Forecasting and Stress Testing Credit Card Default using Dynamic Models

Tony Bellotti and Jonathan Crook

Credit Research Centre
University of Edinburgh Business School

Version 4.5

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 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. Additionally, by simulating extreme economic conditions, we show how these models can be used to stress test credit card portfolios.

JEL: G21, D14.

Keywords: discrete survival models, stress testing, loss distributions, choleski decomposition, credit risk.

Address for correspondence: Credit Research Centre, University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH10 4SW, UK.
{j.crook,t.bellotti}@ed.ac.uk

Acknowledgement: This work was partially funded by the EPSRC grant EP/D505380/1.

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 collected at time of application that then remain fixed. Typically, this is information taken from a completed application form and a credit score for the individual provided by a credit bureau. But such models are restrictive and models which answer more specific questions can be estimated from credit portfolios because the latter provide panel data (Crook and Bellotti: 2009) for a sample 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 included 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 would 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 and this requires the use of dynamic rather than static models (Thomas et al :2001, Ma et al :2009). In this paper, we present dynamic models of default which include behavioural variables (BV) and macroeconomic variables (MV) in addition to application variables (AV).

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. Accurate stress tests are becoming increasingly important in evaluating the risk to banks as is evident by the recent evaluation of US banks (FRS 2009) and the recognition by the FSA (2008) that stress testing is a key tool in helping financial institutions make business strategy, risk management and capital planning decisions. However, very few stress test results for retail loan portfolios have been published. Breedan and Ingram (2009) discuss issues involved in generating scenarios using a model where the default rate of a portfolio over time is explained in terms of a function of 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 and it is debatable whether simulating a parameterised function of calendar time is the same as simulating macroeconomic variables that are related to the probability of default at account level. Rösch and Scheule (2004) assume a Merton one factor model and estimate loss distributions for credit cards, mortgages and other consumer loans in the US. But they use aggregate default rate data and omit variables specific to the obligor. It is also unclear how they preserved the correlation structure between the MVs in their model. In this paper we use Monte Carlo simulation to generate loss distributions of estimated default rates as the basis of a stress test of our credit card data. Boss (2002) uses a similar dynamic model structure for simulation-based stress tests but on corporate loans.

Several modelling techniques have been proposed to develop a dynamic model of credit (see Crook and Bellotti: 2009 for a review). Thomas et al (2001) describe how to use a Markov chain stochastic process 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 allowing modelling of different risk groups. They also describe survival analysis as a means to build dynamic models since this 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 using a survival model. However, in contrast with these papers which use continuous time models, we use *discrete* 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 already occurred on that account. Since credit data is usually in the form of 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.

Discrete survival models have been applied successfully in the analysis of personal bankruptcy and delinquency in the USA (Gross and Souleles 2002), mortgage terminations (Calhoun and Deng 2002) and competing risks of foreclosure and sales in the US subprime market (Gerardi et al 2008). Gross and Souleles (2002) use several BVs and MVs. In particular they included outstanding account balance and repayments and found the former had a positive affect on bankruptcies and the latter had a negative effect. They also found that local unemployment rate had a statistically significant positive effect on bankruptcy, which is what we would expect since an increase in unemployment is likely to affect some obligors adversely. Calhoun and Deng (2002) derive dynamic variables measuring probability of negative equity and mortgage premium. Both change over time and have a positive affect on default. They also include the ratio of 10-year to 1-year Constant Maturity Treasury yield and find it statistically significant for models of early repayment. For fixed-rate mortgages, the coefficient increases for higher ratios; the rationale is that mortgagors are moving to adjustable-rate mortgages to take advantage of the short-term relatively low interest rates. Gerardi et al (2008) found that interest rates (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.

All these studies show that both BVs and MVs are useful explanatory covariates for consumer credit risk. However none of them report using these dynamic models for forecasts or stress testing. 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 adverse conditions. For this reason, we focus primarily on using the models for forecasting PD and default rate. For a large database of UK credit cards we establish the following new results. First, the inclusion of BVs improves model fit and also improves forecasts. The best results are achieved with the most recent behavioural data. Second, several MVs are found to be statistically significant explanatory variables of default, but this does not translate into improved forecasts at the account-level. Third, we show that including MVs does improve estimation of loss (default rate) at a portfolio or aggregate level. Fourthly, we show how models with MVs can be used for stress testing and report a loss distribution 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 treat time as being discrete and adopt the following notation. We denote calendar time as c and a_i is the date that account i was opened. 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 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 each account 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

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

where 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 and in particular, we use the transformation $\boldsymbol{\varphi}(t) = \left(t, t^2, \log t, (\log t)^2\right)$ 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 $F(x) = 1/\left(1 + e^{-x}\right)$.

