Dynamic Models of Default on UK Credit Cards

Tony Bellotti and Jonathan Crook
Credit Research Centre, University of Edinburgh Business School
William Robertson Building, 50 George Square, Edinburgh EH8 9JY, UK

Abstract

Typically models of credit card default are built on static data, often collected at time of application. We consider alternative models that also include behavioural data about credit card holders and macroeconomic conditions across the credit card lifetime, using a discrete survival analysis framework. We find that models that include these dynamic variables give statistically significant improvements in model fit which translate into better forecasts of default at both account and portfolio level when applied to an out-of-sample data set. Additionally, by simulating extreme economic conditions, we show how these models can be used to stress test credit card portfolios.

1. Introduction

Consumer credit scoring models use details about obligors or potential customers that are static. In particular, this is the case for application scoring models that are used to determine whether an applicant should be granted credit based on data collected at time of application that then remain fixed. Typically, this is information taken from the completed application form and a credit score for the individual provided by a credit bureau. Such static models can also be used to assess obligors after they have been issued credit to determine if they are likely to become delinquent on payments. However, such models are restrictive since a credit portfolio is naturally a panel data composed of a collection of obligor accounts, each with its own credit history over time. As such a dynamic model would be more appropriate to determine creditworthiness within a portfolio. In this way, recent time-varying behavioural factors such as credit usage and payments can be used to supplement the basic application data in order to yield more accurate estimates of creditworthiness. Additionally, a dynamic model can include other time-varying components. In particular we may expect common economic risk factors to affect all obligors in a portfolio generally in the same way. For example, we would expect that a large increase in interest rates to cause, ceteris paribus, a general increase in probability of default (PD). Further, static models typically only have value in assessing the riskiness of applicants and obligors. However, if we want a complete picture we should be looking at *return* alongside risk. Thomas et al (2001) argue that this can be done using profitability estimates based on account performance over time for which dynamic models are required and for which static models cannot be used. In this paper, we present dynamic models of default which include behavioural variables (BV) and macroeconomic variables (MV) in addition to application variables (AV). We have a large panel database of UK credit card data and, based on this, assess these models in terms of model fit and predictive performance. The inclusion of MVs also enables us to perform stress tests since extreme economic conditions can be simulated and included in the model to generate a measure of stressed loss. We use Monte Carlo simulation to generate loss distributions of estimated default rates (DR) as the basis of a stress test of our credit card data.

Survival analysis can be used to build dynamic models of default since it readily allows the inclusion of BVs and MVs as time-varying covariates (TVCs). Bellotti and Crook (2008) follow this path using the Cox proportional hazard survival model to model time to default for a large database of credit cards. They include MVs, but not BVs, as TVCs and find a modest improvement in predictive performance in comparison to a static logistic regression. Here, we take a similar approach. However, we differ in that we use *discrete* survival analysis as opposed to continuous time. Discrete survival analysis can also be understood as a logistic regression on a panel data set with the data set up so that default is conditional on no prior default having already occurred, for each account. Since credit data is usually in the form of a panel data, and in particular account records are discrete (eg monthly records), this is a more natural choice than continuous time survival analysis. It also has the advantage of being more computationally efficient since probability forecasts involve simple summations over time periods, rather than an integration which may be complex when TVCs are included in the model.

Ultimately, financial institutions and regulators are interested in consumer credit risk models for estimation of future losses at both account and portfolio levels, either in normal (expected) circumstances or considering stressed conditions. For this reason, we focus primarily on using the models for forecasting PD and DR. For a large database of UK credit cards we establish the following new results:-

- The inclusion of BVs improves model fit and also improves forecasts; and the best results are achieved with the most recent behavioural data;
- Several MVs are found to be statistically significant explanatory variables of default, but this does not translate into improved forecasts at the account-level;
- However, we show that including MVs does improve estimation of loss (DR) at a portfolio or aggregate level;
- Additionally, we show how models with MVs can be used for stress testing and report loss distributions based on Monte Carlo simulation of economic conditions.

In section 2 we outline the methods we use, describing the discrete survival model, test procedures and stress testing methodology. In section 3 we describe our data and present results in section 4. Finally, we give conclusion and discussions in section 5.

2. Methods

2.1 Discrete survival model for dynamic credit scoring

We consider a panel data set of loan or credit card accounts. For each account *i* we have the following data:

- a_i is the date of account opening;
- d_{it} indicates whether the account *i* defaults at time *t* after account opening (0=non-default, 1=default);

- $\mathbf{w}_i$ is a vector of static AVs collected at time of account application;
- $\mathbf{x}_{it}$ is a vector of BVs collected across the lifetime of the account.

