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

This version: July 2013

Mindy Leow & Jonathan Crook
Credit Research Centre
University of Edinburgh Business School

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 after the start of the credit crisis of 2008. Two survival models are developed for accounts that were accepted before and after the crisis, and 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 parameter estimates 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. By estimating survival models for accounts that were accepted during the downturn and non-downturn periods, and by creating test sets which either keep the macroeconomic conditions or the cohort quality unchanged, we are able to test for each effect. The results are displayed in terms of distributions of predicted probabilities, which show that both macroeconomic conditions and cohort quality do affect probabilities of default, with economic conditions having a larger effect.

Keywords: discrete time survival model, macroeconomic variables, credit crisis, time-varying covariates, probability of default

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, and 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 for 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) and Leow et al. (2011). 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 retail 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. As expected, 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 is different to the existing literature.

Firstly, 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 (graphs appended in the Appendix), and it would be interesting to explore how these large and unexpected changes would affect default models and their predictions.

Secondly, we investigate the stability of survival model parameter estimates over time, especially over 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 after the credit crisis started, 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 given a credit account 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 parameter estimates 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.

2. Data

The data is supplied by a major UK bank and is a random sample of credit cards that were given out 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. This means that account observation times range from a minimum of 12 months, up to 110 months. 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 does the account misses a payment over its entire history. Any behavioural variables are included in the model are lagged by three months.

Although default information is available from the dataset, it is not consistent 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[1] 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 did 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 (reference here). As the work here only focuses on the default event, we do not specify the transitions between the preceding months in arrears; further details can be found in Leow and Crook (2012).

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 it was opened are included.

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

First, the dataset is split into two training sets. 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; and the second consisting of accounts that started between January 2008 and February 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 after the start of the credit crisis.

[1] Note that this percentage is different to that used in LEOW, M. & CROOK, J. N. 2012. Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies.

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 observation period.

2.2. Macroeconomic variables

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

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. 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.

3. Methodology

Due to the way the data is gathered, i.e. as regular, discrete monthly points in time, and how the default event is recorded, i.e. the default happens in a particular month with reference to the month the account was open, the survival models are estimated in discrete time. Another advantage is a much lower computational time in model estimation as we deal with a large dataset.

$P_{i\tau}$ is 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.

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

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

In this regression model, the dependence of the hazard on time, α_t, is specified as $\tau + \tau^2 + \ln\tau + (\ln\tau)^2$. By doing so, we do not put any restrictions on the relationship between effect on time and the odds of default 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 12^{th} differences, lagged or otherwise, of each macroeconomic variable were examined.

4. Results

4.1. Parameter estimates

The parameter estimates from the training sets I and II representing accounts that were accepted before the crisis and after crisis started respectively, are given in Table 3. Single (*) and double asterisks (**) are used to represent covariates that are statistically insignificant at the 0.01 and 0.05 level respectively. Due to confidentiality agreements, we are unable to detail all variables used in the model.

The application variables are fairly stable, with most of the younger borrowers not significantly different from each other, which is also why we take the age groups representing the older borrowers in the interaction terms. Variable X is an interesting categorical variable which has its ranking changed before and during the crisis period. Employment status of the borrower does not seem to have much effect on default probability. The behavioural variables are very stable, with very similar estimates over the two models. On the other hand, 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 bust two to three years earlier than that of the credit crisis (c.f. Figure A8 in the Appendix), and thus would not be reflected in 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 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.

Table 3: Parameter estimates for PD model for accounts accepted pre-crisis and after the crisis started. The single asterisk (*) and double asterisk (**) represent variables that are statistically insignificant on the 0.01 and 0.05 levels.

Code	Variable	PRE-CRISIS Estimate			CRISIS Estimate		
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		

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.

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

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

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

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.

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

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 from our analysis[2] show that the test statistic has a value greater than the critical value, so the null hypothesis is rejected, i.e. the parameter estimates of models developed on training sets I and II are statistically significantly different from each other.

Although the test has established that the parameter estimates from the two models are statistically 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.

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

[2] 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.

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

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. 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.

