Forecasting and Stress Testing Credit Card Default using Dynamic Models

Jonathan Crook,
University of Edinburgh - Business School
50 George Square , Edinburgh, Mid Lothian EH8 9JY
United Kingdom
J.Crook@ed.ac.uk

Tony Bellotti
Imperial College - Mathematics and Statistics
London
United Kingdom

Abstract: We present discrete time survival models of borrower default for credit cards that include behavioural data about credit card holders and macroeconomic conditions across the credit card lifetime. We find that dynamic models that include these behavioural and macroeconomic variables give statistically significant improvements in model fit which translates into better forecasts of default at both account and portfolio level when applied to an out-of-sample data set. By simulating extreme economic conditions, we show how these models can be used to stress test credit card portfolios.

1. Introduction

Application consumer credit scoring models use details about obligors or potential customers that are static. Such models are used to determine whether an applicant should be granted credit, based on data which are collected at the time of application and then remain fixed. Behavioural consumer credit scoring models use both information collected at the time of application and behavioural variables, the values of which have changed over time, but which are fixed at the time of estimation. Both are cross-sectional models and allow the prediction of a probability of default within a specified time window, like eighteen months. However, models that answer more specific questions can also be estimated from credit portfolios, since they provide panel

data (Crook & Bellotti, 2009) for a sample of obligor accounts. Panel data allow one to estimate hazard models which predict the probability of an event (such as a default) occurring in the next instant of time, conditional on the event not having occurred before that time, for any future time period one chooses. Unlike cross-sectional models, in a panel model one can include variables whose values change over the estimation period. Of particular relevance here are common economic risk factors that affect all obligors in a portfolio in generally the same way. For example, we would expect that a large increase in interest rates would cause, *ceteris paribus*, a general increase in the probability of default (PD). Time varying behavioural variables may also be included. We call these sorts of models that include time varying covariates (TVCs) ‘dynamic models’. Furthermore, 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 the *return* alongside risk, which requires the use of dynamic rather than static models (Thomas, Ho, & Scherer, 2001; Ma, Crook, & Ansell, 2009). In this paper, we present dynamic models of default which include time varying behavioural variables (BVs) and macroeconomic variables (MVs), in addition to application variables (AVs).

The inclusion of MVs also enables us to perform stress tests, since extreme economic conditions can be simulated and included in the model in order to generate a measure of stressed loss (default rates). Accurate stress tests are becoming increasingly important in evaluating the risk to banks, as is evident from the evaluation of US banks (Board of Governors, 2009) and the recognition by the Financial Services Authority (2008) that stress testing is a key tool in helping financial institutions to make business strategy, risk management and capital planning decisions.

Our paper contributes to the literature in four ways. Firstly, for a large portfolio of credit card accounts, we show that including behavioural variables improves the model fit in a discrete time hazard model, and that their inclusion improves the forecast accuracy. Secondly, we find that, while several MVs are statistically significant explanatory variables of default, this does not translate into improved forecasts at the account level. Thirdly, we show that including MVs can improve the estimation of loss (default rate) at the portfolio level. Fourthly, using account level data, we demonstrate the use of MVs for stress testing and report the distribution of expected default rates based on Monte Carlo simulation of economic conditions.

In Section 2, we provide a literature review. In Section 3, we outline the methods we use, describing the discrete survival model, our test procedures and the stress testing methodology. We then describe our data in Section 4, and present some results in Section 5. Finally, we offer conclusions and a discussion in Section 6.

2. Literature Review

Several modelling techniques have been proposed to develop a dynamic model of credit (see Crook & Bellotti, 2009, for a review). Thomas et al. (2001) describe how a Markov chain stochastic process can be used as a dynamic model of delinquency. However, the approach they describe does not allow for model covariates, although models can be built on separate segments to allow the modelling of different risk groups. They also describe survival analysis as a means of building dynamic models, since this readily allows the inclusion of BVs and MVs as time-varying covariates (TVCs). Bellotti and Crook (2009) follow this path, using the Cox

proportional hazard survival model to model the 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. We take a similar approach using a survival model here, but with *discrete* time survival analysis. Discrete survival analysis can also be understood as a logistic regression on a panel data set, with the data arranged so that default is conditional on no prior default having occurred on that account. Since credit data are usually in the form of panel data, and in particular account records are discrete (e.g. 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.

Discrete survival models have been applied successfully to the analysis of personal bankruptcy and delinquency in the USA (Gross & Souleles, 2002), mortgage terminations (Calhoun & Deng, 2002), and competing risks of foreclosure and sales in the US subprime market (Gerardi, Shapiro, & Willen, 2008). Gross and Souleles (2002) used several different BVs and MVs. In particular, they included the outstanding account balance and repayments, and found that the former had a positive affect on bankruptcies, while the latter had a negative affect. They also found that the local unemployment rate had a statistically significant positive affect on bankruptcy, which is what we would expect, since an increase in unemployment is likely to affect some obligors adversely. Calhoun and Deng (2002) derived dynamic variables measuring the probability of negative equity and the mortgage premium. Both change over time and have a positive affect on default. They also included the ratio of 10-year to 1-year Constant Maturity

Treasury yields, and found it to be statistically significant for models of early repayment. For fixed-rate mortgages, the coefficient increases for higher ratios, with the rationale that mortgagors will be moving to adjustable-rate mortgages, in order to take advantage of the short-term relatively low interest rates. Gerardi et al. (2008) found that interest rates (the 6-month libor rate) and unemployment rate are statistically significant explanatory variables for both mortgage default and sales, with a positive affect on default, as we would expect, and a negative affect on sales. These studies have shown that both BVs and MVs are useful explanatory covariates for consumer credit risk. We therefore extend this work by using these dynamic models for forecasts and stress testing. Ultimately, financial institutions and regulators are interested in consumer credit risk models for the estimation of future losses at both the account and portfolio levels, either in normal (expected) circumstances or when considering adverse conditions. For this reason, we focus primarily on using the models for forecasting PD and the default rate.

The literature on stress testing is growing rapidly. In an early paper, Berkowitz (2000) proposed a stress testing methodology in which two separate forecast distributions are generated using a risk model: one for normal conditions and another reflecting stressed conditions, based on changes in an underlying factor. Our approach is rather different to that of Berkowitz (2000), in that we do not choose an initial distribution under stressed conditions; instead, we generate a single distribution of expected default rates and focus on the upper percentiles for stress testing.

