes and Timmermans (2005) express that NMD volumes usually decrease with high market rates and increase with low market rates. The authors describe two scenarios in the case of increased market rates. Firstly, banks may partially raise deposit rates in response to higher market rates, which might lead to customer withdrawals if alternative investments yield higher return. Secondly, if deposit rates are fully changed in reaction to increased market rates, banks will face major costs related to all existing deposit balances. Dzmuranova and Teply (2016) highlight that in reality, deposit rates are not solely derived from market rates, but also from competitors’ pressures and customers’ expectations. Thus, banks have to incorporate these factors as well in the interest rate risk management of NMDs.

The liquidity risk is caused by the fact that future account balances are unknown, which leads to difficulties regarding deposit volume predictability. The importance of making reasonable estimates of future account balances comes from that banks usually reallocate the core of saving deposits into long-term investments to hopefully earn a positive interest rate margin on this (Dzmuranova and Teply, 2016). Thus, for banks to achieve positive maturity transformation, they must ensure that there are enough available funds to cover unexpected withdrawals of NMDs. This implies that especially banks with high confidence in deposit funding are in a risky position. The liquidity risk management of NMDs can be considered way more complex in comparison to contracts with predetermined maturity.
NMDs have zero contractual maturity, however, the effective duration, i.e. the duration calculation for instruments having embedded options, is much higher (Dzmuranova and Teply, 2016). The effective maturity measures the sensitivity of deposit rates and balances in contrast to market rate changes. Maes and Timmermans (2005) point out that the effective maturity differs a lot between banks. Possible explanations for this are differences in banks’ reasoning when determining deposit rates, customers’ saving behavior, overall interest rate environment and banks’ use of different modeling methods and assumptions.

To estimate saving durations related to NMDs, banks usually apply statistical methods. Straßer (2014) states that the most established approach among banks is static replicating portfolio, which is based on transforming non-maturing liabilities into a portfolio of fixed-income assets with known maturity. However, this type of model uses one historical scenario only as input and therefore it may fail to produce relevant predictions about future deposit behavior. This makes the model especially vulnerable from a risk perspective.

Maes and Timmermans (2005) explain that based on the limitations of static replicating portfolio models, an increased number of institutions have adapted more complex methods, such as dynamic replicating portfolio and net present value Monte Carlo simulation. In contrast to the static versions, these use numerous scenarios to simulate the risk factors, making them considerably flexible and accurate. They can also adapt to future changes in the variables, which makes them even more advantageous. The problematic with these advanced models is that some banks find them complicated to apply.

Despite that statistical methods are continuously being improved, they still have deficiencies. The fact that there are substantially many uncertain factors affecting depositors’ behavior makes it difficult to identify those that are most relevant as well as to make proper assumptions about them. This demands for alternative models for NMDs that can highlight aspects of customer behavior.

2.2 Regulations and guidelines

In 2016, the Basel Committee on Banking Supervision (BCBS) proposed a document regarding the treatment of financial products with behavioral optionality. This paper includes a standardized framework on how banks can measure and manage the interest rate risk arising from the banking book (IRRBB) related to NMDs. The banking book consists of securities that are expected to be held until maturity and includes activities related to customers’ loans and deposits. It also involves the derivatives with the purpose of hedging exposures related to banking book activities, such as IRRBB.

The BCBS divides IRRBB into three sub-components; gap risk, basis risk and
option risk. Regarding option risk, it is further split up into automatic options and behavioral options. These options are embedded in the sense that the bank or its client can change the level and timing of cash flows at any time. Automatic option risk includes the financial aspects related to the risk arising from interest sensitive instruments, whereas behavioral option risk involves the risk associated with customer behavioral optionality.

According to BCBS’s framework, institutions should separate their NMDs and divide them into retail or wholesale categories based on the deposits’ characteristics. Retail deposits cover individual investors and small business customers, where the deposit is less than €1 million, while remaining deposits belong to wholesales. Furthermore, retail deposits are divided into transactional or non-transactional categories, depending on the transactional rate.

The bank should then identify the stable and non-stable portions in each NMD category. Based on historical data that exhibit volume changes over the past ten years, the stable part constitutes of deposits that are not likely to be withdrawn. The stable part is further segmented into core and non-core deposits, where core deposits have little likelihood of being repriced to changes in market interest rates. In contrast to core deposits, the non-core deposits are very likely to reprice to fluctuations in market interest rates. Lastly, the core NMDs should be slotted into the proper time bucket, whereas the non-core deposits should be slotted into the overnight time bucket.

Another contributor to the treatment of NMDs is the European Banking Authority (EBA). In 2018, EBA published guidelines with the aim to advice institutions on how to more practically identify, evaluate and manage IRRBB. In similar manner to BCBS, EBA defines the three sub-components of IRRBB, and further specifies behavioral options as:

“The volume of mortgages, current accounts, savings and deposits where the customer has the option to deviate from the contractual maturity; the volume of commitments with interest rate sensitive customer drawings.” (European Banking Authority, 2018)

This description refers to the fact that account balances and saving durations of contracts with embedded optionality, such as NMDs, are exposed to behavioral option risk. With other words, NMDs are subject to customer behavior. Furthermore, in the context of behavioral modeling assumptions for accounts without specific repricing dates, it is stated that institutions should

“Not exclusively rely on statistical or quantitative methods to determine the behavioural repricing dates and the cash flow profile of NMDs. Further, the determination of appropriate modelling assumptions for NMDs may require the collaboration of different experts within an institution (e.g. risk management and
The guideline indicates that apart from using statistical methods, banks should consider to implement alternative modeling approaches for NMDs. Moreover, institutions are emphasized to involve several departments for the determination of modeling assumptions.

2.3 Saving theories

When studying customer behavior of NMDs, it is of interest to briefly examine the concept of saving. Over the years, various types of saving models have been suggested through theoretical and empirical approaches. Fisher and Montalto (2010) express that the perspectives of saving vary among different disciplines, where one viewpoint is the psychologist’s perspective. It considers saving as a decision making process and as an act of putting aside capital for a certain purpose. Typically, the focus is on a certain type of factors such as socioeconomic variables, e.g. age, education and occupation, or personality traits, e.g. self-control, saving horizon and risk aversion.

Furthermore, Fisher and Montalto (2010) mention the economist’s view of saving, which defines saving as the difference between the net worth at the end of a period and the net worth at the beginning of the period. This perspective usually emphasizes the following three types of saving motives: contractual, discretionary and residual saving. A monthly deposit of money is a type of contractual saving while a onetime transaction is considered as discrete saving. Residual saving is simply the remaining part of the income after consumption expenditures over a specific time period.

Eriksson and Hermansson (2014) express that one of the most commonly used saving theories is the life-cycle hypothesis (LCH), which concentrates on customers’ tendency to consume as well as their liquidity preferences. The theory includes three types of saving motives; the transactions motive, the precautionary motive and the speculative motive. One perspective of LCH is based on consumers’ age and saving horizon, where saving behavior is described as individuals’ attempt to optimize their resources evenly over their expected remaining lifetime. Here, the focus is on current and future income and the present consumption level is considered as a part of lifetime wealth, where individuals behave as if they had perfect knowledge about the future.

The authors continue with the relative income hypothesis, which is another perspective of the LCH (Eriksson and Hermansson, 2014). It concentrates on current and past income or consumption levels, where individuals are believed to compare their present earnings with some moment in the past. Another view of the
relative income hypothesis is that individuals compare their income and consumption levels with that of other people, rather than thinking in absolute terms.

Furthermore, Eriksson and Hermansson (2014) underline one of the key aspects in the LCH, which is the assumption that individuals try to maximize consumption utility over their lifetime. However, the future is still unknown and one disadvantage with the LCH is that it discounts the fact that individuals’ saving motives may vary over time. In fact, people’s saving objectives tend to depend on different stages in life, instead of being based on a complete life cycle.

2.4 Factors affecting bank customer behavior

In banking, customer behavior has been widely studied over time. Consistent among many of the researches is that modeling client behavior is highly complicated since the behavioral analysis reveals a major system of factors that influence customers’ saving decisions. The complexity increases with the realization that these variables are not completely independent of each other. Moreover, it is unclear how to actually translate customers’ intentions and attitudes into real actions.

Next, factors that may affect clients’ saving behavior will be identified and described based on the reviewed literature. According to the nature of the variables, they are divided into either one of the two categories: internal (individual) factors or external factors. While customer behavior is commonly denoted as the dependent variable, the other factors can be referred to as independent variables. The content of the two factor categories are first briefly presented, followed by detailed descriptions of their included components.

2.4.1 Internal factors

Internal factors with regard to saving behavior are related to customers’ individual characteristics. Such variables can further be divided into socioeconomic factors and personality traits. In this context, age, gender, income and education appear to be some of the most relevant socioeconomic variables. Baumann et al. (2007) describe that socioeconomic factors are usually available for banks, which turn them into important components in the behavioral analysis. The authors explain that banks can actually use this knowledge in marketing purposes to actively target specific client segments.

Customers’ personality traits with regard to saving behavior, such as financial literacy, saving motives and saving horizon, can be related to individuals’
intentions and attitudes towards saving. While socioeconomic factors appear to be relatively independent, personality traits are to a large extent influenced by socioeconomic variables. Although, clients’ future actions are unpredictable, previous studies have been able to conclude certain patterns regarding the effect of socioeconomic factors and personality traits on saving behavior. The following paragraphs present a more detailed description of customer behavior related to internal factors.

