The Stability of Survival Model Parameter Estimates for Predicting the Probability of Default: Empirical Evidence over the Credit Crisis

The Stability of Survival Model Parameter Estimates for Predicting the Probability of Default: Empirical Evidence over the Credit Crisis

This version: 15 February 2014

Mindy Leow[a,b] & Jonathan Crook[a,c]

---

[a] Credit Research Centre, University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS
[b] mindy.leow@ed.ac.uk
[c] j.crook@ed.ac.uk

Abstract

Using a large portfolio of credit card loans observed between 2002 and 2011 provided by a major UK bank, we investigate the stability of the parameter estimates of discrete survival models, especially since the start of the credit crisis of 2008. Two survival models are developed for accounts that were accepted before and since the crisis. We find that the two sets of parameter estimates are statistically different to each other. By applying the estimated parameters onto a common test set, we also show that they give different predictions of probabilities of default. The changes in the predicted probability distributions are then investigated. We theorise them to be due to the quality of the cohort accepted under different economic conditions, or due to the drastically different economic conditions that was seen in the UK economy, or a combination of both. We test for each effect.

Keywords: forecasting, robustness and sensitivity analysis, macroeconomic variables, monitoring forecasts, structural change, probability forecasting

1. Introduction

The application of survival analysis models onto credit-related problems is not new (for example, see Banasik et al. (1999), Pennington-Cross (2010)) and is welcomed for its ability to take into account factors that are inherent in the modelling of credit risk and the prediction of credit events, where regression methods are unable to. First, survival models are able to account for censoring, which allows for a realistic and practical model to be developed. Second, they are able to incorporate time-dependent variables with ease, which will allow the inclusion of time-dependent account-specific covariates as well as time-dependent macroeconomic variables in credit models. When this is combined with simulation, a plausible platform for stress testing is created, as proposed by Rodriguez and Trucharte (2007), Leow et al. (2011) and Bellotti and Crook (2013a, 2013b). Third, and most crucially, survival models are able to generate probabilities of how likely an event will occur over time, conditional on the event not having occurred before, and this provides a dynamic framework for the prediction of credit events (e.g. default or customer churn of credit loans, repossession or early-prepayment for mortgage loans). Because the likelihood of the credit event occurring over time can be estimated, the corresponding losses (McDonald et al. (2010)) or profits (Ma et al. (2010)) can also be predicted. In terms of how well survival models predict, there has been some work done specifically to compare its prediction to that of regression models: Stepanova and Thomas (2002) looked at the model performances in the prediction of early prepayment and default of personal loans; Bellotti and Crook (2009) looked at model performances in the prediction of default of credit card loans. Both papers found that survival models are able to predict better than static regression models.

This work does not attempt to revisit the advantages of survival models over their regression counterparts – that much has been established in the literature over different retail products. The work here differs from the existing literature in two ways. First, we have a rich source of credit card loan data that goes from 2002 to 2011, and so encompasses the credit crisis from 2008, which is not commonly available. Macroeconomic indicators over time will show a large difference in values, and it would be interesting to explore how these large and unexpected changes would affect default models and their predictions. Second, we investigate the stability of survival model parameter estimates before and after the credit crisis. Using a portfolio of active credit cards observed between January 2002 and March 2011, we investigate whether parameter estimates change over the crisis period, and whether the inclusion of time-varying covariates representing the economy are able to adequately account for changes to debtors propensity to default. By separately and

independently estimating a survival model for periods before and since the start of the credit crisis, i.e. 2002 to 2007 and 2008 to 2011 respectively, we use the Chow Test (more details in Section 4.1) to check for statistical differences between the two sets of parameter estimates. To illustrate how the two sets of parameter estimates are different, we apply each survival model developed onto a common test set to get the average predicted probabilities over the (duration) time of the loan.

During the course of this work, population drift, and how it might affect parameter estimates, is also considered as a related issue, due to the differing types of debtors securing credit accounts before and during the credit crisis. However, because of the large variations in macroeconomic conditions that was seen in our period of interest, it is also possible that changes in distributions of probabilities are due to the changes in these macroeconomic variables. We investigate the effects of either by selecting two cohorts, representing a set of accounts accepted during a non-downturn period and a downturn period, and estimating a survival model for each period. We then create test sets based on each training set, by holding constant either the cohort quality or the macroeconomic conditions, and compare the distribution of predicted probabilities to see how the distributions shift due to changes in cohort quality or economic conditions. We find macroeconomic conditions do affect probabilities of default, and could affect different groups of debtors in different ways.

2. Methodology

We use data gathered at regular, discrete monthly points in time, and the default event is recorded in a particular month with reference to the month the account was open. Therefore we estimate the survival models in discrete time. Another advantage of discrete time rather than continuous time survival models is a much lower computational time in model estimation. This is important because we deal with a large dataset.

Let $P_{i\tau}$ be the probability that an individual account *i* goes into default at duration time (of loan) τ, given that default has not happened up to time $\tau - 1$, and the final model developed is given in Equation 1.

$$\log\left(\frac{P_{i\tau}}{1-P_{i\tau}}\right) = \alpha_\tau + \beta_1 \mathbf{X}_i + \beta_2 \mathbf{Y}_{i\tau-3} + \beta_3\left(\mathbf{Z}_{\tau-3} - \mathbf{Z}_{\tau-15}\right) + \beta_4 \mathbf{X}_i\left(\mathbf{Z}_{\tau-3} - \mathbf{Z}_{\tau-15}\right) \quad (1)$$

where α_τ represents the effect of time on the odds of default; $\mathbf{X}_i$ is a vector which represents the time-independent, account-dependent covariates, i.e. application variables; $\mathbf{Y}_{i\tau-3}$ is a vector which represents time- and account-dependent covariates, i.e. behavioural variables, lagged 3 months;

$\mathbf{Z}_{\tau-3} - \mathbf{Z}_{\tau-15}$ is a vector which represents time-dependent, account-independent covariates, i.e. macroeconomic variables, at 12th difference and lagged 3 months; and $\mathbf{X}_i\left(\mathbf{Z}_{\tau-3} - \mathbf{Z}_{\tau-15}\right)$ is a vector which represents interaction terms between selected application variables and macroeconomic variables at 12th difference and lagged 3 months.

