Modelling the purchase propensity: analysis of a revolving store card

By G. Andreeva[1], J. Ansell[1] and J.N. Crook[1]

[1] *Credit Research Centre, University of Edinburgh, UK*

Correspondence to: G.Andreeva, Credit Research Centre, University of Edinburgh, 50 George Square, Edinburgh, EH8 9JY, UK.
E-mail: G.Andreeva@ed.ac.uk

Abstract

We investigate the incremental roles of information which becomes available only after a revolving loan has been granted in explaining and predicting the time taken until the borrower makes a second purchase. Using data relating to a store card, granted around the time of first purchase and used in Belgium, we find that characteristics of a first purchase and remaining credit available for use enhance the explanatory and predictive power of application characteristics. The relationship differs between good and poor payers.

Keywords: credit scoring; survival analysis; risk; banking

Introduction

Historically credit scoring has concerned itself with assessing which individuals are a good risk and which are a poor risk, primarily focussing on the probability of default at any point within a given time period after receiving a loan. A number of authors now appreciate that a second aspect of the risk of default which is relevant to profitability is the time between the initial granting of credit and the time to default[1,2]. Some high risk applicants can generate a significant profit if they use the credit product actively and pay interest and charges for long enough before going into default. On the contrary, low risk applicants may pay the full balance every month, thus keeping the revenues from such ‘good’ accounts low.

This observation has led many to use techniques developed in survival analysis to predict the time to default [3-8]. It has been found that survival analysis, especially Cox Regression[9], compares well to logistic analysis in providing an ability to predict default, but it also gives insights into the time to default. This can be particularly helpful in determining the profitability of a client.

Only a small number of studies have used a survival analysis to predict the time to the next purchase of a product. Ansell et al.[10] considered the behaviour of clients of an insurance company. The aim was to characterise their behaviour to decide on the marketing strategy for the individuals. The approach was to segment the population into defined groups and then explore the behaviour of these segments. They found that generally the less sophisticated and younger segments tended to come back reasonably swiftly if they were to return. Hence once past a certain time they were unlikely to return.

Older and more sophisticated customers tended to continue to respond over a longer period and there was no obvious cut off point.

The objective of the current study is to develop a model which explains and predicts the time taken by the holder of a revolving credit product to make a second purchase. The analysis relates to a store card and its use in Belgium. The card is normally taken out at the time of a first purchase. This has two benefits for the store: a purchase is made and a relationship can develop. The credit relationship may also be attractive to the store and the lender. Clearly the store will be keen that the customer will return to make further purchases and this may be in the interest of the lender. The issues then are which individuals return and when do they return. Such knowledge will give the opportunity to plan a strategy to enhance the relationship and to gain mutual benefit from the relationship for the store, the lender and the customer.

In this study we are particularly interested in the explanatory and predictive power of information which becomes available after the card has been issued but which is received in time to make future predictions. Stepanova and Thomas[5] found that the importance of application and behavioural data in predicting time to default varied throughout the age of the loan. In the current study we find that the combination of application, purchase and behavioural characteristics predict time to second purchase well, with behavioural characteristics becoming most important over time. A slight caution has to be read into our analysis since it is based on a limited period.

A noteworthy finding is that the difference between the credit limit and the outstanding balance, therefore the available credit to spend, has a major effect on the time until a second purchase. We also find future defaulters are *less* likely to make use of the card than non-defaulters.

The structure of the paper is as follows. The next section presents the basic model. The following three sections present the effects of introducing information sequentially gained by a lender during and after the application process. Then we consider differences between good and poor payers. The subsequent section examines the predictive performance of the variables and the final section concludes.

Basic Model

One can consider the information available to predict a customer's behaviour as being revealed in a series of sequential stages. First there is the application for credit, which provides the application data. At this stage bureau data may also be available. At the next stage there is extra information in terms of the nature of the purchase and the type of agreement entered into. This provides further insight into the potential behaviour of the customer. Finally there is the customer's behaviour after he/she has been granted the credit. In this study our interest is focussed on further purchases and primarily the second purchase.

Each individual in the analysis applied for a storecard, has been given the card and has made the first purchase. The customer may then within the observed time either make a second purchase, or make no further purchases and/or default. Both 'Second Purchase' and 'Default' can be well defined. Our definition of default is '2 consecutive missed payments'. Figure 1 displays the behaviour of five typical customers. Customer A makes a purchase within the study period and then continues on potentially to make further purchases. Customer B defaults and does not continue. Customer C does not make a further purchase within the study period. Customer D defaults but then makes repayments and so can make a further purchase before the end of the study period. Customer E makes the second purchase and defaults afterwards. The simplest model would be to consider time to the second purchase ignoring the impact of default. In this formulation there is a single form of censoring by time.

FIGURE 1 HERE

Our research strategy is to model the hazard function using predictor variables which are available at each sequential stage of a customer's behaviour. First we consider as predictor variables only information available to the lender at the time of application. Second, the predictor variables are those available at the time the customer makes a first purchase, and thirdly the predictor variables are those available after the first purchase has been made.

