Modeling of non-maturing deposits

NIKITA DOMIN
FREDRIK STAVRÉN
Modeling of non-maturing Deposits

NIKITA DOMIN
FREDRIK STAVRÉN
Royal Institute of Technology
*School of Engineering Sciences*
**KTH SCI**
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci
Abstract

The interest in modeling non-maturing deposits has skyrocketed ever since the financial crisis 2008. Not only from a regulatory and legislative perspective, but also from an investment and funding perspective.

Modeling of non-maturing deposits is a very broad subject. In this thesis some of the topics within the subject are investigated, where the greatest focus is on the modeling of the deposit volumes. The main objective is to provide the bank with an analysis of the majority of the topics that needs to be covered when modeling non-maturing deposits. This includes short-rate modeling using Vasicek’s model, deposit rate modeling using a regression approach and a method proposed by Jarrow and Van Deventer, volume modeling using SARIMA, SARIMAX and a general additive model, a static replicating portfolio based on Maes and Timmerman’s to model the behaviour of the deposit accounts and finally a liquidity risk model that was suggested by Kalkbrener and Willing. All of these models have been applied on three different account types: private transaction accounts, savings accounts and corporate savings accounts.

The results are that, due to the current market, the static replicating portfolio does not achieve the desired results. Furthermore, the best volume model for the data provided is a SARIMA model, meaning the effect of the exogenous variables are seemingly already embedded in the lagged volume. Finally, the liquidity risk results are plausible and thus deemed satisfactory.
**Sammanfattning**


Resultatet är att räntemodelleringen samt replikeringsportföljen inte ger adekvata resultat på grund av den rådande marknaden. Vidare så ger en SARIMA-modell den bästa prediktionen, vilket gör att slutsatsen är att andra exogena variabler redan är inneslutna i den fördröjda volymvariabeln. Avslutningsvis så ger likviditetsmodellen tillfredsställande resultat och antas vara rimlig.
Acknowledgements

We would like to give a special thanks to our supervisor at KTH, Anja Janssen, for always being available and helpful throughout the project. Furthermore, we would also like to give our sincerest thanks to SEB for giving us this opportunity and our supervisor at SEB, Jakob Nuldén, for his support and encouragement. Additionally, we would also like to thank all of the people at SEB who took their time to answer our questions and teach us new things. Finally, a big thank you to our friends and families who have pushed and supported us, without whom this would not have been possible.
# Contents

1 Introduction .......................... 1
   1.1 Background .......................... 1
   1.2 Regulations and guidelines ........ 2
   1.3 Aim and delimitations ............. 3
   1.4 Structure of the report .......... 5

2 Previous works ....................... 6

3 Theoretical background .............. 9
   3.1 Market interest rates ............. 9
      3.1.1 Zero rates .................. 9
      3.1.2 Modeling of the short rate .... 10
      3.1.3 Calibration of Vasicek’s model .. 11
   3.2 Deposit interest rates ............ 12
      3.2.1 Jarrow and Van Deventer ....... 12
      3.2.2 Regression model ............ 12
   3.3 Time series models ............... 13
      3.3.1 ARIMA models ............... 13
      3.3.2 Augmented Dickey-Fuller test .. 16
   3.4 Regression models ................. 16
      3.4.1 Random Forest Regression ....... 16
      3.4.2 Multiple linear regression and OLS .. 18
      3.4.3 Modified generalized additive model .. 19
   3.5 Risk management .................. 22
      3.5.1 Static replicating portfolio: Maes and Timmerman .. 22
      3.5.2 Liquidity risk .............. 23

4 Data .................................. 24
   4.1 Description of the data .......... 24
      4.1.1 Volume data ............... 24
List of Figures

1.1 The different topics that need to be modeled . . . . . . . . . . . 4

3.1 An illustration of a random forest with decision tree one, two
and n shown. Each circle represents a node where a split is done. 18

4.1 Volumes of different deposit types . . . . . . . . . . . . . . 24
4.2 Seasonal plots of scaled volumes for different deposit types . 25
4.3 Swedish market rates . . . . . . . . . . . . . . . . . . . . . 26
4.4 Deposit rates . . . . . . . . . . . . . . . . . . . . . . . . . . 26
4.5 Deposit rates vs market rate . . . . . . . . . . . . . . . . . 29
4.6 Daily aggregated data for the private transaction accounts . . . 30

6.1 Simulations of the market rate using the Vasicek short rate
model for the complete data-set . . . . . . . . . . . . . . . . . . . 38
6.2 Simulations of the market rate using the Vasicek short rate
model for a shorter time period . . . . . . . . . . . . . . . . . . . 38
6.3 Deposit rate of savings accounts (left) and corporate savings
accounts (right) . . . . . . . . . . . . . . . . . . . . . . . . . . 40
6.4 JvD: Deposit rate for savings accounts (left) and corporate sav-
ings accounts (right) . . . . . . . . . . . . . . . . . . . . 41
6.5 Average salaries for private sector (left) and unemployment
(right) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
6.6 Fitted private salaries (left) and predicted private salaries (right) 44
6.7 Fitted unemployment rate (left) and predicted unemployment
rate (right) . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
6.8 ACF plot of residuals: residuals for private salaries (left) and
unemployment rate (right) . . . . . . . . . . . . . . . . . . . . 45
6.9 Decomposition of original time series . . . . . . . . . . . . . . 46
6.10 ACF and PACF of the original time series . . . . . . . . . . . 46
6.11 Plot of differenced time series . . . . . . . . . . . . . . . . . . 47
6.12 ACF and PACF plot of differenced time series 47
6.13 Importance of different lagged volumes for the private transaction accounts 48
6.14 a) and b) is the fit and prediction respectively of the SARIMA models for the private accounts. c) and d) is the fit and prediction respectively of the SARIMAX models for the private accounts 49
6.15 Diagnostic plots of residuals 49
6.16 a) and b) is the fit and prediction respectively of the SARIMA models for the savings accounts. c) and d) is the fit and prediction respectively of the SARIMAX models for the savings accounts 50
6.17 a) and b) is the fit and prediction respectively of the SARIMA models for the corporate savings accounts. c) and d) is the fit and prediction respectively of the SARIMAX models for the corporate savings accounts 51
6.18 ACF plots of the residuals for the SARIMA models 52
6.19 ACF plots of the residuals for the SARIMAX models 53
6.20 Fit and prediction of the private transaction accounts 54
6.21 Fit and prediction of the savings accounts 54
6.22 Fit and prediction of the corporate savings accounts 55
6.23 Fit and prediction of the first vanilla model 56
6.24 Decomposition of the first vanilla model 57
6.25 Diagnostic plot, MAPE 58
6.26 Diagnostic plot, RMSE 58
6.27 Fit and prediction of the second model (improved model) with dummy variables and seasonal adjustment 59
6.28 Diagnostic plots of residuals, MAPE 59
6.29 Diagnostic plots of residuals, RMSE 59
6.30 Efficient frontier of 50 000 randomly generated weight sets 60
6.31 Return of different portfolios 62
6.32 Duration of different portfolios 63
6.33 Liquidity risk quantiles 65
6.34 Liquidity risk quantiles 66
List of Tables

1.1 Caps on core % for investment and maturity for each deposit type by BCBS ................................. 3
4.1 Correlations between different market rates .................... 28
5.1 Models for each deposit type ................................. 32
6.1 Parameters of calibrated Vasicek’s model .................... 37
6.2 Values of parameters for the regression model ................. 39
6.3 Values of parameters for Jarrow and Van Deventer .......... 40
6.4 Result of p-values of different variables for the different deposit types ........................................ 42
6.5 Visualization of the variable selection for each model .......... 43
6.6 Results for the private transaction accounts .................. 51
6.7 Results for the savings accounts ............................. 51
6.8 Results for the corporate savings accounts .................. 51
6.9 Accuracy measures of MGA model on monthly data .......... 55
6.10 Weight distribution for different portfolios ................... 61
6.11 Different measures of the portfolios ........................ 61
Chapter 1

Introduction

1.1 Background

A non-maturing deposit (NMD) represents a liability of the bank which does not have a contractual maturity date. Deposit owners have flexibility in withdrawal timing, meaning customers can withdraw their funds at any time. Meanwhile, a financial institution can change its deposit interest rate to try to regulate the deposit volumes. NMDs are of great interest to the financial institutions due to their historically low cost and the large potential gain, in other words it is easy funding.

There are different outflow types, price driven and idiosyncratic. Price driven outflow means that interest rate risk and liquidity risk are somewhat intertwined. This is an effect of customers being interest rate sensitive, i.e. customers will move their funds to get a better rate, which leads to an outflow of money. Therefore, there is a trade-off between the deposit rate and liquidity. Hence, to attract liquidity, one needs to increase the deposit rate and thus lower the spread (margin of the deposit rate and market rate).

Idiosyncratic outflows are such as bill payments, non-deposit investments and other use of funds, meaning they are unrelated to the market interest rate. Another important factor is the ease of moving money. The easier it gets, the higher the liquidity risk. Sweden has seen a significant increase in fin-tech companies that provide instantaneous deposits and withdrawals at a rate that, more often than not, is better than what the banks can give.

Moreover, the risks inherited in non-maturing deposits are market risk and liq-
uidity risk. Market risk is due to the interest rate risk in the form of a margin between deposit rate and market interest rate, i.e. the bank is overpaying on term deposits. Liquidity risk is the risk from the bank’s side that it will not have enough liquidity at hand to cover their deposit outflows.

Modeling of NMDs is a very complex topic since there are numerous factors that directly and indirectly affect the cash flows in the bank. Due to the nature and complexity of NMDs, the majority of the models are rough approximations that are used to get an indication. Hence, one often refers to historical data and concludes that the future is explained by the past. However, the future is unknown and there can sometimes be large deviations from the past pattern.

Finally, the market interest rates have been low globally since 2008, which in turn leads to low or zeroed deposit rates. In theory, negative market rates should encourage households to invest in other alternatives, e.g. the stock market, instead of savings accounts. However, if the stock market is unstable these deposit accounts are deemed more valuable by some due to their security. Hence, the bank gets an inflow of clients when the market is deemed too volatile which (hopefully) leads to larger deposit volumes. Furthermore, banks are relatively trusted due to regulation and supervision. Also, having all your funds in one financial institution is a very attractive aspect for customers.

1.2 Regulations and guidelines

The 30th of June 2019 numerous rules and regulations issued by the EBA (European Banking Authority) and the BCBS (Basel Committee Banking Supervision) need to be met. Regarding NMDs, they are first to be divided into retail and wholesale accounts. Additionally, the deposits are split into a stable and a non-stable part using a historical data sample of 10 years. Furthermore, the stable part shall be split into a core and a non-core part. Typically, the non-core part is invested in a very short time horizon (overnight or one week), while the core part is sliced and slotted in different time buckets depending on the nature of the deposits. According to the BCBS each type of NMD should be modeled separately and then all the results are to be aggregated. Finally, there are caps on the core percentage available for investment and on the investments’ (average remaining) maturity (see the tables below)[9].
Table 1.1: Caps on core % for investment and maturity for each deposit type by BCBS

<table>
<thead>
<tr>
<th>Deposit type</th>
<th>Cap on core</th>
<th>Cap on maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail/Transactional</td>
<td>90%</td>
<td>5</td>
</tr>
<tr>
<td>Retail/Non-transactional</td>
<td>70%</td>
<td>4.5</td>
</tr>
<tr>
<td>Wholesale</td>
<td>50%</td>
<td>4</td>
</tr>
</tbody>
</table>

Retail deposits are defined as deposits placed with a bank by an individual person (or small businesses). Usually accounts that have a volume of \( \leq 1 \text{ mil} \) are to be considered retail deposits. If regular transactions are carried out in a specific account or if the deposit is non-interest bearing it should be considered as a transactional retail account. Other retail deposits should be considered as held in a non-transactional account. Deposits from legal entities, sole proprietorships or partnerships are placed in the wholesale deposit categories [9].

