MS-E2177 Seminar on Case Studies in Operations Research

S-Bank: Allocation of the Sales Price of the Credit Collection Portfolio

Interim report

Joonatan Honkamaa, Project Manager
Juhana Vehmas
Mattias Österbacka
Petri Koivisto

May 22, 2023

1 Changes in objectives and scope

No changes in neither the objectives nor the scope of the project.

2 Project status

2.1 Completed tasks

Before more comprehensive analysis, the data had to be preprocessed. We were given multiple different Excel spreadsheets, containing information about the defaulted loans since 2015. Using R, we combined these spreadsheets into a one data-frame containing all possibly relevant information needed in the model. This data-frame contains information about the loan type, the recoveries collected, the collection date, the current balance and the age group of the borrower. It also tells if the loan has been sold, fully paid or if the collection is still ongoing or terminated due to other reasons.

A linear model to predict the future cash flows for loans has been developed. When examining the balance for a single loan, an almost linear dependence between the balance and time was noticed. We then determined that a simple

linear model that determines how the balance of a single loan will behave in the future, will be chosen.

The explanatory variables are the the dates of payment for a single loan. The response variables are the balance of the loan at the date of payment. Using this model makes it easy to extrapolate and predict future cash flows.

In addition we chose a period of 5 years from the point of default, as the time where we predict cash flows. The reason why we stop after 5 years is that people rarely pay after that [1]. Thus, there is a hard cut off after 5 years. Figure 1 shows this concept.

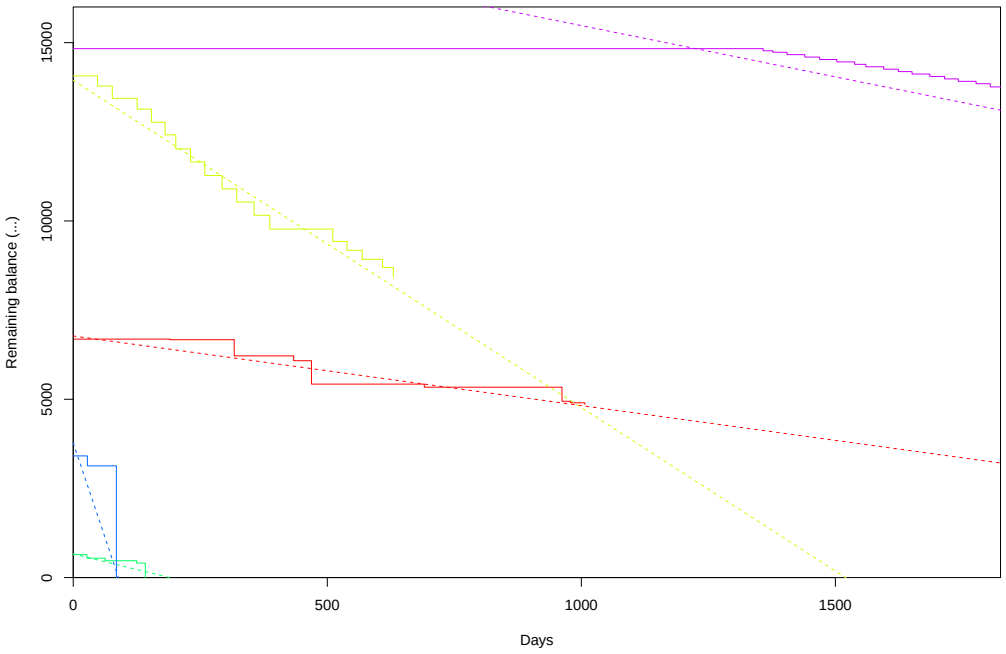

Figure 1: The balance of five randomly chosen loans, and their corresponding linear model, over a five year period.

From Figure 1 we notice that the loan marked with yellow is expected to be paid in full within a five year period, while the loan marked in red will not. To use these cash flows to predict LGD for loans, the predicted cash flows have to be discounted on a yearly basis. This means that even though a loan is predicted to be paid in full, the LGD will not be 0, due to the cash flows being discounted.

2.2 Current tasks

The main goal of this project is to allocate the portfolio sales price for the individual loans in the portfolio. We were given the sales price of a one portfolio sold to the collections agency in the spring of 2022. Now the goal is to use the linear model defined in the previous section, and determine the possible future cash flows for the loans in this portfolio. After the predicted cash flows are calculated and discounted to the time of the sale, we can allocate the sales price for an individual loan i such that:

$$\frac{DFC_i}{DFC_P} = \frac{S_i}{S_P}$$

where DFC_i is the discounted future cash flows for loan i, DFC_P is the sum of the discounted future cash flows for the whole portfolio, S_i is the sales price allocated for loan i and S_P is the total portfolio sales price.