Stress-testing models may be divided into macro stress-testing models and micro stress-testing models. Macro stress-testing models concern the implications of stressed states of the economy

for the capital of *groups* of institutions, and the aim is to examine the capital adequacy of the financial sector of an economy in the event of adverse shocks. In contrast, a micro stress test relates to a specific portfolio of one lender. Sorge and Virolainen (2006) divide macro stress tests into (i) those that relate aspects of banks' balance sheets to macroeconomic activity, and (ii) value at risk models where macroeconomic factors are related to aggregate default rates (not the probabilities of default of individual obligors). For examples of balance sheet models, see Drehmann, Sorensen, and Stringa (2010) and Delgado and Surina (2004).

Macro Value at Risk (VaR) stress tests and micro stress tests follow similar methodologies. First, a model that relates macroeconomic variables to each other, often a Vector Autoregressive Regression model, is estimated. Second, a default rate (macro) or probability of default (PD) (micro) model that incorporates macroeconomic variables is parameterised. Third, a macroeconomic scenario is chosen and its implications for the distributions of PD (or expected loss) are predicted, or impulse response functions are estimated. For examples of such macro stress tests, see Breuer, Jandacka, Mencia, and Simmer (2012), Jokivouille and Viren (2011), Castren, Dees, and Zaher (2010), and Huang, Zhou, and Zhu (2009).

Our work differs from that of others in a number of ways. We are concerned with micro stress testing rather than macro stress testing, and we consider *account level* observations rather than aggregate default rates. The data used in our analysis are therefore more granular than the data used in almost all other studies. In addition, the vast majority of published macro stress tests relate to corporate loans, whereas in this paper we consider credit card loans.

Few stress test results for retail loan portfolios have been published. Breeden and Ingram (2009) discuss the issues which are involved in the generation of scenarios using a model where the default rate of a portfolio over time is explained in terms of a function of the duration time, a function of calendar time, and a function of vintage, but they do not present the results of a stress test for a portfolio. Rösch and Scheule (2004) assume a Merton one-factor model and estimate loss (default rate) distributions for credit cards, mortgages and other consumer loans in the US. Unlike our paper, they use aggregate default rate data and do not include variables which are specific to the obligor. In more recent work, Rösch and Scheule (2008) consider estimating and stressing the correlation between PD and LGD. They present an integrated approach to stress testing using a risk model with unobserved systematic risk factors and observed economic variables, and show that this model can be used to produce plausible stress testing estimates of PD, LGD and loss rates for Hong Kong mortgage loans. Their methodology is similar to ours.

Pesaran, Schuermann, Treutler, and Weiner (2006) and Drehmann, Patton, and Sorensen (2005) are the only studies we know of that use a methodology similar to ours, but their work relates to corporate portfolios and their data are pooled across different lenders.

3. Methods

3.1. *Discrete survival model for dynamic credit scoring*

We treat time as being discrete and adopt the following notation. We denote calendar time by c and the date that an account i was opened by a_i . Let t be the number of months since an account was opened (duration time). The term d_{it} indicates whether account i defaults at time t after account opening (0 = non-default, 1 = default). The term $\mathbf{w}_i$ is a vector of static AVs collected at

the time of account application and $\mathbf{x}_{it}$ is a vector of BVs collected across the lifetime of the account. The term $\mathbf{z}_{it}$ is a vector of MVs, which is the same for all accounts on the same date; that is, for any two accounts i, j having records for duration times t and s respectively, if $a_i + t = a_j + s$ then $\mathbf{z}_{it} = \mathbf{z}_{js}$.

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

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

where k and l are fixed lags on BVs and MVs respectively; and $\boldsymbol{\varphi}$ is a vector transformation function of duration that is used to build a parametric survival model; specifically, we use the transformation $\boldsymbol{\varphi}(t) = (t, t^2, \log t, (\log t)^2)$. α is an 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 cumulative distribution function. We use the logit function $F(x) = 1/(1 + e^{-x})$. The vintage effect is introduced by using indicator variables in $\mathbf{w}_i$.

We ensure that the underlying panel data are constrained by the condition in Eq. (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 defined as a default. It can be estimated using standard maximum likelihood estimation for logistic regression (Allison, 1995).

Coefficient estimates on the duration $\boldsymbol{\varphi}(t)$ give a baseline hazard. If indicator variables are included for each discrete time, then this would parallel the commonly used Cox proportional

hazard model. However, having so many indicator variables would make the model difficult to estimate efficiently, even with a large sample. Published studies have suggested that there is a common shape to the distribution over duration time of default hazard rates: they rise sharply within the first few months, then fall steadily over the remaining duration of the account (Gross & Souleles, 2002, Figure 1; and Andreeva, 2006, Figure 1). We use a parametric form for ϕ which allows us to capture this structure of the hazard over time. Log terms are included in order to allow this structure to take a skewed shape. 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,

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

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

To compare the performances of different model components such as BVs and MVs, we consider the following special cases of Eq. (1):

1. duration only: fix $\beta_2, \beta_3, \beta_4$ to zero;
2. AV only: fix β_3, β_4 to zero;
3. AV and BV only: fix β_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 the period we can forecast forward, the better. However, we would expect that if longer lags were used, the forecast performance would deteriorate. Therefore, we have a trade-off. We expect forecasts of 3 to 12 months ahead to be useful, so we consider lags of 3, 6, 9 and 12 months. It is also possible that some BVs may be endogenous variables. For example, there may be a common underlying factor which causes both an increase in account balance and default. Then, a high balance is not a cause of default, but it may be found to be an important driver of default in the model. The shorter the lag period, the more likely this connection will be, which is a further reason for preferring longer lags. For this reason, we focus on a BV model with 12 month lag. Nevertheless, we note that, although endogeneity affects the identification of the cause, it does not affect the forecasts, which are the main concern of this paper.

The implications of the lag term l for the MVs are 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 can use MV values at the time of default. In particular, since we define an account as being in *default* when it is recorded as having failed to make the minimum payments for three consecutive months or more, where the time window over which the missing payments were recorded changes over time as payments are made, we use a 3-month lag on MVs, to correspond to the beginning of missed payments leading to default.

