OPTIMIZING THE COLLECTIONS PROCESS IN CONSUMER CREDIT

Adiel T. De Almeida Filho[1], Christophe Mues[2] and Lyn C Thomas[2]

(1) Federal University of Pernambuco, UFPE

Cx. Postal 7462, CEP 50.630-970, Recife-PE, Brazil.

E-mail: atalmeidafilho@yahoo.com.br

(2) Quantitative Financial Risk Management Centre

School of Management

University of Southampton

Southampton, UK

E-mail: c.mues@soton.ac.uk

E-mail: l.thomas@soton.ac.uk

OPTIMIZING THE COLLECTIONS PROCESS IN CONSUMER CREDIT

ABSTRACT

This paper builds a dynamic programming model to optimize the collections process in consumer credit. It determines which collections actions should be undertaken and how long they should be performed. Theoretical results about the form of the optimal policy under certain conditions are obtained. Finally a case study is described based on data from the collection department of a European bank.

KEYWORDS:

Collection Process, Optimization, Operations Management, Dynamic Programming, Consumer Credit.

1. Introduction

Consumer credit has been little researched in banking and finance certainly when compared with the level of research directed at corporate credit. In particular there has been no modelling undertaken to improve the management of the operations involved in recovering some or all of the debt incurred when a borrower defaults. Yet the US sub prime mortgage crisis and its knock on effects has made bankers and ordinary citizens realise how important is consumer credit. One important issue which has arisen in consumer credit within the last few years, because of changes in regulation and the increase in the default rate on credit cards to above 5%, is how to maximize the amount recovered from consumer debts. This leads into how best to manage the recoveries and collection process, particularly for unsecured credit since the process in secured credit can be modeled by a two stage process – what is the chance the lender will have to take possession of the security and how much will the lender get for the security.

This paper seeks to model the collections process of unsecured consumer credit debt. In this process the collections department of the lender has a number of actions it can take to secure some repayment of the debt. These can range from telephone calls and gentle reminder letters to more formal letters, getting agreement to rescheduled repayment patterns with some repayment being made immediately. If this fails the collectors occasionally can make home visits but more generally seek legal redress involving court proceedings and the use of bailiffs. The questions are which actions should be taken and how long should a particular course of action be undertaken before trying another action. These problem can be modeled in two different ways. In the first case, the decisions are made each period in the light of the individual debtor's repayment performance up to that point. Such decisions would be modeled using stochastic dynamic programming. In the second case, the decision made is which action to

perform next and how long to undertake it, and this depends only on the debtor's repayment performance on previous actions and assumes a deterministic "average" recovery profile under the current action. This gives rise to a deterministic dynamic programme. This has the advantage that you can easily calculate what are the optimal actions that will be applied to the "average" debtor and what is the average cash flow of recoveries that will follow from these actions. These are exactly the forecasts those who have to manage a collections process need to make prospectively when they are deciding the resources needed in the collections department and reporting to those who have to deal with provisioning (setting aside money to cover lender expected losses in the future) how much of the bad debt they expect to recover. In the former case they want to know how many staff they will require and how many will need familiarity with the legal side of debt recoveries. So knowing which actions will be undertaken and how long on average they will last for the typical debtor makes such resource calculations possible. Similarly knowing the repayment pattern for the average debtor allows estimates of total cash flow from recoveries to be calculated. We concentrate on this second model in this paper.

There has been very little analytic modeling of the collections process for any form of lending until the advent of the new Basel Accord which came into operation in 2007. This changed the way regulators determined how much capital banks have to hold to cover against credit risk. It required banks to estimate for each segment of their loan portfolio three quantities- PD, the probability of default (what proportion of the portfolio will default), EAD the exposure at default (the amount of money that could be defaulted upon within the segment) and, LGD, the loss given default (the percentage of any default that is not recovered eventually), (Bennett et al 2005). The capital the banks have to set aside for the credit risk of this loan segment is then f(PD).EAD.LGD where f(.) is defined in the Basel regulations. LGD is closely related to the

recovery rate, RR, the percentage of the default amount that is recovered since LGD =1-RR. It is partly this regulatory emphasis on percentage of loss recovered that makes lenders' collections departments measure their performance in terms of recovery rates rather than total amount recovered. They also use recovery rate because it is a measure which has the same bounds (apparently 0 and 1) on all loans, and so it is easier to empirically estimate recovery rate distributions. Also, as most countries have limits on the amount of unsecured credit that can be offered, the default amounts of unsecured credit are not that different from one another. Empirically one can have recovery rates less than 0 if interest is charged on defaulted loans and nothing is repaid; recovery rates greater than 1 can occur if all that interest and the original default amount is paid off.

Previously there had been some work on estimating recovery rates in corporate lending since these affect the price of risky bonds. The edited book by Altman et al (Altman et al 2002) outlines, the mainly, regression based models that seek to relate recovery rates to economic factors and characteristics of the loan and the defaulter. The work on modeling the collections process for secured consumer lending (Lucas 2006) is directly motivated by Basel and so is more interested in estimating how much of the debt would be collected rather than optimizing the collections process.

For unsecured consumer credit, Matuszyk et al (Matuszyk et al 2007) have recognized that the recovery rate depends both on decisions by the lender as well as the uncertainty about the borrower's ability and intention to repay. They though look at models which support the strategic level decision of whether to collect the debt in house, use an agent or sell off the debt. Makuch et al (Makuch et al 1992) addressed bad consumer credit management for General

Electric Co in its GE Capital which provides credit card services, building a probabilistic account flow model to optimize resource allocation through linear programming.