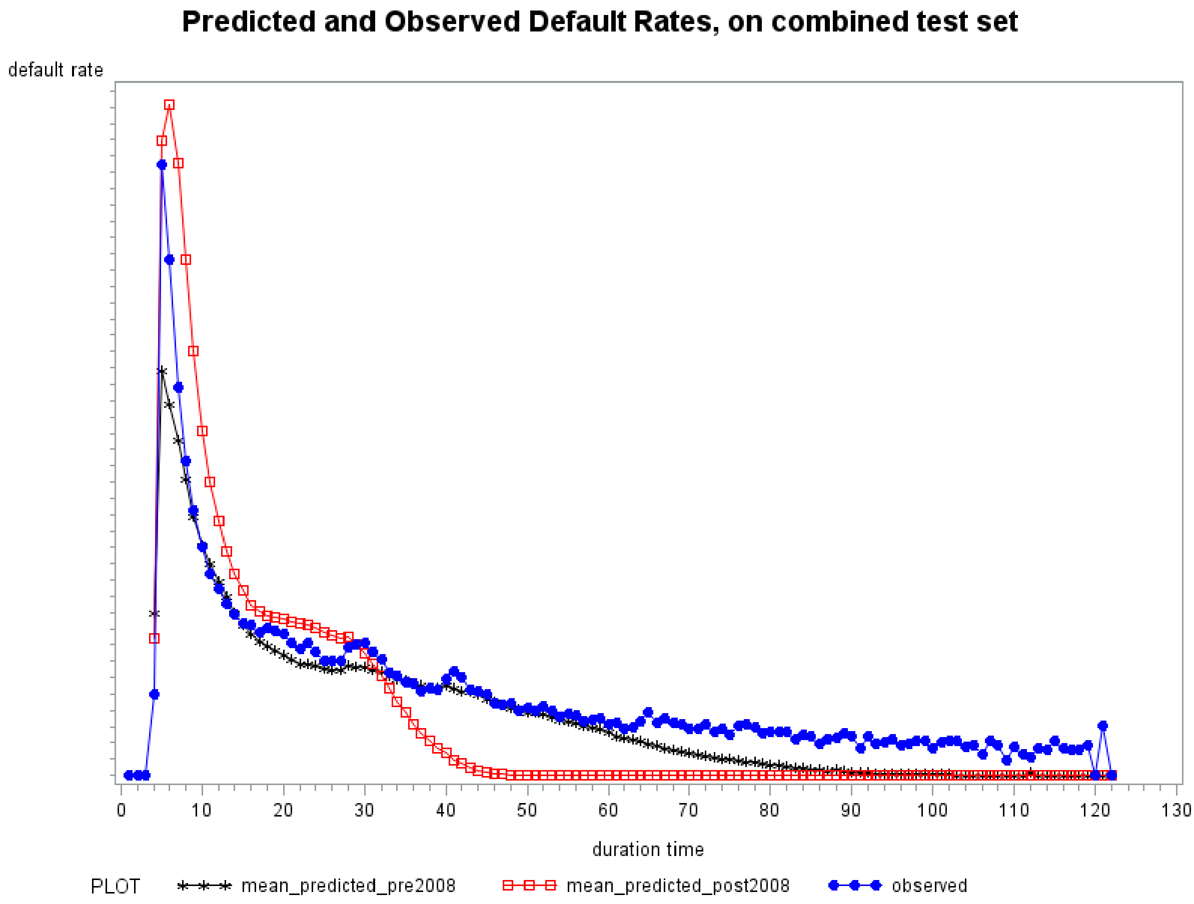

Figure 1: Predicted and observed default rates for combined test set. The solid dots represent the observed default rate; the asterisks represent the predicted default rate from the model developed on training set I, i.e. pre-2008; the squares represent the predicted default rate from model developed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

From Figure 1, we see that there is some difference in the trends of the likelihood of default based on the two models over the duration of the loan. We theorize that the differences in the predictions could be due to either, or a combination of, the quality of the accounts and the macroeconomic conditions they subsequently experienced. The two training sets represent accounts that were accepted at very different economic conditions (2002 to 2007 versus 2008 to 2010), which would imply that cohorts are of differing qualities, since we expect that the acceptance policies of banks would be different during non-downturn and downturn periods, as banks adapt policies to reflect the changing risk profiles over different periods of economic growth. For example, the likelihood of accepting a self-employed debtor would be different in good and poor economic times. Also, after the accounts were accepted, the debtors were then subject to various economic conditions, which again were very different during the periods 2002 to 2007 and 2008 to 2011 (cf. macroeconomic graphs in Appendix). These macroeconomic variables would have some effect on the hazard of default, however the extent of which is unclear, and is investigated in the following sub-section.

4.4 Cohort quality versus macroeconomic situation

In order to investigate if changes in the parameter estimates (cf. Figure 1) are due to the quality of the cohort accepted during that period or the different macroeconomic conditions, we select two cohorts representing accounts that were accepted at different time periods, under different economic conditions.

To represent a cohort that was accepted during normal economic conditions, we use accounts that were opened between January 2002 and February 2004, observed up to March 2005, referred to as the "normal cohort"; and to represent a cohort that was accepted during downturn economic conditions, we use accounts that were opened between January 2008 and February 2010, observed up to March 2011, referred to as the "downturn cohort". The two cohorts are observed for the same length of time since default risk is measured in terms of duration time, as well as aligned such that they are at the same point in time (month) in the year, so as to remove any seasonality effects that default risk might undergo. These cohorts and the macroeconomic conditions they each experience form the training sets, from which a discrete survival model would be estimated. The parameters estimated are then applied onto created test sets but where either the cohort quality (i.e. the accounts and their application variables) remain unchanged or the macroeconomic conditions (i.e. all macroeconomic variables) remain unchanged. Where applicable, interaction terms are updated such that they reflect the interaction effect between application variables and the changed macroeconomic variables. The training sets selected and the test sets created are summarised in Table 4.

Table 4: Training sets and created test sets

Acceptance policy	Training data	Test data 1 (Fixed cohort)	Test data 2 (Fixed macroeconomic conditions)
Normal	Cohort from 2002 to 2004; macroeconomic conditions from 2002 to 2005	Cohort from 2002 to 2004; macroeconomic conditions from 2008 to 2011	Cohort from 2008 to 2010; macroeconomic conditions from 2002 to 2005
Downturn	Cohort from 2008 to 2010; macroeconomic conditions from 2008 to 2011	Cohort from 2008 to 2010; macroeconomic conditions from 2002 to 2005	Cohort from 2002 to 2004; macroeconomic conditions from 2008 to 2011

For each training dataset and its 2 corresponding test sets, we plot the distributions of the predicted probabilities of default at duration time 12 months into the loan, segmented by observed defaults and non-defaults. These are given in Figures 2 and 3, from which we can see how, and by how much, the predicted distributions have shifted when we allow for changes to either cohort quality or macroeconomic conditions. We will note that the differences in the distributions for defaults and non-defaults within each dataset are only slight, which could be a due to a number of reasons. For clearer indication of whether the predicted probabilities are affected by cohort quality or macroeconomic changes, we have excluded all behavioural variables from these models because these variables will be affected by the economy but we would be unable to create the corresponding behavioural variables with changes to macroeconomic variables. Also, the probabilities and their distributions displayed here do not imply any measurement of predictive performance; what we have done here is to compare the distributions of predicted probabilities at a singular and particular time point when the output of the survival model would actually produce a probability of default for each time point, in order to fulfil the purpose of our investigation. Finally, the values of the probabilities would seem to be abnormally small, but this is due to it being the probability of an event (i.e. default) occurring at a particular time point over a horizon of time. The distributions from the two figures are not directly comparable as the underlying survival models for each figure are developed based on different training sets.

