Exploring the Effects of Macroeconomic Variables on Credit Card Delinquency and Default Behaviour

(Preliminary Version, work in progress)

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

---

[a] Credit Research Centre, University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, Scotland, United Kingdom

[b] mindy.leow@ed.ac.uk

[c] j.crook@ed.ac.uk

Abstract

Using a large dataset comprising of credit card loans from 2002 to 2010, and incorporating a mix of application variables, behavioural variables and macroeconomic variables, we use intensity models to estimate probabilities of delinquency and default. We estimate these predicted probabilities for each account for each type of transition over the duration time of the loan, and find interesting insights and differences between groups of accounts, as well as over time. By segmenting the dataset by selected application variables, we find that the predicted probabilities of delinquency and default for different groups of accounts have different trends. Next, we employ two methods of simulation. The first aims to translate the predicted probabilities into predicted events at each time point, whilst the second generates transition rates for the test set based on empirical transition rates from the training set in order to compute the VaR for each transition distribution.

Keywords: intensity models, macroeconomic variables, simulation, loss distributions, delinquency probabilities, default probabilities

1. Introduction

The prediction of transition probabilities between repayment states for a loan has significant implications for the amounts of both regulatory and economic capital a bank has to hold and for loan pricing. A number of papers have tried to model transition probabilities for a loan portfolio. Ho et al (2004) test for the order of a Markov Chain for bank overdrafts and personal to find they are not first, second or third order. Malik and Thomas (2012) use an ordered logistic model to model the probability of an account moving from one behavioural score to another in the next month where a given order of the transition matrix has to be assumed to make predictions and the same model used for each initial state regardless of subsequent state. No application or behavioural variables are included. Grimshaw and Alexander (2011) also propose a method to predict the probability that an account will move between states between two adjacent time periods but do not show any empirical results. However these models do not show the marginal effects of particular variables on specific transition probabilities and in the case of Malik and Thomas assume time homogeneity. To address these issues intensity or doubly stochastic models have been proposed. These models have been parameterised by Lando and Skodeberg (2002), Duffie (2007), Kadam and Lenk (2008) and Stefanescu et al (2009) for corporates loans. For retail loans the only published paper is by Leow and Crook (2014) who estimated models for data between 2001 and 2006 using a large sample of credit card accounts. However this paper did not include data that covers the recent financial crisis and omitted the influence of macroeconomic variables on transition probabilities. The aim of this paper is to extend this work in a number of ways. First, we incorporate the effects of the macro-economy by including macroeconomic variables. These macroeconomic variables are matched to accounts and updated at every duration time point of the loan, making them time-dependent. Second, having access to a more recent dataset encompassing the credit crisis of 2008, we assess and validate the intensity model over a period of time which consists of both downturn and non-downturn periods. Third, we combine simulation and intensity modelling to gain a distribution of predicted transition probabilities between different repayment states, creating a plausible framework for stress testing. Fourth, using these account-level intensity models, we estimate the portfolio Value at Risk (VaR) via a combination of intensity modelling and a simulation.

We find that the probabilities of an account transiting from being up to date to one payment behind decline the longer the borrower has held a card until about 17 months and then increases slightly, that the probability that an account, when delinquent, will become more delinquent or default declines the longer the account is in that state, and that the probability of cure (moving from a delinquency state to a lesser state) declines the longer an account is in that state. We further find that the change in the relative transition probabilities over time for different segments of a portfolio. We find significant relationships between various macroeconomic variables and transition probabilities between different states, for example that when share prices are higher six months later transition probabilities into delinquency and default are lower and transition probabilities from delinquency into less delinquent states (cure) are higher. We find that the distribution of transition probabilities has a VaR nearly seven times the mean for transitions from up to date to one payment behind but a lower ratio to the mean for more delinquent transitions.

This work contributes to the literature in a few ways. First, we believe we are the first to show the relationship between transition probabilities for retail loans as an account descends towards default and as an account cures and various macroeconomic variables. Second we demonstrate our findings using data that covers a period of relative stability and financial difficulty and so relatively robust to macroeconomic circumstances. Third we have used a novel way to simulate transition events over time. Fourth we combine the intensity models with simulation to show the parameters of the distribution of transition probabilities for different transitions and over different duration times and so add to the methodology of micro-stress testing at portfolio level.