The proportional hazards model that we use can be described as follows. Let T be the time until a customer makes a second purchase. The hazard function can then be defined as

$$h(t) = \frac{P(t \le T \le t + \Delta t \mid T \ge t)}{\Delta t}$$

That is, it is the instantaneous potential for an event – second purchase – to occur in the next instant of time, given that it has not previously occurred. The proportional hazard approach assumes

$$h(t) = e^{\boldsymbol{\beta}'\mathbf{x}} h_0(t),$$

where $\boldsymbol{\beta}$ is a vector of parameters to be estimated and h_o is a baseline hazard. As was shown by Cox (1972) the $\boldsymbol{\beta}$ vector can be estimated without knowledge of $h_o(t)$. Cases for which the event did not occur within the observation period are censored and the parameters can only be estimated for non-censored cases.

The data used relates to a store card which is used in 3 European countries, but our data relates to its use only in Belgium. After data cleaning 25792 observations were available for analysis. All card holders in the sample had successfully applied for a card within a 14 month period in the late 1990s. The observation period ranged from 12 to 25 months after the month of first purchase. The behaviour of some cardholders was not observed for the full 25 months because they defaulted or they were issued with a card between periods 2 and 14 or because they closed their account within the observation period. The composition of the sample between those whose observed purchase behaviour was censored and those for whom it was not is shown in Table 1. The predictor variables available at each of the three stages described above are shown in Table 2. Their values/levels were coarse-classified according to similarity in p_j where p_j denotes the probability that those cases within a coarse category, j, make a second purchase, and transformed into binary dummy variables, see Thomas et al. [11] Model parameters were estimated using a randomly selected training sample (18040 cases) and the predictive performances of the estimated models were tested on the remaining 7752 cases.

TABLES 1 AND 2 HERE

Initial Stage

The purpose of the analysis is to make predictions of behaviour to help a lender make profitable strategic decisions concerning the lender-customer relationship. A warning should be given about a potential problem of the consistency in decision-making over time. Since in credit scoring the analysis is based on historic data, the conditions at the time of model development may change by the time of model application. This is part of a general problem, discussed elsewhere [8,12].

At the initial stage the only information available to make predictions derives from the application variables and potentially bureau data. The results of applying stepwise proportional hazards using just the application data are shown in Tables 3 and 4, column 1. Table 3 reports the parameter estimates obtained from fitting the model to the training sample. Table 4 indicates the corresponding hazard ratios. Since a stepwise variable selection procedure is used, the table reports the variables that are significant at the 0.05 level.

TABLES 3 AND 4 HERE

The results are similar to standard credit scoring results with Marital status, Residency type, Employment status, Industrial sector and Time at address entering. The interpretation of the results seems unsurprising. Widows are more likely to make further purchases on the card than other marital groups, the hazard ratio for the Widowed category is 1.118 (see Table 4, column 1), which means that the hazard of making the second purchase is nearly 12% higher for widows than for other categories of Marital status. The self employed are less likely to purchase than other groups and those working part-time are more likely than others. One may say, though, that those renting are more likely to purchase on the card as well as those only recently at the address. These might be explained in terms of the individuals' financial context, i.e. these customers can be credit constrained.

The baseline survival curve in Figure 2 shows that 50% of the customers have made a second purchase within 7 months and about 30% have not made a second purchase after 25 months. These percentages will be affected by those customers who have defaulted and those who have not been observed over the whole period.

FIGURE 2 HERE

From a marketing view it would be helpful to know whether either of these two groups could be segmented. The former group is making use of the card; the latter might be deemed non-users.

Immediately After Purchase

After purchase more information becomes available. This information included the nature of the initial product, its price, the agreement type, date of payment and whether the customer took out insurance on either the card or credit. The result of fitting both application and purchase data together using a stepwise model are presented in Tables 3 and 4, column 2. This model accounts for considerably more variation than the previous model, see Table 5. The previous model accounted for an improvement of 99 in Log Likelihood (which measures the model fit) above the no-covariate model, compared with 1078 for the current model.

TABLE 5 HERE

It is notable that the variables previously included appear again with only minor modification in the parameter estimates. The main changes are the introduction of Age and Spouse's Age, and the downgrading of Time at Address. Obviously these variables are often found to be collinear and this may be the effect seen here. Younger individuals are more likely not to make a second purchase as are those without a spouse, and hence the effect of Time at Address is seen.

All the purchase variables enter the model. Not taking card insurance seems to be an indicator of a smaller likelihood of making a further purchase, whereas no credit insurance implies a greater chance of purchase. Type of product bought also seems to affect the potential for subsequent purchase with those buying computers less likely than those buying phones to make the second purchase. Those who purchase relatively low cost items initially are more likely to make a second purchase on the card. This may be explained in terms of their need for credit. This is reinforced when agreement type is considered. Those using budget plans are much more likely to make further purchases on the card, than those with other agreement types. The payment date has an effect on the likely purchase with those later in the month more likely to use the card again. It may be that payment date is related to the borrower's own salary date.

Beyond First Purchase

Amount to spend

The behaviour of the borrower after the card has been issued is the last piece of information to become available. Variables representing this behaviour vary with time and therefore there is a need to decide how they should be fitted into the model. The difference between the outstanding balance and the credit limit, which we call 'amount to spend', (ATS), when the storecard is issued, represents the credit availability to the customer when he/she first takes out the credit agreement. This will be known at the time of the first purchase, *f*. We denote this value of ATS as $ATS_{t=f}$. The results when using this variable are presented in Tables 3 and 4, column 3. Again there is an improvement in Log Likelihood by 1376 as compared to the no-covariate model (Table 5). Variables similar to those entering the initial model appear again though there are some slight modifications in parameter estimates. The major difference is that Age leaves and Time on Job enters, along with $ATS_{t=f}$. As expected the $ATS_{t=f}$ variables enter with those with the highest difference between outstandings and credit limit being the most likely to make the second purchase. The least likely are those with a negative difference (their outstanding balance is above their credit limit). Technically if ATS is zero, no second purchase should have been possible. Yet the data contained such cases. It is possible that the lender applied a 'shadow' credit limit which exceeded the variable labelled 'credit limit'. The latter may have been the value which the lender led the borrower to believe applied.

