European Journal of Operational Research 249 (2016) 476–486

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Modelling repayment patterns in the collections process for unsecured consumer debt: A case study

Lyn C. Thomas [a,*], Anna Matuszyk [b], Mee Chi So [a], Christophe Mues [a], Angela Moore [a]

[a] Southampton Business School, University of Southampton, Southampton SO17 1BJ, United Kingdom
[b] Warsaw School of Economics, Al. Niepodleglosci 162, 02–554 Warsaw, Poland

ARTICLE INFO

Article history:
Received 14 January 2014
Accepted 8 September 2015
Available online 26 September 2015

Keywords:
OR in banking
Payment patterns
Collection process
Markov chain models
Survival analysis models

ABSTRACT

One approach to modelling Loss Given Default (LGD), the percentage of the defaulted amount of a loan that a lender will eventually lose is to model the collections process. This is particularly relevant for unsecured consumer loans where LGD depends both on a defaulter's ability and willingness to repay and the lender's collection strategy. When repaying such defaulted loans, defaulters tend to oscillate between repayment sequences where the borrower is repaying every period and non-repayment sequences where the borrower is not repaying in any period. This paper develops two models – one a Markov chain approach and the other a hazard rate approach to model such payment patterns of debtors. It also looks at simplifications of the models where one assumes that after a few repayment and non-repayment sequences the parameters of the model are fixed for the remaining payment and non-payment sequences. One advantage of these approaches is that they show the impact of different write-off strategies. The models are applied to a real case study and the LGD for that portfolio is calculated under different write-off strategies and compared with the actual LGD results.

1. Introduction

There are two major reasons to model the collections process for the recovery of defaulted consumer debt. First the regulations, incorporated in Basel II (BCBS, 2004) and Basel III (BCBS, 2011), on the risk capital that banks must hold required banks to estimate Loss Given Default (LGD) for each segment of their loan portfolio. LGD is the percentage of the debt at default that is still not collected at the end of the collection process. Basel Accord II (BCBS, 2004) suggests three ways of modelling LGD: historical average, regression approaches and modelling the recovery process. For consumer debt, the historic average does not make much sense and the regression approaches lead to poor results with models in the literature having *R*-squared between 0.05 and 0.22. One reason for these poor results is the non-normal form of the LGD distributions but another significant reason is that LGD depends partially on the debtor's capacity and willingness to repay but also on the collection strategy. The models in this paper allow incorporation of the lender/collector's write-off strategy, which materially affects the resultant LGD. They also allow lenders to identify which among a set of write-off strategies will be most profitable over the whole debt portfolio. This is a second reason for modelling the collections process since lowering LGD affects who should get credit in the first place and at what price.

Default is defined as borrowers being 90 days overdue or there is evidence to the lender that the borrowers will not repay. Default triggers the collections process as the lender seeks to recover the debt. Most collections processes measure their success by the Recovery Rate (RR) they achieve, where $RR = 1–LGD$.

The recovery rate depends not only on the debtors' capacity and willingness to repay but also on the lenders' actions and their collection policy. Previous models have ignored the lenders' influence in their models. One such collection action is to write off the loan and make no further attempt to collect. Writing-off is determined by the collector's expectation of future recoveries and the effort in collecting them. Such trade-offs can be used to determine whether the future expected recovery amount including recovery costs would be positive. Currently collectors make such write-off decisions subjectively and are often swayed by end of the quarter financial objectives or the pressure on the collections process. There is currently little modelling support for such actions. As well as estimating Recovery Rate (RR), the models presented here support collectors in assessing this trade-off between recoveries and the effort involved. This trade-off is influenced by the way debtors have already been repaying their debts, the cost of the collection process, and the likely

* Corresponding author. Tel.: +44 (0)23 8059 7718; fax: +44 (0)23 8059 3844.
E-mail address: l.thomas@soton.ac.uk (L.C. Thomas).