Otherwise the only analysis tends to be on very specific issues within the process such as how one could use text mining of the recorded conversation between the collector and the defaulter to identify whether the defaulter is likely to repay (Chin and Kotak 2006). The books by McNab and Wynn (McNab and Wunn 2000) and Anderson (Anderson 2007) describe the process and the sorts of actions that can be undertaken but do not model the process. In other areas there have been some attempts to model an operations process so as to optimize the outcome. One of the nearest to the work here is (Yu and Gittins 2008). This looks at how many staff should be put in the different stages of a pharmaceutical Research and development operations. There the time of the different operations depends on the number of staff assigned to it. In our model the decisions are how long to run each different action for.

In section two we introduce the model, while in section three we prove some results concerning the form of the optimal recovery rate and the optimal collections process under certain special conditions. Section four describes the case study of applying this model to real collections data, while the final section draws some conclusions and indicates how one could develop more detailed models of the collections process. These would be aimed at customizing the actions for each particular debtor whereas in this model we assume we are dealing with a homogeneous population.

2. Collection Process Optimization Model

As we described in the previous section, any organization seeking to recover a borrower's defaulted debt has a number of actions it can pursue as part of the collections process. The aim

of the model introduced in this section is to aid the collector determine which actions to undertake and how long to undertake them in order to optimize the amount of the debt that is recovered. A debtor will remain in the collections process until either all the debt is cleared or the collections agency determine that it is not worth pursuing the debt any further and ceases all actions and writes off (charges off) the remaining debt.

Assume that the collections process has $i = 1,...,n$ actions which can be undertaken. We assume there is an obvious ordering of the actions so that one cannot undertake action i, if action $j, j > i$ has already been used on the debtor. This is because there is little point in reverting to "softer" actions once harsher ones have been applied. For example once court action has been commenced to recover the debt, there is no advantage in sending reminder letters - it confuses the debtor and the collector and could lead to legal complications. Time could be considered as continuous, but since most collection agencies review the repayment performance of the defaulter at regular intervals, usually monthly, and make decisions about whether to change or continue with the current collections actions at these review points, we will consider it to be discrete. So let $s = 0,1,2,..$ be the number of periods the current recovery action has been in operation on this debtor.

The aim of the whole collection process is to maximize the recovery rate. The definition of recovery rate can be extended to cover the recovery rate for each particular collections action as follows. Define $F_i(s)$ to be what is recovered by operating with action i for s periods as a fraction of the debt outstanding when action i was started. So $F_i(0) = 0, F_i(\infty) \leq 1$. There is a collection cost for each period action i is undertaken and we define c_i to be this cost expressed as a fraction of the original debt. We do not add this cost to the debt but we will subtract it from the total recovery rate to get the net recovery rate.

The objective is to find the collections strategy which maximizes the net recovery rate from each debtor, allowing for collections costs, where the cash recovered and the costs incurred are discounted by a factor β for each period into the future. We take the approach used by most reasonable lenders of freezing the amount of debt at default but recognizing that the subsequent cash flow of recoveries should be discounted to reflect the time value of money. Thus in almost all cases, the recovery rate will be less than 1. One can easily modify the model so that interest is charged on the unpaid debt, but the results obtained are very similar.

This problem can then be modeled as a dynamic programming problem. The decision epochs are the monthly reviews undertaken on each case to check what recoveries have occurred in the past month and what to do in the next period. We denote these review times as $T=\{1,2,...,\}$. $A = \{a_1,..,a_k\}$ is the set of actions which the lender can use in seeking to recover a debt with the ordering described previously so when using action i the lender cannot in future use actions j, $j<i$. The state space describes the situation at any time of an individual loan and consists of three components. These are the current action, i, being performed on that loan, the number of decision periods, s for which that action has been undertaken so far and r the recovery rate of the loan when this current action was started. With this definition the transitions in the process are deterministic. Either the lender keeps with the existing action and so moves to state $(r,i,s+1)$ or it moves to another action immediately and so instantaneously moves to state $(r+(1-r)F_i(s), j,0)$. Note that though we are modeling this problem deterministically and are interested in the optimal actions on the "average debtor" for resource purposes, the model gives a policy where the time in any action does depend on the recovery rate from that debtor from the previous actions used.

Define $V(r,i,s)$ to be the future net discounted recovery rate given that currently action i has been performed for the last s periods and r is the percentage of the original debt that had been recovered when action i commenced. The net discounted recovery rate allows both for the cost of collections and for discounting both costs and recoveries by β. The decision this period is whether to continue with action i or to move to a more severe action. Hence $V(r,i,s)$ satisfies the following optimality equation of a dynamic programming model (Puterman 1994; Bertsekas 2007; Denardo 1982)

$$V(r,i,s) = \max \begin{cases} (1-r)(F_i(s+1) - F_i(s)) - c_i + \beta \cdot V(r,i,s+1) \\ V(r+(1-r)F_i(s), i+1, 0) \end{cases} \qquad (1)$$

The first expression is the recovery if action i is continued for another period. The amount recovered in that period is $(1-r)(F_i(s+1) - F_i(s))$, since (1-r) is the fraction of the debt still to be recovered when action i starts. c_i is the cost of action i and the recovery rate in subsequent periods is $\beta V(r,i,s+1)$ as the collection process starts the next period in state $(r,i,s+1)$ and the subsequent recoveries need to be discounted by one period. The second expression is what happens if it is decided to start action i+1. The recovery rate at the start of the action is $r+(1-r)F_i(s)$ where the latter is what has been recovered from s periods of action i. Note that one does not need to consider more drastic actions than $i+1$ because in solving for the state (., i+1,0) one will consider moving immediately to action $i+2$. Repeating this argument means one allows the possibility of moving immediately to all states (., j,0), i.e. starting any action $j, j > i$.