We model the probability of default (PD) for each account *i* at time *t* as

$$\begin{aligned} P_{it} &= \Pr\left(d_{it} = 1 \mid d_{is} = 0 \text{ for all } s < t; \mathbf{w}_i, \mathbf{x}_{it}, \mathbf{z}_{it}, k, l\right) \\ &= F\left(\alpha + \boldsymbol{\varphi}(t)^T \boldsymbol{\beta}_1 + \mathbf{w}_i^T \boldsymbol{\beta}_2 + \mathbf{x}_{i(t-k)}^T \boldsymbol{\beta}_3 + \mathbf{z}_{i(t-l)}^T \boldsymbol{\beta}_4\right) \end{aligned} \tag{1}$$

where

- $\mathbf{z}_{it}$ is a vector of MVs which are the same for each account on the same date; that is, for any two accounts *i*, *j* with duration times *t* and *s* respectively, if $a_i + t = a_j + s$ then $\mathbf{z}_{it} = \mathbf{z}_{js}$;
- *k* and *l* are fixed lags on BVs and MVs respectively;
- $\boldsymbol{\varphi}$ is a vector transformation function of duration that is used to build a parametric survival model; in particular, we use the transformation $\boldsymbol{\varphi}(t) = \left(t, t^2, \log t, (\log t)^2\right)$;
- α is intercept and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4$ are vectors of coefficients to be estimated;
- F is a given distribution function. We use the standard logit function $F(x) = 1/\left(1 + e^{-x}\right)$.

We ensure that the underlying panel data is constrained by the condition in (1): that is, no observations are recorded after the first default on any account. Given this condition, the model is a proportional odds discrete survival model with the failure event given as default. It can be estimated using standard maximum likelihood estimation for logistic regression (Allison 1995). Coefficient estimates on duration $\boldsymbol{\varphi}(t)$ form a baseline hazard function. The estimated survival probability of an individual *i* at some time *t* is given as the product of the probability of not failing at each time period conditional on not having failed previously. That is

$$\hat{S}_i(t) = \prod_{s=1}^{t} \left(1 - P_{is}\right). \tag{2}$$

The failure probability $1 - \hat{S}_i(t)$ then gives PD up to time *t* which is the typically used measure of PD and can be used in further analysis at the account or portfolio level for credit scoring and computing capital requirements.

In order to compare performance of different model components such as BVs and MVs we consider the following special cases of model (1):

1. Duration only: fix $\boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4$ to zero.
2. AV only: fix $\boldsymbol{\beta}_3, \boldsymbol{\beta}_4$ to zero.
3. AV and BV only: fix $\boldsymbol{\beta}_4$ to zero.
4. AV, BV and MV: all coefficients are estimated.

The lag *k* on the BVs restricts the range of forecasts that can be made by the model, since a period *k* after our observation date, there will no longer be any behavioural data available to make estimates. For example, if the lag is 6 months then we can only forecast using the BV model up to 6 months ahead. Clearly the longer period we can forecast forward, the better. However, we would expect that if longer lags were used, forecast performance would deteriorate. So we have a trade-off. We expect forecasts of 6-12 months ahead to be useful and so we consider lags of 12, 9 and 6 months. We also include a 3 month lag model, even though this is not a useful forecast period, for comparative purposes over short lag periods. It is also possible that some BVs are endogenous variables. For example, there may be a common underlying factor which causes both an increase in account balance and default. Then high balance is not a cause of default, although it may be found to be an important driver of default in the model. The shorter the lag period, the more likely this connection, which is another reason why longer lags are preferable and we report the BV lag 12 explanatory model. Nevertheless, we note that although endogeneity affects the identification of cause, it does not affect forecasts which are the main concern of this paper. The situation with the lag term *l* on the MVs is different. The MVs can be estimated using standard autoregressive methods (Hamilton 1994) or may be used with simulated values during stress testing. For this reason we use MV values at time of default. In particular, for this exercise, we use a definition of default as 3 months of consecutively missed payments, therefore we take a 3 months lag on MVs.

2.2 Performance measures

We use the log-likelihood ratio (LLR) for model fit for each of the model separately and to test goodness-of-fit for nested models. But this only measures model fit on the training sample and not performance measures for forecasts. Since we are using survival models which model time to failure (default), the usual measures of model fit for classification algorithms, such as error rates and the Gini coefficient, do not naturally apply, nor do the standard residuals for regression such as mean square error. Survival analysis has its own residuals that can be used to measure fit related to how well the estimated survival probability matches the observed time of default. In particular, these residuals take account of censored data too. One useful measure is the deviance residual given by

$$r_{Di} = \mathrm{sgn}(r_{Mi})[-2\{r_{Mi} + \delta_i \log(\delta_i - r_{Mi})\}]^{1/2} \tag{3}$$

where $r_{Mi} = \delta_i - r_{Ci}$ is the martingale residual and $r_{Ci} = -\log \hat{S}_i(t_i^*)$ is the Cox-Snell residual where t_i^* is the last observation available for individual i and $\delta_i = d_{it_i^*}$ indicates whether it failed at that observation. The martingale residual takes account of whether or not an individual fails but unfortunately they are not symmetrically distributed about zero nor are they additive terms. The deviance residuals have the advantage that they are more symmetrically distributed and the sum of their squares forms the statistic

$$D = \sum r_{Di}^2 = -2(\log \hat{L}_C - \log \hat{L}_f) \tag{4}$$

where $\hat{L}_C$ and $\hat{L}_f$ are the maximum partial likelihood under the current and the full model respectively. The full model implies a model with perfect fit to the data therefore D gives a measure of log-likelihood deviance of the model from the best case. Therefore models giving smaller values of D have better fit (Collett 1994).