Age

Eriksson and Hermansson (2014) describe that age is related to saving behavior as people have different motives and needs during their life cycle. Saving motives are in turn influenced by people’s ability and willingness to save, which means that these vary over time as well. For instance, saving for retirement is usually more relevant for middle-aged individuals than for young and elderly people. Furthermore, Wolff (2000) shows that people overall tend to save less the older they get, i.e. deposit balances decrease with aging.

Interestingly, age appears to affect customers’ choice of banking with one or several banks. Baumann et al. (2007) find that clients in their fifties and sixties tend to spread their banking activities, whereas customers in their retirement mainly use one bank only. According to the authors, a possible explanation for this behavior is that individuals gain banking experience and financial literacy, but also tend to be wealthier as they become older, and may find a more optimal economic solution by combining several suppliers. However, at retirement customers might want to simplify their banking errands by using one bank only. Baumann et al. (2007) express that with this knowledge; banks can actively target customers in the ages before they start spreading their deposits by offering attractive products that make them less interested in adding further suppliers.

Gender

Hermansson (2017) discusses that potential saving behavioral differences with respect to gender could be related to that men tend to dominate regarding income level and degree of establishment on the labor market. In general, men have higher and better-paid positions within companies, which increase their abilities to save for growth. On the other hand, women are likely to save out more for daily-expenses due to lower income. Women also appear to prioritize to save for retirement, possibly because they tend to have longer life span than men in general. Hence, the different prerequisites for men and women allow for different types of savings.
Furthermore, Baumann et al. (2007) show that women are more likely to use one bank only, while men tend to spread their banking activities among several institutes. Possible reasons for this behavior could be that women in larger extent lack financial literacy and have lower interest in finance and thus use a single bank for the sake of simplicity, whereas men more actively compare the pricing of products between different banks. Based on this, the authors suggest that banks either focus on maintaining their female clients or put more effort into attracting male customers.

Income

Customer behavior depends on individuals’ income since it influences the lifestyle as well as the ability to save. Fisher and Montalto (2010) describe that consumption and saving decisions depend on the level of income at different points of time in life rather than on lifetime income. Overall, people with high income usually have higher saving rates and save more regularly in comparison to individuals that earn a low income. The authors further describe that the income level also tends to affect saving horizons, where high-income earners have increased ability to save for longer time periods. On the other hand, individuals with low income tend to plan their budgets for shorter saving horizons. In similar manner, high-income earners are more likely to save for growth, while a greater portion of low-income earners’ salary is used for daily consumptions. Middle-income earners show to prioritize to save for potential emergencies. Since individuals’ salaries tend to vary throughout life, saving motives do as well.

Baumann et al. (2007) find that income level is related to the number of banks customers use. Apparently, high-income earners are more likely to have several banks, whereas clients with lower income usually have one bank only. However, the authors mention that customers that are provided with the right type of products might not involve additional banks. Similar to this, Diepstraten and van der Cruijsen (2017) show that customers with higher income switch from their first-ever bank in greater extent than low-income earners.

Education

Education affects customer behavior in a similar way to income. Typically, a longer education is associated with higher income and vice versa, which means that highly educated people usually have greater ability to save. Hermansson (2017) finds that people with higher level of education also tend to be more involved while making financial decisions and are more likely to use professional advice. On the other hand, individuals with low education rely on informal sources of financial advice, e.g. family and friends, to a greater extent. Furthermore, Baumann et al. (2007)
show that customers without a university education are more likely to use one bank only, whereas educated clients spread their banking businesses among several banks in order to improve their financial situation.

Financial literacy

Financial literacy has a significant impact on saving behavior since it tends to affect both the ability and the willingness to save (Eriksson and Hermansson, 2014). People with little financial knowledge are less likely to engage in financial planning, compared to people with high financial literacy. Thus, financial illiteracy may decrease the likelihood of saving, but it can also lead to that individuals make unwise economic decisions, for instance by taking on too high financial risks or unsustainable loans. In turn, such decisions might lead to insufficient savings for investments or pensions, indebtedness or impaired financial security.

Furthermore, low financial knowledge tends to be more common among women, elderly and people with low education and income (Wieselqvist Ekman, 2015). These categories of people are usually more risk-averse, which also has negative impact on potential return. However, customers whose financial knowledge is strengthened can potentially increase their interest in finance, which might overall benefit household economies (Eriksson and Hermansson, 2014). Additionally, Baumann et al. (2007) express that experienced investors who feel like they understand the market dynamics are more likely to spread their banking activities instead of using one bank only.

Saving motives

According to Fisher and Montalto (2010), people tend to have many saving motives during their lifetime. These usually vary over time in line with individuals’ life cycle, rather than being based on a lifetime perspective. Furthermore, Eriksson and Hermansson (2014) describe that saving motives may differ among people depending on personal characteristics and financial situations, but also on individuals’ expectations on market developments. Precaution, retirement and children’s needs tend to be some of the most prioritized saving motives and these significantly increase the likelihood of saving.

Saving horizon

The saving horizon can be used to forecast customers’ saving decisions. Fisher and Montalto (2010) express that medium to long saving horizons tend to increase the
likelihood of saving, compared to short saving horizons, and people who plan their saving needs well in advance usually are able to save more. Starting to save at a late point in life may decrease the probability of achieving certain investing goals in life and may also lead to insufficient savings during retirement.

2.4.2 External factors

External factors’ affect on saving behavior can mainly be linked to the bank and the economic environment. Bank-related factors are associated with customers’ experience at the bank, but also the bank’s behavior and reactions to market events. Moreover, how clients perceive the service quality in terms of products, availability and advice tend to significantly affect customers’ propensity to withdraw funds or switch to another bank. In fact, specific relationship characteristics could be used to predict switching behavior such as length of the relationship, if the customer uses several services at the bank and if the customer has several suppliers.

Furthermore, while the increased use of digital banking improves service availability, it also tends to distance the bank-customer relationship, making it less personal. The use of financial advisory can actively strengthen the link between the bank and the client and also has the possibility to improve customers’ financial literacy. Moreover, in order to attract and maintain clients, banks have to incorporate their competitors’ offer and pricing of products. The bank may for instance focus on a specific customer segment or offer pricing incentives to entice clients to keep their deposits. Switching costs tend to be an efficient strategy to retain customers since they may affect market competition by creating lock-in effects.

Macroeconomic factors may influence customers’ overall ability to save. For instance, when stock index decline, people become more pessimistic about the future, which implies reduced consumption and increased savings. Financial market dynamics also have an impact on saving behavior in the sense that they affect customers’ propensity to invest in securities versus keeping the money on saving accounts. Lastly, seasonal variations tend to affect account balances as well as cash flow volatility. The following paragraphs present a more detailed description of customer behavior related to external factors.

Length of relationship

The length of the relationship, i.e. for how long a person has been a bank’s customer, influences clients’ tendency to switch to another bank. Diepstraten and van der Cruijsen (2017) describe that long relationships indicate strength and trust
and may imply less probability that customers change bank, whereas clients with short relationships are more likely to switch to another bank. Ruzena and Tomas (2014) show that clients who feel that their bank is stable, reliable and congenial are more likely to stay than those who don’t have these feelings. Moreover, Baumann et al. (2007) explain that customers with longer relationships tend to use more services at the bank and perhaps also deposit more money.

**Number of services and suppliers**

The number of current services and suppliers customers have affect their likelihood of withdrawing funds as well as their propensity of switching to another bank. Diepstraten and van der Cruijsen (2017) show that clients who bank with a single bank and use several services at that bank tend to be more loyal and less likely to switch than others. These customers also tend to have higher deposit balances. Furthermore, Baumann et al. (2007) explain that the more convinced a customer is regarding how a potential bank switch could economically benefit his or hers financial situation, the more likely is the client to use several suppliers.

Baumann et al. (2007) further express that banks should be able to determine a customer’s number of suppliers by analyzing the client’s cash flow between different accounts, but also through conversations with the client. This information could be used in order to target customers who have several suppliers, for example by offering them some type of pricing incentive if they consolidate all their banking activities to the main bank.

Recent tendency of behavior for deposits shows that an increased number of clients acquire an additional bank instead of switching from their present one (Svenskt kvalitetsindex, 2018). Although, customers who bank with a single bank usually are more satisfied than those who use several banks, recent statistics prove that client satisfaction is increased among those with several suppliers. This indicates a development of products and services that attract customers that use a variety of banks.

**Financial advisory**

Hermansson (2017) shows that the use of financial advisory tends to significantly affect customer behavior. Apparently, those who seek to consult with a financial advisor typically have certain saving motives and seem to value the future highly. For example, customers who have strong motives to save for wealth and retirement are more likely to use a financial advisor than clients who save for precautionary reasons. The author suggests that customers are more likely to use a financial advisor when it concerns savings with longer horizons since these involve more
uncertainty. Moreover, clients who are middle-aged or retired, high-income earners and have little financial literacy have increased probability of using a financial advisor, compared to customers who are younger, low-income earners and have high financial literacy.

Furthermore, Eriksson and Hermansson (2014) describe that the characteristics of the relationship between the customer and the financial advisor affects saving behavior as well. A strong and close relationship is a result of frequent meetings and good understanding of each other, but also an indication of loyalty and trustworthiness. Customers with high confidence in their advisors and who are committed in the meetings have the possibility to strengthen their financial knowledge and are more likely to stay with their current bank. According to the authors, clients’ loyalty could be measured by the number of advisors or suppliers the customer has simultaneously, which in turn can show the propensity of switching to another bank. Typically, loyalty is higher among customers who are older and who have lower income and education degree. On the other hand, advisors can show loyalty by continuously providing clients with updated financial information as well as by giving customized advice. However, not all advisors act properly due to e.g. provisions, but the risk of misleading can be considered decreased with existing regulations.

