MS-E2177 Seminar on Case Studies in Operations Research
Final Report
S-Bank: Risk Characteristics of Non-maturity Retail Deposits

Joonas Rasa, Project Manager
Iivo Savolainen
Kevin Nguyen
Lauri Turtiainen
Toni Huuhka

May 24, 2022

Contents

1 Introduction

1.1 Background and Motivation

Our client is S-Bank, the first and only "supermarket bank" in Finland with 3,1 million customers. It is the most rapidly growing bank with a reported profit of 24,8 million euros in 2021. S-Bank is widely known as the secondary bank among Finnish consumers.

In this project, the focus is on the risk characteristics of the transactional retail deposits used on a day-to-day basis by regular people. These are called non-maturity retail deposits for buying everyday necessities and receiving salaries. Not only can the risk factor be economical for these deposits but also political and societal. Also, the reputation and credibility of the bank are important for the customers to keep their money in the deposits.

Non-maturity deposits (NMDs) are deposits for which a contractual maturity has not been stated. The depositors have the possibility to deposit or withdraw funds with no penalty, so the balance of these types of funds might suddenly increase or decrease throughout the day. The depositors value non-maturity deposits through two factors, received value and perceived value. This means that higher interest rates paid to the depositors and more significant barriers to withdraw the deposits create longer-term deposits. When the market interest rates increase, the balance of non-maturity deposits tends to fall due to the depositors finding more attractive investment opportunities elsewhere, and thus withdrawing at least a part of their funds. Alternatively, when the interest rates are low, non-maturity deposits are a more profitable option to consider. Cipu and Udriste (2009) These changes may have a significant impact on banks, since in 2012 the proportion of non-maturity deposits out of bank's funds was 58% on average. Thus, non-maturity deposits are the primary funding source for an average bank Blöchlinger (2015).

Due to the aforementioned reason, non-maturity deposits have been predicted based on the interest rate levels. These models have been more helpful in an environment where the interest rates have changed through time. Moreover, in recent years these have been close to zero, and thus these models are not ideal. Interest rates are not the only reason behind account behavior, and thus these models cannot give precise predictions. For this reason, the other factors that explain the account behavior might be easier to spot when there are no significant interest rate changes in the observed period.

Having no agreed-upon maturity, the balance of non-maturity deposits may fluctuate quickly as clients withdraw and deposit money without any penalty. An intuitive example of larger than normal withdrawals could be Christmas, as it could be expected that buying Christmas presents causes some additional expenses for a significant share of the population. Other such dates when one could expect abnormally large withdrawals or deposits could be at the start of and the middle of the month, as rents and wages are typically paid then. However, not all inflows and outflows are as recurring and small as the aforementioned ones. During both the Global Financial Crisis and the European debt crisis several *bank runs* occurred as people rushed to empty their bank accounts in fear of losing all their money (Wikipedia contributors, 2022). During a bank run a large share of the bank's clients attempt to withdraw money from their accounts, which in fact strengthens the bank run further; the likelihood that the bank defaults increases as more people withdraw their money, which again gives an incentive to more clients to withdraw their money. As a result, bank runs may happen suddenly and may result in the bank running out of cash at which point they face bankruptcy.

The risk characteristics of NMDs can be observed by conducting analyses on the historical data of customers' money transactions in non-maturity deposits and find the best and fittest model to predict the customers' consumption behaviours and if different seasonal periods have an effect. One of the critical factors in making credible analyses on NMDs is to perform customer segmentation.

1.2 Objectives

This project aims to analyse the interest rate and liquidity risk characteristics of non-maturing retail deposits. The specific objectives are:

- Developing a justified and documented model based on several variables such as
  - The economic environment
  - Product type
  - Customer segment
  - Calendar effects
- Comparing S-Bank's data to the Finnish aggregate index data
- Finding out whether there is any structural change or regime shift present in the data due to the Covid-19 pandemic

Developing a justified model for non-maturity retail deposits is our main task. Historically, the NMDs have been modeled with respect to interest rates due to the opportunity cost in the investment market. In recent years, the interest rates have been mostly negative, and the interest rate-based models have not been developed for this kind of environment. S-Bank's non-maturity retail deposits do not gain a negative interest rate.

Time series-based models have also been used to predict non-maturity retail deposits. In this project, we shall compare different models to see which of these best suit the S-Bank's data and which parameters sought to be used.

Traditional customer segmentation could improve the model accuracy, but account behavior could serve our client better. We will test how the account segmentation affects the accuracy of our models.

2 Literature Review

2.1 Non-maturity Deposits

Despite scarcity of previous research focusing on non-maturity retail deposits, some frameworks have been developed for their valuation. The first models were based on the present values of expected future payments from retail deposits (O'Brien et al., 1994), (Selvaggio, 1996) after which arbitrage-free pricing models were developed to incorporate the prevailing interest-rate regime (Jarrow and Van Deventer, 1998). The dynamics of non-maturity retail deposits have been previously modelled using *autoregressive-moving-average* (ARMA) models (Hamilton, 1994) and stochastic models (Jarrow and Van Deventer, 1998). Interestingly, the aforementioned studies have focused on modelling all liabilities as an entity - pooling all types of clients into one. Although this clearly simplifies the development of a modelling framework, a lot of information is evidently lost.

The non-maturing deposits are commonly segmented as stated by Kalkbrener and Willing (2004). The authors state that it is common for banks to split the total volume of accounts into two; a stable core part and a more volatile floating part, which is the fundamental idea behind the so-called *non-maturation theory*. This is justified as the number of accounts is large, whereas the average volume per account is small in comparison. Typically, the bulk of the accounts do not face large withdrawals – this is more of an exception. In the unlikely event of a *bank run* this phenomenon would most likely not hold, from a modeling perspective this can be argued to be a fair assumption. Subsequently, the modeling would focus on the core part and the floating part with a long and short maturity, respectively.

The most common, and perhaps also the most naive, approach to modeling non-maturing deposits is to make static assumptions about the maturity of the deposit. By assigning one maturity for every deposit, the clear benefit is simple cash flow analysis. However, this can be problematic regardless of whether the assumed maturity is longer or shorter than the actual maturity of the deposit. If the assumed maturity is longer than the real maturity, the immediate risk is poor liquidity during a time of withdrawals, although the deposits can be invested. On the other hand, if the assumed maturity is shorter than the actual maturity, the deposits cannot be allocated in a productive way, although liquidity is guaranteed in the case of withdrawals.

A clear improvement to this approach is to utilize the *non-maturation theory* and to split the deposits into two parts with two different maturities. Subsequently, the maturity of the floating part can be modeled using different empirical methods such as the *outflow rate method* OeNB (2008) or using a *replicating portfolio approach*, where different fixed income instruments are combined in order to artificially create similar cash flows as the accounts. Combining different methods is also common, both for the stable part and the volatile part. However, as all the aforementioned methods are deterministic, they do not include any randomness. For this purpose, it is also possible to view the NMDs as options and apply *option-adjusted spread* (OAS) models, where stochastic interest rate term structures are modeled.

In addition to the prevailing interest rate environment as a major factor impacting NMDs, the prevailing unemployment rate is also a factor to be taken into account. As stated by Stavrén and Domin (2019), the unemployment rate is expected to have an inverted correlation with respect to NMD volumes. This is theoretically justified as the mean wealth decreases as a result of increased unemployment, subsequently leading to lower

NMD volumes. In their study, Stavrén and Domin (2019) found this relationship to be statistically significant for both savings accounts and corporate savings accounts.

2.2 Segmentation

Customer segmentation is a way to categorize customers into smaller groups to provide targeted services and assess the creditworthiness in banks. This segmentation is done traditionally via demographic or economic criteria such as age or income (Machauer and Morgner, 2001). The problem with this approach is that outliers do not behave as the averages of their segment. More recently, account behavior has been used as part of customers' creditworthiness. These categorizations that have been done for other purposes might give more appropriate approximations for interest rate-based prediction models, or new segmentation might serve this purpose even better. Account activity is also an indicator of how likely the funds are withdrawn – this is one of the features in the data that S-Bank provides us.

2.3 Models

Different methods for forecasting non-maturing liabilities have also been developed and studied in previous research. Ahmadi-Djam and Belfrage (2017) tested several time series models to determine whether they are are suited for modelling deposit volumes. The tested models included a Holt-Winters model, a Stochastic Factor model, an ARIMA model and an ARIMAX model, of which the ARIMAX model was found to be the most suitable one. Using only daily data, the model was found to forecast accurately from 3 months up to 6 months.

One of the most commonly used methods to model non-maturity deposits was developed by Jarrow and Van Deventer (1998). It is a one factor model that uses short rates for forecasting future non-maturity levels. The shortcoming of this model in a negative interest rate environment is that bank's retail customers' rates bottom boundaries are capped to zero. This weakens the precision of the model Stavrén and Domin (2019).

3 Data & Methods

3.1 Data

The data that is used in this study consists of four separate files; each one with the same layout and content, but for different types of accounts. More specifically, the four types of data correspond to four different types of clients in order of importance to the bank. Two are seen as very important: A1 and A2, and two are seen as less important: B1 and B2. Each dataset contains daily observations of the amount of money deposited to 5000 accounts ranging from January 2, 2015 to February 22, 2022. The daily sum of all 5000 accounts of the dataset A1 is presented in Figure 1, and the daily sums of the respective accounts in the other datasets are presented in Figure 17 in Appendix A.

