“Collections policy comparison in LGD modelling”

LC. Thomas, A. Matuszyk, A. Moore,

Abstract

This paper discusses the similarities and the differences in the collection process between in house and 3rd Party collection. The objective is to show that although the same type of modelling approach to estimating Loss Given Default (LGD) can be used in both cases the details will be significantly different. In particular the form of the LGD distribution suggests one needs to split the distribution in different easy in the two cases as well as using different variables. The comparisons are made use two data sets of the collections outcomes from two sets of unsecured consumer defaulters.

Keywords: *credit risk, collection process, LGD modelling*

1. Introduction

When a borrower defaults on a loan some of the debt will be recovered during the subsequent collections process. Loss Given Default (LGD) is the percentage of the exposure at default which it is not possible to recover during this collections and recovery process. Modelling LGD has come to prominence in the last few years because under the internal ratings based regulations of the Basel Accord, (BCBS 2005) on capital adequacy, lenders have to estimate LGD for each segment of their loan portfolio.

There is a literature on LGD modelling for corporate loans, mainly because LGD is a vital factor in the pricing of risky bonds. The main approach to estimating LGD in this case is to use linear or non-linear regression based on a number of factors. These include details of the loan, such as the priority of the bond, details of the borrower, particularly the geographic and industry sector that the firm is part of, and the economic conditions. The book edited by Altman et al (Altman et al 2005) gives details of some of the models developed, though it is worth noting how difficult such estimation seems to be as indicated by the low R^2 values of many of the regressions. One example of a non linear regression is the commercial product

LossCalc (Gupton and Stein 2005) which is based on the fact that the LGD distribution should be approximated by a Beta distribution,

The literature for unsecured consumer credit is much sparser and it is only with the advent of the new Basel Accord in 2007 that there has been a concentrated attempt by practitioners and academics to model LGD for this type of debt. Earlier Makuch et al (Makuch et al 1992) has used linear programming to determine the nest allocation of resources in a collections department, but did not use this to estimate LGD. Thomas et al (2007) pointed out that one of the problems with LGD modelling for unsecured credit is that the outcome depends both on the ability and the willingness of the debtor to repay but also on the decisions by the lender. They used a decision tree approach to model the strategic level decisions of a lender of whether to collect in house, through an agent or to sell off the debt to a third party. They also suggested that LGD estimates for one type of collection might be built using mixture distributions. Caselli et al (Caseeli et al 2008) used data from an Italian banks in house collection process to show that economic effects are important in LGD values. Bellotti and Crook (Bellotti and Crook 2009) also looked at using economic variables as well as loan and borrower characteristics in a regression approach to LGD for in house collection while Somers and Whittaker (Somers and Whitaker 2007) suggested using quantile regression to estimate LGD , but in all case the resultant models had R^2 values between 0.05 and 0.2. It seems estimating LGD is a difficult problem.

This paper concentrates on the fact that recovering unsecured consumer debt is a sequential process with different parties being involved in seeking to recover the debt. Usually the first attempt to recover the debt is by the collections department of the lender (the "in house" process). If this is not proving worthwhile, or for other commercial reasons, such as not wanting the lender's reputation to be affected by it bringing court actions against debtors, the lender can use agents to collect the debt on a commission basis – i.e. they keep x% of what is recovered. Alternatively, or sometimes after using agents, the debt can be sold to third parties for a small fraction of the debt. This paper investigates the differences in the debt characteristics between the debt which is being collected in house and that which is being collected by a third party. Although the general approach to modelling LGD can be applied to both forms of collection, the differences in debt characteristics lead to differences in both the form of the model and the types of characteristics used to estimate LGD.

2. Data description

Normally the first attempt at collections is undertaken by an in house team belonging to the lender. Such a team will have the information the debtor supplied on application, all the details of the loan and the borrower's repayment performance until default. Although the formal Basel definition in the UK for default is that the debtor is 180 days overdue (unlike most other countries which is 90 days overdue) most lenders will freeze the loan or credit card facilities and undertake recovery measures once the loan is 90 days overdue. The representative data set we used for modelling such "in house" collections was provided by a UK financial institution. It consisted of 11,000 defaulted consumer loans which defaulted over a two year period in the 1990s together with their repayment performance in the collection process. We concentrated only on this performance in the first two years in collections to match the information that was available on the third party collections process.