1.3 Aim and delimitations

The goal of the project is to perform an analysis of the non-maturing deposits of the bank, but with a focus on the volume modeling. It should be viewed as a pre-study (for the bank) of the area with some regard to the upcoming BCBS regulations.

Additionally, the idea is to evaluate a basic approach for the modeling of non-maturing deposits, i.e. what needs to addressed when analyzing and modeling NMDs. The purpose of modeling NMDs is risk management. However, in order to model the risk adequately, one needs to model the different areas accurately.
Here, orange represents the areas that are modeled in this thesis. However, interest rate risk is also an important topic but it is only brushed upon in the form of a discussion.

The subject "non-maturing deposits" is very extensive and numerous delimitations are made in order to form a feasible project. First and foremost all of the different areas of the report are restricted to the Swedish market. More precisely, the considered available risk-free assets are the STIBOR and government bonds, which are assumed to be traded frictionless. Furthermore, three different accounts are modeled: private accounts, savings accounts and corporate savings accounts. It should be noted that the data provided for these ac-
count types are monthly volume observations at the beginning of each month. Hence, the data has a limitation in itself. Additionally, some large financial institutions and certain individuals are not included in the data. Furthermore, loans and other credits are not considered. Finally, interest rate risk and valuation of the deposits are not examined in this thesis (even though they are important, especially the former).

1.4 Structure of the report

The structure of the report is as follows. First off, relevant conclusions from previous works on the subject are presented. After that, in Chapter 3, a theoretical background is established on which the models of this thesis are based on. In the next chapter, the data that is used is introduced and discussed. Moreover, in Chapter 5 the theory presented in Chapter 3 is applied to formulate a methodology that is suited for the different problem topics of the thesis. Next, the results are presented in Chapter 6. Finally, in Chapter 7 and 8 the results (and the topic in general) are discussed and also final remarks and conclusions are made.
Chapter 2

Previous works

There has been a significant increase of research on this specific area for the past years, but there is no real consensus on what the best method is when modeling some of the varying topics of non-maturing deposits. First off is the replicating portfolio approach. The objective is to construct a portfolio of vanilla instruments in order to replicate the deposit account. The reason is that it is very difficult, or even impossible, to quantify some of the risk measures of the original non-maturing deposit. There are several different types of replicating portfolios. One of the most used by European banks is the static replicating portfolio proposed by Maes and Timmerman (2005). The reason for its popularity is that it is a transparent method that also often provides a decent fit. In their publication the authors test two different variants of their model. One where they maximize the Sharpe ratio of the portfolio and one where they minimize the standard deviation.

Jarrow and Van Deventer (1998) describes an arbitrage-free approach for the valuation and hedging of non-maturing deposits. They show that non-maturing liabilities are equivalent to particular interest rate swaps. Furthermore, the authors obtained an analytic valuation formula in a simple one-factor model with deposit rate and volumes given by deterministic functions of the short rate, which is modeled by an extended one-factor Vasicek model.

Next, Kalkbrener and Willing (2004) describe different techniques on managing the risks of non-maturing deposits. They provide a framework to assess the liquidity risk, which is based on a stochastic approach. Moreover, in their paper they propose two diffusion volume models. The first one is normal-distributed and the other is log-normal, since they assume the increments of
the volume are normally distributed. Both of these models consist of two components, a deterministic linear trend and a stochastic process (Ornstein-Uhlenbeck process).

Castagna and Manetti (2013) describe their approach which is based on the stochastic factor approach. The stochastic factor approach is based on three building blocks. A stochastic model for the interest rates, based on the Vasicek short rate model, a stochastic model for the deposit rates which is linked to the interest rates and lastly the model for the deposit volumes is based on lagged deposit volumes, deposit rates and interest rates.

Finally, Ahmadi-Djam and Belfrage (2017) tried to apply different time series models to predict daily deposit volumes. The models they investigated were Holt-Winters, a simple stochastic factor model, an ARIMA, a SARIMA and a SARIMAX with dummy variables (certain week days). It is not clear which model yielded the best result, since they have tested the different models on different time windows. However, they concluded that a naive model was the best in their case due to its relatively high predictive power and its simplicity. Furthermore, they also conclude that an interesting future topic would be to come up with a model that incorporates expertise in the model (a mix of an "expert model and a time series model").

Thus, this report will build upon the findings of Achmadi-Djam and Belfrage (2017) when it comes to volume modeling. They concluded that a quantitative model that can incorporate expertise would be interesting to experiment with. Hence, a new model (a modified generalized additive model) is selected for comparison with different ARIMA-models. The reasoning is that this model is a mix between quantitative modeling and added expertise, since it is a model that allows for easy tuning of different factors (such as trends and business-related events). Furthermore, time-series modeling with macroeconomic variables is not something that has been widely researched and that is also something that this report analyses.

Moreover, the deposit rates are zeroed and the market rate is negative which currently makes some of the topics of risk management uninteresting. However, it is still important to model and thus this report will focus on a simpler static replicating portfolio that is modified to fit the regulations set by the European Banking Authority. Additionally, the liquidity risk framework of Kalkbrener and Willing (2004) will be tested by using the volume models obtained
in this thesis. As a final note, all data is different for each financial institution, Hence, it is important to make a pre-study and test already established methods and models, before moving on to more complex ones.
Chapter 3

Theoretical background

3.1 Market interest rates

3.1.1 Zero rates

Coupon bearing bonds are in essence a loan, or a debt obligation, that pays interest payments periodically at a fixed rate or as a percentage of the principal (or face value) F. A price $P_0$ is paid at time $t = 0$ and thereafter a coupon $c_t$ is received at set times $t_0, t_1, ..., t_n$ until maturity $T$, whence also a principal is received. The cash flows are aggregated and the value of the coupon bond is discounted back to today’s value, in order to set a fair price today. Hence, the following formula is the present value of a coupon bond:

$$P_0 = P(0, T) = \sum_{i=1}^{n} c_{t_i} e^{-y(t_i)t_i} + F e^{-y(T)T}. \quad (3.1)$$

Here $y(t_i)$ is the continuously compounded zero rate. Thus, a zero coupon bearing bond’s (henceforth zero coupon bond) present value is defined as

$$P(0, T) = F e^{-y(T)T}. \quad (3.2)$$

There are several financial derivatives and assets from which the market rates can be derived. In this paper, STIBOR and government bonds are used. STIBOR (Stockholm Interbank Offered Rate) is in short the average rate at which a number of Swedish banks are willing to lend out money to each other [10]. They are quoted as annual returns and are considered to be zero rates. Additionally, the same is true about the characteristics of the government bonds.
3.1.2 Modeling of the short rate

There are multiple ways to model the evolution of the market rates. It is most common to assume that the theoretical short rate under the objective probability measure $P$ has the following dynamics

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW(t)$$ (3.3)

where $\mu$ and $\sigma$ are the drift and diffusion terms respectively. Furthermore, $W(t)$ is a Wiener process (or a Brownian motion) with the following properties

1. $W(0) = 0$ a.s
2. $W(t)$ has independent increments: if $v < s < t < u$, then $W(u) - W(t)$ and $W(s) - W(v)$ are independent stochastic variables
3. $W(t)$ has Gaussian increments, $W(s) - W(t) \sim N(0, t - s)$, where it is assumed that $s < t$
4. The process $W(t)$ has continuous trajectories

A money account $B(t)$ is also used, where the dynamics of the process are given by

$$dB(t) = r(t)B(t)dt$$ (3.4)

Under the martingale measure $Q$, the dynamics of $r(t)$ are

$$\begin{cases} dr(t) = (\mu(t, r(t)) - \lambda(t, r(t)))dt + \sigma(t, r(t))d\bar{W}(t) \\ r(0) = r_0 \end{cases}$$ (3.5)

where $\lambda(t)$ is the market price of risk, which is the risk premium per unit of volatility and $\bar{W}(t)$ is a Wiener process under the martingale measure $Q$ [2]. The martingale measure (or risk-neutral measure) is the measure under which each price is exactly equal to the discounted expectation under this measure of the price [7]. The drift term is usually concatenated as $\mu(t, r(t))$ under $Q$.

In this report the Vasicek one factor model is adopted, which is a model that is widely used within the financial sector. The dynamics under the martingale measure $Q$ are specified as
\[ \begin{align*}
dr(t) &= \kappa(\theta - r(t))dt + \sigma dW(t) \\
r(0) &= r_0
\end{align*} \]  
(3.6)

where \( \kappa \) is the mean reversion strength, \( \theta \) is the long term mean and \( \sigma \) is the same diffusion (volatility) as before. All of these aforementioned stochastic differential equations can be solved analytically [6], specifically the strong solution to the linear SDE (3.6) is

\[ r(t) = r_0 e^{-\kappa t} + \theta (1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)}dW(s). \]  
(3.7)

### 3.1.3 Calibration of Vasicek’s model

The Vasicek model’s parameters need to be calibrated to the data which it should model. In the Vasicek model these parameters are kept constant and can be either estimated using least squares regression or maximum likelihood estimation (MLE). Here, the MLE approach will be used. MLE is a statistical approach which estimates the parameters given observed historical data.

Assume that a random sample of \( n \) observations \( x_1, x_2, \ldots, x_n \) is obtained with density function \( f(x_i | \Theta) \), with \( \Theta \) being a vector of the parameters which are to be estimated. The objective is to find a \( \hat{\Theta} \) that matches \( \Theta \) as closely as possible and this is done by using the likelihood function

\[ \mathcal{L}(\Theta; x) = f(x_1, x_2, \ldots, x_n | \Theta) = \prod_{i=1}^n f(x_i | \Theta) \]  
(3.8)

Now, take the logarithm and average of (3.8) to form a new function \( \hat{l} \)

\[ \hat{l}(\theta; x) = \frac{1}{n} \ln \left( \mathcal{L}(\Theta; x_1, x_2, \ldots, x_n) \right) = \frac{1}{n} \sum_{i=1}^n \ln \left( f(x_i | \Theta) \right) \]  
(3.9)

Then, the optimal estimates \( \hat{\Theta} \) that maximize (3.9) can be found. This is the general approach of MLE. To directly apply this method to estimate the parameters of Vasicek, a few characteristics need to be stated. First off, the transition density function of the Vasicek short rate is normal [8]. Hence, the conditional expected value and variance can be expressed as

\[ \begin{align*}
m_s(t) &= \theta \kappa B(s, t) + r_s(1 - \kappa B(s, t)) \\
v_s(t) &= \sigma^2 \left( B(s, t) - \frac{1}{2} \kappa B(s, t)^2 \right)
\end{align*} \]  
(3.10)
where it is presumed that $0 < s < t$. Furthermore, $B(s, t)$ is defined as

$$B(s, t) = \frac{1 - \exp(1 - \kappa(t - s))}{\kappa}$$

(3.11)

hence, 3.9 can then be expressed in the following form

$$\hat{l}(\theta, \kappa, \sigma; \mathbf{x}) = -\frac{1}{2} \sum_{i=1}^{n} \left( \ln \left( 2\pi v_{t-1}(t_i) \right) + \frac{(r_{t_i} - m_{t-1}(t_i))^2}{v_{t-1}(t_i)} \right) \quad (3.12)$$

It should be noted that since the Vasicek model only has three parameters the fit will never be perfect.

### 3.2 Deposit interest rates

#### 3.2.1 Jarrow and Van Deventer

Jarrow and Van Deventer (1998) presents two deposit rate models. One is for discrete time and one is for continuous time [13]. Here, the discrete time model is presented since the other model is not relevant for this thesis.

$$\Delta d_t = d_{t-1} + \beta_0 + \beta_1 r_t + \beta_2 (r_t - r_{t-1})$$

(3.13)

Equation (3.13) is a difference equation and it describes the stochastic change of the deposit rate, where $r_t$ is the market rate and $d_t$ is the deposit rate. The solution to the equation can be obtained by using successive substitution and can then be rewritten in an iterative form as

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

(3.14)

where $\beta_0$, $\beta_1$ and $\beta_2$ are the coefficients that need to be estimated using, for instance, OLS.