In this regression model, the dependence of the hazard on time, α_t, is specified as $\alpha_1\tau + \alpha_2\tau^2 + \alpha_3 \ln\tau + \alpha_4 (\ln\tau)^2$. By doing so, we allow the relationship between effect on time and the odds of default to be very flexible with an added advantage of allowing for prediction beyond the maximum duration time that is observed in the training set.

A number of model variations were considered in the course of this work, mainly experimenting with the way the macroeconomic variables were included in the model. Lags of between 3 and 12 months were considered, and to address the possible correlation between macroeconomic variables, both levels and 12th differences, lagged or otherwise, of each macroeconomic variable were examined.

3. Data

The data is supplied by a major UK bank and is a random sample of credit cards that were issued in the UK between 2002 and 2010. It consists of almost 538,000 unique credit card accounts and each account is tracked monthly up to March 2011, or until the time the credit card account is closed, whichever is earlier. Common application variables are available: type of employment, length of time the debtor has been with the bank, income at application and age at application, among others. Because each account is updated monthly, we also have behavioural variables, including repayment amount, credit limit and outstanding balance, from which further behavioural indicators could be inferred, for example, how frequently the account misses payment(s) over its entire history. Any behavioural variables included in the model are lagged by 3 months.

Although default information is available from the dataset, it is not consistently defined across the entire dataset. Therefore, a monthly minimum repayment amount is defined and is used to define arrears and default. This minimum repayment amount is 2.5% of the previous month's outstanding balance or £5, whichever is higher, unless the account is in credit, in which case the minimum repayment amount is £0, or the account has an outstanding balance of less than £5, in which case the minimum repayment amount would be the full outstanding amount. An account is said to be in

arrears if it does not make the minimum payment. A default is said to occur if and when an account goes three months in arrears (not necessarily consecutive). Note that this definition of default is not the conventional "three consecutive months of missed payment", but is acceptable as financial institutions are not bound to this definition of default (Basel Committee on Banking Supervision (2004), Paragraphs 452-456). As the work here only focuses on the default event, we do not specify the transitions between states of arrears in the preceding months; further details can be found in Leow and Crook (2014).

2.1. Training and test set splits

The dataset is used in a number of ways here. In order to accommodate the lagged behavioural covariates, only accounts that are observed for longer than three months since each was opened are included.

First, the dataset is split into two training sets (see Table 1). The first consists of accounts that started between January 2002 and December 2007 inclusive, with an observation period up to December 2007, i.e. any remaining active accounts are censored in December 2007. The second consists of accounts that started between January 2008 and July 2010 inclusive, with an observation period up to March 2011, i.e. accounts are censored in March 2011, if the account has not been closed earlier. Note that the two training sets are completely separate. The creation of these two training sets represent portfolios of loans that were accepted before and during the credit crisis, since we expect bank policies and acceptance decisions to change slightly over the years, with distinguished differences before and since the start of the credit crisis.

Table 1: Dataset splits

Dataset	**Acceptance period**	**Observation period**
Training set I	January 2002 to August 2007	May 2002 to December 2007
Training set II	January 2008 to July 2010	May 2008 to March 2011
Combined / "Test"	January 2002 to July 2010	May 2002 to March 2011

Due to the split of the training sets, it is not sensible to try and reduce the length of either training set further to get a test set. In order to get an indication of how similar (or different) the models of each training set would predict, we apply the respective models onto the entire dataset, i.e. combining training sets I and II, as a test set. Doing so would mean that a common test set is used without any further loss to both training sets in terms of observations and observation period.

2.2. Macroeconomic variables

The macroeconomic variables considered are given in Table 2. The main source of macroeconomic variables is the Office of National Statistics (ONS), supplemented by data from Bank of England (BOE), Nationwide and the European Commission (EC) where appropriate. The non-seasonally adjusted series is selected unless unavailable because the default indicator is not seasonally adjusted. Based on commentary from key industry contacts, UK banks increase their market share by lowering cut-off thresholds on application scorecards and extending credit for current borrowers, so this is taken into account with the inclusion of total consumer credit outstanding. In order to reduce the impact of trends, the macroeconomic variables are included in the model as its 12th difference, lagged 3 months. Interaction terms between selected macroeconomic variables and application variables are also considered.

Table 2: Table of macroeconomic variables

Variable	**Source**	**Description**
AWEN	ONS	Average earnings index, including bonus, including arrears, whole economy, not seasonally adjusted
CIRN	BOE	Monthly weighted average of UK financial institutions' interest rate for credit card loans to households, not seasonally adjusted
CLMN	ONS	Claimant count rate, UK, percentage, not seasonally adjusted
CONS	EC	Total consumer confidence indicator, UK, seasonally adjusted
HPIS	Nationwide	All houses, seasonally adjusted
IOPN	ONS	Index of production, all production industries, not seasonally adjusted
IRMA	BOE	Monthly average of Bank of England's base rate
LAMN	ONS	Log (base e) of total consumer credit, amounts outstanding, not seasonally adjusted
LFTN	ONS	Log (base e) of FTSE all share price index, month end, not seasonally adjusted
MIRN	BOE	Monthly weighted average of UK financial institutions' interest rate for loans secured on dwellings to households, not seasonally adjusted
RPIN	ONS	All items retail price index, not seasonally adjusted
UERS	ONS	Labour Force Survey unemployment rate, UK, all, ages 16 and over, percentages, seasonally adjusted

4. Results

4.1. Parameter estimates

The parameter estimates from training sets I and II representing accounts that were accepted before the crisis and since the crisis started respectively, are given in Table 3 Due to confidentiality agreements, we are unable to detail all variables used in the model.

Table 3: Parameter estimates for PD model for accounts accepted pre-crisis and after the crisis started.