To find the optimal overall recovery rate and the optimal collections strategy one must solve $V(0,1,0)$, i.e. start with nothing recovered and having performed action 1 for 0 periods. Computationally this can be solved using value iteration, outlined in the next section.

We have assumed a homogeneous population in the model but of course one could apply this at a subpopulation level and allow different types of subpopulations of debtors depending on their socio demographic characteristics, the type and amount of debt and their previous credit history. The case study we discuss later has two such subpopulations.

3. Properties of the Optimal Collections Policy

There are some obvious properties one might expect of the relationship between the recovery rate and the duration of a specific collections action. Trivially one should expect

A1: $F_i(s)$ is non decreasing in s

A more debatable assumption is

A2: $F_i(s)$ is concave in s (i.e., $\nabla F_i(s) = F_i(s+1) - F_i(s)$ is non increasing in s)

This is reasonable if i is a repetitive action like telephoning, where as time goes by the amount recovered from each telephone call gets less. It is the law of diminishing returns. The assumption may be less obvious for actions, which take some time to set up, like undertaking legal action for recovery of the debt. Even there one the repayment pattern is agreed the increasing possibility of the debtors reneging on the agreement means the assumption has some validity.

The standard way of solving dynamic programming is to use value iteration where the iterates $V_n(r,i,s)$ are defined by

$$V_{n+1}(r,i,s) = \max_i \begin{cases} (1-r)(F_i(s+1) - F_i(s)) - c_i + \beta \cdot V_n(r,i,s+1) \\ V_n(r + (1-r)F_i(s), i+1, 0) \end{cases} \qquad (2)$$

$$V_0(r,i,s) \equiv 0$$

$V_n(r,i,s)$ converges to the solution of the optimality equation and this gives a way of solving numerically to find the optimal collections policy and the optimal recovery rate. To carry out this computation one discretizes the r values (we took 100 values, i.e. every 0.01 and used interpolation to get estimates of the values in between), while i and s are already discrete. Results in Puterman(1994) shows that value iteration can solve problems with more than 1,000,000 states relatively easily, whereas our state space is in the low thousands. One can also use induction on the iterates of value iteration to prove results about the form of the value function and the optimal policy.

The first result explains how the optimal value function depends on the state descriptors r,i,s.

Lemma 1:

i) $V(r,i,s)$ is non increasing in r

ii) $V(r,i,0)$ is non increasing in i

iii) If A1 and A2 hold, then $V(r,i,s)$ is non increasing in s.

Proof

ii) follows trivially because from (1) $V(r,i,0) \geq V(r,i+1,0)$ The proofs of (i) and (iii) use induction on n the iterate of the value iteration procedure in equation (2). Trivially all three conditions hold at n=0 since $V_0(.,.,.) \equiv 0$

i) Assume i) holds for $V_n(r,i,s)$ then as r increases $(1-r)(F_i(s+1)-F_i(s))$ decreases and the induction assumption implies $V_n(r,i,s+1)$ is non increasing. Summing these two gives that the first expression on the RHS of (2) is non increasing in r . As r increases

$r + (1-r)F_i(s) = r(1-F_i(s)) + F_i(s)$ increases   and   so $V_n(r + (1-r)F_i(s), i+1, 0)$ is   non

increasing by the induction hypothesis. Hence $V_{n+1}(r,i,s)$ is non increasing in r and the

induction is proved.

iii) From A2 we have $(F_i(s+1) - F_i(s))$ is non increasing in s. $V_n(r,i,s)$ is non increasing in s

because of he induction hypothesis . A1 implies $F_i(s)$ is non decreasing in s and so i) shows

that $V_n(r + (1-r)F_i(s), i+1, 0)$ is non increasing in s. Hence all the terms on the RHS of (2) are

non increasing in s and the induction hypothesis holds. $Q.E.D.$

The optimal collections policy can be defined by a set of functions $s^*(r,i)$, i=1,2,..,n, $0 \leq r \leq 1$,

so that if the collection process starts action i , with the recovery rate at r, the action should be

continued for $s^*(r,i)$, periods, where

$$s^*(r,i) = \min\{s \mid V(r,i,s) = V(r + (1-r)F_i(s), i+1, 0)\}$$

Note that for some actions j, $s^*(r,j) = 0$ which implies that action j will not be used if the

recovery at the point at which it was considered for use is r .

Lemma 2:

i) If $(1-r)F_i(s) - sc_i \leq 0, \forall s \geq 0$, then

$$s^*(r,j) = 0, \forall j \geq i$$

Proof

Assume the contrary namely $s^*(r,j) = s\text{*} > 0$,then as $F_i(s) - sc_i \leq 0, \forall s \geq 0$

$$V(r,i,0)=(1-r)F_i(s^*)-s*c_i+\beta^{s^*}V(r+(1-r)F_i(s^*),i+1,0)\leq\beta^{s^*}V(r+(1-r)F_i(s^*),i+1,0)$$
$$<V(r+(1-r)F_i(s^*),i+1,0)\leq V(r,i+1,0)$$

,

where the first inequality comes from the assumption in (i) and the third inequality from Lemma 1 (i). This contradicts the definition of $V(r,i,0)$ in (1) and so $s^*(r,i)=0$. *Q.E.D.*

Note that if A2 holds and $F_i(1^*)\leq c_i/(1-r)$, then condition (i) will also hold.

