Modelling LGD for unsecured personal loans: Decision tree approach

Lyn C. Thomas
Christophe Mues
Anna Matuszyk
University of Southampton[1]

Abstract

The Basel New Accord which is being implemented throughout the banking world on 1 January 2007 has made a significant difference to the use of modelling within financial organisations. In particular it has highlighted the importance of Loss Given Default (LGD) modelling.

We propose a decision tree approach to modelling LGD in the consumer credit area and using real data from the financial organisation in UK model the components that make up this tree.

Key words: Basel II, consumer credit, LGD

1. Introduction

The New Basel Accord allows a bank to calculate credit risk capital requirements according to either of two approaches: a standardized approach which uses agency ratings for risk-weighting assets and internal ratings based (IRB) approach which allows a bank to use internal estimates of components of credit risk to calculate credit risk capital. Institutions using IRB need to develop methods to estimate the following components for each segment of their loan portfolio:

– PD (probability of default in the next 12 months);
– LGD (loss given default);
– EAD (expected exposure at default).

Modelling PD, the probability of default has been the objective of credit scoring systems for fifty years but modelling LGD is not something that had really been addressed in consumer credit until the advent of the Basel regulations.

[1] L.Thomas@soton.ac.uk; cm1@soton.ac.uk; amatuszy@soton.ac.uk;
University of Southampton, Management Department, Highfield, SO17 1BJ, Southampton, UK;

What LGD modelling had been done was mainly in the corporate lending market where LGD (or its opposite Recovery Rate RR, where RR=1-LGD), was needed as part of the more sophisticated bond pricing formulae. Even there, for over twenty years LGD was often set at around 40% because of some historical analysis on a subset of bonds done in the 1960s. It was only in the last decade that its dependence on economic conditions, type of loan and type of borrower we recognised as important and the book by Altman et al:"Recovery Risk".

Such modelling is not appropriate for consumer credit LGD models since there is no continuous pricing of the debt as is the case on the bond market. The purpose of this paper is to address this issue by looking at how modelling the collections process may give a handle on a LGD model.

The idea of using the collection process to model LGD was suggested for mortgages by Lucas (2006). The collection process was split into whether the property was repossessed and the loss if there was repossession. So a scorecard was built to estimate the probability of repossession and then a model used to estimate the "haircut" the percentage of the estimated sale value of the house that is actually realised at sale time.

In the remaining section of the paper, we introduce a model for estimating LGD for revolving and unsecured consumer credit which we then tested out using personal loans data. The important aspect of the model is to allow one to include both the decisions made by the lender and the risk of the borrower not being willing or able to meet the debt obligations. This is important since the Basel Accord is interested in estimating downturn LGD and in such circumstances both the lenders' collection decisions and the borrowers ability to repay may change. In section two we describe the overall model while in section three we show how to model the recovery rate for given collection decision, by using data in a case study of modelling the recovery rate for an in-house collection process on personal loans debt. Section four draws some conclusions.

2. Decision tree LGD model

Default occurs when an obligor fails to meet a financial obligation (Frye 2004). For a defaulted loan, loss given default (LGD) is the proportion of exposure that is lost. For a loan that has not (yet) defaulted, LGD is a random variable.

There are many reasons why customer fails to make a payment (McNabb, Wynn 2000). The majority are related to a customer's change in circumstances, and/or an inability to manage finances. Typical examples include:

- loss of job or marital breakdown – such changes cause disruption to the normal pattern of life, and in most cases reduction in income, which results in temporary or longer-term payment problems;
- occasional missed payments, either because the customer forgot or was on holiday – these customers generally pay within the month and rarely require or experience any action from the collections department;
- incorrect set-up of a direct debit facility or set-up too close to the payment date – these are sometimes called "technical arrears" cases and typically affect the first payment made on a new account; and
- Disputes either with the lender, or more frequently with a retailer from whom they have purchased goods using the lender's credit facility – in these cases the customer may withhold payment until the dispute is resolved.

LGD is an outcome of a mix of random events and decisions made by the lender to decide what kind of collection strategy should be used; whether to collect in house, give to an agent at a certain percentage or sell off to a third party at a fixed price. These decisions affect the outcome but so does the risk of non-repayment.

2.1. Collection models on macro level

Generally, it can be said, that companies collect the debt mainly in house and have their own collection departments. However some companies do use outside agents and from time to time they sell off their debt to third parties. Accordingly, collection process was divided into 3 phases:

1. Collection process in house;
2. Collection process using agent;
3. Selling off the debt.

This illustrates one of the important issues in LGD modelling namely that LGD depends not only on the uncertainty of whether a defaulter will repay, or how much they can afford to repay, but also on the lender's collection policy. For example the three macro-level strategies identified above put different bounds on the possible LGD values, e.g.