Code	**Variable**	**PRE-CRISIS**			**CRISIS**		
		Estimate	**p < 0.05**	**p < 0.01**	**Estimate**	**p < 0.05**	**p < 0.01**
Intercept		-3.9076		**	-27.4414		**
Application variables							
NOCards	Number of cards	0.0341		**	0.0805		**
INC_L	Income, ln	-0.4172		**	-0.0836		**
INC_M0	Income, missing or 0	-4.0706		**	-0.4216		**
ageapp_1	Age group 1						
ageapp_2	Age group 2	0.0071			-0.0668		**
ageapp_3	Age group 3	0.0231			-0.0700		**
ageapp_4	Age group 4	0.0304			-0.0515		
ageapp_5	Age group 5	0.0271			-0.1126		**
ageapp_6	Age group 6	-0.0438	*		-0.1622		**
ageapp_7	Age group 7	-0.0693		**	-0.2637		**
ageapp_8	Age group 8	-0.0287			-0.3339		**
ageapp_9	Age group 9	-0.3979		**	-0.3823		**
ageapp_10	Age group 10	-0.4241		**	-0.6817		**
X_A	Variable X, group A						
X_B	Variable X, group B	0.1830		**	-0.0693	*	
X_C	Variable X, group C	0.1524		**	-0.0502		**
X_D	Variable X, group D	0.0219			-0.1117		**
X_E	Variable X, group E	0.0370		**	-0.1643		**
ECode_A	Employment code group A						
ECode_B	Employment code group B	-0.0303	*		-0.0122		
ECode_C	Employment code group C	-0.0112			-0.1468		

ECode_D	Employment code group D	0.1945	*	0.0806	
ECode_E	Employment code group E	0.1434	**	0.2490	**
Behavioural variables, lagged 3 months					
LPAY_lag3	Repayment amount, ln	-0.1053	**	-0.1161	**
LCRL_lag3	Credit limit, ln,	0.5198	**	0.5770	**
PARR_lag3	Proportion of months in arrears	2.2486	**	2.1021	**
PRDR_lag3	Proportion of credit drawn	3.6999	**	4.2920	**
Macroeconomic variables, differenced 12 months, lagged 3 months					
CIRN_d12_lag3m	Credit card interest rate	-0.1714	**	0.0267	
RPIN_d12_lag3m	Retail price index	0.1228	**	-0.0061	
AWEN_d12_lag3m	Average wage earnings	-0.2911	**	-0.0322	
LFTN_d12_lag3m	FTSE Index, ln	3.8616	**	-0.3210	**
UERS_d12_lag3m	Unemployment rate	-0.3190	**	-0.0182	
IOPN_d12_lag3m	Index of production	-0.0008		-0.0003	
HPIS_d12_lag3m	House price index	0.0026	**	0.0114	**
CONS_d12_lag3m	Consumer confidence	-0.0452	**	-0.0080	**
LAMN_d12_lag3m	Total credit outstanding, ln	3.4840	**	5.7486	**
Interaction terms					
INCL_RPd3	Income, ln * Retail price index	0.0056		-0.0030	**
INCL_AWd3	Income, ln * Average wage earnings	0.0238	**	0.0033	
INCM_RPd3	Income, missing * Retail price index	0.0695		-0.0211	
INCM_AWd3	Income, missing * Average wage earnings	0.2746	**	0.0159	
AAP8_CId3	Age group 8 * Credit card interest rate	-0.0260		0.0018	
AAP8_AWd3	Age group 8 * Average wage earnings	-0.0216		0.0038	
AAP9_CId3	Age group 9 * Credit card interest rate	0.0072		-0.2876	
AAP9_AWd3	Age group 9 * Average wage earnings	0.0184		0.0014	
AAP10_CId3	Age group 10 * Credit card interest rate	-0.0232		-0.1651	

AAP10_AWd3	Age group 10 * Average wage earnings	-0.0084		0.0080	
ECC_CId3	Employment group C * Credit card interest rate	-0.0305		-0.0602	
ECC_RPd3	Employment group C * Retail price index	-0.0017		0.0171	**
ECD_CId3	Employment group D * Credit card interest rate	0.0548		0.3505	**
ECD_RPd3	Employment group D * Retail price index	0.0004		0.0127	*
Variables for dependence of hazard on time					
ctime	calendar time (in months, referenced from December 2000)	0.0017		0.0167	**
t	duration time	0.1518	**	3.0308	**
tsq	duration time, squared	-0.0013	**	-0.0248	**
lnt	duration time, ln	-1.8893	**	17.1773	**
lntsq	duration time, ln, squared	-0.0894		-9.6266	**

The single asterisk (*) and double asterisk (**) represent variables that are statistically significant at the 0.05 and 0.01 levels, i.e. * $p < 0.05$ and ** $p < 0.01$.

The application variables are fairly stable, with most of the younger borrowers not significantly different from each other before the crisis, but becoming significant since the crisis. Variable X is an interesting categorical variable which has its sign changed before and during the crisis period. Employment status of the borrower does not seem to have much effect on default probability. The parameters on the behavioural variables are very stable, with very similar estimates over the two models. On the other hand, the parameters on the macroeconomic variables are not, and vary in terms of statistical insignificance over the two periods, as well as in terms of parameter estimates signs. Given the instability of these macroeconomic variables, it is not surprising to see that most of the interaction terms are statistically insignificant. Based on information from key industry contacts, we know the credit cards portfolio experienced a macroeconomic downturn two to three years earlier than that of the credit crisis (c.f. Figure A1 in the Appendix), and this would not be reflected in the significance of general macroeconomic variables. However, perhaps due to the way the training sets were defined, this effect is not obviously captured by the covariates that were used in the model. Although it is possible to try and include more economic variables that are relevant to the type of loan here, e.g. economic indicators on a household level or retail loans write off rates,

most of these variables are either not available for as far back as our dataset period, or are only available on an annual basis.