3.2. *Forecasting procedure*

Credit risk models can be used to explore causal hypotheses of consumer credit behaviour; for example, Calhoun and Deng (2002) explore the dynamics and causes of mortgage terminations. However, for financial institutions and regulators, these models typically have value for estimating the risk to individual accounts or losses on credit portfolios. In this way, banks can assess possible future losses and calculate capital requirements as buffers against adverse loss (Board of Governors, 2009). It is in this forecasting capacity that we assess these models. We divide the panel credit data set into an in-sample *training data set* T , and a post, out-of-sample *test data set* S . Models of default are built on T and forecasts of default are evaluated on S . The procedure is as follows:

1. Given a ratio r , accounts are sampled randomly, ensuring that the ratio of the number of accounts in T to S is approximately r .
2. Given an observation date Ω , accounts in T are right censored so that only account records prior to Ω are included (i.e., $t + a_i \leq \Omega$).
3. Accounts in S are included only if they were opened prior to Ω (i.e., $a_i \leq \Omega$), but they are left-censored so only post-observation date records are included (i.e., $t + a_i > \Omega$).

This procedure is intended to mimic the operational situation, where a model is built at a particular time (corresponding to the observation date) using past training data, and forecasts ahead for accounts that already exist on the books. A single model is built based on the fixed training data T , even though the forecast period may be long enough to allow us to refresh with

new data, since this reflects the industry practice of using a credit scoring model for several years before renewal; it also allows us to report consistently on a single fixed model structure.

This procedure does result in a large number of records being removed, but the random sampling ensures that no bias is introduced when generating the out-of-sample test set, whilst the censoring ensures that all predictions are forecasts. As a practical point, financial institutions could also use post, *in-sample* data sets for forecasts, and these may well give more accurate results. However, for this exercise, in order to avoid the introduction of bias, the forecasts are restricted to an out-of-sample data set (Granger & Huang, 1997).

3.3. *Performance measures*

Since we are using survival models which model the time to default, the usual predictive performance measures for classification algorithms, such as error rates and the Gini coefficient, do not naturally apply, and nor do the standard residuals for regression, such as mean square errors. Therefore, we use the deviance (the log-likelihood ratio) to measure the model fit for each model separately, and also to test the goodness-of-fit for nested models. The contribution of each individual i to the log-likelihood function for discrete survival analysis using the logistic function is

$$\begin{aligned} L_i &= \sum_{s=1}^{t_i^*} d_{is} \log P_{is} + (1 - d_{is}) \log(1 - P_{is}) \\ &= \delta_i \log P_i^* + (1 - \delta_i) \log(1 - P_i^*) + \sum_{s=1}^{t_i^*-1} \log(1 - P_{is}) \end{aligned} \quad , \quad (5)$$

where t_i^* is the last observation available for account i and $P_i^* = P_{it_i^*}$ denotes the hazard probability of this last observation, remembering that only the last observation can fail within the survival analysis framework. From Eqs. (2) and (5), it then follows that the *deviance residual* is

$$-L_i = r_{Ci} - \delta_i \log(P_i^* / (1 - P_i^*)), \quad (6)$$

where $r_{Ci} = -\log \hat{S}_i(t_i^*)$ is the Cox-Snell residual and $\delta_i = d_{it_i^*}$ indicates default.

The deviance residual is used to assess predictions at the account level. However, our models can also be used to forecast at an *aggregate* level, e.g., across accounts within a single portfolio. The observed default rate for an aggregate of N accounts at a particular calendar date c is given by

$$D_c = \frac{1}{N} \sum_{i=1}^N d_{i(c-a_i)}, \quad (7)$$

which implies that the estimated default rate forecast given by a particular model is

$$E(D_c) = \frac{1}{N} \sum_{i=1}^N P_{i(c-a_i)}. \quad (8)$$

The difference $E(D_c) - D_c$ then gives a measure of performance for aggregate forecasts.

3.4. Stress testing

We consider a simulation-based stress test of default rate on an aggregate of accounts using Monte Carlo simulation (see for example Marrison, 2002). Over m iterations, the procedure is as follows.

1. Build a dynamic model with MVs from a training data set.

2. Generate a simulation of economic conditions using values of MVs based on a distribution of historic macroeconomic data.
3. Simulate default events on test data by substituting the simulated MV values into the model.
4. Repeat steps 2 and 3 m times, to build a distribution of estimated default rates (DR) over different economic scenarios.
5. Use the DR distribution to compute the estimated DR, given extreme economic circumstances.

Stress tests should consider *unexpected* but *plausible* events. When m is large, a sufficient number of extreme events can be simulated to meet the first criteria; basing the simulations on historical data ensures the second. Notice that if we assume a constant loss given default and exposure at default for each account, the DR distribution is proportional to the distribution of expected losses. Further explanations of the steps are given below.

In step 1, we use the discrete survival model described in Section 3.1, although this is a general stress testing method, and other dynamic model structures are possible.

In step 2, simulated values of MVs are drawn randomly from the available time series of historic MVs. However, if the structure of dependencies between the MVs is not taken into account, this will yield implausible scenarios and lead to misleading results. Therefore, the Cholesky decomposition is used (Marrison, 2002) in order to preserve the covariance structure between MVs. If $\mathbf{V}$ is a matrix of covariances between time series of historic macroeconomic data, then it is decomposed by a lower triangular matrix $\mathbf{L}$, such that $\mathbf{V} = \mathbf{L}\mathbf{L}^T$. Then, if $\mathbf{u}$ is a sequence of

values generated independently from the standard normal distribution, $\mathbf{z}^* = \mathbf{L}\mathbf{u}$ will follow the covariance structure of $\mathbf{V}$, and so can be used as plausible economic simulations. The Cholesky decomposition assumes that the variables are normally distributed. However, this is not usually the case for MVs, and so we apply a transformation to MVs prior to simulation, if required. Therefore, a Box-Cox transformation is used, since this often produces an approximately normal distribution (Box & Cox, 1964). Alternatively, an empirical probit transformation is used to impose a normal distribution on the historical data prior to using a Cholesky decomposition.

