Credit scoring with macroeconomic variables using survival analysis

Citation for published version:
Bellotti, T & Crook, J 2009, 'Credit scoring with macroeconomic variables using survival analysis', Journal of the Operational Research Society, vol. 60, no. 12, pp. 1699-1707. https://doi.org/10.1057/jors.2008.130

Digital Object Identifier (DOI):
10.1057/jors.2008.130

Published In:
Journal of the Operational Research Society

Credit Scoring with Macroeconomic Variables using Survival Analysis

Tony Bellotti and Jonathan Crook

Credit Research Centre

Management School and Economics

University of Edinburgh

8 September 2008

Research funded by EPSRC grant EP/D505380/1, working as part of the

Quantitative Financial Risk Management Centre

Abstract

Survival analysis can be applied to build models for time to default on debt. In this paper we report an application of survival analysis to model default on a large data set of credit card accounts. We explore the hypothesis that probability of default is affected by general conditions in the economy over time. These macroeconomic variables cannot readily be included in logistic regression models. However, survival analysis provides a framework for their inclusion as time-varying covariates. Various macroeconomic variables, such as interest rate and unemployment rate, are included in the analysis. We show that inclusion of these indicators improves model fit and affects probability of default yielding a modest improvement in predictions of default on an independent test set.

Keywords: credit scoring; survival analysis; time-varying covariates; risk; banking; macroeconomic variables.

1. Introduction

A credit application scoring model involves predicting the probability that an applicant will default over a given future time period in terms of characteristics of the applicant measured at the time of application. Yet after the time of application the ability of an applicant to repay may change due to factors which credit scoring models typically assume are constant over time. The predictive accuracy of such models might be improved if the lender could incorporate into the prediction equation additional variables, which are predictable and which are correlated with the changing circumstances of a borrower. Macroeconomic variables (MVs) are an example of such variables and the inclusion of MVs into the prediction equation is an attempt to do just this. There is accumulating econometric evidence that aggregate delinquency and write-off rates vary with the state of the macroeconomy (Crook & Banasik 2005). However there has been no published work that incorporates states of the macroeconomy into credit scoring models that predict the probability of default (PD) for an individual applicant.

In this paper we test the hypothesis that the PD of an individual applicant is affected by macroeconomic conditions as measured by MVs such as bank interest rates, unemployment index and earnings. The novelty of this paper is that we test which of these macroeconomic conditions have a statistically significant effect and provide quantification for the level of the effect for modelling and predicting PD for individual applicants. Since MVs are given as time series data, this cannot be done so easily using the usual regression models used to build credit scoring models such as logistic regression (LR). One approach is to use survival analysis. Survival analysis is an expanding area of research in credit scoring (Banasik et al 1999, Stepanova and Thomas 2002, Andreeva et al 2007). It enables macroeconomic time series data to be incorporated naturally into the survival model as time-varying covariates (TVCs) as suggested by Banasik et al 1999. We conduct experiments to test the effect of MVs on the PD of individual credit card account holders. We are interested in assessing MVs in terms of their explanatory and predictive power in models of default. Survival models with TVCs are constructed and contrasted with standard LR models to determine any uplift in predictive performance. We show that the inclusion of MVs gives a statistically significant explanatory model of the data and a statistically significant uplift in predictive performance. This suggests that lenders would, on average, gain a more accurate prediction of an applicant's PD if the lender used a survival model which includes MVs rather than a LR model which omits them. In addition, our results imply that when MVs are included in a survival model a lender can simulate the effects of downturns in the macroeconomy on the future PD for an applicant and can also do so for future PDs for a portfolio of applicants.

In the next section we outline the Cox Proportional Hazards (PH) model with TVCs. In section 3 we discuss the expected signs of the MVs and in section 4 we explain the details of our implementation. Section 5 presents the results and in section 6 we discuss our results and draw conclusions.

2. Cox Proportional Hazard Model with Time-varying Covariates

Although LR has become a standard method for estimating applicant scoring models (Thomas et al 2002), there has recently been interest in using survival analysis for credit scoring. This allows us to model not just if a borrower will default, but when. The advantages of this method are that (i) survival models naturally match the loan default process and so incorporate situations when a case has not defaulted in the observation period, (ii) it gives a clearer approach to assessing the likely profitability of an applicant and (iii) survival estimates will provide a forecast as a function of time from a single equation (Banasik et al 1999). Survival analysis has been applied in many financial contexts including explaining financial product purchases (Tang et al. 2007), behavioural scoring for consumer credit (Stepanova & Thomas 2001), predicting default on personal loans (Stepanova & Thomas 2002) and the development of generic score cards for retail cards (Andreeva 2006).

Survival analysis is used to study time to *failure* of some population. This is called the *survival time*. Survival analysis is able to facilitate the inclusion of observations that have not failed. These are treated as *censored* data and an observation time can be given for censored cases indicating the last time they were observed. In the context of consumer credit, the population comprises individuals applying for credit in the form of loans or credit cards. When a consumer defaults on a loan or credit card payment then this is a failure event. Survival time is measured from the date the account was opened. If a consumer never defaults during the lifetime of their account then they are censored and observation time is the period of time the account was open or, if the account was never closed, the time from when the account was opened to the date of data collection.