4.2. Chow test

The Chow Test is a test of equality between parameter estimates of two linear regression models developed on different datasets, first developed by Chow (1960). An equivalent test for use in logistic regression models is the Chow Test Analogue, given in DeMaris (2004). Logistic regression models are each developed for training sets I and II, and the combined dataset, given in Equations 2 to 4.

$$\text{combined:} \quad \log\left(\frac{P_{i\tau}}{1\text{-}P_{i\tau}}\right) = \alpha_\tau + \beta_1 \mathbf{X}_i + \beta_2 \mathbf{Y}_{i\tau-3} + \beta_3\left(\mathbf{Z}_\tau - \mathbf{Z}_{\tau-12}\right) + \beta_4 \mathbf{X}_i\left(\mathbf{Z}_\tau - \mathbf{Z}_{\tau-12}\right) \tag{2}$$

$$\text{training set I:} \quad \log\left(\frac{P_{i\tau}}{1\text{-}P_{i\tau}}\right) = \alpha_\tau + \gamma_1 \mathbf{X}_i + \gamma_2 \mathbf{Y}_{i\tau-3} + \gamma_3\left(\mathbf{Z}_\tau - \mathbf{Z}_{\tau-12}\right) + \gamma_4 \mathbf{X}_i\left(\mathbf{Z}_\tau - \mathbf{Z}_{\tau-12}\right) \tag{3}$$

$$\text{training set II:} \quad \log\left(\frac{P_{i\tau}}{1\text{-}P_{i\tau}}\right) = \alpha_\tau + \varsigma_1 \mathbf{X}_i + \varsigma_2 \mathbf{Y}_{i\tau-3} + \varsigma_3\left(\mathbf{Z}_\tau - \mathbf{Z}_{\tau-12}\right) + \varsigma_4 \mathbf{X}_i\left(\mathbf{Z}_\tau - \mathbf{Z}_{\tau-12}\right) \tag{4}$$

The null hypothesis states that the parameter estimates from training sets I and II are equal, i.e. $\gamma_1 = \varsigma_1, \gamma_2 = \varsigma_2, \gamma_3 = \varsigma_3$. For two groups, the test statistic follows a chi-squared distribution, with degrees of freedom calculated to be the total number of parameters in the models of training sets I and II less the number of parameters in the combined dataset, given in Equation 5.

$$\chi^2 = -2\ln L_c - \left[-2\ln L_1 + \left(-2\ln L_2\right)\right] \tag{5}$$

where $-2\ln L_c$, $-2\ln L_1$, $-2\ln L_2$ are the fitted log-likelihood of the combined model and models from training sets I and II respectively.

The results[1] reject the null hypothesis, i.e. the parameter estimates of models developed on training sets I and II are statistically significantly different from each other.

[1] Besides performing the Chow Test on the final model, we also repeat the test for two other models: without interactions terms and behavioural variables, and without interactions terms. All three test statistics indicate that the parameter estimates from the two training sets are significantly different.

We also look at the predicted probabilities of default as predicted by the two models. By applying the parameter estimates onto the test set, predicted probabilities of default for each discrete time point of each account can be calculated. The predicted probability of default at each time point is then calculated to be the mean probability of default for all accounts that are at risk of default at that time, given in Equation 6.

$$\hat{D}_\tau = \frac{\sum_{j \in R_\tau} p_{j\tau}}{\sum_{j \in R_\tau} n_j} \tag{6}$$

where R_τ denotes the risk set, i.e. all active accounts, at time τ, and $n_j = 1$ if account $j \in R_\tau$, 0 otherwise.

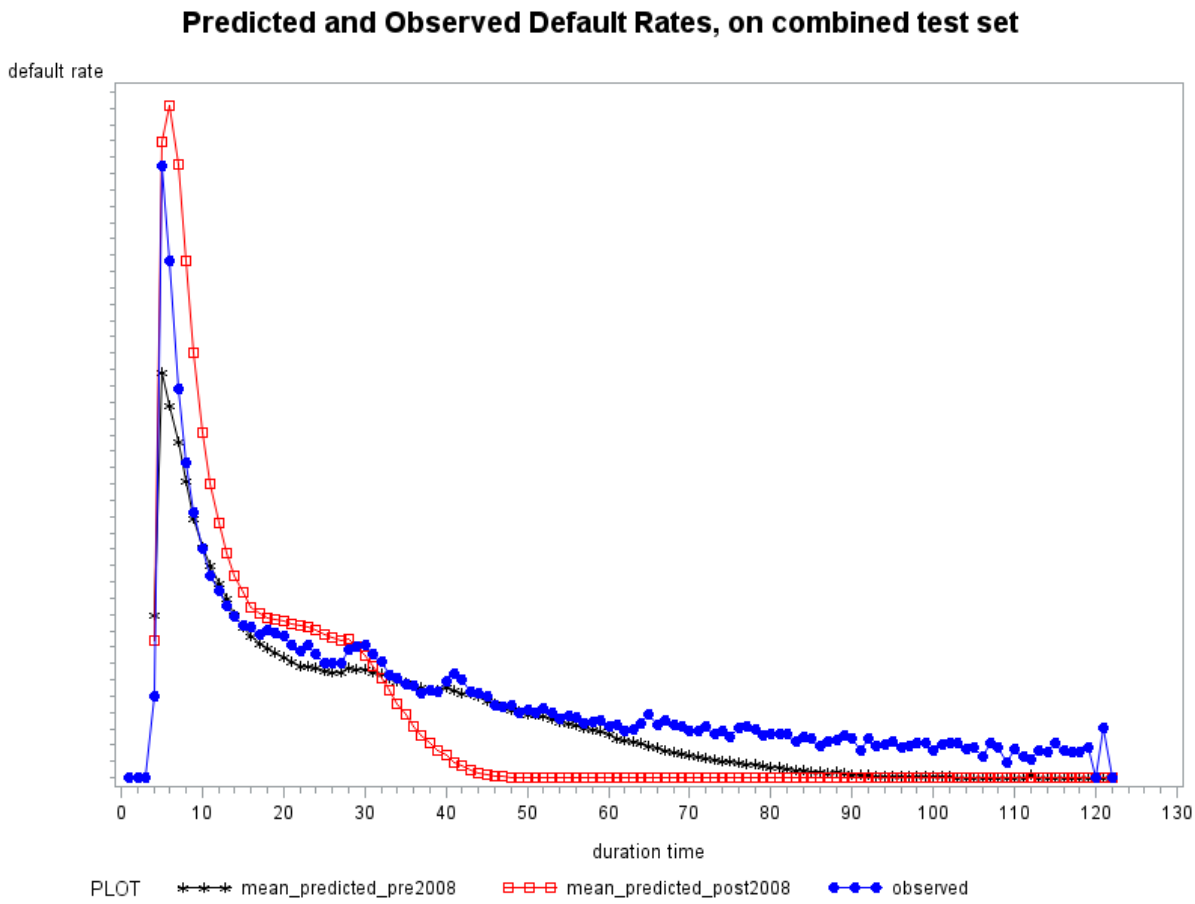

Figure 1: Predicted and observed default rates for combined test set. The solid dots represent the observed default rate; the asterisks represent the predicted default rate from the model developed on training set I, i.e. pre-2008; the squares represent the predicted default rate from model developed on training set II, i.e. 2008 and onwards. Due to confidentiality issues, the values of the vertical axis are omitted.