#### 3.2.2 Regression model

Another deposit rate model that will be used in this thesis is a model that is based on regression. In theory, the bank sets a deposit rate in relation to the market rate(s). Hence, an adequate model of the deposit rate would be a linear combination of a set of STIBORs with different maturities.
\[ d_t = \max (\beta_0 + \sum_{i=1}^{n} \beta_i r_{i,t}, 0) \] (3.15)

where \( d_t \) is the deposit rate, \( \beta_i \) are the coefficients that need to be estimated, \( r_{i,t} \) are the different market rates at time \( t \) and \( n \) is the number of available market rates.

### 3.3 Time series models

#### 3.3.1 ARIMA models

There are many different methods and techniques to model and forecast time series data. One of the most popular is the ARIMA model, which stands for Auto Regressive Integrated Moving Average, introduced by Box and Jenkins [5]. By analysing a time series of univariate historical data, its trends and seasonalities, an ARIMA model can be constructed in order to forecast a future path. Furthermore, the AR-component in ARIMA stands for autoregression and is a component that uses the dependent relationship between an observation and some number of lagged observations. Additionally, MA stands for moving average and is a component that uses the dependency between an observation and the residual error from a moving average model applied to lagged observations.

ARIMA is the generalization of simpler ARMA models with the added notion of integration. The integration is expressed as \( I(d) \), where \( d \) stands for the number of differences applied to the time series. Differencing is often a useful technique to make a time series stationary. Stationarity means that a times series’ mean and variance do not change over time, i.e. its short term patterns are consistent through time. The stationarity aspect is crucial, since the forecasting method is based on the assumption that the time series are (approximately) stationary.

The standard notation is ARIMA(p,d,q), where the arguments are substituted with numeric values depending on how many AR, I and MA-terms that are used. More precisely, the parameters are defined as: \( p \) number of lag observations included, also called lag order, \( d \) is the number of times the time series has been differentiated to produce a stationary time series, \( q \) represents the number of lagged forecasted errors in the prediction equation.
When a time series exhibits seasonality, a SARIMA (seasonal ARIMA) model might be more suitable. The SARIMA model is applied to the times series \( y_t \) as follows

\[
\Phi(B^s)\phi(B)\nabla^d\nabla_s^D y_t = \Theta(B^s)\theta(B)\epsilon_t. \tag{3.16}
\]

The standard notation for these models is SARIMA(p,d,q)(P,D,Q)s, where \( s \) is the seasonal component, for example \( s = 12 \) for monthly data. \( B \) stands for the lag operator and \( \epsilon_t \) is the Gaussian white-noise process with mean zero and variance \( \sigma^2 \). Furthermore, the difference operator is \( \Delta^d \) where \( d \) stands for the number of differences applied to the time series and the seasonal difference is \( \Delta_s^D \) where \( D \) stands for the number of seasonal differences. The first difference operator is simply the difference between current value and the previous value i.e. \( y_t - y_{t-1} \) and the seasonal difference is the series of changes from one season to the next, i.e. for the monthly data it is \( y_t - y_{t-12} \). Finally, the difference operators are applied to transform the non-stationary time series \( y_t \) to the stationary time series \( y_t^* \) in the following way

\[
y_t^* = \nabla^d\nabla_s^D y_t. \tag{3.17}
\]

The polynomials \( \phi(B) \) and \( \theta(B) \) in the lag operator are defined as

\[
\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p, \tag{3.18}
\]

\[
\theta(B) = 1 - \theta_1 B - ... + \theta_q B^q. \tag{3.19}
\]

The seasonal polynomials \( \Phi(B^s) \) and \( \Theta(B^s) \) are defined as

\[
\Phi(B^s) = 1 - \Phi_1 B^s - ... - \Phi_p B^{ps}, \tag{3.20}
\]

\[
\Theta(B^s) = 1 - \Theta_1 B^s - ... + \Theta_Q B^{qs}. \tag{3.21}
\]

Another statistical model is the SARIMAX which extends the SARIMA model by including exogenous covariates in order to take advantage of the potential linear effect that one or more exogenous series have on the stationary time series \( y_t \) [3]. The general condensed form in the lag operator notation is as follows

\[
\Phi(B^s)\phi(B)\Delta^d\Delta_s^D y_t = c + x_t'\beta + \Theta(B^s)\theta(B)\epsilon_t. \tag{3.22}
\]
where the vector $x'_t$ holds the values of the $r$ exogenous variables, time-varying predictors at time $t$, with coefficients denoted $\beta$ and $c$ is the intercept.

In order to evaluate the parameters for the SARIMA model one can investigate the Autocorrelation Function (ACF) which tells how the time series is correlated with itself. The ACF of the stationary time series is as follows

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad (3.23)$$

where $\gamma(h) = Cov(X_{t+h}, X_t)$ is the autocovariance function of the time series for the different lags $h$.

The Partial Autocorrelation Function gives the partial correlation of a stationary time series with its own lagged values, regressed on the values of the time series at all shorter lags. The PACF of the stationary time series is as follows

$$\alpha(0) = 1 \quad (3.24)$$

and

$$\alpha(h) = \phi_{hh}, \quad (3.25)$$

for the $h \geq 1$, where $\phi_{hh}$ is the last component of

$$\phi_h = \Gamma_h^{-1} \gamma_h, \quad (3.26)$$

where $\Gamma_h = [\gamma(i-j)]_{i,j=1}^{h}$ and $\gamma_h = [\gamma(1), \gamma(2), ..., \gamma(h)]'$.

To evaluate the accuracy of the fit and the prediction one can use different measures. Here, the root mean squared error (RMSE) is presented which is an average of the squared forecast error values. Squaring the values of the forecast errors makes them positive and also puts more weight on large errors. The RMSE is defined as follows

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2}. \quad (3.27)$$

MAPE, which stands for Mean Absolute Percentage Error, is also a common accuracy measure which expresses accuracy as a percentage and is defined as
Another useful metric is the Akaike information criterion (AIC), which is an estimator of the relative quality of a statistical model for a given set of data. The mathematical definition is as follows

\[ AIC = n_l \ln \left( \frac{SSE}{n_t} \right) + 2(n_p + 1) + \frac{2(n_p + 1)(n_p + 2)}{n_t - n_p - 2} \]  

(3.29)

where \( n_t \) is the number of data points, \( n_p \) is the number of estimated model parameters and \( SSE \) is the error sum of squares.

To summarize, an ARIMA-model can be viewed as a “filter” that tries to separate the time series data from the noise and thereafter the time series can be extrapolated into the future to obtain forecasts.

### 3.3.2 Augmented Dickey-Fuller test

The augmented Dickey-Fuller test is used to get an estimation of whether or not a time series is likely to be stationary. The null-hypothesis is that there is a unit root present in a sample of the time series (which implies that the time series is non-stationary). The form of the model which the test is carried out on is[5]

\[ \Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 y_{t-1} + ... + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t \]  

(3.30)

where \( \alpha, \beta, \gamma \) and \( \delta \) are coefficients. Under the null hypothesis, \( \gamma \) is equal to zero and for the alternative hypothesis \( \gamma < 0 \). The test statistic \( \frac{\hat{\gamma}}{SE(\hat{\gamma})} \) is then compared to a relevant critical value, which is usually at a 95%-confidence level. Meaning, that if the p-value is less than 0.05 we can reject the unit root assumption. However, it should be noted that just because the test is successful does not mean the time series is actually stationary, but it is some indication.

### 3.4 Regression models

#### 3.4.1 Random Forest Regression

Since the random forest regression is just a minor part of the thesis, it will be kept rather short. Random Forest is a supervised learning algorithm that
can be used to rank the feature importance of lagged dependent variables. It is a meta estimator that fits a number of classifying decision trees on various sub-samples and uses averaging to improve the predictive accuracy and control over-fitting.

One first need to understand how a decision tree is formed. It is worth noting that it can be used both as a classification and as a regression algorithm. Since regression is used in this thesis, that is the method that will be briefly explained here. There are in essence two steps

- Dividing the predictor space, the set of possible values for the observations, into J distinct regions $R_1, R_2, \ldots, R_J$
- Every observation that falls into the region $R_j$ is assigned the same prediction value, which is the mean of the response values for the training observations in $R_j$.

These regions are constructed in the form of high-dimensional boxes where the residual sum of squares (RSS) are minimized [12]. Here, the RSS are defined as

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$$

(3.31)

where $y_i$ is the response variable and $\hat{y}_{R_j}$ is the mean response for the training observations within the j:th box. At each node the best split is done, without respect to future splits that might have lead to a better tree in a future step. Hence, this is a greedy approach and it is called recursive binary splitting. Recursive binary splitting is performed by selecting a predictor $X_j$ and a cut-point $s$, such that the predictor space is partitioned into regions $(X | X_j < s)$ and $(X | X_j \geq s)$, which leads to the greatest reduction of the RSS.

Now, to create a random forest model one first creates a significant amount (400-1200) of decision trees by randomly sampling from the training dataset. For each tree, samples are drawn with replacement, meaning the trees are bagged. Hence, not every tree "sees" all the features and the consequence is that the trees are de-correlated, which avoids the risk of over-fitting. Furthermore, a regression model is fitted of the selected features in order to predict the target variable. Additionally, for all the features available to the tree the best features are chosen by the impact that feature has on the impurity and that
is how the split in the node is done. In regression, the impurity is quantified as variance. Hence, the features importance can be computed by measuring how much the feature decreases the weighted impurity. These results are then averaged for the full forest to get the feature importance score, which is on the interval $[0, 1]$ [4].

![Random Forest](image)

**Figure 3.1:** An illustration of a random forest with decision tree one, two and $n$ shown. Each circle represents a node where a split is done.

Mathematically, a random forest regression can be expressed as

$$
\bar{r}_n(X, D_n) = \mathbb{E}_{\Theta}[r_n(X, \Theta, D_n)]
$$

(3.32)

where $r_n(X, \Theta, D_n)$ is a randomized base regression tree, $\Theta_1, \Theta_2, ..., \Theta_n$ are i.i.d outputs of a randomizing variable $\Theta$ that is used to determine how the successive cuts are performed when building the individual trees, $\bar{r}_n(X, D_n)$ is the aggregated regression estimate and $D_n$ is the data-set which can be written as $D_n = \{(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)\}$. Here, $X_i$ are the features that seek to explain the endogenous variable $Y_i$.

### 3.4.2 Multiple linear regression and OLS

In general, the dependent variable $y$ may be linearly related to some $k$ number of predictor variables. The model

$$
y(t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon_t
$$

(3.33)

is called multiple linear regression with $k$ regressors and $\beta_j, j = 0, 1...k$ are called regression coefficients. The $\epsilon_t$ is the Gaussian white-noise process with mean zero and variance $\sigma^2$ [16]. The multiple linear regression model can be interpreted as a hyper-plane in the $k$-dimensional space of the regressor variables $x_j$. 
The fitted regression equation or model is then typically used in order to forecast future observation of the response variable. The method of least squares (OLS) can be used to estimate the regression coefficients. The least squares function is defined as follows

\[
S(\beta_0, \beta_1, ..., \beta_k) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2
\]  

(3.34)

After minimizing the OLS function with respect to \( \beta_0, \beta_1, ..., \beta_k \) one obtains the least squares normal-equations’ solution which yield the least squares estimators \( \hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k \).

Finally, it should be noted that the inclusion of lagged dependent variables i.e. \( y(t - 1), y(t - 2), ..., y(t - k) \) can suppress the explanatory power of other variables. Meaning that variables that are interpreted as insignificant when including the lagged variable might actually be significant [1].