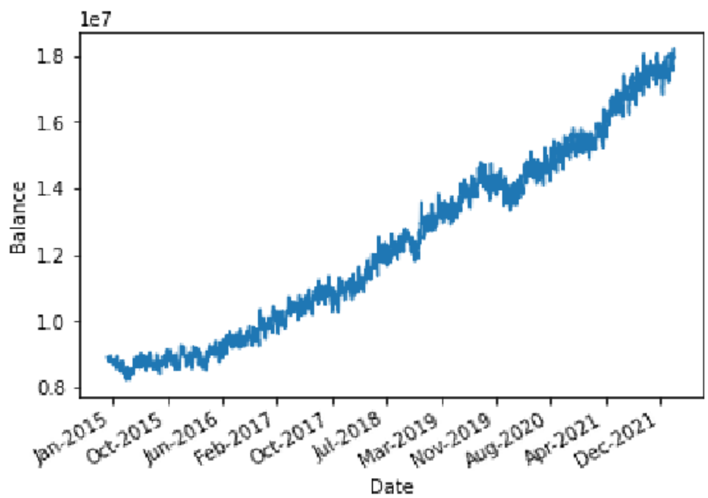


Figure 1: Daily total balance of the accounts in dataset A1.

As the data contains deposits from real people and wealth distributions are notoriously skewed (Benhabib and Bisin, 2018), we expected extreme values in all datasets. Summary statistics from the beginning and the end of the period of the A1 dataset are presented in Table 1 below:

	Start of the period	End of the period
Mean	1783.59	3589.16
Median	201.99	400.1
Std Dev	7375.3	14067.69
Minimum	−146.57	−39.76
Maximum	242585.48	432305.94
Skewness	14.41	12.28
Kurtosis	330.92	245.04

Table 1: Summary statistics of the data in the A1 dataset

As can be seen from Table 1 above, the data is extremely skewed positively and leptokurtic. This means that the data is unevenly distributed which is caused by the mean being significantly larger than the median. Furthermore, the data contains negative values due to irregular spending by some accounts. Moreover, the share of accounts that have had at some point a negative balance is high, roughly 42.3%. The summary statistics of the other datasets are presented in Tables 31, 32, and 33 in Appendix B. In Figure 2, we show the performance comparison of the four datasets, we mentioned earlier. Specifically, the plot displays the scaled version of monthly total balance of each segment. In addition, we added the corresponding data from Bank of Finland (SP) which consist data of the total deposits of all Finnish working households. The total balances are scaled with the respect to its first total balance value and thus makes the comparison valid.

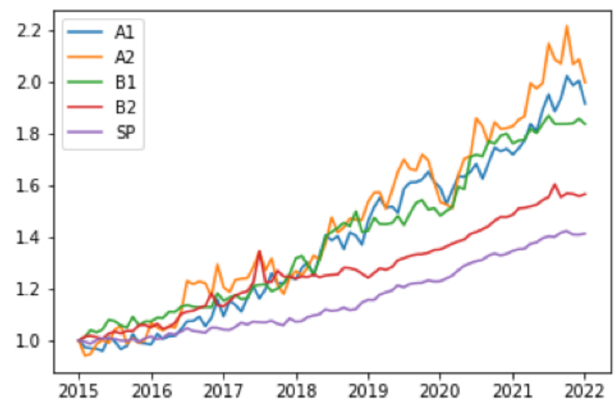


Figure 2: Performances of total balances of the datasets and Bank of Finland (SP).

Figure 2 shows the segments A1 and A2 are very important and the main NMD holding capitals comes from them in the perspective of S-Bank. The total deposit of Bank of Finland seems to underperform compared to the segments. This is simply because the total deposits of the segments of S-Bank are fraction of the sizes of the total deposits of the Bank of Finland.

3.1.1 How Covid-19 Affected the Data

As we can see in Table 2, there were significant changes in the variance of the total NMD balance within every segment. January 29th 2020 was used as the beginning date of COVID-19 in the context of this project, because the first infection in Finland was diagnosed on that day. The variance within segment A1 has increased by a factor of 1.74. This means that the uncertainty of the segment has grown to be 1.74 times larger after COVID-19 started, compared to the uncertainty level before COVID-19. For all other segments, the variance for the time after COVID-19 started is smaller than the variance before said time. This can be interpreted as the uncertainty of these segments decreasing, which seems counter-intuitive. The reason for this is unclear.

Segment	Variance pre-covid	Variance post-covid	$\frac{Var_{Post\ covid}}{Var_{Pre-covid}}$
A1	73 479 529.31	128 009 810.71	1.74
A2	2 089 594 571 865.98	1 461 989 452 614.01	0.70
B1	268 036 278 812.63	124 882 605 604.41	0.47
B2	52 956 732 963.65	21 454 011 423.96	0.41

Table 2: The variances of the daily total NMD balances between different segments, pre-covid and post-covid

3.2 Segmentation

In order to improve the accuracy of our model, we first divide our data into different data sets based on S-Banks's own segments, use a clustering algorithm to divide the data further and then fit the model for each of these segments separately. After this, we sum each of the segmentation based predictions.

We are testing four different segmentation algorithms in this project. The goal is to divide accounts into groups based on the account behaviour. Thus, we are selecting the segmentation algorithm that creates groups that differ the most from the original data set on average based on mean squared error, MSE. The difference measure:

$$\sum_{s \in S} \frac{MSE(a(d) - k(s,d)) \times S(s)}{\sum_{d \in D} a(d) \times S(d)},$$

where S is the set of segments, D is the set of days, $S(s)$ is the size of the segment s, $S(d)$ is the size of the data set, $a(d)$ is the account balance at day d and $k(s,d)$ is the average account balance of a segment s at day d.

3.2.1 Density-based Spatial Clustering of Applications with Noise

Density-based Spatial Clustering of Applications with Noise, also known as DBSCAN, is a popular data clustering method where given a set of observation points in some specific area, it segments together the points that are close to each other based on a distance with some minimum number of points in this area. Hence, why the method is called density-based. Generally it is measured with Euclidean distance. DBSCAN only requires 2 parameters. First parameter is $epsilon$ which is the maximum distance between some two points to be considered neighbors or in a same segment. Second parameter is $minPoints$ which is the minimum number of points to form a dense area. For example, for $minPoints = 5$, if there are five points or more within the $epsilon$ distance of each other, then it is called its own separate segment. We received the following difference measures when performing DBSCAN with parameters $epsilon = 0.5$ and $minPoints = 100$ on four different datasets:

Table 3

	A1	A2	B1	B2
Difference measure	2078.71	1241.76	621.03	116.99

After performing DBSCAN clustering on the existing data, we concluded that this clustering method was not suitable for day-to-day NMD balance data when the balance can stay the same for weeks when the customer does not withdraw or deposit cash in it for a while. This makes DBSCAN very non-robust for the analysis of short periods when the variation of data is not spread enough.

3.2.2 K-means clustering

K-means algorithm divides the data into a given number of clusters, k, by minimizing the means of these groups.

$$\arg\min \sum_{i=1}^{k} \sum_{x \in S_i} (x - \mu_i)^2.$$

At the first iteration the algorithm gives the data k arbitrary selected means $\mu_i,\ i \in [1,..,k]$. Each data sample is assigned to the group with nearest mean. New means are then calculated for each of these groups and each sample is again assigned to the group with nearest mean. This process is continued until the iterations do not alter the groups anymore. If the algorithm does not converge, it is stopped when given number of iteration is reached. K-means method is the most widely used centroid-based segmentation algorithm. The standardized algorithm was published by Lloyd (1982).

For data sets B1 and B2, the K-means clustering provided the best results. Thus, we are testing how it performs with our model. The difference measures we received with this method and seven clusters were:

Table 4

	A1	A2	B1	B2
Difference measure	3378.47	2705.08	1214.21	629.51

3.3 Gaussian Mixture

Similar to the K-means clustering, the Gaussian Mixture Model (GMM) divides the data into predefined number of clusters. However, Gaussian Mixture Model performs the division based on an assumption that all the data points are generated with some number of Gaussian distributions. The segments maximize the likelihood that these are generated by the same set of these distributions.

Gaussian mixture clustering provided us the best results for data set A2 based on our experiments. Thus, we are testing how it performs with our model. The difference measures of the Gaussian mixture with seven clusters were:

Table 5

	A1	A2	B1	B2
Difference measure	3713.95	2779.52	1200.6	494.16

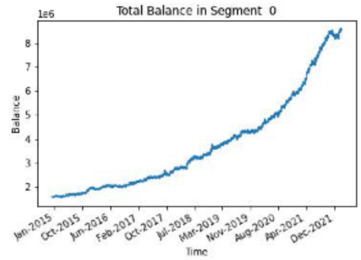
3.4 Affinity Propagation

Affinity Propagation does not take the number of clusters as an input like the previous two. The algorithm first picks an arbitrary number of samples from the data set as exemplars and then calculates how similar the samples that have not been picked are to these exemplars. Each of these remaining samples are assigned to the exemplar to maximize the total similarity. The set of exemplars and to which examplar the remaining samples are assigned is updated at each iteration until the algorithm converges.