Whilst the ATS at application provides insight, it seems more likely that the individual will be influenced in second (and future) purchases by their *most recent* ATS. Hence the second model contemplated was to include the ATS for the period preceding the second purchase into the model as a time-varying variable, ATS_{t-1}. The model to be estimated then took the form:

$$h(t) = e^{\boldsymbol{\beta}'\mathbf{x} + \gamma z_{t-1}} h_0(t),$$

where z is time-varying ATS and **x** is a vector of static application variables.

The results are presented in Tables 3 and 4, column 4. The improvement in Log Likelihood over the no-covariate model is 2281 emphasising the importance of recent ATS over initial ATS. Marital status and credit insurance leave the model, and whilst other variables are affected by inclusion of $ATS_{t-1,}$ their coefficients change only slightly.

To check the character of the relationship with ATS, the model was fitted with ATS two periods before the purchase, $ATS_{t-2,}$ and both ATS one and two periods before the purchase. Table 6 compares the Log Likelihood statistics for the model with ATS lagged and not lagged, applied to the sample of cardholders that make the second purchase, starting from period 3.

The inclusion of both lagged and current ATS results in a quite notable increase in the amount of variation accounted for, and both variables are significant (Tables 3 and 4, column 5). This suggests that ATS cannot be used simply in the model as a first order markovian variable.

Investigating the dynamics of ATS

The importance of behavioural variables in describing the time to the second purchase called for closer investigation of the dynamics of ATS. A dynamic model was developed, where parameters were re-estimated for each period of time[5]. The model which was estimated can be written as

$$h^s(t) = e^{\boldsymbol{\beta}'(s)\mathbf{x} + \gamma(s) z_{t-1}} h_0^s(t),$$

where s indicates the number of time periods elapsed since the first purchase.

Since the number of people falling into the category of 'No amount to spend' was rapidly decreasing with time, it was decided to group this category with 'Over credit limit'. The results for the first six months are shown in Figure 3. The same value of application score was included for each period.

As can be seen from Figure 3 all groups, except the group 'No amount to spend' demonstrate a proportional effect at each time point. The 'No amount to spend' group includes those individuals who are at or beyond their credit limit. The figure suggests this group is less likely to make the second purchase over time which is as expected.

FIGURE 3 HERE

Incorporating additional behavioural information

Starting from period two, more behavioural information becomes available: delinquency status and repayments dynamics. Therefore additional information was included into the analysis. For each period of time three more variables were added: an indicator of 1 missed payment, an indicator of two or more missed payments (our definition of default), and the percentage of the outstanding balance that was repaid during the period preceding the purchase.

Figure 4 shows a slow decline in the impact of the application score. This is expected since the information becomes more historic. This highlights the difference between this investigation and previous work on defaults on fixed terms loans, (see Stepanova and Thomas[5]). The behavioural aspects become of greater importance over time.

Percent repaid becomes significant at period 4 but loses its significance temporarily at period 7. Delinquency status, both 1 and 2 periods, remains significant through all 10 periods. Period 10 was the last period for which a model was estimated because the risk set became too small for subsequent periods.

The model with time-dependent behavioural variables was fitted to the risk set of cardholders who make a second purchase at period 3 and beyond. The parameter estimates and hazard ratios are shown in Tables 3 and 4, column 6. It is interesting to note that percent repaid was not selected by the stepwise procedure. This reflects its unstable behaviour in the dynamic model where it was initially not significant and later became significant.

FIGURE 4 HERE

Good versus Bad

Given the results of the last section it is worth exploring whether there is a difference between those who default during the observation period after first purchase and those that do not. Figure 5 gives baseline survival curves for Goods and Bads. Goods have a higher chance of making the second purchase, as expected.

The sample was split into those who did not default in the observation period after first purchase (Good) and those who did default (Bad). The Bads were divided into Bads after second purchase, 1213 cases, and Bads before second purchase, 422 cases, (the lender did not close an account when two payments were missed). Separate models were applied to each of the samples using application, purchase and ATS variables. One should, however, treat the results with a degree of caution since, as usual, the number of defaulters is low. This will affect the robustness of estimation and of the significance tests of the variables. Since the groups were small the estimation was carried out using the training and hold-out samples combined.

The results are presented in Tables 3 and 4, columns 7 to 9. There are differences between the models arising from the samples. As might be expected the model for the Good is similar to that previously seen. For Bads though the new models appear with the entry of whether a person has a phone for both categories of Bads, whilst number of children appears for those that default before purchase. Also Residential status, Industrial sector, Card and Credit insurance do not appear in the models for the Bad sample. Hence there does seem to be clear differentiation between Good and Bad, and such information may be useful in credit control. It may suggest that spending pattern be taken into account when assessing risk.

Equally there seems to be a difference between the models for those who default before purchase and who default after. Again a slight warning is applicable that those who default before second purchase will be expected to have longer time to second purchase. This result emphasises that it would be inappropriate to treat the problem as one of competing risks.