In step 3, given a simulated vector of MVs, $\mathbf{z}^*$, and using the latent variable model of logistic regression (Verbeek: 2004, Section 7.1.3), a default event for observation i is simulated as

$$d_{it}^* = \mathbf{I}\left(\hat{\alpha} + \boldsymbol{\varphi}(t)^T \hat{\boldsymbol{\beta}}_1 + \mathbf{w}_i^T \hat{\boldsymbol{\beta}}_2 + \mathbf{x}_{i(t-k)}^T \hat{\boldsymbol{\beta}}_3 + \mathbf{z}^{*T} \hat{\boldsymbol{\beta}}_4 + e_{(i)} > 0\right),$$

where $\mathbf{I}(\cdot)$ is the indicator function, and the error $e_{(i)}$ is simulated from the distribution of the link function F . Then, using Eq. (7), the simulated DR for a calendar time period c is computed as

$$\hat{D}_c(\mathbf{z}^*, \mathbf{e}) = \frac{1}{N} \sum_{i=1}^N \mathbf{I}\left(\hat{\alpha} + \boldsymbol{\varphi}(c - a_i)^T \hat{\boldsymbol{\beta}}_1 + \mathbf{w}_i^T \hat{\boldsymbol{\beta}}_2 + \mathbf{x}_{i(c-a_i-k)}^T \hat{\boldsymbol{\beta}}_3 + \mathbf{z}^{*T} \hat{\boldsymbol{\beta}}_4 + e_{(i)} > 0\right), \quad (9)$$

where $\mathbf{e} = (e_{(1)}, \dots, e_{(N)})$ is a vector of N independent error terms, each of which is cumulatively distributed as F . The simulation takes into consideration the error in the model, $\mathbf{e}$, along with changes in macroeconomic conditions. This is natural, since otherwise the point predictions of Eq. (8) are assumed to be exactly correct. Notice also that the fixed $\mathbf{w}_i$ and lagged $\mathbf{x}_{it}$ variables are not simulated, since they are known and are available from the account data.

In step 4, multiple simulations of $\mathbf{z}'_j$ and $\mathbf{e}'_j$ are made for $j = 1$ to m , leading to a series of m estimates of DR, $\hat{D}_c(\mathbf{z}'_j, \mathbf{e}'_j)$, which form an empirical distribution. For a further discussion, suppose that the simulations are ordered by descending values of the estimated DR; i.e., for all $h \leq j$, $\hat{D}_c(\mathbf{z}'_h, \mathbf{e}'_h) \geq \hat{D}_c(\mathbf{z}'_j, \mathbf{e}'_j)$.

In step 5, extreme values are computed for stress testing. We consider two measures: *Value at Risk* (VaR) and *expected shortfall*. VaR is defined as the maximum expected loss, within a certain time period, for a given percentile, q , and the VaR (at $100(1 - q)\%$) of DR here is approximated as $\hat{D}_c(\mathbf{z}'_{[qm]}, \mathbf{e}'_{[qm]})$. However, VaR captures the worst loss in normal circumstances, whereas stress tests should consider losses during unusual circumstances. Therefore, VaR may not be an appropriate measure of loss during adverse conditions, and may be too conservative (BIS, 2005). For this reason, we also consider the expected shortfall as a measure of loss. This is given as the expected DR in the upper q percentile of the DR distribution, for a given q . Therefore, the Monte Carlo approximation of the expected shortfall DR is

$$S_q \approx \frac{1}{qm} \sum_{j=1}^{[qm]} \hat{D}_c(\mathbf{z}'_j, \mathbf{e}'_j). \quad (10)$$

The number of iterations m is chosen so that both the VaR and expected shortfall converge to stable values.

4. Data

4.1. *Credit card data*

We have three large data sets of UK credit card data covering the period from 1999 to mid-2006 and comprising over 750,000 accounts. All of the data sets include AVs taken at the time of application, along with monthly account behavioural records. Most of the data are collected in the same way and have the same objective meaning between credit card products, although the distributions vary, since different products will have different demographic and risk profiles. Some of the data have already been used for analysis by Bellotti and Crook (2012). Variables that may be defined differently for each product have not been used. A list of the variables used is given in Table 4. Categorical variables for employment and payment status are included as a series of indicator variables. Age is divided into a series of age category indicator variables, since age has a non-linear relationship with default. All of the monetary values, such as income and balance, are given as log values, in order to normalize their distributions. There is a small proportion of missing values in the monthly payment amount, so an indicator variable for these values is also included. Also, there is a large proportion of zero values for some BVs, such as payment amount, sales amount and APR, so indicator variables are included for those cases too. In these experiments, we define an account as being in *default* when it is recorded as having failed to make the minimum payments for three consecutive months or more, where the time window over which the missing payments were recorded changes over time as payments are made. This is a common definition in the industry, and follows the Basel II convention of 90 days delinquency for consumer credit (BCBS 2006). The data which we use for our analysis are commercially sensitive, and therefore we cannot provide further details or data description statistics, or report the observed default rates. It is possible that two or more credit card accounts

may relate to the same person. However, we expect this to be very rare, and unlikely to affect model specification significantly.

To assess the forecasts, an observation date of 1 January 2005 is set. Since the dataset runs to mid-2006, this provides up to 18 months of test data, which is a good period for forecasts, whilst allowing for a long run of training data. After censoring, using the procedure described in Section 2.2, this gives over 400,000 and 150,000 accounts in the training and test sets respectively, thus providing sufficient observations for training whilst leaving a good number of accounts out-of-sample for forecasting.

4.2. Historic UK macroeconomic data

We consider several UK MVs which we expected to have an effect on the PD. These are listed in Table 1. Bellotti and Crook (2009) found that bank interest rates, earnings, the production index and the house price were statistically significant explanatory variables of UK defaults, so we include these. The production index is used instead of GDP because it is available monthly, whereas the GDP figures are only provided quarterly. Gerardi et al (2008) also found the unemployment rate to be significant for US defaults, and in a study of a number of stressed credit markets worldwide, using a dynamic model, Breeden and Thomas (2008) found that variables for consumer sales and prices were correlated with default and bankruptcy. Therefore, we also include MVs for these risk factors. In addition, the FTSE index and a consumer confidence index are also included, since they may be good indicators of confidence in the economy.

TABLE 1 HERE

To remove seasonality, we took twelve-month differences of the MVs. In addition, log values are taken for the MVs with clear exponential growth: earnings, FTSE and house prices. For stress testing, historical values of MVs from 1986 to 2004 are taken; i.e., only MV data prior to the observation date are included.

The observation date of 1 January, 2005, gives training data between 1999 and 2004. During this period, there were moderate changes in the economy: the production index rose in 1999 and peaked in 2000 (104.7), after which time it dropped (to 97.3), then stayed steady over the remaining period (ending at 101); the unemployment rate was generally falling in 1999 (from 1,800,000 to 1,431,000), after which it rose slightly in 2003 (to 1,561,000), before falling again (to 1,400,000); bank interest rates fell from 6% in 1999 to 3.5% in 2003, before rising back to 4.75% in 2004. Table 2 summarizes the MV statistics for this model build period only. Inevitably, in contrast to Table 1, the standard deviations on the variables are smaller (by approximately half), but still show some movement in values over the model training period.