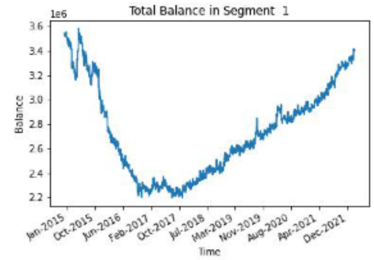
Affinity propagation had problems finding a reasonable number of clusters in our data set. When dividing the 5000 accounts from data set A1 into segments, it found 166 different segments. Computationally, this would not have been an issue with our data sample, but S-Banks 3,1 million customers could create issues. For data sets A1, B1, and B2, the algorithm did not converge. The difference measures for Affinity propagation were:

Table 6

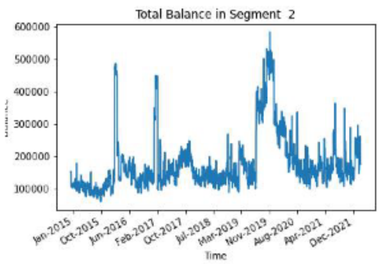
	A1	A2	B1	B2
Difference measure	4797.58	Did not converge	Did not converge	Did not converge



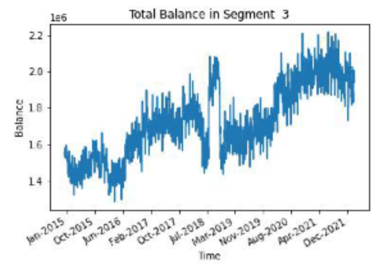
(a) Segment 0, segment size: 663



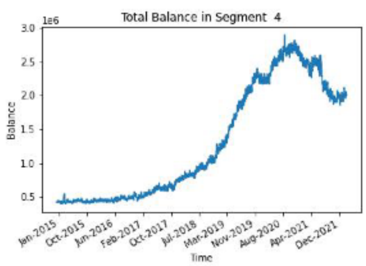
(b) Segment 1, segment size: 546



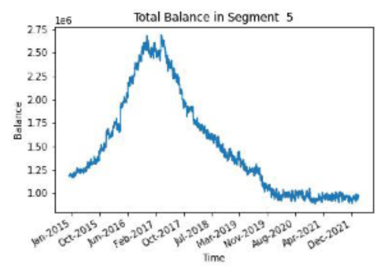
(c) Segment 2, segment size: 673



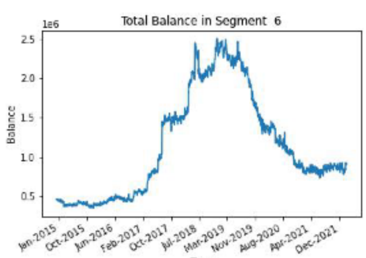
(d) Segment 3, segment size: 1235



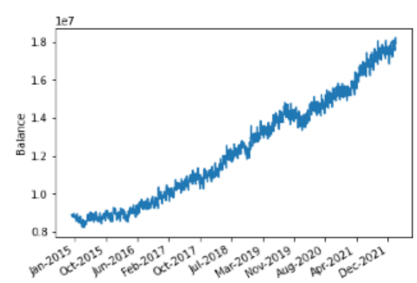
(e) Segment 4, segment size: 742



(f) Segment 4, segment size: 594



(g) Segment 6, segment size: 547



(h) Total balances in data set, size: 5000

Figure 3: Results of Gaussian Mixture segmentation on data set A1.

3.5 Forecasting Models

The Jarrow and Van Denventer model can be used to forecast non-maturity deposits based on the relationship between deposit rates, d, and interest rate levels, r. The interest rates are predicted with, for example, Vasicek-model or ARMA(1,1). The total deposits from time 0 to t can be estimated with the following formula:

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

where the coefficients $\beta_0, \beta_1, \beta_3$ are obtained by minimizing the estimation error over the training period. Since the retail customers can not have negative interest rates in their accounts, we alter this model as:

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

There are only 55 one month interest rate observations and predictions that are positive in our time data set. Those are in the beginning of year 2015. For this reason, this model shrinks down to $d_t = d_0 + \beta_0 t$. Thus, we are not expecting this model to be very accurate.

ARMA (Autoregressive Moving Average) models are used for statistical analysis of time series. ARMA models have two orders - one of the autoregressive part and one for the moving average part. ARMA models are usually denoted as ARMA(p, q), where p is the order for the autoregression and q is the order for the moving average. Mathematically, an ARMA(p,q) model is denoted as follows:

$$X_t = \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + c,$$

where X_t is the predicted quantity at the moment of time t, ϵ_t is the error in the prediction, $\phi_{1...p}$ are the autoregression coefficients, $\theta_{1...q}$ are the moving average coefficients, $\epsilon_{t-1...t-q}$ are the error terms for the past predictions and c is a constant.

ARMA models are fitted by using least squares regression to minimize the error term in order to find the correct coefficients $\phi_{1...p}$ and $\theta_{1...q}$. SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) models are quite similar to ARMA models, but they account for seasonality in the time series, as well as take into account an exogenous variable in order to describe the time series better. A SARIMAX(p,d,q)(P,D,Q,s) model is denoted as follows:

$$X_t = \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \sum_{i=1}^{r} \beta_i x_{i_t} + \sum_{i=1}^{P} \alpha_i X_{t-si} + \sum_{i=1}^{Q} \rho_i \epsilon_{t-si} + c,$$

where d is the integration order, P is the autoregressive order for the seasonal component, D is the seasonal integration order, Q is the moving average order for the seasonal component, s is the season length, x_{i_t} is the exogenous variable at time t, $\beta_{1...r}$ are the coefficients for the exogenous variable values, $\alpha_{1...P}$ are the autoregressive coefficients for the seasonal component, X_{t-si} are the past variable values offset by the season length, $\rho_{1...Q}$ are the moving average constants for the seasonal component and ϵ_{t-si} are the error terms for the past predictions offset by the season length.

The modelling was begun by first considering multiple ARMA models to fit the time series formed by the sum of account balances for each day. The fitted models were MA(1), ARMA(1,1), ARMA(1,2), ARMA(2,1) and ARMA(2,2). These models were first used to build the modelling functionality, not for actual predictions so the results provided by said models will not be discussed in depth. The MA(1) model gave the most pessimistic predictions, i.e. the lowest values, and the ARMA(2,1) model gave the most optimistic predictions. After fitting the ARMA models, an interest rate data set for years 2015-2022 was formed from multiple smaller 6 month data sets. The data set contains daily values for 1 week, 1 month, 3 month, 6 month and 12 month Euribor rates. Utilizing this data set, the ARMAX models were fitted using different Euribor rates as an exogenous variable. A separate ARMAX(1,1) model was created for each Euribor rate. Further discussion with the client had to be had to determine the most sensible interest rate to use in the modelling, and it was determined that the 3 month and 12 month Euribor rates were the most sensible. The Euribor rates were used to build the functionality of the ARMAX models since there was daily interest rate data available on them, and the results provided by the ARMAX models will not be discussed.

After fitting these simple models, the Bayesian information criterion (BIC) was calculated for every ARMA model from ARMA(1,1) to ARMA(10,10) in order to find the model that has the best balance between fitting the data and complexity. The lower the BIC, the better the model is in terms of finding this balance. The Bayesian information criterion is

$$\text{BIC} = k\ ln(n) - 2l,$$

where k is the number of parameters in the model, n is the number of samples used to fit the model and l is the maximized value of the log-likelihood function of the model.

The best ARMA model in terms of the BIC was ARMA(7,9). However, this model was deemed too complex in the modelling phase, and it was opted to use ARMA(3,3) instead, as it was simple enough and minimized the BIC when the orders were limited to be at most 3. Table 7 shows that the variations between the BIC values of different models were quite small. The table contains the quartiles of BIC values calculated for every model from ARMA(1,1) to ARMA(10,10). Thus the table is constructed based on the BIC values of 100 ARMA models. From the perspective of BIC, the model selection from the available models was quite inconsequential.

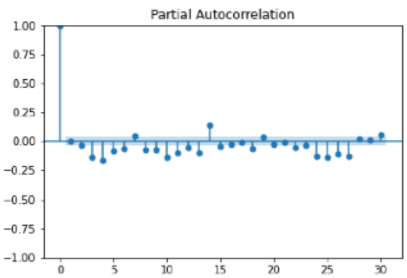
Percentile	value
0	72 830.79
25	73 094.55
50	73 195.95
75	73 248.89
100	73 619.99

Table 7: Quartiles of the BIC values, ARMA(1,1) to ARMA(10,10)

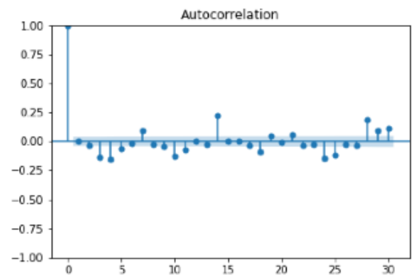
The parameters of the SARIMAX model were selected in a similar way as the parameters of the ARMA model, the main difference being that the season length, integration order and seasonal integration order had to be determined first. The season length used in this project is 14 days, and the selection was performed using intuitive knowledge of customer behaviour supported by observations based on the autocorrelation function 4b and the partial autocorrelation function 4a of the stationarised time series. The integration order d was set to 1 and seasonal integration order D was set to 0, as a single non-seasonal differencing was sufficient to make the time series stationary.

The remaining parameters for the SARIMAX model, the seasonal and non-seasonal autoregressive and moving-average orders p, q, P and Q, were determined by calculating the BIC of every possible model with the aforementioned parameters ranging from 0 to

2, resulting in 81 tested models in total. The number of tested models was limited by the testing process being computationally heavy and time-consuming. The model that yielded the lowest BIC was SARIMAX(0,1,1)(2,0,0,14), which was then selected for forecasting.



(a) Partial autocorrelation of the differenced A1 series



(b) Autocorrelation of the differenced A1 series

Linear regression was also used to model the relationship between the total NMD balance and the interest rates. The 3 month Euribor rate was first used to perform the regression, as our client deemed this to be one of the most prominent interest rates to use in the modelling, alongside the 12 month Euribor rate. Later, the regression was performed using every Euribor rate separately.

Linear regression is a linear method for modelling the relationship between one or more explanatory variables and one dependent variable. The used regression method was linear least-squares regression, which means that the sum of squared errors between the fitted and actual values is minimized. Mathematically, the regression equation is

$$y_i \approx \beta_0 + \beta_1 x,$$

where β_0 is a constant β_1 is the constant for the interest rate and x is the interest rate .

From Figure 5 we can see the regression line for the 3 month Euribor rate, and its relationship to the scattered data points. We observe a downwards trend. We also observe, that the data points are mostly above the regression line on the left side of the plot, then below the regression line approximately when the interest rate is from -0.3 % to -0.1 %, and then above the regression line again after that. From Table 8 we can see the statistics of the regression. The slope was -12 198 989.20, which indicates that the total NMD balance will decrease by the said amount when the interest rate increases by one percentage point. The y-intercept is 5 442 722.35, which implies that the total NMD balance would be that amount when the interest rate is 0 %. The $|R^2|$ value is 0.85, which means that 85 % of the variability of the total NMD balance is explained with the 3 month Euribor rate. The minus sign indicates that the relationship is negative. The p-value of 0 implies that there is no chance that there is no actual relationship between the total NMD balance and the 3 month Euribor rate, and thus the 3 month Euribor rate can be declared to be a statistically significant variable in the regression.