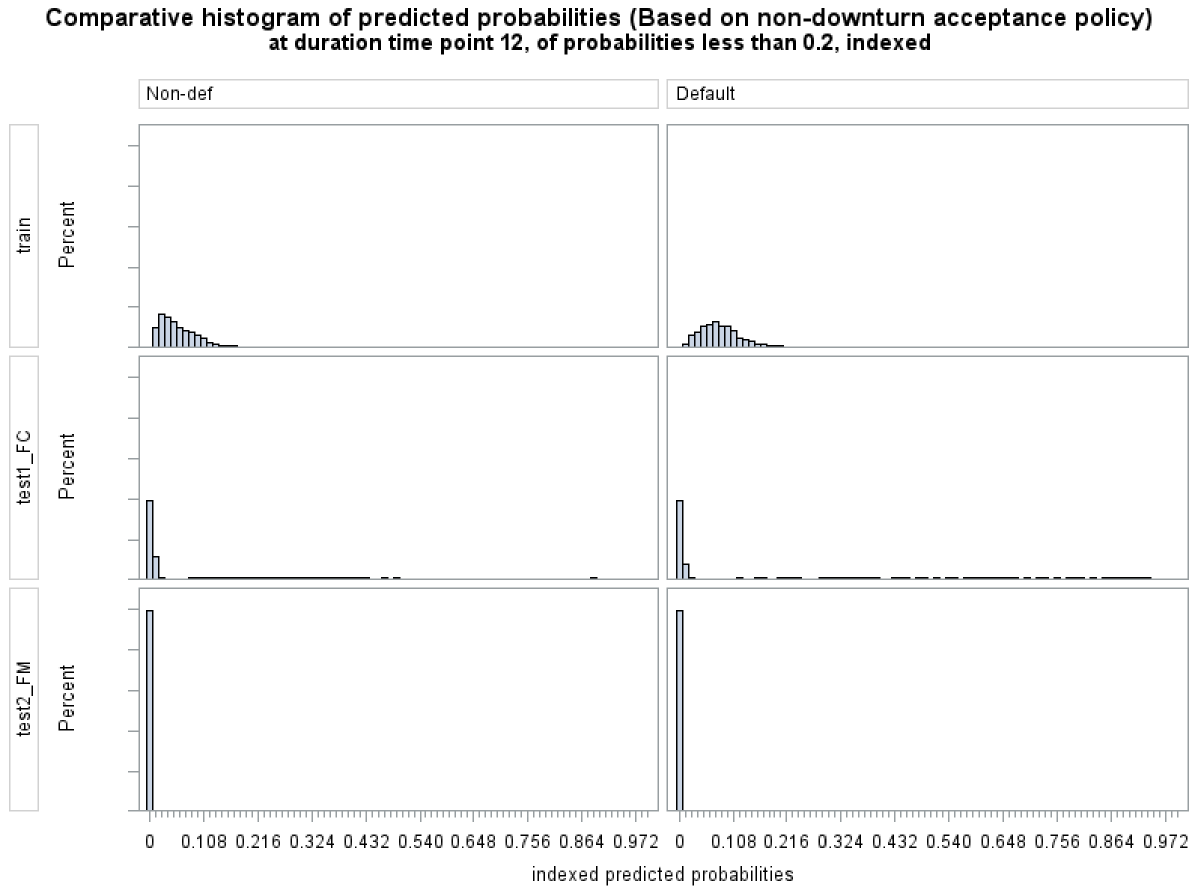

Figure 2: Comparative histogram of predicted probabilities, during a normal acceptance policy, for predicted probabilities less than 0.2, indexed, at duration time 12. The panel on the left (right) shows the distributions of predicted probabilities for observed non-defaults (defaults); the top row shows the distributions of predicted probabilities based on the training set, i.e. cohort from 2002-2004, macroeconomic conditions from 2002-2005; the second row shows the distributions of predicted probabilities for the created test set where cohort is fixed, i.e. cohort from 2002-2004, macroeconomic conditions from 2008-2011; the third row shows the distributions from predicted probabilities for the created test set where macroeconomic conditions are fixed, i.e. cohort from 2008-2011, macroeconomic conditions from 2002-2005.

Figure 2 consists of a comparative histogram of predicted probabilities where the model was estimated based on a set of accounts that were accepted during a non-downturn period. Each row gives the distribution of the predicted probability for observed defaults (on the left panel) and non-defaults (on the right panel). The first row gives the distributions of the predicted probabilities, scored on the training set itself, in this case, accounts accepted between 2002 and 2004 and observed up to March 2005, with macroeconomic conditions from 2005 to 2005. We note that there is a large difference in the distributions of default probabilities although the graphs only take into account probabilities less than 0.2. This would be due to the large differences between probabilities of the training set, from which the underlying survival model is based, and the two test sets.

The second row gives the distributions of predicted probabilities, scored on the test set where the cohort was kept unchanged, i.e. the test set consists of the same accounts in the training set (accepted between 2002 and 2004) but where the macroeconomic conditions at each duration point of the loan is now said to be that experienced during the downturn (from 2008 to March 2011). We see an interesting shift is the distribution. Given the same cohort quality, the poorer macroeconomic conditions that have been imposed here brings about much higher predicted probabilities for a portion of the accounts (for both defaults and non-defaults) as observed by the obvious long right tail, compared against the distributions of the top row. As expected, poorer economic conditions lead to higher probabilities of default. However, another portion of accounts are now predicted to have much lower probabilities of default. It would seem that the poor economic conditions have affected different debtors differently. The third row gives the distributions of the predicted probabilities, scored on the test set where the macroeconomic conditions were kept unchanged, i.e. the test set consists of accounts that were accepted during the downturn period (between 2008 and 2010) and said to experience the macroeconomic conditions of the non-downturn period (from 2002 to 2005). We see an obvious shift in the probabilities to the left, indicating that under the same macroeconomic conditions, accounts that were accepted during the downturn period would have lower default probabilities than those accepted during the non-downturn period. This is an intuitive result because we expect accounts that were accepted during the downturn period to be of a higher quality than those accepted during the non-downturn period due to more stringent acceptance policies.