Finally we prove the intuitive result that in the situation when $\beta=1,c_i=0\;\forall i$, one might as well run with each collection action for as long as possible before then moving onto the next hardest action, since there is no cost for operating any action and no penalty if the recovered amounts are very late in being paid.

Lemma 3: If $\beta=1,c_i=0\;\forall i$, $s^*(r,i)=\infty,$ i=1,2,..,n, $0\leq r\leq1$

Proof

In this case, we can write $V(r,i,s)=(1-r)W(i,s)$. Substituting this into equation (1) gives an optimality equation for $W(i,s)$, namely

$$V(r,i,s)=(1-r)W(i,s)=\max\begin{cases}(1-r)(F_i(s+1)-F_i(s))+(1-r)W(i,s+1)\\(1-r-(1-r)F_i(s))W(i+1,0)\end{cases}$$
$$\Rightarrow W(i,s)=\max\begin{cases}(F_i(s+1)-F_i(s))+W(i,s+1)\\(1-F_i(s))W(i+1,0)\end{cases}$$

(3)

Note that $0\leq W(i,s)\leq1$, since it is a recovery rate with no costs or discounts involved. To prove the theorem, it is enough to show that the first term in (3) is greater than the second term for all s. Now

$$F_i(s+1)-F_i(s)+W(i,s+1)\geq F_i(s+1)-F_i(s)+(1-F_i(s+1))W(i+1,0)$$
$$>W(i+1,0)-F_i(s)>(1-F_i(s))W(i+1,0)$$

where the first inequality holds because in state (i,s+1) choosing to start the i+1th action result in a recovery less than or equal to the optimal action in that state. The last two inequalities hold because $W(i+1,0) \leq 1$. Thus one keeps on using action i all the time and so $s^*(r,i) = \infty$, i=1,2,..,n, $0 \leq r \leq 1$.   Q.E.D.

4 Case Study

The model developed in the previous sections was applied to collections data from an European Bank. The sample consisted of 3084 consumer loans that had defaulted over a three year period. For each loan there were some details of the borrower, and loan details including when it was taken out, how much it was for, when the borrower defaulted and how much was owed at default. There were also monthly details (covering 150,658 loan/months in total) of the collection process including which collection action was applied to the defaulter in the month and how much was recovered in that month.

There were 5 different collections actions recorded but in fact only 3 actions were ever considered for each debtor. This was because the collections process segmented the debtors into two groups depending on whether the defaulted amount was a low percentage of the original loan (which we will classify as G, the Goods) or a high percentage of the original loan (which we classify as B, the Bads). The policy was to allow actions 1G, 2G and 3 to be applied to those in the Good group and action 1B, 2B and 3 were applied to those in the Bad group. Although not the same actions, 1G and 1B both concentrated on communicating with the defaulters and arranging repayment schedules. 2G and 2B are also different actions but both use legal procedures to recover the debt. The collection procedure has to begin with action 1G or 1B, because the regulations prevented collectors using legal procedures without first seeking

to agree a repayment schedule. Since the collectors' objective is to maximize the recovery rate, allowing for collection process expenses they would want to move to stronger actions in due course Action 3 is essentially passive in that the debt is kept on the books but no effort ($c_3 = 0$) is made to recover the debt. The legal position in that country is that someone with an outstanding debt is not allowed to take out further credit and therefore many debtors do pay off their debt unsolicited (sometimes several years after the debt occurred) when they want to obtain further credit. Note that although action 3 is the weakest action, it still satisfies the modeling assumption that having moved to it, and written off the loan, the collector will not move subsequently to actions 1 or 2. We solve the problem separately for the two segments-Good and Bad.

We plotted the relationship between recovery rate and maturity for each action case by case and from that felt that the exponential function

$$F_i(s) = A_i(1 - e^{-B_i s}), \quad A_i, B_i > 0 \tag{4}$$

fitted action 1 and 2 well. This exponential function also satisfied Assumptions A1 and A2, since $F(s+1) - F(s) = Ae^{-Bs}(1 - e^{-B})$ ↓ as s ↑. Coefficient A_i is the recovery rate if the action was allowed to continue indefinitely while B_i is related to the rate of recovery ($B_i = -\ln(0.5)/\bar{s}_i$ where $\bar{s}_i$ is the time until half the maximum possible recovery under this action will have occurred. The parameters A and B were estimated by minimising the square of the errors of the average amount collected in each period s, i.e. $F_i(s+1) - F_i(s)$.