The regression statistics for the linear models using the other Euribor rates as an explanatory variable are in Table 8. The absolute value of the slope decreases when moving from shorter term Euribor rates to longer term Euribor rates. Alternatively, the y-intercept grows in magnitude, so the longer term the Euribor rate, the more optimistic the predicted value of the total NMD balance is when the interest rate is at 0 %. The difference between the best and worst R^2 value is four percentage points. The 12 month Euribor rate explains 88 %, and the 6 month Euribor rate 87 % of the variability in the total NMD balance, and thus it seems that the longer term Euribor rates do a slightly better job in explaining the aforementioned variability.

Even though linear regression seems to fit the data well, there is a problem. From Table 9 we can see the predicted total NMD balances for interest rates 0...0.5 %. As we can see from the table, the predicted NMD balance turns negative when the 1 month Euribor rate is somewhere between 0.3...0.4 %. The problem arises due to the used data being so monotonic. The interest rates have been low, and even negative for the whole time period the data was obtained from. From Figure 6 we can see that the variation range for each of the Euribor rates is under one percentage point. As mentioned, the data is simply too monotonous for the linear regression to have predictive power when the interest rates change into larger positive numbers. Thus, the linear regression approach is documented, but will not be pursued further.

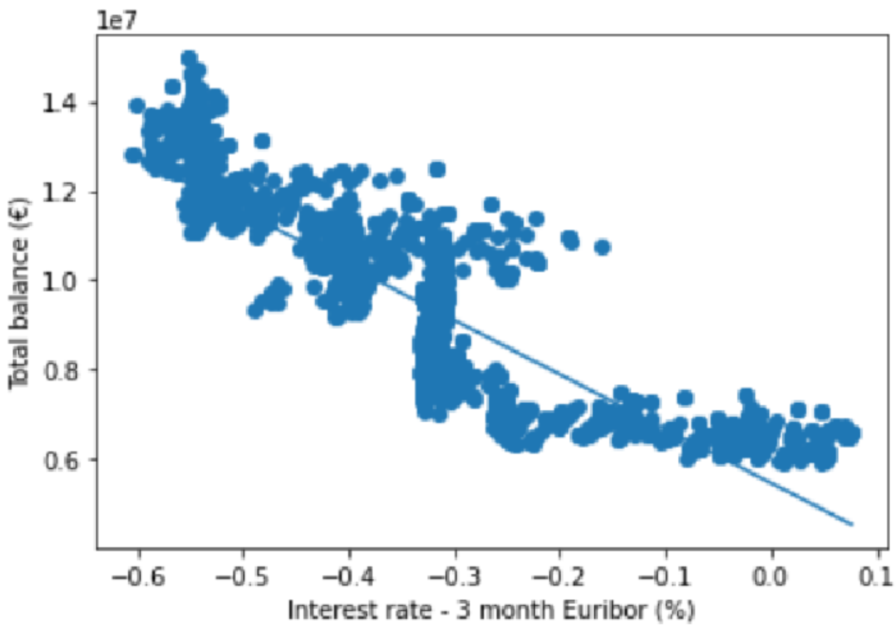


Figure 5: Linear regression using the 3 month Euribor

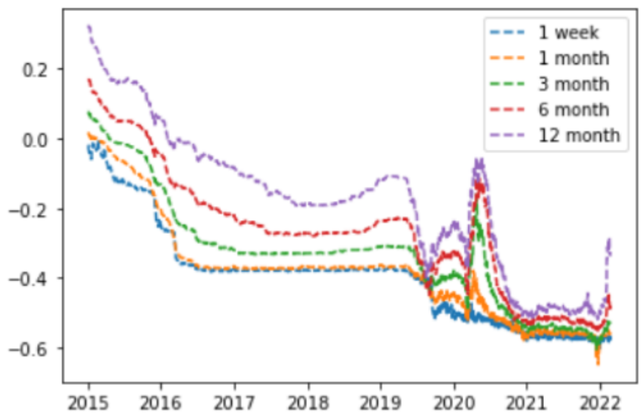


Figure 6: A visual representation of the variation of the Euribor rates.

Interest rate	Slope	y-intercept	R^2	p-value	Standard error
1 week Euribor	-13 675 273.28	4 056 118.05	-0.85	0	181 933.91
1 month Euribor	-12 825 006.28	4 629 308.00	-0.84	0	179 939.18
3 month Euribor	-12 198 989.20	5 442 722.35	-0.85	0	162 857.55
6 month Euribor	-11 089 127.57	6 481 854.00	-0.87	0	124 017.64
12 month Euribor	-9 687 983.60	7 745 787.60	-0.88	0	112 788.48

Table 8: Descriptive data of the regression models.

1 month Euribor (%)	Predicted NMD balance (€)
0	4 629 309.00
0.1	3 346 808.37
0.2	2 064 307.74
0.3	781 807.12
0.4	-500 693.51
0.5	-1 783 194.14

Table 9: Predicted total NMD balances using linear regression, 1 month Euribor

3.6 Segments Application on the Forecasting Models

In the final model, we combine the best performing segmentation algorithm and forecasting model. The forecasting model is fitted for each segment's training data separately and then it counts the result as a sum of each of these NMD predictions. The pseudo code of this process is as follows:

Algorithm 1 Segmentation Based Forecasting Model

Input: prediction length PL, number of segments k, training data D
Output: prediction P

```
P ← [ ]
use k-means to the data D and add a column to D indicating the segment
for data segment i in D do
    S ← time series of summed daily account balances in segment i
    M ← model obtained by fitting time series model to S
    P ← P + predictions the model M gives for days PL
end for
return P
```

4 Results

The results we compared are obtained by creating predictions for 90 day intervals with selected models. For each prediction, the model parameters are fitted based on training data of 365 days prior to the predictions. The results are analyzed by 30 day interval averages as well as the total 90 day interval. We focus on the Mean Absolute Percentage Error (MAPE) and the standard deviation of the error terms.

The results presented in this section include baseline results using the Jarrow and Van Deventer model as well results using an ARMAX(1,0,1) model, an ARMA(3,3) model and a SARIMAX(0,1,1)(2,0,0,14) model. Furthermore, the results of each model with the exception of the Jarrow and Van Deventer model are presented without segmentation, with K-means segmentation and with Gaussian Mixture segmentation. The number of parameters for each model was determined using the Bayesian Information Criterion.

4.1 Jarrow and Van Deventer

In this project, the Jarrow and Van Deventer model was used to benchmark our other models. We did not expect this model to perform well in test settings during a negative interest rate environment. The restriction that interest rates cannot be negative makes the model simple and linear and thus the accuracy seems to decline when predicting further. The mean absolute percentage errors for each data set are presented in Table 10 and summary statistics of the standardized residuals are presented in Table 11.

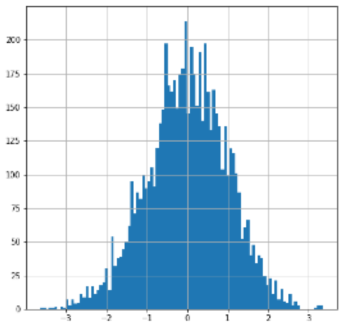
Table 10

	A1	A2	B1	B2
MAPE - total	2.304	6.134	2.066	1.652
MAPE - [0-30]	2.145	5.756	2.597	1.096
MAPE - [30-60]	2.274	6.251	1.896	1.689
MAPE - [60-90]	2.493	6.395	1.704	2.170

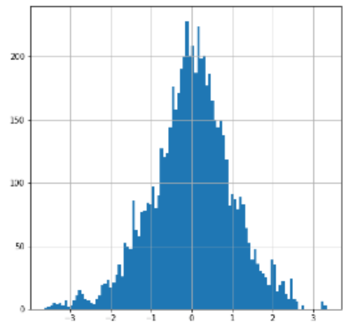
Table 11

	A1	A2	B1	B2
mean	-1.85e-16	-1.53e-15	1.49e-16	-6.20e-16
min	-3.63	-3.16	-3.63	-4.55
25 %	-6.33e-01	-5.78e-01	-5.269e-01	-3.95e-01
50 %	2.02e-02	2.87e-02	7.677e-02	6.33e-02
75 %	6.88e-01	6.23e-01	5.942e-01	5.23e-01
max	3.35	3.34	4.455	5.72

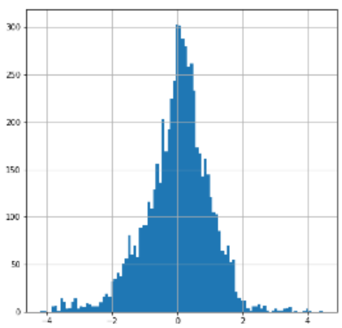
Distributions of the standardized residuals are shown in Figure 7. From the distributions we can see that the residuals are seemingly more normally distributed for datasets A1 and A2 than for B1 and B2, indicating that the Jarrow and Van Deventer model is better suited for forecasting using datasets A1 and A2.



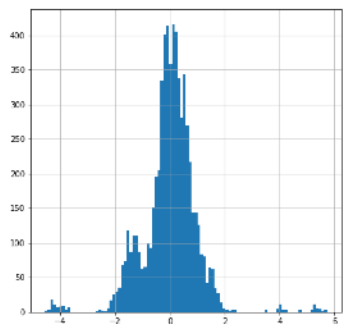
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 7: Plotted standardized residuals for all data sets using the Jarrow and Van Deventer model.

4.2 ARMAX(1,0,1) with 1-month Euribor

In this subsection the results of the ARMAX(1,0,1)-models with the 1-month Euribor rate as an exogenous variable are presented. First the results without segmentation are presented, then the results with K-means clustering are presented and lastly the results with Gaussian Mixture Segmentation are presented.