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 defined as default. It can be estimated using standard maximum likelihood estimation for logistic regression (Allison 1995).

Coefficient estimates on duration $\boldsymbol{\varphi}(t)$ give a baseline hazard. If it included dummy variables for each discrete time then the coefficients would form a non-parametric baseline hazard and model (1), overall, would be a semi-parametric model paralleling the commonly used Cox proportional hazard model. Published studies suggest there is a common shape to the distribution over duration time of default hazard rates: they rise sharply within the first few months before they begin to fall steadily over the remaining duration of the account (Gross and Souleles 2002, Figure 1, and Andreeva 2006, Figure 1). We use a parametric form for $\boldsymbol{\varphi}$ since this allows us to capture this structure of hazard over time. Log terms are included 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

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

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

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 consider a 3 month lag model, even though this is not such 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 a further reason why longer lags are preferable, and so 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 implications of the lag term l on 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 time of default. In particular, since we define default as 3 consecutively missed payments, we use 3 months lag on MVs to correspond with the beginning of missed payments leading to default.

2.2 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 estimation of 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 (FRS 2009). It is in this forecast capacity that we assess these models. Following Granger and Huang (1997) 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 estimates of default are measured on S. In detail, accounts are randomly sampled so that the ratio r of number of accounts in T to S is fixed. Then given a calendar date at the time of observation,

Ω, accounts in T are right censored so they were opened prior to the observation date and only those records prior to the observation date are included (ie $t + a_i \leq \Omega$). Set S also includes accounts opened prior to the observation date (since in a real-world situation accounts opened after would be unknown) but only the post-observation date records are in the test set (ie $t + a_i > \Omega$). That is they are left censored. This procedure unfortunately results 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 all predictions are forecasts. As a practical matter, financial institutions could use post, *in-sample* data sets for forecasts and these may well give more accurate results. However, for this exercise, to avoid over-fitting and the introduction of bias, forecasts are restricted to an out-of-sample data set (Granger and Huang 1997).

2.3 Performance measures

We use the log-likelihood ratio (LLR) to measure model fit for each model separately and also to test goodness-of-fit for nested models. But LLR only measures model fit on the training sample and not accuracy of forecasts. Since we are using survival models which model time to default, the usual predictive performance measures 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 related to how well the estimated survival probability matches the observed (true) time of default. In particular, these residuals take account of censored data. One such useful measure is the deviance residual given by

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

where $r_{Mi} = \delta_i - r_{Ci}$ is the martingale residual and $r_{Ci} = -\log \hat{S}_i\left(t_i^*\right)$ is the Cox-Snell residual where t_i^* is the last observation available for account *i* and $\delta_i = d_{it_i^*}$ indicates whether it failed. The martingale residuals 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 approximately symmetrically distributed and the sum of their squares forms the statistic

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

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 *R* gives a measure of log-likelihood deviance of the estimated model from the best case. Therefore models yielding smaller values of *R* give better fit (Collett 1994).

Many of the properties of martingale deviance residuals are proved under the assumptions that, firstly, the survival model is a Cox proportional hazards model and, secondly, that no TVCs are included in the model (Therneau et al 1990). Unfortunately neither assumption is true in our analysis. Nevertheless, the deviance residual is a typical residual for survival analysis and so, supposing robustness, we report deviance calculated over the test set as one of our performance measures. However, we also derive an alternative residual based directly on the log-likelihood of survival for each test case given a model. The residual for each account is the negative of the log of the probability of the series of events for the account, given the model. The lower this is, the better the model's prediction matches outcome. The contribution of each individual i to the log-likelihood function for discrete survival analysis using the logistic function is

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

where $P_i^* = P_{it_i^*}$ denotes the hazard probability of the last observation, remembering that only the last observation can fail within the survival analysis framework. From (2) and (5) it then follows that the *log-likelihood residual* is

$$-L_i = r_{Ci} - \delta_i \log\left(P_i^* / \left(1-P_i^*\right)\right) \tag{6}$$

which, interestingly, is similar to the Martingale residual, except for a change of sign and the account-specific scaling term on defaults $\log\left(P_i^* / \left(1-P_i^*\right)\right)$. The advantage of (6), however, is that it is meaningfully additive since its sum over all individuals is the negative of the log-likelihood statistic.