When comparing the use of a financial advisor, Hermansson (2017) finds no statistically significant difference with regard to gender. However, the author notes that the probability of using financial advisory differs between men and women based on the significance level of different saving motives. For men, saving for wealth, retirement and rainy days are motives that increase the likelihood of using a financial advisor, whereas for women, only the retirement motive is important.

_Digital banking_

Banking is becoming more and more digitalized, which increases customers’ self-service opportunities (Svenskt kvalitetsindex, 2018). Internet banking enables clients to take more control over their economy and makes it possible to transfer money between different accounts anytime and anywhere. The online sites are also great sources of information, giving clients easy access to financial product data. This enables customers to effectively compare the pricing of instruments set by different banks. Hence, digital banking increases customers’ sensitivity towards changes in deposit rates.

In general, people’s attitudes towards Internet banking is becoming more positive, even among elderly people, since online services are continuously being improved and become easier to manage (Svenskt kvalitetsindex, 2018). However, for some services, e.g. taking out mortgage loans, plenty of customers still prefer physical
meetings. This implies that a major challenge for banks is how to provide clients with technical services that facilitate digital banking and at the same time keep the relationship with the customer close and personal. Especially banks that focus solely on online services tend to attract people with positive attitudes towards digital banking.

**Switching costs**

Switching costs can be thought of as a onetime cost that applies to customers who change from one bank to another. Although, monetary costs may be the dominated form of switching costs, there exist other types based on psychosocial-, effort- and time aspects. Sanjeepan (2017) describes that switching costs can be related to the time and effort customers have to spend in the switching process and to learn a new system. There are also risks associated with bank switching, for instance that the new bank does not fulfill the customer’s expectations or potential disturbance of current business operations during the transition period.

The author also expresses that switching costs is an important aspect concerning banking competition, since their presence can lead to that customers stay with their current bank even though they are dissatisfied. Consequently, they can create a lock-in effect for clients, which in turn will have an impact on market competition.

**Pricing incentives**

Schlueter et al. (2015) show that banks can intentionally try to affect customers’ saving behavior through the offering of different pricing incentives, e.g. interest bonuses. A contract could be constructed such that the customer receives a reward if he or she keeps the money on the account for a certain time period. However, the customer can freely withdraw the deposit within the time period since the contract represents an NMD, but then he or she will not receive the bonus. The authors explain that the main reason behind the introduction of pricing incentives is to increase clients’ saving persistence. This can in turn lead to reduced cash flow volatility and thus provide the bank with more stable deposit funding.

**Deposit rates**

Changes in deposit rates in reaction to market rate developments may affect customers’ decisions and have a major impact on account balances and saving durations. Although, depositors’ behavior is affected by idiosyncratic events, Maes and Timmermans (2005) describe that these are for the most part diversifiable
across clients. In fact, deposit volume dynamics are mostly driven by overall market conditions. However, Dzmuranova and Teply (2016) express that NMDs are more or less interest rate sensitive depending on the type of account. Market interest rate movements tend to affect deposit rates on saving accounts, but not on current accounts.

Furthermore, changes in deposit rates and in deposit balances tend to be relatively stable and behave in a sluggish way. Maes and Timmermans (2005) describe that deposit rates are fairly sticky since banks usually do not adjust them fully and immediately in accordance with market rate changes. The authors explain this by when deposit rates are being repriced, it applies to all outstanding balances. The increased deposit rates will lead to higher costs for the bank, which can turn out quite expensive if there is a large volume of deposits. On the other hand, if alternative investments become sufficiently attractive, depositors might withdraw their balances partially or fully, which is also costly for the bank. Therefore, banks have to consider both volumes and repricing effects at the same time in the interest rate risk management of NMDs.

Maes and Timmermans (2005) continue with explaining that the sluggish behavior of deposit balances comes from that a substantial part of these volumes are usually not very rate sensitive. This portion is commonly referred to as the core deposit since it is not expected to be withdrawn in the short-term. Reasons for the sluggish behavior of deposit balances is mainly linked to switching costs.

**Macroeconomic factors**

Macroeconomic factors, such as economic outputs, inflation, tax levels, gross domestic product and unemployment rates broadly affect the economy on national level. Customers’ expectations about the general macroeconomic situation influence their attitudes towards saving and consuming (Eriksson and Hermansson, 2014). Negative macroeconomic events are likely to imply economic instability, which reduces individuals’ motivation to consume, while saving increases. On the other hand, positive macroeconomic factors indicate economic growth, which escalate consumer spending and make other investments than saving more attractive.

**Financial markets**

The outlook of financial markets, where the trading of equities, bonds, currencies and derivatives take place, influences individuals’ overall motivation of saving money on deposit accounts versus investing in securities (Eriksson and Hermansson, 2014). Although, financial market prices may have uncertainties and
perhaps do not reflect true values of instruments, they still significantly affect customers’ economic decisions. For example, when macroeconomic factors points towards economic growth, individuals are usually willing to take on more risk and invest in securities, instead of keeping the money on saving accounts.

**Competitive environment**

The financial industry is highly competitive, which puts great pressure on banks. Similar to that customers differ from one another, banks are specialized in different areas and attract clients through that. Dzmuranova and Teply (2016) describe that larger banks, also known as traditional banks, typically offer a broad range of services that customers can assess through digital tools and at local bank branches. These customers are usually tied up to several products at the same bank, perhaps because they get better terms that way or to spare time by having one bank relation only. In turn, these contracts provide the banks with stable variations of funding sources and the relatively high fees these customers are charged with are also an important source of income.

In recent years, a new type of bank, known as low-cost banks, has gained market share (Dzmuranova and Teply, 2016). Compared to traditional banks, these operate completely or almost completely via the Internet and typically do not offer services at branches. In that way, they can save on major costs and attract customers by offering low-cost deposit products with higher interest rates. The funding of these banks typically relies heavily on NMDs in liabilities, which result in a higher cost of funding. Consequently, these banks need to attain appealing levels of deposit rates in order to attract new customers as well as to retain the existing ones.

**Seasonality**

Wolff (2000) describes that account balances of NMDs are affected by seasonal variations. At certain times of the year, for example during Christmas holiday, people tend to spend more money such that more transactions are being made. This implies higher cash flow volatility, which usually requires more liquidity and higher account balances. However, Dzmuranova and Teply (2016) mention that variations of NMD balances also depend on the specific type of account. Current accounts, which are mostly used for transactional purposes, exhibit a pattern where balances usually increase when salaries come, followed by a decrease during the remaining part of the month. Thus, these accounts are not that difficult to predict. On the other hand, saving accounts are way harder to forecast since these are more long-term deposits. Whether a customer adds or withdraws an amount next month, next year or in ten years is unknown. This obviously makes liquidity risk management more complicated for saving accounts compared to current accounts.
2.5 Mathematical theory

This section presents the mathematical theory used in the study. The focus is on linear regression analysis and the statistical outcomes that are used for the evaluation of the fitted model.

2.5.1 Simple linear regression

Simple linear regression means that the model has a single explanatory variable. Suppose there are \( n \) observations of data points \( \{y_i, x_i\}_{i=1}^{n} \). The general form of the simple linear regression model is then given by

\[
y_i = \beta_0 + \beta_1 x_i + \varepsilon_i
\]

where \( y_i \) is the dependent, response or target variable and \( x_i \) is the predictor or independent variable. The meaning of a linear model refers to the linearity in the predictor variable. Note that the dependent variable \( y_i \) is a sum of two parts, a deterministic part and a random error term, \( \varepsilon_i \), which is assumed to be \( N(0, \sigma^2) \)-distributed with mean zero and variance \( \sigma^2 \). By using an ordinary least squares (OLS) method, the aim is to find the estimated values \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) of the unknown regression coefficients \( \beta_0 \) and \( \beta_1 \) such that the sum of squares of deviation

\[
\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2
\]

is minimized. This means that the obtained regression equation minimizes the differences between the observed and predicted values of the dependent variable \( y \). The fitted line is defined as

\[
y_i = \beta_0 + \beta_1 x_i
\]

where \( \hat{y}_i \) denotes the predicted value of \( y_i \) for a given \( x_i \) when \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) are determined (Draper and Smith, 1998).

2.5.2 Multiple linear regression

Multiple regression is an extension of simple linear regression since it involves two or more explanatory variables. Suppose there are \( n \) observations of data points \( \{y_i, x_{i1}, x_{i2}, \ldots, x_{ik}\}_{i=1}^{n} \) and \( k \) number of predictor variables. The general form of the multiple linear regression model is then given by

\[
y_i = f(x_{i1}, x_{i2}, \ldots, x_{ik}) + \varepsilon_i = \beta_{i1} x_{i1} + \beta_{i2} x_{i2} + \ldots + \beta_{ik} x_{ik} + \varepsilon_i
\]

where \( y_i \) is the response variable, or regressand, and \( x_{i1}, x_{i2}, \ldots, x_{ik} \) are the predictor variables, or regressors (Green, 2003). Thus, the notation \( x_{ik} \) refers to a predictor \( k \).
with input \( i \). As in the first order model, the dependent variable \( y_i \) is a sum of two parts, a deterministic part and a random error term, \( \varepsilon_i \), which is assumed to be \( N(0, \sigma^2) \)-distributed with mean zero and variance \( \sigma^2 \). By using an OLS method, the aim is to find the estimated values \( \hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k \) of the unknown regression coefficients \( \beta_0, \beta_1, ..., \beta_k \) such that the sum of squared residuals

\[
\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2
\]  

is minimized. The fitted line is defined as

\[
y_i = \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}
\]

where \( \hat{y}_i \) denotes the predicted value of \( y_i \) for given values of \( x_{i1}, x_{i2}, ..., x_{ik} \) when \( \hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k \) are determined (Chatterjee and Simonoff, 2013).