4.2.1 ARMAX(1,0,1) without Segmentation

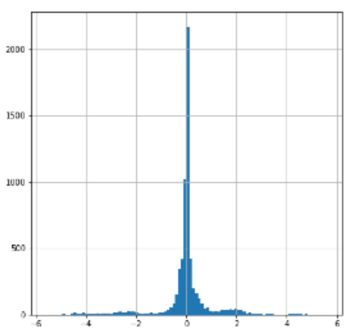
The ARMAX(1,0,1) model without segmentation is clearly the least suited for forecasting as indicated by the high MAPE shown in Table 12. Furthermore, by examining the distribution of standardized residuals in Table 13 and Figure 8 it is clear that the residuals are more dispersed than for the distributions resulting from the Jarrow and van Deveter model.

Table 12

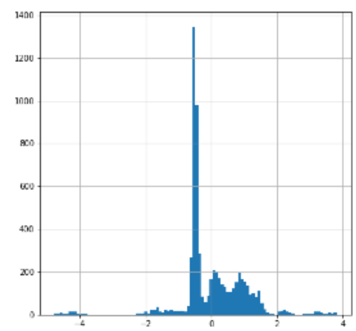
	A1	A2	B1	B2
MAPE - total	3.274	27.195	14.308	15.811
MAPE - [0-30]	3.205	27.200	13.906	15.671
MAPE - [30-60]	3.349	27.262	14.100	15.692
MAPE - [60-90]	3.269	27.124	14.919	16.071

Table 13

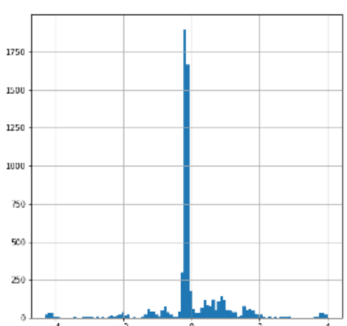
	A1	A2	B1	B2
mean	-2.21e-16	1.45e-15	7.17e-16	-2.50e-16
min	-5.63	-4.79	-4.29	-3.84
25 %	-1.03e-01	-4.94e-01	-1.87e-01	-2.33
50 %	1.63e-02	-3.69e-01	-1.47e-01	-2.17
75 %	1.18-e01	6.33e-01	5.69e-02	-1.27
max	5.63	3.82	3.99	3.39



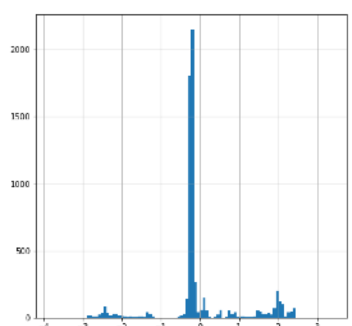
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 8: Plotted standardized residuals for all data sets for the ARMAX(1,0,1) model without segmentation.

4.2.2 ARMAX(1,0,1) with K-means Segmentation

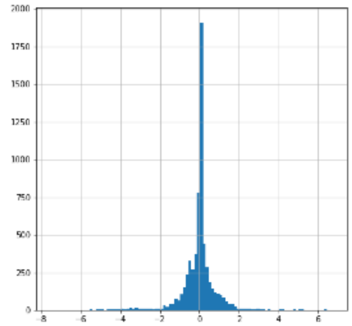
The mean average percentage errors of the ARMAX(1,0,1) model with K-means segmentation are presented in Table 14 and descriptive statistics of the standardized residuals are presented in Table 15. A visual representation of the residuals is shown in Figure 9.

Table 14

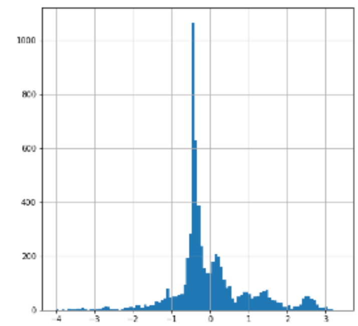
	A1	A2	B1	B2
MAPE - total	1.567	15.755	11.807	15.695
MAPE - [0-30]	1.703	15.747	11.807	15.689
MAPE - [30-60]	1.548	15.826	11.936	15.618
MAPE - [60-90]	1.449	15.693	11.676	15.776

Table 15

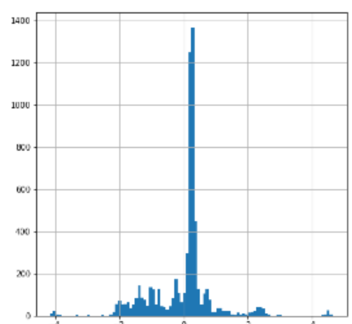
	A1	A2	B1	B2
mean	-2.71e-16	-7.04e-16	2.01e-16	6.26e-17
min	-8.73	-4.00	-4.15	-3.14
25 %	-2.36	-4.50e-01	-2.81e-01	-4.94e-01
50 %	4.07	-3.00e-01	2.27e-01	-4.67e-01
75 %	2.07	3.17e-01	3.09e-01	-6.79e-02
max	9.07	3.34	4.63	2.90



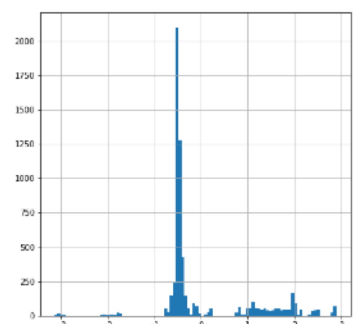
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 9: Plotted standardized residuals for different data sets for the ARMAX(1,0,1) model with K-means segmentation.

From the results presented above it is clear that the K-means segmentation improved the forecasting accuracy, however, the JVD model that was used as a baseline still shows more accurate forecasts.

4.2.3 ARMAX(1,0,1) with Gaussian Mixture Segmentation

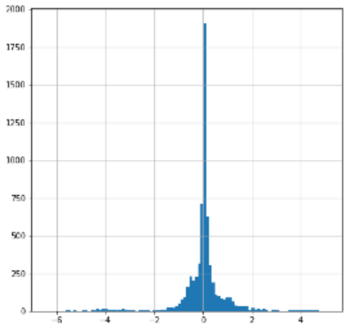
Lastly, the results of the ARMAX(1,0,1) model with Gaussian Mixture Segmentation are presented in this section. The MAPE of the model presented in Table 16 indicates that this model is slightly less suitable for forecasting than the model using K-means segmentation. However, based on the distribution of residuals presented in Table 17 and Figure 10, the distribution is less disperse. Nonetheless, this model is also clearly worse than the JVD model.

Table 16

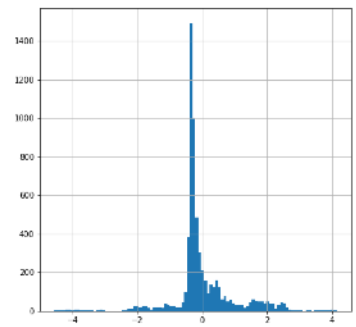
	A1	A2	B1	B2
MAPE - total	1.565	16.888	12.848	17.467
MAPE - [0-30]	1.484	16.534	12.847	17.424
MAPE - [30-60]	1.585	16.734	12.949	17.490
MAPE - [60-90]	1.625	17.397	12.748	17.489

Table 17

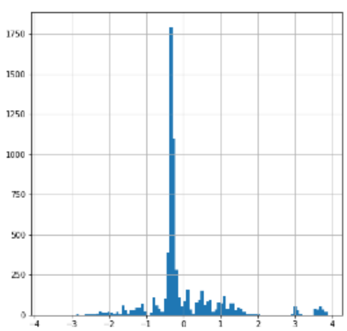
	A1	A2	B1	B2
mean	-2.03e-16	-4.20e-16	7.63e-16	2.44e-16
min	-6.46	-4.58	-3.71	-4.97
25 %	-1.41e-01	-3.27e-01	-3.48	-1.63e-01
50 %	5.97e-02	-2.56e-01	-2.91e-01	-1.45
75 %	1.97e-01	2.36e-01	1.44e-01	-2.79e-02
max	5.10	4.15	3.88	2.10



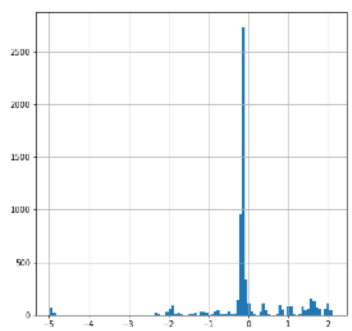
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 10: Plotted standardized residuals for all data sets using the ARMAX(1,0,1) model with GMM segmentation.

4.3 ARMA(3,3)

In this section the results using an ARMA(3,3) model are presented. Here too, the results are presented first when no segmentation was conducted, then with K-means segmentation and lastly, with Gaussian Mixture segmentation.

4.3.1 ARMA(3,3) without Segmentation

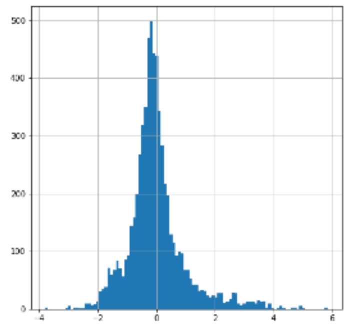
The MAPE of the ARMA(3,3) model are presented in Table 18 below. Comparing these values with the MAPE of the JVD model in Table 10 it is clear that the out-of-sample prediction error is significantly smaller using the ARMA(3,3) model for every data set. However, by examining the distribution of residuals in Table 19 and Figure 11 we can see that despite the prediction error being smaller, the residuals are more dispersed.