Figures 1 shows the form of the cumulative recovery rate function for 1G and 2G where the estimate for 2G ignores how a debtor had performed on 1G. The estimated parameters for the Good segment were

$$F_{1G}(s) = 0.243(1 - e^{-0.147s}) \text{ with } R^2 = 0.9755 \text{ and}$$

$$F_{2G}(s) = 0.425(1 - e^{-0.0577s}) \text{ with } R^2 = 0.9924.$$
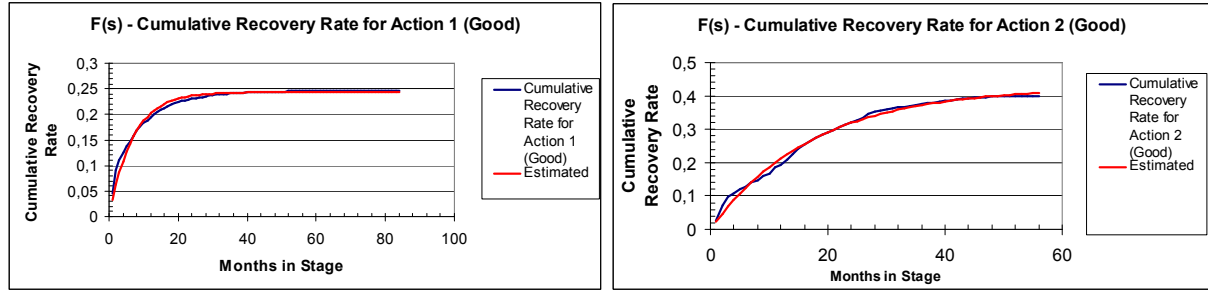
Figure 1 – Cumulative Recovery Rate for Action 1G and 2G (Good)

It is clear that in the long run action 2G collects more than 1G in that it would eventually recover 42% of the debt while 1G would only recover 24%. However 1G is quicker in recovering debt than 2G recovering 14.2% in the first 6 months compared with 12.4% for 2G recovered in the first 6 months.

In the Bad segment the cumulative recovery functions are similarly estimated. They are shown in Figure 2, where

$$F_{1B}(s) = 0.0188(1 - e^{-0.770s}) \text{ with } R^2 = 0.8073 \text{ and}$$

$$F_{2B}(s) = 0.252(1 - e^{-0.0681s}) \text{ with } R^2 = 0.9976.$$
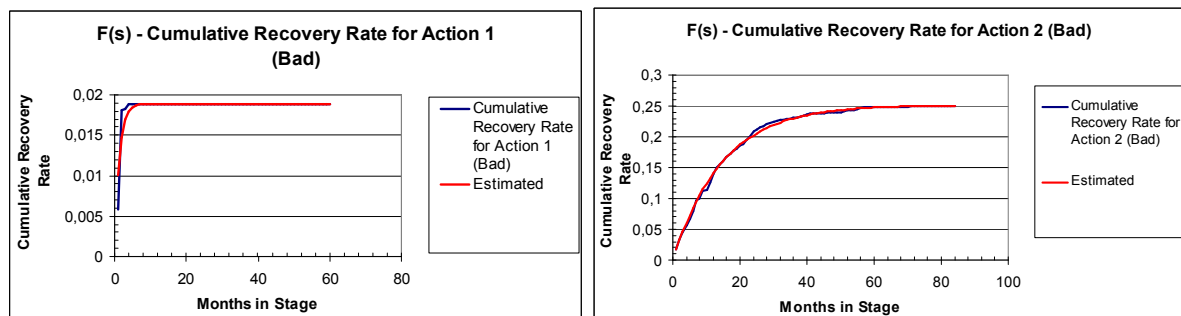
Figure 2 – Cumulative Recovery Rate for Action 1B and 2B (Bad)

In this case action 1B on average only recovers 1.8% of the debt even if it was used forever, whereas action 2B could recover 25% of what is owed if it was operated for ever. So really action 1B is extremely unsuccessful and 2B is far less successful than 2G, though that has more to do with the difference between the debtor types rather than any difference in the actions.
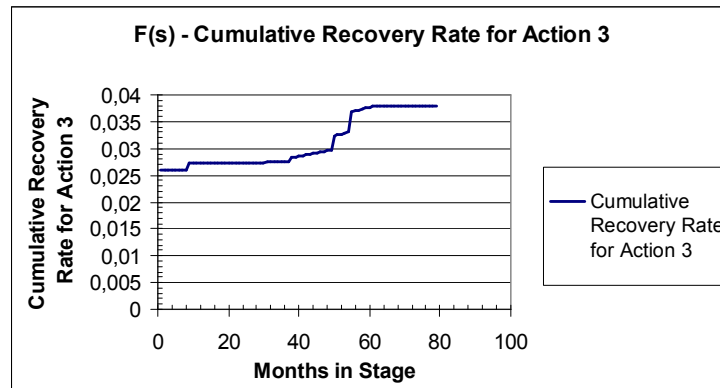

Figure 3 – Cumulative Recovery Rate for Action 3

Action 3 which involved no proactive collection procedures, was the only action where plotting the data did not give a good fit to the exponential function. Investigation also suggested only minute differences between the results for the two groups Good and Bad and so it is considered that the same action (or perhaps non action is a better description) is applied to both groups. As there are no costs in this action and no "harsher" action can be undertaken once a debt reaches this state, there is no point in ever stopping doing this action. Figure 3 shows the debtors' repayment behavior under action 3. We sought to fit it using a combination of linear functions and the exponential function which fitted the other actions so well. The result was a curve of the form

$$F_3(s) = \begin{cases} 0.0259 & 1 \le s \le 8 \\ 0.0272 & 8 < s \le 30 \\ 0.000009s + 0.0272 & 30 < s \le 37 \\ 0.0314(1 - e^{-0.05924s}) & 37 < s \end{cases}$$

In calculating the cost c_i of operating the various actions one has to use expert judgment. The monthly cost of each operation was determined having examined the whole collections budget and after discussions with the collections managers. c_i though is this cost expressed as a fraction of the debt outstanding on default and to do this we took the average balance at default of those who were in the Good group (for c_{1G}, c_{2G}), and in the Bad group (for c_{1B}, c_{2B}).This resulted in estimates of the costs of

$$c_{1G} = .00048656, \ c_{2G} = .00398703, \ c_{1B} = .00003769, \ c_{2B} = .00242082 \qquad (5)$$

There was data for up to 80 months in the collection process with up to 60 months available on individual actions. It was clear that almost all collection actions achieved no further recoveries after 48 months (some of them such as 1B well before). So it was decided to let each action takes place for a maximum period of 48 months, i.e. 4 years.