TABLE 2 HERE

5. Results

We present our results in five subsections. Firstly, we present the underlying hazard rate for default. Secondly, we discuss coefficient estimates from the model build. Thirdly, we present the model fit and forecasting results at the account level. Fourthly, we give forecast results at the aggregate level. Lastly, we present results for stress testing.

5.1. Hazard rate for default

The duration-only model provides initial baseline hazards. Figure 1 shows the shape of the hazard probability over time. It has the typical survival profile for consumer credit: PD peaks early, at 8 months, then slowly declines over time, as those who are highly likely to default drop out. This structure has been reported previously by others; see for example Gross and Souleles (2002) and Andreeva (2006). Figure 1 also shows a second small increase in the hazard, peaking at around 36 months. This is because, for all credit card products, accounts with no recent usage are removed from the portfolios after two years. Since these are typically low risk accounts, their removal leads to a small overall increase in the default risk. It is therefore important to realize that the hazard rate is not just an indication of the obligor's propensity to default, it will also be influenced by periodic operational decisions made by portfolio managers. The hazard probability estimate in Figure 1 is based on a sample size of over 60,000 observations for all months of account age (3 to 48) given on the horizontal axis. For the sake of commercial confidentiality, the precise number of observations for each account month cannot be given.

FIGURE 1 HERE

5.2. Model and coefficient estimates

Many AVs and BVs and several MVs are statistically significant explanatory variables. We focus on the model for the BV lag 12 months, since this is the most practically valuable model in terms of the forecast range. Table 3 shows coefficient estimates for this model. We find the following key outcomes.

1. The signs on the current balance (log) and its square are opposite, but the positive sign on the square term dominates. Therefore, the balance outstanding on the account has an increasing positive effect on the default hazard. This is unsurprising, since a larger balance will be more difficult to clear. Second, an increase in the credit limit reduces the hazard. Initially, this may appear surprising, since we might argue that a high credit limit encourages a higher balance, and therefore a greater risk. However, at least in the short term, a high credit limit provides the obligor with a buffer, enabling the obligor to build up debts before reaching default.
2. The bank sets the credit limit based on their own assessment of the obligor's behaviour, so the credit limit acts partly as a proxy for a behavioural score.
3. The amount paid back each month, indicated by the payment status and amount, has a negative effect on default. This is to be expected, since a greater ability to repay implies that default is less likely.
4. The number of transactions has a positive effect on default. This is to be expected, since it indicates a greater card usage, and hence a rising balance. However, interestingly, the transaction sales amount has a negative effect. A possible explanation is that the sales amount acts as an indicator of wealth, when taken together with the 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 are missing, PD decreases considerably. However, since no duration times up to 12 months will have BVs (because of the lag), this is mainly a joint effect with duration.

6. Indicator variables have been added for vintage, indicating the year of account opening. These are significant, implying that cohorts explain some of the effects over time for default rates within the data set. This is natural, since lenders will allow new accounts with higher or lower levels of risk onto their books at different times, depending on their changing attitude to risk at different times in the business cycle.
7. Concerning macroeconomic variables, the interest rate has a positive effect on default. This is to be expected, since rising interest rates imply a greater demand for repayment on outstanding loans and mortgages, which will have an adverse effect on people who are more highly indebted. The unemployment rate also has a positive effect on default. The unemployment rate is an indicator of the direct economic stress on individuals. In particular, obligors who become or remain unemployed will find it more difficult to repay their debt. Conversely, if unemployment decreases, then we would generally expect unemployed obligors to find jobs, making it easier for them to repay. Therefore, the effect of this MV on default is as expected. Only the interest and unemployment rates are statistically significant MVs in the model, so the model can be rebuilt without the other MVs, to determine how they are affecting the parameter estimates of the former. We find similar results: the interest rate remains significant ($p < 0.0001$), although its effect size is smaller (0.0615), and the unemployment rate is also significant ($p < 0.0001$), but its effect size is approximately unchanged (0.000621).

The first three findings corroborate the results of Gross and Souleles (2002), who built dynamic models of default for US credit card data. They found that the risk of default rises with the balance and falls with repayments. They used *utilization* (the outstanding balance divided by the

credit limit) instead of the raw value of the balance, which is sensible, given the relationship discussed in point 2.

The shape of the relationship which we find between the baseline hazard and the duration of holding the card is consistent with that of other studies, but we find the baseline hazard to be at a maximum at a duration time of 8 months, whereas Gross and Souleles (2002), who also consider credit cards, found it to be around 18 months (see their Figure 1). Calhoun and Deng (2002, Exhibit 2) and Willen et al (2007, Figure 12) both consider mortgages and find that the maximum hazard occurs at between 16 and 28 quarters. One possible reason for the difference between our results and those of Gross and Souleles may be that American card holders hold more cards than UK card holders, and so spread their outstanding balance over more cards.

The findings for the MVs are generally consistent with those of other comparable studies. As in our study, significant positive effects of the unemployment rate have been found by both Gross and Souleles (2002) when modelling the probability of bankruptcy, and Willen et al. (2008) when modelling the probability of mortgage default. Bellotti and Crook (2009), using a different data set, also found a positive sign for unemployment, but it was not significant. Gross and Souleles (for bankruptcy), Willen et al., and Bellotti and Crook all found a positive and significant effect for interest rates, consistent with our result in point 7. However, when Gross and Souleles (2002) modelled delinquency by credit card holders using a definition very similar to that which we use for default, they did not find that either the unemployment rate or house prices were significant predictors. There are a range of differences between our study and that of Gross and Souleles that might account for this difference in findings. Gross and Souleles had a

different set of covariates, and their estimation period was much shorter than ours. Their training data covered only two years, whereas ours extended over at least five years, and so potentially included more information about the relationship between the default probability and the unemployment rate (they did not include an interest rate). Our training data also covers a different period (1999–2006) to their data (mid-1995 to mid-1997).

The correlations we find with the unemployment rate are also consistent with the study by Breeden and Thomas (2008), who use several world-wide data sets, although they also found the GDP to be significant in many cases. We did not include this variable because we were using monthly data and the GDP is not available monthly. They also included vintage effects in their models.

Univariate associations for each MV are also reported in Table 4. This indicates that almost all MVs have an association with default. However, these MVs are correlated with one another, so they do not all come out as significant variables in the multivariate model.

These results are for the model with BVs lagged 12 months. We found similar results for models with shorter lag periods, and the comments made above also hold in these cases, except that the effects and statistical significance tend to be stronger for models with shorter lags.