A common means to analyze survival data is through the *hazard function* which gives the rate of change of probability of failure at time *t*:

$$h(t) = \lim_{\delta t \to 0} \left(\frac{P(t \le T < t + \delta t \mid T \ge t)}{\delta t} \right) \qquad (1)$$

where T is a random variable associated with survival time. The probability of survival at time t can be given in terms of the hazard function:

$$S(t) = P(T \ge t) = \exp\left(-\int_0^t h(u)du \right). \qquad (2)$$

This is the probability of survival up to time t (Collett 1994, Section 1.3). For credit data, this is the probability that an account has not defaulted by some time t after the account has been opened, ie 1-PD at time t. A series of n observations i=1 to n is given in survival analysis in terms of observation times t_i and indicators c_i where c_i=0 for a censored observation and c_i=1 for a failure event, in which case t_i is the survival time. In addition, each observation will include a vector of covariates that may be associated with survival time. Some of these may be time-varying so, in general, they are given as functions of time, $\mathbf{x}_i(t)$. Application data are fixed with respect to time. However, MVs change over time and the value of the covariate is given as the value of the MV at time of failure. Several models of the hazard function are available, but in this paper we use the Cox PH model since it allows for the inclusion of MVs as time-varying covariates (TVCs). This model is semiparametric, depending partly on a vector of coefficients $\boldsymbol{\beta}$ that are linear multiples of the covariates and a nonparametric baseline hazard function h_0 dependent on time but not the covariates. With TVCs, the Cox PH model is given by

$$h(t, \mathbf{x}(t), \boldsymbol{\beta}) = h_0(t) \exp(\boldsymbol{\beta} \cdot \mathbf{x}(t)) \qquad (3)$$

where $h(\cdot)$ is the hazard at time t for an application where, for credit scoring, $\mathbf{x}(t)$ is a vector of covariates containing the following elements:

- α application elements $a_1, \ldots, a_\alpha$ fixed at the date of application *d*;
- *m* macroeconomic variables dependent on time *t* following application, but otherwise independent of application data, $z_1(d+t), \ldots, z_m(d+t)$;
- several interaction terms between application and macroeconomic variables $a_i z_j(d+t)$ for $(i, j) \in S \subseteq \{1, \ldots, \alpha\} \times \{1, \ldots, m\}$ where S is the set of interactions to be included, determined by the variable selection method described later in Section 4.1.

The coefficients $\boldsymbol{\beta}$ are estimated using the partial likelihood function on the training observations,

$$l_p(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta} \cdot \mathbf{x}_i(t_{(i)}))}{\sum_{j \in R(t_{(i)})} \exp(\boldsymbol{\beta} \cdot \mathbf{x}_j(t_{(i)}))} \right)^{c_i} \tag{4}$$

where $t_{(i)}$ are ordered survival times and the risk set $R(t) = \{j : t_{(j)} \geq t\}$. This allows the use of maximum likelihood estimation to estimate $\boldsymbol{\beta}$ without needing to know the baseline hazard (Hosmer & Lemeshow 1999, Section 7.3). However, in order to use the model for estimation of survival probabilities, the baseline hazard is needed. This can be estimated based on the parameter estimates $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ given by the maximum likelihood estimation and using an estimate for integrated baseline hazard given by Andersen (1992),

$$\hat{H}_0(t) = \sum_{t_{(i)} \leq t} \frac{c_i}{\sum_{j \in R(t_{(i)})} \exp(\boldsymbol{\beta} \cdot \mathbf{x}_j(t_{(i)}))}. \tag{5}$$

Numeric integration is used to compute survival probability for each individual using Equations (2) and (3), following Chen et al (2005). Notice from Equation (3) that variations in the MVs have two effects. Firstly, the same effect on the hazard for all applicants by the same amount, and secondly, through the interaction terms that change the hazard differentially depending on the value of characteristics at the time of application.

3. Macroeconomic Variables

Several MVs are used and are described in Table 1. These MVs were selected since monthly time series data are available for them and they are the most likely to affect default. Table 1 shows the effect we expect each MV to have on risk of default. A positive value means that as the value of the MV rises, this is likely to be linked to a rise in risk of default. Conversely a negative expectation means that an increase in the value is likely to be linked to a decrease in risk. So an increase in interest rates is likely to increase interest repayments relative to disposable income and so is likely to increase the mean PD across applicants. It may also increase the PD for certain applicants more than others, for example if an applicant has a high income at application they may be more able to maintain payments when interest rates rise than if their income is low. A rise in unemployment rate may be expected to increase the mean PD of a population of applicants but may also have a disparate effect on those unemployed at the time of application because it may be even harder for them to subsequently gain employment than it would be for those who were already in work at application. However, increases in earnings may reduce the ratio of interest payments to earnings and so reduce the mean PD, but it may also differentially impact some groups rather than others. For example it may lead some groups, such as those who own their own homes, to wish to borrow more than those who do not own their own homes and so be more at risk. An increase in house prices would increase a borrower's wealth which, from economic theory (the Permanent Income Hypothesis: Friedman 1955, Deaton 1992) would lead to greater credit being taken. Some econometric evidence supports this (Crook 2001) and some does not (Crook & Hochguertel 2007). This would increase the debt payments relative to current income and so increase mean PD. Of course, there may be a counter effect, which is that as wealth increases households are more able to pay credit card debt from their assets. So the sign on house prices is difficult to predict. Again the effect of changes in house prices may differ between applicants; for example if the applicant owns his own house at application, a rise in house prices may reduce his PD whereas if one is a tenant the effect may be the opposite. An increase in average consumer confidence may be expected to increase the average demand for debt if this confidence is consistent with individuals expecting their income to rise (Friedman 1955). Increasing debt may lead to increased mean PD. Similarly increases in the FTSE index, which implies an increase in mean wealth, and growth in production, correlated with increased income, are also indicators of the state of the economy providing conditions for a reduction in the mean risk of default with possible differential effects.

