Retail Credit Stress Testing Using a Discrete Hazard Model with Macroeconomic Factors

Citation for published version:
Bellotti, T & Crook, J 2013, 'Retail Credit Stress Testing Using a Discrete Hazard Model with Macroeconomic Factors' Journal of the Operational Research Society, vol. 65, no. 3, pp. 340–350. DOI: 10.1057/jors.2013.91

Digital Object Identifier (DOI):
10.1057/jors.2013.91

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
Journal of the Operational Research Society

Retail Credit Stress Testing Using A Discrete Hazard Model With Macroeconomic Factors

Tony Bellotti and Jonathan Crook

Credit Research Centre
University of Edinburgh Business School

Version 2c

Abstract

We present a stress testing approach based on a dynamic model of default. Retail credit models are implemented using discrete survival analysis which enables macroeconomic conditions to be included directly as time-varying covariates. In consequence, these models can be used for stress testing by determining changes in default given downturn economic scenarios. In particular Monte Carlo simulation is used to generate a distribution of estimated default rates from which extreme Value at Risk and expected shortfall are computed. Several macroeconomic variables are considered and factor analysis is employed to model the structure between these variables. Two large UK data sets are used to test this approach, resulting in plausible dynamic models and stress test outcomes.

Credit Research Centre Working Paper 10/2

Keywords: discrete survival models, stress testing, loss distributions, principal components analysis, credit risk.

Address for correspondence: Credit Research Centre, University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS
{j.crook,abellott}@uun.ed.ac.uk

1. Introduction

Stress tests have become increasingly important in evaluating the riskiness of bank loan portfolios and they are recognised as a key tool in helping financial institutions make business strategy, risk management and capital planning decisions (FSA 2008). They allow us to ask what level of losses we can expect given worst case scenarios when taking a number of risk factors into account. Stress tests should take into consideration unexpected but also *plausible* events from which unexpected losses can be computed. In turn regulatory and economic capital can be computed as required by the Basel II Accord (BCBS 2005). Most models of default for retail credit are either point in time (PIT) or through the cycle (TTC), neither of which are able to give good estimates of default rate on a portfolio through the business cycle since in the first case, PIT will reflect the conditions of a particular point in the business cycle and in the second case, TTC models will reflect average conditions. Unlike previous literature we consider a form of dynamic model of consumer default that includes time varying macroeconomic conditions and we use it to consider losses during downturn periods and as the basis of a principled approach to stress testing a portfolio. We present such a stress test of a portfolio.

Bellotti and Crook (2009) use a Cox Proportional Hazards survival model of time to default. This model has the advantage that macroeconomic time series data can be included in a principled way into the model as time varying covariates. They show that the inclusion of macroeconomic variables such as bank interest rates and earnings was significant and had the expected effect: that is, an increase in interest rates tends to raise risk of default whilst a rise in earnings tends to lower risk of default. We develop this method by building discrete time survival models of default with macroeconomic conditions on two different large credit card data sets. The advantages of this approach are that (1) the model reflects the discrete nature of accounts data, in our case it is monthly, (2) the model build is quicker and (3) the procedure for forecasting using this model is less complex. Discrete time survival models have been used to model corporate bankruptcy without macroeconomic variables (Shumway 2001, Cheng et al 2009), and personal bankruptcy (Gross and Souleles 2002) and mortgage foreclosures (Gerardi et al 2008) with macroeconomic variables, but none of these papers show how the models can be used for stress testing. Breedon and Thomas (2008) use a dynamic model to stress test over several scenarios from past economic crises. They identify a number of important macroeconomic indicators of default such as interest rate and GDP but do not build distributions of estimated default rates.

The use of scenario-based stress tests is now common in the regulation of banks (Hoggarth and Whitley 2003, FSA 2008, FRS 2009). This approach is based on selecting hypothetical economic scenarios using judgements supported by prior economic knowledge and considering plausible developments of the economy. However, this approach is problematic since it allows a high degree of subjective judgement in the selection process. For example, recent stress tests of major US banks (FRS 2009) have been criticized since the “more adverse” conditions it uses are considered too weak. For example, the estimate for adverse 2009 unemployment rate was already exceeded within the year. Baker (2009) estimates that the US stress test could have under-estimated losses by $120 billion. Clearly, the recent financial crisis has shown that past stress tests failed, since banks and regulators were left surprised by levels of losses. Haldane (2009) gives several reasons for this failure. One is that the banking system has not taken all risk factors into account or has under-estimated their effect (disaster myopia). A second is that banks never had internal incentives to conduct stress tests seriously (misaligned incentives). It is clear that further rigour is required in the stress testing process.

In this paper we consider Monte Carlo simulation for stress testing of consumer credit portfolios as an alternative to a scenario-based approach. This is a statistical and computational method which is less subjective than scenario selection, since scenarios are automatically simulated based on historical distributions of risk factors. There remains a subjective judgement in the selection of risk factors themselves. Monte Carlo simulation is a standard approach for stress testing of corporate credit (Marrison 2002), although we have seen no published account of its use for retail credit. Monte Carlo simulation generates a distribution of estimated loss. It is common to use Value at Risk (VaR) to compute extreme loss based on this distribution. However, there is a distinction between VaR and the requirements of stress testing since VaR captures worst case *in normal circumstances*, whereas stress testing attempts to capture losses given unusual circumstances. There is a connection between the two, but a noticeable difference in value can emerge when considering non-linear exposures or fat-tail loss distributions (BIS 2005). For this reason we also report expected shortfall for worst case scenarios. To generate economic simulations, it is necessary to model the structure of the macroeconomic data. We use principal component analysis (PCA) to derive key macroeconomic factors (MF) which are used in the default model. Factor analysis has been used successfully to model macroeconomic conditions; for example, the Chicago Fed National Activity Index is a highly regarded and reliable factor representing the US economy (Federal Reserve Bank of Chicago 2001).

Dynamic models of default including macroeconomic conditions are built for two large UK portfolios of credit cards and used to conduct plausible stress tests using a simulation approach. In section 2 we describe our modelling and stress testing methods, in section 3 we describe our data and give results and in section 4 we discuss our conclusions.

2. Method

We employ a discrete time logit survival model to estimate a dynamic model of default. We then use Monte Carlo simulation to generate distributions of estimated default rate across an aggregate of accounts. We discuss each of these techniques in the following subsections.