The paper is structured as follows. We describe

2. Data

The data is from a portfolio of credit card accounts provided by a major UK bank. It consists of variables collected at the time of application, e.g. length of time at address, income and employment code; and behavioural variables collected at monthly time points, e.g. credit limit, repayment amount. Further behavioural indicators are computed, e.g. rate of transitions or proportion of credit drawn. All behavioural variables, as are the macroeconomic variables, used here are lagged 6 months[1].

These accounts were accepted on books between 2002 and 2010, and observed monthly up to the first quarter of 2011, thus encompassing the period of the credit crisis which started in 2008. The length of the dataset also means that we are able to have a test set that is entirely out-of-sample and out-of-(calendar)-time. Accounts that were opened before (and up to) June 2008 and observed up to (or censored at) December 2008 belong to the training set, containing about 385,000 unique accounts. Accounts that were opened from January 2009 and observed up to March 2011 make up the independent test set with about 89,000 unique accounts.

3. Methodology

3.1. Definition of state and default

Following the work in Leow and Crook (2014), we define 4 states: up-to-date (state 0), one month in arrears (state 1), two months in arrears (state 2) and default (state 3), where movements between the states depend on whether the borrower makes the minimum repayment for that month. The rules for transition between states remain are as in Leow and Crook. These are as follows. All accounts start in state 0 that is up to date with repayments. If at any time during the observation period the repayment amount made is less than the minimum required the borrower advances to the next immediate state. A borrower who has missed a repayment before and is in states 1 or 2 but makes a repayment of some amount in the following month(s) will (a) remain in that state if the repayment made is greater than the minimum required but less than the sum of the amounts required in the current and previous month or (b) be moved to a one lower state if the repayment made exceeds the sum of the minimum required in the current and previous months but is less than the outstanding

[1] The lag used in LEOW, M. & CROOK, J. N. 2014. Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies. *European Journal of Operational Research,* 236**,** 685-694. was 3 months. As part of robustness testing, we estimated the model with lags of 3 months and found little difference in terms of parameter estimates, so decided to go with 6 month which would provide a longer prediction horizon.

balance or (c) be moved to state 0 (up to date) if the repayment amount exceeds the balance outstanding in the previous month. Note that the months in arrears are not necessarily consecutive and default is said to occur once an account goes into 3 months in arrears, i.e. default will occur when an account goes into 3 months in arrears where the 3 missed payments may not happen in consecutive months. In this work, we do not consider other pathways to default, e.g. where an account can go into default because of its unlikeliness to repay even though it has not missed 3 payments. Also note that the definition of delinquency and default applied here is different to that used by the data provider. The minimum repayment amount applied in this work is lower than in our previous work to better represent banking practices; if the account is in debit, the minimum repayment is 2.5% or £5, whichever higher, of the outstanding balance from the previous month.

3.2. Intensity modelling

A summary of the intensity model methodology, is given in the Appendix. For more details including likelihood functions see Leow and Crook (2014), Andresen et al (1991) and Andresen et al (1993). In essence there are three components. The first is the transition matrix, usually denoted as $\alpha_{hji}(\tau)$, representing the transition intensities, i.e. the rate of change in the number of observations, between states *h* and *j* for account *i* at time τ. These are a function of covariates, $Z_{hji}(\tau - l)$, which denotes a vector of time varying covariates lagged *l* months, specific to the transition from state *h* to state *j*, and having values for case *i*, the parameters of which are to be estimated. Next is the generator matrix, usually denoted as $\mathbf{A}$, which is a square matrix of dimensions given by the number of possible states and consists of the intensity of experiencing an event cumulated up to any time τ. Finally, the transition matrix, usually denoted $P_i\left(s, t, Z_{hji}(\tau - l)\right)$, which gives the probabilities of individual *i* being in each state at time *t* given the state it was in at time *s*.

Based on the intensity models detailed in Leow and Crook (2014), we use the equivalent[2] application and behavioural variables (listed in Table 1, some omitted due to non-disclosure agreements), as well as include macroeconomic variables, to estimate the 6 transition

[2] Jump rate and employment type have been defined slightly differently in this work.

intensities for delinquency (state 0 to state 1, state 1 to state 2), default (state 2 to state 3) and recovery (state 1 to state 0, state 2 to state 1, state 2 to state 0). The transition intensities are then used to compute the generator matrices and the transition matrices, where the final output of interest are the individual probabilities of the 6 possible types of transitions over duration time of the loan. In this work, we compute transition probabilities for all accounts in the test set up to the maximum observed time in the test set. To be specific and to underscore the magnitude of the output, this means that we have a set of predicted probabilities for each of the 89,000 individual accounts, for each of the 6 possible transitions, and for each time point of between 8 to 26 months of the loan.