TABLE 1 HERE

4. Implementation

4.1 Model Selection

We expected that the inclusion of interactions between application variables and MVs may lead to better models since some categories of credit consumers would be more prone to changes in economic conditions than others. The following model selection method is employed to determine which interactions to include, based on the strategies described by Collett (1994, Section 3.6). Each MV is interacted with an application variable and added to the basic application survival model. The uplift of model fit is then measured using the log-likelihood ratio (LLR) derived from the maximum likelihood procedure used to estimate the model. For each MV, the interaction giving the lowest p-value based on its LLR is included in the optimal model. Due to the large size of the training set, processing time to fit each model was long. This meant constraining the model selection phase and, in particular, it was judged that forward selection or backward elimination methods would be too time consuming to use.

4.2 Data

Sample application and monthly performance data were used for a credit card product provided by a UK bank. This sample spanned a period of credit card accounts opened from 1997 to mid-2005. Accounts opened between 1997 and 2001 were used as a training data set, and those opened between 2002 and 2005 were used as an out-of-sample test data set. Each data set contained over 100,000 accounts with application variables such as income, age, housing and employment status along with a bureau score taken at the time of application.

For this experiment, an account is defined as being in *default* if three consecutive payments are missed. The usual classification credit scoring methods, such as LR, are restricted to considering default within a certain time frame, say 12 or 18 months. However, survival analysis does not have this restriction since it models time to default rather than whether default occurs within a time horizon. In this sense, survival analysis provides a more general method than LR and other classification algorithms for modelling default. Nevertheless, since we want to compare performance with LR we will need to focus attention on predicting default within a specific time period. We use default within 12 months of opening an account since this is typical for credit scoring (see eg Banasik et al 1999, Stepanova & Thomas 2002). We refer to an account that defaults within 12 months as a *bad* case and otherwise as a *good* case.

For this data set the proportion of bad cases is small.[1] Therefore we decided to over-sample them for training the survival analysis model. The problem of imbalanced classes has been discussed within the data mining community, eg see Chawla et al. [2004]. It has been shown that the natural distribution is not necessarily the optimal one for building classifiers [Weiss and Provost 2003]. A number of solutions have been proposed and perhaps the simplest and most common are under-sampling the majority class or over-sampling the minority class. For example, Burez and Van den Poel [2008] consider both methods for prediction of customer churn, as do Schuermann and Matthews [2005] for fraud detection. There are some concerns about using over-sampling, in particular the problem of over-fit [Weiss 2004], but with our data set we found, when tested on an independent hold out sample, over-sampling gave good results. We over-sampled bad cases in the training set so that total numbers of bad and good cases were approximately equal and this produced a good predictive model. In contrast, when training on a data set with the natural distribution we found this gave a model with poor performance for predicting bad cases. Although we over-sampled for survival analysis, we found this was not

[1] For reasons of commercial confidentiality, we cannot reveal the exact figures for default rate or volumes of data.

necessary for LR since we found that the best predictive results for LR were achieved when no over-sampling was used.

4.3 Assessment

We assess our optimal model in terms of both its explanatory power on the training data and its predictive power on the independent test set.

Explanatory Model

The Cox PH model is assessed as an explanatory model by reporting its fit to the training data with and without MVs, additional to inclusion of the usual application variables. The log-likelihood ratio (LLR) is used to assess model fit. Since over-sampling of bad cases artificially alters the distribution of training cases, we do not use the Wald statistic to generate p-values on coefficient estimates as would normally be the case. Instead we use the bootstrap method to compute percentile confidence intervals for each of the coefficient estimates from which we then report p-values (ie achieved significance levels) to test the null hypothesis that each coefficient is zero (Efron and Tibshirani 1993).