In Figure 3, we present the results of a similar analysis, but where the survival model is estimated on the cohort accepted during the downturn period (between 2008 and 2010), and which then went on to experience very poor economic conditions (between 2008 and 2011). Similarly, the top row displays the distributions of predicted probabilities, segmented by defaults (on the right panel) and non-defaults (on the left panel), for the training set.

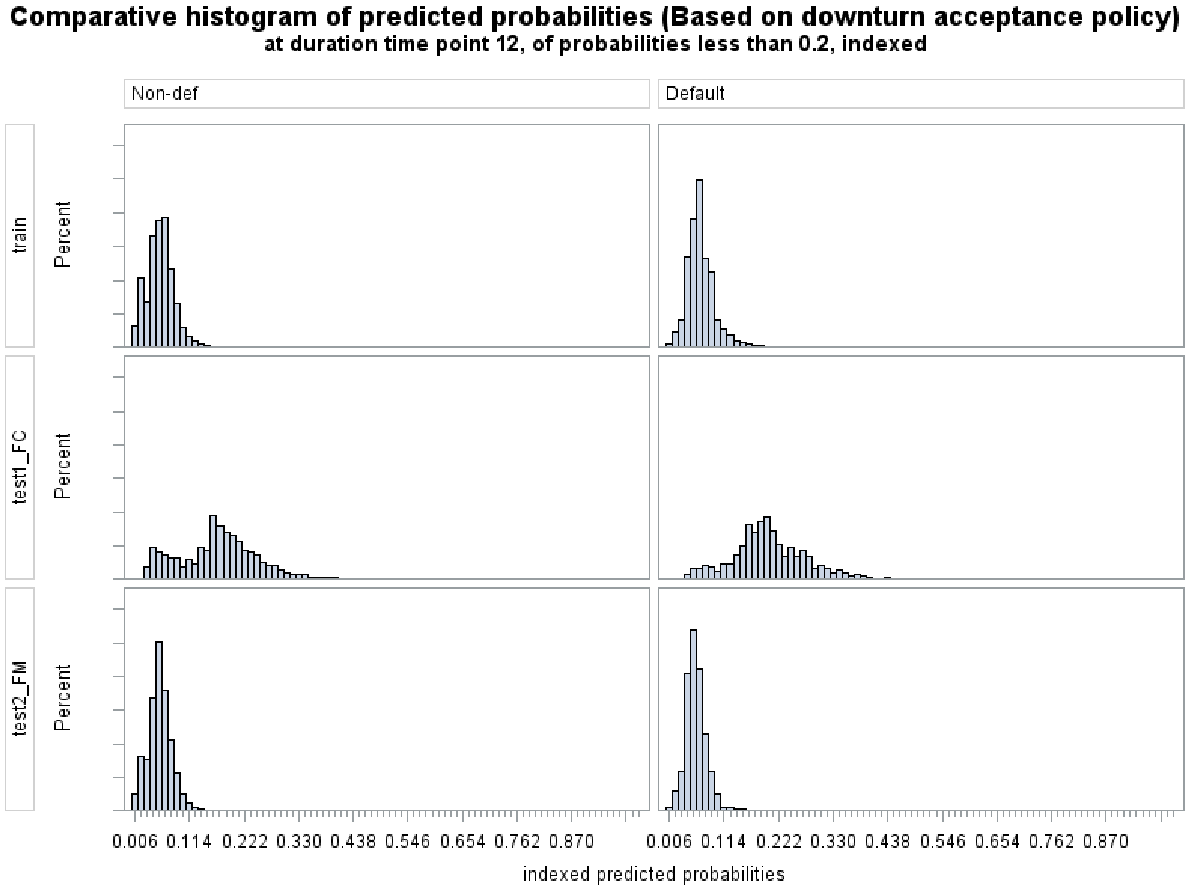

Figure 3: Comparative histogram of predicted probabilities, during a downturn acceptance policy, for predicted probabilities less than 0.2, indexed, at duration time 12. The panel on the left (right) shows the distributions of predicted probabilities for observed non-defaults (defaults); the top row shows the distributions of predicted probabilities based on the training set, i.e. cohort from 2008-2011, macroeconomic conditions from 2008-2011; the second row shows the distributions of predicted probabilities for the created test set where cohort is fixed, i.e. cohort from 2008-2011, macroeconomic conditions from 2002-2005; the third row shows the distributions from predicted probabilities for the created test set where macroeconomic conditions are fixed, i.e. cohort from 2002-2004, macroeconomic conditions from 2008-2011.

The second row gives the distributions of predicted probabilities for the test set where the cohort quality is kept unchanged, i.e. where the cohort is that accepted during the downturn period (2008 to 2010) but where the macroeconomic conditions are better (from 2002 to 2005). With the improvement in macroeconomic conditions, we expect to see the likelihood of default to decrease compared to that observed in the training set, instead, we see a shift to the right in the distributions. This could imply that accounts that were accepted during the downturn period are higher quality accounts such that their probabilities of default are not expected to improve by much in good economic times. The third row gives the distributions of the predicted probabilities for the test set where the macroeconomic conditions are kept unchanged, i.e. where the cohort accepted is that during the non-downturn period (2002 to 2004) and where the macroeconomic conditions remain as that seen during the downturn (2008 to 2011). Here, we see that the distribution has shifted to the left slightly, which again is surprising as we expect accounts accepted during the downturn period to be better than those accepted during the non-downturn period, leading to higher probabilities of default. Since any changes in probabilities in this panel would be purely due to application variables (and interaction terms between application and macroeconomic variables), this would imply that there has been a shift in the population; and that the effects of characteristics during the downturn and non-downturn periods are different.

From this analysis and their results, we see that the both the quality of the cohort that was accepted at the time of application and the subsequent macroeconomic conditions they experience would affect the probability of default, with effects subtle and not easily quantified. It should be noted that we were only able to analyse borrowers that have applied for loans in each time period, whereas there might be applicants who refrained from applying for, or accepting credit card accounts, at different economic conditions due the effects on the terms of the loan. Overall, we see a larger effect on distributions of probabilities when macroeconomic conditions are changed.