We can use the linear model only for those loans that have received payments after the event of default. However almost 60% of the loans in the portfolio, sold in the spring of 2022, are loans without any payment history. The next step in the allocation is to decide how to deal with these types of loans.

Another important aspect to consider is how to validate the results of our model. This affects both the reliability and quality of our results, reliability for obvious reasons and quality because we are still to choose the exact form of our model. We have two main ideas for validation. First idea is to construct the model based on earlier payment history and compare the results of that model to later payment history of the loan. By repeating this process for each loan we get a distribution of errors which we can use to see how well the model performs.

Second idea is similar but it is based on a special data set: In spring 2022 S-bank had a portfolio of loans they decided not to sell yet. By now we have information about these loans from a time period of almost a year, and we can use this information for validation by constructing the model based on information available when the portfolio was decided not to be sold. The data sample is smaller, but assumably a reliable representation of a typical portfolio. We aim to use the both validation ideas and compare the results.

2.3 Next steps

The next steps to be completed in the project are

1. deciding how to treat loans with little or no payment history,
2. performing comparison of validation methods,
3. improving and possibly parameterising the model based on validation,
4. calculating DFC values for the dataset,
5. allocating the sales prices of portfolios based on the DFC values,
6. analysing the results of sales price allocation,
7. final report writing.

3 Schedule

We have some minor changes in the dates of our client meetings. The updated schedule for the project is presented below in Figure 2.

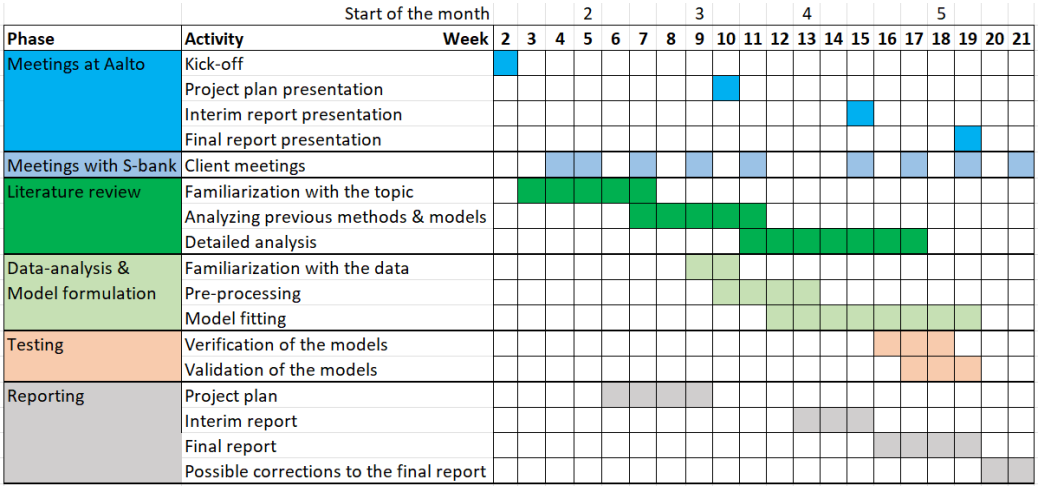

Figure 2: Updated Gantt chart for the project.

4 Risks

The updated risk management plan is presented in Table 1. Many risks have become less likely in comparison with the project plan because we have managed to follow our schedule without encountering major problems. However, the risk of poor validation of the model is a new recognized risk and has been added to the plan.

Risk	**Probability**	**Effect**	**Impact**	**Prevention**
Poor data quality	High	Created models are not predictive	Medium	Active communication with S-bank experts, identification of outliers and justified assumptions
Team member inactivity	Low	Increased workload of other members	Medium	Clear allocation of tasks and investing to building team spirit
Model overfitting	Medium	The model reacts too strongly to data features	Medium	Careful analysis of model performance
Validation methods for the model are poor	Medium	Estimating the accuracy and usability of the model is difficult	Medium	Comparison of different validation methods and communication with S-bank experts
Macroeconomic phenomena have affected the data	Medium	The model does not represent current behavior of loan cash flows	Medium	Recognition and clear communication about the issue
Insufficient communication with the client	Low	The model does not satisfy the requirements	Medium	Clear and regular communication between the team members and the client
Predictive power of the model is poor	Medium	The results are not useful	Medium	Analysis of model ideas within our team and with the client

Table 1: Updated risk management plan.

References

[1] Burkhard Heppe (2022):Valuation of Non-Performing Loans: Calibration of unsecured recovery curves; https://www.nplmarkets.com/es/research/article/valuation-of-non-performing-loans-calibration-of-unsecured-recovery-curves-/; retrieved 2.4.2023