The lender can also decide to use a third party to try and collect the defaulted amounts usually on a percentage fee basis so the third party will keep x% of what is collected. Alternatively or sometimes after using agents, the lender can sell the debt to a third party who then has the right to seek recovery of the outstanding debt. Our second data set consisted of such loans which had been purchased by a third party from several of the UK banks. This data set consisted of the information on 70,000 loans where the outstanding debts varied from £10 to £40,000. These debts were purchased in 2000 and 2001 and so most of the defaults had occurred in the late 1990s. The repayments of the debtors for the first 24 months in this "third party" collections process were available at an individual loan level.

It is clear when examining the "third party" data that there is less information available on the debtor than was available to the in house collectors. The details of the debt, including the amount outstanding, when default occurred and when last there was a payment was available. Also in order to set the purchase price, the history of how many different parties had sort to collect the debt is reported. There was some information available about the debtor including details of address and telephone numbers when available, and some demographic information. However there was little information on the default risk scores of the borrower- either application score or behavioural score- or on the borrower's performance before default. Thus in comparing the data we have restricted it to the details that were available both in the "in house and in the "third party" data sets.

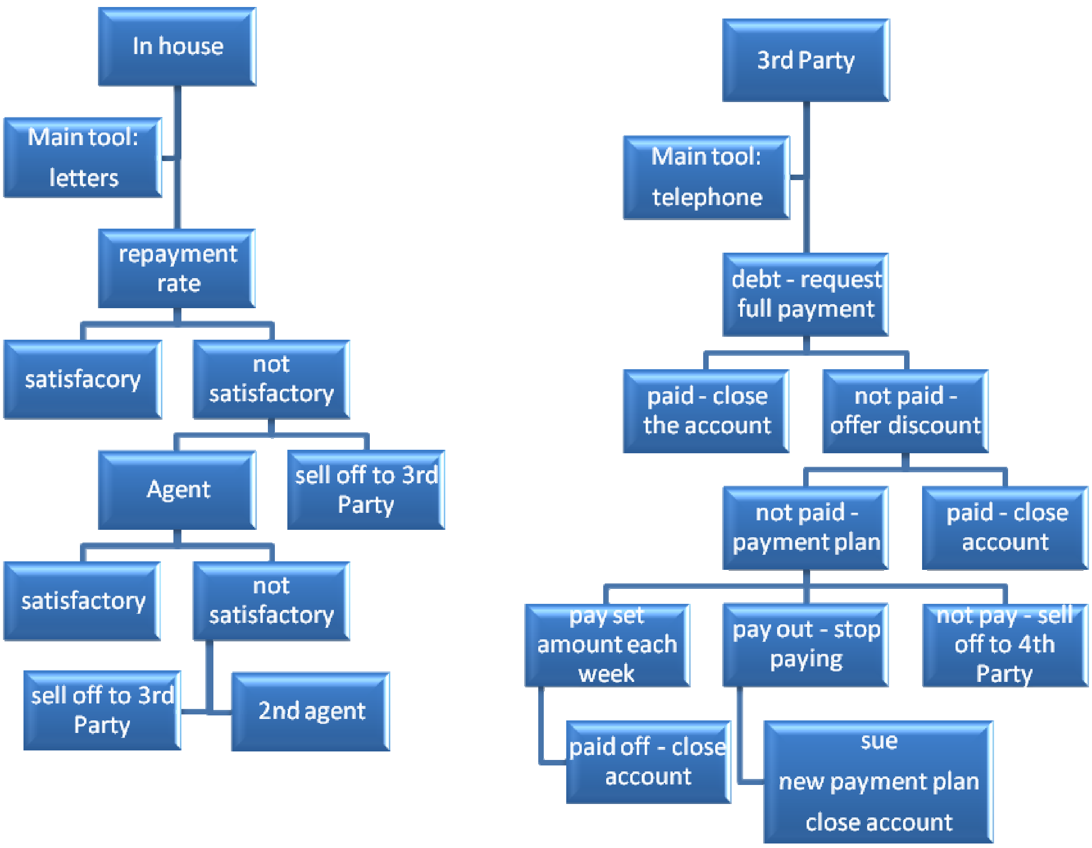
3. Collection strategies

The information available to the in-house collection department is different to that data available to the 3rd Party which later has the reflection on the ways the debt is collected. That is why we can distinguish the following sequences of events:

1. Recovery process – internal collection tries to save person
2. Collection process – internal collection tries to save money
3. Collection process – 3rd Party tries to save money

The main tool used in the in house recovery process is letter. There are different types of letters and sending them depends on the status of the customers and the characteristics of the debt. The debt sold to the 3rd Party will normally be debt, which has proven hard for the lender to collect in house. Since this is the case the distribution for LGD shows that the majority of the debt has not been paid. In fact over 80% of the 3rd Party's debts have had no payments made on them at all.

Figure 1: Collection trees