3.3. Simulation

For the purposes of the simulations, we specify the elements in $Z_{hji}(\tau - l)$ such that instead of $P_i(s,t,Z_{hji}(\tau - l))$, we re-write the transition matrix of probabilities as $P_i(s,t,X_i,Y_{hji}(\tau - l),Z_{hj}(\tau - l))$, where X_i represents the account-dependent, time independent variables, i.e. application variables; $Y_{hji}(\tau - l)$ represents account-dependent, time-dependent variables, i.e. behavioural variables, of the *hj* transition, lagged *l* months; and $W_{hj}(\tau - l)$ represents the account independent, time-dependent variables, i.e. macroeconomic variables, of the *hj* transition, lagged *l* months. Its elements are represented by $p_{i,hj}(s,t)$, and when the transition matrix is such that $t = s + 1$, then $p_{i,hj}(s,t)$ can be shorted to $p_{i,hj}(\tau)$.

3.3.1. For predicted events

Given the transition probabilities, we need to find a way to translate the predicted probabilities into predicted events. We observe that each of the empirically observed transition rates in the training set, $r_{hj}(\tau)$, is a form of time series data, and assume it to be a random walk, given in Equation 5.

$$r_{hj}(\tau) = r_{hj}(\tau - 1) + \varepsilon_{hj}(\tau) \tag{5}$$

where $\varepsilon_{hj}(\tau) \sim N(\mu_{hj}, \sigma_{hj}^2)$.

As we are investigating delinquency as well as defaults, we are interested in the movements between states that an account might experience during its lifetime. Given an initial known state and probabilities of transitions, we make a series of cascading predictions of states for each account at each time point, i.e. the predicted state at one time point will affect the state predicted at the following time point. Naturally, this set of predicted states is only one possible outcome, and where a different predicted state early on in the duration time of the loan could mean different predicted states in the time following that. By running a simulation[3], we are able to introduce some variability. Based on the single set of predicted transitions probabilities (for each individual account, for each unit of time over the specified observation time), we repeat this procedure an adequate number of times to eventually get a distribution of transitions and predicted events.

Based on the empirical transition rates in the training set, we generate transition rates over the duration time of the loan for the test set, denoted $\tilde{r}_{hj}^{\,r}(\tau)$, where the subscript *hj* represents the transition from state *h* to state *j* and the superscript *r* represents the r^{th} run in the simulation, given in Equation 6.

$$\tilde{r}_{hj}^{\,r}(\tau)=\tilde{r}_{hj}^{\,r}(\tau-1)+\left(\tilde{u}(0,1)\times\hat{\sigma}_{hj}\right)+\hat{\mu}_{hj} \qquad \tilde{r}_{hj}^{\,r}(\tau)>0 \tag{6}$$

where $\tilde{u}(0,1)$ is a random number between 0 and 1, $\hat{\sigma}_{hj}$ and $\hat{\mu}_{hj}$ are the estimated standard deviation and mean of $\varepsilon_{hj}(\tau)$.

These generated transition rates are then applied onto the ranked predicted probabilities in the test set to get a cut-off, $c_{hj}^{\,r}(\tau)$. Comparing these cut-offs against the predicted probabilities, $p_{i,hj}(\tau)$, we assign predicted states at each run *r*, $s_i^{\,r}(\tau)$, according to the

[3] Other methods of simulation were explored: (1) using the Fleishman power transformation to get randomly generated numbers to have the same distribution as observed transitions rates; and (2) generating random numbers between 0 and 1 as cut-off values – need to explain further, e.g. why were they not used in the end?

algorithm described in Equations 7 to 9. We get predicted states for each account *i* at each duration time τ for each run *r*. In this work, we did a total of 50 runs[4].