### 3.4.3 Modified generalized additive model

Another statistical model that will be used is a modified generalized additive model (GAM), which will be referred to as the MGAM in this report. The MGA models are inherently different from time series models such as ARIMA, since MGA models are based on curve fitting while ARIMA models try to filter the time series from the noise [17].

The MGA model consists of three main components: trend, seasonality and a so called holiday component. These three components are combined into the following equation

\[
y(t) = g(t) + s(t) + h(t) + \epsilon_t,
\]

(3.35)

where \( g(t) \) is the trend function which models non-periodic changes, \( s(t) \) is the seasonal function which models periodic changes such as weekly and yearly seasonality, \( h(t) \) is the holiday effect component, which models irregular occurring events in the time series \( y(t) \) and \( \epsilon_t \) is the Gaussian distributed error term.
The trend changes are incorporated by explicitly defined changepoints where the growth rate is allowed to change. First, let $S$ be the number of changepoints at times $s_j$, for $j = 1, \ldots, S$. Next, define a vector of rate adjustments $\delta$, where each element $\delta_j$ in the vector is the change of rate that occurs at time $s_j$. The rate at any time $t$ is expressed as the base rate $k$ plus all the previous adjustments up to time $t$, i.e. $k + \sum_{j:t>s_j} \delta_j$. This is expressed more neatly on vector form as $k + a(t)^T \delta$, where the vector $a(t) \in \{0, 1\}^S$ can be expressed as

$$a_j(t) = \begin{cases} 1, & \text{if } t \geq s_j, \\ 0, & \text{otherwise}. \end{cases} \quad (3.36)$$

When the rate $k$ is adjusted then an offset parameter $m$ must also be adjusted in order to connect the endpoints of the segments. The right adjustment at changepoint $j$ is computed by

$$\gamma_j = (s_j - m - \sum_{l<j} \gamma_l) \left( 1 - \frac{k + \sum_{l<j} \delta_l}{k + \sum_{l\leq j} \delta_l} \right). \quad (3.37)$$

Moreover, since it is assumed that the time series does not exhibit a saturated growth (it is not expected to reach a ceiling in the dependent variable $y(t)$), the trend component $g(t)$ is modeled by a piece-wise constant rate of growth. The trend model is expressed as

$$g(t) = (k + a(t)^T \delta) t + (m + a(t)^T \gamma), \quad (3.38)$$

where as before the constant $k$ is the rate of growth, $\delta$ has the rate adjustments, another constant $m$ is the offset parameter and $\gamma$ is set to $-s_j \delta_j$ to make the function continuous. The changepoints $s_j$ can be specified automatically by constructing a sparse prior on $\delta$. For example, the sparse prior can have a Laplace distribution, $\delta_j \sim \text{Laplace}(0, \tau)$. Hence, the parameter $\tau$ controls the flexibility of the trend component by altering its rate.

The seasonal component relies on Fourier series to flexibly model periodic effects in the time series. A standard Fourier series is defined as

$$s(t) = \sum_{n=1}^{N} \left( a_n \cos \left( \frac{2\pi nt}{P} \right) + b_n \sin \left( \frac{2\pi nt}{P} \right) \right) \quad (3.39)$$
where $P$ is the regular period we expect in the time series (i.e. $P = 365.25$ for yearly data or $P = 7$ for weekly). Moreover, fitting the seasonal component to the series requires estimating the $2N$ parameters $\beta = [a_1, b_1, ..., a_N, b_B]$, which is done by forming a matrix of seasonality vectors for each time $t$ both in the historical and future data. For example, if seasonality is $N = 10$ then

$$X(t) = \left[ cos\left(\frac{2\pi (1)t}{365.25}\right), ..., sin\left(\frac{2\pi (10)t}{365.25}\right) \right].$$

(3.40)

The seasonal component is then

$$s(t) = X(t)\beta$$

(3.41)

where $\beta \sim \mathcal{N}(0, \sigma^2)$.

Finally, the holidays or events component is included in the model to provide somewhat predictable "shocks" to the time series, since these kind of abnormalities do not follow a seasonal pattern. These events can be seen as dummy variables. The effect of repeating events in the data is approximately similar from year to year. Mathematically, the dummy variables in this model are incorporated by the use of an indicator function. For each event $i$, let $D_i$ contain the past and future dates for that specific event. Each event is assigned a parameter $\kappa_i$, which is the corresponding change in the forecast. Then, one can construct a matrix of regressor parameters as $Z(t) = [(t \in D_1), ..., (t \in D_L)]$. Hence, the events matrix is defined as follows

$$h_t = Z(t)\kappa$$

(3.42)

where $\kappa \sim \mathcal{N}(0, v^2)$.

The MGAM has a lot of degrees of freedom and is based on many ad hoc assumptions, specially in the case of the prior parameters. Therefore, this model should be applied with caution, especially for forecasting purposes. However, at the same time the MGAM offers a lot of flexibility and tuning possibilities, as well as interesting decomposition results of the times series as part of the analysis.
3.5 Risk management

3.5.1 Static replicating portfolio: Maes and Timmerman

In order to evaluate some of the risk measures for NMDs a static replicating portfolio method is employed. Another use of the portfolio is that it can be used to model the deposit rate.

The idea is to create a portfolio of real (or fictional) risk-free assets with fixed weights that replicates the behaviour of the deposits as close as possible. In other words, one can transform the non-maturing deposits into a portfolio of vanilla instruments that makes it possible to quantify some of the sought after risk measures. Maes and Timmerman (2014) proposed a model with two different objective functions. The first objective function is to minimize the standard deviation of the spread between the return of the portfolio and the deposit rate, i.e minimize the standard deviation of the margin.

\[
\text{min} \left( \text{std}(r_p^t - d_t) \right)
\]

(3.43)

Subject to the constraints:

1. \( \sum_{i=1}^{n} w_i r_i = r_p^t \)
2. \( \sum_{i=1}^{n} w_i = 1 \)
3. no short sales are allowed, i.e. \( w_i \geq 0 \)

where \( r_p^t \) is the return of the portfolio, \( d_t \) is the deposit rate, \( n \) is the number of assets, \( w_i \) is the weight for the \( i \)th asset and \( r_i \) are the market rates. The first condition is technically not a constraint, but it rather explains how the return is derived. It simply states that the sum of the weights multiplied by the corresponding risk-free asset is equal to the return of the portfolio. Moreover, the second constraint ensures that all of the available volume for investment is invested. Finally, the third constraint affirms that no short-selling is allowed.

For the second approach the objective function is to maximize (a slightly modified) Sharpe ratio:

\[
\text{max} \left( \frac{r_p^t - d_t}{\text{std}(r_p^t - d_t)} \right)
\]

(3.44)
under the same constraint as for the minimization objective function. In other words, to maximize the risk adjusted average return of the portfolio [15]. Both of these non-linear optimization problems are convex since the objective functions are convex and the constraints are linear [11].

When the weights are computed one can then obtain the duration of the replicated deposit by taking the sum of the weighted time to maturity for each asset

\[ D_p = \sum_{i=1}^{n} w_i m_i. \] (3.45)

Here \( D_p \) is the duration of the portfolio and \( m_i \) is the time to maturity for the \( i:th \) asset. In other words, the duration can be viewed as the average time to repricing of the deposits.

### 3.5.2 Liquidity risk

There are essentially two types of liquidity risk: funding risk and market liquidity risk. Funding risk is the risk of not being able to meet certain payment obligations without incurring a substantial cost. Meanwhile, market liquidity risk is the risk of not being able to immediately sell a financial instrument without a substantial loss. In this thesis the former is discussed. In order for the bank to meet their payment obligations the deposit volumes must obviously be large enough. Hence, it is of interest to know what the lowest available volume will be in a specified time interval. In this report, the liquidity risk analysis will be based on the framework suggested by Kalkbrener & Willing (2004).

Let \( V(s) \) be a stochastic volume process on the interval \([0, T]\), where \( s < T \), then the minima process \( M(t) \) is simply

\[ M(t) := \min_{0 \leq s \leq t} V(s) \] (3.46)

which can be interpreted as the worst case scenario (or the lowest volume) in that period of time. Several volume paths are simulated in order to estimate the \( p \)-quantile of \( M(t) \), where \( p \) is the desired level of confidence. Thus, the \( p \)-quantile of \( M(t) \) is the term-structure of liquidity, which is denoted as \( TSL(t, p) \). Hence, the probability that the volume drops below \( TSL(t, p) \) for each fixed \( t \) is \( p \). In other words, the available volume at the bank during the period \([0, t]\) is at least \( TSL(t, p) \) with probability \((1-p)\). Furthermore, it is evident by equation (3.46) that \( TSL(t, p) \) is non-increasing [14].
Chapter 4

Data

4.1 Description of the data

4.1.1 Volume data

In this project the analysis will mainly be done on the aggregated monthly volume observations of three different deposit accounts: private transaction accounts, savings accounts and corporate savings accounts. The data is observed from the end of January 2010 to the end of November 2018 (a total of 107 observations), which have been provided by the bank.

![Figure 4.1: Volumes of different deposit types](image)

In Figure 4.1 the volumes have been scaled due to business secrecy.
4.1.2 Market rate data

From Riksbanken.se the following market rates were acquired: STIBOR O/N, 1W, 1M, 2M, 3M, 6M and Swedish government bonds with maturities 2Y, 5Y, 7Y and 10Y. They are quoted as the monthly average rate for each maturity. Moreover, the obtained sample has the same period as the volumes.
4.1.3 Deposit rate data

The Bank also provided monthly average deposit rate data for every deposit type (same period as for the volumes).

![Deposit rates graph](image)

*Figure 4.3: Swedish market rates*

*Figure 4.4: Deposit rates*
4.1.4 Additional data

Some models that are analyzed in this thesis also utilize macro-economic variables. In addition to the market and deposit rate, it also elected to include monthly unemployment rate for the population between the ages 20-64 and the average monthly salary in the public and private sector in Sweden (SCB).

Intuitively, all of these variables should have an instant impact on the volumes except perhaps the market rate, since the general individual is rather risk averse by nature and would wait with an investment until the market is rather stable. It is expected that the unemployment rate has an anti-correlation with the volumes, since in theory the average amount of wealth that people have would decrease if unemployment increases and all other variables are kept constant. Furthermore, individuals and businesses would intuitively move money to the market if the market is good. Hence, STIBOR 1M would also be anti-correlated with the volumes. In this report, the STIBOR 1M is a very rough estimation of how good the "alternative" market currently is in relation to the deposit accounts. However, it should be noted that another interesting alternative would be to incorporate the volatility of the stock market, i.e the instability/stability of the stock market. Finally, an increase of salaries would have a positive impact on the volumes, especially for the savings accounts, since many customers move a portion of their salaries into their savings accounts each month.

4.2 Analysis of the data

4.2.1 Trends and seasonality in the volumes

Since the volume observations are monthly and recorded at the beginning of a new month a lot of information is lost. For example one would have seen a large increase of the volumes around the 25:th due to salaries and then a large decrease around the same time, since people are paying their bills among other things.

Ever since 2010 the GDP of Sweden has had a steady increase (see SCB), which might be one of the underlying reasons as to why all of the volumes have an up-going trend. Another explanatory reason could simple be an increase of new customers. Moreover, one can clearly see peaks at June of 2012 and June of 2013 for the private transaction accounts. This is further empha-
sized by the seasonal plots, where one can observe a peak at June almost every year. Furthermore, for each of the seasonal plots the volumes seem to behave rather similar each year, except for some peaks which might be due to local or worldwide events. For instance, the change of mortgage rates that took place in the last quarter of 2014 had a large impact on the transactional accounts. However, what is interesting to note is that the BREXIT announcement 2016, which caused the stock market to plummet, did not seem to have an effect on the deposit accounts.

