

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

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

May 2, 2022

1 Changes in Objectives and Scope

The main goal has remained the same in our project. That is, to learn more about the risk characteristics of S-Bank's non-maturity retail deposits. However, the scope has become narrower.

We have left out the account segmentation based on the payees of the transactions. This type of clustering would have made the size of data files harder to work with and could have made the workload too large for the course. Therefore, in this project, we cluster the accounts via S-Bank's readymade segmentation and algorithms that divide these accounts into smaller groups based on the relative daily changes in the account balances.

2 Status of Project

2.1 Completed Tasks

Since the delivery of the project plan the group has focused on the two major tasks of the project, namely segmenting the data using various methods and fitting the data to different forecasting models. In addition to this we have familiarized ourselves with the data; we have computed descriptive statistics for every dataset and found some outliers in the data which are being discussed further with the client.

In order to properly segment the data, the original data files were manipulated to omit data that was redundant with respect to the segmentation. Subsequently, several methods for segmenting the data have been used: S-Bank's own segment labels for each account, a K-Means clustering algorithm, a DBSCAN clustering algorithm, a Gaussian Mixture Model, and an Affinity Propagation clustering algorithm. Some initial results have been obtained using the segmentation techniques, although none of them have yet been used in

the subsequent forecasting. Based on the initial results, the K-Means clustering algorithm seems very promising although the tentative results are not conclusive. The segmentation methods cannot be verified until they can be used together with the forecasting models in order to compute out-of-sample error metrics.

The task completed this far have been completed by dividing the tasks between the group members. By doing so we have been able to allocate tasks that fall in line with the respective members' interests. As a result, we are on schedule with the segmentation and with applying existing forecast models to the data despite the somewhat late start.

The data has been used to fit several ARMA and ARMAX models. 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). The models were used to predict future daily values for the sum of account balances, but the results have not yet been analyzed 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 must be had to determine the most sensible interest rate to use in the modelling, the Euribor rates were used to build the functionality of the models since there was daily interest rate data available on them.

We have familiarized ourselves with the data that was provided by the client. The balances of the bank accounts are negative in the data. However, there are a few nonnegative values. When the additive inverse of each balance value was taken, the numbers made more sense. It was confirmed from the client that this is fine, and no information is lost, since the negative values signify money leaving from the bank, thus representing the bank's perspective. Now there were some negative values, which means that some people seemingly have spent money in a way that leaves their bank account with a negative balance. This could be due to said customers making small transactions that do not contain checking the balance of the account, thus making it possible to result a negative balance. What ever the cause, this needs to be studied further.

2.2 Current tasks

We plan to account for seasonal effects in our model by switching to a SARIMA (Seasonal ARMA) model and experimenting with various seasonal lengths, as it is highly probable that deposit behaviour is subject to weekly, monthly, yearly or by public holidays (Christmas, Midsummer etc.). Work on customer segmentation continues and, as mentioned before, the exact interest rate to be used in modelling is to be determined.

2.3 Next steps

After segmenting the data with clustering methods and fitting ARMA and ARMAX models to predict the future behaviour of customer bank deposits, it is time to proceed in the project with following next steps. So far the implementations have been fairly straightforward, but now to recognize the best customer based model for NMD risk characteristics, we must apply the obtained segments on the relevant ARMA and ARMAX models which were considered in subsection 2.1.

In addition, we will proceed on the uncompleted tasks in the project plan report such as comparing the model efficiencies and start the writing of the final report and decide how to report the results. The results of the best fitting model that best explains the risk characteristics will be then compared to the appropriate interest rate aggregate, mainly the different periods of Euribor rates.

3 Schedule

There have not been significant changes in the tentative schedule. We have decided to change our biweekly team meetings into weekly meetings.

Main Task	Activity	Week	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Classroom	Course kick-off																				
	Presentation of project plans																				
	Presentation of interim reports																				
	Presentation of final reports																				
Team	Team meeting																				
Client interaction	Client meetings																				
Literature review	Familiarizing to the topic																				
	Searching references																				
	Analyzing previous models and methods																				
Reporting	Writing project plan																				
	Writing intermien report																				
	Writing final report																				
Modeling	Familiarizing to the data set																				
	Applying already existing models																				
	Customer segmentation																				
	Implementing customer based model																				
Testing	Comparing model accuracies																				

Figure 1: Tentative schedule.

4 Risks

Table 1

Risk	Probability	Effect	Impact	Prevention
Collinearity of explanatory variables	High	Errors in explanatory variable significance	High	Instead of including collinear variables into the same model, creating separate models for each variable that has collinearity with other variables
Overly complex model	Medium	Loss of predictive power due to overfitting	High	Not losing sight of what is important, not including unnecessary variables or features
Inner functional errors in Python packages	Medium Low	Incorrect calculations on the auto-correlation of the residuals and the parameters of SARIMA-model	High	Alternatively use R language for data analysis
Data breach	Low	Violation of the NDA contract	High	Making sure we understand which parts of the data are private by discussing it with the client
Lack of client involvements	Low	The final project does not meet the expectations of the client due to lack of professional knowledge	High	Taking initiative in communicating with the client
Team member inactivity	Low	Workload increases for other members	Medium	Active communication, clear allocation of tasks, clear schedule
Complete change in the existing economic environment and monetary policy	High	Changes in interest rate trends diminish the predictive power of our model	High	Acknowledging the issue informing the client. Future adjustments might be needed.