Modeling and Forecasting Customer Behavior for Revolving Credit Facilities

Radoslava Mirkov[1], Holger Thomae [1], Michael Feist [2], Thomas Maul [1], Gordon Gillespie [1], Bastian Lie [1]

[1] TriSolutions GmbH, Hamburg, Germany,
[2] DZ BANK AG, Frankfurt am Main, Germany

E-mail for correspondence: `radoslava.mirkov@trisolutions.de`

Abstract: Revolving credit facilities are studied with an aim to forecast customer behavior and thus minimize liquidity and income risk. Historical data provide information about characteristics and withdrawal patterns of borrowers, which enables a deeper insight into their behavior. The withdrawal patterns are expressed in terms of the relative withdrawal and its variance over the contractual period, and the relative duration. Possible explanatory variables are for e.g. a customer's rating, a credit line limit, a lifespan of the credit facility, the economic sector of the customer, etc. We study the dependence of the target and explanatory variables utilizing statistical modeling techniques integrated in data mining methodology. Multivariate recursive partitioning enables recognition of the customer behavior, and the prediction of future withdrawals for new customers.

Keywords: revolving credit facilities; customer behavior; data mining; multivariate recursive partitioning; prediction.

1   Introduction and Motivation

Lines of credit are the focus of our investigation. In this paper we concentrate on revolving credit facilities (RCF) and want to forecast customer behavior taking the RCF. The RCF, often referred to as a revolver, is a flexible loan which allows the borrower to use the funds when they are needed. It can be taken out by both corporations and individuals. The bank guarantees the customer a loan up to the credit limit during a lifespan of the credit facility, without having to reapply each time the cash is needed. As the borrower repays the money, it is available to be borrowed again. This is a flexible loan available to clients which has been designed so clients can repay their loan balance to zero but still have the facility available to borrow again. This facility is useful for clients who may not want to use a loan facility at all times but appreciate the possibility that the agreed limit will be available up to a contractual maturity of the facility. In other

words, the borrower is under no obligation to actually take out a loan at any particular time, but may take part or all of the funds at any time over a period of several years. This agreement is common in situations in which a business must pay obligations but its operating cashflow varies, for e.g. seasonally. At any given time, the balance due may fluctuate from zero to the maximum credit limit. The interest rate paid on this kind of credit is usually floating, i.e. it is linked to EURIBOR or LIBOR so that customers receive market rate. In general, it is available in all major currencies.

In the case a bank grants the RCF, it is faced to several very specific risks beside the common credit risks. The most obvious risk is the uncertainty up to which extent of the facility will the borrower draw a loan. This depends usually on the operational needs of the customer, so the RCF might be used just for back up reasons to bridge liquidity gaps in stressed market situations, or it might be permanently drawn to some extent, due to his general operational needs. If the bank would hedge the liquidity requirements fully, the RCF wojust in time when the money is drawn, it might happen that there is no sufficient liquidity available in the markets, especially in times of financial crisis, and it is not warranted that this liquidity is accessible at reasonable prices, and the loan can cause high losses when interest rate levels are higher than they were at the time the RCF was granted and its conditions were fixed. These risks are also be hedged completely, but the conditions would no longer be competitive for the customer. If the bank decides, on the other hand, to refinance the loans under the RCF two characteristics of liquidity risk: availability and term transformation.

Beside these liquidity risks the customer is granted several options that have the potential to have a high impact on the revenues of the bank. For e.g. the withdrawal date can be chosen, or in which currency and for what term the loan is drawn. The hedging of these options is also very difficult for the banks and causes additional opportunistic costs. We will also focus on the rights of cancellation of the RCF. Usually (and especially in the German jurisdiction) every borrower has the right to cancel the facility at the end of each thereunder withdrawn loan. In the case the RCF was hedged for a longer term and cancelled prematurely the bank might be faced with the problem to reinvest the former acquired liquidity at lower interest rate which would again imply losses. All this issues influence the loan conditions, and imply the risk adjusted conditions for customers. One might suggest to develop an option pricing model to deal with all of the above mentioned customer options. But option pricing models are not adequate when one takes into account that the behavior of the customer does not strictly depend on the current market conditions (interest rates) and not only on the specific credit prices (credit spreads) for liquidity. One could assume that most of the small and medium sized customers behave just according to their specific operational needs. Figure 1 illustrates the described situation.