Overall it is difficult to conclude which kind of events may affect the volumes by only looking at the monthly aggregated data set.

### 4.2.2 Market and deposit rates

In Figure 4.3 one can observe that the market rates all have a similar pattern but are somewhat shifted upwards, depending on the maturity of the bond. As expected the bonds with a longer maturity have a greater yield. Furthermore, it is also worth noting that from April 2015 all of the STIBORs are negative.

Next, the correlations between the market rates are investigated in order to find a market rate from which we will derive our (amongst other things) simulated market rate.

| Table 4.1: Correlations between different market rates |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | T/N 1.0         | STIBOR 1W 0.999445 | STIBOR 1M 0.998243 | STIBOR 2M 0.996994 | STIBOR 3M 0.994833 | STIBOR 6M 0.991364 | GVB 2Y 0.864318 | GVB 5Y 0.730346 |
| T/N 1.0         | 1.0             | 0.999445          | 0.998243          | 0.996994          | 0.994833          | 0.991364          | 0.864318        | 0.730346        |
| STIBOR 1W 0.999445 | 0.999445        | 1.0               | 0.999088          | 0.998045          | 0.996519          | 0.994915          | 0.993915        | 0.992232        |
| STIBOR 1M 0.998243 | 0.999088        | 1.0               | 0.999039          | 0.997562          | 0.995832          | 0.994182          | 0.992232        | 0.984043        |
| STIBOR 2M 0.996994 | 0.998045        | 1.0               | 0.999200          | 0.998061          | 0.996012          | 0.994002          | 0.992232        | 0.984043        |
| STIBOR 3M 0.994833 | 0.999039        | 1.0               | 0.999200          | 0.998061          | 0.996012          | 0.994002          | 0.992232        | 0.984043        |
| STIBOR 6M 0.991364 | 0.997562        | 1.0               | 0.998493          | 0.987647          | 0.978872          | 0.973193          | 0.972866        | 0.972866        |
| GVB 2Y 0.864318  | 0.995832        | 1.0               | 0.998493          | 0.987647          | 0.978872          | 0.973193          | 0.972866        | 0.972866        |
| GVB 5Y 0.730346  | 0.994182        | 1.0               | 0.998493          | 0.987647          | 0.978872          | 0.973193          | 0.972866        | 0.972866        |
| GVB 7Y 0.766080  | 0.992232        | 1.0               | 0.998493          | 0.987647          | 0.978872          | 0.973193          | 0.972866        | 0.972866        |
| GVB 10Y 0.997486 | 0.972866        | 1.0               | 0.998493          | 0.987647          | 0.978872          | 0.973193          | 0.972866        | 0.972866        |

As can be seen in Table 4.1 the different STIBORs are most correlated with STIBOR 1M. Hence, this particular rate will be used as the basis in the analysis.

At last, the relationship between the market rate and the deposit rate for the different deposit types is investigated.
The deposit rates have a staircase pattern, which is obvious since the rate is set as constant between two time points. Additionally, the deposit rates seem to follow the market rate almost perfectly in its "sudden" increases and decreases. If the deposit rates had been observed daily, we would probably have seen a lag in the increase of the deposit rates as the market rate increase. That is because the bank does not want to prematurely increase the deposit rates, since if the increase of the market rate was just a random fluctuation and not a steady increase the bank would loose a lot of money. Finally, as can be seen in Figure 4.5 the deposit rates are floored at zero as soon as the market rate hits a certain threshold. That is because administration costs and other costs need to be covered. Thus, the bank would like to set a negative rate as seen from a cost perspective. However, for the majority of the accounts the bank chooses not to set a negative rate. That is because that would likely deter customers from deposit their funds at the bank.

### 4.3 Extracurricular volume data

Later in the project, daily aggregated data, see Figure 4.6, was provided for the private transaction accounts. It should be noted that since this data was not included in the original data set it has not been under the same rigorous analysis as the other monthly data set. The observation period is from 2014-04-01 to 2018-11-30.
Figure 4.6: Daily aggregated data for the private transaction accounts

By looking at the graph for the aggregated daily data one can, as expected, see that the time series exhibit a clear seasonal behaviour with a distinct up-going trend.
Chapter 5

Methodology

Here, it is explained how the theory presented in Chapter 3 is applied to model the different areas of this thesis. The general approach is to model each area in a specific order, since some results are needed for models that are used later in the project. First off, the market rate will be modeled using Vasicek’s model, which is a model that needs to be calibrated as explained in Section 3.6. When a model for the market rate is obtained, one can then continue to make an approximation of the deposit rate. These two variables are usually the foundation that is needed when modeling non-maturing deposits. After having sufficient models for the aforementioned variables, one can then move on to model the volumes. In the literature it is often proposed to use a stochastic factor model to predict the volumes, but it is omitted in this report due to spurious results which are most likely because of the negative market rate and zeroed deposit rates.

The next step is basically the objective when modeling non-maturing deposits: risk management. First off, the estimated duration of the deposits will be computed using the static replicating portfolio approach. In order to use this approach, one must also estimate the core part of the deposits. Consequently, the core part is estimated using both historical data and using the worst outcome of several volume predictions. Furthermore, these volume predictions are used to gauge the liquidity risk for each of the deposit types.

Since the methodology of how the market and deposit rates are modeled is clear, these are not further explained in the methodology. Hence, emphasize is put on how the volume models and risk management are constructed.
5.1 Volume models

The idea is to get a reliable volume model which can be used to estimate future volumes. Additionally, one would then potentially be able to use this model to compute the (future) core part and liquidity.

First, an econometric analysis will be performed in order to find potentially impactful variables that are to be included and tested in the models. A test of correlation and significance level is computed, together with previous research and our own intuition, to decide which variables to carry on with.

The next step is to model the actual volumes. Two different approaches will be compared (autoregressive integrating moving average and a general additive model). Each of these models will be evaluated and validated by their RMSE and MAPE for a training and test set.

5.1.1 ARIMA models

Several different ARIMA-models will be tested with respect to seasonality and the exogenous parameters that are the most significant. In other words a SARIMA model and a SARIMAX model are compared. Hence, two SARIMA(X) models for each deposit type are investigated and compared. As for the choice of exogenous variables, there are many potential alternatives. However, the variables investigated are based on a short econometric analysis.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Private transaction accounts</th>
<th>Savings accounts</th>
<th>Corporate savings accounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARIMA</td>
<td>Model 1.1</td>
<td>Model 2.1</td>
<td>Model 3.1</td>
</tr>
<tr>
<td>SARIMAX</td>
<td>Model 1.2</td>
<td>Model 2.2</td>
<td>Model 3.2</td>
</tr>
</tbody>
</table>

Above are the names of the models which will be referenced later on. To clarify, the first model type is a fitted SARIMA model and the second is a SARIMAX model with only the most significant variable(s).

Next, the parameters of each model need to be found. This is done in multiple steps, as explained in Section 3.3.1. Furthermore, a random forest regression is used in order to find the impact that different lags of the endogenous variable have on the sought after variable. This is done to further emphasize the results interpreted from the ACF and PACF plots. The models are trained on a sample
consisting of 80% of the total available observations and are then tested on the remaining 20%.

All of the models presented in Table 5.1 are evaluated based on three different measures in addition to the ACF and QQ plot of the residuals. In this report, the chosen measures are AIC, RMSE and MAPE, where RMSE and MAPE are computed for both the fitted and the forecasted volumes.

Finally, the significance of the exogenous variables are evaluated based on their p-value and cross-correlation. Hence, the models are based on the same SARIMA-terms in order to evaluate the exogenous variables’ impact.

5.1.2 Modified generalized additive model

The modified generalized additive model (MGA model) is applied both to the monthly and extracurricular daily data sets. The model is calibrated as follows:

First off, the model is fitted to the data without any tuning of the parameters. If the out of sample prediction is showing too much seasonality or if its too tightly fit to the historical data, one has to tune the prior (which is Laplace-distributed) that controls the number of change points to include. In other words, the flexibility of the model with respect to the trend. Moreover, the seasonal flexibility can be tuned by changing the constant $N$, which controls the number of sinusoids and cosinusoids terms in the Fourier series. The higher $N$, the faster the model adapts to changes in seasonality. However, the drawback is that the risk of overfitting increases. Furthermore, the seasonality can also be adjusted by changing $\sigma$ in $\beta$’s prior, which is normal distributed.

Now, a model is fitted to the daily aggregated data without any adjustments of the monthly seasonality to investigate the behaviour of the model. The model, as expected, did not capture any of the abrupt jumps in the data. Therefore, dummy variables are introduced around the $23th – 26th$ day for each month due to the increase of the volumes around this time (which is most likely due to salaries). These dummy variables are calibrated by looking at historical calendar days for each month. Another dummy variable is introduced around the $12th – 14th$ (depending on when weekends occur) for each month. That is, because on average, it is observed that the majority of the withdrawals are done in this time frame.
Moreover, since seasonality in this particular model is modelled by using Fourier series it is possible to change the sensitivity of the prior scale of the seasonal behaviour, which has been done since it improved the accuracy. However, one has to be careful to not overfit the model.

Finally, other tweaks and tests did not dramatically improve neither MAPE nor RMSE, therefore no more improvements were made.

## 5.2 Risk management

Here, the method proposed by Maes and Timmerman (2005) is used with the alteration of an additional restriction on the duration of the portfolio. The added constraint is due to EBA’s new regulations that will come in effect this summer. Thus, the optimization problem can be formulated as

\[
\min \left( \text{std}(r^p_t - d_t) \right) \tag{5.1}
\]

Subject to the constraints

1. \[ \sum_{i=1}^{n} w_i r_i = r^p_t \]
2. \[ \sum_{i=1}^{n} w_i = 1 \]
3. no short sales are allowed, i.e. \( w_i \geq 0 \)
4. a \[ \sum_{i=1}^{n} w_i m_i \leq 5 \]
   b \[ \sum_{i=1}^{n} w_i m_i \leq 4.5 \]
   c \[ \sum_{i=1}^{n} w_i m_i \leq 4 \]

where \( r^p_t \) is the return of the portfolio, \( d_t \) is the deposit rate, \( n \) is the number of assets, \( w_i \) is the weight for the \( i \):th asset, \( r_i \) are the market rates and \( m_i \) is the time to maturity for the \( i \):th asset. The first of the last three conditions is for retail/transactional accounts, the second is for retail/non-transactional accounts and the third one is for wholesale accounts.

Additionally, another objective function will be tested which is the maximization of the (slightly edited) Sharpe ratio:

\[
\max \left( \frac{r^p_t - d_t}{\text{std}(r^p_t - d_t)} \right) \tag{5.2}
\]
under the same constraint as for the minimization problem. It should be noted that a total of four portfolios will be evaluated for each deposit type. For each objective function one will be tested without the duration constraint.

First off, 50 000 random weights are simulated to visualize the efficient frontier of the portfolios to get a good starting guess for the optimal weights. Furthermore, it also makes it easier to interpret the trade-off between the volatility and the margin. Then, it is just a matter of solving the convex optimization problems (5.1) and (5.2).

When the weights are obtained the duration of the portfolio (and in effect, the NMD it tries to replicate) can be computed as:

$$D_p = \sum_{i=1}^{n} w_i m_i$$  \hspace{1cm} (5.3)

The portfolios will also be calibrated on a window of three years that is continually moving. This is to evaluate the stability of the model (and the deposits’ interest rate sensitivity), but also to see how the portfolios evolves over time.