2.1 Dynamic model of default

We consider a panel data set of credit card accounts. For each account *i* we have the following data: a_i is the date of account opening; d_{it} indicates whether the account *i* defaults at some time *t* after account opening (0=non-default, 1=default); $\mathbf{w}_i$ is a vector of account-level static variables; and $\mathbf{x}_{it}$ is a vector of lagged behavioural variables. Additionally, we have macroeconomic variables which have the same value for all accounts on the same date: $\mathbf{z}_{it}$ is a vector of macroeconomic variables such that for any two accounts *i*, *j* with duration times *t* and *s* respectively, if $a_i + t = a_j + s$ then $\mathbf{z}_{it} = \mathbf{z}_{js}$. Probability of default (PD) for each account *i* at time *t* is modelled using

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

where α_t is a fixed effect for time t and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ are coefficients that need to be estimated. We use a logit link function $F(x) = 1/\left(1 + e^{-x}\right)$. We denote a specific model parameterization by $\boldsymbol{\theta} = \left(\alpha_1, \cdots, \alpha_T, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3\right)$. A standard software package can be used to estimate this model using maximum likelihood estimation. Since we model default conditional on no previous default for the same account, this is the Cox discrete survival model and the series of fixed effects α_t form a baseline hazard function. It follows from this conditionality that dependency of observations within each account is not a problem since probabilities of events factor out (see Allison 1995, chapter 7).

2.2 Default rate estimates, Value at Risk and expected shortfall

For a given calendar date c, the default rate for an aggregate of N accounts that remain open on that date is given by

$$D_c = \frac{1}{N} \sum_{i=1}^{N} d_{i(c-a_i)} \tag{2}$$

which, assuming independence between default events implies that the expected default rate is

$$E(D_c) = \frac{1}{N} \sum_{i=1}^{N} P_{i(c-a_i)} . \tag{3}$$

This is our usual point prediction of default rate. However, for stress testing we are interested in a distribution of estimated default rate given changes in the economy, so we consider the cumulative probability distribution over default rates given by

$$P\left(D_c \le y \mid \boldsymbol{\theta}\right) = \int_{\mathbf{z}} P\left(D_c \le y \mid \boldsymbol{\theta}, \mathbf{z}\right) p(\mathbf{z}) dz \tag{4}$$

for some density function p across economic conditions $\mathbf{z}$. Distribution (4) can be used to compute extreme estimates of default rate. In particular, Value at Risk (VaR) for percentile q is given by the smallest value V_q such that $P\left(D_c \le V_q \mid \boldsymbol{\theta}\right) \ge q/100$. Then expected shortfall is computed as the mean value for the worst case scenarios in the distribution above percentile q:

$$S_q = E\left(D_c \mid \boldsymbol{\theta}, D_c \ge V_q\right) = \tfrac{1}{1-q/100} \int_{y=V_q}^{\infty} y P\left(D_c = y \mid \boldsymbol{\theta}\right) dy \tag{5}$$

In addition to macroeconomic risk factors, we also need to consider noise in the data, relative to the model, as a risk factor in estimating default rates, since this will effect the distribution of outcomes. These are introduced by considering the model as a latent model with a residual term ε_{it} independent of all covariates and independently distributed in F:

$$\begin{aligned} d_{it}^* &= \alpha_t + \mathbf{w}_i{}^T \boldsymbol{\beta}_1 + \mathbf{x}_{i(t-l)}^T \boldsymbol{\beta}_2 + \mathbf{z}_{i(t-l)}^T \boldsymbol{\beta}_3 + \varepsilon_{it}, \\ d_{it} &= \mathrm{I}\left(d_{it}^* > 0\right) \end{aligned} \tag{6}$$

where $\mathrm{I}(\cdot)$ is the indicator function (see Verbeek 2004, section 7.1.3). Then substituting (6) into (2) we have default rate in terms of the model, macroeconomic conditions $\mathbf{z}$ and a vector of N residual terms $\boldsymbol{\varepsilon} = \left(\varepsilon_{(1)}, \cdots, \varepsilon_{(N)}\right)$ as

$$D_c'\left(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\varepsilon}\right) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{I}\left(\alpha_{c-a_i} + \mathbf{w}_i{}^T \boldsymbol{\beta}_1 + \mathbf{x}_{i(c-a_i)}{}^T \boldsymbol{\beta}_2 + \mathbf{z}^T \boldsymbol{\beta}_3 + \varepsilon_{(i)} > 0\right). \tag{7}$$

Then $P\left(D_c \le y \mid \boldsymbol{\theta}, \mathbf{z}\right) = P\left(D_c'\left(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\varepsilon}\right) \le y \mid \boldsymbol{\theta}, \mathbf{z}\right) = \int_{\boldsymbol{\varepsilon}} \mathrm{I}\left(D_c'\left(\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\varepsilon}\right) \le y\right) p\left(\boldsymbol{\varepsilon}\right) d\boldsymbol{\varepsilon}$ where $p\left(\boldsymbol{\varepsilon}\right)$ is the probability given that each residual is drawn independently from F. Substituting into (4) and assuming independence between $\mathbf{z}$ and $\boldsymbol{\varepsilon}$ gives

$$P(D_c \leq y \mid \boldsymbol{\theta}) = \int_{\boldsymbol{\varepsilon},\mathbf{z}} \mathrm{I}(D'_c(\boldsymbol{\theta},\mathbf{z},\boldsymbol{\varepsilon}) \leq y) p(\boldsymbol{\varepsilon}) p(\mathbf{z}) d\boldsymbol{\varepsilon} d\mathbf{z} \tag{8}$$

2.2 Monte Carlo simulation

Monte Carlo simulation is a means to compute integrals across distributions of values. In general, if $\mathrm{E}_f(h(X)) = \int_\chi h(x) f(x) dx$, where χ is uni- or multi-dimensional, then given large *m*, $\bar{h}_m = \sum_{j=1}^{m} h(x_j)$ is a good approximation, and converges in the limit, to $\mathrm{E}_f(h(X))$, where each x_j is a random draw (simulation) from the density function *f* (Robert and Casella 1999). Suppose, then, that for *j*=1 to *m*, $\mathbf{z}'_j$ and $\boldsymbol{\varepsilon}'_j$ are randomly generated from distributions for $p(\mathbf{z})$ and $p(\boldsymbol{\varepsilon})$ respectively and indexed such that simulated default rates are in ascending order: that is, for all $h \leq j$, $D'_c(\mathbf{z}'_h, \boldsymbol{\varepsilon}'_h) \leq D'_c(\mathbf{z}'_j, \boldsymbol{\varepsilon}'_j)$. Then, by Monte Carlo simulation, (8) is approximated as

$$P(D_c \leq y \mid \boldsymbol{\theta}) \approx \frac{1}{m} \sum_{j=1}^{m} \mathrm{I}(D'_c(\boldsymbol{\theta},\mathbf{z},\boldsymbol{\varepsilon}) \leq y). \tag{9}$$