When a MV z_j interacts with multiple application variables, it is difficult to immediately determine the effect of z_j on the PD. However, it is possible to determine the marginal effect on *log-hazard*, $\log h(t, \mathbf{x}(t), \boldsymbol{\beta}) = \log h_0(t) + \boldsymbol{\beta} \cdot \mathbf{x}(t)$, from Equation (3). Since log-hazard is linear in $\boldsymbol{\beta} \cdot \mathbf{x}(t)$, the marginal effect of z_j on log-hazard, conditional on the interaction terms, is

$$\gamma_j = \beta_{(j)} + \sum_{(i,j)\in S} \beta_{(i,j)} a_i \qquad (6)$$

where $\beta_{(j)}$ and $\beta_{(i,j)}$ are coefficient estimates for z_j and each interaction term $a_i z_j$ respectively (see Brambor et al 2005) and S is the set of selected interactions. In this paper we report a single figure for marginal effect by substituting the mean values of each application variable $a_1, \ldots, a_n$ in Equation (6), thus providing a value of marginal effect of z_j for the mean observation. One way to determine the relative *importance* of each MV in the model is to measure the magnitude of the standardized marginal effect; ie the absolute value of the marginal effect multiplied by the standard deviation of the MV over the period of time of the MV data. This provides an indication of the relative importance of each MV in the model.

Predictive Performance

To determine its usefulness as a credit scoring system, the Cox PH model is tested as a predictor of default within a time period as discussed in Section 4.2. Predictions are made using survival probabilities computed using the Cox PH model. Given a cut-off threshold, the survival probabilities are used as scores to predict default. Thus if a case has a survival probability at 12 months that is greater than the cut-off then it is predicted as good, otherwise it is predicted as a bad case. Predictions are made with LR in a similar way using a cut-off on PDs computed using the LR model.

Notice from Equation (2) that the survival probability at time t is based on the integral of the hazard values, each being evaluated over a successive time period, from the point in time at which the account was opened until period t, and each of which, from Equation (3), is evaluated using the values of the MVs at each time. So the predicted survival probabilities over a 12 month period depend on the values of the MVs as measured over those same 12 months. To use the model to make out of sample predictions, we need to forecast future values for the macroeconomic values. In conducting our tests we have used ex post observed values of the macroeconomic variables and so have assumed that, had the model been used to make ex ante predictions, the values of the macroeconomic variables were predicted with complete accuracy.

It is typical in credit data for there to be a large imbalance between good and bad cases. This is particularly the case for the data we use which has very low default rates. It is easy to achieve a very high success rate in predictions due to the large proportion of good cases in the data with an algorithm which actually gives poor discrimination between good and bad cases. Additionally, we know that for a financial institution, the relative loss for a rejected good account will be much smaller than that for an accepted account that eventually goes bad. For this reason, errors on bads have a higher cost than those on goods and a cost function is used to determine the value of a prediction: (1) a correctly classified case has a cost of 0, (2) a good case wrongly predicted as bad incurs a cost of 1 and (3) a bad case wrongly predicted as good incurs a cost of 20. A relative cost penalty of 20:1 is chosen since the low default rate means that a cost of 10 or lower would be so low that the most cost effective policy for this data would simply be to accept all credit card applications whilst, on the other hand, a cost much greater than 20 is unlikely to reflect a realistic ratio of costs between decisions made based on wrongly predicted good and bad cases. Nevertheless, to demonstrate robustness, relative costs of 15 and 25 are also reported. We have chosen not to use receiver operating characteristic (ROC) curves to assess the models since they are insensitive to the relative costs that we can expect between errors on good and bad cases in consumer credit and, therefore, can possibly give rise to misleading conclusions. Gini coefficients also measure discrimination over all cut-offs but a lender is typically interested in cut-offs around a narrow region which gives an acceptable good rate (Hand 2005). The same weakness applies to the Kolmogorov-Smirnov statistic and the Brier score.

For each model, the cut-off threshold is computed to minimize the total cost of errors on the training set. This cut-off is then applied to make predictions on the test set. Therefore, the predictions made on the test set are completely independent of the training data. However, the cut-off computed in this way is unlikely to be optimal for the test set and there will be a degree of fluctuation between the computed cut-off and the cost optimal cut-off for the test set. This will affect the relative performance of the models. In order to determine that improvement in performance is due to the model, rather than a fluctuation in the cost effectiveness of the cut-off on the test set, the analysis is repeated with cut-offs computed so as to minimize total cost on the test set. This is likely to introduce a bias, but it does, however, allow us to discount fluctuations in the cut-off term as a cause of improvement in performance. If a particular model performs well with both cut-offs derived from training and test sets, then it shows that the model is both an unbiased good predictor and that the results are not due to fluctuations in the optimality of the cut-off threshold.

Assessment is made on the independent test set. The *mean cost* per observation is computed for each model as the mean cost of errors across all cases in the test set. Models giving a lower mean cost have performed better. To see how relative performance between models changes over time, differences in cost will also be reported over the time period of the test set by year and quarter. The significance of any improvement in performance of one model over another is measured using a paired *t*-test on the sequence of costs incurred between two models on the sequence of independent test cases (Witten & Frank 2005, Section 5.5). *Sensitivity* and *specificity* will also be reported for the optimal Cox PH model with MVs. These are the proportions of good cases in the test set predicted as good and bad cases predicted as bad, respectively. These figures allow us to contrast with results using other credit models to ensure our model's behaviour is typical (Baesens et al 2003).

5. Results

All the models are statistically significant as measured using LLR, at significance level 0.0001. In particular, treating the Cox PH model with MVs and selected interactions as a nested model (see Collett 1994) within the Cox PH model without MVs, we find the inclusion of MVs improves LLR at significance level 0.0001.