Looking at the case where the costs are given by equation (5), but keeping the discount factor $\beta = 1$, we get the results in Table 1 for the Bad segment of debtors and the results in table 2 for the Good segment. Of course for resource allocation purposes we are only interested in the case when the recovery rate before the first action is 0 in which case the second action starts after 8 months with an expected recovery rate of 0.0188 from the first action. We have added the results for how long the collectors should undertake the second action 2B if the recovery rate in the first action is not this average value for sensitivity analysis purposes. The column "Recovery rate after 1B applied" shows reference values of recovery before the start of action 2B, and the rest of the row described how long 2B should be undertaken and the effect it has on the net recovery rate. Table 2 describes the same results for the Good segment of debtors.

Table 1 – Optimal policy for Bad segment with Costs and No discounting

Recovery rate before 1B applied	Optimal duration of 1B	Recovery rate after 1B applied	Optimal duration of 2B	Recovery rate after 2B applied	Optimal duration of 3	Total recovery rate	Optimal total net recovery rate
0	**8**	**0.0188**	**29**	**0.231**	**∞**	**0.254**	**0.183**
Sensitivity analysis on recovery rate under action 1B							
		0	29	0.217	∞	0.24	0.17
		0.1	28	0.293	∞	0.314	0.246
		0.2	26	0.367	∞	0.386	0.323
		0.3	24	0.442	∞	0.458	0.4
		0.4	22	0.517	∞	0.531	0.478
		0.5	19	0.591	∞	0.603	0.557
		0.6	16	0.667	∞	0.677	0.638
		0.7	11	0.74	∞	0.747	0.721
		0.8	6	0.817	∞	0.822	0.808
		0.9	0	0.9	∞	0.903	0.903

The results are that for the Bad segment one can expect an optimal net recovery rate of 18.3% (including costs) by applying action 1B for 8 months followed by action 2B for 29 months before moving the debt into Action3 (the passive action) where it can stay indefinitely. 22% of the debt should be expected to be recovered by action 2B though. Figure 4 and 5 shows this optimal recovery rate as a function of the months into the collections process for the type of customers.

The results in the Good segment are quite different as Table 2 shows. One would expect to recover 51% of the debt eventually and the optimal policy is to operate with action 1G for 29

months followed by action 2G for another 27 months before moving the debt to the passive state. In this case almost equal amount of debt are recovered by the two types of actions.

Table2 – Optimal policy for Good segment with Costs and No discounting

Recovery rate before 1G applied	Optimal duration of 1G	Recovery rate after 1G applied	Optimal duration of 2G	Recovery rate after 2G applied	Optimal duration of 3	Optimal total recovery rate	Optimal net recovery rate
0	**29**	**0.240**	**27**	**0.495**	**∞**	**0.510**	**0.388**
		Sensitivity analysis on recovery rate under action 1G					
		0	32	0.358	∞	0.377	0.25
		0.1	30	0.415	∞	0.432	0.313
		0.2	28	0.473	∞	0.488	0.377
		0.3	26	0.531	∞	0.545	0.441
		0.4	23	0.588	∞	0.6	0.508
		0.5	20	0.646	∞	0.656	0.576
		0.6	16	0.703	∞	0.711	0.647
		0.7	11	0.76	∞	0.767	0.723
		0.8	4	0.818	∞	0.823	0.807
		0.9	0	0.9	∞	0.903	0.903

This policy could be compared to other, such as using action 1G for one year, followed by action 2G for another year and action 3 indefinitely. This policy would collect 39.0% of the debt instead of the 51.0% recovered using the optimal policy. The net recovery rate would be 33.6% using this policy and 38.8% using the optimal policy. Note that the optimal policy can recover 12% more of the debt than this other policy but half of the "extra" amount recovered is used up by the extra collection effort so the net improvement is just over 5% of the outstanding debt.
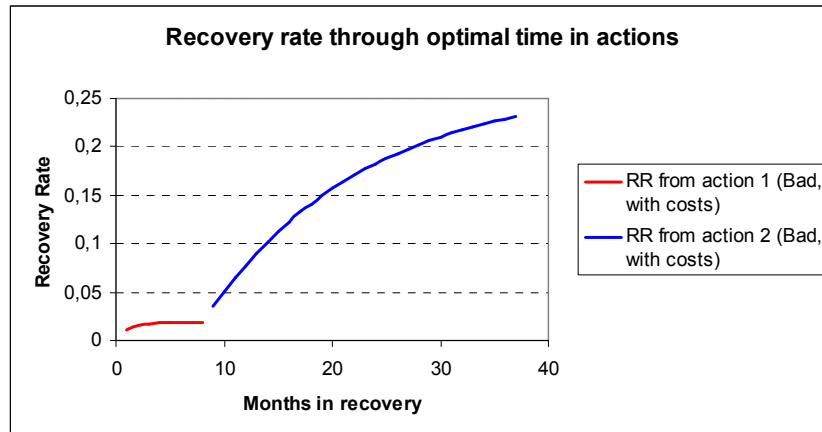
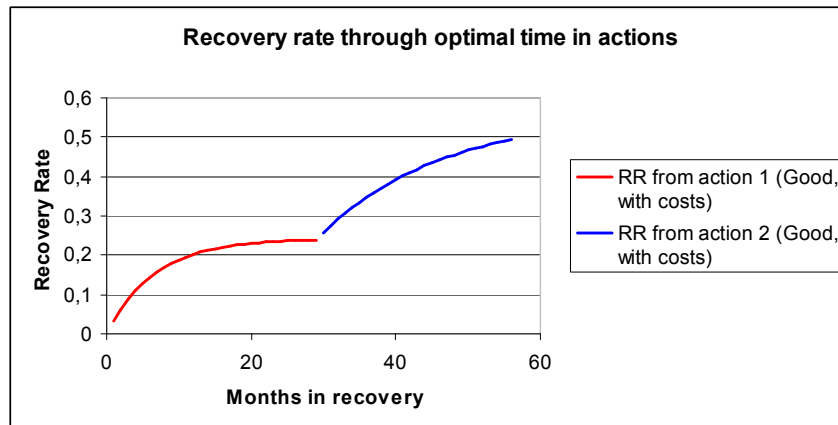*Figure 4 – Recovery rate through optimal time in actions 1B and 2B*