The deviance residual is designed for assessing forecasts at the account level. However, our models can also be used to forecast at an *aggregate* level: eg across accounts within a single portfolio. The observed DR for an aggregate of N accounts at a particular calendar date c is given by

$$D_c = \frac{1}{N}\sum_{i=1}^{N} d_{i(c-a_i)} \quad (5)$$

which, assuming independence between default events, implies that the estimated DR forecast by a particular model is

$$E(D_c) = \frac{1}{N}\sum_{i=1}^{N} P_{i(c-a_i)} \ . \quad (6)$$

The difference between expected and observed DR then gives a measure of performance for aggregate forecasts.

2.3 Stress testing

We consider a simulation-based stress test of DR on an aggregate of accounts using Monte Carlo simulation (see eg Marrison 2002). The procedure is as follows.

1. Build a dynamic model with MVs from a training data set.
2. Generate a simulation of economic conditions as values of MVs based on historic macroeconomic data.
3. Substitute the simulated MV values into the model for all cases in a test data set to simulate defaults.
4. Repeating steps 2 and 3, *m* times to build a loss distribution of estimated DR over different economic scenarios.
5. Use the loss distribution to compute estimated DR for extreme circumstances.

Stress tests should consider *unexpected* but *plausible* events. When *m* is large, sufficient extreme events can be simulated to meet the first criteria; basing the simulations on historical data ensures the second.

Often Value at Risk (VaR) is used to compute stressed values in step 5. However, VaR captures worst case in normal circumstances, whereas stress tests should consider unusual circumstances. Therefore VaR may not be an appropriate measure of stressed loss (BIS 2005). For this reason we consider *expected shortfall* as a measure of loss. We define this as the expected (mean) loss in the upper $(1-\alpha)$ percentile of the loss distribution given a significance level α. Using the latent variable model of logistic regression (Verbeek 2004, section 7.1.3), we can simulate DR for some calendar time period *c*, given a model, a vector of macroeconomic conditions $\mathbf{z}$ and a vector of *N* independent residual terms $\boldsymbol{\varepsilon} = (\varepsilon_{(1)}, \cdots, \varepsilon_{(N)})$, each distributed in *F*, as

$$D_c^*(\mathbf{z}, \boldsymbol{\varepsilon}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{I}\left(\alpha + \boldsymbol{\varphi}(c-a_i)^T \boldsymbol{\beta}_1 + \mathbf{w}_i^{\ T} \boldsymbol{\beta}_2 + \mathbf{x}_{i(c-a_i-k)}^{\ \ T} \boldsymbol{\beta}_3 + \mathbf{z}_{i(c-a_i-l)} > \varepsilon_{(i)}\right). \quad (7)$$

where I is the indicator function. For example, if our panel data is monthly and *c* is taken as June 2005 then (9) gives an estimate of DR within June 2005. Monte Carlo simulation can then be used to approximate expected shortfall with

$$S_\alpha \approx \frac{1}{\alpha m}\sum_{j=1}^{\lceil \alpha m \rceil} D_c^*\left(\mathbf{z}_j^*, \boldsymbol{\varepsilon}_j^*\right) \quad (8)$$

where *j*=1 to *m*, each $\mathbf{z}_j^*$ is generated by macroeconomic simulation and $\boldsymbol{\varepsilon}_j^*$ are generated randomly from F^N and both are indexed so that the simulated DRs are in descending order; ie for all $h \le j$, $D_c^*\left(\mathbf{z}_h^*, \boldsymbol{\varepsilon}_h^*\right) \ge D_c^*\left(\mathbf{z}_j^*, \boldsymbol{\varepsilon}_j^*\right)$. The number of iterations *m* is chosen so that (8) converges to a stable value. This simulation takes into consideration the error in the model represented by the residual terms, along with changes in macroeconomic conditions. This is natural, since otherwise the point predictions of equation (6) are wrongly assumed to be exactly correct.