### 2.5.3 Regression statistics

**Regression coefficients**

The regression coefficients describe the correlation between the dependent variable and respective independent variable, i.e. in what way changes in the predictors are associated with changes in the target variable. For every incremental change in the independent variable, the coefficient tells how much the mean of the dependent variable increase or decrease (Green, 2003).

**R-squared, \( R^2 \)**

R-squared, which is also referred to as the coefficient of determination, is a goodness-of-fit measure for linear regression models. It measures how well the regression line fits the data and is defined as the percentage of the variance in the dependent variable that is explainable by the independent variables,

\[
R^2 = 1 - \frac{SS_{Res}}{SS_T}
\]

The total sum of squares for the dependent variable, \( SS_T \), is given by

\[
SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2
\]

where the values of \( y_1, ..., y_n \) denote the observations of the dependent variable and the term \( (y_i - \bar{y}) \) are the deviations from its mean. Furthermore, the mean of the observed data, \( \bar{y} \), is defined as

\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.
\]
The residual sum of squares, $SS_{Res}$, is given by

$$SS_{Res} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where the difference between the observed value $y_i$ and the fitted value $\hat{y}_i$ denotes the residual.

R-squared indicates the strength of the relationship between the model and the independent variables by taking values between 0% and 100%. 0% means that the model can explain none of the variability in the response variable around its mean, which would be a very bad fit. 100% indicates that the model can explain all of the variability in the response variable around its mean and thus would be a very good fit.

R-squared is a relative measure that does not tell anything about the actual precision of the estimations. Thus, a low value of R-squared does not necessarily have to imply a bad model and a high value does not automatically indicate a good model. Also, the number of independent variables in the model affects the value of R-squared. The more predictors that are included, the more R-squared increases, but it never decreases. Consequently, a model might seem to have a better fit than it actually has because it involves many terms (Chatterjee & Simonoff, 2013).

**Adjusted R-squared, $R^2_a$**

In contrast to R-squared, the adjusted R-squared takes the number of predictor variables in a model into account. This means that the adjusted R-squared defines the percentage of variation that is explainable by only the independent variables that actually affect the dependent variable. Thus, the adjusted R-squared is always less than or equal to R-squared. It is defined as

$$R^2_a = R^2 - \frac{k}{n-k-1}(1 - R^2)$$

where $n$ is the size of the sample, $k$ is the number of predictors and $R^2$ is obtained from equation (7). The adjusted R-squared is only meaningful besides R-squared when the number of independent variables, $k$, is large relatively the size of the sample, $n$ (Chatterjee & Simonoff, 2013).

**Standard error of the estimate**

The standard error of the estimate is another goodness-of-fit measure besides R-squared. It is the estimate of the standard deviation of the sampling distribution of
the regression line and is determined by the difference between the observed data points and the regression line. The measure is defined as

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}}
\]  

(12)

where \(y_i\) is an observed value, \(\hat{y}_i\) is a predicted value, \(n\) is the size of the sample and \(k\) is the number of independent variables. Lower standard errors are preferable since it means that the probability of smaller errors is higher. It does not guaranty that it will always be smaller. (Chatterjee & Simonoff, 2013).

**Standard errors of the estimated coefficients**

The standard errors of the estimated coefficients, \(s.\hat{e}.(\hat{\beta})\), is given by the square roots of the diagonal elements of

\[
\sqrt{\Psi(\beta)} = \sigma \sqrt{(X'X)^{-1}}
\]

(13)

where \(\sigma\) is the standard error of the estimate in equation (12). The term \(\Psi(\beta)\) denotes the variance of the predicted coefficient vector

\[
\beta = \begin{pmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_k
\end{pmatrix}
\]

and the matrix notation of \(X\) is given by

\[
X = \begin{pmatrix}
1 & x_{11} & \cdots & x_{1n} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{1k} & \cdots & x_{nk}
\end{pmatrix}
\]

where \(n\) is the size of the sample and \(k\) is the number of independent variables (Chatterjee & Simonoff, 2013).

**F-value**

An \(F\)-test is used to determine whether any of the independent variables provide predictive power of the target variable. It tests the overall significance of the regression by comparing a model with no predictors, i.e. an intercept-only model, to the fitted model,

\[
H_0: \beta_1 = \cdots = \beta_k = 0
\]
versus

\[ H_a: \text{some } \beta_j \neq 0, \quad j = 1, ..., k. \]

The null hypothesis, \( H_0 \), indicates that the intercept-only model fits the data as well as the predicted model, whereas the alternative hypothesis, \( H_a \), states that the fitted model provides a better prediction than the intercept-only model. In comparison to a \( t \)-test, which examines the regression coefficients individually, an \( F \)-test treats all of the coefficients jointly. If the \( t \)-tests show that none of the independent variables are statistically significant, neither will the \( F \)-test. The \( F \)-value is given by

\[
F = \frac{SS_{\text{Reg}}/k}{SS_{\text{Res}}/(n-k-1)}
\]  

(14)

where the regression sum of squares, \( SS_{\text{Reg}} \), is defined as

\[
SS_{\text{Reg}} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2
\]  

(15)

and the residual sum of squares, \( SS_{\text{Res}} \), is given by equation (10) (Chatterjee & Simonoff, 2013).

The significance \( F \), which is obtained from the \( F \)-value, is the probability that the null hypothesis cannot be rejected, i.e. an indication of the probability that all the estimated coefficients are zero. For a confidence level of 95%, a model is considered statistically significant if the significance \( F \) is less than 0.05.

**P-value**

The \( p \)-value of the respective estimated coefficients indicates how statistically significant the relationship between the response variable and each predictor is. A \( p \)-value tests the null hypothesis that the independent variable has no correlation with the dependent variable, i.e. the coefficient is equal to zero.

\[ H_0: \beta_j = 0, \quad j = 1, ..., k \]

versus

\[ H_a: \beta_j \neq 0. \]

The values of \( \beta_j \) can be tested in the null hypothesis by a \( t \)-test,

\[
t_j = \frac{\hat{\beta}_j}{s.e.(\hat{\beta}_j)}
\]  

(16)

where \( t_j \) is the \( t \)-statistic and \( s.e. (\hat{\beta}_j) \) denotes the standard error of the estimated coefficient given by equation (13). However, the \( t \)-statistic is not very meaningful,
but it is needed for the computation of the $p$-values. For a confidence interval of 
95%, a $p$-value less than 0.05 indicates that the null hypothesis can be rejected, 
which means that changes in the independent variable are associated with changes 
in the response variable. In that case, the variable is likely to be a meaningful 
addition to the regression model. A $p$-value higher than 0.05 indicates that there is 
insufficient evidence to conclude that a non-zero correlation exists, which means 
that changes in the independent variable are not associated with changes in the 
target variable (Chatterjee & Simonoff, 2013).

**Confidence interval for the regression coefficients**

A confidence interval can be interpreted as an estimation of a range of values that 
with a certain confidence is going to overlap with the true value. A $100 \times (1 - \alpha)\%$ 
confidence interval for $\beta_j$ is given by

$$
\beta_j \pm t_{a/2}^{n-k-1} \times e. (\beta_j)
$$

where $t_{a/2}^{n-k-1}$ has a $t$-distribution on $n - k - 1$ degrees of freedom and represents 
the appropriate critical $t$-value at the two-sided level of $\alpha$. With this notation, a 
confidence level of 95% implies that $\alpha = 0.05$ (Chatterjee & Simonoff, 2013).
3. Building a model of customer behavior for NMDs

This chapter presents the development of a model for customer behavior regarding NMDs. By using the reviewed literature and the author’s own assumptions, the factors in the model are studied based on their believed integration with each other and significance level. The model is built on three steps: identification, integration and significance level of the factors. These parts are described in sections 3.1-3.3 and a summary is given in 3.4. Lastly, an introduction of a mathematical model for customer behavior of an NMD is presented in 3.5.

3.1 Identification

This section introduces the factors that are included in the model. The relationships between the variables and customer behavior are briefly described and motivated. Depositors’ saving behavior can be considered the dependent variable, whereas the other factors are the independent variables. The factors are categorized into either internal (individual) factors (table 1) or external factors (table 2).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age is related to saving behavior since customers have different abilities and willingness to save depending on their current life cycle phase.</td>
</tr>
<tr>
<td>Gender</td>
<td>Saving behavior may differ with respect to gender since men and women tend to have different prerequisites regarding saving.</td>
</tr>
<tr>
<td>Income</td>
<td>The level of income is linked to saving behavior in the sense that it influences the lifestyle and ability to save.</td>
</tr>
<tr>
<td>Education</td>
<td>Education level is usually closely associated with income and thus affect saving behavior similarly to income. Moreover, education tends to be linked to financial literacy.</td>
</tr>
<tr>
<td>Financial literacy</td>
<td>Financial literacy affects customers’ understanding of financial concepts, which in turn may affect their saving decisions.</td>
</tr>
<tr>
<td>Saving motives</td>
<td>Saving motives are related to individuals’ aims of their savings.</td>
</tr>
<tr>
<td>Saving horizon</td>
<td>Saving horizons are the time period during which individuals plan to keep the money on the account.</td>
</tr>
</tbody>
</table>
### Table 2: Identification of external factors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of relationship</td>
<td>The length of the customer relationship may be an indication of client loyalty.</td>
</tr>
<tr>
<td>Number of services and suppliers</td>
<td>The number of current services at a particular bank and the number of suppliers customers have may show their propensity to switch to another bank.</td>
</tr>
<tr>
<td>Financial advisory</td>
<td>The use of financial advisory may have an impact on customers’ saving decisions and also has the possibility to increase individuals’ financial literacy.</td>
</tr>
<tr>
<td>Digital banking</td>
<td>Digital banking enables clients to more efficiently manage their accounts and compare the pricing of products.</td>
</tr>
<tr>
<td>Deposit rates</td>
<td>Changes in deposit rates may affect customers’ decision of transferring money from one account to another.</td>
</tr>
<tr>
<td>Switching costs</td>
<td>Switching costs have the possibility of creating lock-in effects, which may decrease customers’ propensity to switch to another bank.</td>
</tr>
<tr>
<td>Pricing incentives</td>
<td>Banks’ offer of pricing incentives may influence customers’ saving persistence.</td>
</tr>
<tr>
<td>Macroeconomic variables</td>
<td>Macroeconomic factors tend to influence the overall direction of consumer behavior.</td>
</tr>
<tr>
<td>Financial markets</td>
<td>Financial market dynamics may affect customers’ attitudes towards saving versus investing in securities.</td>
</tr>
<tr>
<td>Competitive environment</td>
<td>Banking competitors’ product range, pricings and service quality may affect customers’ tendency to switch bank.</td>
</tr>
<tr>
<td>Seasonality</td>
<td>Customers’ consumption levels may change with seasonality, which in turn affect account balances.</td>
</tr>
</tbody>
</table>