The number of iterations *m* is chosen such that (9) converges to a stable value which is data dependent. Since the definition of D'_c (7) involves an indicator function over a threshold term for default, this simulation can be interpreted as simulating default or non-default events for each account in the data set, depending on the risk factors. In this sense it can be viewed as following the final simulation step used by Jokivuolle et al (2008) in their work on stress testing capital requirements for corporate data. From the definition of VaR and the ordering of simulated default rates it follows that

$$V_q \approx D'_c(\mathbf{z}'_{\lceil mq/100 \rceil}, \boldsymbol{\varepsilon}'_{\lceil mq/100 \rceil}) \tag{10}$$

Similarly (5) gives

$$\begin{aligned} S_q &= \tfrac{1}{1-q/100}\int_{\boldsymbol{\varepsilon},\mathbf{z}}\int_{y=V_q}^{\infty} yP\left(D_c'(\boldsymbol{\theta},\mathbf{z},\boldsymbol{\varepsilon})=y \mid \boldsymbol{\theta}\right)p(\mathbf{z})p(\boldsymbol{\varepsilon})dyd\mathbf{z}d\boldsymbol{\varepsilon} \\ &= \tfrac{1}{1-q/100}\int_{\boldsymbol{\varepsilon},\mathbf{z}} D_c'(\boldsymbol{\theta},\mathbf{z},\boldsymbol{\varepsilon})\mathrm{I}\left(D_c'(\boldsymbol{\theta},\mathbf{z},\boldsymbol{\varepsilon})\geq V_q\right)p(\mathbf{z})p(\boldsymbol{\varepsilon})d\mathbf{z}d\boldsymbol{\varepsilon} \end{aligned}$$

which is expressed as the Monte Carlo simulation

$$S_q \approx \frac{1}{m(1-q/100)}\sum_{j=\lceil mq/100\rceil}^{m} D_c'\left(\mathbf{z}_j',\boldsymbol{\varepsilon}_j'\right). \quad (11)$$

It is easy to generate the residual terms by repeatedly sampling from F, which here is the standard logit distribution. However, the distribution over macroeconomic conditions $p(\mathbf{z})$ requires that the structure amongst the macroeconomic variables is modelled. This can be done using Cholesky decomposition which preserves the covariance structure between simulated variables (Marrison 2002). However an alternative is to apply principal component analysis (PCA) to the macroeconomic series prior to including them in the model. PCA is a well-known technique which generates a series of components that are a linear combination of a set of random variables such that the first component accounts for as much of the variability in the data as possible, the second component is orthogonal to the first whilst accounting for as much of the remainder of the variance, and so on. The problem is well-posed and is solved by finding eigenvalues and eigenvectors of the matrix of data. For factor analysis, it is conventional to retain all components with eigenvalues greater than 1. For details, see Joliffe (2002). Hence, instead of including raw macroeconomic time series, we use macroeconomic factors (MF) instead. This is suitable since the factors will not be correlated with one another, therefore they can be generated independently in the simulation process. The macroeconomic values are drawn from the historical distribution of the factors. These are not necessarily normal, so we use the Box-Cox transformation to model each factor distribution and convert to normal. Box and Cox

(1964) show how to transform a given random variable x with the goal of producing an approximately normal distribution. They use the general form

$$x^* = \begin{cases} \dfrac{(x+k)^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x+k) & \text{if } \lambda = 0 \end{cases} \tag{11}$$

where k is a fixed parameter to allow for negative values of x and λ is a parameter estimated by maximizing the likelihood of the transformed values, assuming they are generated from a normal distribution.

2.3 Validation of stress tests

To compare and contrast using different risk factors for stress testing, three discrete survival models are built: **No MF**: A simple model without any MFs, **All MF**: A model including all MFs and **Selected MF**: A model including only MFs selected using stepwise variable selection. These three models form the basis of three stress tests using different risk factors. The No MF model is used for stress tests when only noise in accounts is assumed; the All MF model tests for when all macroeconomic conditions are included; and the Selected MF model tests for when macroeconomic conditions are selectively included in the stress test. The Selected MF model is required since not all MFs are likely to be relevant risk factors and for stress testing, their inclusion may therefore lead to inaccurate estimates of extreme values. This is particularly true if any MFs are correlated within the period of the training data which may lead to multicollinearity and therefore poor estimates for coefficient estimates [1]. This is unlikely to be a problem for prediction, when distributions of risk factor values in the forecast data would be expected to follow those in the training data, but for stress testing, when extreme values of risk factors are considered, it will more

[1] Although factors will be uncorrelated over the period of PCA, this does not imply they are uncorrelated for any sub-period.

likely have a noticeable effect. By including both the All MF and Selected MF models, this hypothesis may be tested. Additionally, the coefficient estimates of the model can themselves be a risk factor, since they are not exact estimates but have a distribution governed by a covariance matrix which is the outcome of the maximum likelihood procedure. During simulation, these estimates can be adjusted to determine the effect of estimation uncertainty on the loss distribution. For this reason, we also include a fourth stress test using the Selected MF model with estimation uncertainty.

A serious weakness of stress tests is that they cannot be easily validated since extreme events are both rare and unique. Their rarity means we are unlikely to have past extreme events for testing and even if we do their uniqueness means it is difficult to apply statistical analysis. This is particularly the case for scenario-based stress tests which analyze a risk model in relation to a specific set of extreme events. However, an advantage of the approach we take is that it generates a loss distribution from which stressed values are then derived as VaR or expected shortfall. This means that although the stressed events themselves cannot easily be tested the underlying loss distribution can be back-tested against historical data. Most financial institutions should have sufficient historical data in their retail portfolios to do this. In particular, the loss distributions can be validated given a time series of default rates for a post out-of-sample data set (Granger and Huang 1997) using a binomial test. To determine if the stress test is unrealistically conservative we can check if the number of observed default rates that exceed estimated VaR is likely to occur by chance. For example, if we are considering 99% VaR, we would expect only 1% of observations to exceed VaR on average and the distribution of such cases is governed by the binomial distribution which allows us to test the significance of outcomes (Marrison 2002, chapter 8). The application of the binomial test to this problem is well-known and forms the basis of the traffic-light validation system used by industry and regulators (Blochwitz and Hohl 2007). A potential difficulty with this approach is that the binomial test assumes independence between observations and this may not be the case for default rates on a portfolio over time. Therefore, this assumption should be tested prior to using the binomial test. Autoregression can be used to test dependency over the time series.