The predicted probability of default from the models based on training sets I and II are applied onto the test set to see how the predictions differ. Together with the observed default rate from the test set, all three are plotted on the same graph, given in Figure 1. We note that although there are potentially, two very different models within the period of the test set, we are not comparing how well each model predicts, but how differently the two models predict for each other. An alternative would be to have two separate test sets for each training period, but that would not provide the same level of comparison which is achieved here.

The differences between the two predicted hazard rate plots in Figure 1 are due to parameterising the same model specification using two different training sets; the test set is the same. In other words the differences are due to differences in the estimated parameters between the two training periods: pre- and post- crisis. We see that the model based on pre-crisis accounts significantly under-estimate default rates in the first 12 months of a loan but estimates well for the rest of the loan, while the model based on crisis accounts slightly over-estimates default rates in the first 30 months of the loan and under-estimate default rates after that.

4.4 Hazard Distributions

We now investigate the sources of changes in the distributions of hazard rates before and after the crisis. The models that have been estimated using training data are equations 3 and 4. Simplifying, these may be represented as equations 7 and 8 as follows:

$$\hat{y}_{it} = \hat{\boldsymbol{\beta}}_{1a}^{T}\mathbf{x}_{i1} + \hat{\boldsymbol{\beta}}_{1b}^{T}\Delta_{12}\mathbf{z}_{1t-l} + \hat{\beta}_{1c}^{T}\mathbf{x}_{i1}\Delta_{12}\mathbf{z}_{1t-l} \tag{7}$$

$$\hat{y}_{it} = \hat{\boldsymbol{\beta}}_{2a}^{T}\mathbf{x}_{i2} + \hat{\boldsymbol{\beta}}_{2b}^{T}\Delta_{12}\mathbf{z}_{2t-l} + \hat{\beta}_{2c}^{T}\mathbf{x}_{i2}\Delta_{12}\mathbf{z}_{2t-l} \tag{8}$$

where y_{it} denotes the logit of the hazard probabilities, $\mathbf{x}_{i1}(\mathbf{x}_{i2})$ denotes the vector of application variables for individuals *i* from period 1 (2), that is in the pre- (post-) crisis period and $\mathbf{z}_{1,t-l}$ ($\mathbf{z}_{2,t-l}$) denotes the vector of macroeconomic variables measured in period 1 (2) respectively, lagged *l* months. $\boldsymbol{\beta}_1(\boldsymbol{\beta}_2)$ represent the vector of parameters $\hat{\boldsymbol{\beta}}_{1a}, \hat{\boldsymbol{\beta}}_{1b}, \hat{\boldsymbol{\beta}}_{1c}$ ($\hat{\boldsymbol{\beta}}_{2a}, \hat{\boldsymbol{\beta}}_{2b}, \hat{\boldsymbol{\beta}}_{2c}$)that have been estimated using period 1 (2) data.

Table 4: Training sets and created test sets, and their corresponding statistics.

Acceptance Policy	**Training set**	**Parameters**	**Test sets**							
			A		B		C		D	
Normal	cohort 2002-2004 ($\mathbf{X}_1$) macro 2002-2005 ($\Delta\mathbf{Z}_1$)	$\boldsymbol{\beta}_1$	cohort 2002-2004 ($\mathbf{X}_1$) macro 2008-2011 ($\Delta\mathbf{Z}_2$)		cohort 2008-2010 ($\mathbf{X}_2$) macro 2002-2005 ($\Delta\mathbf{Z}_1$)		cohort 2002-2004 ($\mathbf{X}_1$) macro 2002-2005 ($\Delta\mathbf{Z}_1$)		cohort 2008-2010 ($\mathbf{X}_2$) macro 2008-2011 ($\Delta\mathbf{Z}_2$)	
			ND	*D*	*ND*	*D*	*ND*	*D*	*ND*	*D*
		mean	0.1975	0.2047	0.0011	0.0012	0.1173	0.1696	0.0112	0.0098
		standard deviation	0.2970	0.3166	0.0014	0.0014	0.0807	0.0925	0.0272	0.0208
		median	0.0131	0.0049	0.0008	0.0009	0.0975	0.1556	0.0032	0.0024
		mode	0.0002	0.0002	0.0007	0.0007	0.2054	0.2054	0.0057	0.0057
			E		F		G		H	
Downturn	cohort 2008-2010 ($\mathbf{X}_2$) macro 2008-2011 ($\Delta\mathbf{Z}_2$)	$\boldsymbol{\beta}_2$	cohort 2008-2010 ($\mathbf{X}_2$) macro 2002-2005 ($\Delta\mathbf{Z}_1$)		cohort 2002-2004 ($\mathbf{X}_1$) macro 2008-2011 ($\Delta\mathbf{Z}_2$)		cohort 2008-2010 ($\mathbf{X}_2$) macro 2008-2011 ($\Delta\mathbf{Z}_2$)		cohort 2002-2004 ($\mathbf{X}_1$) macro 2002-2005 ($\Delta\mathbf{Z}_1$)	
			ND	*D*	*ND*	*D*	*ND*	*D*	*ND*	*D*
		mean	0.3421	0.4033	0.1116	0.1183	0.1132	0.1350	0.3427	0.3532
		standard deviation	0.1598	0.1483	0.0518	0.0502	0.0572	0.0527	0.1420	0.1203
		median	0.3415	0.3906	0.1089	0.1130	0.1127	0.1271	0.3375	0.3404
		mode	0.3204	0.3204	0.0941	0.0941	0.0899	0.0899	0.2753	0.2753

"ND" refers to non-default accounts, "D" refers to default accounts.