### 3.2 Integration

To increase the understanding of the factors’ impact on customer behavior, it is sufficient to analyze how they integrate with each other. In table 3, the believed influence that the variables have on each other is presented. To clarify the integration analysis, it is built on the factors that are considered to have the highest degree of dependence, which means that not all factors will be included in this part. In each of the following cases, one variable is denoted as dependent and the factors that are believed to influence the particular variable are referred to as independent.
Table 3: Integration analysis

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Independent variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saving motives</td>
<td>▪ Age ▪ Gender ▪ Income ▪ Education ▪ Macroeconomic factors ▪ Financial markets</td>
<td>Customers’ saving motives tend to vary in line with their life cycle and thus with age. Individuals who are women, low-income earners and low educated are likely to prioritize to save for daily consumption and precaution, whereas people who are men, high-income earners and highly educated tend to focus on the growth motive. The overall macroeconomic situation as well as financial market dynamics affect individuals’ saving motives based on the ability and willingness to save.</td>
</tr>
<tr>
<td>Saving horizon</td>
<td>▪ Age ▪ Gender ▪ Income ▪ Education ▪ Macroeconomic factors ▪ Financial markets ▪ Deposit rates ▪ Pricing incentives</td>
<td>Different ages may allow for different saving horizons in the sense that individuals' abilities and willingness to save vary over their lifetime. For example, young adults and retired may not be motivated or do not have enough money to save for longer time horizons, compared to middle-aged people. Individuals who are men, high-income earners and highly educated tend to have longer saving horizons than people who are women, low-income earners and low educated. Macroeconomic factors and financial market dynamics may have an impact on customers’ ability and willingness to save for different time periods. Attractive deposit rate levels in comparison to other products, but also pricing incentives offered by the bank, may entice customers' to save for longer time horizons.</td>
</tr>
<tr>
<td>Financial literacy</td>
<td>▪ Age ▪ Gender ▪ Income ▪ Education</td>
<td>Individuals who are middle-aged, men, high-income earners and highly educated are likely to have more financial literacy than people who are younger or retired, women, low-income earners and low educated.</td>
</tr>
<tr>
<td>Financial advisory</td>
<td>▪ Age ▪ Income ▪ Education</td>
<td>The use of financial advisory tends to be more common among individuals who are middle-aged or retired, high-income earners and highly educated, compared to those who are younger, low-income earners and low educated.</td>
</tr>
<tr>
<td>Number of suppliers</td>
<td>▪ Age ▪ Gender ▪ Income</td>
<td>Customers who are middle-aged, men, high-income earners and highly educated tend to be more likely to use several</td>
</tr>
<tr>
<td>Education</td>
<td>Digital banking</td>
<td>Switching costs</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>suppliers simultaneously, whereas individuals who are retired, women, low-income earners and low educated usually have one bank only. Digital banking improves customers' ability to manage and compare banks' products at the same time. Switching costs have the possibility to tie up clients to several services at a bank, which in turn may lower the likelihood that customers acquire additional suppliers. The attractiveness of products offered by banking competitors may also affect customers' choices of adding another supplier.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1:** The figure illustrates the integration analysis of the selected factors.

Furthermore, it is very likely that the dependent variables influence each other as well. For instance, saving motives affect saving horizons since a specific saving motive will require a certain period of saving. The other way around applies too, where customers might be restricted to some time horizon, which in turn will allow for certain motives. Moreover, the probability of using financial advisory tends to be influenced by saving motives and saving horizons. Customers with clear and specific saving motives and who have longer saving horizons involving more uncertainty usually have increased likelihood of using financial advisory. The use of a financial advisor also tends to be highly integrated with financial literacy in the sense that customers' financial literacy may affect the probability of using an advisor. Individuals with low financial literacy are more likely to have a financial advisor, whereas the opposite applies to people with high financial literacy.
Additionally, the use of financial advisory has the possibility to improve customers’ financial literacy.

The number of suppliers customers have is likely to depend on financial literacy, but also on the bank-customer relationship. Customers with high financial literacy are more likely to have several suppliers, while customers with low financial literacy mainly use one bank only. Furthermore, if a customer’s confidence in the financial advisor increases, it is possible that he or she deposits more money into the bank, but also assigns the advisor more freedom to take decisions.

### 3.3 Significance level

The identified factors in section 3.1 are believed not to be equally significant, where some of these variables are likely to have greater impact on saving behavior than others. In table 4 and 5, a low, moderate and/or high significance level has been assigned to each factor together with a motivation of the chosen level or combination of levels.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Significance level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Moderate</td>
<td>Individuals’ saving needs are highly associated with their current stage of life, which in turn is linked to age. However, it is rather the specific circumstances in people’s lives at different ages, than the actual age, that are the influencers of saving behavior.</td>
</tr>
<tr>
<td>Gender</td>
<td>Moderate</td>
<td>Potential saving behavior differences between the genders are linked to that men and women tend not to have the same prerequisites on the labor market, but also due to personality differences. Thus, it is unlikely that the actual gender is a determinant of saving behavior, but rather the underlying factors that usually distinct men from women.</td>
</tr>
<tr>
<td>Income</td>
<td>High</td>
<td>Saving behavior is to a great extent dependent on income since it affects individuals’ ability to save as well as their life style.</td>
</tr>
<tr>
<td>Education</td>
<td>High</td>
<td>Education is considered a significant determinant of saving behavior since it tends to influence individuals’ knowledge, involvement and interest in finance, but it is also typically linked to income.</td>
</tr>
</tbody>
</table>
Financial literacy is an important factor of saving behavior due to that it affects individuals’ fundamental understanding of household economy, but also since it controls for the logic behind the decisions being made.

Saving motives have high impact on saving behavior since they are the foundation of what customers plan to do with their money.

Customers’ ability and willingness to save allow for different saving horizons.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Significance level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of relationship</td>
<td>Low/Moderate</td>
<td>While longer relationships may be an indication of trust from individuals’ side, they do not necessarily have to imply higher customer loyalty. To get a picture of clients’ trustworthiness they need to be studied in more detail.</td>
</tr>
<tr>
<td>Number of services and suppliers</td>
<td>Moderate/High</td>
<td>The number of current services customers use at a particular bank and the number of different suppliers they have can be important measures of customer loyalty as well as their propensity to switch to another bank.</td>
</tr>
<tr>
<td>Financial advisory</td>
<td>Moderate/High</td>
<td>The impact on saving behavior by the use of financial advisory depends on the relationship between the customer and the advisor, but also on the individual’s attitude.</td>
</tr>
<tr>
<td>Digital banking</td>
<td>High</td>
<td>Digital banking increases customers’ ability to easily access and manage accounts, which implies that decisions and transfers can be made more rapidly.</td>
</tr>
<tr>
<td>Deposit rates</td>
<td>Low/Moderate/High</td>
<td>Deposit rates affect on saving behavior vary between different account types, but also depend on how actively the customer compares rates between different products and banks.</td>
</tr>
<tr>
<td>Switching costs</td>
<td>Moderate/High</td>
<td>Switching costs is a significant factor that may prevent customers from withdrawing their money. However, the impact of this variable depends on case-to-case, where the new bank must be evaluated against the current bank as well as the different costs that will apply to the customer during a potential switch.</td>
</tr>
</tbody>
</table>
Pricing incentives | Low/Moderate | While pricing incentives may increase customers’ saving persistence, not all individuals have the economic ability to extend their saving periods.

Macroeconomic factors | High | Macroeconomic factors are significant determinants of saving behavior since they influence customers' overall attitudes towards saving, consuming and other investment objectives.

Financial markets | High | Financial market prices affect the attractiveness of saving money on deposit accounts versus investing in securities.

Competitive environment | Low/Moderate/High | Other banks’ pricing of products and service quality are important determinants of customers’ propensity to switch bank. However, not all customers actively compare their current bank to others.

Seasonality | Low/Moderate/High | Seasonality variations affect deposit balances, however, to what extent depends on the specific account type. Transactional accounts tend to be influenced to a greater extent than saving accounts.