Prediction

The predictive ability of some of the models was tested on hold-out samples. Models 1 to 4 (Tables 3 and 4, columns 1-4) were tested on a hold-out sample from the full risk set. The dynamic models were tested on separate hold-out samples corresponding to the period of time they were developed for. Thus the size of hold-outs was shrinking with time, and became too small for robust results after period 6. Table 7 summarises the results. The hold-out sample number indicates the time period, the number in brackets gives the size of the hold-out. The table entries report the area under the ROC curve, which provides a measure of classification accuracy, not dependant on any threshold or acceptance rate[13]. It corresponds to the Wilcoxon or Mann-Whitney or *U* statistic, which estimates the probability that a ranking of a randomly selected bad account will be less than or equal to a ranking of a randomly selected good account[14].

The results in Table 7 indicate that at the period of the first purchase the best prediction is obtained by including the Amount to Spend available right after the purchase. This may be helpful in identifying customers that are going to make the second purchase immediately in the next period. As time progresses, behavioural information gives a lot of additional predictive power. For example, the difference in the area under the ROC curve between each application, product and ATS model and, for the same time period, the same variables plus behavioural information, increases as we consider future time periods. Thus for holdout 2 the difference between AUROC for the model including application, product and all behavioural variables and AUROC for the model including application, product and initial ATS is (0.643 – 0.652)= -0.009, which means that adding behavioural variables at this stage does not improve the prediction, but makes the model less parsimonious. Whereas for holdout 6 the same difference is (0.648-0.586)=0.062, which means that a probability of randomly selecting a good account with a higher or equal ranking as compared to a randomly selected bad account, has increased by 6.2%. The best results can be achieved by incorporating all behavioural information dynamically. The model with the largest area under the ROC curve for each holdout is identified in bold. Four out of five of these occur when a dynamic model is used. This reinforces the view that over time the behavioural information becomes more important in determining actions.

Conclusion

This paper presents an exploration of customer behaviour in using a retail card. Modelling the behaviour of the customer will help to establish the customer's profitability and hence the benefit for the lender. The paper has demonstrated how the data which is available at various early stages of the life cycle of a card can be used. The results obtained are similar in a large part to those seen in using survival techniques for predicting default. There are differences that reflect the need for certain customers to use the card to enhance their spending power.

The new information at each stage does enhance the modelling with a greater volume of variation accounted at each stage. It is also notable that the coefficients are generally only slightly modified. Thus it is important that a lender builds subsequent models after the application model because the incorporation of subsequent data would enhance the predictive performance of the model.

The variables that seem to have a major effect are the contract type and ATS. In this paper it has been seen that ATS does not have a first order markovian property with the most recent value accounting for the whole effect of the variable. Whilst the customer requires a positive credit balance before repeatedly using the card, it is also important to know how the customer got to this stage and if they were generally in surplus before. It is important to note that, increasingly throughout the period, behavioural aspects have major impacts on the behaviour of a customer.

The use of the card also indicates the likely behaviour of the individual. Defaulters who may be credit constrained are less likely to use the card than non-defaulters. There seems to be differentiation between models for prediction for the Good and for the Bad. Hence again indicating different behaviour. For the Bads there are also differences between those that default before the second purchase and those that default after second purchase. Such difference may also help in identifying patterns so that action can be taken to protect the lender.

References

1 Hopper MA, Lewis EM (1992). Development and use of credit profit measures for account management. *IMA J Math Appl Business Ind* **4:** 3-17.

2 Leonard KJ (1997). Behavior Scores to Predict Profitability. In: *Proceedings of the Credit Scoring and Credit Control V conference,* Credit Research Centre: University of Edinburgh.

3 Narain B (1992). Survival analysis and the credit granting decision. In: Thomas LC, Crook JN, Edelman DB (eds). *Credit Scoring and Credit Control,* Oxford University Press**:** Oxford, 109-122.

4 Banasik J, Crook JN, Thomas LC (1999). Not if but when will borrowers default. *J Opl Res Soc* **50:** 1185-1190.

5 Stepanova M, Thomas LC (2001). PHAB scores: proportional hazards analysis behavioural scores. *J Opl Res Soc* **52:** 1007-1016.

6 Stepanova M, Thomas LC (2002). Survival analysis methods for personal loan data. *Opns Res* **50:** 277-289.

7 Stepanova M . Using survival analysis methods to build credit scoring models. PhD thesis, 2001, University of Southampton.

8 Hand DJ, Kelly MG (2001). Lookahead scorecards for new fixed term credit. *J Opl Res Soc* **52:** 989-996.

9 Cox DR, Oakes D (1984). Analysis of Survival Data. Chapman & Hall**:** London.

10 Ansell JI, Archibald T, Harrison T (2001). Lifestage, Lifestyle and the Impact on Client Propensity in Financial Services. *Working Paper* 01/3, Credit Research Centre: University of Edinburgh

11 Thomas LC, Edelman DB, Crook JN (2002). Credit scoring and its applications. Society for Industrial and Applied Mathematics**:** Philadelphia, PA.

12 Hand DJ (1998) Consumer credit and statistics, in *Statistics in Finance*, ed. Hand DJ and Jacka SD, Edward Arnold, London.