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 after the crisis started, we use the chow test, a statistical test to test for differences between two sets of parameter estimates, to show that there are significant differences between the two sets of estimated parameters. We also apply the estimated parameters onto a test set to show how each set of parameters would give different predictions for probabilities of the default. Based on our results, we find that the parameter estimates do change over time, especially since the start of the credit crisis of 2008.

We then investigated whether these changes in the parameter estimates are 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. 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). Survival models were estimated for each period separately and independently. Based on these two selected cohorts, we then created four test sets holding constant either cohort quality or the economic conditions. Our results show that macroeconomic conditions do affect probabilities of default, and could affect different groups of debtors in different ways. Overall though, we see probabilities of default increase when macroeconomic conditions worsen. On the other hand, probabilities are also subtly affected by the quality of the cohort, but this is less straightforward to define.

There is much further work to be done in this area. 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.

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

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. 2012. Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies.

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

Graphs of macroeconomic variables (12th differenced)

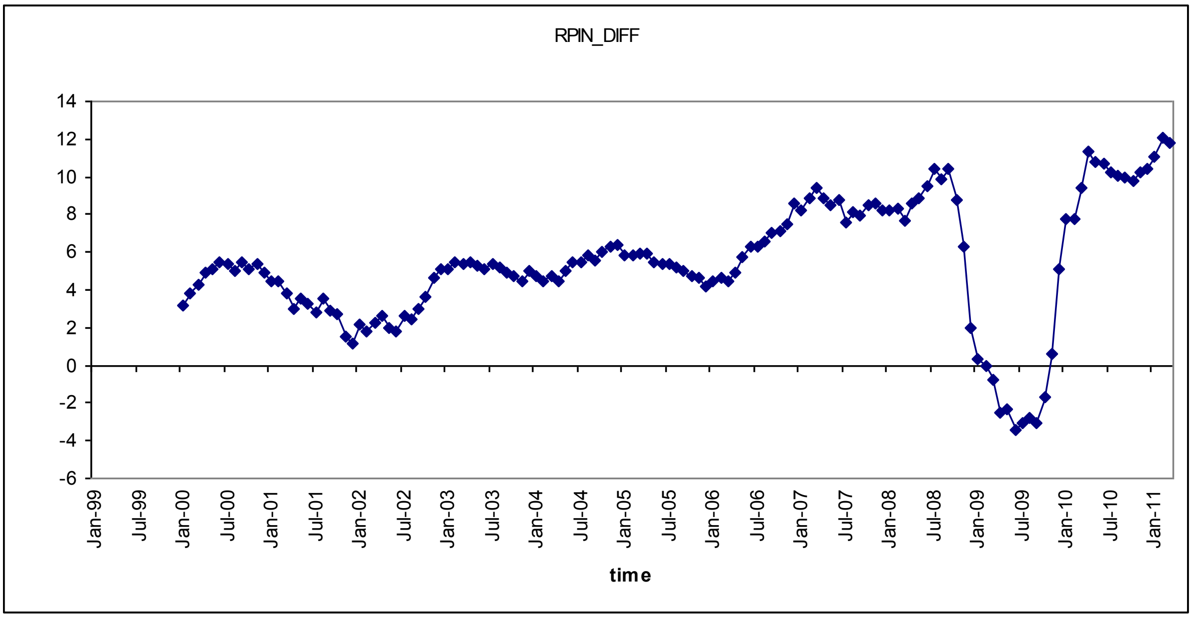

Figure A1: Retail price index, non-seasonally adjusted, 12th differenced.

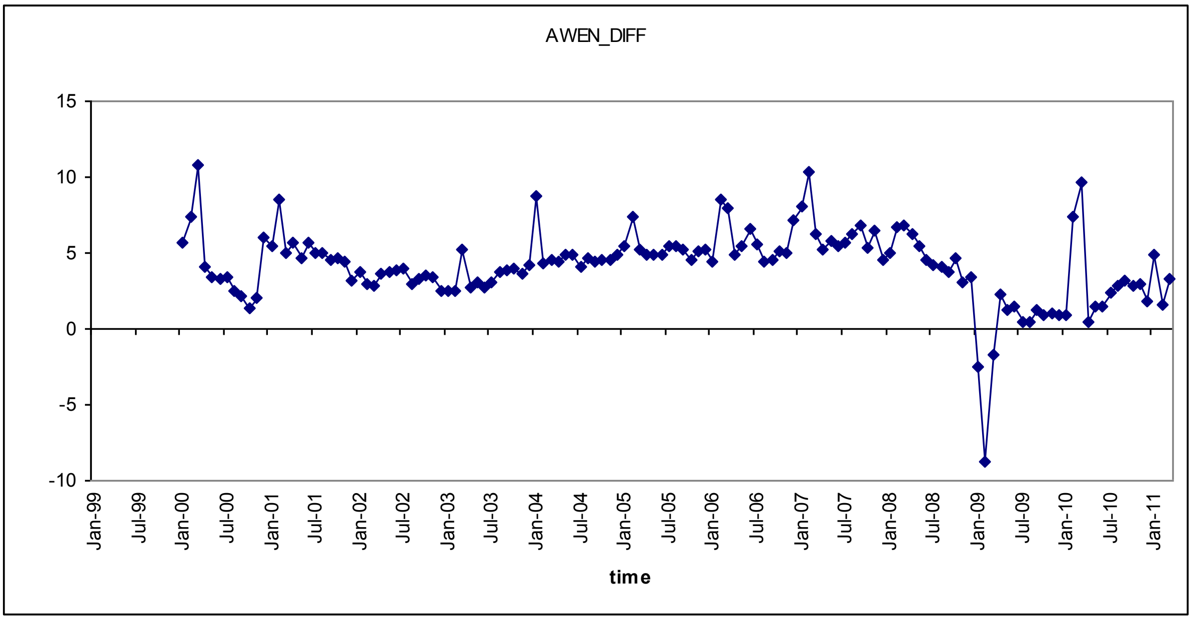

Figure A2: Average wage earnings, non-seasonally adjusted, 12th differenced.