Decision which action to take in in-house collection, is made on the basis of different conditions. Ususally, the first step is to send the letters at the beginning of every month. There are diffferent types of letters and sending them depends in which areears the customers are. If this method is not sufficient the company must use other possible methods : calling the client, paying the visit for the client, trying to set up an agreement and find possible solution like rearranging the mortgage, selling the property etc.

When either a 3rd Party or in-house collections department takes over an account, they have to decide how to collect the debt. Their first step will be to always collect the full outstanding debt. If debtor pays then they close the account. If not then a discount is offered for a lump sum payment. If the debt is paid then the account is closed, otherwise the payment plan is set up (most likely outcome). If the full amount is paid at x£ per week the account is closed. If the customer pays and stops then the lender will have to decide to either close the account if the total amount paid is satisfactory. If it is not – they may try to sue or start up new payment plan. If they don't pay the payment plan at all then they will either sell the debt or close the account. The primary method of debt collection, used by the 3rd Party from which the data was acquired, is telephone with written communication in support. The telephone is used because it can lead to fast recovery of debt, as it is a direct line of communication with the debtor and can result in a payment from the first conversation. The telephone is also very cost effective compared to face-to-face communication but is just as personal. There is also the element of surprise and the debtor and collector can negotiate to achieve a mutually satisfactory result.

Table 1: Debt comparison

Factor	In house data set	3rd Party data set
Main tool	Letter	Telephone
Age of Debt	New	Old
Type of Debt	Unsecured	Unsecured
Average Debt Amount	£3,609	£562
Percentage Who Paid Back Whole Debt	30%	0.7%
Percentage Who Paid Back Part of the Debt	60%	16.3%
Percentage Who Paid Nothing	10%	83%
Mean value of LGD	0.544	0.95
Collection model	Decision tree model with sub-models	Agent's sub-model
LGD model	2-step model	2-step model
Information available	All details of loan and customer	Restricted data since not original lender

4. Distribution of LGD

Analyzing the distribution of LGD, in Figure 1, it can be seen that 30% of the debtors paid in full and so had LGD=0. Less than 10% paid off nothing. For some debtors the LGD value was greater than 1 since fees and legal costs had been added. This is not the case usually in 3rd Party collection where almost 90% of the population have LGD=1 (Figure 2). 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 2: Distribution of LGD in the sample for in house collection (collection for 24months: January1991-dec 1992)