3.4 A framework for customer behavior of NMDs

The information and analysis presented in sections 3.1-3.3 explain the likely behavior of NMD customers. In summary, the below framework describes the structure and significant components of saving behavior.
3.5 A mathematical approach to model customer behavior of an NMD

An attempt to model customer behavior of an NMD mathematically is made in addition to the framework described in sections 3.1-3.4. The factors to be included in the mathematical model are selected based on available data provided by a large Swedish bank. Consequently, the involved variables are customers’ age, gender and their respective daily account balances for some time interval. Due to the limited data, only a small part of the framework is modeled and tested mathematically.

It is of interest for banks to look into potential relations between changes in deposit volume on a certain transactional account and different customer characteristics. Since the data available for this analysis is based on customers’ age and gender, a suggestion would be to analyze if men respectively women in different ages tend to exhibit certain patterns regarding their withdrawal and depositing of money on this transactional account. Such information could be useful in the bank’s liquidity risk management, i.e. in the identification of stable deposit volumes. To structure this analysis, individuals are divided into different customer categories based on their age and gender.
Changes in deposit volumes can be examined in absolute or relative terms. This study will incorporate both of these perspectives in the evaluation. The absolute volume change calculates the difference in the volume over two periods in time, whereas the relative volume change expresses the absolute volume change as a percentage of for example the average volume or the volume in the beginning of the period. Since the absolute volume change varies from client to client, this measure has very little generalizability. On the other hand, the relative volume change can describe customer behavior regardless of account balances.

The aim of the mathematical model is twofold. Firstly, to estimate the long-term linear trend of the time series of NMD volumes for specified customer categories and to analyze how these differ from each other in absolute and relative terms. Secondly, by the use of multiple linear regression, to examine the relationship between relative volume changes and customers’ age and gender. The two objectives are treated separately and the mathematical approaches are described in the next chapter.
4. Methodology

This chapter presents the methodology used in the study. A description and analysis of the data is presented in section 4.1, followed by an explanation of the implementation in 4.2.

4.1 Description and analysis of the data

The data used in this study is provided by a large Swedish bank. It consists of anonymized information regarding 10,235 randomly selected retail customers’ age and gender together with their daily NMD balances from 04-11-2014 to 01-04-2019. The NMD is a transactional account, which has had a deposit rate of zero since the financial crisis of 2008. Customers typically use this account for daily consumption and for receiving income.

Based on the given data, the randomly selected customers are clustered into segments based on their age and gender. Four age categories have been chosen for this analysis and are similar to the ones in the study by Schlueter et al. (2015). Hence, the customers are grouped into the following age categories: 18-24, 25-44, 45-64 and ≥65. Since individuals’ age obviously vary during the time interval, the allocation of customers to the different categories is based on their year of birth together with their current age. Furthermore, the reason for excluding people below the age of 18 is due to that their deposits are likely to be influenced by their parents and thus the actual depositor might not control the account.

The combination of gender with the four age categories results in a total of eight customer segments. To distinguish male and female categories from each other, the notations ‘M’ respective ‘F’ are used followed by the age category denoted ACj, where j refers to the number of the age category. This means that the categories corresponds to the following age intervals: AC1: 18-24, AC2: 25-44, AC3: 45-64 and AC4: ≥65. Table 7 shows the distribution of customers in respective category.
Table 6: Distribution of individuals in the customer categories

<table>
<thead>
<tr>
<th>Customer category</th>
<th>Number of customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAC1</td>
<td>551</td>
</tr>
<tr>
<td>FAC2</td>
<td>1522</td>
</tr>
<tr>
<td>FAC3</td>
<td>1641</td>
</tr>
<tr>
<td>FAC4</td>
<td>1272</td>
</tr>
<tr>
<td>MAC1</td>
<td>557</td>
</tr>
<tr>
<td>MAC2</td>
<td>1652</td>
</tr>
<tr>
<td>MAC3</td>
<td>1820</td>
</tr>
<tr>
<td>MAC4</td>
<td>1220</td>
</tr>
</tbody>
</table>

4.2 Implementation

In section 3.5, the two aims of the mathematical model were stated. The first purpose is to estimate the long-term linear trend of the time series of NMD volumes for the eight customer categories and to analyze how these differ from each other in absolute and relative terms. The second objective is by the use of multiple linear regression, to examine the relationship between relative volume changes and customers’ age and gender. A description of the mathematical methodology is presented below.

To treat the first objective, individuals are, based on their age and gender, allocated to an appropriate customer category. This means that each category contains the aggregate deposit volumes for the belonging individuals. To obtain the average volume per client in each segment, the total volume is divided by the number of individuals in that specific category. Thereafter, time series plots of the average deposit volume evolution per customer and category are made. By using simple linear regression, as explained in section 2.5.1, a linear trend over the time interval is estimated with the volume as the dependent variable and time as the independent variable.

Let $V_i(t)$ be a unary function that represents the linear trend of the average deposit volume per individual in customer category $i$ at time $t$, where $t$ is measured in days. Then, $V_i(t)$ can be written on the form

$$V_i(t) = a_i + b_i t + \epsilon$$

where $a_i$ is the average initial deposit volume per individual, $b_i$ represents the daily average absolute volume change per individual and $\epsilon$ is the error term. The estimates $\hat{a}_i$ and $\hat{b}_i$ of the unknown coefficients $a_i$ and $b_i$ are obtained by using an OLS approach. Thus, the fitted regression equation can be written
\[ \hat{V}_i(t) = \hat{\alpha}_i + \hat{\beta}_i t. \] 

(18)

This means that the number of regressions made corresponds to the number of customer categories, i.e. eight, and the number of observations in each regression corresponds to the number of individuals in respectively category. The slope coefficient of the fitted line, \( \hat{\beta}_i \), represents the daily average absolute volume change over the time period. Thus, a positive trend would suggest an overall daily increase in deposit volumes during the time interval, whereas a negative trend would indicate a decrease.

An alternative way of expressing the average deposit volume for an individual in category \( i \) at time \( t \) is

\[ f_i(t) = f_i(0)(1 + \Delta f_i \times t) = f_i(0) + f_i(0) \times \Delta f_i \times t \]

(19)

where \( f_i(t) \) corresponds to \( \hat{V}_i(t) \) in equation (18), \( f_i(0) \) is the average initial deposit volume per individual in category \( i \) and \( f_i(0) \times \Delta f_i \) is the daily average absolute volume change per individual. By comparing the equations (18) and (19), the estimated coefficients \( \hat{\alpha}_i \) and \( \hat{\beta}_i \) can be identified by

\[ \hat{\alpha}_i = f_i(0) \] 

(20)

\[ \hat{\beta}_i = f_i(0) \times \Delta f_i \] 

(21)

Substituting \( f_i(0) = \hat{\alpha}_i \) into equation (21) and rewriting gives the following expression for \( \Delta f_i \),

\[ \Delta f_i = \frac{\hat{\beta}_i}{\hat{\alpha}_i} \] 

(22)

Thus, \( \Delta f_i \) denotes the daily average relative volume change with respect to the initial volume per individual in category \( i \), which refers to the daily average absolute volume change divided by the average initial volume. This means that \( \Delta f_i \) is expressed in decimal form and per day.

The second objective of the mathematical model is treated by first applying simple linear regression, followed by multiple linear regression with a special case.

Compared to the case of the first objective, where individuals were organized into customer categories, they are now handled separately in the simple linear regression. From the time series of individuals’ deposit volume evolutions, linear trends are estimated using an OLS approach. Again this means that simple linear regression is applied with the volume as the dependent variable and time as the independent variable. The number of regressions made, as well as the number of fitted equations obtained, corresponds to the total number of customers in the sample.
To obtain the relative deposit volume change with respect to the average volume per customer, the estimated slope coefficients are divided by the individuals’ average deposit volume over the time interval. One way to think of this ratio is by looking at the following example. Consider two individuals who save 1000 SEK per month respectively. One of them has on average 10,000 SEK on the account, whereas the other one has an average deposit of 100,000 SEK. The measure then tells that the person with lower average volume has higher saving rate compared to the other individual. Furthermore, it turns out that 234 individuals’ average deposit volume sums up to zero, which implies that the division becomes undefined. These cases are excluded from the multiple regression, which implies that the sample size is 10,001 instead of earlier 10,235.

A technique using binary variables, also known as dummy variables, is applied to the multiple linear regression. Such variables are useful when a predictor does not take any real numerical value, but instead it represents an attribute that has at least two distinct categories, such as age and gender. A dummy variable adopts the value ‘1’ for some observation to indicate membership in a group and ‘0’ for observations not indicating membership.

An important aspect regarding this method is that the number of binary variables in the model should equal the number of categories in an attribute minus one. The reason for excluding one category is to avoid the so-called ‘dummy variable trap’, which occurs when all categories in one attribute are included in the model (Green, 2003). Such a mistake would result in that the dummy variables sum to one at every observation, which in turn would reproduce the constant term and lead to a situation of perfect multicollinearity. In this case, gender has two categories, which implies that a single dummy variable is needed to represent this attribute. Similarly, the four age categories imply that three dummy variables are needed for this attribute. The oldest age category is chosen to be excluded from the regression analysis; however, the correctness of the fitted equation is not affected by which one of these that is being removed. Note that the categories are mutually exclusive and exhaustive, which imply that the dummy variable technique is applicable. The multiple linear regression equation for this specific case can be written on the form

$$
\Delta f_i = \beta_0 + \beta_1 G + \beta_2 AC1 + \beta_3 AC2 + \beta_4 AC3 + \epsilon_i
$$

(23)

where $\Delta f_i$ is the daily relative volume change with respect to average volume for customer $i$, $G \in \{0, 1\}$ designates gender and $ACj \in \{0, 1\}$ defines age category. As before, $\epsilon_i$ refers to the error term and estimates of the unknown coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ and $\beta_4$ are determined by an OLS approach. Thus, the fitted equation can be written

$$
\hat{\Delta f}_i = \hat{\beta}_0 + \hat{\beta}_1 G + \hat{\beta}_2 AC1 + \hat{\beta}_3 AC2 + \hat{\beta}_4 AC3.
$$

(24)

Table 6 illustrates the regression technique with dummy variables. Here, the value ‘0’ is assigned to women, which means that ‘1’ refers to men. Regarding the age categories, a ‘1’ indicates that an individual belongs to some age category and ‘0’
otherwise. The values of $\Delta f_i$ represent male and female individuals that belong to some age category. As can be seen in the table, there are 10,001 observations of $\Delta f_i$.