Finally, the liquidity risk will be modeled as outlined in Section 3.5.2. First, using Monte Carlo, 10 000 volume paths will be simulated using the model which gave the most accurate volume prediction. Let $V(t)^{(i)}$ denote the volume path for the $i$:th simulation. Then, the minima for each simulation is per definition expressed as:

$$M(t)^{(i)} := \min_{0 \leq s \leq t} V(s)^{(i)}$$  \hspace{1cm} (5.4)

$M(t)^{(i)}$ is thus a vector of length 10 000 that contains the minimum volume on the interval $[0, t]$, meaning that a matrix $M(t)$ for an end time $T$ and simulations $N$ can be formed as:

$$
\begin{bmatrix}
M(1)^{(1)} & M(1)^{(2)} & \ldots & M(1)^{(N)} \\
M(2)^{(1)} & M(2)^{(2)} & \ldots & M(2)^{(N)} \\
\vdots & \vdots & \ddots & \vdots \\
M(T)^{(1)} & M(T)^{(2)} & \ldots & M(T)^{(N)}
\end{bmatrix}
$$  \hspace{1cm} (5.5)

Next, for each time $t$, each row of matrix (5.5) is ordered in ascending order creating a new row vector $m(t)$. Effectively meaning that the worst case sce-
enario is the first element in the vector. Now, a $p$-quantile $TSL(t, p)$ can be formed by taking the $Np$:th element of $m(t)$ such as

$$TSL(t, p) = m(t)_{[Np]}$$

(5.6)

So that $TSL(t, p)$ is then with probability $(1-p)$ the least available amount on the account. Typically, the 95% and 99%-quantiles are of interest and these levels are also used in this thesis. Hence, $p$ is set to 0.05 and 0.01 respectively.
Chapter 6

Results

6.1 Short rate

The calibration of Vasicek’s model is first carried out on the data from January 2010 to November 2018. However, due to the change in the trend after year 2015 a calibration on the period of January 2016 to the end of the sample period is also conducted. The result is:

Table 6.1: Parameters of calibrated Vasicek’s model

<table>
<thead>
<tr>
<th></th>
<th>$\kappa$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All observations</td>
<td>0.0088</td>
<td>-0.0976</td>
<td>0.0035</td>
</tr>
<tr>
<td>Narrowed time-frame</td>
<td>3.8214</td>
<td>-0.0052</td>
<td>0.0030</td>
</tr>
</tbody>
</table>

Since the market rates have been negative since 2015 there is a downward trend and mostly negative rates, which is also reflected in the mean reverting parameter. Now Monte Carlo is used to simulate 10 000 paths of market rates 10 years into the future.
CHAPTER 6. RESULTS

Figure 6.1: Simulations of the market rate using the Vasicek short rate model for the complete data-set

Figure 6.2: Simulations of the market rate using the Vasicek short rate model for a shorter time period
It is noted that the calibration on the shorter time period yields a more "stable" result in the sense that the predicted short rate will fluctuate around the same value the upcoming years. Meanwhile, when calibrated on the complete data set the short-rate has a constant decline for the next 10 years.

### 6.2 Deposit rate

The deposit rate is, in this report, modeled in two different ways. The first one is a multiple regression model and the second is a method implemented by Jarrow and Van Deventer (1998). As for the first model, different variables (different STIBOR maturities) are tested in order to find a simple and suitable model of the deposit rate. The significance of adding more variables is not very high and the fit is only marginally increased. Hence, these variables are omitted in the final model. This result is further emphasized by looking at the correlation in Table 4.1. Thus, the final model is:

\[ d_t = \max(\beta_0 + \beta_1 r_t, 0) \]  \hfill (6.1)

where \( d_t \) is the deposit rate at time \( t \), \( \beta_0 \) is a constant, \( \beta_1 \) is a coefficient and lastly \( r_t \) is the STIBOR 1M.

<table>
<thead>
<tr>
<th>Deposit type</th>
<th>Intercept</th>
<th>( \beta_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Savings accounts</td>
<td>0.0011</td>
<td>0.5601</td>
</tr>
<tr>
<td>Corporate savings accounts</td>
<td>0.0010</td>
<td>0.5627</td>
</tr>
</tbody>
</table>

*Table 6.2: Values of parameters for the regression model*
40 CHAPTER 6. RESULTS

**Figure 6.3:** Deposit rate of savings accounts (left) and corporate savings accounts (right)

Here, the red line represents the market rate (MIR), blue is the true deposit rate (DIR) and the lighter blue line is the predicted deposit rate (Predicted). The plots of the deposit rate for the transaction accounts are omitted since the rate has been zero during the observation period. Furthermore, the plots of the out of sample tests are also omitted due to the market rate being negative, which just leads to zeroed deposit rates in that particular period. It is concluded that the most simple formula produces an adequate fit for the purposes of this thesis.

Next, the method of Jarrow and Van Deventer (1998), as explained in Section 3.2.1, is used to compute the deposit rates.

**Table 6.3:** Values of parameters for Jarrow and Van Deventer

<table>
<thead>
<tr>
<th>Deposit type</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Savings accounts</td>
<td>0.0002</td>
<td>-0.0110</td>
<td>0.7217</td>
</tr>
<tr>
<td>Corporate savings accounts</td>
<td>0.0002</td>
<td>-0.0113</td>
<td>0.7575</td>
</tr>
</tbody>
</table>
CHAPTER 6. RESULTS

(a) Fitted deposit rate of savings accounts (b) Fitted deposit rate of corporate savings accounts

Figure 6.4: JvD: Deposit rate for savings accounts (left) and corporate savings accounts (right)

The fit is slightly better at the beginning of the period, which is due to the added variables. However, since this model was proposed in a market free of negative rates it handles negative rates very poorly, which can be seen at the end of the sample period. However, this could easily have been remedied by just adding a max function where the rate is floored at zero. Nonetheless, here it is elected to keep the model in its original form.

6.3 Volume modeling

6.3.1 Econometric analysis

The volumes will be modeled in several different ways. One of these models is considered to be an econometric model which will be compared to a SARIMA model. Hence, it is important that the relationship between the volumes and a selected few variables is established and understood. In other words, we will first conduct an analysis to see how the volumes are affected by different economic variables. In this project, it is elected to analyze the impact of STIBOR 1M, unemployment rate, private and public salaries, lagged volume (i.e. $V_{t-1}$) and lastly the volume observation from the year before (i.e. $V_{t-12}$).

One cannot simply test the correlation by taking the original time series, but one must make them stationary and then test the cross-correlation. However, no clear result could be discerned by investigating the cross-correlations and so they are excluded. Thus, the variables’ p-value is investigated using a multi-linear regression, which is solved by using OLS. Furthermore, only the rela-
tion between the volumes and different explanatory variables is sought after. Hence, because the scale and units are different, all of the data is standardized. The p-values are presented below.

**Table 6.4: Result of p-values of different variables for the different deposit types**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Private transaction accounts</th>
<th>Savings accounts</th>
<th>Corporate savings accounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.001</td>
<td>0.082</td>
<td>0.117</td>
</tr>
<tr>
<td>Market rate</td>
<td>0.004</td>
<td>0.458</td>
<td>0.124</td>
</tr>
<tr>
<td>Volume(_{t-1})</td>
<td>0.001</td>
<td><strong>0.000</strong></td>
<td><strong>0.000</strong></td>
</tr>
<tr>
<td>Public salaries</td>
<td>0.853</td>
<td>0.192</td>
<td>0.209</td>
</tr>
<tr>
<td>Private salaries</td>
<td>0.863</td>
<td><strong>0.028</strong></td>
<td>0.681</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.068</td>
<td><strong>0.030</strong></td>
<td><strong>0.027</strong></td>
</tr>
<tr>
<td>Volume(_{t-12})</td>
<td><strong>0.000</strong></td>
<td>0.094</td>
<td>0.185</td>
</tr>
</tbody>
</table>

Here, the variables that are significant on a level of 95% or above are shown in bold. As expected, the lagged volumes are very significant. Logically, many individuals have a standing transaction from their transaction account to their savings’ account each month (around the 25:th). That would explain why private salaries are so significant for the savings accounts but not for the transaction accounts. Additionally, the market rate seems to be significant for the transaction accounts, which may be due to people moving money from their transaction accounts to the market instead of to their savings accounts. Although, intuitively clients would instead move money from their savings account to other risk-free assets in the market. However, the market rate (STIBOR 1M) has been negative for quite some time which may skew the variables’ importance. Lastly, the unemployment rate is also significant to some degree for all deposit types, which is expected. Still, another problem with this analysis is that the lagged volume may suppress the explanatory power of the remaining variables (as discussed in Section 3.4.2). Hence, one needs to be mindful of this when choosing the variables for the different models.

With the above in mind, two models for each deposit type are tested.
Table 6.5: Visualization of the variable selection for each model

<table>
<thead>
<tr>
<th>Model</th>
<th>Private salaries</th>
<th>Public salaries</th>
<th>Market rate</th>
<th>Unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>2.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Here, 1.1, 2.1 and 3.1 are SARIMA models and 1.2, 2.2 and 3.2 are SARI-MAX models.

All variables mentioned can be modeled as continuous time-series. Still, there are other factors and odd events that have a large impact on the volumes. Such as the previously mentioned change of mortgage rates at the bank. These types of events are sporadic and are often times not recurring on the same days. However, in the case of changing mortgage rates, that is something the bank knows some time in advance and this can be used in the modeling of the volumes.

6.3.2 Forecasting of exogenous variables

In order to get a legitimate prediction of the SARIMAX models, the test set consists of the forecasted selected variables. The market rate is modeled with Vasicek’s model and the result is presented in Section 6.1. Thus, it remains to model the unemployment rate and private salaries.

![Figure 6.5: Average salaries for private sector (left) and unemployment (right)](image)

There is a strong seasonality for the unemployment rate and for the private salaries there is a strong positive linear trend and perhaps some seasonality.
Hence, a SARIMA model seems to be well in place. The same steps and procedures as explained in Section 3.3.1 are used. Since the focus of the thesis is not to model macro-economic variables, most of the results of the steps in the procedure are omitted (so are the error measures). The models with the lowest AIC are chosen.

(a) SARIMA(1,1,1)(1,1,0)_{12}  
(b) SARIMA(1,1,1)(1,1,2)_{12}  

Figure 6.6: Fitted private salaries (left) and predicted private salaries (right)

(a) SARIMA(1,1,1)(1,1,2)_{12}  
(b) SARIMA(1,1,1)(1,1,2)_{12}  

Figure 6.7: Fitted unemployment rate (left) and predicted unemployment rate (right)

Now, a check of the residuals is done to see if a satisfactory model has been achieved.
The results are satisfactory, since the residuals are within the confidence interval without a specific pattern. Thus, one can conclude that the residuals are seemingly white noise.

## 6.3.3 ARIMA

In this section results for different SARIMA/SARIMAX volume models are presented. Focus has been on private transaction accounts, savings accounts and corporate savings accounts. For each deposit type the best model based on the lowest AIC-score and the error measures mentioned in Section 3.3.1 have been obtained and tested for the fit and forecast. Furthermore, all volumes have been transformed by taking the natural logarithm in order to minimize scale differences and avoid heteroscedasticity. Finally, the result is also scaled due to secrecy.

First off, the private transaction accounts’ aggregated volume is analyzed. The first step is to decompose the time series in order to inspect trends and seasonalties, see Figure 6.9

**Figure 6.8: ACF plot of residuals: residuals for private salaries (left) and unemployment rate (right)**
Next, PACF and ACF plots are investigated to find potential seasonality and trends in the time-series, see Figure 6.10.

Now, the original time series is differentiated to make it stationary. An augmented Dickey-Fuller test is also executed in addition to the plot of the differentiated signal, where the p-value is computed to $5.6 \cdot 10^{-22}$.
From the PACF and ACF plot respectively one can infer that an AR term of two and an MA term of one seems reasonable. Moreover, we have seemingly a significant lag at lag 12 in both the ACF and PACF plot. Hence, a seasonal AR and MA term of one is chosen. To further check for seasonality and how the volume correlates with its own lagged states, random forest regression is used.
Now, the parameters for the SARIMA model can be chosen. The result is a $(2, 1, 1)(1, 0, 1)_12$ SARIMA, which was further emphasized when a grid search was employed. The grid search tests all possible combinations between zero and three for each parameter and the corresponding AIC-score. All of the variables are significant on a 95%-level or higher, except for the MA term which is significant at 93%. However, the fit and prediction was much better with this additional MA term. Furthermore, the AIC was also lower with this term included. Since the model does not have that many variables there is low risk of overfitting.