The deviance and log-likelihood residuals above are 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 default rate for an aggregate of N accounts at a particular calendar date c is given by

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

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

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

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

2.4 Stress testing

We consider a simulation-based stress test of default rate 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 using values of MVs based on 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 loss distribution of estimated DR over different economic scenarios.
5. Use the loss distribution to compute estimated DR for extreme economic 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.

Value at Risk (VaR) is defined as the maximum expected loss, within a certain time period, for a given percentile, q. Sometimes VaR is used to compute stressed values in step 5. However, VaR captures worst loss in normal circumstances, whereas stress tests should consider losses during unusual circumstances. Therefore VaR may not be the appropriate measure of loss during adverse conditions (BIS 2005). For this reason we also consider *expected shortfall* as a measure of loss. This is defined as the expected (mean) loss in the upper q percentile of the loss distribution, for a given q.

Using the latent variable model of logistic regression (Verbeek 2004, section 7.1.3), we can simulate default rates for some calendar time period *c*, given a model, a vector of macroeconomic conditions **z,** and a vector of *N* independent residual terms $\mathbf{e} = (e_{(1)}, \cdots, e_{(N)})$, each cumulatively distributed as *F*, as

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

where $\mathrm{I}(\cdot)$ is the indicator function. Monte Carlo simulation can then be used to approximate expected shortfall default rate with

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

where *j*=1 to *m*, each $\mathbf{z}'_j$ is generated by macroeconomic simulation and $\mathbf{e}'_j$ are generated randomly from F^N and both are indexed so that the simulated default rates are in descending order; ie for all $h \le j$, $\hat{D}_c(\mathbf{z}'_h, e'_h) \ge \hat{D}_c(\mathbf{z}'_j, e'_j)$. The number of iterations *m* is chosen so that (10) converges to a stable value. This simulation takes into consideration the error in the model represented by the residual terms, $\mathbf{e}_j$, along with changes in macroeconomic conditions. This is natural, since otherwise the point predictions of equation (8) 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 $\mathbf{V}$ and so can be used as plausible 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 750,000 accounts. All data sets include AVs taken at time of application, along with monthly account behavioural records. 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. A list of variables used is given in Table 2. 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 to PD. All 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 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. In these experiments, we define an account as in *default* when it 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 default rates.

To assess forecasts, an observation date of 1 January 2005 is set. Since the data 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. We set the training/test data set split ratio r=2/1. 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, providing sufficient observations for training whilst leaving a good number of accounts out-of-sample for forecasts.

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 on a different data set Bellotti and Crook (2008) 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 figures are only provided quarterly. Gerardi et al (2008) also found unemployment rate was significant for US defaults, and Breedon and Thomas (2008) found variables for consumer sales and prices were correlated to default and bankruptcy in a study of a number of stressed credit markets worldwide, using a dynamic model. Therefore, we also include MVs for these risk factors. Additionally, FTSE index and a consumer confidence index are also included since they may be good indicators of confidence in the economy.

TABLE 1 HERE

Many of the MVs have a time trend. For example, earnings and housing price have an obvious upward trend, whereas interest rates have an overall downward trend from 1999-2006. This could be a problem since default rate also has a time trend and so model fit could simply be due to fitting this trend, rather than the macroeconomic condition itself. Therefore, to reduce this possibility, all MVs are included in the model as difference variables over 12 months. Additionally, log values are taken for those MVs with clear exponential growth: earnings, FTSE and house prices. For stress testing, historical values of MVs are taken from 1986 to 2004; ie only MV data prior to the observation date is included.

We experimented using interaction terms between MVs and BVs and AVs, since different risk groups may be more susceptible to economic changes than others. However, as with Bellotti and Crook (2008), we did not find their inclusion improved model fit or forecasts, and indeed made them worse. For this reason we do not report results using interaction terms.