$$s_i^r(\tau) \mid s_i^r(\tau-1)=0=\begin{cases} p_{i,01}(\tau)>c_{01}^{\ r}(\tau), & s_i^r(\tau)=1 \\ p_{i,01}(\tau)\le c_{01}^{\ r}(\tau), & s_i^r(\tau)=0 \end{cases} \tag{7}$$

$$s_i^r(\tau) \mid s_i^r(\tau-1)=1=\begin{cases} p_{i,10}(\tau)>c_{10}^{\ r}(\tau), & s_i^r(\tau)=0 \\ p_{i,10}(\tau)\le c_{10}^{\ r}(\tau)\wedge p_{i,12}(\tau)>c_{12}^{\ r}(\tau), & s_i^r(\tau)=2 \\ p_{i,10}(\tau)\le c_{10}^{\ r}(\tau)\wedge p_{i,12}(\tau)\le c_{12}^{\ r}(\tau), & s_i^r(\tau)=1 \end{cases} \tag{8}$$

$$s_i^r(\tau) \mid s_i^r(\tau-1)=2=\begin{cases} p_{i,21}(\tau)>c_{21}^{\ r}(\tau), & s_i^r(\tau)=1 \\ p_{i,21}(\tau)\le c_{21}^{\ r}(\tau)\wedge p_{i,20}(\tau)>c_{20}^{\ r}(\tau), & s_i^r(\tau)=0 \\ p_{i,21}(\tau)\le c_{21}^{\ r}(\tau)\wedge p_{i,20}(\tau)\le c_{20}^{\ r}(\tau)\wedge p_{i,23}(\tau)>c_{23}^{\ r}(\tau), & s_i^r(\tau)=3 \\ p_{i,21}(\tau)\le c_{21}^{\ r}(\tau)\wedge p_{i,20}(\tau)\le c_{20}^{\ r}(\tau)\wedge p_{i,23}(\tau)\le c_{23}^{\ r}(\tau), & s_i^r(\tau)=2 \end{cases} \tag{9}$$

where $s_i^r(\tau)$ represents the state of individual *i* at time τ in run *r*, $p_{i,hj}(\tau)$ represents the predicted probability of individual *i* making a transition from state *h* to state *j* at time τ, and $c_{hj}^{\ r}(\tau)$ represents the cut-off value for the *hj* transition at duration time τ in run *r*.

3.3.2. For distribution of probability of default

One of the issues of a simulation that involves macroeconomic variables are the relationships and correlations amongst them. Naïve simulations from independent historical distributions is not appropriate as different macroeconomic variables have different relationships with each other, which may change over time. Preserving these relationships becomes more complex with a larger number of macroeconomic variables, and when behavioural variables are included, as it is in our work. One could model these relationships or extract factors or use a Choleski decomposition (see Bellotti and Crook 2012 and). But including the behavioural as well as the macroeconomic variables makes this tasks very complex.

We propose linking the macroeconomic variables with the behavioural variables experienced by each individual account at each single duration time, i.e. $Y_{iu}Z_u$, where *u* represents that particular time point account *i* is at, but not duration time. Due to the large number of

[4] We plan to increase this number to 1000. The results in this paper are very preliminary.

accounts we have and the length of time they were observed (January 2002 to December 2007[5], giving up to 84 months for each unique account), we create a large database from which we can draw unique combinations of observed values of behavioural and macroeconomic variables. This method does not upset any relationship between macroeconomic variables themselves, as well as any possible correlations between macroeconomic variables and behavioural variables. We also preserve the possibility of macroeconomic variables and/or behavioural variables having different relationships over time.

We then draw randomly and with replacement from this database (consisting of $Y_{iu}Z_u$ observations) to be applied onto test set accounts and their application variables, i.e. X_i, for each time-step in a specified *s* to *t* time period. This means that for each account and at each time and for each run *r*, we get simulated behavioural and macroeconomic variables, $\{X_i[Y_{iu}Z_u]^s, X_i[Y_{iu}Z_u]^{s+1}, \ldots, X_i[Y_{iu}Z_u]^t\}^r$, which we score (according to the model described in Section 3.2) and get the associated generator matrix, followed by the probability matrix, $P_i(s,t,X_i,Y_{hj}(u),Z_{hj}(u))$.

A possible application of this simulation would be to get a predicted distribution of (probability of) default and its 99.9 percentile value (the probability of default equivalent of Value at Risk). Given a reasonable period of time *s* to *t*, we have three possible transitions to default and we consider the predicted probabilities, $p_{i,03}(s,t)^r$ for transition 03, $p_{i,13}(s,t)^r$ for transition 13 and $p_{i,23}(s,t)^r$ for transition 23 at each run *r*, i.e. the probability of being in state 3 at time *t* given its state in time *s*. Over *r* runs, we get a distribution of predicted probabilities for each transition, as well as its 99.9 percentile value.

Due to the size of the dataset and the computational time, we run this simulation on a sample of the test set consisting of 1000 unique accounts, for duration time 9 to 24, and for100 runs.