3. Experimental Results

3.1 Data

We have two large data sets for two UK credit card products, consisting of over 200,000 accounts each and spanning a period from 1999 to mid-2006. The data consists of (1) data collected at time of application such as the applicant's age, income, employment status, housing status and credit bureau score, (2) account open date and (3) monthly behavioural data including credit limit, outstanding balance, card usage and payment history: amount paid and minimum payment required. We define account *default* as the case when minimum payment is missed on an account for 3 consecutive months[2]. This definition is typical in the industry and matches the standard specified by Basel II (BCBS 2005) of 90 days delinquency.

A validation set consisting of one year of data is produced by randomly dividing each product data set into a training and validation data set in a 2:1 ratio of accounts, then discarding all records after an observation date of June 2005 from the training data set

[2] For reasons of commercial confidentiality, we cannot reveal descriptive statistics or default rates for these data.

and keeping only accounts opened prior to the observation date, but considering only records after the observation date, for the test set. This procedure ensures (1) there is no selection bias, since accounts in the validation data set are selected randomly and independently, (2) the validation data set is both out-of-sample and post-training data and (3) the validation is realistic in the sense that only accounts that are known (ie already open) prior to the observation date are included in the validation set.

Breedon and Thomas (2008) use several macroeconomic variables in their models of default such as GDP, interest rate, unemployment rate, house price and consumption variables, such as retail sales, which they argue could impact consumer delinquency. Bellotti and Crook (2009) also find bank interest rates, earnings, production and house prices significant in explaining default for different credit card data. We therefore follow with a similar set of variables described in Table 1. Notice that production index is used instead of GDP since production index is available monthly in the UK whereas GDP is quarterly. The difference in the value over 12 months is used for all variables to avoid inadvertently including a time trend or seasonal variation in the time series. The log value of FTSE, house prices and earnings are used in the model since these follow an obvious exponential trend. Since both the behavioural and macroeconomic data is monthly we use discrete monthly time in the survival model.

TABLE 1 HERE

3.2 Factor analysis

Table 2 shows that PCA applied to the macroeconomic time series from 1986 to June 2006 returns 4 factors with an eigenvalue greater than 1. MF1 is loaded on a broad range of economic variables, but not including FTSE index. MF2 is loaded mainly on consumption variables: RPI, consumer confidence and earnings. MF3 is driven by the FTSE index, although a mix of other variables also contribute to the factor. MF4 mirrors MF2, picking up consumption variables. For several variables we have a prior expectation of sign of effect on the probability of default and these are shown in Table 2. In particular, we expect greater values of interest rates, unemployment rate and RPI to represent increased stress on retail obligors, whereas greater values for production, earnings and house price should indicate improved economic conditions and hence a reduction of likelihood to default. The expected effect of each MF on the probability of default is then also given, based on the sign of the loading of the variable within the factor. We observe that only MF2 is expected to have an overall positive effect on default since all variables with a prior expectation are expected to have a positive effect in MF2 except for earnings. For other variables, we can hypothesize economic effects on the likelihood to default having either sign and we are unable to judge *a priori* which effect would be stronger. So increases in equity prices (FTSE) are indicators of economic health which we expect to reduce default rates; however, they are also linked to greater consumer activity which implies greater use of consumer credit and possible indebtedness. Similarly, the effect of consumption variables, retail sales and consumer confidence, are less easy to predict. It is possible that increases in these variables imply a greater load on credit card accounts as a consequence of higher sales. However, an opposite effect is possible since they also indicate improved economic confidence among consumers. For this reason, we do not state a prior overall expected effect for these factors.

TABLE 2 HERE

The movement of MFs is shown in Figure 1, extrapolating into the period of the financial crisis of 2008. The previous major recession in the UK began with the stock market crash in October 1987. All MFs show a large movement following this date. In particular, MF3 shows a sharp decline at this time and this is unsurprising given that the main contributor to MF3 is the FTSE index. However MF2 has the most sustained upward trend for several years after the beginning of the crisis, indicating strain on consumption after the stock market crash. Extrapolating into 2008, we see the dramatic effect of the financial crisis during this period on all MFs. This is evidence that these factors are good indicators of economic stress and could be used for stress testing.

FIGURE 1 HERE

We apply the Box-Cox transformation to model the factors. Both MF1 and MF3 require no transformation since their Box-Cox $\lambda = 1$. However, both MF2 and MF3 have long tails and require a transformation with $\lambda = -1$ and 2 respectively as shown in Figure 2.

FIGURE 2 HERE

3.3 Model fit

Along with MFs, the models also include application and behavioural variables and annual indicator variables for vintage, but these will not be reported in our results since the main focus of this paper is the inclusion of macroeconomic conditions and stress testing. For further details about building and assessment of a survival model using application and macroeconomic variables for retail credit cards, see Bellotti and Crook (2009). Models were built with behavioural variables lagged by 12 months in order to reduce the possible effect of endogeneity between behavioural data and default event (eg a rise in account balance and default may have a common external cause) and to allow for forecasts up to 12 months ahead. MFs were included with lag 3 months since we anticipate that economic conditions contribute to default at the time when payments first begin to be missed. It is possible that earlier lags could be used but our preliminary experiments indicated that a 3 month lag is sufficient. For the Selected MF model, a significance level of 1% was chosen for stepwise variable selection. This was sufficient for selection of factors whilst ensuring highly correlated factors are not included together.

Table 3 shows MF coefficient estimates for models built on training data for each product. Several factors are statistically significant in the models. In particular, MF2 is a strong driver in all models and has the direction of effect as expected overall, as shown in Table 2. Figure 2 showed that MF2 was the strongest signal of the effect of the October 1987 stock market crash, hence its inclusion in the models is promising for stress testing. MF3 is also a driver for product B although the size of effect is not as large as MF2. There was no overall expected direction of effect for MF3: as shown in Table 2, three loaded variables have a positive expected effect and three with a negative expected effect. However FTSE is the largest contributor to MF3 which implies that the strongest effect of FTSE on likelihood to default is not as an indicator of economic health but as an indirect indicator of consumer spending and possible over-indebtedness.

TABLE 3 HERE

The size of effect of MF2 is much larger in the All MF model than the Selected MF model. This is a consequence of the inclusion of MF4 which has an opposite effect to MF2 over the period of training. This is evident in Table 4 which shows surprisingly high local correlations between factors over the period 1999 to mid-2006. In particular, there is strong negative correlation between MF2 and MF4. Overall, this is evidence of multicollinearity between MF2 and MF4 in the training time period which inflates the coefficient estimate of MF2. As discussed in Section 2.3, this is not a problem for predictions but may be for stress testing. Variable selection resolves this problem by excluding MF4 from the model. There is no correlation between MF2 and MF3 so the model for product B which includes both these factors is not affected by multicollinearity.