4. Results

We present 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 model fit and forecasting results at the account level. Fourthly, we give forecast results at the aggregate level. And lastly we present results for stress testing.

4.1 Hazard rate for default

The duration only model provides initial baseline hazards. Figure 1 shows the shape of 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 highly likely to default drop out. This structure has been reported by others; see eg Gross and Souleles (2002) and Andreeva (2006). Figure 1 also shows a small second rise in hazard, peaking 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 tend to be low risk accounts, their removal leads to a small overall increase in default risk[1].

FIGURE 1 HERE

4.2 Model and coefficient estimates

Many AVs and 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. We find the following key outcomes. First, the signs on current balance (log) and its square are opposite but the positive sign on the square

[1] It is therefore important to realize that the hazard rate is not just an indication of obligor’s propensity to default but will also be influenced by periodic operational decisions by portfolio managers.

term 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. Second, An increase in credit limit reduces the hazard. Initially this may be surprising since we might argue that a high credit limit encourages higher balance and therefore greater risk. However, firstly, at least in the short term, a high credit limit enables the obligor to have a buffer to build up debts before reaching default. Secondly, 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. Third, the amount paid back each month, indicated by payment status and payment amount, has a negative effect on default. This is expected since a greater ability to repay implies that default is less likely. Fourth, 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. Fifth, 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. Sixth, indicator variables have been added for vintage, which indicates year of account opening. These are significant and therefore imply that cohorts explain some of the effect over time for default rates within the data set. This is natural since lenders will allow greater or less risky new accounts onto their books at different times, depending on their changing attitude to risk at different times in the business cycle.

Concerning macroeconomic variables, interest rate has a positive effect 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. Unemployment rate also has a positive effect on default. Unemployment rate is an indicator of direct economic stress on individuals. In particular, obligors who become or remain unemployed will find it more difficult to repay debt. Conversely, if unemployment decreases, then we would generally expect unemployed obligors to find jobs, therefore making it easier for them to repay. Therefore, the effect of this MV on default is as expected.

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 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 and unemployment described in points 7 and 8. Bellotti and Crook (2008) found similar results for interest rates on a different UK credit card data set, although this study discovered earnings to be a more important MV than unemployment. The macroeconomic effects also corroborate the study by Breedon and Thomas (2008) across several world-wide data sets, although they also found GDP to be significant in many cases. They also included vintage effect in their models.

These results are for the model with lag 12 month BVs. 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 tends to be stronger for models with shorter lags.

4.3 Model fit and forecasts of time to default

Model fit is shown in Figure 2 for several alternative models. 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. 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, this improvement comes at the price of a much shorter range of forecasts. We see in Figure 2 that, although some of the MVs are statistically significant, their contribution to model fit is weak. Nested model fit is also assessed with results shown in Table 3. This shows that adding BVs to the model gives a statistically significant improvement in fit and also adding MVs to the model gives a statistically significant improvement, even though this is small.

FIGURE 2 HERE

TABLE 3 HERE

Figure 2 also shows results of 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. Also, both the conventional deviance residual and the log-likelihood residual follow each other closely, implying that either measure is sufficient for this problem domain.

4.4 Estimation of Default Rates

Figure 3 shows estimated default rates for different models along with the observed (or true) default rates for each month of the test data set. The monthly observed default rates have high variance but there is a general trend of high values beginning in 2005, falling during 2005, then rising again in 2006. The AV model is able to model the general fall in default rates. However BVs are required to forecast the overall trend including the rise in 2006. However, the best forecasts indicating 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 aggregate DR, achieving better results than even BV models with shorter lags as demonstrated in Table 4.

FIGURE 3 HERE

TABLE 4 HERE

4.5 Stress test results

We ran Monte Carlo simulations using the MV model given in Table 2. Estimated DR was simulated on the test data set, 12 months following the observation date; ie for December 2005. A stable loss distribution was generated after m=25,000 simulations and is shown in Figure 4. The right-hand tail shows risk for more adverse conditions. In particular we have included the figure for expected shortfall at the 99% percentile. This shows that for the worst 1% of economic scenarios we consider, the expected DR is 1.73 times greater than normal conditions (ie median estimated DR).