Figure 5 – Recovery rate through optimal time in actions 1G and 2G

Note that as the recovery rate from action 1 increases the time in action 2 drops. This is because there is less debt still to recover and so the operating costs are a larger proportion of the possible amount to recover. Of course the total net recovery rate increases as the amount recovered in action 1 increases but the amount of this that comes from the actual collection actions decreases. If a sufficient amount has been recovered it may not be worth while undertaking the second collection actions at all. This is the case for action 2B in the Bad

segment if at least 90% of the debt has been collected by action 1B. Similarly for the Good segment one would not use 2G if 84% of the debt had been recovered before its operation.

Table 3 – Optimal policy for Bad segment with Costs and discounting

Recovery rate before 1B applied	Optimal duration of 1B	Recovery rate after 1B applied	Optimal duration of 2B	Recovery rate after 2B applied	Optimal duration of 3	Optimal total recovery rate	Optimal total net discounted recovery rate
0	**5**	**0.0184**	**29**	**0.231**	**∞**	**0.254**	**0.149046**
Sensitivity analysis on recovery rate under action 1B							
		0	29	0.217	∞	0.24	0.139
		0.1	27	0.290	∞	0.311	0.221
		0.2	26	0.367	∞	0.386	0.302
		0.3	24	0.442	∞	0.458	0.383
		0.4	21	0.515	∞	0.529	0.466
		0.5	19	0.591	∞	0.603	0.548
		0.6	15	0.664	∞	0.674	0.632
		0.7	11	0.72	∞	0.747	0.718
		0.8	5	0.814	∞	0.820	0.807
		0.9	0	0.9	∞	0.903	0.903

If we introduce discounting with the monthly discount factor set at $\beta = 0.99746$ (corresponding to an annual inflation rate of 3.2%- the European average for the last five years), the results are very similar as Tables 3 and 4 show. The main difference is that the optimal recovery rates are slightly lower because of the time before one starts to recover some of the amounts. Also the durations of the first action drop a little because, especially in the Bad segment, one wants to

get more quickly to action 2B which is where the main recoveries occur so that the recoveries are not too heavily discounted.

Table 4 – Optimal policy for Good segment with Costs and discounting

Recovery rate before 1G applied	Optimal duration of 1G	Recovery rate after 1G applied	Optimal duration of 2G	Recovery rate after 2G applied	Optimal duration of 3	Optimal total recovery rate	Optimal total net discounted recovery rate
0	**25**	**0.237**	**27**	**0.493**	**∞**	**0.508**	**0.301016**
Sensitivity analysis on recovery rate under action 1G							
		0	32	0.358	∞	0.377	0.244
		0.1	30	0.415	∞	0.432	0.313
		0.2	28	0.473	∞	0.488	0.382
		0.3	26	0.531	∞	0.545	0.451
		0.4	23	0.587	∞	0.6	0.520
		0.5	20	0.646	∞	0.656	0.590
		0.6	16	0.702	∞	0.711	0.662
		0.7	11	0.76	∞	0.767	0.735
		0.8	4	0.817	∞	0.823	0.812
		0.9	0	0.9	∞	0.903	0.903

5 Conclusions

It is surprising that so little research has been undertaken into the operations management of the collections process in consumer credit, so that this is, we believe, the first model to seek to optimize aspects of the process. This may well be psychological in that banks want to "bury their mistakes" but it is also partly the accounting conventions in that when a bank has written

off all or part of a loan , senior managers are only concerned that the write off covers the loss and not to minimize the loss by optimizing the subsequent collections process. Moreover sometimes collections is contracted out or sold on to other debt recovery organizations and so the data on collections is not available. Even when collections is kept "in house", collections departments tended to register only the most basic information on the process.

The model introduced here is essentially a homogeneous one in that differences between debtors is only allowed for by segmentation and within a segment, the population is assumed to be homogeneous. This deterministic model is useful as a way of identifying what the cash flow from recoveries would be under the optimal collections process. Such information together with estimates of numbers of new debtors moving into collections each month is what is required for provisioning. The model is also useful in determining the allocation of resources within the collections department to obtain this optimal average recovery rate under the different actions.