13 Bamber D (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* **12:** 387-415.

14 Hanley JA, McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143:** 29-36.

Table 1. Sample used in the analysis

Performance	Total	2-nd purchase	Censored	%Censored
Good	22708	16648	6060	26.69%
Bad	3084	1634	1450	47.02%
Total	25792	18282	7510	29.12%

Table 2. Variables used in the analysis

Application variables	
Home telephone	Time in employment
Residential status	Type of business (Manufacturing, banking, catering, etc.)
Marital status	Employer's phone
Occupation (Full-time, part-time, self-employed, etc.)	Spouse age
Age	Number of dependants
Time at address since 18 years old	
Initial Purchase Variables	
Product code	Card insurance
Product price	Credit insurance
Payment date	Contract type
Credit Behaviour Variables	
Difference between the outstanding balance and credit limit at period *i*	Delinquency status at period *i*
Percent of outstanding balance repaid at period *i*	

Table 3. Parameter estimates of proportional hazard models with stepwise selection

Variable	Model								
	Application	App+ Purchase	App+ Purchase+ $ATS_{t=f}$	App+ Purchase+ ATS $_{t-1}$	App+ Purchase+ + ATS $_{t-1}$ +ATS $_{t-2}$	From 2nd period- more variables	App+ Purchase+ ATS $_{t-1}$ – Goods	App+ Purchase+ ATS $_{t-1}$ - Bads after 2nd p	App+ Purchase+ ATS $_{t-1}$ – bads before 2nd p
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Phone number given								0.264	0.615
1 child									0.329
Widowed	0.111	0.118	0.135						
Renting room, parents	0.086	0.079	0.086		0.104	0.119	0.114		
Renting house/flat	0.104	0.084	0.096	0.081	0.137	0.161	0.142		
Retired		-0.078							
Part-time	0.105								
Self-employed	-0.226	-0.122	-0.102	-0.126	-0.133			-0.465	
Type of business – Unknown	0.139	0.089			0.121				
Type of business – 21	0.188	0.115	0.082	0.109	0.147	0.117	0.077		
Age : under 21		-0.129							
Age : 22-27						0.069			
No spouse		-0.077	-0.069		-0.059			-0.197	
Time at address : 6 months	0.061		0.062		0.119	0.136	0.086	-0.136	
Time at address : 6 mths – 1 yr	0.104	0.052	0.105	0.045	0.151	0.166	0.116		
Time on job: 1 yr						0.062			
Allowance			0.120	0.088		0.175	0.136		
No card insurance		-0.076	-0.078	-0.061	-0.107	-0.124	-0.083		
No credit insurance		0.121	0.088		0.128	0.128	0.106		
Pay date 01		-0.319	-0.314	-0.300			-0.328	-0.376	
Pay date 08		0.169	0.137	0.158	0.334	0.328	0.133		-0.407
Pay date 14,15		0.372	0.351	0.360	0.221	0.206	0.363	0.284	
Product type- computers		-0.118		-0.117	-0.217	-0.201		0.280	
Product type- TV					-0.061				
Product type- household1									0.364
Product type- household2									0.369
Product type- phones		0.056	0.047	0.055		0.073	0.090		
Price = 0		0.280		0.149	0.589	0.483			
Price < =10,000 BEF		0.630	0.208	0.487	0.706	0.695	0.261		0.341
10,000 < Price <= 16,000 BEF		0.487	0.085	0.367	0.529	0.471	0.110		
16,000 < Price <= 20,000 BEF		0.328		0.224	0.316	0.274			
20,000 < Price <= 40,000 BEF		0.205		0.153	0.246	0.227	0.058		
Contract type – budget		0.726	0.854	0.775	0.452	0.675	0.878	1.209	
ATS $_{t-1}$ – over credit limit			-1.095	-3.867	-4.607	-2.540	-1.107	-1.251	
ATS $_{t-1}$ = 0			-0.235	0.574	0.624	1.223	-0.171	-0.631	
ATS $_{t-1}$ <= 5000			-0.423	-0.318	-1.156	-0.451	-0.359	-0.554	
5000< ATS $_{t-1}$ < =10,000			-0.222		-0.315		-0.194	-0.305	
10,000< ATS $_{t-1}$ <= 20,000			-0.128				-0.140		0.282
ATS $_{t-2}$ – over credit limit					0.878				
ATS $_{t-2}$ = 0					0.956				
ATS $_{t-2}$ <= 5000					0.517				
5000< ATS $_{t-2}$ < =10,000					0.142				
1 missed payment						-4.855			
2 missed payments						-4.339			