Table 18

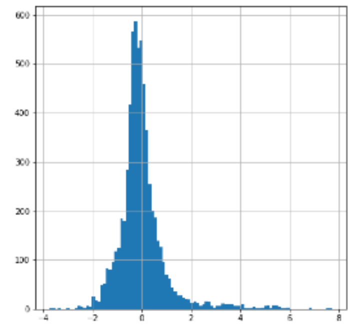
	A1	A2	B1	B2
MAPE - total	0.882	1.954	0.517	0.482
MAPE - [0-30]	0.888	1.970	0.519	0.416
MAPE - [30-60]	0.879	1.940	0.500	0.482
MAPE - [60-90]	0.879	1.953	0.532	0.548

Table 19

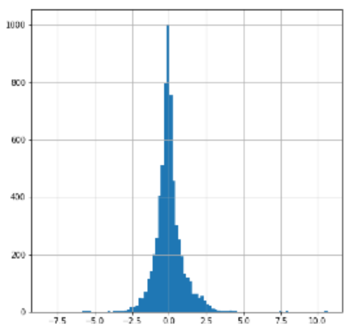
	A1	A2	B1	B2
mean	-4.42e-16	4.74e-16	6.54e-16	-3.57e-16
min	-3.79	-3.75	-8.32	-9.43
25 %	5.00e-01	-4.63e-01	-4.54e-01	-2.44e-01
50 %	-1.26e-01	-1.34e-01	-6.73e-02	-7.57e-02
75 %	3.03e-01	2.59e-01	3.52e-01	1.53e-01
max	5.83	7.72	1.07e+01	1.07e+01



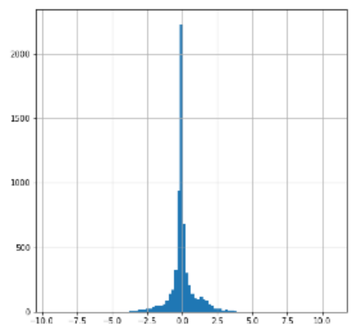
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 11: Plotted standardized residuals for all data sets using the ARMA(3,3) model without segmentation.

4.3.2 ARMA(3,3) with K-Means Segmentation

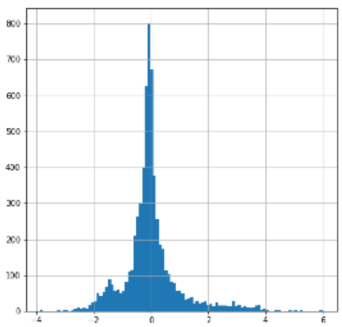
Using the ARMA(3,3) model with K-means segmentation resulted in slightly better out-of-sample performance than without segmentation as shown in Table 20. Nonetheless, the residuals are dispersed with this model as shown in Table 21 as well and moreover, for data sets A1 and A2 the distribution of residuals is clearly skewed as shown in Figure 12.

Table 20

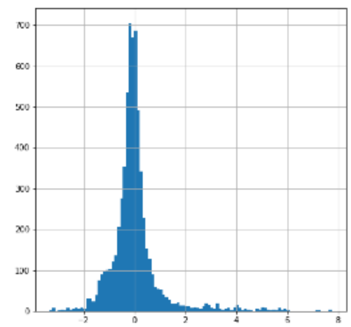
	A1	A2	B1	B2
MAPE - total	0.802	1.834	0.529	0.449
MAPE - [0-30]	0.807	1.831	0.509	0.362
MAPE - [30-60]	0.804	1.838	0.510	0.448
MAPE - [60-90]	0.796	1.833	0.567	0.537

Table 21

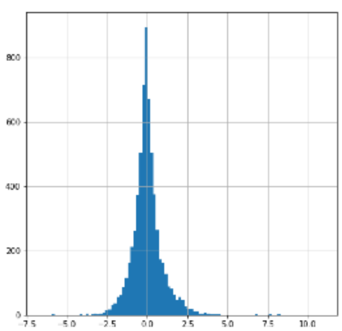
	A1	A2	B1	B2
mean	-5.32e-18	-4.00e-16	7.63e-17	-2.60e-16
min	-3.88	-3.34	-6.66	-1.02e+01
25 %	-3.89e-01	-3.91e-01	-4.65e-01	-2.72e-01
50 %	-7.98e-02	-9.00e-02	-6.17e-02	-6.70e-02
75 %	2.13e-01	1.94e-02	3.85e-01	2.01e-01
max	6.00	7.75	1.09e+01	1.14e+01



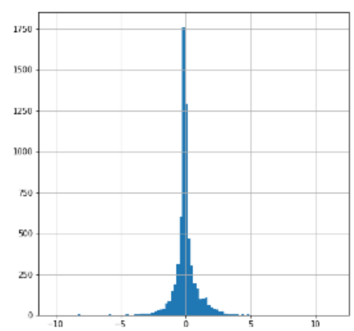
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 12: Plotted standardized residuals for all data sets using the ARMA(3,3) model with K-means segmentation.

4.3.3 ARMA(3,3) with Gaussian Mixture Segmentation

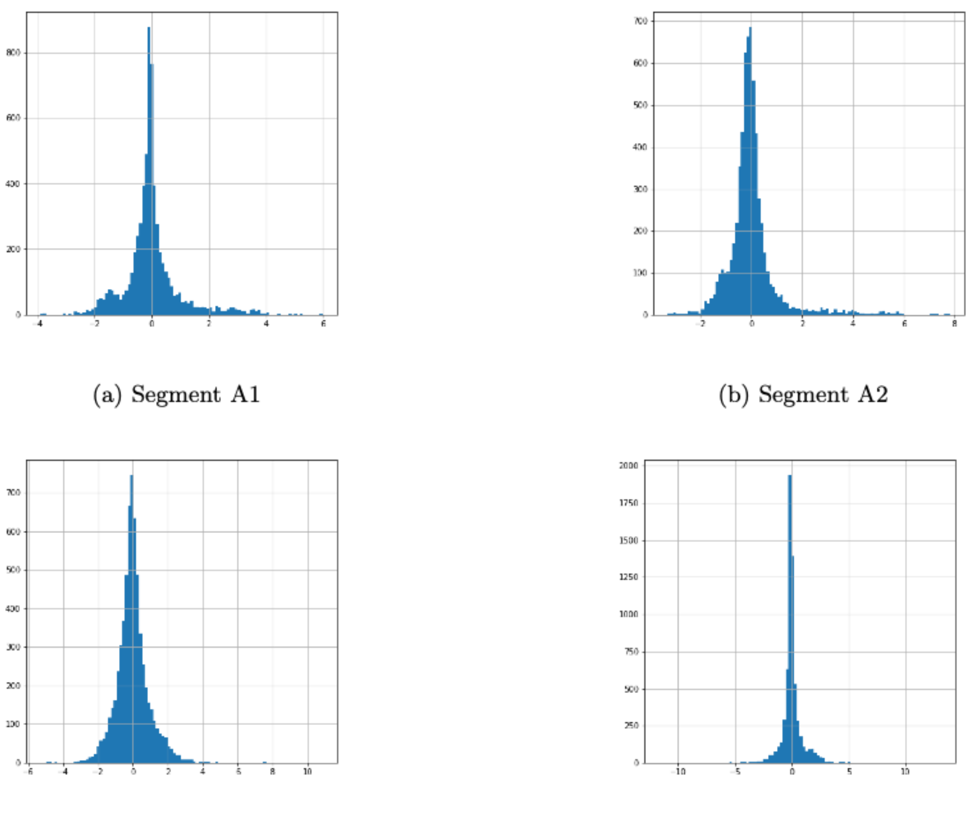
In the last subsection the results of the ARMA(3,3) model with Gaussian Mixture segmentation are presented. By comparing the MAPE in Table 20 with the MAPE in 22 no conclusive difference is indicated with the exception of data set B2, for which the model using GMM seems more suitable. However, by examining the residuals in Table 23 and Figure 13, it is clear that no significant change is seen in the behaviour of the residuals.

Table 22

	A1	A2	B1	B2
MAPE - total	0.804	1.833	0.559	0.377
MAPE - [0-30]	0.801	1.835	0.516	0.323
MAPE - [30-60]	0.806	1.830	0.527	0.370
MAPE - [60-90]	0.805	1.834	0.634	0.439

Table 23

	A1	A2	B1	B2
mean	5.31e-17	-2.72e-16	-1.36e-16	-5.51e-16
min	-3.82	-3.13	-5.32	-1.17e+01
25 %	-3.90e-01	-4.05e-01	-4.89e-01	-2.75e-01
50 %	-7.90e-02	-9.38e-02	-5.88e-02	-7.85e-02
75 %	2.09e-01	2.04e-01	3.95e-01	1.88e-01
max	5.88	7.57	1.09e+01	1.31e+01



(a) Segment A1

(b) Segment A2

(c) Segment B1

(d) Segment B2

Figure 13: Plotted standardized residuals for all data sets using the ARMA(3,3) model with GMM segmentation.

4.4 SARIMAX(0,1,1)(2,0,0,14)

In this section the results using a SARIMAX(0,1,1)(2,0,0,14) model are presented. Similarly to the previous subsections, the results are presented first when no segmentation was conducted, then with K-means segmentation and lastly, with Gaussian Mixture segmentation.

4.4.1 SARIMAX(0,1,1)(2,0,0,14) without Segmentation

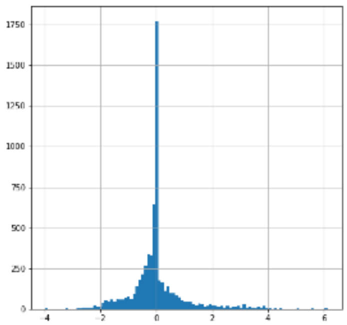
The MAPE of the SARIMAX(0,1,1) model without segmentation are presented in Table 24 below. By comparing these values with the MAPE of the previous models it is evident that the out-of-sample prediction error is significantly smaller using the SARIMAX(0,1,1) model. However, as with the previous autoregressive models the distribution of residuals presented in Table 25 and Figure 14 are more dispersed and more skewed than the residuals of the JVD model.