Table 2 shows coefficient estimates for all MVs along with the selected interaction terms. The model also includes application variables but these estimates are not reported firstly for reasons of commercial confidentiality, but also because the focus of this paper is on MVs. Several of the MVs proved to be important explanatory variables. Interaction terms were selected using the automated model selection process described in Section 4.1. Two of these proved to be significant (at a 5% level). Firstly, the term IR * income with a negative sign implies that as income increases so an applicant is less sensitive to the effect of IR; eg the marginal effect of IR was +0.12 for an income of £10,000 but this decreases to +0.06 for £40,000. Secondly, the term Prod * bureau score implies different sensitivity to economic changes by credit risk; ie the sign of the marginal effect of the production index is negative for accounts with low credit bureau scores, whereas it is positive for high scores.

TABLE 2 HERE

Table 3 shows the estimated marginal effect for each MV, taking the interaction terms into consideration. They are calculated using Equation (6) with figures for coefficient estimates given in Table 2. The prior expected sign for each of the marginal effects taken from Table 1 is also shown. The coefficients are positive for IR and Unemp, indicating a marginal increase in hazard (risk of default) with increases in bank interest rates and levels of unemployment. This is what we would expect since higher interest rates mean generally higher repayments on credit and higher levels of unemployment mean less economic stability. Conversely, hazard decreases with increases in the FTSE index and levels of real earnings. Production index does not have the sign we expected, in the average case. However, since it interacts with the bureau score significantly, we note that for low bureau scores the marginal effect is negative which is what we expect. This may indicate that high risk applicants are more sensitive to economic changes related to changes in production. Table 3 also shows the relative importance of each MV, given as the magnitude of standardized marginal effect. These are also shown graphically in Figure 1. Interest rate is by far the most important MV influencing default risk as we would expect, followed by earnings.

TABLE 3 HERE

FIGURE 1 HERE

Table 4 shows prediction results on the test set for each model when the optimal cut-off is computed from the training data set. These results reveal that survival analysis improves performance in terms of reduced mean cost and that this is largely due to the inclusion of MVs. This is most noticeable for a relative cost of 20 or higher. For example, the cost reduction when comparing LR with Survival Analysis with MVs was 1.7% at cost ratios 20 and 25. Comparison between the Cox model with and without MVs also shows a reduction in costs when using MVs which demonstrates that most of the effect is due to the addition of MVs. Significance tests given in Table 5 demonstrate that the improvement in performance is significant at a 0.001 level for a cost ratio of 20 and that this is largely due to the inclusion of MVs.

TABLE 4 HERE

TABLE 5 HERE

Table 4 also shows test results when the experiments are repeated with the cut-off computed using the test set to yield optimal performance. Again, these results reveal an improvement in performance when MVs are included, indicating that the results are not related to fluctuations due to the method of computing the cut-off. Table 5 also shows that the performance uplift in this case is also statistically significant and due mainly to the inclusion of MVs. P-values are higher when the cut-off is optimized on the test set but this is natural since this process introduces a bias into the model assessment. Nevertheless a significance level of 0.05 is achieved.

The sensitivity and specificity using the Cox PH model with MVs on the test data set, with cost on bad cases=20, are 96.7% and 16.9% respectively. Sensitivity is much higher than specificity but this is typical of credit data (see eg Baesens et al 2003) which demonstrates that our model is behaving normally as a credit model.

The effect of interest rates on the performance of the survival model is apparent in Figure 2 where the difference between predictions from the survival model including MVs and one that does not is shown over time by the unbroken line. The MV values that enter the predictions for a cohort occur over the four quarters following the account open date, which is shown on the horizontal axis. It can be seen that when interest rates change, from about 2003Q1, the performance of the model with MVs is better than the model without. The exception is 2004Q3 where the large contribution of MVs can be explained by the steep rise in unemployment rate over the following 12 months. The broken line shows differences between the survival model with MVs and the LR model. This follows a similar pattern but with a random perturbation over time. There is a sharp dip in the performance of the survival models in 2005Q2 but overall the survival models still perform best on average as is evident from Table 4.

FIGURE 2 HERE

Note that several alternative applications of MVs were also considered in our experiments, such as lagged values over 3, 6 or 12 months or taking the difference in values over 3, 6 or 12 months. We found that they did not lead to better performance so we report just the simplest model, taking MV values at the point of default.

6. Discussion and Conclusions

These results demonstrate that survival analysis is competitive in comparison with LR as a credit scoring method for prediction. The inclusion of MVs gives a statistically significant improvement in predictive performance. We show that model fit improves significantly and that the direction of the marginal effect on log-hazard rate of MVs mostly matches our prior expectations. Additionally, Figure 1 indicates that interest rate is the more important MV for estimation of risk of default as we would expect given its direct connection with credit. Although model fit and improvement in predictive performance is shown to be statistically significant, it also emerges that the effect is modest, in the order of a 1.7% reduction in loss across the whole portfolio and varies with the cost ratio chosen. When we weigh risk of default more heavily, the value of the MV model as an estimator becomes more pronounced.