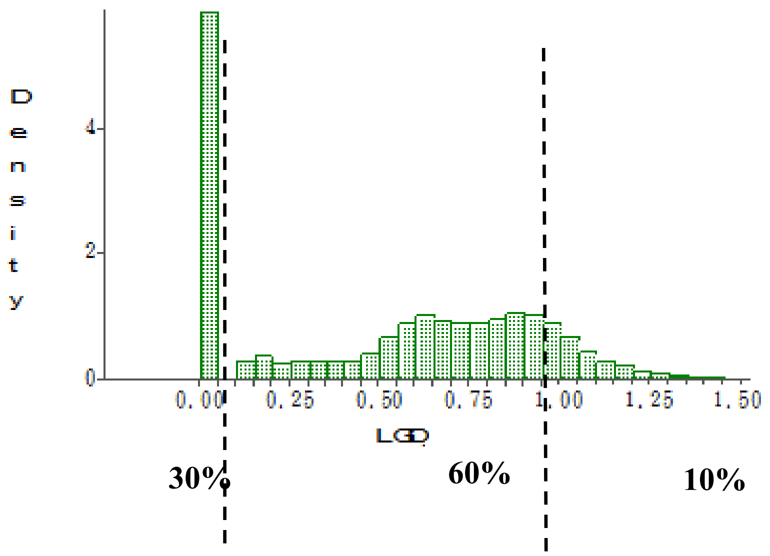
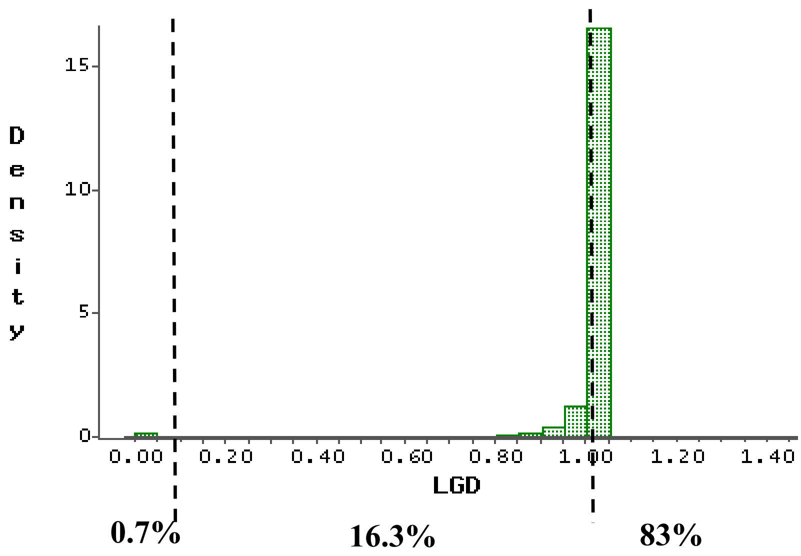


Figure 3: Distribution of LGD for credit card debt sold to a 3rd Party



↑↓Figure 3 shows the Loss Given Default (LGD) for the credit card debt collected by the 3rd Party. The x-axis shows the LGD, the column above 1 represents the number of debtors who failed to pay back any of their debt hence LGD=1. The column above 0.95 represents all of the debtors who paid back up to 5% of their debt (0.95<=LGD<1). The column above 0 represents all of the debtors who paid bake more than 95% of their debt (0<=LGD<0.05). The y-axis shows the number of debtors within each LGD bracket. The majority of the debtors (83%) failed to pay back any of their debt.

The recovery rates or loss given default for the two samples are very different. The majority of loans collected in-house have an LGD < 1, where as the loans collected by the 3rd Party have LGD = 1. There are several factors contributing to this difference. Firstly the debt collected in-house is new debt, no one else has previously tried to collect the debt and they have only recently defaulted at the time of collecting. On the other hand the 3rd Party debt is most likely old and has been collected before, this makes it harder for the 3rd Party to collect further. Secondly the in-house collection department will have access to more data and that data will have more details. This means that they can look at past behaviour, the original loan details in some cases they may also have access to data connected with their bank account and income. The 3rd Party will not have any of this data, in some cases the debtor may even need to be traced because they have moved or are deliberately trying to hide from the debt collections 3rd Party so that they cannot collect the debt.

5. Analysis of the common variables

The variables available for analysis and common in both data sets are as follows: age, amount of debt and residential status[1].