http://dx.doi.org/10.1016/j.ejor.2015.09.013
0377-2217/

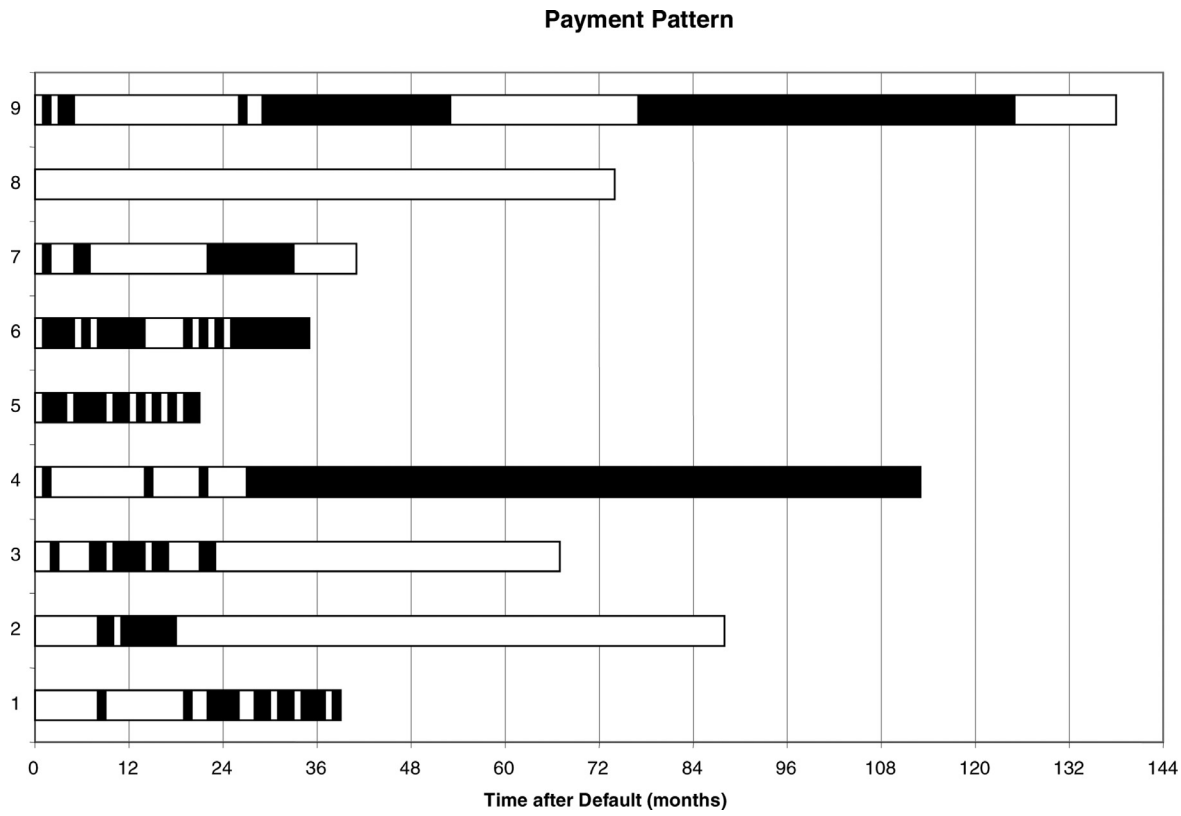

Fig. 1. State space description of the payment sequences for 9 different defaulters. (Note: Black when payment occurs; white when no payment).

duration until the debt will be repaid. The models allow collectors to have some data driven indication of which write-off policies are most profitable.

Provided the debtor is contactable, collections start with an agreement for the debtor to repay a fixed amount every period or to pay off the debt in one payment. What subsequently occurs is that there is an initial sequence of periods of non-payment while the agreement is put into place, followed by a sequence of periods of payment. This may stop and then a sequence of non-payment periods occurs until repayment restarts again. This can be repeated several times throughout the collection process as some of the real data examples in Fig. 1 show. Alternatively the debt may be "cured" in that the repayments made cover the defaulted amount. In this paper we take "cured" to mean the debt is fully repaid, but a minor adjustment of the models would allow "cured" to mean a satisfactory percentage of the debt is repaid or a sufficient number of repayments has been made.

This paper introduces two modelling approaches to describe these patterns of repayment and non-repayment. The first is a payment sequence approach which looks at the movements at sequence level between a sequence of payments and a sequence of non-payments. The second is a survival analysis approach, which looks at whether there is a repayment or no repayment in each time period (usually a month). It models how many payments are made in a sequence until the debtor stops paying and how many missed periods occur before they start paying again. Using the average repayment rate per sequence for the first approach and the average repayment rate per period in the second approach, one can calculate the distribution of the overall repayment rate. The models are appropriate for portfolio level decisions and overall LGD rates. To estimate LGD for an individual, one needs to extend the models so the parameters are functions of the individual debtor's characteristics.

These approaches allow one to calculate the repayment rate under different write-off strategies as well as the average duration a debtor is in the collection process. This would allow the lender to decide on a suitable trade-off between the future recovery rate and the amount of future effort expended to reach that rate under the different write-off strategies. The results are relevant at the portfolio level since they involve the average recovery rate and the average extra effort involved. The models are not intended to identify the optimal write-off strategy but can be seen as a progress to optimising such decisions.

The next section gives an informal description of the data from the case study on which the models will be built. This is the type of data that collectors are now recording on a regular basis. Section 3 discusses the literature on collections processes as well as the use of Markov chain models in consumer lending. Section 4 describes the sequence based Markov chain model where the debtor moves between payment and non-payment sequences. Section 5 applies this model to the case study data to estimate recovery rate, and hence LGD, under simple write-off strategies. Section 6 describes a hazard rate model of the collections process. This involves more estimation than the sequence based model but allows much more complex write-off strategies. In both cases, a full model is outlined together with simplifications of the model which require fewer parameters as they assume that after an initial period the parameters of subsequent payment (and non-payment sequences) are the same. The models in Section 6 are applied to the case study data in Section 7. Finally conclusions are drawn from the models and their results.

2. Description of the collections data set

The data we use in this case study describes the repayment history of 10,000 defaulted personal loans from a UK bank's loan book. These are loans that defaulted between 1988 and 1999 where default was defined as 90 days in arrears. The performance of the loans in the collection process was recorded from the start of 1988 until the end of 2003. The collections policy of the lender was to agree where possible with the debtor an amount that should be repaid each month until the debt was fully paid off. The data recorded whether there was a