Table 4. Hazard ratios of proportional hazard models with stepwise selection

Variable	Model								
	Applica-tion	App+ Purchase	App+ Purchase+ $ATS_{t=f}$	App+ Purchase+ ATS $_{t-1}$	App+ Purchase+ + ATS $_{t-1}$ +ATS $_{t-2}$	From 2nd period- more variables	App+ Purchase+ ATS $_{t-1}$ – Goods	App+ Purchase+ ATS $_{t-1}$ - Bads after 2nd p	App+ Purchase+ ATS $_{t-1}$ – bads before 2nd p
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Phone number given								1.302	1.850
1 child									1.390
Widowed	1.118	1.125	1.144						
Renting room, parents	1.090	1.082	1.090		1.110	1.127	1.120		
Renting house/flat	1.109	1.088	1.101	1.085	1.147	1.174	1.152		
Retired		0.925							
Part-time	1.111								
Self-employed	0.798	0.885	0.903	0.881	0.876			0.628	
Type of business – Unknown	1.149	1.093			1.129				
Type of business – 21	1.206	1.122	1.086	1.115	1.158	1.124	1.080		
Age : under 21		0.879							
Age : 22-27						1.071			
No spouse		0.926	0.934		0.942			0.821	
Time at address : 6 months	1.063		1.063		1.126	1.146	1.090	0.873	
Time at address : 6 mths – 1 yr	1.110	1.054	1.111	1.046	1.163	1.181	1.123		
Time on job: 1 yr						1.064			
Allowance			1.128	1.092		1.191	1.146		
No card insurance		0.926	0.925	0.941	0.899	0.883	0.921		
No credit insurance		1.129	1.092		1.136	1.136	1.111		
Pay date 01		0.727	0.731	0.741			0.721	0.687	
Pay date 08		1.184	1.146	1.171	1.397	1.389	1.142		0.666
Pay date 14,15		1.450	1.421	1.433	1.247	1.229	1.438	1.328	
Product type- computers		0.888		0.890	0.805	0.818		1.324	
Product type- TV					0.941				
Product type- household1									1.440
Product type- household2									1.446
Product type- phones		1.058	1.048	1.057		1.076	1.094		
Price = 0		1.324		1.161	1.802	1.621			
Price < =10,000 BEF		1.878	1.231	1.628	2.026	2.003	1.298		1.407
10,000 < Price <= 16,000 BEF		1.627	1.089	1.444	1.697	1.602	1.116		
16,000 < Price <= 20,000 BEF		1.388		1.251	1.372	1.315			
20,000 < Price <= 40,000 BEF		1.227		1.166	1.279	1.255	1.059		
Contract type – budget		2.068	2.349	2.171	1.571	1.963	2.406	3.350	
ATS $_{t-1}$ – over credit limit			0.334	0.021	0.010	0.079	0.331	0.286	
ATS $_{t-1}$ = 0			0.790	1.775	1.867	3.399	0.842	0.532	
ATS $_{t-1}$ <= 5000			0.655	0.727	0.315	0.637	0.699	0.575	
5000< ATS $_{t-1}$ < =10,000			0.801		0.730		0.823	0.737	
10,000< ATS $_{t-1}$ <= 20,000			0.880				0.870		1.326
ATS $_{t-2}$ – over credit limit					2.406				
ATS $_{t-2}$ = 0					2.601				
ATS $_{t-2}$ <= 5000					1.677				
5000< ATS $_{t-2}$ < =10,000					1.153				
1 missed payment						0.008			
2 missed payments						0.013			

Table 5. Log Likelihood statistics for models with different information levels

Model	Log L	df	Log L difference
No covariates	-120280		
Application	-120181	9	99
Application+product	-119202	23	1078
Application+product+ATS at period1	-118904	23	1376
Application+product+ATS time-varying	-117999	20	2281

Table 6. Log Likelihood statistics for the ATS lagged and not lagged

Model	Log L	df	Log L difference
No covariates	-72778		
Application+product+ATS time-varying 1 period before the purchase	-71590	23	1188
Application+product+ATS time-varying 2 periods before the purchase	-71678	24	1100
Application+product+ATS time-varying 1 and 2 periods before the purchase	-71183	22	1595

Table 7. Area under the ROC-curve for models with different information levels

Model	**Hold-out**					
	1 (7752)	**2 (5725)**	**3 (5120)**	**4 (4688)**	**5 (4405)**	**6 (4167)**
Application	0.553					
Application+product	0.689					
Application+product+ATS at period 1	**0.715**	0.652	0.626	0.608	0.596	0.586
Application+product+ATS time-varying	0.710	0.662	0.638	0.621	0.610	0.600
Application+product+ all behavioural info time-varying		0.643	0.638	**0.658**	0.647	0.648
Dynamic model, ATS $_{t-1}$		**0.670**	0.654	0.649	0.641	0.635
Dynamic model, all beh. variables		0.669	**0.658**	0.654	**0.654**	**0.650**

Figure 1. Time sequences

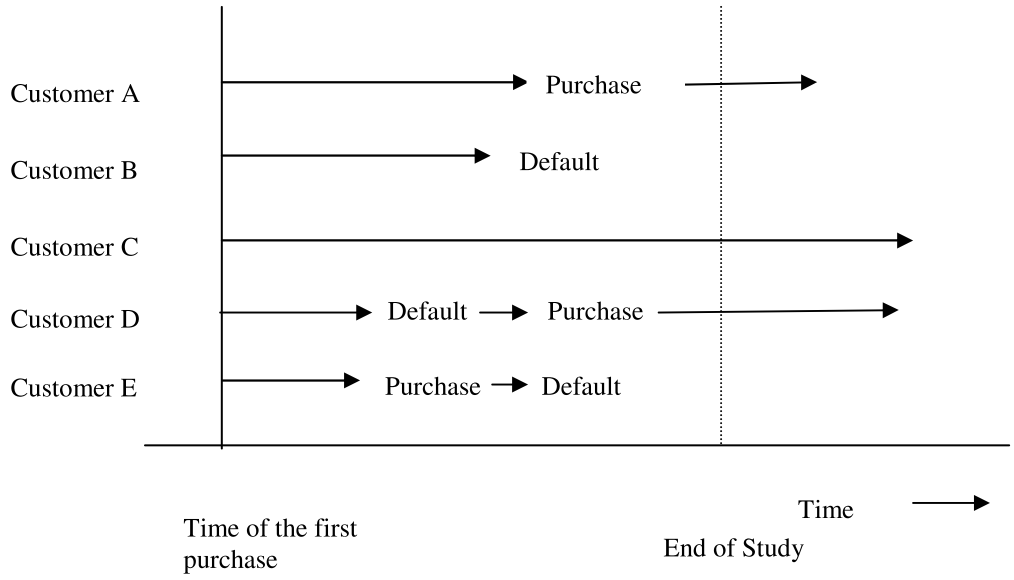


Figure 2. Baseline Survival Distribution Function with 95% Confidence Intervals, application data