TABLE 4 HERE

3.4 Stress test results

Models are built with an observation date of June 2005. Stress tests are then performed for one year ahead, June 2006. Figures 3 to 5 show distributions of estimated default rates following Monte Carlo simulation on the test set for each product and for stress tests using different risk factors. We found these distributions converged after 50,000 simulations. For reasons of commercial confidentiality we cannot report the precise default rates. Instead we report estimated default rate as a ratio of the median estimated default rate computed using the Selected MF model. Table 5 shows statistics for each of these distributions in terms of median default rate, VaR and expected shortfall. A level of 99.9% is used since this reflects the standard level recommended for use in the Basel II Accord.

FIGURE 3 HERE

FIGURE 4 HERE

FIGURE 5 HERE

TABLE 5 HERE

For product A we notice that the shape shown in Figure 3 is approximately the same for both the No MF and Selected MF models except that the No MF model tends to give higher estimates. However in Figure 4 the right tail is shown in detail and clearly shows that the Selected MF model has the longer tail, accounting for the much higher estimates of VaR and expected shortfall. For product B, the distribution is much broader for the Selected MF model than the No MF model and, again, the Selected MF model has a long tail, as shown in Figure 5. The long tail is typical of loss distributions and we would expect to observe it (BIS 2005). In the case of these experiments, the long tail is a consequence of including MF2 in the stress test which itself has a long tail (see Figure 2). Comparing the All MF to the Selected MF models, we find that for both products, the distributions based on the All MF models are much broader, leading to relatively extreme VaR and unexpected shortfall. This is a consequence of the inflated coefficient estimates caused by multicollinearity between MFs, rather than a genuine warning of greater risk. Finally, we note that excluding estimation uncertainty as a risk factor makes very little difference to the distribution, even at extremes, and so for practical purposes it can be safely excluded.

Validation is performed using a binomial test on the post, out-of-sample validation data consisting of 12 months of data. Figures 6a to 6d show time series plots of observed default rates, along with percentiles for the simulated distributions for each month. The binomial test is based on the number of observations that are above the 99^{th} or 99.9^{th} VaR. Figures 6a and 6c show that for both products, the Selected MF model gives plausible distributions and the binomial test gives probability of outcome greater than 10%. However, the No MF model gives implausible outcome with probability of outcomes above VaR being less than 1%, based on Figures 6b and 6d. The binomial test assumes independence between observations. We tested this assumption for our data using autoregression (AR(1) and AR(2)) and found no significant correlation for default rates over time .

FIGURES 6a-6d HERE

Figures 6a to 6d also show that observed default rates are higher than the median of the predicted values estimates for all models and approaching the 99% VaR even for the MF model. This is unsurprising since there has been a generally rapid rise in credit card delinquency from 2005 (Bank of England 2008, Chart 2.7 shows increase in write-offs) and this is the case for both products that we used. Hence, in reality, hitting 99% VaR is reasonable given the increased risks on many credit card portfolios during this period, relative to previous periods.

The results suggest that the Selected MF model gives the most plausible stress test outcomes. From Table 5 we see that this stress test yields a stressed value of monthly default rate about double the median. A doubling of default rates is a large increase but not implausible given an economic crisis. Indeed, based on data given by the Bank of England (2008, Chart 2.7), average write-off rates for credit cards in the UK, generally, were 2.5% during the relatively benign period of our training data, but rose to 7% during the recent financial crisis by 2007: a multiple of 2.8. Further, in our experiments we find expected shortfall is 15-20% greater than VaR. As argued earlier, expected shortfall is the more reasonable value to use for stress testing. The observed difference is sufficiently large to support this argument empirically and to show that VaR should not be used as a substitute for stress testing for typically long-tailed loss distributions.

5. Conclusion

We present an approach for stress testing retail credit portfolios using a dynamic model of default that includes macroeconomic conditions. We use PCA to generate MFs based on several macroeconomic time series that we believe could affect consumer delinquency. Since the MFs are uncorrelated[3], simulated values can be generated for them independently and used as economic scenarios for stress testing, employing Monte Carlo simulation to build a distribution of estimated default rates. This simulation approach has the advantage that it is potentially less subjective than scenario based approaches and, since it generates a loss distribution, it enables an empirical validation step through back-testing.

Our experimental results based on two large credit card portfolios show that dynamic models including macroeconomic conditions can be built successfully with one or two statistically significant MFs. MF2 is connected to consumption variables and is found

[3] Over the period for which they were estimated.

to be the strongest macroeconomic driver of default. The inclusion of MFs is sufficient to produce a long tail on the simulated loss distributions. Without MFs, the tail is too short and, consequently, VaR and expected shortfall are too low to be plausible estimates of stressed loss rates. A binomial test is used to check the plausibility of loss distributions based on an out-of-sample validation data set. We also discovered that although the MFs are generally uncorrelated, over a local time period they may be highly correlated, leading to multicollinearity in the model. Although this problem does not affect forecasts of expected (mean) default rates it does affect the use of the model for stress testing, generating much broader loss distributions and, consequentially, much larger VaR and expected shortfall. We successfully employed variable selection to avoid this problem.

The use of expected shortfall is contrasted with VaR as a measure of stressed loss. We find a sufficiently large difference in the two values to effect risk management decisions and capital requirement calculations. Certainly, the use of expected shortfall is the more principled approach to calculating expected loss given worst case and this result reinforces the point that VaR may not be the most suitable measure for stress testing (BIS 2005). We found that 99.9% expected shortfalls for the two credit card products A and B were 2.11 and 2.43 respectively, which are high but not implausible, in the light of evidence from the recent financial crisis.