Table 24

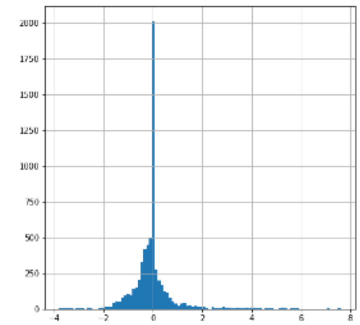
	A1	A2	B1	B2
MAPE - total	0.750	1.680	0.370	0.144
MAPE - [0-30]	0.751	1.689	0.375	0.144
MAPE - [30-60]	0.751	1.674	0.366	0.145
MAPE - [60-90]	0.749	1.676	0.369	0.144

Table 25

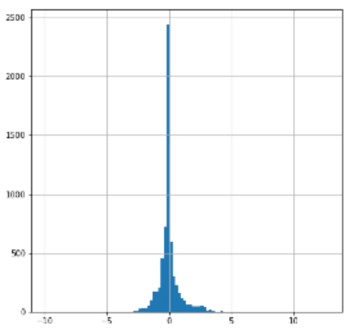
	A1	A2	B1	B2
mean	3.70e-17	-8.74e-18	1.19e-16	1.15e-17
min	-3.98	-3.81	-9.97	-1.84e+01
25 %	-3.55e-01	-3.47e-01	-3.35e-01	-2.04e-01
50 %	-2.21e-02	-8.60e-03	-4.47e-02	-6.45e-02
75 %	9.04e-02	5.56e-02	1.24e-01	-3.96e-02
max	6.13	7.61	1.27e+01	1.94e+01



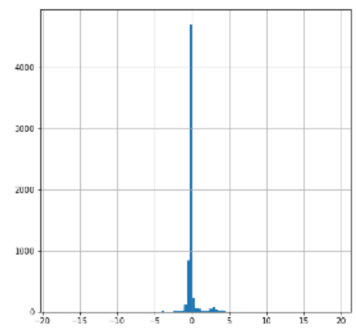
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 14: Plotted standardized residuals for all data sets using the SARIMAX(0,1,1) model without segmentation.

4.4.2 SARIMAX(0,1,1)(2,0,0,14) with K-Means Segmentation

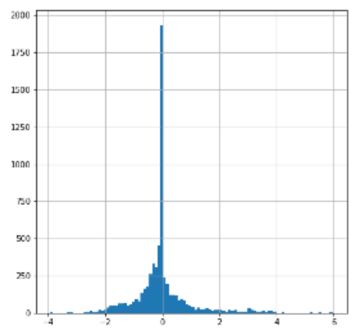
Using K-means segmentation did not significantly improve the prediction accuracy using the SARIMAX model as indicated by the values in Table 26 and the residuals in Table 27 and Figure 15.

Table 26

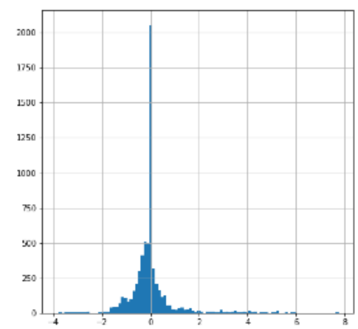
	A1	A2	B1	B2
MAPE - total	0.755	1.678	0.361	0.146
MAPE - [0-30]	0.756	1.684	0.366	0.147
MAPE - [30-60]	0.755	1.675	0.359	0.147
MAPE - [60-90]	0.753	1.675	0.357	0.144

Table 27

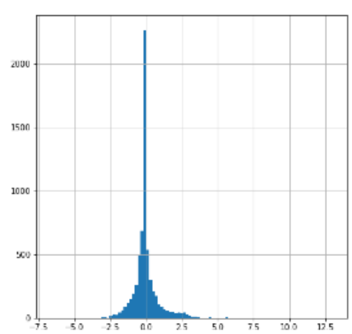
	A1	A2	B1	B2
mean	-2.17e-17	1.93e-17	-1.85e-18	2.31e-18
min	-3.93	-3.90	-6.66	-1.86e+01
25 %	-3.61e-01	-3.37e-01	-3.41e-01	-2.08e-01
50 %	-2.55e-02	-1.37e-02	-4.52e-02	-6.71e-02
75 %	9.53e-02	4.98e-02	1.12e-01	-3.85e-02
max	5.97	7.76	1.30e+01	1.95e+01



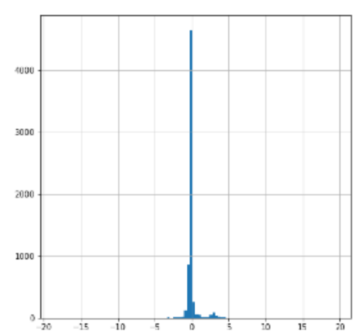
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 15: Plotted standardized residuals for all data sets using the SARIMAX(0,1,1) model and K-means segmentation.

4.4.3 SARIMAX(0,1,1)(2,0,0,14) with Gaussian Mixture Segmentation

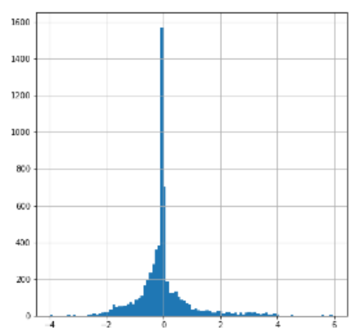
Similarly to using K-means segmentation, using Gaussian Mixture segmentation did not improve either the prediction accuracy using the SARIMAX model as indicated by the values in Table 28 and the residuals in Table 29 and Figure 16.

Table 28

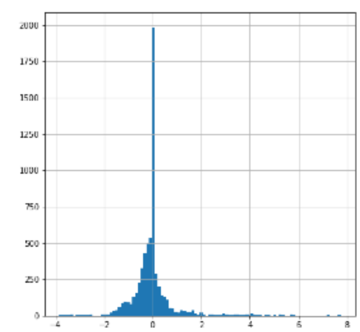
	A1	A2	B1	B2
MAPE - total	0.750	1.678	0.358	0.145
MAPE - [0-30]	0.750	1.683	0.365	0.145
MAPE - [30-60]	0.752	1.675	0.356	0.146
MAPE - [60-90]	0.747	1.677	0.353	0.144

Table 29

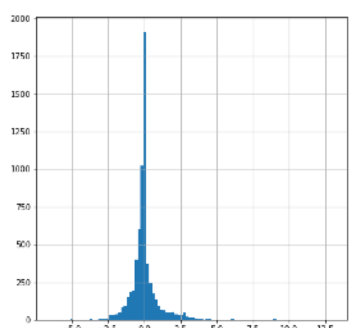
	A1	A2	B1	B2
mean	-2.65e-17	1.43e-18	-6.34e-19	3.98e-18
min	-3.98	-3.87	-6.49	-1.85e+01
25 %	-3.74e-01	-3.48e-01	-3.55e-01	-2.07e-01
50 %	-2.36e-02	-1.27e-02	-4.65e-02	-6.80e-02
75 %	8.20e-02	5.46e-02	1.17e-01	-3.85e-02
max	5.94	7.74	1.30e+01	1.94e+01



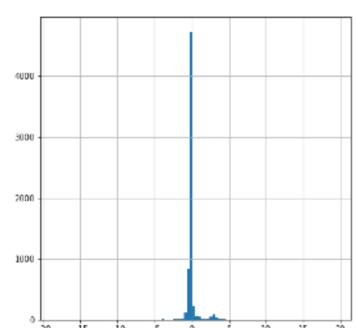
(a) Segment A1



(b) Segment A2



(c) Segment B1



(d) Segment B2

Figure 16: Plotted standardized residuals for all data sets using the SARIMAX(0,1,1) model and GMM.

5 Discussion

5.1 Reflection on Literature

Due to the effects of negative interest rates, the Jarrow and Van Deventer model was not ideal for predicting the behavior of non-maturity deposits. Stavrén and Domin (2019) obtained the same results. This method is inaccurate when used with the period used in this research but to study how much the accuracy drops when interest rates turn negative, a longer data sample would be needed.

Ahmadi-Djam and Belfrage (2017) considered an ARIMAX to be the best model to forecast non-maturity deposits. In this project, the SARIMAX model was deemed the best. The apparent reason that the SARIMAX did not fit well for Ahmadi-Djam and Belfrage (2017) was that the data set they used did not have information about the weekends and holidays. This also causes issues with fitting seasonality and predicting the results at one-day intervals. For this reason, we can not compare the accuracies directly. Furthermore, the segmentation used in their experiment was based on customer attributes and was not based on account behavior. For this reason, the segmentation did not increase the accuracy. In this research, the segmentation was done based on S-Bank's own criteria, and then these segments were split further using segmentation algorithms. This increased the predictive power for our models, disregarding SARIMAX. Parameters for the final SARIMAX were optimized based on the whole data set. The comparison between optimized SARIMAX and SARIMAX where the parameters are optimized for each cluster separately could be conducted for further research.

The settings in research by Stavrén and Domin (2019) also differ from ours. The models compared in said research used only monthly aggregated data observing the volumes. Also, this aggregate data is from Sweden. In their research, unlike in the research conducted by the Ahmadi-Djam and Belfrage (2017), SARIMA model was deemed to be the best model for modeling non-maturity deposits, and the exogenous variable did not increase the accuracy notably.

5.2 Assessment of the Results

The results obtained in this study clearly indicate that using the SARIMAX(0,1,1)(2,0,0,14) model both with and without segmentation provides the best forecasting accuracy out of any of the models tested. Although segmentation clearly improves the forecasting accuracy when the ARMAX and ARMA models are used, no improvement is seen with the SARIMAX model. The total MAPE of every model for each data set are presented in Table 30 where the best overall models have been bolded.

Model \ Data set	A1	A2	B1	B2
JVD (baseline)	2.304	6.134	2.066	1.652
ARMAX(1,0,1)	3.274	27.195	14.308	15.811
ARMAX(1,0,1) K-means	1.567	15.755	11.807	15.695
ARMAX(1,0,1) GMM	1.565	16.888	12.848	17.467
ARMA(3,3)	0.882	1.954	0.517	0.482
ARMA(3,3) K-means	0.802	1.834	0.529	0.449
ARMA(3,3) GMM	0.804	1.833	0.559	0.377
SARIMAX(0,1,1)	0.750	1.680	0.370	0.144
SARIMAX(0,1,1) K-means	0.755	1.678	0.361	0.146
SARIMAX(0,1,1) GMM	0.750	1.678	0.358	0.145

Table 30: The total MAPE of every model for each data set.