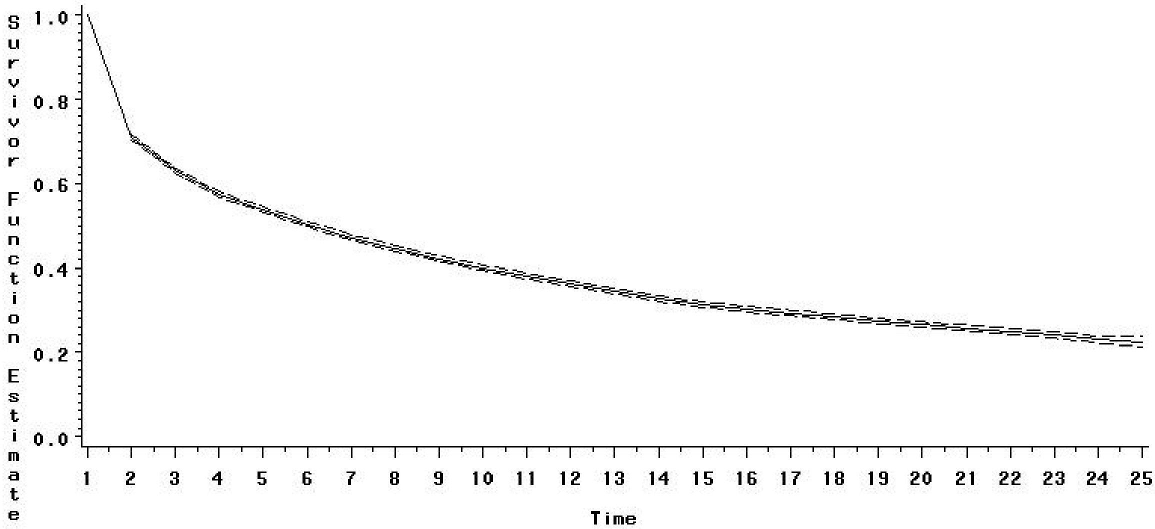


Figure 3. Parameter Estimates From Dynamic Model For Amount To Spend

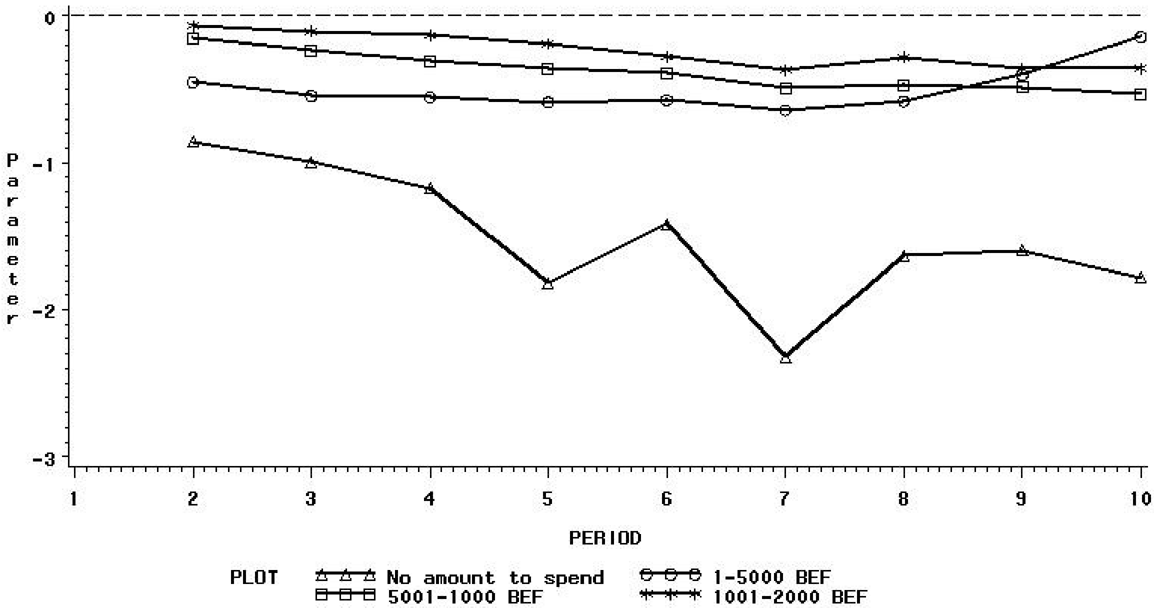


Figure 4. Parameter Estimates From The Dynamic Model. Application Score, Delinquency, Percent Repaid