VaR is also shown for comparison. We see that this gives a lower estimate of loss (1.59) which may not reflect extreme circumstances sufficiently. These figures are slightly higher than those suggested as part of the US stress testing exercise by FRS (2009). In particular the FRS study estimates a more modest rise between 20% and 55% in DR when contrasting a normal "baseline" figure to "more adverse" conditions[2]. But our results appear lower than those of Rösch (2004) who found a VaR for US credit cards to be 2.31 times the mean, although he used aggregate data, not account level data as we do.

FIGURE 4 HERE

5. Conclusion

Dynamic models are a flexible approach to model and forecast consumer credit risk. They have a number of well known advantages over static models including modelling the conditional probability of default in a specific time period rather than in a time window and enabling the prediction of the profitability of specific loans (Bellotti and Crook 2009).

We have used *discrete-time* survival analysis to model credit card risk. 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 commonly used *continuous-time* survival models. This is important when model builders use large databases of credit accounts.

---

[2] 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%.

We have used a large data set of UK credit card accounts to test the effectiveness of dynamic survival models with BVs and MVs as models of default. Unlike previous literature we explore these models as tools for risk measurement, forecasting and stress testing. We conclude that many BVs are statistically significant explanatory variables of default and including them gives improved model fit. Important BVs are account balance, repayments, number of transactions within each month and credit limit. We find model fit 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.

Second bank interest rates and unemployment rate significantly affect the hazard. Whilst their inclusion gave only a modest improvement in model fit and no noticeable improvement in forecasts of 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 affect will only become noticeable at the aggregate level where accounts are taken together. Where comparable our results corroborate results given by others (Gross and Souleles 2002, Calhoun and Deng 2002, Gerardi et al 2008).

Third, The inclusion of MVs enables stress tests which generate credible results indicating that adverse conditions may raise DR by around 79%. We used a simulation-based approach for our experiments 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.

Andreeva G (2006). European generic scoring models using survival analysis. *Journal of the Operational Research Society* **57** pp 1180-1187.

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

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.

Bellotti T and Crook J (2009). Credit scoring with macroeconomic variables using survival analysis. *The Journal of the Operational Research Society*, vol 60, no 12, December, pp1699-1707.

Boss M (2002). A macroeconomic credit risk model for stress testing the Austrian credit portfolio. *Financial Stability Review* 4, Oesterreichische Nationalbank.

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.

Crook J and Bellotti T (2009) *Asset Correlations for Credit Card Defaults*. Working paper, Credit Research Centre, University of Edinburgh Business School (March 2009).

Financial Services Authority FSA (2008). *Stress and scenario testing*. Consultation paper 08/24 FSA: UK.

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

Crook J and Bellotti T (2009). Time Varying and Dynamic Models of Consumer default. *Journal of the Royal Statistical Society, Series A* (forthcoming).

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*. Princeton University Press.

Ma P, Crook J and Ansell J (2009) Modelling take-up and Profitability. *Journal of the Operational Research Society* (1 April 2009) doi:10.1057/jors.2009.33 Special Feature.

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

Rösch D and Scheule,T. (2004). Forecasting Retail Portfolio Credit Risk. *Journal of Risk Finance*, Winter/Spring, pp16-32.

Rösch D (2003). Correlations and business cycles of credit risk: evidence from bankruptcies in Germany. *Swiss Society for Financial Market Research* vol 17(3) pp 309-331.

Therneau TM, Grambsch PM and Fleming TR (1990). Martingale based residuals for survival models. *Biometrika* **77** pp 147-160.

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).

Tables

Table 1.

Descriptive Statistics for Macroeconomic Variables (MVs)

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	Halifax House Price index	LBG	-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

The data is from1986 to 2004. Sources: UK Office of National Statistics (ONS), Lloyds Banking Group (LBG) and the European Commission (EC). Data is monthly and may be seasonally adjusted (SA).

Table 2.