It can now be seen that there are at least three sources of differences between the distributions of hazards before and after the crisis. These are differences in the estimated parameters, differences in the distributions of the application variables (X values) and differences in the distributions of the macroeconomic variables (Z values). To isolate the effects of each we hold two constant and vary the third, for each source, in turn. Note that for these predictions we omitted the behavioural variables as we are not able to predict how these variables would react to the changes in macroeconomic variables. The interaction terms are also updated with the changed macroeconomic conditions correspondingly.

The specific set-up is shown in Table 4. In the top panel, the model is estimated using cohort 2002-2004 as the training sample yielding parameters denoted by $\boldsymbol{\beta}_1$. This cohort was accepted under a 'non-crisis', that is normal conditions, cut-off rule. In the lower panel the model is estimated using cohort 2008-2010 as the training sample yielding parameters denoted by $\boldsymbol{\beta}_2$. This cohort was accepted during the crisis period, that is 2008-2010. These identifiers are given in columns 1 and 2. The remaining columns identify the combinations of the application and macroeconomic variables that are used to yield predicted hazard rates in the test sets. For example, test set A refers to an out-of-sample test set with values of application variables measured in 2002-2004 ($\mathbf{X_1}$) and values of macroeconomic variables taken during 2008-2011 ($\Delta\mathbf{Z}_2$), scored using the $\boldsymbol{\beta}_1$ values; test set E predictions are made using the $\boldsymbol{\beta}_2$ parameter values, values of application variables for the cohort 2008-2010 ($\mathbf{X}_2$) and values of macroeconomic variables for 2002-2005 ($\Delta\mathbf{Z}_1$). Immediately below each test set are some statistics based on the predicted probabilities, further segmented by defaults (D) and non-defaults (ND). The remaining cells are to be interpreted similarly. Since the distributions of the hazards are skewed we concentrate on the differences in the median and modal values so of the distributions[2]. Table 5 summarises the results.

First, we examine the effects of differences in the training cohorts holding the application and macroeconomics conditions constant. Four comparisons are possible since we have two possible values for the application variables and two possible values for the macroeconomic conditions (see Panel A of Table 5). A comparison of the distributions generated by values in test sets A (pre-crisis) and F (post-crisis) isolate the effects of the parameter differences between the crisis and non-crisis periods by holding the X values at the 2002-2004 values and the Z values at 2008-2011 values. It can be seen that the changes in the parameters from the pre to the post-crisis periods led to an increase

[2] Due to the number of comparisons we are making in this work and the similarities of the graphs, we have chosen not to include the graphs of distributions of each pair-wise comparison.

in the median and modal hazards for both the non-default cases and the default cases. The spread, measured by the standard deviation, decreased. If we compare test sets B (pre-crisis) and E (post crisis) we see the effects of changes in the parameters holding X at the 2008-2010 values and the Z at the 2002-2005 values. Again the median and modal values of the hazards increased after the crisis, but the spread actually increased. Comparing C with H fixes the application and the macroeconomic variables values at the pre-crisis levels and again moving from the pre- to the post-crisis models results in increases in the median and modal values of the hazards. The comparison D with G shows the same qualitative results. In conclusion, the parameters of survival models of default changed between the pre-crisis and the post-crisis periods, where the model developed using downturn data ($\boldsymbol{\beta}_2$) consistently gives higher predictions of hazards across different macroeconomic conditions or cohorts. It would seem that hazards from $\boldsymbol{\beta}_1$ are under-estimated, but it is not clear whether the predicted hazards from $\boldsymbol{\beta}_2$ have compensated enough.

Next, we examine the effects of changes in the distributions of the application characteristics of credit card holders on predicted hazards. Again four comparisons are possible: test sets A with D, C with B, and also F with G and E with H (see Panel B of Table 5). If we compare test sets A and D, both test sets have the same parameter values ($\boldsymbol{\beta}_1$) and both have the same values of the macroeconomic variables (2008-2011), so the differences in predicted hazards are due to differences in the application variables (**X**). The results show that moving from the pre- to the post-crisis values the median (and the mean) of the hazard rates decreased considerably for the default and non-default samples whilst the mode actually increased for both groups. The spreads also fell. Comparing test sets C and B, we again condition on $\boldsymbol{\beta}_1$ but now fix the fix the macroeconomic variables at pre-crisis levels (2002-2005). Moving from the pre- to the post-crisis cohorts, the median and the modal hazards both fell. We also make the comparisons by fixing the model to be the post-crisis model ($\boldsymbol{\beta}_2$) by comparing test sets F and G. Now conditioning on the parameters gained from the crisis cohort and moving from the pre- to the post-crisis cohorts, we see only very slight changes in all values of mean, median, mode and spread. This is the same case when we compare test sets E and H. In conclusion, the effects of changes in cohorts in the test set depend on the cohort used to train the model. Assuming the bank does become more stringent with its acceptance policy since the crisis, it is likely that the post-crisis cohort is less risky, which the model developed on non-downturn data ($\boldsymbol{\beta}_1$) is suggesting. However, the model developed on downturn data ($\boldsymbol{\beta}_2$) is not able to differentiate between cohorts even when macroeconomic conditions are held constant.