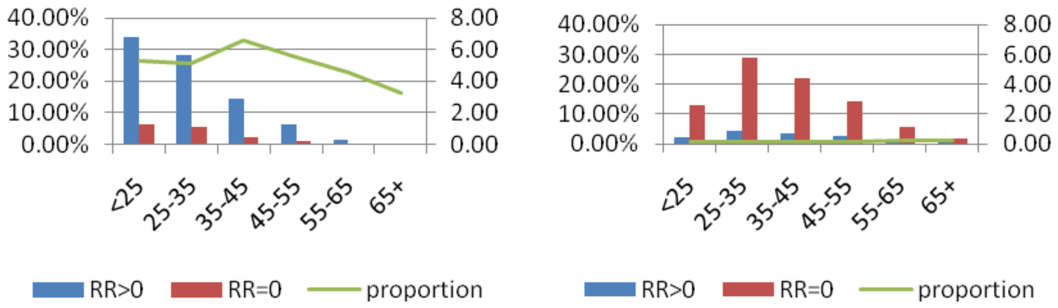
a) **Age**

Majority of debtors from in house data set, are in the "<25" and "25-35" brackets, minority in "65+" one. Most of the customers from 3rd Party data set are in the "25-35" and "35-45" brackets. In the 3rd Party case, the trend of proportion is rather stable, slightly increasing for the last two buckets, where in the in-house case, the higher RR is in group 35-45, then the older the debtor the lower the proportion.

[1] Because of the in house data set distribution, we took the following assumption: if RR<=0, then RR=0.

Figure 4: RR distribution by age for in-house collection and 3rd Party collection

In house   3rd Party



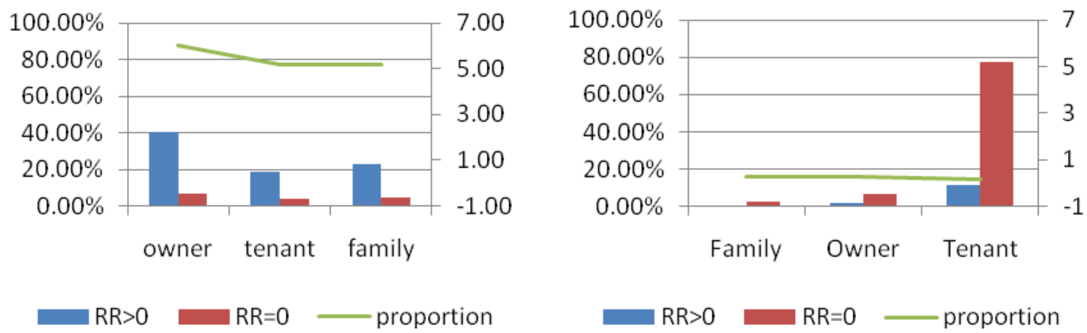
*) proportion=RR>0/RR=0

b) Residential status

Homeownership is divided into the following classifications: family, owner, joint ownership tenant and other. If the debtor is known to reside in a property owned by a member of their family, but not themselves or live with parents, then their homeownership is classified as Family. If the debtor resides in a property owned solely by them then their homeownership status is Owner. Joint status is recorded if the debtor and another own their residence and Tenant status if they are renting and finally, Other if the details are unknown. The vast majority of the debtors in 3rd Party data set are recorded as Tenants, over 85%. In the in house data set, majority of the clients have the Owner status (40%). This can also explain the behaviour of customers. Owners are more likely to pay off the debt where tenants belong to the most risky group.