---

[5] Although this period represents only non-downturn macroeconomic conditions, this simulation can be easily repeated using a different period.

4. Results

4.1. Parameter estimates

The signs of the parameter estimates of the intensity models are included in Table 1. By classifying the transitions into "delinquency transitions" and "recovery transitions", we see some overall trends. These are mostly aligned with previous work, although the inclusion of macroeconomic variables does change the parameter estimate signs, as well as the statistical significance of some variables.

In terms of application variables, borrowers who are more settled, e.g. have a longer relationship with the bank or been staying at the same address for a longer time, or have higher income, have a lower risk of delinquency and are more likely to recover should it go into delinquency. We find behavioural variables to be stable indicators of delinquency and recovery. Reassuringly, we find borrowers with higher credit limit have lower risk of delinquency, and borrowers who are able to make higher repayments amounts have higher risk of recovery. We also find borrowers who have had previous transitions are more likely to experience some type of transition again (delinquency and recovery) but are less likely to default, which seems to indicate that some borrowers either accidentally or habitually go into 1 or even 2 months in arrears but keep themselves out of default. For the macroeconomic variables, we find it difficult to pick out any specific trends. Higher values of FTSE and consumer confidence are associated with subsequent lower risk of delinquency and higher probability of recovery. Generally when unemployment is higher there is a higher chance of an account becoming delinquent and of defaulting and a lower chance of recovery six months later. Higher house prices and mortgage interest rates increase the chance of delinquency and default whilst higher house prices increase the chance of recovery, but higher mortgage rates do the opposite. Since house prices indicate greater average wealth these results are plausible. As house prices increase households may on average borrow more and on average have less discretionary income so over stretching themselves by borrowing more than they can afford. Higher mortgage rates would have the same effect. Correspondingly households would be less able to recover from delinquency. Puzzlingly, higher credit card interest rates and total debt outstanding reduce the chance of delinquency and increase the chance of recovery . Perhaps the additional debt supply is being used to repay credit card debt.

Table 1: Parameter estimate signs for intensity models.

Code	Variable	Delinquency Transitions						Recovery Transitions					
		TRANSITION 01		TRANSITION 12		TRANSITION 23		TRANSITION 10		TRANSITION 21		TRANSITION 20	
Application variables													
NOCards	number of cards	-	**	+	**	+	**	-	**	-	**	-	**
LLine	indicator for presence of landline	+	**	-		-	**	-	**	+	*	-	
TAAdd	time at address, in years	+		-	**	-	**	+		+		-	**
TWBank	time with bank, in years	-	**	-	**	-	**	+	**	+	**	+	**
TWBank MU	indicator for missing time with bank	-	**	-	**	-	**	+	**	+		-	
INC_L	income, ln	-	**	-	**	+	**	+	**	+	**	+	**
INC_M0	indicator for missing or 0 income	-	**	-	**	+	**	+	**	+	**	+	**
X_A	variable X, group A												
X_B	variable X, group B	+	**	+	**	+	**	-	**	-	**	-	**
X_C	variable X, group C	+	**	+	**	+	**	-	**	-		-	**
X_D	variable X, group D	+	**	+	**	+	**	-		+	**	-	**
X_E	variable X, group E	+	**	+	**	+	**	-	**	+		-	**
ageapp_1	age at application, group 1												
ageapp_2	age at application, group 2	-	**	-	**	+	**	+	**	-	**	-	**
ageapp_3	age at application, group 3	-	**	-	**	+	**	+	**	-	**	-	**
ageapp_4	age at application, group 4	-	**	-		+	**	-	**	-	**	-	**
ageapp_5	age at application, group 5	-	**	-		+	**	-	**	-	**	-	**
ageapp_6	age at application, group 6	-	**	-	**	+	**	-	**	-	**	-	**
ageapp_7	age at application, group 7	-	**	-	**	+	**	-		-	**	-	**
ageapp_8	age at application, group 8	-	**	-	**	+	*	+	**	-	**	-	**