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

Covariate	Estimate
Intercept	n/a**
Duration	1.35**
“ (squared)	-0.00698**
“ (log)	16.4**
“ (log squared)	-6.42**
Selected application variables (AV)	
Time customer with bank (years)	-0.00250**
Time with bank unknown +	-0.342**
Income (log)	-0.146**
Income unknown +	-1.46**
Number of cards	-0.0610**
Time at current address	-0.00129
Employment + :	
Self-employed	0.303**
Homemaker	0.072
Retired	0.111
Student	-0.035
Unemployed	0.231
Part time	-0.365**
Other	-0.037
Excluded category: Employed	
Age + : 18 to 24	0.074
25 to 29	-0.058
30 to 33	0.010
34 to 37	0.100**
38 to 41	0.046
48 to 55	-0.108**
56 and over	-0.243**
unknown	-2.74**
Excluded category: 42 to 47	
Credit bureau score	-0.00322**
Product + : A	0.535**
B	0.371**
Excluded category: C	
Vintage (+): 1999-2003	n/a **
Behavioural variables (BV) lag 12 months	
Payment status + :	
Fully paid	-0.390**
Greater than minimum paid	-0.090**
Minimum paid	0.149**
Less than minimum paid	0.714**
Unknown	-0.148*
Excluded category: No payment	

Current balance (log)	-1.58**
“ (log squared)	0.517**
“ is zero +	-1.05**
“ is negative +	-0.802**
Credit limit (log)	-1.22**
Payment amount (log)	-0.154**
“ is zero +	-0.133
“ is unknown +	-0.452**
Number of months past due	0.134*
Past due amount (log)	0.0795
“ is zero +	-0.623**
Number of transactions	0.00663**
Transaction sales amount (log)	-0.350**
“ is zero +	-0.567**
APR on purchases	-0.00487
“ is zero +	-0.482**
Behavioural data is missing +	-3.73**
Macroeconomic variables (MV) lag 3 months	
Bank interest rate	0.113**
Unemployment rate	0.000672**
Production index	-0.0101
FTSE all 100 (log)	0.0591
Earnings (log)	1.57
Retail sales	0.00929
House price (log)	-0.218
Consumer confidence	-0.00217
Retail price index (RPI)	-0.0298

Indicator variables are denoted by a plus sign (+). Statistical significance levels are denoted by asterisks: ** is less than 0.001 and * is less than 0.01 level. 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 3.

Model fit

Line	Nested model	Compared to base model	Difference in 2×LLR	Number of added covariables	P-value
1.	AV only	Duration only	22531	34	<0.0001
2.	AV & BV lag 12	AV only	9237	22	<0.0001
3.	AV, BV lag 12 & MV lag 3	AV & BV lag 12	47	9	<0.0001
4.	AV & BV lag 9	AV only	7946	22	<0.0001
5.	AV & BV lag 6	AV only	12458	22	<0.0001
6.	AV & BV lag 3	AV only	26559	22	<0.0001

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

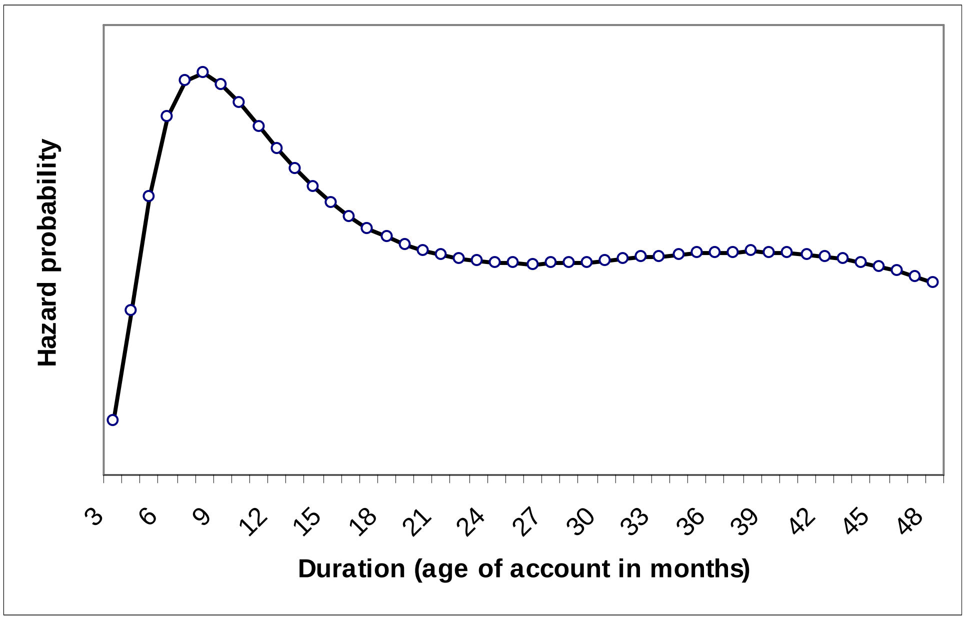
Table 4.