The figure contains the following labels:

Maximum amount of withdrawal

Expected maturity due to historic cancellations

Confidence Level for Liquidity Hedge Determines the Funding costs

Expected Average Withdrawal

Confidence Level for Earning Risk Provisions to compensate Underestimated Withdrawals

Realized Withdrawals

Today

Future Withdrawal Period with unknown
- Withdrawal Period Length
- Amount
- Drawn Currency
- Cancellation Date
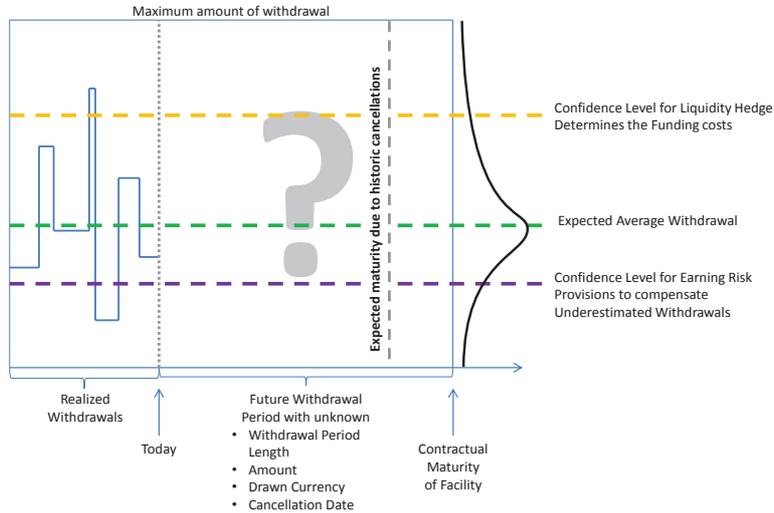
Contractual Maturity of Facility

FIGURE 1. The Principle of the RCF.

This drives the authors to investigate the historic behavior of a large portfolio of RCFs with the intention to identify behavioral clusters for the drawings under RCFs that allows the unique assignment of a cluster to each RCF. The aim is to predict up to a certain degree the likely future behavior of a specific client in a way that allows the bank to find an appropriate strategy for refinancing and therefore minimize liquidity risks and earnings risks.

2   Data Description

Data for this study is provided by one of the largest German banks. It contains historical data related to the RCF customers and their behavior. Mean withdrawals over the lifetime of the RCF and their variance as well as the relative duration of the actual withdrawal compared to the contractual duration of $n = 128$ agreements from one specific business branch with different customers are included. Further facts about these agreements, such as the duration of the contract, the agreed limit and the currency are also provided. As we are interested in the customer behavior, the information about customers like the rating, the probability of default, the type of

financing, their economic and risk group, is also relevant. Additionally, the data describes the purpose and some properties of the loan.

3   Data Mining Modeling Approach for the RCF Forecasting Customer Behaviour

Data mining, the process of automatically discovering patterns in large data sets, includes predictive modeling approach. Models for the continuous target variable as a function of the explanatory variables based on the regression techniques are built. We chose recursive partitioning (RPART) to build regression models of a very general structure

$$y_i = \sum_{i=1}^{n} x_i^k, \qquad n = 128, \ \ k = 15,$$

based on the ANOVA method, as y is a numeric vector. The splitting criteria is $SST - (SSL + SSR)$, where SST is the sum of squared errors for the node, and SSR and SSL are the sums of squared errors of the right and the left child-node, respectively.

In De'ath, G. (2002) the generalization of this approach to multivariate target variables, which also includes cross-validation of the results is proposed. Thus, the solution with the best predictive power is retained. Multivariate regression trees (MRT) turn out to be a powerful and robust method that handles a variety of situations, such as missing values, non-linear or higher-order interactions among explanatory variables, as well as categorical or quantitative predictors.