Our research has raised several issues that require further investigation. Firstly, the model and simulation assumes that accounts are independent, conditional on economic circumstances. This may not be the case and we may need to assume or calculate an asset correlation. The inclusion of MFs explains some of the asset correlation, but may not account for all of it. Rösch (2003) shows a considerable reduction in asset correlation between accounts once macroeconomic risk factors are included in a model of corporate bankruptcy and Crook and Bellotti (2009) show the same effect for retail credit card delinquency. Nevertheless even a small asset correlation, additional to economic effects, could be included as a risk factor during simulation and may produce more accurate distributions. Secondly, with our approach we have attempted to generate economic scenarios based on random sampling from the historic distribution of economic conditions. However, the structural development of the economy could also be modelled and used to generate simulations. For example, Jiménez and Mencía (2007) use vector autoregression to construct a structured model of the macroeconomy and other risk factors, whilst Hoggarth and Whitley (2003) report using a dynamic medium-term macroeconomic model for stress tests of UK banks. Thirdly, empirical validation remains a problem for stress testing due to the inherent rarity and uniqueness of the events we are attempting to model. We give a binomial test approach on the underlying loss distributions. However, it is primarily a test to reject loss distributions that are clearly implausible. The binomial test still cannot test the validity of extreme value estimates (above the 99^{th} percentile) and therefore cannot be used to positively validate the plausibility of VaR and expected shortfall. Further development of empirical tests of extreme values would be valuable. For example, Wong (2009) suggests an alternative approach using a size of tail loss statistic instead of a binomial test and finds this has more statistical power. Nevertheless, a simple binomial test gives plausibility to the overall loss distribution, providing further confidence in their use for stress tests.

References

Allison, P.D. (1995). *Survival analysis using SAS*. SAS Press.

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

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.

Bank of England (2008). *Financial Stability Review*, April 2008.

Baker, D. (2009). Background on the Stress Tests: Anyone Got an Extra $120 Billion?, Beat the Press Blog, *The American Prospect*, 8 May 2009

Bellotti, T. and Crook J. (2009). Credit scoring with macroeconomic variables using survival analysis. *The Journal of the Operational Research Society* (2009) **60**, 1699-1707.

Blochwitz, S. and Hohl, S. (2007). Validation of Banks' Internal rating Systems – a Supervisory Perspective in *The Basel Handbook 2nd edition* ed. Ong M (Risk Books).

Box, G. E. P. and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B* **26**: pp211–246

Breedon, J. and Thomas, L. (2008). A common framework for stress testing retail portfolios across countries. *The Journal of Risk Model Validation* **2**(3), 11-44.

Cheng, K.F. Chu, and Hwang, R.C. (2009) Predicting bankruptcy using discrete semi-parametric hazard models. *Quantitative Finance* Doi 10.1080/14697680902814274.

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

Federal Reserve Bank of Chicago (2001). *CFNAI Background Release*, working paper.

FRS: Board of Governors of the Federal Reserve System (2009). *The supervisory capital assessment program: overview of results*. White paper, May 7 2009， FRS.

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

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

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

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

Haldane, A.G. (2009), *Why the banks failed the stress test*, basis of a speech given at the Marcus-Evans Conference on Stress-Testing, 13 Febuary 2009, Bank of England.

Hoggarth, G. and Whitley, J. (2003). Assessing the strength of UK banks through macroeconomic stress tests. *Financial Stability Review*, June 2003, Bank of England.

Jiménez, G. and Mencía, J. (2007). *Modelling the distribution of credit losses with observable and latent factors*. Banco de Espana Research Paper.

Joliffe, I. T. (2002). *Principal component analysis*. (2nd ed, Springer).

Jokivuolle, E., Virolainen, K. and Vähämaa (2008). *Macro-model-based stress testing of Basel II capital requirements*. Bank of Finland research discussion papers 17/2008.

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

Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods* (Springer).

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.

Shumway, T. (2001) Forecasting bankruptcy more accurately, a simple hazard model. *Journal of Business* vol 74(1), 101-124.

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

Wong, Woon, K. (2009), Backtesting Value-at-Risk based on tail losses, *Journal of Empirical Finance*, doi: 10.1016/j.jempfin.2009.11.004

Tables and Figures

Table 1.

UK Macroeconomic Variables With Descriptive Statistics.

MV	***Description and source***	***Date available***	***Descriptive statistics*** *(for difference over 12 months)*			
			Min	***Mean***	***SD***	***Max***
IR	UK bank interest rates (ONS)	Jun 1985	-4.5	-0.43	1.90	6.5
Unemp	UK unemployment rate (in '000s) SA (ONS)	Feb 1972	-535	-94	238	575
Prod	UK production index (all) (ONS)	Jan 1968	-5.2	1.10	2.30	6
FTSE	FTSE 100 all share index (ONS)	Jan 1975	-822	81	286	682
Earnings	Earnings (log) all including bonus (ONS)	Jan 1990	0.008	0.019	0.006	0.038
House price	House price (Halifax)	Jan 1986	-6.5%	+7.9%	7.6%	+26%
Retail sales	Retail sales value (ONS)	Jan 1986	0.3	3.92	1.49	8.5
RPI	Retail price index (all items) (ONS)	Jan 1987	1.2	4.96	2.36	12.8
Cons conf	Consumer confidence index (EC)	Jan 1985	-20.3	0.7	24.2	186.8

Sources are the UK Office of National Statistics (ONS), Nationwide Building Society (Nat) and the European Commission (EC). Data is monthly and may be seasonally adjusted (SA).

Table 2.

Macroeconomic Factors (MF)

		Macroeconomic factors with expected effect (EE)							
	EE	**MF1**	**EE**	**MF2**	**EE**	**MF3**	**EE**	**MF4**	**EE**
Eigenvalue		2.47		1.83		1.44		1.08	
Variables:									
IR	+	0.80	+	0.21	+	-0.03	−	-0.30	−
Unemp	+	-0.85	−	0.33	+	0.02	+	0.22	+
Prod	−	0.57	−	-0.42	+	0.40	−	-0.24	+
FTSE		0.01		-0.11		0.84		0.01	
Earnings	−	0.36	−	0.60	−	0.33	−	0.48	−
House price	−	0.64	−	-0.13	+	-0.36	+	0.28	−
Retail sales		0.36		-0.26		-0.35		0.58	
RPI	+	0.34	+	0.85	+	0.19	+	0.08	+
Cons conf		-0.05		-0.56		0.42		0.48	

The factors are derived from macroeconomic data from 1986 to June 2006 using principal component analysis. Only the four factors with eigenvalues greater than 1 are shown. The direction of expected effect (EE) is shown for each variable, and its effect within each MF given the direction of the loading.

Table 3.

Coefficients estimates for MFs for the two MF models.

Model	**Variable**	**Product A**		**Product B**	
		Estimate	**Chi-sq**	**Estimate**	**Chi-sq**
All MF	MF1	-0.0325	1.1	0.00931	0.1
	MF2	0.1429	15.3 **	0.1659	20.1 **
	MF3	0.0329	5.3	0.0711	23.3 **
	MF4	0.0494	6.0	0.0473	4.8
Selected MF	MF1	-		-	
	MF2	0.0796	9.8 *	0.1064	16.8 **
	MF3	-		0.0638	26.2 **
	MF4	-		-	

Statistical significance is shown at 1% level (*) and 0.01% level (**). Note that all models also included application and behavioural variables but coefficient estimates for these are not shown.