TABLE 3 HERE

TABLE 4 HERE

5.3. *Model fit and forecasts of time to default*

Model fits for several alternative models are shown in Figure 2. This shows 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. In addition, we also observe that the model fit improves with shorter lags on the BVs, with a relatively large improvement at a 3-month lag. However, as we have discussed, this improvement comes at the price of a much shorter range of forecasts. We can see in Figure 2 that, although some of the MVs are statistically significant, their contribution to model fit is weak. The fit of the nested model is also assessed, with the results shown in Table 5. This shows that all model extensions lead to a statistically significant improvement in fit, particularly the basic duration time only model and the inclusion of BVs and MVs.

FIGURE 2 HERE

TABLE 5 HERE

Figure 2 also shows the results of the forecasts. These follow the model fit results very closely. They show a marked improvement in fit for the BV models, improving with shorter lags. However, there is no noticeable change in forecast accuracy when MVs are included.

5.4. *Estimation of default rates*

Figure 3 shows the estimated monthly default rates for various different models, along with the observed (or true) monthly default rates for each month of the test data set. The observed default rates have high variance, but there is a general trend of high rates beginning in 2005, falling during 2005, then rising again in 2006. The AV model is able to model the general fall in default rates, but BVs are required to forecast the overall trend, including the rise in 2006. However, the

best forecasts indicating a high DR in early 2005 and mid-2006, whilst also forecasting the dip in default rates at the end of 2005, are only made when MVs are included in the model. The BV model, lag 3 months, also performs well, but this is not surprising, given the short forecast period, using behavioural data just one month before accounts begin missing payments. Overall, the BV lag 12 month model with MVs performs best at forecasting the aggregate DR, achieving better results than even BV models with shorter lags, as demonstrated in Table 6.

FIGURE 3 HERE

TABLE 6 HERE

5.5. *Stress test results*

We ran Monte Carlo simulations using the MV model given in Table 3. The estimated DR was simulated on the test data set, 12 months after the observation date; i.e., for December 2005. A stable default rate distribution was generated after $m = 25,000$ simulations, and is shown in Figure 4. Notice, firstly, that the observed DR is positioned centrally within the distribution, close to the median. This demonstrates that this period did not represent extreme default rates. The right-hand tail shows the risk for more adverse conditions. In particular, we have included the figure for expected shortfall at the 99th percentile. This shows that, for the worst 1% of economic scenarios we consider, the expected DR is 1.73 times greater than normal conditions (i.e., the median estimated DR). The VaR is also shown for comparison. We see that this gives a lower estimate of DR (1.59 times the median), which may not reflect extreme circumstances sufficiently. These figures are slightly higher than those suggested as part of the US stress testing exercise by the Board of Governors (2009). In particular, the Board of Governors study estimates a more modest rise in DR of between 20% and 55%, when contrasting a normal

“baseline” figure to “more adverse” conditions.¹ However, our results appear lower than those of Rösch and Scheule (2004), who found the VaR for US credit cards to be 2.31 times the mean, although they used aggregate data, not account level data.

FIGURE 4 HERE

6. Conclusion

Dynamic models form a flexible approach to modelling and forecasting consumer credit risk. They have a number of well-known advantages over static models, including modelling the conditional PD in a specific time period rather than in a time window, and enabling the prediction of the profitability of specific loans (Bellotti & Crook, 2009). In particular, we have used *discrete-time* survival analysis to model credit card risk. This has two main advantages. Firstly, it is a principled means of building dynamic models of default using accounting records; and, secondly, modelling and forecasting are computationally efficient when compared to *continuous-time* survival models. This is important when model builders use large databases of credit accounts.

We have used a large data set of UK credit card accounts to test the effectiveness of discrete survival models with BVs and MVs as models of default. Unlike the previous literature, we explore the use of these models as tools for risk measurement, forecasting and stress testing. We

¹ The Board of Governors (2009) gives baseline two-year loss rates as 12–17%, and “more adverse” rates as 18–20%. Taking the lower and upper bounds of each range and converting to an average monthly DR gives a 20–55% expected increase in loss. Taking a mean value for the baseline and more adverse rates (14.5% and 19% respectively) gives a mean increase of 34%.

conclude that many BVs are statistically significant explanatory variables of default, and including them gives improved model fit. Important BVs are the account balance, repayments, the number of transactions within each month, and the credit limit. We find that an improved model fit translates into improved forecasts of the time to default. The performance improves with shorter lags on BVs. This is to be 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 levels of endogeneity between BVs and the default event. For this reason, we focus on lag 12 month BVs. This gives an improved performance relative to the AV only model, and also allows for useful forecasts up to 12 months ahead.

We found that bank interest rates and the unemployment rate affect the hazard significantly. Whilst their inclusion gave only a modest improvement in model fit and no noticeable improvement in forecasts of the time to default at the account level, their inclusion improves forecasts of default rate at the aggregate level. This is understandable, since MVs affect all predicted PDs, rather than at the individual account level. Hence, their effect will only become noticeable at the aggregate level where accounts are taken together. Where they are comparable, our results corroborate the results obtained by others (Gross & Souleles, 2002; Calhoun & Deng, 2002; Gerardi et al., 2008).

Third, the inclusion of MVs enables us to use stress tests, which generate credible results, indicating that adverse conditions may raise the default rate by around 79%. We used a simulation-based approach for our experiments, but scenarios could also be designed and used with these models.

The following further issues are suggested as developments from this study:

1. The data cover a period of relatively benign economic conditions (1999–2006), although there is some evidence of UK default rates increasing toward the end of this period. It would be interesting to follow up this study using new data through the period of the credit crunch (2008–2010) and beyond. Nevertheless, this study shows that credible models with macroeconomic variables can be built using data from less extreme economic conditions. It may well be that a model change occurs after the credit crunch period, in which case it would be beneficial to have separate models built across the pre- and post- credit crunch periods.
2. The inclusion of MVs may explain some of the correlations between accounts over calendar time. However, it is possible that not all borrower correlations are explained, and a latent variable over time may be of benefit. This was suggested by Bellotti and Crook (2012), who implement an asset correlation model and show that including MVs explains some of the correlation, but not all.
3. In this paper, we have investigated the problem of forecasting and stress testing default rates. However, economic capital is often defined as the difference between the VaR and the mean of the Expected Loss distribution. The calculation of the Expected Loss for an obligor involves a model of loss given default (LGD) and exposure at default (EAD), as well as PD, along with their correlations. In effect, this paper assumes a constant LGD and EAD. An interesting development of this work would be to explore the role of a discrete survival model within the scope of this larger problem.