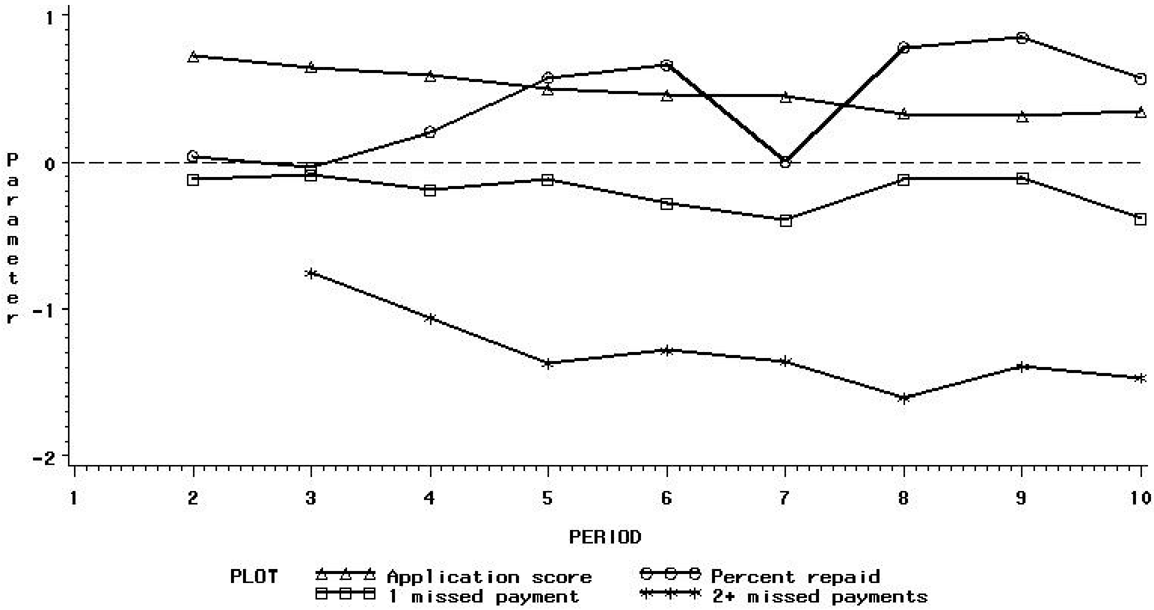
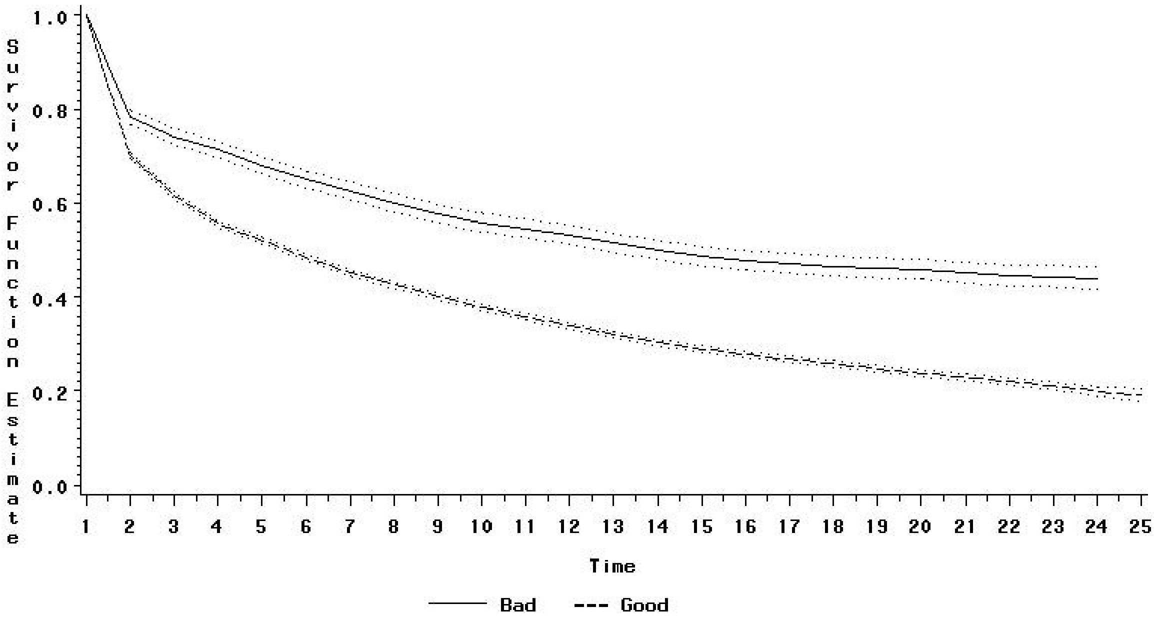


Figure 5. Baseline Survival Distribution Function for Good/Bad with 95% Confidence Intervals



List of Tables

Table1. Sample used in the analysis

Table 2. Variables used in the analysis

Table 3. Parameter estimates of proportional hazard models with stepwise selection

Table 4. Hazard ratios of proportional hazard models with stepwise selection

Table 5. Log Likelihood statistics for models with different information levels

Table 6. Log Likelihood statistics for the ATS lagged and not lagged

Table 7. Area under the ROC-curve for models with different information levels

List of Figures

Figure 1. Time sequences

Figure 2. Baseline Survival Distribution Function with 95% Confidence Intervals, application data

Figure 3. Parameter Estimates From Dynamic Model For Amount To Spend

Figure 4. Parameter Estimates From The Dynamic Model. Application Score, Delinquency, Percent Repaid

Figure 5. Baseline Survival Distribution Function for Good/Bad with 95% Confidence Intervals