Figure 5: RR distribution by homeownership for 3rd Party collection

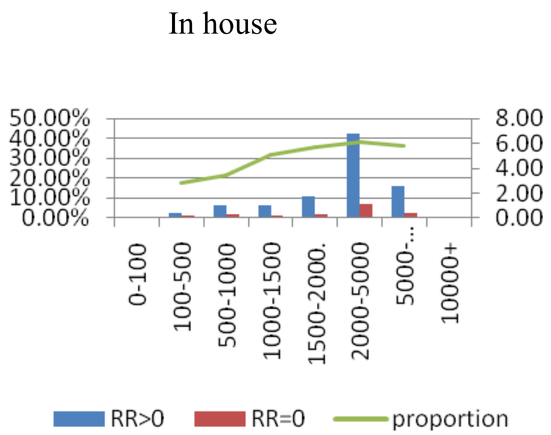
In house   3rd Party

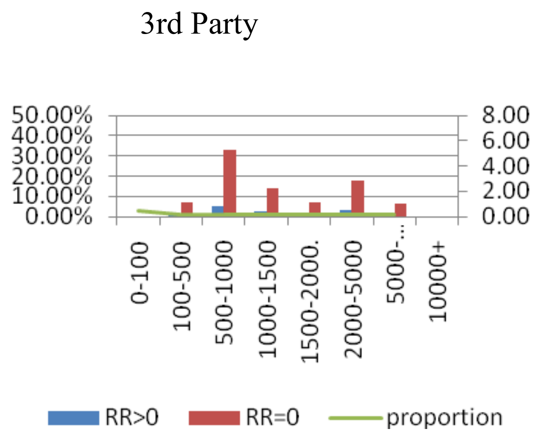


c) Debt Amount

The amount of the debt was from few pounds to 50000£. The variable was divided into eight groups. What is surprising clients behave in a slightly different way in both data sets. For in-house collection the recovery rate is growing with the amount of debt, in case of 3rd Party the trend is stable with the only exception for the first bucket (0-100£) where the repayment rate is the highest.

Figure 6: RR distribution by debt amount for in-house collection





This analysis demonstrates that some debtor properties like their age, debt amount and residential status have a clear effect on the recovery rate.

6. LGD models

For both data sets, models built consisted of two steps. In the first one step we tried to estimate the spike in the distributions. So for in-house we were concerned with LGD: LGD≤0 and LGD>0 and LGD=1 or LGD<1 for 3rd Party collection. The splits were necessary in case of the shapes of LGD (Figures 6 a and b). Logistic regression models were built for both data sets to split them into two groups. The predicted value for those in the first class should be either LGD=0 (In house) Or LGD=1 (3rd Party). For those who paid back part of their debt, the LGD was estimated using a number of different variants of linear regression. These included using ordinary linear regression, applying Beta and log normal transformations to the data before applying regression, the Box-Cox (Box, Cox, 1964) approach to "normalising" the data and using linear regression with weight of evidence (WOE) approach.