References

- Allison, P. D. (1995). *Survival analysis using SAS*. SAS Press.
- Andreeva, G. (2006). European generic scoring models using survival analysis. *Journal of the Operational Research Society*, 57, 1180-1187.
- Bank for International Settlements (BIS) (2005). *Stress testing at major financial institutions: survey results and practice*. Working report from Committee on the Global Financial System.
- Basel Committee on Banking Supervision (BCBS) (2006). *Basel II: international convergence of capital measurement and capital standards*. www.bis.org/publ/bcbsca.htm.
- Bellotti, T., & Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60(12), 1699-1707.
- Bellotti, T., & Crook, J. (2012). Asset correlations for credit card defaults. *Applied Financial Economics*, 22, 87-95.
- Berkowitz, J. A. (2000). Coherent framework for stress testing. *The Journal of Risk*, 2(2), 5-15.
- Board of Governors of the Federal Reserve System (2009). *The supervisory assessment program: overview of results*. Washington DC: Board of Governors.
- Boss, M. (2002). A macroeconomic credit risk model for stress testing the Austrian credit portfolio. *OeNB Financial Stability Review*, 4, 64-82.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211-246.
- Breeden, J., & Ingham, D. (2009). Monte Carlo scenario generation for retail loan portfolios. *Journal of the Operational Research Society*, 61, 399-410.
- Breeden, J., & 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.
- Breuer, T., Jandacka, M., Mencia, J., & Simmer, M. (2012). A systematic approach to multi-period stress testing of portfolio credit risk. *Journal of Banking and Finance*, 36, 332-340.
- Calhoun, C. A., & Deng, Y. (2002). A dynamic analysis of fixed- and adjustable-rate mortgage terminations. *Journal of Real Estate Finance and Economics*, 24(1/2), 9-33.
- Castren, O., Dees, S., & Zaher, F. (2010). Stress-testing euro area corporate default probabilities using global macroeconomic model. *Journal of Financial Stability*, 6, 64-78.

- Collett, D. (1994). *Modelling survival data in medical research*. London: Chapman & Hall.
- Crook, J., & Bellotti, T. (2010). Time varying and dynamic models for default risk in consumer loans. *Journal of the Royal Statistical Society, Series A*, 173, 283-305.
- Delgado, J., & Saurina, J. (2004). *Credit risk and loan loss provisions. Analysis with macroeconomic variables*. Bank of Spain: Working Paper.
- Drehmann, M., Sorensen, A., & Stringa, M. (2010). The integrated impact of credit and interest rate risk on banks: a dynamic framework and stress testing application. *Journal of Banking and Finance*, 34, 713-729.
- Drehmann, M., Patton, A. J., & Sorensen, S. (2005). *Corporate defaults and large macroeconomic shocks*. Bank of England Working paper.
- Financial Services Authority (FSA) (2008). *Stress and scenario testing*. Consultation paper 08/24 FSA: UK.
- Gerardi, K., Shapiro, A. H., & Willen, P. S. (2008). *Sub-prime outcomes: risky mortgages, homeownership experiences, and foreclosures*. Working paper 07-15, Federal Reserve Bank of Boston.
- Granger, C. W. J., & Huang, L. L. (1997). *Evaluation of panel data models: some suggestions from time series*. Discussion paper 97-10, Department of Economics, University of California, San Diego.
- Gross, D. B., & Souleles, N. S. (2002). An empirical analysis of personal bankruptcy and delinquency. *The Review of Financial Studies*, 15(1), 319-347.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton University Press.
- Huang, X., Zhou, H., & Zhu, H. (2009). A framework for assessing the systemic risk of major financial institutions. *Journal of Banking and Finance*, 33, 2036-2049.
- Jokivouille, E., & Viren, M. (2011). Cyclical default and recovery in stress testing loan losses. *Journal of Financial Stability*, 9(1), 139-149.
- Ma, P., Crook, J., & Ansell, J. (2009). Modelling take-up and profitability. *Journal of the Operational Research Society*, 61, 430-442.
- Marrison, C. (2002). *Fundamentals of risk measurement*. New York: McGraw-Hill.
- Peseran, M. H., Schuermann, T., Treutler, B.-J., & Weiner, S. M. (2006). Macroeconomic dynamics and credit risk: a global perspective. *Journal of Money, Credit and Banking*, 38(5), 1211-1261.

- Rösch, D. (2003). Correlations and business cycles of credit risk: evidence from bankruptcies in Germany. *Swiss Society for Financial Market Research*, 17(3), 309-331.
- Rösch, D., & Scheule, T. (2004). Forecasting retail portfolio credit risk. *Journal of Risk Finance*, Winter/Spring, 16-32.
- Rösch, D., & Scheule, H. (2008). Integrating stress-testing frameworks. In D. Rösch & H. Scheule (eds.), *Stress testing for financial institutions*. Risk Books.
- Sorge, M., & Virolainen, K. A. (2006). Comparative analysis of macro-stress testing methodologies with application to Finland. *Journal of Financial Stability*, 2, 113-151.
- Therneau, T. M., Grambsch, P. M., & Fleming, T. R. (1990). Martingale based residuals for survival models. *Biometrika*, 77, 147-160.
- Thomas, L. C., Ho, J., & Scherer, W. T. (2001). Time will tell: behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, 12, 89-103.
- Verbeek, M. (2004). *A guide to modern econometrics* (2nd ed). Chichester:Wiley.

Table 1.

Descriptive statistics and sources for macroeconomic variables for the period 1986 to 2004.

Macroeconomic Variable	Source	Descriptive statistics (for difference in value over 12 months)			
		Min	Mean	SD	Max
UK bank interest rates	ONS	−4.5	−0.387	1.89	6.5
UK unemployment rate (in '000s) SA	ONS	−535	−91.2	232	575
UK production index (all)	ONS	−5.2	1.09	2.26	6
Retail sales value	ONS	0.3	3.92	1.46	8.5
FTSE 100 all share index	FTSE	−822	85.1	280	682
Halifax House Price index	LBG	−9.37%	+8.64%	9.98%	+34.7%
Retail price index (all items)	ONS	1.2	5.02	2.23	12.8
Earnings (log) all including bonus	ONS	0.00820	0.022	0.00786	0.0384
Consumer confidence index	EC	−20.3	0.725	7.61	18.1

Sources: UK Office of National Statistics (ONS), Lloyds Banking Group (LBG) and the European Commission (EC). The data are monthly and may be seasonally adjusted (SA).

Table 2.

Descriptive statistics for macroeconomic variables for the period of model build: 1999 to 2004.