Simulated values for MVs could be drawn naively directly from historic values. However, this would not preserve the structure of dependencies between the MVs and so will yield implausible scenarios and lead to misleading results. To preserve the covariance structure between MVs we use Cholesky decomposition (Marrison 2002). If **V** is a matrix of covariances for historic macroeconomic data then it is decomposed by a lower triangular matrix **L** such that $\mathbf{V} = \mathbf{L}\mathbf{L}^T$. Then, if $\mathbf{u}_j$ is a sequence of independently generated values from the standard normal distribution, $\mathbf{z}_j = \mathbf{L}\mathbf{u}_j$ will follow the covariance structure of **V** and so can be used as economic simulations. Cholesky decomposition assumes the variables are normally distributed. However, this is not usually the case for MVs and so we apply a transformation to MVs if this is required, prior to simulation. A Box-Cox transformation is used since this often produces an approximately normal distribution (Box and Cox 1964). Alternatively, we use an empirical probit transformation to impose a normal distribution on the historical data.

3. Data

3.1 Credit card data

We have three large data sets of UK credit card data covering a period from 1999 to mid-2006 comprising over 500,000 accounts. All data sets include AVs taken at time of application for a credit card then fixed, along with monthly account behaviour records. The AVs include housing and employment status, length of customer relationship with bank, income, age and a credit bureau score. Most data are collected in the same way and have the same objective meaning between credit card products, although distributions vary since different products will have different demographic and risk profiles. Variables that may be defined differently for each product have not been used; eg the bank derived behavioural score is not used. There are a small proportion of missing values for monthly payment amount so an indicator variable is also included for these variables. Also, there are a large proportion of zero values for some BVs, payment amount, sales amount and APR, so indicator variables are included for those cases too. *Default* is defined as when an account goes three consecutive months delinquent on payments. This is a common definition in the industry and follows the Basel II convention of 90 days delinquency for consumer credit (BCSC 2006). The data we use for our analysis is commercially sensitive and therefore we cannot provide further details, data description statistics or report the observed DRs.

Following Granger and Huang (1997) we divide the panel credit data set into an in-sample *training data set* and a post, out-of-sample *test data set* based on a given observation date. Accounts from the full data set are randomly allocated to the training and test data sets in the ratio 4:1. This split of the data allows sufficient observations for training whilst leaving a good number of accounts out-of-sample for independent tests. Additionally, only records prior to the observation date are included in the training data set, whilst only accounts that are open on the observation date are included in S. This procedure enables us to perform forecast tests which mimic the way the model may be used in practice; ie to forecast ahead for an existing set of accounts. We take 1 July 2004 as our observation date since this provides up to 24 months of test data.

3.2 Historic UK macroeconomic data

We consider several UK MVs for which we had a prior expectation of their having an effect on PD. These are listed in Table 1. In a previous study (Bellotti and Crook 2008) on a different data set we found that bank interest rates, earnings, production index and house price were statistically significant explanatory variables of UK default so we include these. Production index is used instead of GDP since it is available monthly whereas GDP is only provided quarterly. Gerardi et al (2008) also found unemployment rate was significant for US defaults. Additional to these, Breedon and Thomas (2008) also found variables for consumer sales and prices were also correlated to default and bankruptcy in a study of a number of stressed credit markets worldwide. We also include these MVs and additionally the FTSE index and consumer confidence index since these may be good indicators of confidence in the economy. All variables are taken as difference over 12 months in order to reduce effects of a time trend and spurious regression.

Table 1. UK macroeconomic variables (MVs) with descriptive statistics for values from 1986 to 2004. Sources are the UK Office of National Statistics (ONS), Nationwide Building Society (Nat.) and the European Commission (EC). Data is monthly and may be seasonally adjusted (SA).

MV	***Description***	***Source***	***Descriptive statistics*** *(for difference in value over 12 months)*			
			Min	***Mean***	***SD***	***Max***
IR	UK bank interest rates	ONS	-4.5	-0.43	1.90	6.5
Unemp	UK unemployment rate (in ‘000s) SA	ONS	-535	-94	238	575
Prod	UK production index (all)	ONS	-5.2	1.10	2.30	6
RS	Retail sales value	ONS	0.3	3.92	1.49	8.5
FTSE	FTSE 100 all share index	FTSE	-822	81	286	682
HP	House price	Nat.	-6.5%	+7.9%	7.6%	+26%
RPI	Retail price index (all items)	ONS	1.2	4.96	2.36	12.8
Earnings	Earnings (log) all including bonus	ONS	0.008	0.019	0.006	0.038
CC	Consumer confidence index	EC	-20.3	0.7	24.2	186.8

4. Results

We present results in four subsections. Firstly, we discuss coefficient estimates from the model build. Secondly, we present model fit and forecasting results at the account level. Thirdly, we present forecast results at the aggregate level. And lastly we present stress test results.