Collection in house $\Rightarrow$ $0 \leq LGD \leq 1$

Collection by agent on 40% commission $\Rightarrow$ $0.4 \leq LGD \leq 1$

Sell off at 5% of face value $\Rightarrow$ $LGD = 0.95$

One way of modelling a problem where the outcome is a mix of decisions and randomness is by a decision tree and the decision tree to represent the collections process is shown in Figure 1.

The proposed tree starts with the division of the group into two sub-groups, according to whether the details of the debtor's address and telephone number are known and accurate: trace and no trace.

- If there is **no trace**, there is a little point in collection in house (one may wish to undertake some effort to trace the customer initially but the no trace outcome would be the result after this initial effort). The lender must decide whether to sell off the debt or use an external collection agency (the first one). If the agency is not able to recover the debt, the lender has again two choices: sell off the debt or sell to the second collection agency. The second agency will demand a higher commission for recovering the debt (since it is older and more difficult to recover).
- If there is **a trace** (i.e. the address and contact details are correct), different strategies can be used. Should an attempt be made to recover the debt in house or not. Normally this decision is based on subjectively chosen rules but one could also develop models to estimate what is the likely recovery rate (RR = 1-LGD) if collected in house, and separately what would be the recovery rate if collection by an agent. The rules used (and the model if built) can depend on several factors:

- How old is the debt,
- Amount of the debt,
- Type of the product,
- In which area is the client,
- Maturity of the customer.

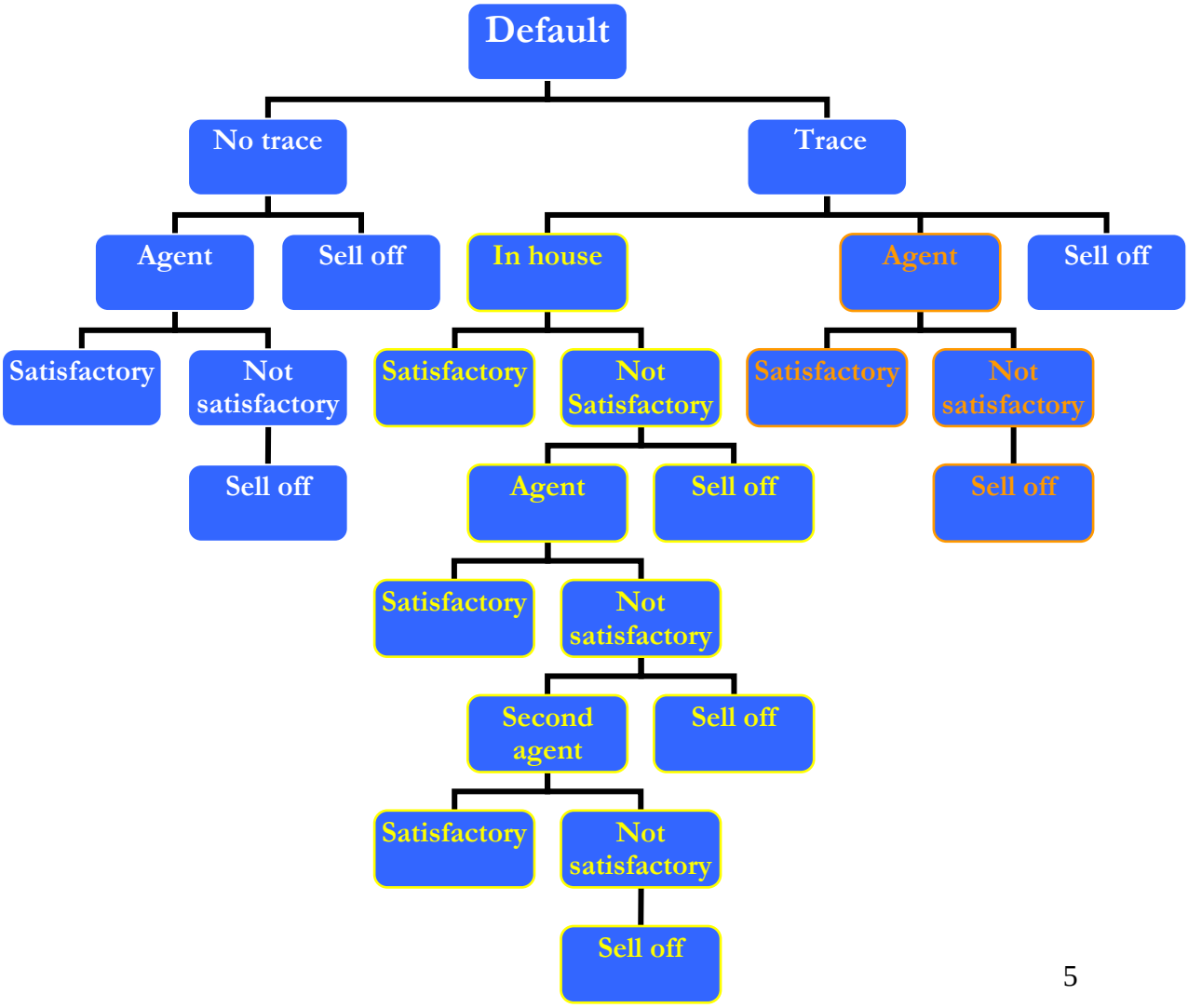
If the collections department is not able to recover a satisfactory amount of the debt within a given time, the lender can sell off the debt or send it to collection agency. If the collection agency has not been successful, the debt can be sold off or sent to a second collection agency. Usually there is a higher commission for the second agency