Figure 6: LGD models

a. In house

b. 3rd Party

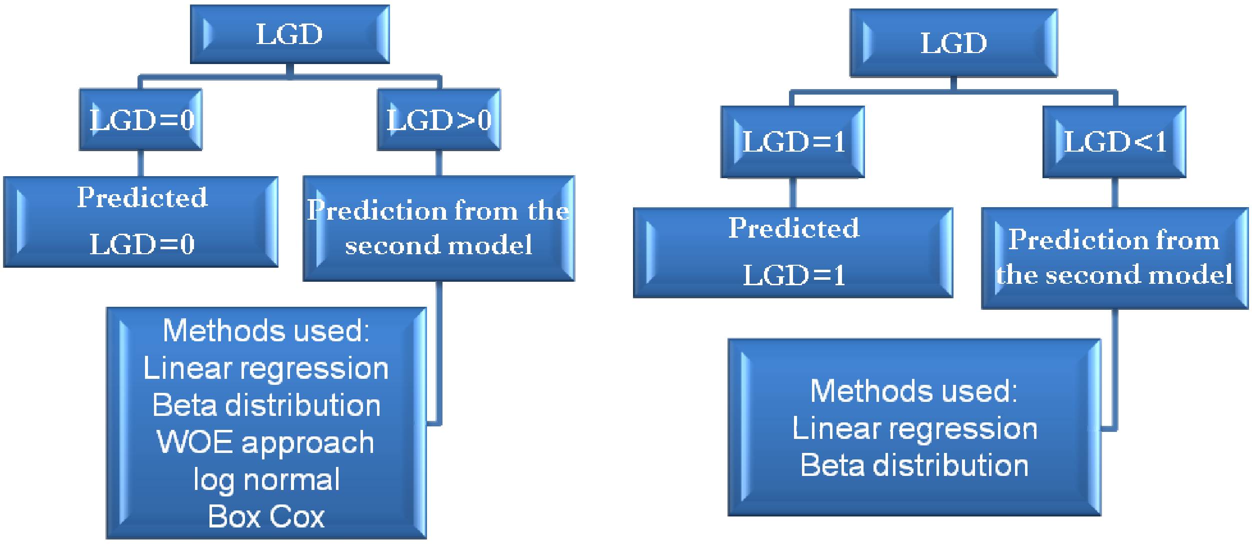


Table 2: Variables and results from modelling LGD

In house	**3rd Party**
1st stage	
LGD=0 versus LGD>0	**LGD=1 versus LGD<1**
The higher the ***loan amount*** the lower the chance of paying off everything The longer the ***lifetime of the loan*** the higher the chance of paying off everything The higher the ***application score*** the higher the chance of paying off everything The more ***time spent in arrears*** during the loan the higher the chance of paying off everything. However those who were in arrears for more than 2/3 of the time, had lower chance of paying off everything The more the customer was ***in arrears recently*** (in the last 12 months) the higher the chance of paying off everything	Having a ***work telephone*** number increases the likelihood of paying back part of the debt Having a ***mobile telephone*** number increases the likelihood of paying back part of the debt Having more ***telephone*** number increases the likelihood of paying back part of the debt Owing ***less than £100*** at default increases the likelihood of paying back part of the debt.

2nd stage predicting: 0<LGD<1	
LGD>0	**LGD<1**
The higher the ***loan amount*** the higher the expected loss	The younger the ***debtor's age*** the lower the expected loss
The higher the ***application score*** the lower the expected loss	The lower the ***default amount*** owed the expected loss
The longer the ***lifetime of the loan*** the lower the expected loss	***Owners*** will have lower expected loss Having a ***mobile*** decreases the expected loss
The more the customer was ***in arrears recently*** (in the last 12 months) the lower the expected loss	Not having a ***contact number*** decreases the expected loss
The more ***time spent in arrears*** during the loan the lower the expected loss	

Table 2 contains the variables and results achieved during the LGD modelling for both data sets. As can be seen, different variables were used because of the information available. In-house collections have more data available to them because they have access to the original loan details and behaviour variables from monitoring the loan throughout its lifetime. Whereas the 3rd Party is limited to information given by the lender. This information is limited due to lender policy and lack of requirements on the lender to provide useful debtor information.

Stage one for in-house and 3rd Party is focused on different extreme LGD results. For in-house we were concerned with paying off the whole loan whereas for 3rd Party we were concerned with not paying of any of the loan because these were the spikes in the LGD distributions. The in-house model found that the higher the *loan amount* the lower the chance of paying off everything and the 3rd Party model found that the higher the *loan amount* the lower the chance of paying of part of the debt. The rest of the In-house model was based on behaviour and application variables chich were unavalible to the 3rd Party. Therefore the 3rd Party model's variables were more facused on haw to contact the debor therefore the telephone number avaible.

The second stage model is focused on predicting the LGD between 0 and 1and trying to fit a distribution. Different methods were tried (see table y), the best method in in-house was weight of evidence with an R^2 of 0.23 and the best method for 3rd Party was beta with R^2 of 0.12.

Table 3 shows the fits of the different approaches used in both data sets with R^2 value. It can be noticed that R^2 values are not very different and in both cases not very high. These results suggest that LGD values seem difficult to forecast. All of the models for 3rd Party and in-house except weight of evidence gave a narrow distribution focused around the mean. Only weight of evidence gave a distribution covering the whole range 0-1 for which the LGD observed results covered.

The variables used by the in-house model and the 3rd Party model are again very different due to the information available. The in-house collections were privy to application and behaviour variables whereas the 3rd Party were limited to personal variables and contact information. Yet despite these different variables and the greater information held in-house the results of the models are very similar. Both the linear regression and the beta distribution models gave R^2 values around 0.1, where the predicted results were a poor representation of the observed results since in all cases the predictions were clustered around the means.