4.1 Model and coefficient estimates

Many BVs and several MVs were statistically significant explanatory variables. We focus attention on the model for BV lag 12 months, since this is the most practically valuable model in terms of forecast range. Table 2 shows coefficient estimates for this model. Only BVs and MVs are reported since for this study they are the ones of interest. We have the following key outcomes:-

1. The signs on current balance (log) and its square are opposite but the positive sign on the square dominates. Therefore, balance outstanding on the account has an increasing positive effect on default hazard. This is unsurprising since a larger balance will be more difficult to clear.
2. An increase in credit limit reduces the hazard. Initially this may be surprising since a high credit limit encourages higher balance and therefore greater risk. However, at least in the short term, it enables the obligor to have a buffer to build up debts before reaching default. Also, the bank sets the credit limit based on their own assessment of the obligor's behaviour, so credit limit is acting partially as a proxy for a behavioural score.
3. The amount paid back each month, indicated by payment status and payment amount, has a negative affect on default. This is expected since a greater ability to repay implies that default is less likely.
4. Number of transactions has a positive effect on default. This is expected since it indicates greater card use and hence a rising balance. However, interestingly, the effect of transaction sales amount is negative. A possible explanation is that sales amount is acting as an indicator of wealth when taken together with number of transactions. That is, people who make a few big purchases are more likely to be wealthier and therefore more able to repay than those who make many small purchases.
5. When behavioural data is missing, PD decreases considerably. However, since all duration times up to 12 months will not have BVs (because of the lag) this is mainly a joint effect with duration.
6. Interest rate has a positive affect on default. This is expected since rising interest rates imply greater demand for repayment on outstanding loans and mortgages which will adversely affect those people who are more highly indebted.
7. Production index has a negative affect on default. Since this is an indicator of economic prosperity, we would expect this as good economic conditions should tend to lead to lower DRs.
8. Increases in UK retail sales leads to an increase of PD. This may be an indication of over-spending which will then tend to place stress on indebted credit card holders.
9. Surprisingly, the FTSE 100 is positively significant. As an indicator of economic health, we would expect it to have the opposite sign, rather like the

production index. However, it is better to interpret it as a measure of *confidence*. In that sense it may behave like the retail sales variable as an indicator of consumer confidence that may lead to over-indebtedness.

Table 2. Coefficient estimates for model with all AVs, BVs lag 12 months and MVs. Only estimates on BVs and MVs are shown. Indicator variables are denoted by a plus sign (+). Statistical significance levels are denoted by asterisks: *** is less than 0.001, ** is less than 0.01 and * is less than 0.05 level.

Covariate	Estimate
Behavioural variables (BV) lag 12 months	
Payment status + :	
Fully paid	-0.360***
Greater than minimum paid	-0.130***
Minimum paid	0.0983**
Less than minimum paid	0.669***
Unknown	-0.0946*
Excluded category: No payment	
Current balance (log)	-1.39***
" (log squared)	0.475***
" is zero +	-0.948***
" is negative +	-0.567***
Credit limit (log)	-1.126***
Payment amount (log)	-0.140***
" is zero +	-0.100*
" is unknown +	-0.418***
Number of months past due	-0.0290
Past due amount (log)	0.0928
" is zero +	-0.722***
Number of transactions	0.00737***
Transaction sales amount (log)	-0.372***
" is zero +	-0.620***
APR on purchases	-0.00295
" is zero +	-0.440***
Behavioural data is missing +	-3.50***
Macroeconomic variables (MV) lag 3 months	
Bank interest rate	0.0568*
Unemployment rate	0.000328
Production index	-0.0144**
FTSE all 100	0.000083**
Earnings (log)	0.969
Retail sales	0.0161**
House price (log)	0.496
Consumer confidence	-0.00019
Retail price index (RPI)	0.00419

Points 1 to 3 corroborate the results of Gross and Souleles (2002) who built dynamic models of default for US credit card data. They found risk of default rises with balance and falls with repayments. They used *utilization* - outstanding balance divided by credit limit - instead of the raw value of balance, which is sensible given the relationship discussed in point 2. Also they had the same outcome for interest rates described in point 6. They also found unemployment rate significant, and although we do not, we do have the same positive sign. Bellotti and Crook (2008) found similar results as points 6 and 7 on a different UK credit card data set and although the studies do not correspond with regard to which MVs are significant, overall, the signs on coefficients are the same. Points 6 and 7 also corroborate the study by Breedon and Thomas (2008) across several world-wide data sets. They find, consistently, GDP negatively correlated and interest rates positively correlated with default and bankruptcy rates.

4.2 Model fit and forecasts of time to default

Model fit is shown in Figure 1 giving a general improvement in model fit as we move from the simple duration only model to the AV only model to the AV and BV model. Additionally, we also observe that model fit improves with shorter lag on BVs with a relatively large improvement at 3 month lag. However, as we have discussed, the improvement comes at the price of a much shorter range of forecasts. We see that although some of the MVs are statistically significant, their contribution to model fit is weak. Nested model fit is also assessed and the results given in Table 3. This shows that adding BVs to the model gives a statistically significant improvement in fit. Adding MVs to the model also gives a statistically significant improvement, even though the improvement is small.