Table 5: Comparisons of test sets

Test set comparisons	Holding constant (test set characteristics)	Change from pre-crisis	Change to post-crisis	Change (in median)	Possible Interpretations
PANEL A: Change in parameter estimates					
C vs H	macro effect ($\Delta\mathbf{Z}_1$), cohort effect ($\mathbf{X}_1$)	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	increase in hazards	$\boldsymbol{\beta}_2$ consistently gives higher predicted hazards across different macroeconomic conditions and cohorts
B vs E	macro effect ($\Delta\mathbf{Z}_1$), cohort effect ($\mathbf{X}_2$)	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	increase in hazards	
A vs F	macro effect ($\Delta\mathbf{Z}_2$), cohort effect ($\mathbf{X}_1$)	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	increase in hazards	
D vs G	macro effect ($\Delta\mathbf{Z}_2$), cohort effect ($\mathbf{X}_2$)	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	increase in hazards	
PANEL B: Change in cohort					
C vs B	parameters ($\boldsymbol{\beta}_1$), macro effect ($\Delta\mathbf{Z}_1$)	$\mathbf{X}_1$	$\mathbf{X}_2$	decrease in hazards	$\boldsymbol{\beta}_1$ gives lower predictions of hazards when the cohort changes
A vs D	parameters ($\boldsymbol{\beta}_1$), macro effect ($\Delta\mathbf{Z}_2$)	$\mathbf{X}_1$	$\mathbf{X}_2$	decrease in hazards	
E vs H	parameters ($\boldsymbol{\beta}_2$), macro effect ($\Delta\mathbf{Z}_1$)	$\mathbf{X}_1$	$\mathbf{X}_2$	almost no change	$\boldsymbol{\beta}_2$ is not able to differentiate between cohorts even when macroeconomic conditions are held constant
F vs G	parameters ($\boldsymbol{\beta}_2$), macro effect ($\Delta\mathbf{Z}_2$)	$\mathbf{X}_1$	$\mathbf{X}_2$	almost no change	
PANEL C: Change in macroeconomic conditions					
C vs A	parameters ($\boldsymbol{\beta}_1$), cohort effect ($\mathbf{X}_1$)	$\Delta\mathbf{Z}_1$	$\Delta\mathbf{Z}_2$	decrease in hazards	$\boldsymbol{\beta}_1$ gives conflicting trends depending on the cohort tested
B vs D	parameters ($\boldsymbol{\beta}_1$), cohort effect ($\mathbf{X}_2$)	$\Delta\mathbf{Z}_1$	$\Delta\mathbf{Z}_2$	increase in hazards	
H vs F	parameters ($\boldsymbol{\beta}_2$), cohort effect ($\mathbf{X}_1$)	$\Delta\mathbf{Z}_1$	$\Delta\mathbf{Z}_2$	decrease in hazards	$\boldsymbol{\beta}_2$ gives lower predictions of hazards when macroeconomic conditions changes
E vs G	parameters ($\boldsymbol{\beta}_2$), cohort effect ($\mathbf{X}_2$)	$\Delta\mathbf{Z}_1$	$\Delta\mathbf{Z}_2$	decrease in hazards	

Finally, we compare the effects of changes in the macroeconomic conditions holding the training model and the application variables constant. This can be done for the pre-crisis model ($\boldsymbol{\beta}_1$) by comparing test sets C and A and B with D and when using the post-crisis model ($\boldsymbol{\beta}_2$) by comparing H with F and E with G (see Panel C of Table 5). Holding the application variables at their pre-crisis values and comparing C with A, we see that moving from the pre- to the post-crisis test sets, the median and the modal values both decreased. Holding the application values at their post-crisis levels, comparing B with D, we see that moving from the pre- to the post-crisis conditions led to an increase in median and modal hazards. Now we make similar comparisons for the post-crisis model ($\boldsymbol{\beta}_2$). We see that moving from the pre-crisis conditions to the crisis conditions decreases the median and modal hazards whether we condition on pre-crisis characteristics (test sets H and F) or post-crisis characteristics (test sets E and G). In conclusion, the model based on non-downturn data

($\boldsymbol{\beta}_1$) gives conflicting trends in predicted hazards depending on the cohort tested; whilst the model based on downturn data ($\boldsymbol{\beta}_2$) estimates a large and obvious drop in predicted hazards when post-crisis macroeconomic conditions are used. Both observations imply that the parameter estimates of the two survival models are not able to adequately take into account macroeconomic effects such that when these change (in the test sets) significantly from what were observed (in the training set), they give very different hazard estimates, and this could be further complicated by changes in the cohort.

5. Concluding remarks

This work investigates the stability of parameter estimates of discrete survival models developed on a large portfolio of credit card loans provided by a major UK bank, consisting of accounts that were accepted between 2002 and 2010, and observed up to early 2011. By developing two survival models, one based on data from before the crisis and the other based on data from since the crisis started, we use the chow test, a statistical test to test for differences between two sets of parameter estimates, and show that there are statistically significant differences between the two sets of estimated parameters, leading to different distributions of predicted probabilities of default. We also apply the estimated parameters onto a common test set to show how each set of parameters would give different predictions for probabilities of the default, and find that the models underestimate and overestimate default rates at different duration times of the loan.

We then investigated the three possible sources leading to the change in the distributions of hazards before and after the crisis: the difference in the quality of the cohort accepted under different economic conditions, the drastically different economic conditions that were seen in the UK economy, or the different estimated parameters. This was done by selecting two cohorts, one representing a cohort of loans accepted during a non-downturn period (i.e. loans that were accepted during 2002 to 2004, observed up to 2005 under 2002 to 2005 macroeconomic conditions), and the other representing loans that were accepted during the downturn period (i.e. loans that were accepted during 2008 to 2010, observed up to 2011, under 2008 to 2011 macroeconomic conditions), and estimating two survival models for each period separately and independently. Based on these two selected cohorts, we then created four related test sets holding constant either cohort quality or the economic conditions, and by applying either set of estimated parameters, we get 8 different scored test sets, which we then compare pair-wise.

We find that changes in cohort, macroeconomic conditions and the estimated parameters all contribute towards the change in the distributions of predicted probabilities of default. Our results show that the model developed on downturn data consistently gives higher predicted probabilities of default, and this held across different macroeconomic conditions and cohorts, perhaps suggesting that using non-downturn data in model development would lead to a falsely optimistic view and underestimate probabilities of default. Depending on whether the model was developed on non-downturn or downturn data, differences in cohort might or might not be picked up by the estimated parameters – our results show that whilst parameter estimates from non-downturn data suggest lower hazards for the post-crisis cohort in line with expectations, the parameter estimates from downturn data was not able to pick it up. However, it is the effects of changes in macroeconomic conditions that are most difficult to unravel. Moving from pre- to post-crisis macroeconomic conditions, the model developed on non-downturn data predicts that hazards would increase for the post-crisis cohort yet decrease for the pre-crisis cohort, suggesting that different macroeconomic conditions affect different people at different times differently. Yet, doing the same comparison using a model developed on downturn data, we find a large decrease in estimated hazards. All these imply that even though we have taken into account macroeconomic conditions and possible interactions between macroeconomic and application variables, the models are still unable to adequately model the various effects coming from type of borrower, the time during which borrowing takes place, and how macroeconomic conditions would affect different individuals differently.