In practice, this model can be used for credit scoring by incorporating *forecasts* of macroeconomic conditions into the assessment of credit card applications. For model training, values for MVs are used at exact time of default, or lagged. For prediction, obviously this exact time is not known. However, importantly, when using the model for prediction, MVs are included as a whole time series estimate across the entire observation period being considered using Equation (2). In this way the survival probability is estimated by integrating the hazard rate and incorporating estimates of the macroeconomic time series across the whole observation period, not at a particular point in time. This method of estimation also makes this model suitable for stress testing by including macroeconomic conditions that simulate a depressed or booming economy. This makes it valuable for the implementation of the requirements of the Basel II Accord (BCSC 2006 paragraph 415).

The inclusion of interaction terms between MVs and application variables means the model also incorporates expected individual responses to the economy. That is, changes in PD in response to change in MVs are not the same for everyone. For example, the inclusion of the interaction term IR * income with a negative coefficient estimate (see Table 2) suggests that individuals with higher income are less sensitive to a change in interest rates in terms of their effect on PD. An implication of the use of this model for acceptance or rejection of credit applicants is that an applicant may be accepted if he applies in one month but may be rejected if he applies in another month, possibly the next month. The reason is that if one is making an acceptance decision based on the PD over the following 12 months, the predicted values of the MVs over the 12 months beginning in, say, January may differ from those beginning a few months later, say, February. So the PD of an applicant may be lower over the next 12 months than over a 12 month period that starts a month or more later. The cut-off might conceivably be between the two PDs and so lead to different conclusions. This allows more accurate prediction of an applicant's PD over a defined time period than the conventional LR method which, effectively, assumes the state of the economy is static.

The inclusion of MVs could also prove useful in anticipating the *volatility* of an applicant's PD to changes in economic conditions in the future and therefore could help give a further insight into risk when assessing applications. That is, instead of assessing applicants simply on PD, with the MV model they can also be assessed based on the volatility of PD given possible movements in the economy.

Future lines of research will focus on further application of these methods to other credit card and fixed loan products and to mortgages. Also, although the analysis of the explanatory model gives an understanding of how each MV contributes to modelling the data, further extensive experimental work is required to determine the separate affect of each of the MVs on the prediction of PD.

Acknowledgements

We would like to thank Professors David Hand, Lyn Thomas and other members of the Quantitative Financial Risk Management Centre for discussion of survival analysis and credit scoring during the course of this research. We are grateful for funding through an EPSRC grant EP/D505380/1. We also thank our anonymous referees for their useful comments that allowed us to make significant improvements to this paper.

References

Andersen P.K. (1992). Repeated assessment of risk factors in survival analysis. *Statistical Methods in Medical Research* **1**, 297-315.

Andreeva G (2006). European generic scoring models using survival analysis. *J Opl Res Soc* **57**(10): 1180-1187

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

Baesens B, van Gestel T, Viaene S, Stepanova M, Suykens J and Vanthienen J (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *J Opl Res* **54**: 1082-1088.

Basel Committee on Banking Supervision (BCBS 2006). *Basel II: International Convergence of Capital Measurement and Capital Standards* at www.bis.org/publ/bcbsca.htm

Brambor T, Clark WR, Golder M (2005). Understanding interaction models: improving empirical analyses. *Political Analysis* **14**: 63-82

Burez J and Van den Poel D [2008] Handling class imbalance in customer churn prediction. *Working paper 2008/517 Faculteit Economie en Bedrijfskunde, Universiteit Gent.*

Chawla NV, Japkowicz N and Kolcz A [2004] Editorial: Special Issue on Learning from Imbalanced Data Sets. *SIGKDD Explorations*, 6(1): 1-6.

Chen L-S, Yen M-F, Wu H-M, Liao C-S, Liou D-M, Kuo H-S, Chen T H-H (2005). Predictive survival model with time-dependent prognostic factors: development of computer-aided SAS Macro program. *Journal of Evaluation in Clinical Practice*, **11**, 2, 181-193.

Collett D (1994). *Modelling Survival Data in Medical Research.* Chapman & Hall, London.

Crook J (2001). The demand for household debt in the USA: evidence from the 1995 Survey of Consumer Finance. *Applied Financial Economics***, 11**: 83-91.

Crook J, Banasik, J *Explaining Aggregate Consumer Credit Delinquency Behaviour over Time.* University of Edinburgh Credit Research Centre Working Paper, 05/03

Crook J, Hochguertel S (2007). *Household Debt And Credit Constraints: Comparaticve Microevidence From Four OECD Countries*. University of Edinburgh Credit Research Centre Working Paper 05/02

Deaton A (1992) *Understanding Consumption*. Oxford: Oxford University Press.

Efron B and Tibshirani RJ [1993] *An Introduction to the Bootstrap*. Chapman and Hall NY.

Friedman M (1955) *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Hand DJ (2005) Good practice in retail credit scorecard assessment. *J Opl Res Soc* **56:** 1109-1117.

Hosmer Jr. DW and Lemeshow S (1999). *Applied Survival Analysis: regression modelling of time to event data*. Wiley.

Scheurmann E and Matthews C [2005] Neural network classifiers in Arrears Management. *ICANN 2005, LNCS 3697*, pp 325-330 *(eds. W. Duch et al.) Springer-Verlag*.