Table 3: Comparison of the results for the 2nd stage models

Method	In-house R^2	3rd Party R^2
Box Cox	0.1299	
Linear regression	**0.1337**	**0.1097**
Beta distribution	**0.0832**	**0.1161**
Log Normal transformation	0.1347	
WOE approach	**0.2274**	**0.1496**

In the WOE approach we defined the target variable - LGD to be above or below the mean. Then for each used characteristic, we split them into ten groups and looked at the ratio of above mean to below mean in each group and combined adjacent groups with similar odds, so as to divide the values of each characteristic into a number of “bins”. Then we defined WOE modifications for each characteristic which took the weight of evidence value for each bin that the corresponding variable had been classed into. Generally, 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:

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

These previous models were all focused on predicting the final LGD, but when looking at wither to sell the debt or collect in-house; it might be useful to predict what will happen over shorter time periods. The next model is a simple linear regression based on what was collected in the first 12 months in-house to see what would happen in the second 12 months. These models estimate the recovery rate (RR) at 24 months and 36 months after default; RR_{24} and RR_{36} respectively.

$$RR_{24}=0.047+1.205RR_{12}$$

This model had an $R^2=0.58$ and a Root MSE=0.13. Expanding the model to see what would happen in the 3rd year gave an $R^2=0.80$ and a Root MSE=0.11:

$$RR_{36}=0.037-0.258RR_{12}+1.233RR_{24}$$

Using the above models a lender can make more informed decisions about when to sell and how much to sell for. The reason these results are so superior to the previous models is because there is a dependence on both sides of the equation. RR_{24} and RR_{36} are dependent upon RR_{12} since they cannot be smaller than RR_{12} by definition. This artificially inflates the R^2 results.

Conclusions

Although both analysed data sets are about debt recovery, the information available to each party is quite different and the success rate for recovering the loan is quite difficult. Despite the fact, that two stage model is appropriate for both even though the spikes are different at LGD values (opposite ends). This is not surprising because 3rd Party debt will usually go through several collection processes, so by definition must be harder to collect.

These models can be used by both sites to determine the price at which to buy a debt. The 3rd Party model gives an indication of recovery rate so the 3rd Party can set an internal upper limit for the price of buying the debt.

For the in-house collection; the question is how much more would they get by keeping debt in their collection process for some further time? To get a feel for this one needs to estimate RR in the next year using their information and current recovery. The models above which estimate RR_{24} and RR_{36} could help the in-house lender set a minimum price at which to sell the debt and determine which debts to sell and which to continue with. However internal politics and procedures are more likely to determine when to sell of the debt.

References

1. Altman, E. I., Resti, A., Sironi, A.(2005): Recovery Risk, Risk Books, London
2. Basel Committee on Banking Supervision (BCBS), (2005): International Convergence of Capital Measurements & Capital Standards- A Revised Framework
3. Bellotti, T., and Crook, J., (2009): Calculating LGD for credit cards, QFRMC Conference on Risk Management in the Personal Financial Services Sector
4. Box, G.E, Cox, D.R., (1964): An analysis of transformation, J. Royal Statistical Society Series B, 26, 211-246
5. Caselli, S., Gatti, S., Querci, F., (2008): The Sensitivity of the Loss Given Default Rate to Systemic Risk: New Empirical Evidence on Bank Loans
6. Gupton, G.M., Stein, R.M., (2005), LossCalc v2; Dynamic Prediction of LGD, Moody KMV,
7. Makuch, W.M., Dodge, J. L., Ecker, J.G., Granfors, D. C., Hahn, G. J. (1992): Managing Consumer Credit Delinquency in the US Economy: A Multi-Billion Dollar Management Science Application
8. Somers, M., Whittaker, J., (2007): Quantile Regression for Modelling Distributions of Profit and Loss
9. Thomas, LC, Mues, C., Matuszyk, A., (2007): Modelling LGD for unsecured personal loans: Decision tree approach", CORMSIS WP 07/07, School of Management, University of Southampton, to appear in J. Operational Research Society