One could expand the problem by seeking to tailor the collections process to the individual debtor, but one would then need to connect the cumulative recovery rate function for each action to the characteristics of the debtor. This will require experimentation by collections departments to acquire this data, but will be worthwhile doing once such data becomes available. One might also want to expand the model by including the information on the recoveries to date from an individual debtor from the current action being applied in deciding whether to continue that action or not. For example if someone is keeping to the agreed restructured repayment pattern, one would not want to change this; if someone has been repaying more than the average so far under a particular action one would want to keep with that action even if for the average debtor it is time to change to another approach. This is like

adjusting the way one operates with normal borrowers (increasing their credit limit for example) in the light of movements in their behavioural score. Since it has taken fifty years for this to happen with credit scoring, namely operating decisions based on behavioural scores, it should not be expected to happen overnight with the collections equivalent problem. The need is to acquire collections data in this level of detail.

Another extension will be to move from the deterministic way of modeling recoveries repayment to modeling it as a stochastic process. Part of this would mean checking what impact the previous recovery actions have on the efficacy of the current action. Our model assumes that they have none but that one can never return to a previous action. Checking whether the recovery process is really "memoryless" in this sense and that returning to previous actions is pointless would be part of developing such stochastic models.

Data is the key. Without it no useful model of the collections process can be built. Some collectors are investigating the use of text mining to identify the key phrases in the recorded conversations between debtor and collector which will categorise those who will subsequently repay the debt and those who will not. Most collections departments though are nowhere near this level of sophistication and a useful first step would be to collect and keep the data which was needed in this case study – the records of what collections actions were performed on the debtor that month and how much of the debt the debtor repaid in that month. With that collections departments would be in a position to improve the operations of their processes by models akin to that developed here.

REFERENCES

Altman, E., Resti, A., Sironi, A. 2005. Recovery Risk. Risk Books, London.

Anderson, R. 2007. The Credit Scoring Toolkit Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford University Press, Oxford.

Bennett, R.L., Catarineu, E., Moral G. 2005. Loss Given Default Validation, Studies on the Validation of Internal Rating Systems. Working Paper 14, Basel Committee on Banking Supervision, Basel, pp60-76.

Bertsekas, D. P. 2007. Dynamic Programming and Optimal Control. Athena Scientific, Vol. I, 3rd edition.

Chin, A.G., Kotak, H. 2006. Improving the debt collection process using rule-based decision engines: a case study of Capital One. Int J. Information Management, 26, 81-88.

Denardo, E.V. 1982. Dynamic Programming. Prentice-Hall, Englewood Cliffs.

Lucas, A. 2006. Basel II problem solving,

http://www3.imperial.ac.uk/portal/pls/portallive/docs/1/7287866.PDF

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. Interfaces, 22 (1), 90-109.

Matuszyk, A., Mues, C., Thomas, L. C. 2007. Modelling LGD for unsecured personal loans; Decision Tree approach. Working Paper CORMSIS 07-07, School of Management, University of Southampton.

McNab, H., Wynn, A. 2000. Principles and Practice of Consumer Credit Risk Management. CIB Publishing, Canterbury.

Puterman, M. L. 1994. Markov Decision Processes. Wiley, New York.

Yu, J-Y, Gittins, J. C. 2008. Models and software for improving the profitability of pharmaceutical research. European journal of Operational Research (in Press)

Authors:

Adiel Teixeira de Almeida Filho

Federal University of Pernambuco – UFPE, Recife, PE, Brazil

E-mail: ataf@ufpe.br; atalmeidafilho@yahoo.com.br

Christophe Mues

University of Southampton, Southampton, UK

E-mail: C.Mues@soton.ac.uk

Lyn Thomas

University of Southampton, Southampton, UK

E-mail: L.Thomas@soton.ac.uk

(Cx. Postal 7462, Recife – PE, CEP: 50.630-970, Brazil)

Biography

Adiel T. de Almeida Filho is a lecturer at the Federal University of Pernambuco (UFPE) in Brazil since 2008. He received his PhD degree in production engineering from UFPE. His research interests include Decision Analysis, Multi Criteria Decision Aid, Credit Scoring and Behavior Scoring, Production Planning and Logistics.

Christophe Mues is a lecturer at the University of Southampton, prior to this appointment at the University of Southampton he was employed as a researcher and teaching assistant at K.U.Leuven (Belgium), where he obtained the degree of Doctor in Applied Economics in November 2002. In September, 2004, he joined the Information Systems group of the School

of Management. His research interests span areas such as knowledge engineering and business intelligence, where his is investigating e.g. the use of decision tables and (binary) decision diagram (BDD) techniques in a variety of problem contexts, most notably the verification and validation of knowledge-based systems, business rule modelling (cf. the 'Prologa' tool home page), and knowledge discovery and data mining. In addition, he has developed a strong interest in applying data mining techniques to financial risk management and credit scoring in particular. Further interests include: web analytics, information systems development methodologies and software metrics.

Lyn Thomas is a Professor of Management Science at the University of Southampton since 2000. Formerly he was Professor of Management Science at the University of Edinburgh, 1985-2000, and President of the Operational Research Society 1994 - 1995. He was elected a Fellow of the Royal Society of Edinburgh, 1992, and was awarded the Beale Medal of the Operational Research Society in 2008. His research interests areas include credit scoring and credit control, and use of management science techniques in financial and banking areas, applications of game theory and logistics, including inventory and replacement.