– the commission depends on how difficult it is to recover the debt. The price the debt can be sold at and the commission rates required are usually agreed by negotiation. They depend on the type and quality of the debt, how long it has been outstanding, but also the current debt market and economic conditions. Hence when estimating downturn LGD for Basel purposes one needs to recognise that one would need to be more conservative in these estimates than the historic long run average.

2.2. Collections Model: Operational Level

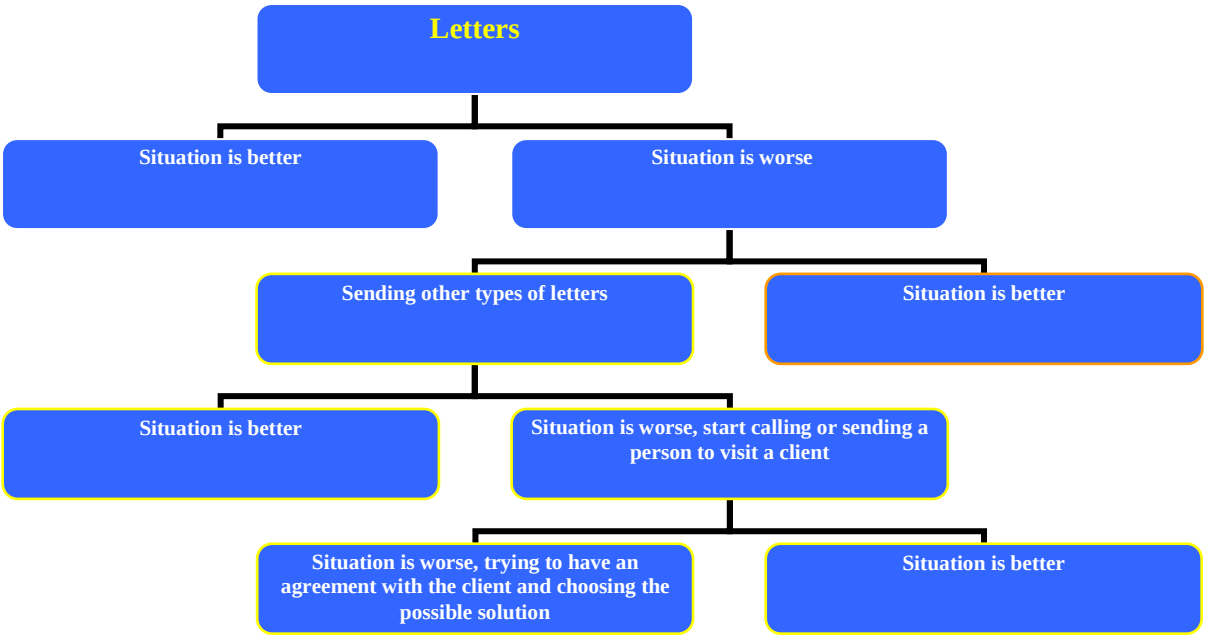
A collection department has the range of tools to use in the recovery process: from the gentlest to the strongest. Usually it starts with contacting the debtor by telephone or letter and trying to arrange for immediate repayment of the debt or some repayment arrangements to be agreed. There are different types of letters and sending them depends on the status of the customers and the characteristics of the debt. So within the in house collection node of Figure 2 is another decision tree which seeks to identify what sequence of actions to undertake and what outcomes make one decide to change the course of action.