Mean absolute difference between 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

Results relate to models based on the 18 months of test results shown in Figure 3.

Figure 1.
Hazard rate function for parametric duration only model of default.



The hazard probability scale is not shown for reasons of commercial confidentiality.

Figure 2.

Model fit and forecast performance for different models.

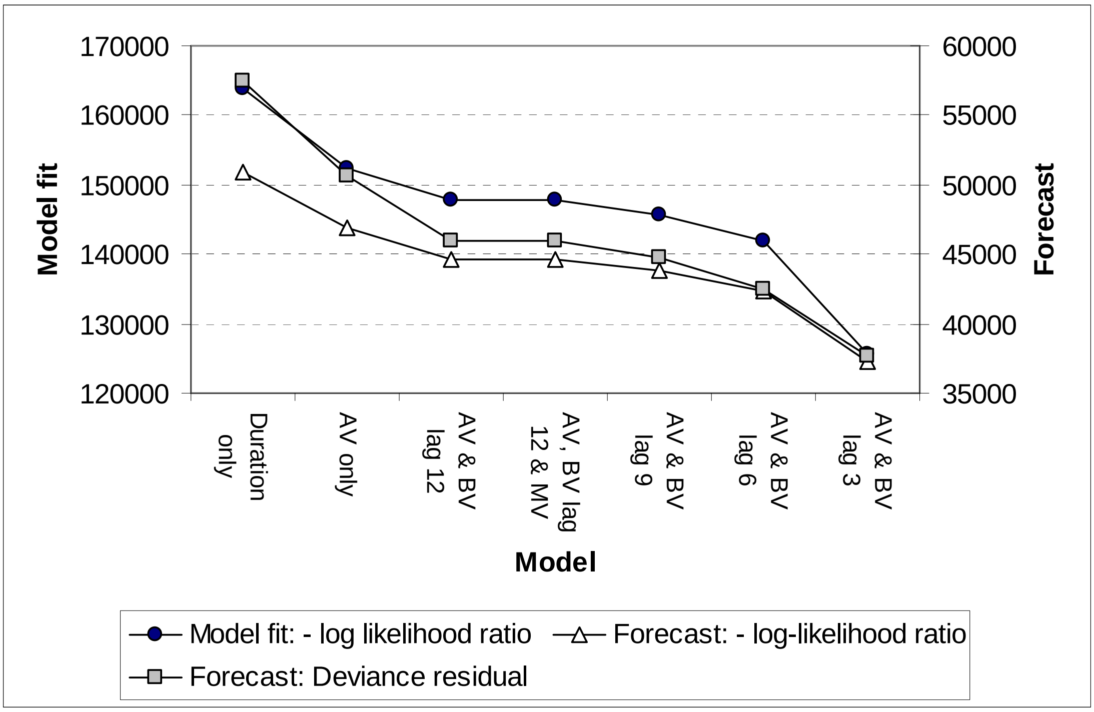
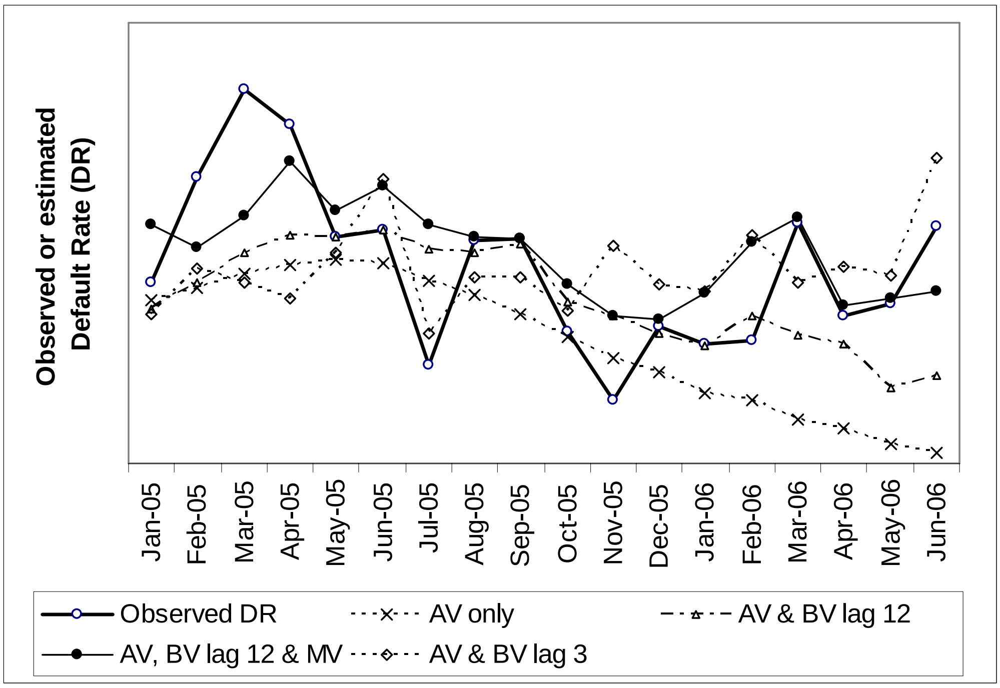