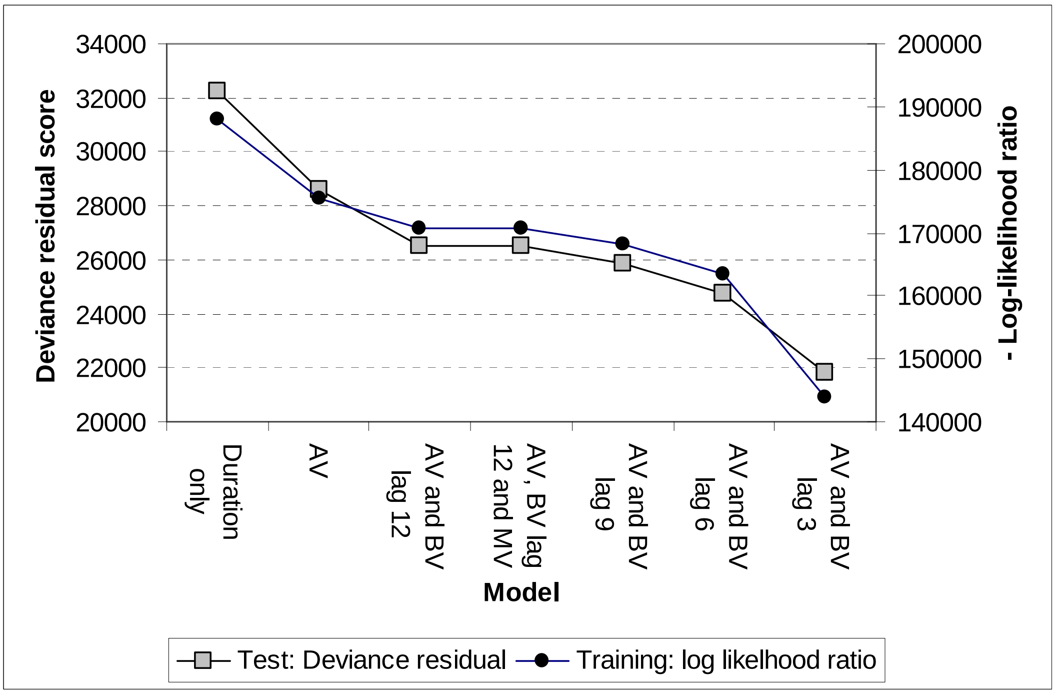

Figure 1. Model fit and forecast performance for different models.

Table 3. Model fit results for nested models using difference in LLR and a chi-square significance test.

Nested model	Compared to base model	Difference in 2×LLR	Number of added covariables	P-value
AV only	Duration only	12413	29	<0.0001
AV & BV lag 12	AV only	5289	21	<0.0001
AV, BV lag 12 & MV lag 3	AV & BV lag 12	18	9	<0.0001
AV & BV lag 9	AV only	7946	21	<0.0001
AV & BV lag 6	AV only	12458	21	<0.0001
AV & BV lag 3	AV only	26559	21	<0.0001

Figure 1 also shows results of forecasts. These follow the model fit results very closely and show a marked improvement in fit for the BV models, improving with shorter lags, but no noticeable change in forecast accuracy when MVs are included.

4.3 Estimation of Default Rates

Figure 2 shows estimated DR for different models along with the observed (true) DR for each month of the test data set. The monthly observed DR has high variance but there is a general trend of high values in 2005 then falling in 2006. All models successfully follow this pattern to some extend, including the AV model. This implies that the changes in DR may largely be due to different levels of risk amongst different cohorts of accounts at time of application. However, the model with MVs clearly follows the observed trend much more closely. Interestingly, although the BV model with lag 3 is the superior forecast model at the account level, it tends to over-estimate DR and the fit is weak at the aggregate level.