The total values are computed using the entire 90-day forecast for each forecast period. As the rolling window method was used by rolling one month forward, a total of 71 periods were used, which translates to 6390 observations in total. By examining the table above we can clearly distinguish that the ARMAX(1,0,1) model has the lowest predictive power both with and without segmentation. It is noteworthy that the MAPE decreases significantly for data sets A1, A2, and B1, although it slightly increases for data set B2. The ARMAX(1,0,1) model is the only model with less overall predictive power than the Jarrow and Van Deventer model that was used as a baseline, as can be seen by the fact that it only has a lower MAPE for data set A1 when segmentation is used.

Unlike the ARMAX(1,0,1) model, the predictive power of the ARMA(3,3) model is evidently higher than the predictive power of the JVD model. Moreover, for the more important data sets A1 and A2 segmentation seems to improve the predictive power roughly 10% as indicated by the values in Table 30, and for data set B2 it seems to slightly improve the predictive power. Interestingly, using the ARMA(3,3) model with either segmentation seems to yield the most accurate forecasts for the third month for data sets A1 and A2, as can be seen by comparing Tables 18, 20 and 22.

The overall best model is clearly the SARIMAX(0,1,1) model. The predictive power of all three versions was significantly better for every data set compared to any of the other models. To distinguish between the different SARIMAX models, despite only a slight improvement compared to the other two models the SARIMAX model using Gaussian Mixture segmentation proved to be the best overall model in terms of predictive power. As stated earlier, segmentation did not improve the predictive power to the same degree as it did with the other two models. As the first model also included the Euribor rate as an exogenous variable, it is unlikely that this behaviour stems from that. Therefore, the difference could be a result from either the ARMA(3,3) model being of a higher order or from the seasonality in the SARIMAX models.

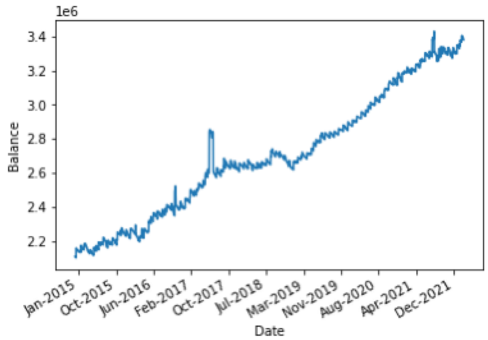
Although the number of observations used in the MAPE computation was high and thus, the results credible, the underlying data causes some uncertainty. This is most evident by examining the distributions of the standardized residuals of the models. As can be seen from Figures 8 through 16, the majority of the histograms present both significant kurtosis and skewness as the right side tails are wider than the left side tails, whereas the left side tails are fatter than the right side tails. This indicates that outliers are present which the forecast is not able model properly. By cleaning the data using, for example, winsorization, the models could possibly show more accurate forecasts and the residuals could behave more coherently. Furthermore, as the autocorrelations of the residuals was not examined the models still leave room for improvement.

6 Conclusions

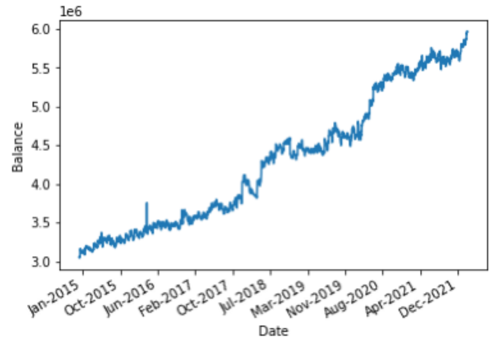
Non-maturity deposits are an essential part of banks' funds, and thus it is vital to understand their behavior well. These balances have seasonality and drift that can be modeled with different time series models. The most commonly used and most prominent exogenous parameter in these models has been the short interest rate due to its effect on the investment market's opportunity cost. However, the prediction power of this variable has decreased with negative interest rates.

The main objective of this project was to provide S-Bank information about the behavior of their non-maturity retail deposits and suggest a model that can predict the future values in a negative interest rate environment. We compared different models and concluded that the SARIMAX is the most accurate model for these. K-means and Gaussian Mixture clustering methods divided the customers into similar segments, which increased the accuracy of the models in most of the cases. However, clustering did not seem to affect the results of the final SARIMAX-model that was selected based on the accuracy measures for the whole data set. The presented method that uses clustering to improve the accuracy could be developed further by defining the SARIMAX parameters for each cluster separately.

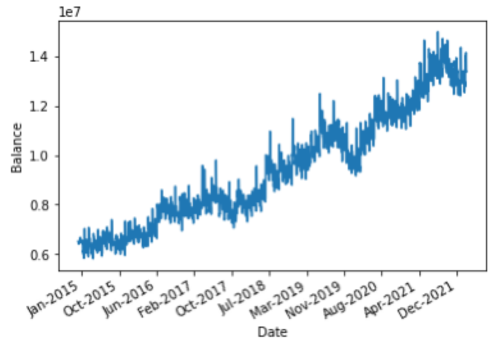
A Daily Total Balance



(a) Daily total balance of the account in dataset B1.



(b) Daily total balance of the account in dataset B2.



(c) Daily total balance of the account in dataset A2.

Figure 17: The daily total balance of each dataset.

B Summary Statistics

	Start of the Period	End of the Period
Mean	421.97	675.7
Median	114.98	168.55
Std Dev	1335.78	2431.81
Minimum	−235.31	−24.87
Maximum	42514.08	100405.87
Skewness	14.63	19.55
Kurtosis	330.18	633.26

Table 31: Summary statistics of the data in the B1 dataset

	Start of the Period	End of the Period
Mean	610.15	1191.68
Median	71.75	83.05
Std Dev	2797.69	6750.86
Minimum	−136.42	−49.85
Maximum	77483.42	213695.17
Skewness	13.46	17.26
Kurtosis	247.11	420.65

Table 32: Summary statistics of the data in the B2 dataset

	Start of the Period	End of the Period
Mean	1303.28	2671.29
Median	153.06	334.08
Std Dev	4852.84	11749.24
Minimum	−14526.97	−53.41
Maximum	113889.58	405507.1
Skewness	10.36	17.31
Kurtosis	148.24	451.5

Table 33: Summary statistics of the data in the A2 dataset

C Self Assessment

C.1 Project Implementation

The scope of the work has been mostly unchanged throughout the project, with the implementation largely following the initial project plan. The main departure from the initial plan is that the analysis of the effect of the Covid-19 pandemic, initially thought of as one of the main project goals, has been kept relatively brief as the project has mainly focused on modeling and forecasting. The Covid-19 effects have been investigated by comparing the variances before and after the start of the pandemic. However, the suggested model is selected such that the accuracy of the model before and during the pandemic is taken into account.

None of the risks considered in the initial project plan or the interim plan were realized, with the exception of the risk “Complete change in in the existing economic environment and monetary policy”, which was added to the interim plan in retrospect following the invasion of Ukraine. In addition, the interest rates have risen positive and are expected to increase further in the future. Historically, changes in interest rate levels have effected the model selection.

Throughout the project, we have met with the client organization regularly. During these meeting have have received guidance and answers to our questions regarding the topic.

C.2 Successes and Failings

The main project goal of developing a justified and documented model for analysing the risk characteristics of non-maturity retail deposits, has been achieved and the feedback from the client has been positive. The suggested model and its parameters have been derived with S-Bank-specific data. We have also provided the codes used in this project to allow further testing, research or implementation.

On the other hand, our team has not been able to gain any unforeseen insight, pandemic-related or otherwise, based on the data. The results suggest that the best model for non-maturity deposits would not get more accurate with segmentation. This might be due to the seasonality of this model or the low degree of its variables. We did not have time to investigate aforementioned things further due to the large workload towards the end of the project.

C.3 Takeaways

In hindsight, the project plan could be revised as the workload of the last weeks of the project has been somewhat demanding with the writing of the final report and the final stages of modeling coinciding with holidays and other team-member arrangements, such as work. In particular, our team could have begun analysing the project results earlier, possibly even before the interim report presentations. These results could have guided the aim of the project. However, our team could not have begun the modeling process earlier as the data was not available to us immediately, which made the uneven schedule to some extent unavoidable.

The presentation schedules and details, particularly the presentation lengths, could have been announced earlier.

References

A. Ahmadi-Djam and S. Belfrage. Forecasting non-maturing liabilities, 2017. KTH, Mathematics (Dept.).

J. Benhabib and A. Bisin. Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature*, 56(4):1261–91, 2018.

A. Blöchlinger. Identifying, valuing and hedging of embedded options in non-maturity. *Journal or Banking & Finance*, 50:34 51, 2015.

E. Cipu and S. Udriste. Estimating non-maturity deposits. *Simulation, Modelling and Optimization*, 1:369–375, 2009.

J.D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, N.J, 1994. ISBN 9780691042893.

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

M. Kalkbrener and J. Willing. Risk management of non-maturing liabilities. *Journal of Banking & Finance*, 28(7):1547–1568, July 2004. ISSN 03784266. doi: 10.1016/S0378-4266(03)00131-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0378426603001316.

S.P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129 137, 1982.

A. Machauer and S. Morgner. Segmentation of bank customers by expected benefits and attitudes. *International Journal of Bank Marketing*, 19(1):6–18, 2001.

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

OeNB. *Guidelines on Managing Interest Rate Risk in the Banking Book*. Number 7. 2008.

R. Selvaggio. Using the oas methodology to value and hedge commercial bank retail demand deposit premiums. *The Handbook of Asset/Liability Management*, pages 363–373, 1996.

F. Stavrén and N. Domin. Modeling of non-maturing deposits. Master's thesis, KTH Royal Institute of Technology, 2019.

Wikipedia contributors. List of bank runs, 2022. URL https://en.wikipedia.org/wiki/List_of_bank_runs. [Online; accessed 12-May-2022].