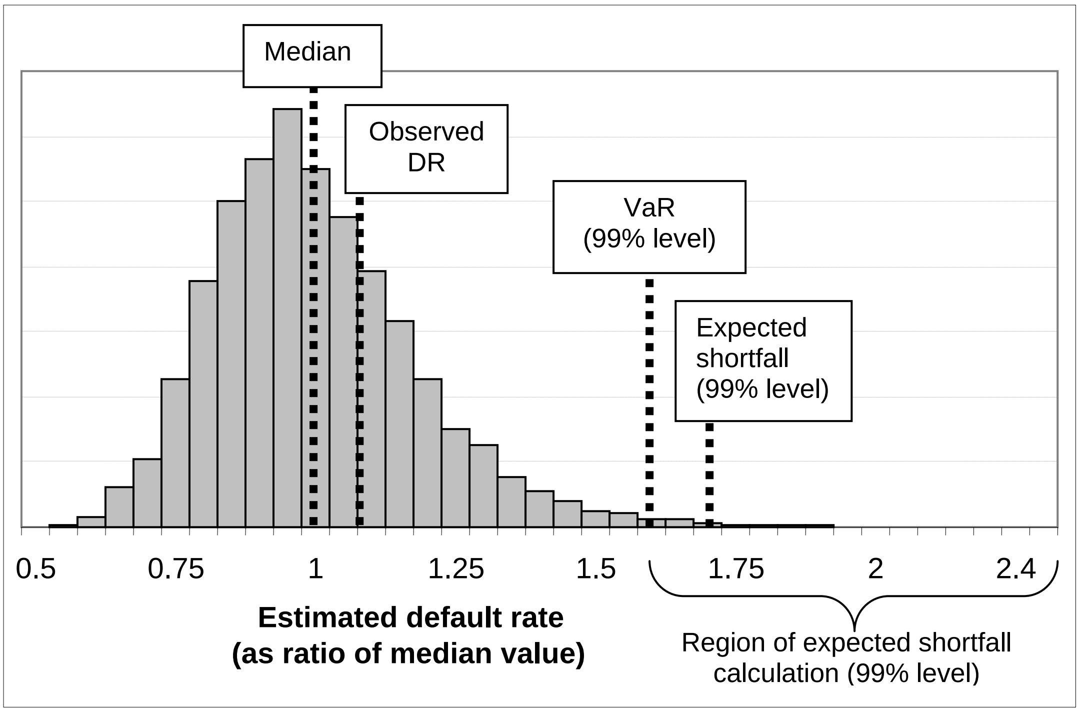
Figure 3.
Comparison of estimated and observed default rates for each month of the test data set.



The scale on the default rate axis is not shown for reasons of commercial confidentiality.

Figure 4.

Distribution of estimated default rates



The distribution is based on 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 Value at Risk (VaR) and expected shortfall at 99% probability. All values are expressed as a ratio of the median estimated DR.