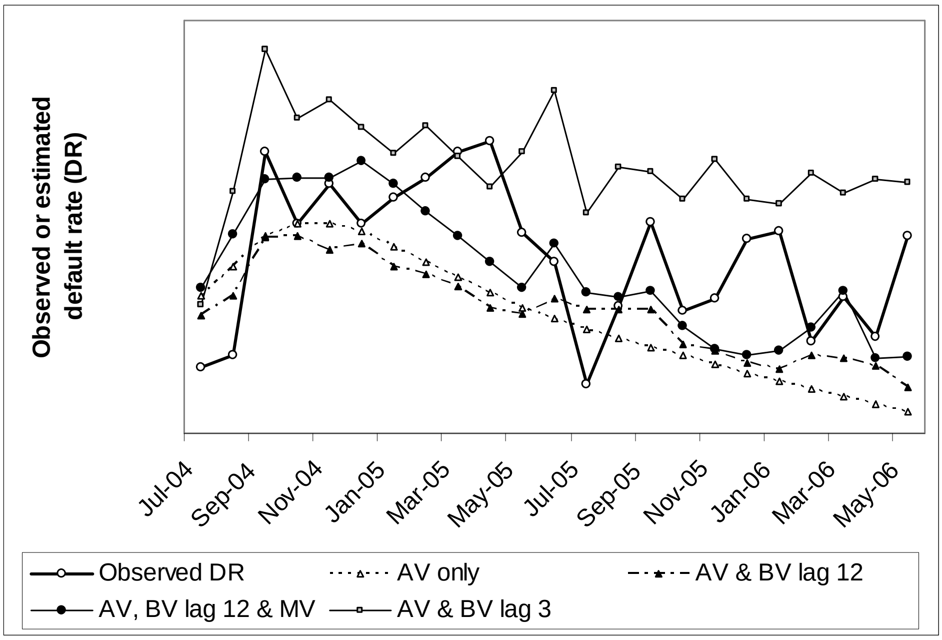

Figure 2. Estimated DR given each model compared with observed DR for each month of the test set data. The scale on DR is not shown since it is commercially sensitive.

Table 4 gives overall performance for each model across the whole test data set. This shows that the model with MVs is clearly the best estimator giving the minimal deviation from observed DR (2.4%). We use a paired t-test to check that the difference in performance is statistically significant. We consider the deviation (ie the absolute difference) each month between forecast and observed DR for each model, using data shown in Figure 2. We find the series of deviations of the MV model is statistically significantly less than both the AV only and BV lag 12 month models with p-values of 0.0016 and 0.022 respectively.

Table 4. Difference between estimated and observed DR across test set data for each model.

Model	Difference between estimated and observed DR
AV only	-9.4%
BV lag 12	-8.6%
BV lag 12 & MV lag 3	-2.4%
BV lag 12, MV lag 3 & interaction terms	+17.2%
BV lag 3	+14.5%

4.4 Stress test results

We ran Monte Carlo simulations using the MV model given in Table 2. A stable loss distribution was generated after m=25,000 simulations and is shown in Figure 3. The right-hand tail shows risk for more adverse conditions. In particular we have included the figure for expected shortfall at the α =1% level. This shows that for the worst 1% of economic scenarios we consider, the expected DR (1.45) is 45% greater than normal conditions (ie median estimated DR). VaR is also shown for comparison. We see that this gives a lower estimate of extreme loss (1.37) which may not reflect extreme circumstances sufficiently. These figures are credible and, in particular, match the figures for credit card loss for the US stress testing exercise by FRS (2009). In particular they estimate a rise between 20% and 55% when contrasting a normal "baseline" figure to "more adverse" conditions[1].

5. Conclusion

We have argued that dynamic models are a flexible approach to model and forecast consumer credit risk. They have several distinct advantages over static models.

(1) We can naturally model time to default (or any other event such as repayment) rather than just whether an account will default (at some time).

(2) They allow the inclusion of BVs which enables better forecasts of default using recent behavioural data.

(3) MVs can be included to give better model fit and improved forecasts at aggregate (portfolio) level and enable stress testing.

(4) Dynamic models potentially allow us to build profitability models which consider the return on loans as well as the risk.

[1] FRS (2009) gives baseline two-year loss rates as 12-17% and "more adverse" as 18-20%. Taking the lower and upper bounds on each range and converting to an average monthly DR gives 20-55% expected increase in loss. Taking a mean value for baseline and more adverse (14.5% and 19% respectively) gives a mean increase of 34%.