Table 4.

Correlation Coefficients Between Macroeconomic Factors (MF) Within Training period

	MF1	**MF2**	**MF3**
MF2	-0.01		
MF3	0.32 *	0.023	
MF4	0.15	-0.6 *	-0.38 *

Training period is 1999 to June 2005. Statistical significance is shown at 1% level (*).

Table 5.

Estimated Default Rates For The Loss Distributions Shown In Figures 3 To 5.

		Model and risk factors			
	Observed	No MF	All MF	Selected MF	Selected MF (no est. unc.)
Product A:	1.41				
Median		1.09	1.04	1	1
VaR		1.50	2.84	1.83	1.79
Expected shortfall		1.54	4.66	2.11	2.15
Product B:	1.32				
Median		0.95	1.03	1	1
VaR		1.24	3.22	2.01	1.98
Expected shortfall		1.27	4.59	2.43	2.31

Estimated default rates are shown as a ratio to the median default rate given the Selected MF model. VaR and expected shortfall are given for the 99.9%ile.

Figure 1.
Movement of Macroeconomic Factors Over Time.

Figure 2. Distribution Of Historic MF2 (Left) And MF3 (Right) Values Along With Optimal Box-Cox Transformation Distribution.

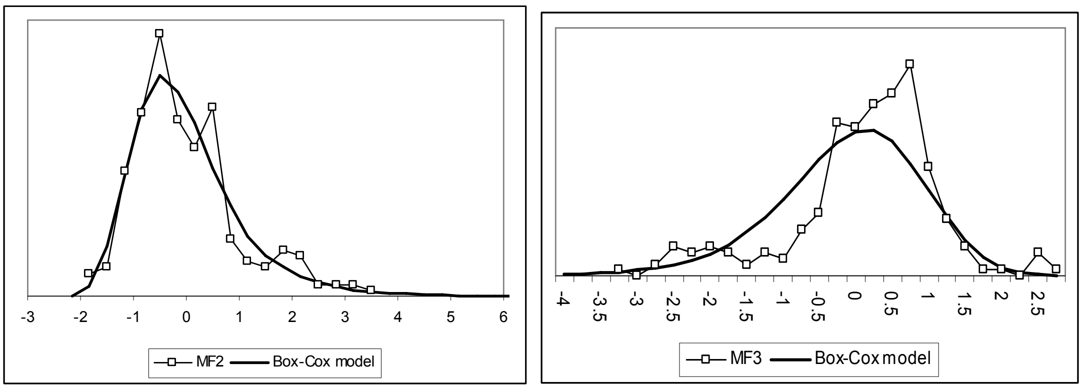


Figure 3.
Loss distributions for product A for June 2006.

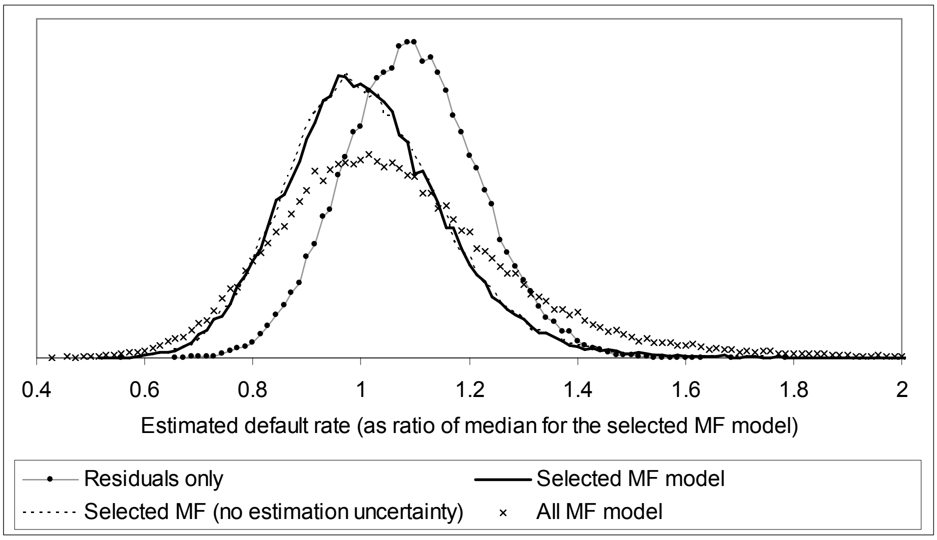


Figure 4.
Extreme Right Tail Of Loss Distributions For Product A For June 2006.

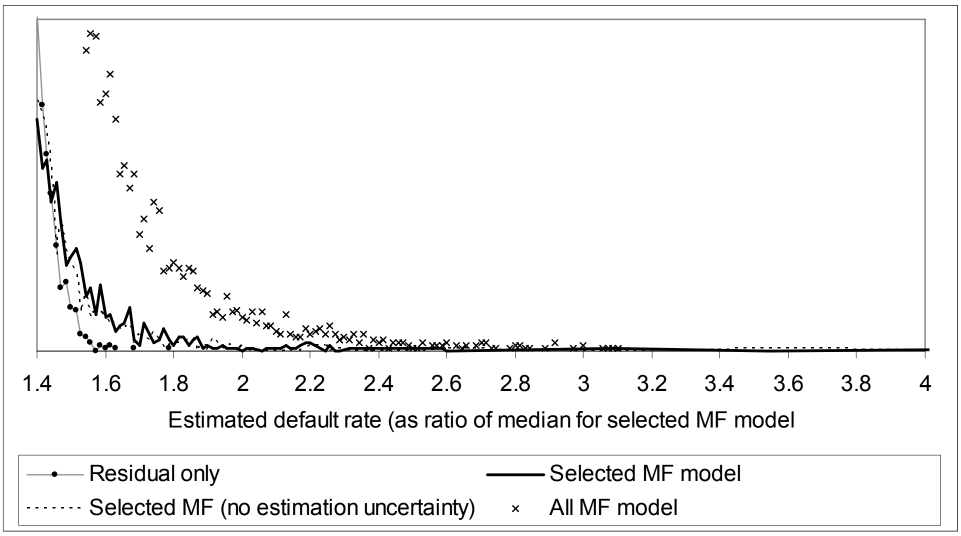


Figure 5.
Loss Distributions For Product B For June 2006.