ageapp_9	age at application, group 9	-	**	-	**	+	**	+	**	-	**	-	**
ageapp_10	age at application, group 10	-	**	-	**	+		+	**	-	**	-	**
Ecode_A	employment code, group A												
ECode_B	employment code, group B	+	**	+	**	-		+	**	+	*	-	
ECode_C	employment code, group C	-	**	-	**	+		+	**	-	**	+	**
ECode_D	employment code, group D	-	**	+	**	+	**	+	**	-	**	+	
ECode_E	employment code, group E	+	*	+	**	+		-		-	**	-	
Behavioural variables, lagged 6 months													
CLI_L_lag6	credit limit, ln	-	**	-	**	-	**	+	**	-	**	+	**
PAY_L_lag6	repayment amount, ln	+	**	-	**	-	**	+	**	+	**	+	**
PDRlag6	proportion of credit drawn	+	**	+	**	-	**	-	**	-	**	-	**
RJTOlag6	rate of total jumps	+	**	+	**	-	**	+	**	+	**	+	**
RSTDlag6	indicator for improvement in state from 3 months previous	-	**	+	**	+	**	-	**	+	**	-	**
Macroeconomic variables, lagged 6 months													
RPIN_lag6	retail price index, non-seasonally adjusted	-	**	-	**	-		-	**	-	**	-	**
AWEN_lag6	average wage earnings, non-seasonally adjusted	-	**	+	**	+	**	-	**	-	**	-	
FTSN_lag6	FTSE index, non-seasonally adjusted	-	**	-	**	-	**	+	**	+		+	**
UERS_lag6	unemployment rate, seasonally adjusted	+	**	-	**	+		-		+	**	-	
IOPN_lag6	index of production, non-seasonally adjusted	-	**	-	**	+		+	**	-		+	
HPIS_lag6	house price index, non-seasonally adjusted	+	**	+	**	+	**	+		-	**	+	
CONS_lag6	consumer confidence, non-seasonally adjusted	-	**	-	**	-	**	+	**	+	**	-	
CIRN_lag6	credit card interest rate, non-seasonally adjusted	-	**	-	**	-	**	+	**	+	**	+	
MIRN_lag6	mortgage loan interest rate, non-seasonally adjusted	+	**	-	**	+	**	-	**	+	**	-	**
LAMN_lag6	total credit outstanding, non-seasonally adjusted, ln	+	**	-	**	-	**	+	**	+	**	+	**

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

4.2. Graphs of probabilities of transitions

We look at the distribution of the predicted probabilities of transitions for each possible transition, segmented by certain borrower characteristics[6]. Each transition is identified by the label on the vertical axis of each graph thus 'index_m_ele12' indicates the (indexed) mean transition probability from state 1 to state 2. Overall, we see a similar trend over the entire portfolio: the probability of staying in state 0 (see transitions 0 to 0) increases as duration time of the loan increases; the probability of going into 1 or 2 months in arrears (see transitions 0 to 1 and 1 to 2 respectively) drop over time but might increase slightly in the second year of the loan. The probability of recovery, whether full recovery or into a lower state (see transitions 1 to 0, 2 to 0 and 2 to 1) is higher near the beginning of the loan rather than later; and the probability of going into default falls over time but increases and peaks around the 18 month mark.

Figure 1 displays the indexed mean probabilities of transitions segmented by age group at application. Indexed values are used for commercial confidentiality reasons. The probabilities of moving from up to date to 1 month in arrears are very small and with similar values across all age groups, with the middle-aged borrowers having a slightly higher probability to do so. This could imply that missing a single payment could happen quite easily and to anyone, and not necessarily a signal of distress. The youngest borrowers (borrowers below 22 years old represented by navy blue, category 1) have the highest probabilities of moving from 1 to 2 months in arrears (cell 12) and into default (cell 23). They also have the lowest probabilities of recovering from 1 or 2 months in arrears (cells 10 and 20). Interestingly, they also have the highest probabilities of recovering from 2 to 1 month in arrears, which seems to imply that they are able to get some help if required. The oldest borrowers (borrowers older than 62 and borrowers between ages 57 to 62 represented by green and mustard respectively) have the lowest risk of all groups with the lowest probabilities of making any delinquency transitions (cells 01, 12 and 23) and the highest probabilities of remaining up to date (cell 00) or recovering (cells 20 and 10).

[6] The indexed average predicted probabilities are segmented according to one particular borrower characteristic at any one time, where other characteristics of the borrower does still contribute towards that probability, i.e. other characteristics of the borrowers are not held constant.