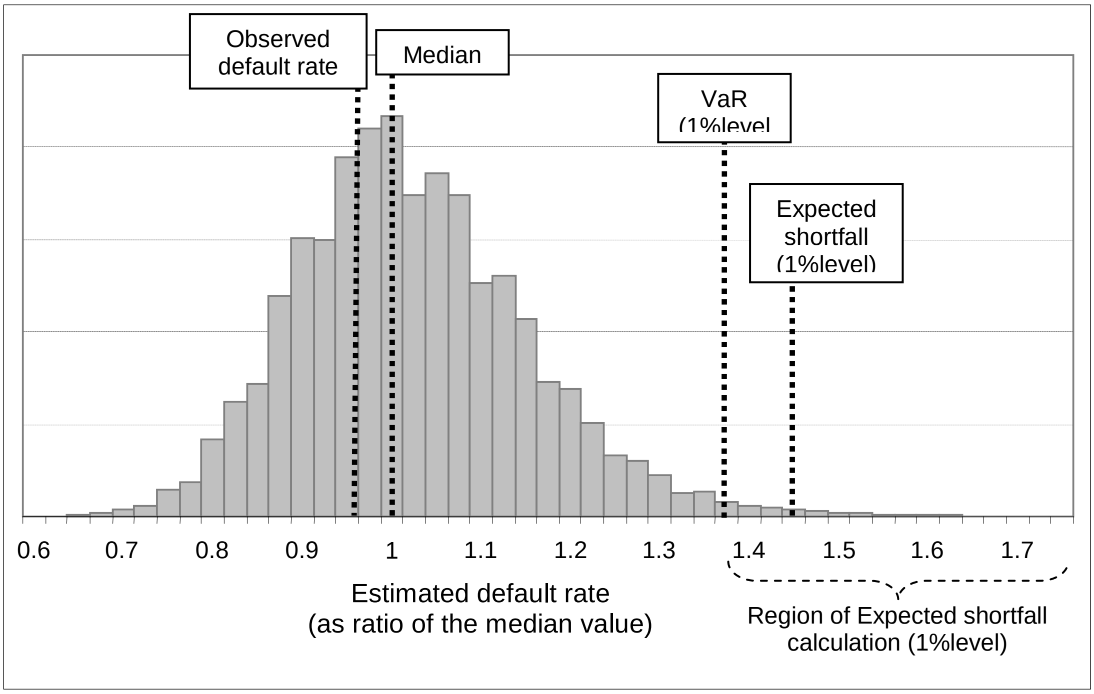

Figure 3. Distribution of estimated DR based on simulation of economic scenarios for test data during July 2004, shown as a histogram. The observed DR for the data set is shown along with Value at Risk (VaR) and expected shortfall at 1% level. All values are expressed as a ratio of the median estimated DR.

We have used discrete survival analysis using the logit model. This has two main advantages. Firstly, it is a principled means to build dynamic models of default and, secondly, modelling and forecasting is computationally efficient when compared to continuous-time models such as the Cox proportional hazard model. This is important when we consider large data sets of credit accounts as we do. We have tested the effectiveness of dynamic survival models with BVs and MVs as models of default. We explore these models as tools for risk measurement, forecasting and stress testing with the following key findings.

- Many BVs are statistically significant explanatory variables of default and including them give improved model fit. Important BVs are account balance, repayments, number of transactions within each month and credit limit. We find this translates into improved forecasts of time to default. Performance improves with shorter lags on BVs. This is expected since shorter lags imply that the model is using more recent information about the obligors. However, we also note that shorter lags imply shorter ranges of forecasts and greater endogeneity between BVs and the default event. For this reason we focus on lag 12 month BVs. This gives improved performance, relative to the AV only model, and also allows for useful forecasts up to 12 months ahead.

- Several MVs were also found to be statistically significant: bank interest rates, production index, FTSE 100 and retail sales. However, their inclusion gave only a modest improvement in model fit and no noticeable improvement in forecasts of time to default at the account level. Nevertheless, the MV model gives significantly better forecasts of DR at the aggregate level. This is understandable since MVs are systematic “macro” variables that affect the population as a whole. Hence their affect will only become noticeable for forecasts at the aggregate level where accounts are taken together.

- The inclusion of MVs enables stress tests which generate credible results. We use a simulation-based approach but scenarios could also be designed and used with these models.

Acknowledgements

We are grateful for funding through an EPSRC grant EP/D505380/1, working as part of the Quantitative Financial Risk Management Centre.

References

Allison PD (1995). *Survival analysis using SAS*. SAS Press.

Basel Committee on Banking Supervision BCBS (2005). *Basel II: International Convergence of Capital Measurement and Capital Standards* at www.bis.org/publ/bcbsca.htm

Bellotti T and Crook J (2008). Credit scoring with macroeconomic variables using survival analysis. *The Journal of the Operational Research Society* (online).

Box GEP and Cox DR (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B* **26**: pp211–246

Breedon J and Thomas L (2008). The relationship between default and economic cycles for retail portfolios across countries. *The Journal of Risk Model Validation* **2**(3), 11-44.

Calhoun CA and Deng Y (2002): A dynamic analysis of fixed- and adjustable-rate mortgage terminations. *Journal of Real Estate Finance and Economics* 24:1/2 pp 9-33.

Collett D (1994). *Modelling survival data in medical research*. Chapman & Hall: London.

Board of Governors of the Federal Reserve System FRS (2009). *The supervisory assessment program: overview of results*. FRS: USA.

Gross DB and Souleles NS (2002). An empirical analysis of personal bankruptcy and delinquency. *The Review of Financial Studies* Vol 15, no 1, pp319-347.

Gerardi K, Shapiro AH, Willen PS (2008). *Subprime outcomes: risky mortgages, homeownership experiences, and foreclosures*. Working paper 07-15 Federal Reserve Bank of Boston.

Granger CWJ and Huang LL (1997). *Evaluation of panel data models: some suggestions from time series*. Discussion paper 97-10, Department of Economics, University of Caliafornia, San Diego.

Hamilton JD (1994). *Time Series Analysis*. Princetoin University Press.

Marrison C (2002). *Fundamentals of Risk Measurement* (McGraw-Hill NY).

Thomas LC, Ho J and Scherer WT (2001). Time will tell: behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics* (2001) **12** pp 89-103.

Verbeek (2004). *A Guide to Modern Econometrics* (2nd ed, Wiley).