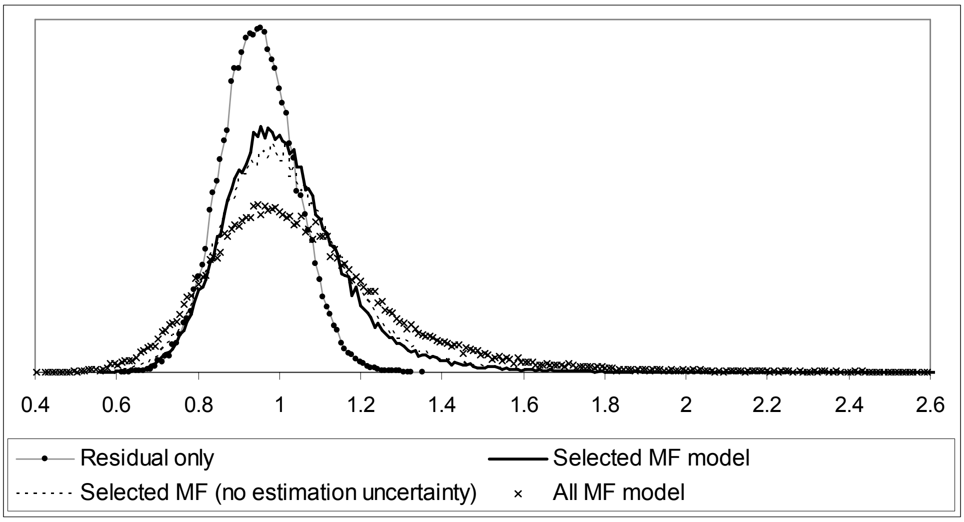
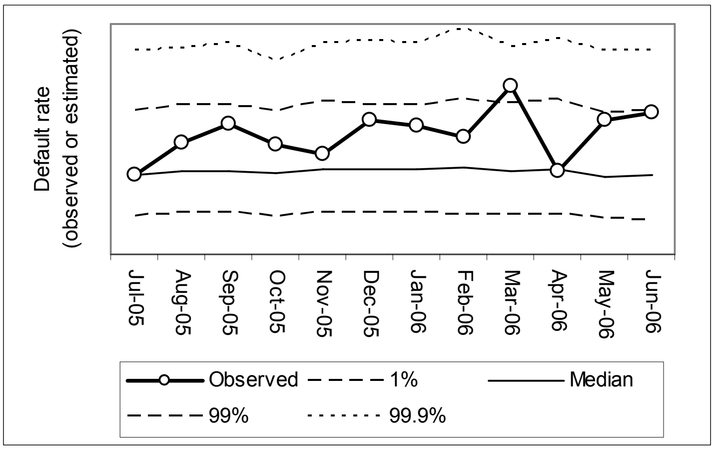


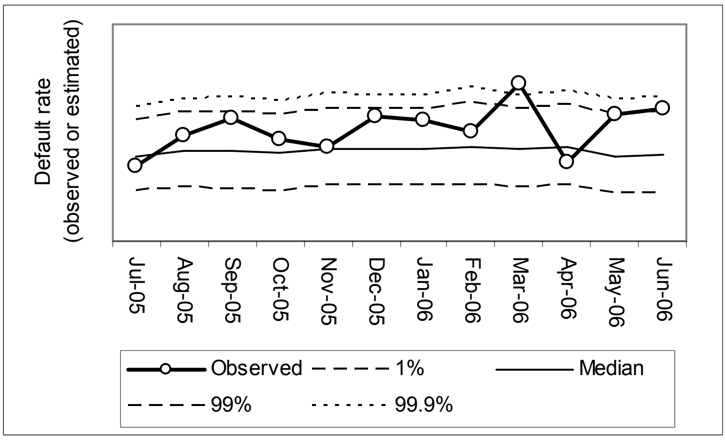
Figure 6a.
Time Series Of Loss Distribution For Product A Given The Selected MF Model.



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

Figure 6b.

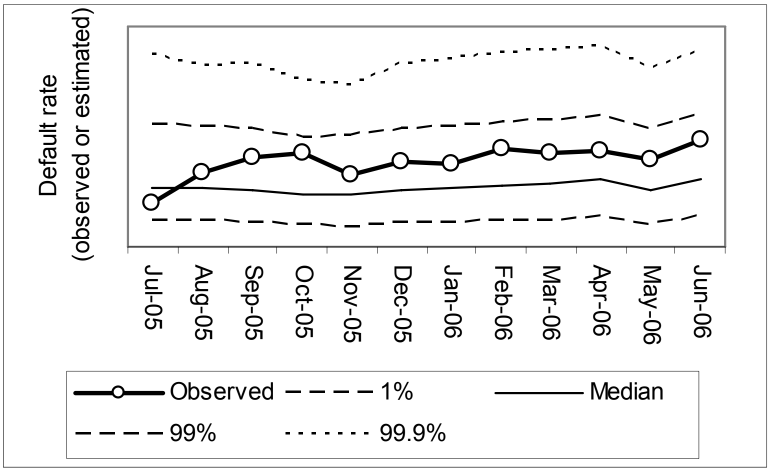
Time Series Of Loss Distribution For Product A Given The No MF Model.



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

Figure 6c.

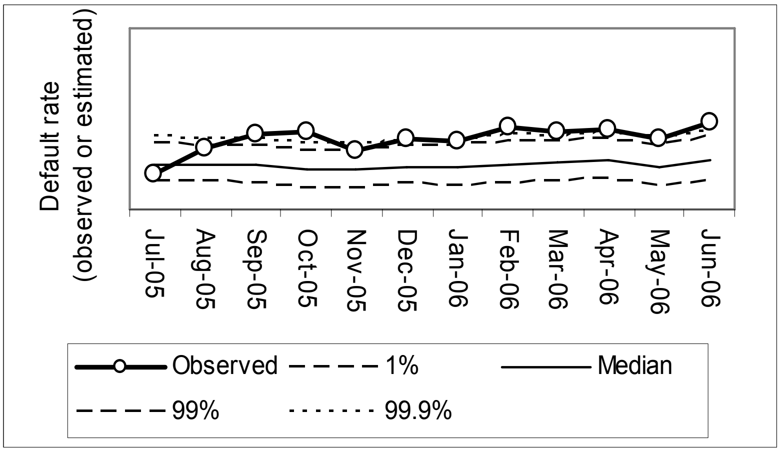
Time Series Of Loss Distribution For Product B Given The Selected MF Model.



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

Figure 6d.

Time Series Of Loss Distribution For Product B Given The No MF Model.



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