As we are interested in the mean relative withdrawal and its standard deviation simultaneously, we choose the MRT method to analyse the data. The measure of predictive error is also of great value.

In the RCF we investigate which of the above mentioned explanatory variables influence the mean relative withdrawal over the lifespan of the loan and its variance, i.e. standard deviation. The type of financing of the RCF customers and the risk group seem to influence the target variable to the great extent.

In our model

$$y_i = (w_i, sd_i), \qquad n = 128,$$

where w_i stands for the mean relative withdrawal, and sd_i for its standard deviation. The variables x_i^k with the best predictive value turn out to be the type of financing, denoted by $X1$, and the risk group of the borrower, given by the variable $X2$. Both explanatory variables are categorical. $X2$ describes how risky is a specific group of RCF consumers, whereas each level of the variable implies different levels of risk. The levels of the financing type X1 are coded and define project (1) and acquisition financing (2),
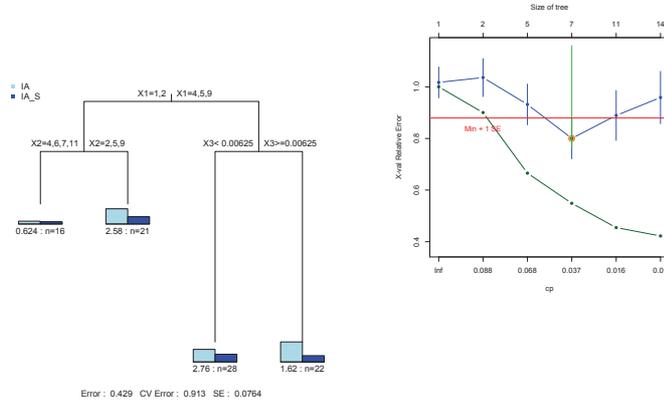
FIGURE 2. Multivariate Decision Tree and its Cross-Validation Error for the Mean Relative Withdrawal and its Standard Deviation.

TABLE 1. Mean Relative Withdrawal and its Standard Deviation for Obtained Cluster.

cluster mean	C1	C2	C3
w	0.10	0.50	0.60
sd	0.07	0.25	0.20

credits for trading (4) and sturctured products (5), and other (9) types. The default probablity is denoted by $X3$.

The results of the MRT procedure are shown in Figure 2. The minimal cross-validtion error is attained for seven nodes, as shown in 2 (left). The relative and cross-validation error in dependence of the tree size can be seen in Figure 2 (left). The red dot denotes the optimal solution based on the cross-validation. We note that the cross-validation error drops to 0.8, whereas the relative error drops below 0.5. We refer to Table 1 for mean values of the possible resulting cluster. The models are implemented in R using the packages RPART and MVPART.

4   Conclusions

We investigate prediction of the customer behavior based on the multivatite data mining models. This seems to be the best available method for our purpose, as the model with the best predictive properties is selected and

the multivariate target variables can be analysed. It turns out that the type of financing and the risk category of the consumer are of particular importance for the behavior of the loan takers. The probability of default seems to influence their behavior, too.

The cross-validation error, which indicates the stability of the prediction drops to 0.8. This means that the predictive power of the model for the given subset of data could be improved. Further steps include more detailed analysis based on a complete data set of better quality, and an investigation of other explanatory variables.

References

Tan, P-N., Steinbach, M., and Kumar, V. (2006). *Introduction to Data Mining*. Pearson Education.

Atkinson, E.J., and Therneau, T.M. (2000). An Introduction to Recursive Partitioning Using the RPART Routines. *Technical Report*, Mayo Foundation.

Kordichev, A., Powell, J.G. and Trippe, D.W. (2005). Structural Models of Revolving Consumer Credit Risk. *Credit Scoring and Credit Control Conference Proceedings*, 1 – 13.

De'ath, G. (2002). Multivariate Regression Trees: A New Technique for Modeling Species-Environment Relationships. *Ecology*, **83(4)**, 1105 – 1117.

Borcard, D., Gillet, F., and Legendre, P. (2011). *Numerical Ecology with R*. Springer.