There is much further work to be done. Kelly et al. (1999) theorized that a model which is able to take into account all known and unknown predictor variables would be able to adapt to changes in the underlying population but that this is not always possible to achieve. While we have considered most major economic indicators (for which data was available), these variables were still unable to adequately represent all of the required predictor variables, hence the significantly different parameter estimates after the credit crisis. Further variables that can be considered include random variables to account for unknown heterogeneity, perhaps that are either or both of individual-specific and time-specific. More work is also required in the exploration and quantification of the effects of macroeconomic variation and cohort quality on probabilities of default, as well as the other components of risk in the calculation of loss.

References

BANASIK, J., CROOK, J. N. & THOMAS, L. C. 1999. Not if but When will Borrowers Default. *The Journal of the Operational Research Society,* 50**,** 1185-1190.

BASEL COMMITTEE ON BANKING SUPERVISION 2004. International Convergence of Capital Measurement and Capital Standards: A Revised Framework.

BELLOTTI, T. & CROOK, J. 2013a. Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting,* 29**,** 563-574.

BELLOTTI, T. & CROOK, J. N. 2009. Credit Scoring with Macroeconomic Variables Using Survival Analysis. *Journal of the Operational Research Society,* 60**,** 1699-1707.

BELLOTTI, T. & CROOK, J. N. 2013b. Retail Credit Stress Testing using a Discrete Hazard Model with Macroeconomic Factors. *Journal of the Operational Research Society*.

CHOW, G. C. 1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica,* 28**,** 591-605.

DEMARIS, A. 2004. *Regression with Social Data: Modelling Continuous and Limited Response Variables*, John Wiley & Sons, Inc.

KELLY, M. G., HAND, D. J. & ADAMS, N. M. 1999. The Impact of Changing Populations on Classifier Performance. *Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San Diego, California, United States.

LEOW, M. & CROOK, J. N. 2014. Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies. *European Journal of Operational Research,* In Press.

LEOW, M., MUES, C. & THOMAS, L. C. 2011. Competing Risks Survival Model for Mortgage Loans with Simulated Loss Distributions. *Credit Scoring and Credit Control XII.* Edinburgh.

MA, P., CROOK, J. N. & ANSELL, J. 2010. Modelling Take-Up and Profitability. *Journal of the Operational Research Society,* 61**,** 430-442.

MCDONALD, R. A., MATUSZYK, A. & THOMAS, L. C. 2010. Application of Survival Analysis to Cash Flow Modelling for Mortgage Products. *OR Insight,* 23**,** 1-14.

PENNINGTON-CROSS, A. 2010. The Duration of Foreclosures in the Subprime Mortgage Market: A Competing Risks Model with Mixing. *The Journal of Real Estate Finance and Economics,* 40**,** 109-129.

RODRIGUEZ, A. & TRUCHARTE, C. 2007. Loss coverage and stress testing mortgage portfolios: A non-parametric approach. *Journal of Financial Stability,* 3**,** 342-367.

STEPANOVA, M. & THOMAS, L. C. 2002. Survival Analysis Methods for Personal Loan Data. *Operations Research,* 50**,** 277-289.

Appendix

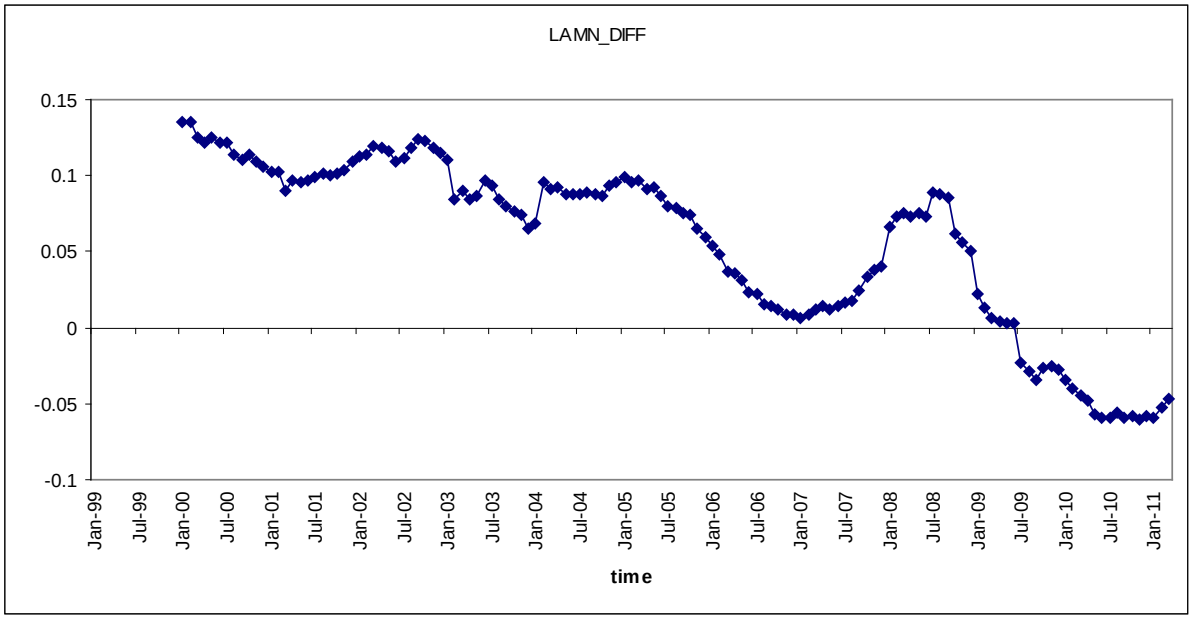

Figure A1: Total consumer credit amount outstanding, ln, non-seasonally adjusted, 12th differenced.