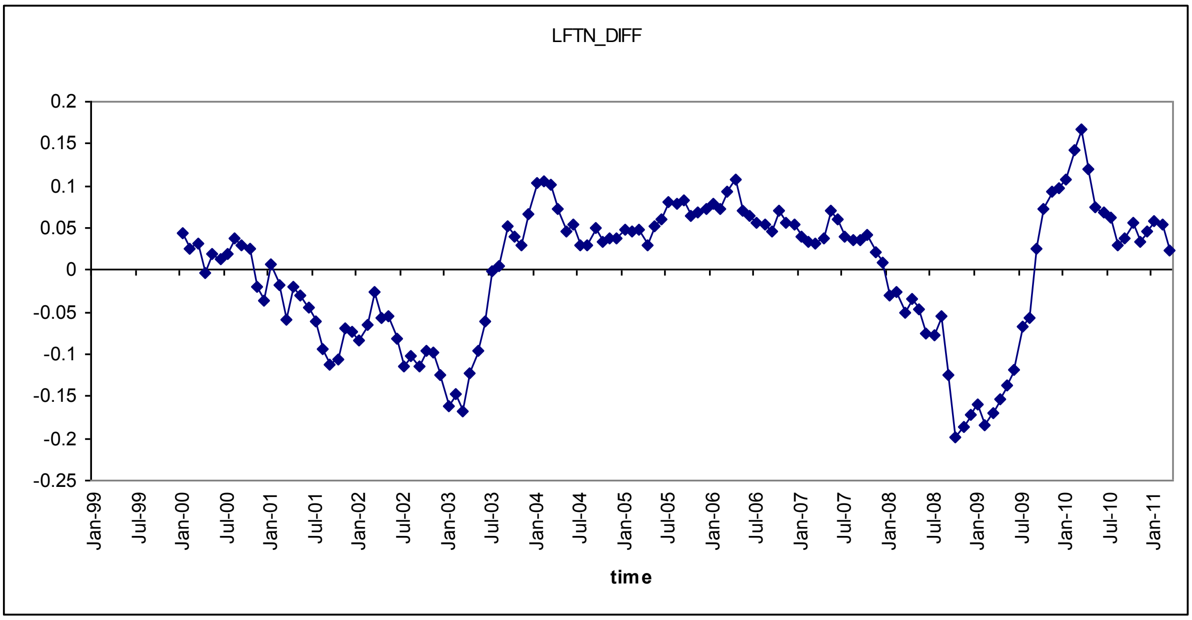

Figure A3: FTSE Index, ln, non-seasonally adjusted, 12th differenced.

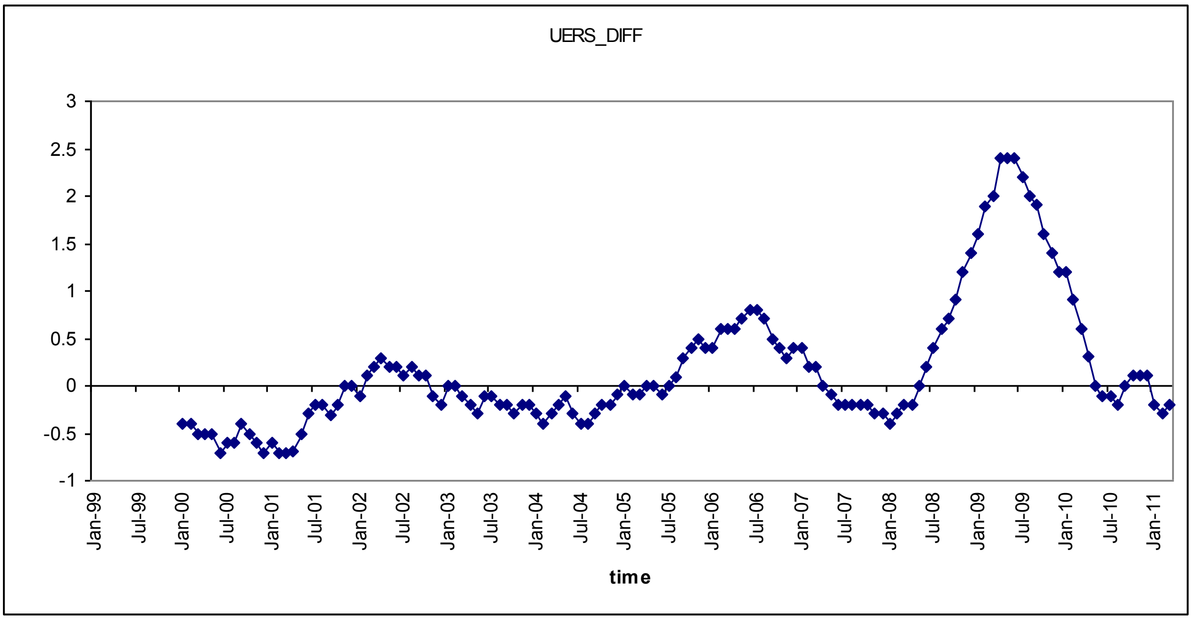

Figure A4: Unemployment rate, seasonally adjusted, 12th differenced.