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

Stepanova M and Thomas LC (2002). Survival analysis for personal loan data. *Opl Res* 50: 277-289.

Tang L, Thomas LC, Thomas S, Bozzetto J-F (2007). It's the economy stupid: modelling financial product purchases. *International Journal of Bank Marketing*. Vol.25, issue 1, pp.22-38.

Thomas LC, Edelman DB and Crook JN (2002). *Credit Scoring and its Applications*. SIAM Monographs on Mathematical Modeling and Computation. SIAM: Philadelphia, USA.

Weiss G and Provost F [2003] Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Researc*h, 19:315-354.

Weiss G [2004] Mining with Rarity: A Unifying Framework. *SIGKDD Explorations*, 6(1): 7-19.

Witten IH and Frank E (2005). *Data Mining*. 2nd ed. Morgan Kaufmann, San Francisco USA.

Figure 1. Importance of each MV in Cox PH model.

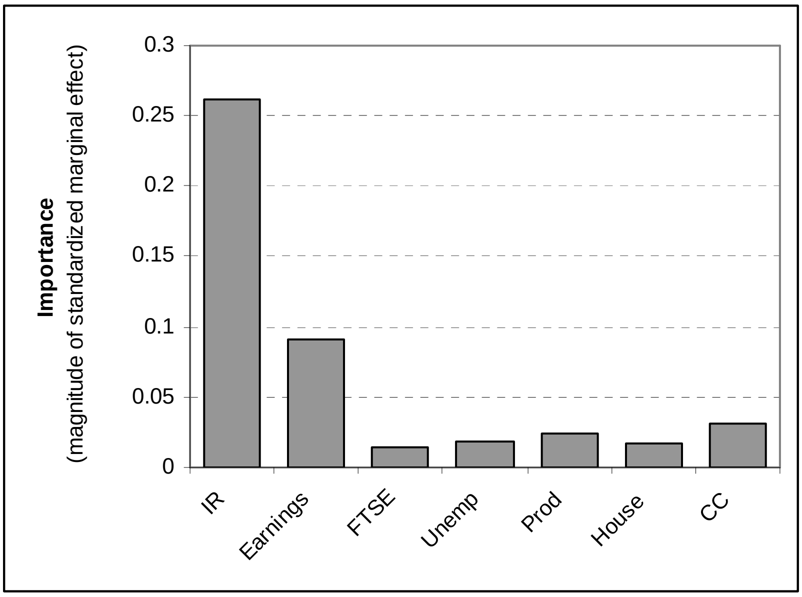
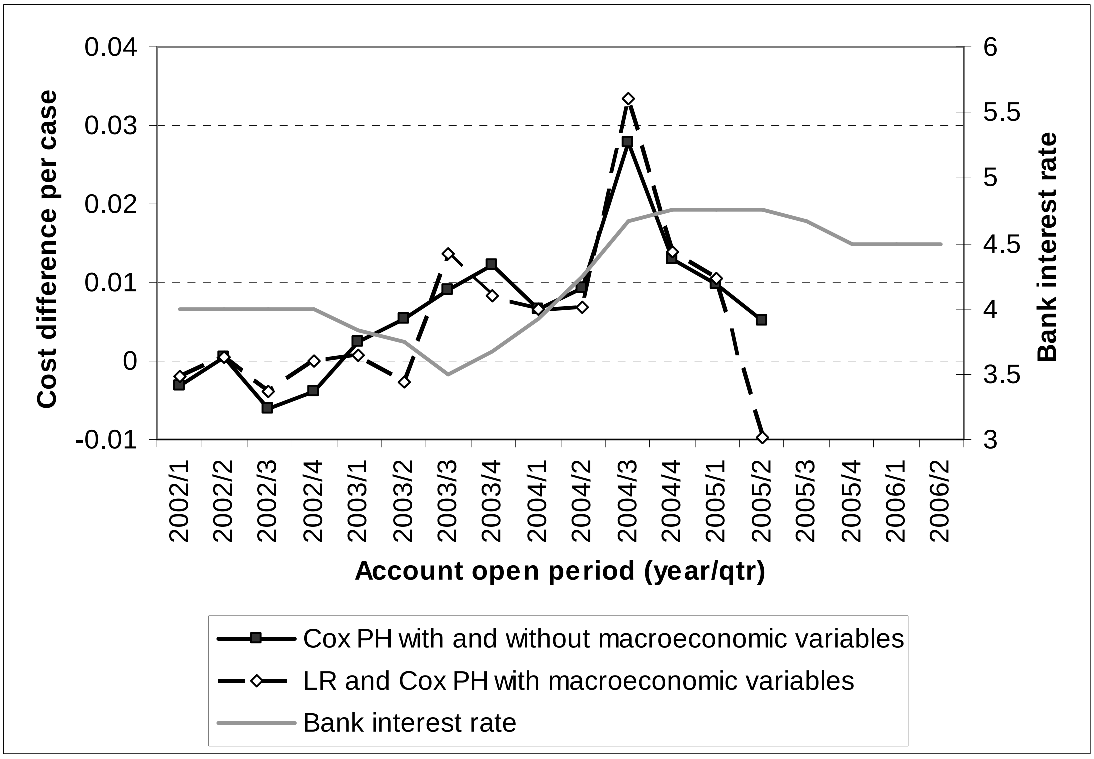


Figure 2. Mean cost differences on test data between models per quarter.