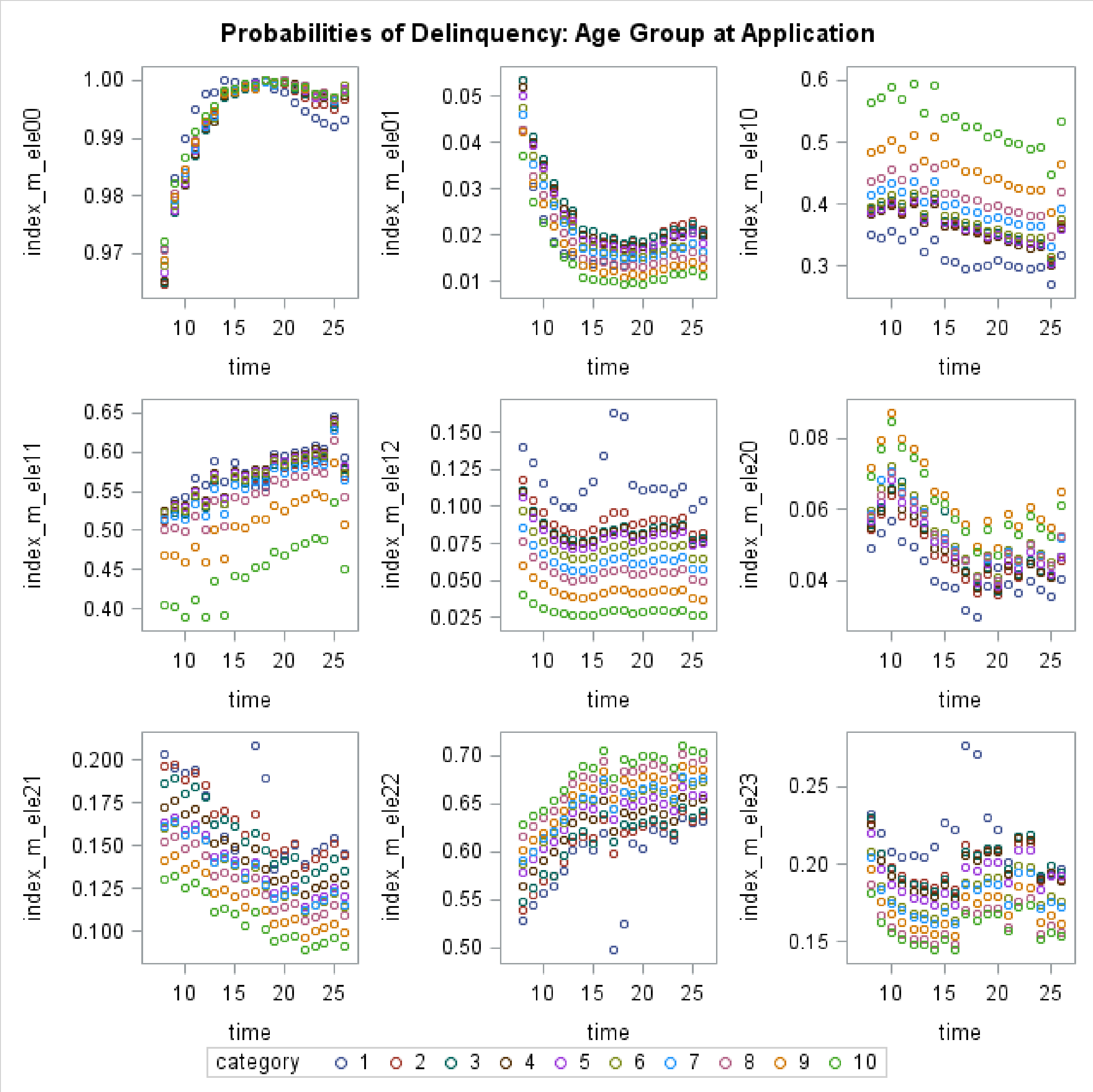

Figure 1: Mean indexed probabilities of transitions segmented by age group at application for test set. Indexed probabilities of transition from L-R for top row: transition 00, transition 01, transition 10; for second row: transition 11, transition 12, transition 20; for bottom row: transition 21, transition 22, transition 23. Details about the groups are given in the appendix.

Figure 2 shows the mean indexed transition probabilities by income group. Borrowers with the lowest incomes (those with 0 income, or income between 0 and £7,335 represented by blue and brown respectively) stand out from the rest of the income groups as having a higher risk of delinquency (cells 12 and 23) and a lower risk of recovery (cells 10, 21, and 20). As expected, borrowers with the highest income (above £60,000 represented by light purple) has high probabilities of making a full recovery when they are in arrears, with the highest probabilities of moving from 1 or 2 months in arrears to up to date (cells 10 and 20).