Figure 1 Collection process



For example, one might have a simple decision tree on the policy with which letters to use as in Figure 2. As well as deciding which sequence of action to undertake, the operational strategy has also to decide what repayment agreement is acceptable to them. Initially the collections policy will seek to recover all the debt, but it may be that when the debt has proved difficult to recover partial repayment may be acceptable. With the advent of IVAs Individual Voluntary Arrangements such decisions will be increasingly common

Figure 2: Letters diagram



3. Repayment model

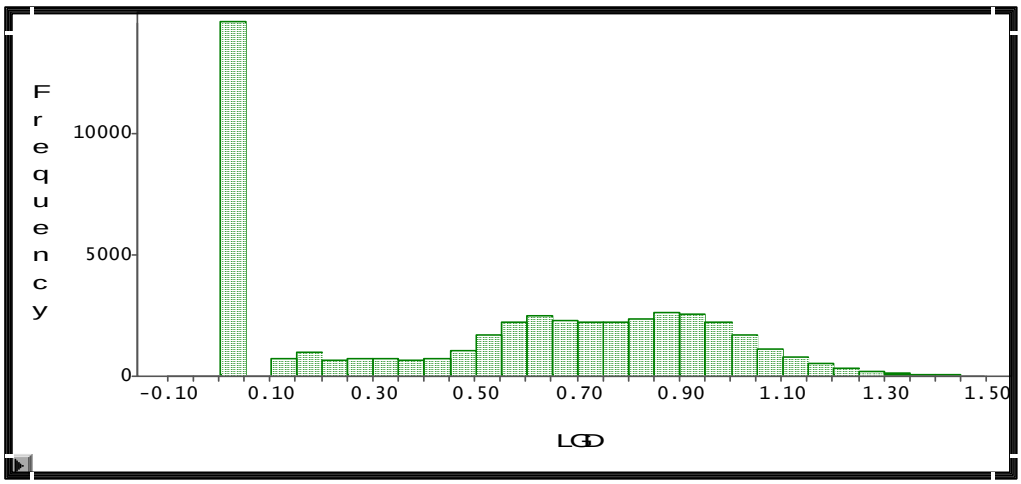
In this section we describe a repayment model. We do so by using the methodology to build an in house repayment rate model using personal loan data from a financial organisation in UK. It consists of almost 50K cases (all defaults) on personal loans granted between 1989 and 2004. Data was provided on individual level. In this case the lenders policy was to seek to collect everything in house and not to use agents. Here the decision tree simplified essentially to the repayment rate model.

Distribution of LGD for in – house collections

Analyzing the distribution of LGD, in Figure 3, it showed that 30% of the debtors paid in full and so had a LGD=0. Less than 10% paid off nothing. Note that for some such debtors the resulting LGD was greater than 1 since fees and legal costs had been added. This is not the case usually in agency collection where in some of the data sets

considered almost 90% of the population have LGD=1. It is clear that the more attempts that have been made to collect from the debtor in the past, the higher the likely LGD will be.

Figure 3: Distribution of LGD in the whole sample.



3.1. Repayment model: identifying class of repayer

So there are two parts of building this model. In the first part one we tried to estimate which of the two groups an individual debtor would be expected to be in, where the split is according to the values of LGD: LGD≤0 and LGD>0. This division into two was partly data driven and partly from knowledge of the lender's collection policy. We are investigating whether one can use algorithms that identify the number of different distributions in a mixture of distributions to assist in this choice of number of classes. A logistic regression model was built to try and separate out the two groups.

The following variables were used in the logistic model (LGD=0 or >0)

- Amount of the loan at opening;
- Number of months with arrears within the whole life of the loan;
- Number of months with arrears in the last 12 months;
- Time at current address;
- Joint applicant.

Results for the LGD=0 model

- The higher the amount owned the lower the chance of LGD=0;
- The longer the client lives at the same address the higher the chance that LGD=0;
- LGD more likely to be 0 if there is a joint applicant;

But

- The more the customer was in arrears in the whole life of the loan the higher the chance that LGD=0, however these who were in arrears a lot i.e. triple the average of time being in the arrears had lower chance of paying off everything;
- The more the customer was in bad arrears recently (in the last 12 months) the more chance LGD=0.