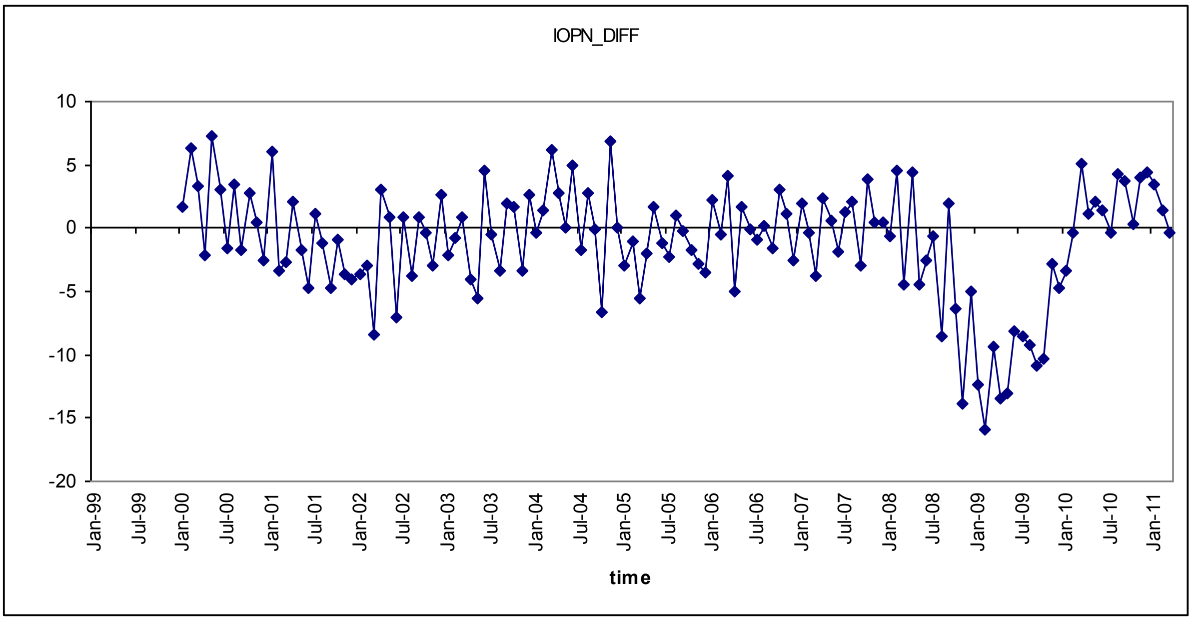

Figure A5: Index of production, non-seasonally adjusted, 12th differenced.

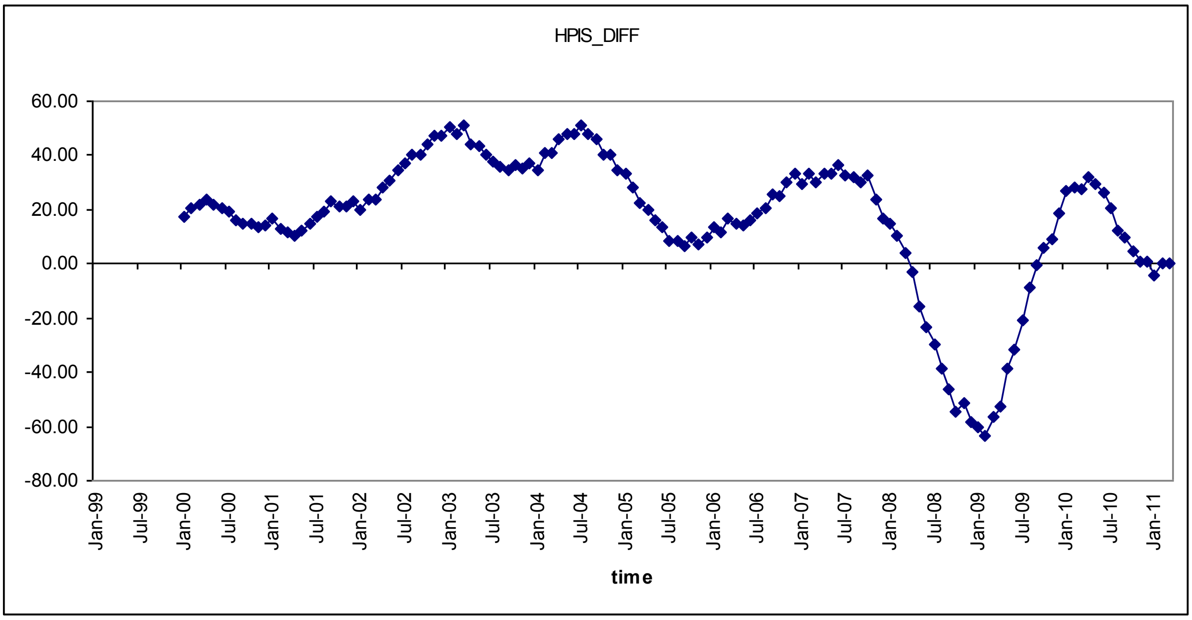

Figure A6: House price index, seasonally adjusted, 12th differenced.

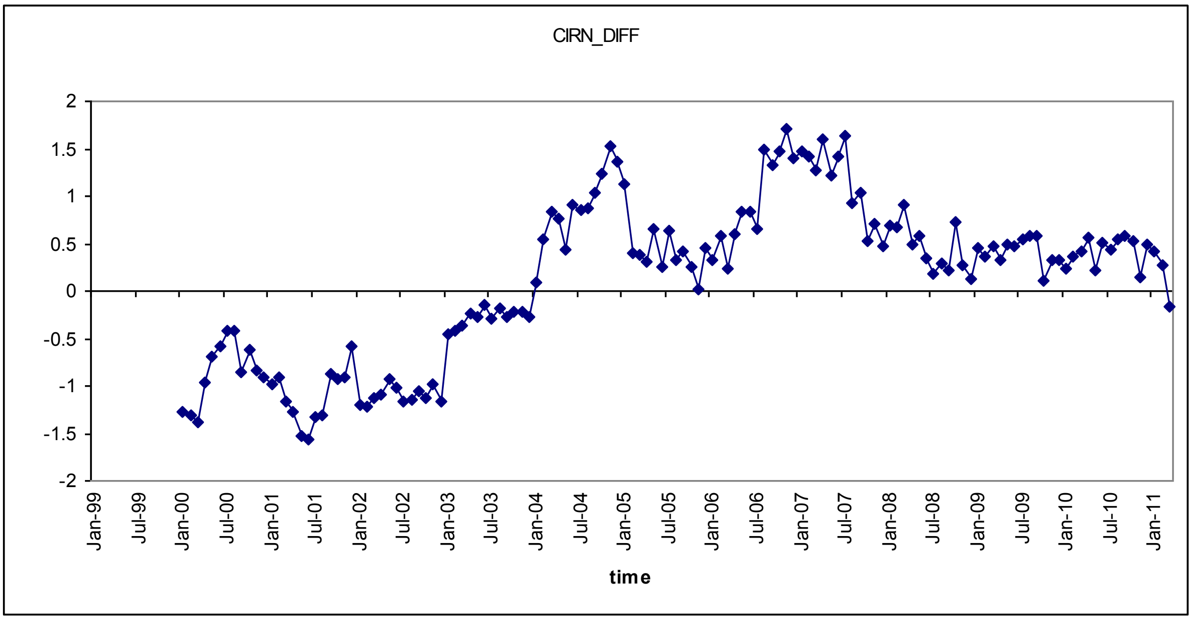

Figure A7: Credit card interest rate, non-seasonally adjusted, 12th differenced.