**Table 7: Multiple regression with dummy variables**

<table>
<thead>
<tr>
<th>$\Delta f_i$</th>
<th>$G$</th>
<th>$AC1$</th>
<th>$AC2$</th>
<th>$AC3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta f_1$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{539}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Delta f_{540}$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{2024}$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$\Delta f_{2025}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{3634}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\Delta f_{3635}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{4889}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Delta f_{4890}$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{5436}$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Delta f_{5437}$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{7035}$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$\Delta f_{7036}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{8005}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\Delta f_{8006}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta f_{10001}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
By performing the multiple regression, as described in section 2.5.2, estimates of the unknown coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ and $\beta_4$ together with some other statistical measures are obtained. It would then be possible to use this equation to predict the daily relative volume change with respect to average volume for a particular customer whose gender and age are known. By evaluating the regression outcomes, conclusions about the relevance of the model can be made and thus whether banks should apply this model in their management of NMDs or not.
5. Results

This chapter presents the results obtained from the mathematical approach to model customer behavior of an NMD. Firstly, the time series of the deposit volumes are shown in 5.1. Secondly, figures of the trend estimation for each client category are displayed in 5.2, followed by calculations of their respective relative volume change in 5.3. Lastly, the results from the multiple regression analysis are presented in 5.4.

5.1 Deposit volumes

Figure 3 shows the time series of the average customer’s deposit volume for the eight client segments. The time interval ranges from 04-11-2014 to 01-04-2019. In absolute terms, the deposit balances clearly vary among the different categories and over time. The time series exhibit monthly seasonal variations, as seen by the systematic and repetitive pattern over the time period, and long-term trends.

The time series of MAC1 and FAC1, which represent men respectively women in the ages of 18-24, are found at the bottom. In the middle part of the graph, the deposit volumes of FAC2, FAC3, FAC4 and MAC2 i.e. women in the ages of 25-44, 45-64 and ≥65 and men 25-44 years old, are located. The categories representing men in the ages of 45-64 and ≥65, i.e. MAC3 and MAC4, exhibit the highest deposit volumes.
5.2 Trend estimation

In figure 4 and 5, the estimated linear trends for the time series are displayed. The intercept and inclination of each trendline are shown in table 8. Consistent for all the time series, which can be seen by the positive sign of the slopes, is that the trends are increasing. This indicates an overall volume growth for the average customer in the sample over the time interval. The intercepts show the average initial deposit volume per individual for the eight customer categories.

The slopes displayed in table 8 indicate that the time series of FAC1 and MAC1 have the flattest lines, whereas FAC2 and MAC2 exhibit somewhat more increasing trends. The categories FAC3, FAC4 and MAC3 show the overall largest daily volume growths in absolute terms. Interestingly, the difference in slopes between MAC3 and MAC4 is significant, whereas the inclinations for FAC3 and FAC4 are almost the same. The male categories MAC1 and MAC2 have greater inclinations than the corresponding female groups, whereas FAC3 and FAC4 have greater slopes than the corresponding male categories. Furthermore, the only occasion where women on average have higher intercept than men is for the youngest age category. The difference between the intercepts with regard to gender increases for older age categories.
Figure 4: The figure shows the graphs of the estimated linear trends for the female categories from 04-11-2014 to 01-04-2019. Volumes are displayed in thousand SEK.
Figure 5: The figure shows the graphs of the estimated linear trends for the male categories from 04-11-2014 to 01-04-2019. Volumes are displayed in thousand SEK.
Table 8: Trend coefficients

<table>
<thead>
<tr>
<th>Category</th>
<th>Intercept (10^4)</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAC1</td>
<td>1.684</td>
<td>0.940</td>
</tr>
<tr>
<td>FAC2</td>
<td>2.604</td>
<td>3.991</td>
</tr>
<tr>
<td>FAC3</td>
<td>3.274</td>
<td>14.937</td>
</tr>
<tr>
<td>FAC4</td>
<td>3.870</td>
<td>14.595</td>
</tr>
<tr>
<td>MAC1</td>
<td>1.622</td>
<td>3.557</td>
</tr>
<tr>
<td>MAC2</td>
<td>3.335</td>
<td>4.537</td>
</tr>
<tr>
<td>MAC3</td>
<td>5.202</td>
<td>14.144</td>
</tr>
<tr>
<td>MAC4</td>
<td>6.036</td>
<td>7.083</td>
</tr>
</tbody>
</table>

Note: The intercepts are displayed in SEK per individual and the slopes are shown in SEK per day and individual.

5.3 Relative deposit volume changes

In table 9, the results of the daily average relative deposit volume change with respect to initial volume per individual for the eight categories are displayed. These are obtained by dividing the slope by the intercept for each category. It can be seen that the female categories show the lowest but also the greatest daily average relative volume changes. In contrast to women, the daily average relative volume changes for the male categories are less varied between the different age intervals.
Table 9: Relative deposit volume changes

<table>
<thead>
<tr>
<th>Category</th>
<th>Δf</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAC1</td>
<td>5.584 \cdot 10^{-5}</td>
</tr>
<tr>
<td>FAC2</td>
<td>1.532 \cdot 10^{-4}</td>
</tr>
<tr>
<td>FAC3</td>
<td>4.563 \cdot 10^{-4}</td>
</tr>
<tr>
<td>FAC4</td>
<td>3.771 \cdot 10^{-4}</td>
</tr>
<tr>
<td>MAC1</td>
<td>2.193 \cdot 10^{-4}</td>
</tr>
<tr>
<td>MAC2</td>
<td>1.360 \cdot 10^{-4}</td>
</tr>
<tr>
<td>MAC3</td>
<td>2.719 \cdot 10^{-4}</td>
</tr>
<tr>
<td>MAC4</td>
<td>1.174 \cdot 10^{-4}</td>
</tr>
</tbody>
</table>

Note: Δf is expressed per day and individual.

5.4 Multiple regression with dummy variables

Table 10 shows the results from the multiple regression, where the purpose is to analyze the relationships between the daily relative volume change with respect to average volume (dependent variable) and customers’ age and gender (independent variables). Note that only three out of the four age categories are listed in the table since setting AC1, AC2 and AC3 to ‘0’ indicates AC4. Important to have in mind is that the interpretations of the statistical outcomes rely heavily on the underlying model assumptions.

The signs of the coefficients tell how changes in the predictor variables are linked to changes in the response variable. First of all, the constant intercept by itself indicates the value of the daily relative volume change a woman in AC4 has. Furthermore, the negative sign in front of the gender coefficient denotes that men on average have lower daily relative volume change than women. With regard to the age categories, AC1, AC2 and AC3 decrease the daily relative volume change. As mentioned before, setting these three age categories equal to zero indicates the presence of AC4.

R-squared based on 10,001 observations is measured to 0.202%, which means that only 0.202% of the variation in daily relative volume changes can be explained by age and gender. Furthermore, the adjusted R-squared results in 0.162%, which is relatively close to R-squared. The similarity between these measures is likely to be explained by that the number of predictors in the model is only four, whereas the number of observations is 10,001. Thus, the great difference between these values implies that the adjusted R-squared is not a meaningful addition for this model.
The significance $F$ is measured to less than 0.05, which means that the fitted model overall is statistically significant. However, only one of the predictors, i.e. the AC2 variable, results in a $p$-value less than 0.05. This suggests that the relationship between the dependent variable and the AC2 predictor is statistically significant for a 95% confidence interval, whereas the other independent variables are not. Moreover, this means that only the null hypothesis for AC2 can be rejected, while there is insufficient evidence for the remaining predictors to conclude that non-zero correlations exist.

The standard error of the estimate is calculated to approximately 0.00312. This value is considered high in comparison to the values of $\Delta f_i$ and the great size of the sample relatively few number of predictors, where the difference between these two values is found in the denominator of the formula. Also, this value influences the standard errors of the estimated coefficients, which appear to be relatively high as well compared to the size of the coefficients.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Standard error</th>
<th>$t$-stat</th>
<th>$p$-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$1.345 \cdot 10^{-4}$</td>
<td>$6.996 \cdot 10^{-5}$</td>
<td>1.922</td>
<td>0.0546</td>
<td>$-2.648 \cdot 10^{-6}$</td>
<td>$2.716 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>Gender</td>
<td>$-4.003 \cdot 10^{-6}$</td>
<td>$6.240 \cdot 10^{-5}$</td>
<td>-0.0642</td>
<td>0.949</td>
<td>$-1.263 \cdot 10^{-4}$</td>
<td>$1.183 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>AC1</td>
<td>$-1.272 \cdot 10^{-4}$</td>
<td>$1.137 \cdot 10^{-4}$</td>
<td>-1.119</td>
<td>0.263</td>
<td>$-3.500 \cdot 10^{-4}$</td>
<td>$9.558 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>AC2</td>
<td>$-3.433 \cdot 10^{-4}$</td>
<td>$8.440 \cdot 10^{-5}$</td>
<td>-4.068</td>
<td>4.775 $\cdot 10^{-5}$</td>
<td>$-5.088 \cdot 10^{-4}$</td>
<td>$-1.779 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>AC3</td>
<td>$-6.231 \cdot 10^{-5}$</td>
<td>$8.275 \cdot 10^{-5}$</td>
<td>-0.753</td>
<td>0.451</td>
<td>$-2.245 \cdot 10^{-4}$</td>
<td>$9.990 \cdot 10^{-5}$</td>
</tr>
</tbody>
</table>

Note: The table shows the statistical outcomes from the multiple regression.
6. Discussion

In this chapter, the findings from the mathematical approach are discussed with regard to customer behavior of the studied transactional NMD. Firstly, the time series of deposit volumes are evaluated in 6.1. Secondly, the results from the multiple regression representing the effect of age and gender on relative volume changes are discussed in 6.2. Thirdly, a reconnection to customer behavior of NMDs is made in 6.3.