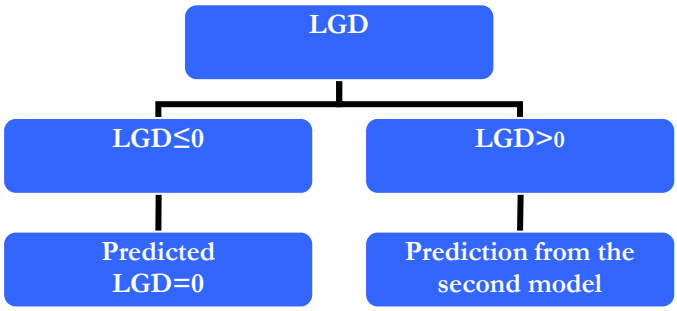
These last two remarkable results were confirmed by looking at two other data sets. It would appear that those have been often in arrears before defaulting are more likely to repay than those who have never been in arrears and for who it appears some very serious event changes their ability to repay. We liken it to "falling off a cliff". Those who have been just keeping their "head above water" are more likely to survive if they go under the waves, than those who have never been close to the water and drop down from a considerable distance above. But there is a limit to this analogy in that, these who are persistently in arrears (more than three times the mean value) have a lower chance of paying off in full.

3.2. Repayment model: estimate of LGD within a class

In this pilot model, it was decided that the predicted value for those in the first class should of course be LGD=0. For those in the second group (LGD>0) the LGD was estimated using a linear regression with weight of evidence approach. In the WOE approach we classified these variables as the target variable whether the LGD value was above or below the mean. In each case we started with ten roughly equally sized groups and combined adjacent groups with similar odds. We then defined WOE modifications for each variable which took the weight of evidence value for each bin that the corresponding variable had been classed into. Crudely if N_a and N_b are the total number of data points with LGD values above or below the mean and n_a(i) n_b(i) are the number in bin i with LGD values above or below the mean. The bin is given the value:

$$\log\left(\frac{n_a(i)}{n_b(i)}\Big/\frac{N_a}{N_b}\right)$$

Figure 4: Model building process



Using univariate analysis we identified five variables which were the strongest predictors of the LGD value for those with LGD>0.

- Number of months with arrears in the whole life of the loan;
- Number of months with arrears in the last 12 months;
- Application score ;
- Loan amount;
- Time of the loan until default.

Table 1 shows relative fits of the different approaches used with R^2 value. Note the R^2 values are not very high but such LGD values do seem difficult to predict. One of the advantages of the WOE approach was that the predicted values spanned 0 to 1, while in some of the other methods, because one was trying to estimate a skewed distribution with values mainly between 0 and 1, the predictions were often all within the 0.4 to 1 range.

Table 1 Comparison of the methods

Method	R^2
Box Cox	0.1299
Linear regression	0.1337
Beta distribution	0.0832
Log Normal transformation	0.1347
WOE approach	0.2274

Summary

The paper discusses a way of modelling LGD for unsecured consumer loans by modelling the collection process. There is obviously some overlap between these models and collection scoring models. But the LGD models need to look at the whole collections strategy not just what to do initially, since for example failed attempts at

collection will diminish the price of the debt if it is subsequently sold off. We suggest that the decisions by the lenders and the repayment risks of the debtors are both vital to model as part of this system. Hence we feel "decision tree" approach is the ideal methodology to use. We have shown how the repayment models may be built by recognising that the resultant distribution is a mixture distribution. This suggests one should use a two stage process to obtain estimates. Firstly use logistic regression (or cumulative logistic regression if more than two underlying distributions) to estimate which class a debtor is in and then a regression type approach for each class (ours is based on weight of evidence) to estimate the LGD values of debtors in that class. In the case study we presented, the collection's policy was straightforward – only in house collections, but for other cases we would need to identify what were the rules that decided on whether to collect in house, via agents or sell off the debt. One would also need to model the sale prices and agency commission as functions of the debt portfolio being sold or managed. Thus there is considerable development still needed in this area, but we do believe this approach addresses the fundamental difficulty in such LGD modelling of separating out the lender's recovery policy from the debtors' willingness and ability to repay the debt.

References

1. Altman E. I., Resti A., Sironi A.: Analyzing and Explaining Default Recovery Rates A Report Submitted to The International Swaps & Derivatives Association, December 2001
2. Andrade F., Thomas L.: Structural models in consumer credit, July 2004
3. Claessens S., Krahnen J., Lang W.W.: The Basel II Reform and Retail Credit Markets
4. Frye J.: Loss Given Default and Economic Capital, Capital and Market Risk Insights, 2/2004
5. Kordichev A., Powel J., Tripe D.: Structural model of revolving consumer credit risk, 2005
6. Lucas A.: Basel II Problem Solving, 2006,
7. McNabb H., Wynn A.: Principles and Practice of Consumer Credit Risk Management, 2000