The same procedure as above is used for all of the three different deposit accounts. However, these first few steps are omitted for the two other deposit accounts and only the end results are shown.
Figure 6.14: a) and b) is the fit and prediction respectively of the SARIMA models for the private accounts. c) and d) is the fit and prediction respectively of the SARIMAX models for the private accounts.

Diagnostics of the time series model can now be performed.

Figure 6.15: Diagnostic plots of residuals
From top left to bottom right: a plot of the residuals, a histogram of the residuals compared to a normal distribution, a normal Q-Q plot and finally a correlogram of the residuals. It is concluded that the residuals seem to originate from a normal distribution and are resembling white noise "enough" to be an adequate result.

**Figure 6.16:** (a) and (b) is the fit and prediction respectively of the SARIMA models for the savings accounts. (c) and (d) is the fit and prediction respectively of the SARIMAX models for the savings accounts.
Figure 6.17: a) and b) is the fit and prediction respectively of the SARIMA models for the corporate savings accounts. c) and d) is the fit and prediction respectively of the SARIMAX models for the corporate savings accounts.

Tables for tested SARIMAX models with different accuracy measures are presented in Tables 6.6, 6.7 and 6.9.

**Table 6.6: Results for the private transaction accounts**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>RMSE fitted</th>
<th>RMSE predicted</th>
<th>MAPE fitted (%)</th>
<th>MAPE predicted (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1.1</td>
<td>-270.29</td>
<td>0.0411</td>
<td>0.0426</td>
<td>3.0165</td>
<td>3.7964</td>
</tr>
<tr>
<td>Model 1.2</td>
<td>-270.25</td>
<td>0.0416</td>
<td>0.0504</td>
<td>2.8860</td>
<td>4.5638</td>
</tr>
</tbody>
</table>

**Table 6.7: Results for the savings accounts**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>RMSE fitted</th>
<th>RMSE predicted</th>
<th>MAPE fitted (%)</th>
<th>MAPE predicted (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2.1</td>
<td>-555.63</td>
<td>0.0086</td>
<td>0.0113</td>
<td>0.5938</td>
<td>0.9103</td>
</tr>
<tr>
<td>Model 2.2</td>
<td>-552.31</td>
<td>0.0087</td>
<td>0.0162</td>
<td>0.5968</td>
<td>1.3535</td>
</tr>
</tbody>
</table>

**Table 6.8: Results for the corporate savings accounts**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>RMSE fitted</th>
<th>RMSE predicted</th>
<th>MAPE fitted (%)</th>
<th>MAPE predicted (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 3.1</td>
<td>-365.25</td>
<td>0.0265</td>
<td>0.0287</td>
<td>2.1279</td>
<td>2.1114</td>
</tr>
<tr>
<td>Model 3.2</td>
<td>-364.23</td>
<td>0.0264</td>
<td>0.0290</td>
<td>2.1345</td>
<td>2.2581</td>
</tr>
</tbody>
</table>
To check if the solutions are satisfactory, the ACF plot of the residuals are inspected both for the SARIMA and SARIMAX.

**Figure 6.18:** ACF plots of the residuals for the SARIMA models

(a) ACF plot of the residuals for private transaction accounts

(b) ACF plot of the residuals for savings accounts

(c) ACF plot of the residuals for corporate savings accounts
6.3.4 Modified generalized additive model

In this section results of the MGA models are presented that have been applied to the monthly data set, its fits and predictions are presented in the figures below.

(a) ACF plot of the residuals for private transaction accounts

(b) ACF plot of the residuals for savings accounts

c) ACF plot of the residuals for corporate savings accounts

Figure 6.19: ACF plots of the residuals for the SARIMAX models
Figure 6.20: Fit and prediction of the private transaction accounts

Figure 6.21: Fit and prediction of the savings accounts
Figure 6.22: Fit and prediction of the corporate savings accounts

Result of the error analysis are summarized in Table 6.9

Table 6.9: Accuracy measures of MGA model on monthly data

<table>
<thead>
<tr>
<th>Type of Account</th>
<th>RMSE fitted</th>
<th>RMSE predicted</th>
<th>MAPE fitted (%)</th>
<th>MAPE predicted (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private transaction accounts</td>
<td>0.0290</td>
<td>0.0564</td>
<td>2.1255</td>
<td>5.3225</td>
</tr>
<tr>
<td>Savings accounts</td>
<td>0.0079</td>
<td>0.1211</td>
<td>0.4947</td>
<td>11.8366</td>
</tr>
<tr>
<td>Corporate savings accounts</td>
<td>0.0131</td>
<td>0.0986</td>
<td>0.9846</td>
<td>9.4141</td>
</tr>
</tbody>
</table>
**MGA model on aggregated daily data**

The MGA model has been applied on the aggregated daily data set. Results below represent the scaled down analysis where accuracy measures, such as MAPE and RMSE, are taken into consideration to track the performance of the model. The first "vanilla" model, without any adjustments, is presented in Figure 6.23.

![Figure 6.23: Fit and prediction of the first vanilla model](image)

In Figure 6.24 the decomposition of the time series, i.e. the trend, weakly, yearly and monthly seasonalties, are presented. As a reminder, seasonal components are modeled by the Fourier series.
Moreover, a cross validation procedure with a rolling window is used in order to evaluate the MAPE, see Figure 6.25 and the RMSE, see Figure 6.26. A training test of 730 days and a prediction horizon of 365 days into the future is used. Grey dots show the absolute percentage error for each prediction. The blue line shows the MAPE where the mean is taken over a rolling window of the dots. One can see that the errors in the one month prediction are on average around 8.31% and increase up to 9.5% in the one year forecast.
Furthermore, instead of modeling the monthly seasonality by using the Fourier series in order to account for the monthly changes, dummy variables are introduced in the event component of the model. The first dummy variable is introduced for deposits (majority of salaries are around the 25th day for each month of the year) and another dummy variable which will represent the majority of the withdrawals that occurs (around the 13th day of every months year). Moreover, flexibility of the yearly seasonality has been increased meaning faster changing cycles. The improved model is presented in the Figure 6.27. Residual diagnostics yielded an improvement of the MAPE, see Figure 6.28 and the RMSE, see Figure 6.29. With the improved model one can see that on average the errors are around 3.95% in the one month prediction and increase up to 4.98% in the one year forecast.
Figure 6.27: Fit and prediction of the second model (improved model) with dummy variables and seasonal adjustment

Figure 6.28: Diagnostic plots of residuals, MAPE

Figure 6.29: Diagnostic plots of residuals, RMSE
6.4 Risk management

6.4.1 Static replicating portfolio

To get a better understanding of the weight’s impact on the portfolio’s behaviour the efficient frontier is plotted. 50 000 different portfolio weight sets are randomly generated for each of the different deposit types.

![Efficient frontier of 50 000 randomly generated weight sets](image)

**Figure 6.30:** Efficient frontier of 50 000 randomly generated weight sets

Initially the idea was to test four different types of portfolios. Two where the Sharpe ratio is maximized, whereas one has a cap on its duration, and two where the standard deviation (or volatility) is minimized, where also one of them has a cap on its duration.
Table 6.10: Weight distribution for different portfolios

<table>
<thead>
<tr>
<th>Portfolio type</th>
<th>Deposit type</th>
<th>STIBOR 1M</th>
<th>STIBOR 2M</th>
<th>STIBOR 3M</th>
<th>STIBOR 6M</th>
<th>GVB 2Y</th>
<th>GVB 5Y</th>
<th>GVB 7Y</th>
<th>GVB 10Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capped duration SR</td>
<td>Private transaction accounts</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No duration SR</td>
<td>Private transaction accounts</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Capped duration Vol</td>
<td>Private transaction accounts</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>52.63%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>47.37%</td>
</tr>
<tr>
<td>No duration Vol</td>
<td>Private transaction accounts</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
</tbody>
</table>

Capped duration SR is the portfolio with a duration constraint that maximizes the Sharpe ratio. No duration SR is without a duration constraint. Capped duration Vol is the portfolio that minimizes the standard deviation with a duration constraint and finally No duration Vol is without the duration constraint. As can be noted in Table 6.10 above, the results for the minimization of the volatility portfolios are the same. Hence, the constraint on these particular portfolios is deemed unnecessary and thus discarded later on. Thus, the reader shall not be confused and wonder if these results are missing.

Table 6.11: Different measures of the portfolios

<table>
<thead>
<tr>
<th>Portfolio type</th>
<th>Deposit type</th>
<th>Return (%)</th>
<th>Standard deviation</th>
<th>Sharpe ratio</th>
<th>Duration (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capped duration SR</td>
<td>Private transaction accounts</td>
<td>1.10</td>
<td>0.00913</td>
<td>1.1809</td>
<td>5</td>
</tr>
<tr>
<td>No duration SR</td>
<td>Private transaction accounts</td>
<td>1.50</td>
<td>0.00916</td>
<td>1.6366</td>
<td>10</td>
</tr>
<tr>
<td>Capped duration Vol</td>
<td>Private transaction accounts</td>
<td>0.92</td>
<td>0.00898</td>
<td>1.0239</td>
<td>5</td>
</tr>
<tr>
<td>No duration Vol</td>
<td>Private transaction accounts</td>
<td>0.92</td>
<td>0.00898</td>
<td>1.0239</td>
<td>5</td>
</tr>
<tr>
<td>Capped duration SR</td>
<td>Savings accounts</td>
<td>1.60</td>
<td>0.00560</td>
<td>1.1321</td>
<td>4.5</td>
</tr>
<tr>
<td>No duration SR</td>
<td>Savings accounts</td>
<td>1.50</td>
<td>0.00779</td>
<td>1.3737</td>
<td>10</td>
</tr>
<tr>
<td>Capped duration Vol</td>
<td>Savings accounts</td>
<td>0.07</td>
<td>0.00529</td>
<td>0.4220</td>
<td>1.22</td>
</tr>
<tr>
<td>No duration Vol</td>
<td>Savings accounts</td>
<td>0.07</td>
<td>0.00529</td>
<td>0.4220</td>
<td>1.22</td>
</tr>
<tr>
<td>Capped duration SR</td>
<td>Corporate savings accounts</td>
<td>1.02</td>
<td>0.00261</td>
<td>1.0683</td>
<td>4</td>
</tr>
<tr>
<td>No duration SR</td>
<td>Corporate savings accounts</td>
<td>1.50</td>
<td>0.00791</td>
<td>1.3585</td>
<td>10</td>
</tr>
<tr>
<td>Capped duration Vol</td>
<td>Corporate savings accounts</td>
<td>0.06</td>
<td>0.00533</td>
<td>0.4124</td>
<td>1.11</td>
</tr>
<tr>
<td>No duration Vol</td>
<td>Corporate savings accounts</td>
<td>0.06</td>
<td>0.00533</td>
<td>0.4124</td>
<td>1.11</td>
</tr>
</tbody>
</table>

In Table 6.11 above the return is the expected annual return of the portfolio (and not the net gain). Furthermore, these are the optimal portfolios if they were to be set up “today”, calibrated using the available historical data since 2010. It is evident that the portfolios that maximize the Sharpe ratio have a larger margin (return) but at the same time a greater standard deviation, which reflect a more unstable portfolio. Lastly, the duration is not constant at the beginning of the period which indicates a lot of fluctuations in the deposit’s volume. The higher the duration, the more sensitive the portfolio (or the deposit account it replicates) is to changes in the interest rate.
Figure 6.31: Return of different portfolios

The portfolio mimics the behaviour of the deposit rate to some extent but the fit is very poor. This is likely due to the deposit rates being zero for such a long period and the lack of available assets.
The portfolios seem to behave in a similar manner for all of the different deposit types. Specifically the portfolios for corporate savings accounts and savings accounts are almost identical except for some minor differences in, among others, the upward parallel shift of the duration curve for the portfolio with capped duration. This is obviously due to the different duration constraints. As to why these particular portfolios behave in the same fashion is due to their deposit rates being the same for the sample period (except for early 2010).