6.1 Time series of deposit volumes

Overall, the deposit volume evolutions for the different customer categories coincide with the reviewed literature. To begin with, the fact that the time series of FAC1 and MAC1 in figure 3 are found below the other categories seems logical since people in these ages usually are students or in the beginning of their careers. Thus, it makes sense that these individuals have not been able to accumulate money to the same extent as people in the older categories. Also, the relatively flat inclinations of FAC1 and MAC1 indicate a low growth of the deposit balances during the time interval. This could be linked to the theory of Eriksson and Hermansson (2014) regarding that individuals’ ability and willingness to save can be related to their current life cycle phase. Thus, a possible explanation for a rather flat trend could be that people at this early stage in life are less likely to obtain significant salary increases.

Furthermore, the deposit volumes of the age interval AC1 do not differ significantly in absolute terms between men and women. An explanation for this could be that student grant, i.e. CSN, is the same amount regardless of gender. It is also likely that potential labor market inequalities based on gender become more distinguishable for higher positions and thus might not appear for entry-level jobs to the same extent.

For higher deposit volumes, the time series of the three older female categories appear relatively close to each other, where FAC3 and FAC4 overall show somewhat higher volumes than FAC2. The fact that the deposit balance of FAC3 is higher than FAC2 is likely to be explained by increased income in general, whereas the similarity between FAC3 and FAC4 could perhaps be linked to the description by Hermansson (2017) regarding women’s tendency to prioritize the retirement saving motive. The category MAC2 is located somewhat above FAC2, which indicate that men are slightly financially superior over women for this age interval.

In contrast to women, there is a more significant difference in men’s deposits between AC2 and AC3. By connecting to the theory of Hermansson (2017), this could be related to that any gender inequalities on the labor market lead to that
men in general achieve higher and better-paid positions within companies in comparison to women. Women are also likely to give birth to children in the ages of category AC2, which may reduce their chances of being promoted. Furthermore, similar to that of women, men’s deposit volumes for the age categories AC3 and AC4 are relatively comparable.

Consistent for both genders is that the slopes increase from AC1 to AC2 and from AC2 to AC3, followed by a decrease between AC3 and AC4. This seems rather logical based on a life cycle perspective since individuals’ ability to accumulate money usually increase with aging. However, similar to the description by Wolff (2000), the absolute growth of deposits tends to slow down at retirement, possibly due to that pension payouts are lower than the income received before retirement.

With regard to gender, it is interesting that the most significant difference in inclinations occurs for the oldest age category. As mentioned earlier, this behavior can possibly be explained by that women want to maintain a steady buffer at retirement and thus uphold a stable volume growth, whereas men in greater extent spend their money.

The intercepts, which indicate the average initial deposit volumes, increase towards older ages for all categories, which are most likely explained by that people tend to accumulate money over the years. However, the fact that the increases between the male categories are much greater than the respective ones for the female groups indicate that the absolute differences in deposit volumes are significantly distinct between men and women.

By relating to the study by Baumann et al. (2007), age and gender appear to be significant determinants of whether individuals spread their money between different accounts and banks or not. This obviously affects deposit volumes but has not been taking into account here since it would require a detailed study of the cash flows. Other important factors that have shown to be closely associated with age and gender are customers’ financial literacy and the use of financial advisory.

6.2 The effect of age and gender on relative deposit volume changes

The evaluation of the fitted regression equation is made with regard to the statistical outcomes. Based on the low value of R-squared, i.e. 0.212%, there is very little association between daily relative deposit volume changes with respect to average volume and customers’ age and gender. The significance F indicates that the overall fitted model is statistically significant. However, this is likely to be a consequence of that the AC2 predictor, whose p-value is very low compared to the p-values of the other independent variables, significantly decreases it.
Based on the results of the multiple regression analysis, the factors age and gender do not considerably affect customers’ savings with respect to their average volume. By relating to the framework in chapter 3, this strengthens the moderate significance level, which was assigned to the variables age and gender. This suggests that it is not customers’ actual age or gender that are important determinants of saving behavior, but rather the underlying dynamics connected to these factors.

Another important aspect to discuss is whether linear regression is an appropriate method for analyzing individuals’ deposit volumes. The studied transactional account exhibits marked seasonal variations, probably since people receive their income monthly, however, these are not taken into account when estimating a long-term linear trend. This implies that important information concerning customer behavior is disregarded with this mathematical model. It is however likely that this method is more applicable for saving deposits, since those usually do not exhibit as rapid fluctuations as transactional accounts.

6.3 Reconnecting to customer behavior of NMDs

In chapter two, the characteristics of NMDs and the risk management challenges that banks experience related to these products were described. There is no doubt that the most important features of NMDs are clients’ options to withdraw money at any time and banks’ right to adjust the deposit rate. In this study, the focus has been on the behavior of customers, which refers to the former of these characteristics.

In parable with previous researches, this study can verify that it is relatively easy to theorize about the relationships between precisely defined variables. On the other side, it is more challenging to obtain accurate measures of these variables. With regard to internal factors, socioeconomic factors are usually more accessible for banks and thus easier to make predictions about in comparison to personality traits. However, through conversations and financial advisory meetings with customers, banks should be able to get a better understanding of their clients’ needs and personalities. Important to note is that the use of financial advisory services tends to be more common for certain clienteles, which means that some customers will be more difficult to access in the context of personality traits. When banks know their customers, they can work on actively targeting different clienteles, as suggested by Baumann et al. (2007).

Furthermore, the accessibility of bank customers’ personal data has decreased since the new and stricter GDPR was enforced in 2018. This regulation implies that banks have to be extremely careful when handling clients’ privacy data, which might impede the analyzing of customer behavior in some aspects. One of the
limitations in this study is that the accessible data only covers customers’ age and gender in relation to their NMD volumes over a time period of approximately five years. The mathematical part of this study proves the difficulties in creating a model for analyzing client behavior based on these factors since there are substantially many variables that have an impact on individuals’ lives. It is also noteworthy that the results might have been completely different if the model would have been applied to another data set.

The European Banking Authority states the importance of that banks should not completely rely on the quantitative results regarding NMDs. However, there is no further description of how an alternative approach could be done. It is therefore more or less up to the banks themselves to construct and develop new modeling methods for NMDs.
7. Conclusion and future research

7.1 Conclusion

NMDs represent a significant portion of many banks’ funding, which makes the modeling of these products highly interesting. Additionally, the increased focus from authorities on these financial instruments highlights the importance of proper and robust models. One of the distinguishing features of NMDs is customers’ right to withdraw money, which has been the focus in this thesis. This study has qualitatively and mathematically analyzed the behavior of bank customers and in that way contributed to the development of NMD models.

This thesis suggests a framework for customer behavior regarding NMDs that is based on the included factors’ identification, significance level and integration with each other. The framework has the opportunity to increase banks’ understanding of which variables that are important to involve in the modeling of NMDs, but also how these might interact with each other. Furthermore, based on the data provided by the bank, a mathematical approach to describe customer behavior of an NMD is made. The analysis of the specified customer categories based on age and gender shows that the eight client segments act differently from each other in absolute and relative terms.

A multiple regression using dummy variables has been performed to examine the relationship between customers’ daily relative deposit volume changes with respect to average volume and their age and gender. The evaluation of the fitted model has been done based on the regression statistics and the conclusion is that age and gender are not significant determinants of the dependent variable. This strengthens the idea that customer behavior of NMDs might not depend on actual age and gender but rather on the underlying dynamics connected to these factors. However, the conclusion is based on this specific model, customer sample and NMD, but might have been completely different under other circumstances. To examine the model’s relevance in general, it needs to be tested for different cases, i.e. for other types of NMD accounts and for other factors besides age and gender.

Although the purpose of this study was to analyze the long-term trend of an NMD, it can be realized that short-term variations would have been interesting to examine as well, especially for a transactional account. Therefore, it might not be optimal to estimate a linear trend using regression since this approach will not allow for the seasonal variations. However, the use of dummy variables to represent the non-numeric attributes age and gender is considered an effective method.
7.2 Future research

NMDs offer a wide range of interesting features to study. This thesis has concentrated on customers’ impact on NMDs, however, the behavior of the bank would also have been relevant to study. Such an approach could focus on the modeling of deposit rates, which is not included in this thesis. Furthermore, this study has analyzed customer behavior based on a few important factors. However, there are a lot of other variables that would have been relevant to examine as well, for example customers’ location, residence, marital status, household size, place of birth and risk tolerance.

The mathematical model in this thesis only involves two independent variables, but if possible, it would have been interesting to incorporate additional factors. Moreover, the used data set concerns a transactional account, which exhibits marked monthly variations. Besides from studying the long-term trend, it would have been relevant to examine the seasonality, for example by using a time series model. Lastly, there exist a variety of different NMDs that can be analyzed, whereas this study only uses data from a transactional account.
References