Notes.

1. Positive differences show improved performance with Cox PH model with MVs.
2. Bank interest rates are superimposed to show relationship with performance of MV models.
3. Cost on bad cases = 20.

Table 1. Macroeconomic variables

Code	*Macroeconomic variable*	*Data source*	*Expected effect on default risk*
IR	Interest rates: Selected UK Retail Banks Base Rate.	ONS	+ve
Earnings	Ratio of UK earnings including bonuses and retail price index on all items, not seasonally adjusted.	ONS	-ve
FTSE	FTSE all-share index.	Publicly available	-ve
Unemp	Unemployment rate for males unemployed for 6 to 12 months, seasonally adjusted.	ONS	+ve
Prod	Index of all UK production, not seasonally adjusted.	ONS	-ve
House	House price index.	Nationwide building society	
CC	UK consumer confidence index, not seasonally adjusted.	ONS	

ONS = Office of National Statistics

Table 2. Cox PH model coefficient estimates.

MVs	*Coefficient estimate*	*P-value*
IR	**0.133**	**0.001**
Earnings	**-7.08**	**0.046**
FTSE	0.000216	0.130
Unemp	0.0044	0.292
Prod	**-0.0491**	**0.050**
House	**0.821**	**0.048**
CC	0.0331	0.124
Selected interactions	*Coefficient estimate*	*P-value*
IR * income	**-0.00000174**	**0.028**
IR * unknown income (y/n)	-0.107	0.130
Earnings * home owner (y/n)	3.25	0.232
Earnings * private tenant (y/n)	8.59	0.057
Earnings * home council (y/n)	6.63	0.183
FTSE * home owner (y/n)	-0.00025	0.128
FTSE * private tenant (y/n)	-0.00045	0.058
FTSE * home council (y/n)	-0.00029	0.205
Unemp * employed (y/n)	-0.00349	0.323
Unemp * self-employed (y/n)	-0.00687	0.223
Unemp * unemployed (y/n)	0.0955	0.214
Prod * bureau score	**0.0000582**	**0.046**
House * home owner (y/n)	-0.721	0.083
House * private tenant (y/n)	-0.839	0.106
House * home council (y/n)	-0.282	0.339
CC * employed (y/n)	-0.0217	0.226
CC * self-employed (y/n)	-0.0457	0.060
CC * unemployed (y/n)	0.0789	0.304

Notes:

1. P-values computed for coefficient estimate equal to 0 using bootstrap percentile confidence intervals with 1,000 bootstrap iterations. Covariates that are significant at 0.05 level are highlighted.
2. (y/n) = yes/no indicator variable (yes=1, no=0).
3. Many variables which were included in the models are not reported in the table for confidentiality reasons.

Table 3. Marginal effects of MVs.

MV	*Marginal effect*	*Expected sign*		*Standard deviation of MV*	*Relative importance**
IR	+0.34	+	✓	0.767	0.261
Earnings	–3.56	–	✓	0.0257	0.0917
FTSE	–0.0000347	–	✓	405	0.0141
Unemp	+0.0017	+	✓	10.5	0.0179
Prod	+0.00465	–	×	5.29	0.0246
House	+0.136			0.130	0.0177
CC	+0.0108			2.90	0.0314

*relative importance = magnitude of standardized marginal effect

Table 4. Prediction results on test data set.

Cost on bad case	*Model*	*Mean cost of predictions on test data* (*when cut-off computed from* training *data)*	*(when cut-off computed from* test *data)*
15	LR	**0.2367**	0.2349
	Cox PH without MVs	**0.2364**	0.2342
	Cox PH with MVs	**0.2365**	0.2328
20	LR	**0.3067**	0.2987
	Cox PH without MVs	**0.3077**	0.2978
	Cox PH with MVs	**0.3014**	0.2940
25	LR	**0.3580**	0.3523
	Cox PH without MVs	**0.3532**	0.3510
	Cox PH with MVs	**0.3514**	0.3463

Table 5. Significance of uplift on test data set due to MVs

when Cost on bad cases=20.

	Costs on models compared	*Mean cost difference*	*N*	*t*	*p-value*
(when cut-off computed from training *data)*	Cox PH with MVs v. LR	**0.0053**	**Over 100000**	**3.54**	**0.0004**
	Cox PH with MVs v. Cox PH without MVs	**0.0062**	**Over 100000**	**4.23**	**<0.0001**
(when cut-off computed from test *data)*	Cox PH with MVs v. LR	0.0048	Over 100000	2.52	0.0119
	Cox PH with MVs v. Cox PH without MVs	0.0038	Over 100000	2.52	0.0117

Figure 1. Importance of each MVs in Cox PH model.

Figure 2. Mean cost differences on test data between models per quarter.

Table 1. Macroeconomic variables.

Table 2. Cox PH model coefficient estimates.

Table 3. Marginal effects of MVs.

Table 4. Prediction results on test data set.

Table 5. Significance of uplift on test data set due to MVs when Cost on bad cases=20.