It should be noted that these portfolios are trying to replicate the core part of the deposits. The non-core part is invested in risk free assets with overnight/weekly maturities. Additionally, one should investigate how the non-maturing deposits should be divided into stable and non-stable parts to then further divide the stable part into a core part and a non-core part. Here, two different methods are presented.

The first one is:

\[ V_C \leq \bar{V} - 2\text{std}(V) \]  \hspace{1cm} (6.2)
where \( V_C \) is the core part of the volume, \( \bar{V} \) is the average volume for the sample period, \( \text{std}(V) \) is the empirical standard deviation of the volume and \( V \) is a vector of the observed volumes during the sample period. Since it is not allowed to show this level due to secrecy it will be quoted as the percentage of the mean of the sample period. This is a very conservative method and thus the levels will be quite low.

When using the first approach, which is used by a few European banks, it is inferred that the core volume is about 36% of the mean for private transaction accounts, 38% for savings accounts and 33% for corporate savings accounts during the sample period. The second method is defined as

\[
V_c \leq \frac{\text{min}(V)}{\bar{V}} \tag{6.3}
\]

where \( \text{min}(V) \) is the minimal observed volume during the sample period and \( \bar{V} \) is the average volume. The results are that 65% of the private transaction accounts’ volumes are core, 68% for savings accounts and 57% for corporate savings accounts. The results are not surprising since companies are intuitively more prone to move larger percentages of their money than individuals.

### 6.4.2 Liquidity risk

For each of the different deposit types, Monte Carlo is used to simulate 10 000 volume paths ten years into the future based on a stochastic factor approach. The methodology is explained in Section 5.2. For all of the three deposit types, the 95%- and 99%-quantiles are computed. It should be noted that each of the paths are simulated from "today’s" volume and onwards, meaning that the first point visualized is one month from "today"

100 paths of the volumes are visualized down below
(a) 100 paths of the private transaction accounts’ volumes

(b) 100 paths of the savings accounts’ volumes

(c) 100 paths of the corporate savings accounts’ volumes

**Figure 6.33: Liquidity risk quantiles**
It is evident that there is a sharp decline at the start of the period for the transaction accounts, but that it evens out as time progresses. Furthermore, corporate savings accounts has an almost linear decline throughout the period but seems to start to flatten out at the end. The curves can be viewed as a rough estimation of the core part, since these are the volumes that will theoretically stay in the bank in a worst case scenario with 95% and 99% confidence respectively. This also seem to match the historically computed core parts somewhat.
Chapter 7

Discussion

First off, the analysis, reasoning and conclusions throughout the report have been made without the access to the daily data set, since this set was provided much later in the project. However, this data set is included in the report because of the curiosity of the general additive model’s performance and potential.

Overall the topic of modeling non-maturing deposits is extensive and includes many important areas to have in mind. As such, this thesis has only touched on the majority of these subjects to evaluate different relevant techniques.

There have been several difficulties that need to be addressed before an analysis of the results can be deducted. First off, the market rates are (and have been) negative in Sweden for quite some time. Consequently, banks set their deposit rates to zero for the accounts that have been investigated in this thesis. Hence, this results in models that are quite (or sometimes completely) independent of the deposit rates. Thus, one must be very cautious and mindful when evaluating these models since an increase of the market rate would probably cause some of the models to be insufficient in their current state. Additionally, the negative market rates and zeroed deposit rates greatly restrict how one can interpret the behaviour of the deposits. As is shown in Section 6.4.1, the "replicated" deposits are not very well replicated by the portfolios. In the case when the Sharpe ratio is maximized it is obvious that it will not fit the (zeroed) deposit rate in a satisfactory manner. That is because the problem can basically be rewritten as "maximize the return of the portfolio divided by its standard deviation", meaning there is no component from the deposit account involved. However, it is worth noting that, in theory, this approach would work
well as a pointer or indication when the market switches to something more familiar. It is also a method that is employed by many large European banks.

Secondly, a more technical note, is which programming language to use. In this report Python was used until it was revealed that some of the packages do not perform some of the computations correctly, for instance the computation of the auto-correlation of the residuals and the parameters of the SARIMA model. Hence, the volume modeling and liquidity risk modeling had to be remade in R.

Finally, the data in itself has proven to be a major restriction. A lot of information is lost in the days that are in-between the monthly observations. Accordingly, a pattern or a perceived seasonality in the monthly data might be something completely different in the daily data. Another important note is that the cleaning and segmentation of the data is much more effective for daily data, where for instance specific interesting outliers can be located and treated. Lastly, daily data would provide more data points which in turn would allow for better calibrated models. It would also make it possible to test more advanced machine learning algorithms.

**Volume modeling**

Regarding the volume modeling for the monthly data, the best model obtained for all of the different deposit types is a SARIMA model. Thus, it is concluded that the exogenous variables do not have a significant impact on the current volume. For this reason, one can perhaps conclude that these economic variables are embedded in the lagged volume component. Furthermore, the general additive model is, in theory, much better suited for daily data. That is because it has a lot of degrees of freedom. Furthermore, it is easy to incorporate re-occurring holidays and events that are on different days each year, which have an effect on the volumes.

This theory was also tested when the other (daily) data set was obtained. The aggregated daily data exhibits a strong seasonal pattern, which was utilized in the model by using dummy variables. It was concluded that the addition of these variables significantly lowered the error scores measured by RMSE and MAPE.

The SARIMA(X) models provides a very good historical fit, but also a good
prediction for the out of sample period. However, the confidence interval is quite large and is rapidly expanding as time progresses, which is natural. Furthermore, the model in itself is quite transparent. However, as was the case in this report, the added exogenous variables often needs to be modeled as well. That means that there will be an error in the modeled exogenous variables, which is then causing added uncertainty in the final volume model. Finally, the extensive time series analysis and the knowledge required to create an adequate model makes the methods a bit more complex. Still, it is preferred over other models, such as most machine learning algorithms, where there is basically a black box that outputs values.

**Static replicating portfolio**

The main issue is that the portfolio does not really replicate, for instance, the private transaction accounts account, since the deposit rate has been zero throughout the observation period. Hence, it can instead be seen as an optimal investment strategy in the current market with respect to the maximization of the risk-adjusted return or minimization of the variance. Some of the portfolios invest in bonds with negative interest rate, which can be seen as a cost of carry. However, in reality the bank would seek other investment opportunities, such as foreign bonds and exchange derivatives.

Moreover, one should also note that the part that is replicated is the core part of the volumes, meaning that the rest of the volume is invested in much shorter maturities. Therefore, the total estimated duration of the deposits will be significantly lower than the duration computed for the portfolio. Another note on the duration is that when the portfolio with a duration constraint was compared to the portfolio without a constraint (in the case of the maximization of the Sharpe ratio) it is concluded that the maximum allowed duration is almost always the result as time progresses. Hence, the portfolio is rather sensitive to interest rate changes due to the relatively high duration.

Furthermore, another reason as to why replicating portfolios in general will always have a drawback is that there does not exist risk-free assets for every maturity at each time point. Hence, the portfolio will most likely never trade at par. Additionally, the market price of the bonds can change over time which is not the case for the deposit accounts (they always trade at par). Thus, a mismatch will most likely occur.
**Liquidity risk**

The core part for the transaction account is quite large in comparison to what was expected. That is probably because the majority of the transactions happens in-between the observation points. For example salaries are deposited the 25:th and bills and other payments are payed around this date. Consequently, the account will be "stabilized" when the volume is observed. As for corporate savings accounts, the peaks and troughs are greater meaning that greater quantities, percentage-wise, are deposited and withdrawn. Hence, the core part is substantially lower. Finally, the core part of the savings account is also very considerable. That might be the result of the strong stable up-going trend, which is also emphasized by the volume path simulation.

### 7.1 Evolvement and future work

First of all, the modeling of the market rate can be done in a more realistic way than using the Vasicek’s short rate model. As it is, it might not be very realistic to assume that the market rate will continue being negative or having a downward trend for the next 10 years. However, an upward change of the trend of the market rate will only occur if Riksbanken raises the key rate, which is quite difficult to predict in a model. Thus, expertise in addition to the model is needed. Furthermore, Vasicek’s model is by nature not the most accurate model because of the few (constant) parameters. Thus, another short-rate model should preferably be used (that allow negative rates). Secondly, modeling using daily data for each customer might be a better approach since a lot of information is lost in-between months. However, it requires a lot of computational power.

The greatest concern of time series prediction is the future uncertainty in the trend. It is impossible for the model to know exactly when the trend is going to change, since the models often "learn" from the pattern of the historical data. Hence, the future is highly dependent on the history, which is not always the case. Thus, expertise knowledge is always required. The general additive model would be able to concatenate modeling and expertise in a smooth manner, if the expert understands or has a priori information about the volumes, which then can easily be incorporated and calibrated in the model.

Next, the replicating portfolio can be further developed by using moving averages or a dynamic approach. Yet, with the current market this would not
improve the end result much and thus it would be unnecessary to apply these changes now. Moreover, these modifications would increase the complexity of the model.

Finally, the liquidity risk can also be improved by incorporating potential bank runs in the model. This is a topic that is becoming more and more relevant due to Sweden being a leading country when it comes to fintech. Consequently, individuals have more and more options to seamlessly transfer money to more beneficial deposit accounts (with respect to the deposit rate). Furthermore, stressed scenarios by shocking the models with an up and down shift of the yield curve are also crucial to implement in the future.

As a last note, this thesis has not touched on the subjects of valuation and barely anything on interest rate risk. The traditional way of valuing deposits is not sufficient in the current market, as noted by Kördel (2017), who analyzed modeling of deposit rates and valuation of NMDs.

Regarding the interest rate risk, a parallel shift downwards would intuitively not change the models and the results much, since they are essentially independent of the market and deposit rate as it is. However, an upward shift would make the deposit rates above zero (depending on the magnitude of the shift), which would then make the deposit rate an important variable to include in the models. Nonetheless, since the models are calibrated on historical data where the deposit rate has been zero throughout the period, the deposit rate variable is not included. Thus, the models would, as previously mentioned, give very faulty predictions if a sufficient up-shift occurs. Furthermore, an up-shift would also impact the bonds in the replicating portfolio. As such, the optimal assets to include in the portfolio would likely change as well. Finally, one should note that this potential shift would be something that the bank would know in advance and the "risk" could be somewhat mitigated.
Chapter 8

Conclusions

Modeling of non-maturing deposits is a very broad but important area to banks and other financial institutions. In this thesis an overview of the different topics has been done where a focus has been on the volume modeling. It is concluded that most models are not, in their current form, suitable since the deposit rate is zero and thus they do not truly reflect the deposit accounts. However, the volume models and then consequently the liquidity risk modeling, all give satisfactory results. Furthermore, the SARIMA model without exogenous variables has the best out of sample performance across all the different deposit types.

Additionally, an extracurricular analysis has been partly performed on daily data due to a curiosity of the MGA model’s anticipated performance. Since the daily data has a somewhat predictable pattern in terms of constantly repeating events (such as salaries), one can model that behaviour by using dummy variables. Another reason is that the general model is fairly easy to implement and can incorporate expertise knowledge in an effortless manner. This approach yielded good results.
Bibliography