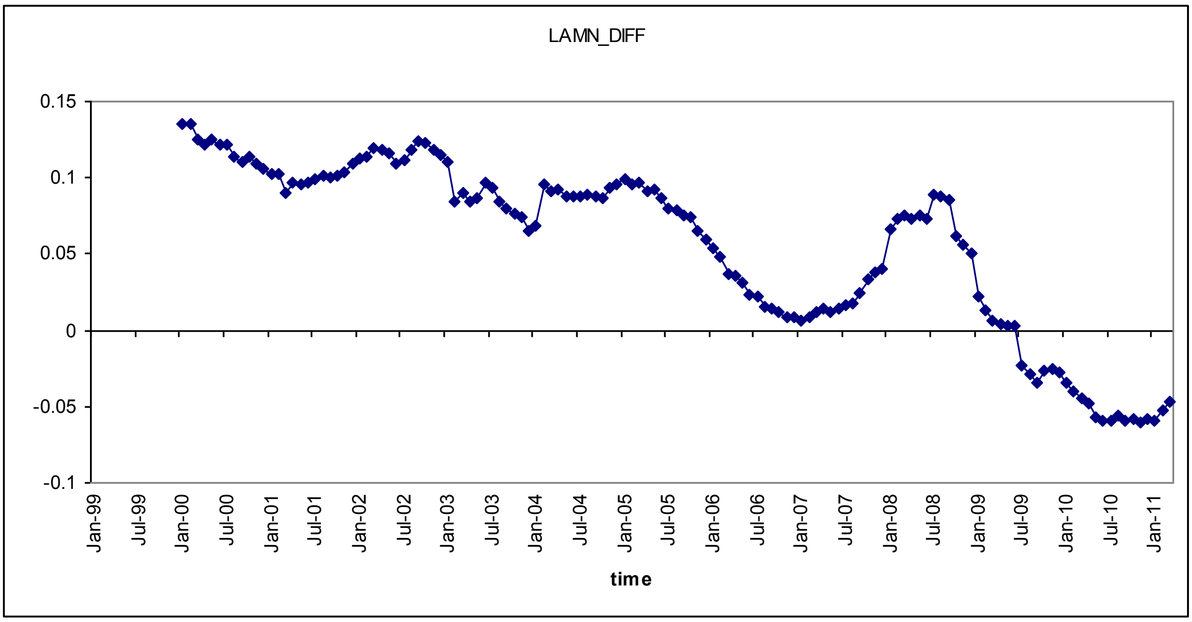

Figure A8: Total consumer credit amount outstanding, ln, non-seasonally adjusted, 12th differenced.

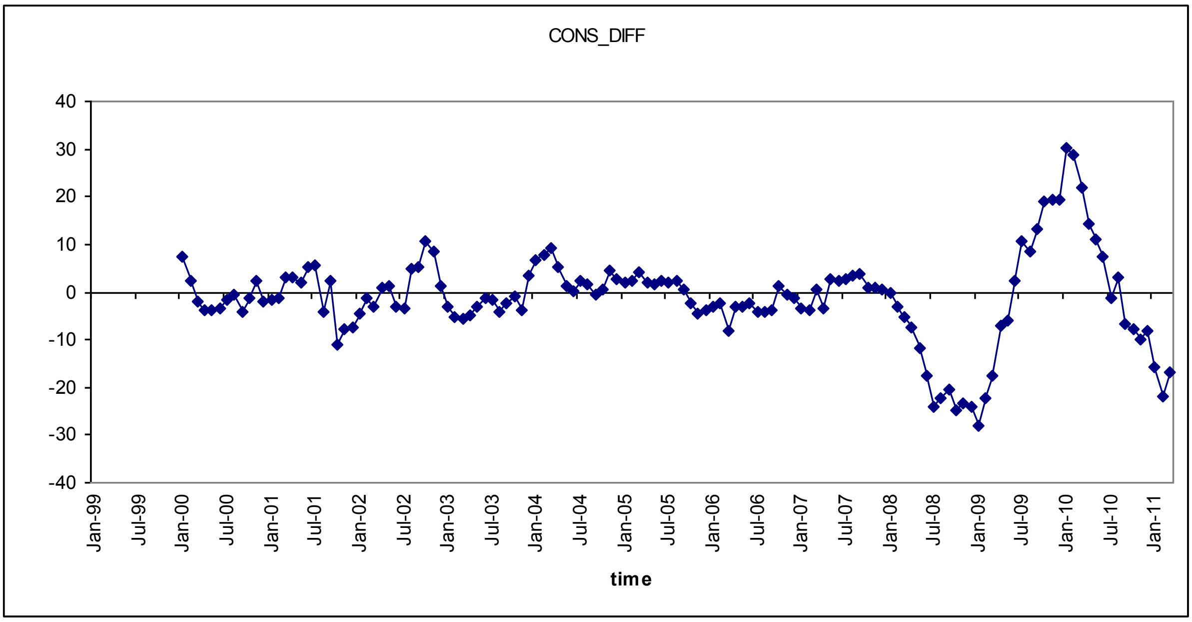

Figure A9: Consumer confidence, seasonally adjusted, 12th differenced.