Macroeconomic Variable	Descriptive statistics (for difference in value over 12 months)			
	Min	Mean	SD	Max
UK bank interest rates	−2.5	−0.469	1.06	1.25
UK unemployment rate (in '000s) SA	−207	−61.8	70.6	102
UK production index (all)	−4.7	0.0458	2.01	3.7
Retail sales value	0.7	4.21	1.79	8.5
FTSE 100 all share index	−822	−64.4	380	568
Halifax House Price index	+1.63%	+14.3%	6.83%	+31.5%
Retail price index (all items)	1.2	3.97	1.4	6.4
Earnings (log) all including bonus	0.00895	0.0178	0.00398	0.0316
Consumer confidence index	−12.6	−0.0667	4.96	10.4

Table 3.

Coefficient estimates for the model with all AVs, BVs lag 12 months and MVs.

Covariate	Estimate	Standard error
Intercept	n/a**	
Duration	1.35**	0.0467
“ (squared)	−0.00698**	0.0003
“ (log)	16.4**	0.467
“ (log squared)	−6.42**	0.198
Selected application variables (AV)		
Time customer with bank (years)	−0.00250**	0.000084
Time with bank unknown +	−0.342**	0.0321
Income (log)	−0.146**	0.0127
Income unknown +	−1.46**	0.123
Number of cards	−0.0610**	0.00782
Time at current address	−0.00129	0.000973
Employment + :		
Self-employed	0.303**	0.0244
Homemaker	0.072	0.0512
Retired	0.111	0.0452
Student	−0.035	0.0431
Unemployed	0.231	0.113
Part time	−0.365**	0.0383
Other	−0.037	0.0253
<i>Excluded category: Employed</i>		
Age + : 18 to 24	0.074	0.0301
25 to 29	−0.058	0.0283
30 to 33	0.010	0.0288
34 to 37	0.100**	0.0287
38 to 41	0.046	0.0302
48 to 55	−0.108**	0.0307
56 and over	−0.243**	0.0381
unknown	−2.74**	0.0548
<i>Excluded category: 42 to 47</i>		
Credit bureau score	−0.00322**	0.000043
Product + : A	0.535**	0.0273
B	0.371**	0.0189
<i>Excluded category: C</i>		
Vintage (+): 1999-2003	n/a **	
Behavioural variables (BV) lag 12 months		
Payment status + :		
Fully paid	−0.390**	0.0535
Greater than minimum paid	−0.090**	0.0271

Minimum paid	0.149**	0.0345
Less than minimum paid	0.714**	0.0445
Unknown	−0.148*	0.0496
<i>Excluded category: No payment</i>		
Current balance (log)	−1.58**	0.0923
“ (log squared)	0.517**	0.0184
“ is zero +	−1.05**	0.123
“ is negative +	−0.802**	0.137
Credit limit (log)	−1.22**	0.0273
Payment amount (log)	−0.154**	0.025
“ is zero +	−0.133	0.0535
“ is unknown +	−0.452**	0.0706
Number of months past due	0.134*	0.0459
Past due amount (log)	0.0795	0.0605
“ is zero +	−0.623**	0.0981
Number of transactions	0.00663**	0.00193
Transaction sales amount (log)	−0.350**	0.0246
“ is zero +	−0.567**	0.0489
APR on purchases	−0.00487	0.00194
“ is zero +	−0.482**	0.0438
Behavioural data is missing +	−3.73**	0.123
Macroeconomic variables (MV) lag 3 months		
Bank interest rate	0.113**	0.019
Unemployment rate	0.000672**	0.000246
Production index	−0.0101	0.0063
FTSE all 100 (log)	0.0591	0.0743
Earnings (log)	1.57	2.11
Retail sales	0.00929	0.00504
House price (log)	−0.218	0.656
Consumer confidence	−0.00217	0.00197
Retail price index (RPI)	−0.0298	0.0151

- a) Indicator variables are denoted by a plus sign (+).
- b) Statistical significance levels are denoted by asterisks: ** is less than 0.001 and * is less than 0.01.
- c) Coefficient estimates on the intercept and vintages are not shown, for reasons of commercial confidentiality. Only selected application variables are included, for the same reason.

Table 4.

Coefficient estimates for the univariate association between the MVs and default.

Macroeconomic variables (MV) lag 3 months		
	Estimate	Standard error
Bank interest rate	0.0865 **	0.00791
Unemployment rate	0.000287 **	0.000098
Production index	0.000849	0.00352
FTSE all 100 (log)	0.334 **	0.0372
Earnings (log)	-0.280	1.67
Retail sales	0.0349 **	0.00372
House price (log)	3.25 **	0.300
Consumer confidence	0.00947 **	0.0014
Retail price index (RPI)	0.0311 **	0.00528

Statistical significance levels are denoted by asterisks: ** is less than 0.001 and * is less than 0.01.

Table 5.

Model fit.

Nested model	Compared to base model	Difference in 2 × LLR	Number of added covariates	p-value
Duration only	Null model	2438	4	<0.0001
AV only	Duration only	22531	34	<0.0001
AV & BV lag 12	AV only	9237	22	<0.0001
AV, BV lag 12 & MV lag 3	AV & BV lag 12	47	9	<0.0001
AV & BV lag 9	AV only	7946	22	<0.0001
AV & BV lag 6	AV only	12458	22	<0.0001
AV & BV lag 3	AV only	26559	22	<0.0001

The results are for nested models using the difference in LLR and a chi-square significance test.

Table 6.

Mean absolute difference between the estimated and observed default rates across the test set.

Model	Mean absolute difference between estimated and observed DR
AV only	0.087
BV lag 12	0.058
BV lag 12 & MV lag 3	0.049
BV lag 9	0.062
BV lag 6	0.070
BV lag 3	0.068

These results relate to models based on the 18 months of test results shown in Figure 4.

Figure 1.
Baseline hazard rate function for the parametric duration model of default.

For reasons of commercial confidentiality, the hazard probability scale is not shown.

Figure 2.

Model fit and forecast performance for different models.

Figure 3.
Comparison of estimated and observed default rates for each month of the test data set.

For reasons of commercial confidentiality, the scale on the default rate axis is not shown.

Figure 4.
Distribution of estimated default rates

The distribution is based on a simulation of economic scenarios for credit card accounts during December 2005, based on a model with MVs trained on data prior to January 2005, shown as a histogram. The observed DR for the test data set is shown, along with the Value at Risk (VaR) and expected shortfall at 99% probability. All values are expressed as a ratio of the median estimated DR.