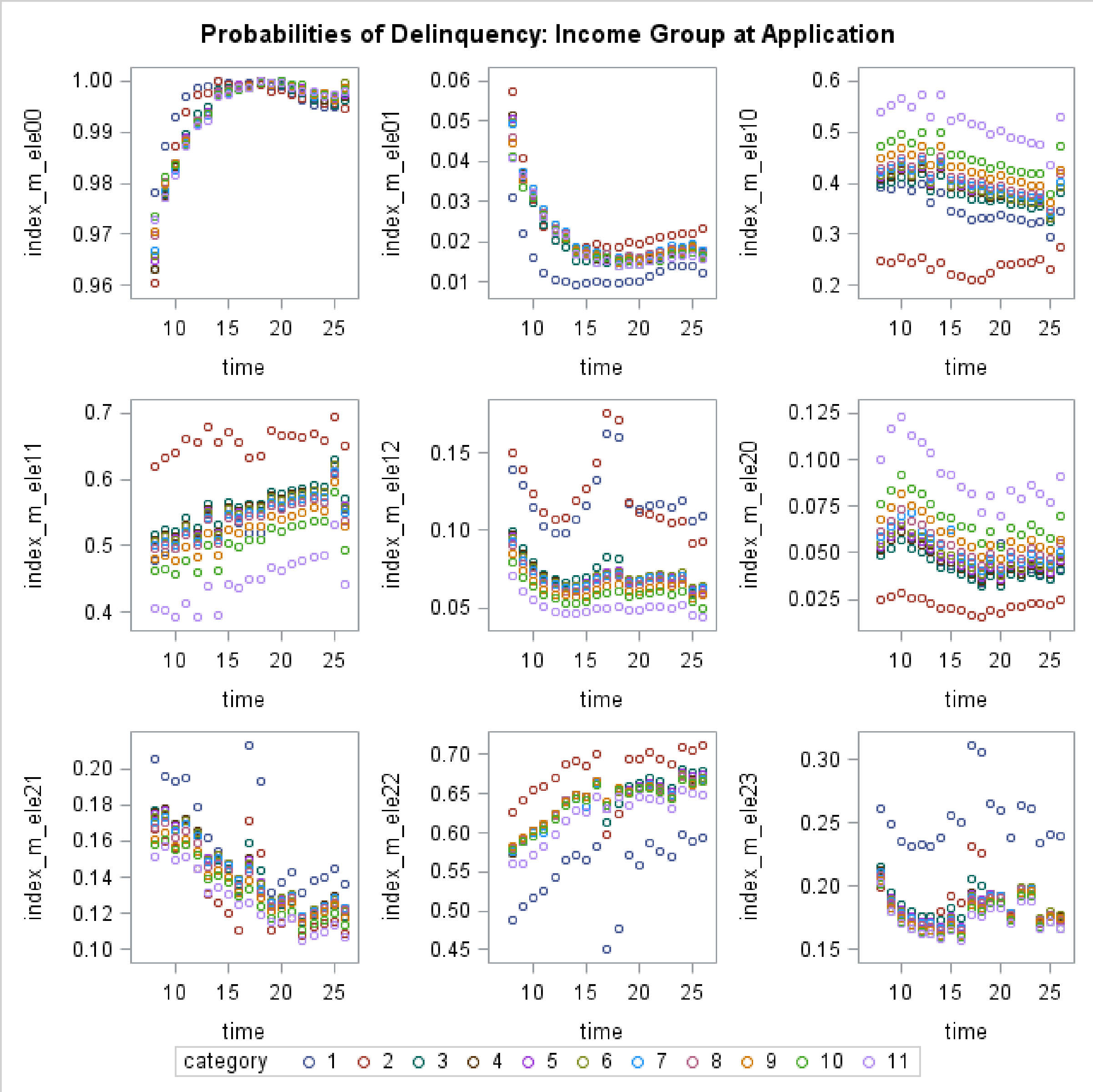

Figure 2: Mean indexed probabilities of transitions segmented by income group at application for test set. Indexed probabilities of transition from L-R for top row: transition 00, transition 01, transition 10; for second row: transition 11, transition 12, transition 20; for bottom row: transition 21, transition 22, transition 23. Details about the groups are given in the appendix.

Figure 3 shows indexed mean probabilities of transition by employment status. The retired and unemployed group have the lowest probabilities of moving into arrears (cells 01 and 12) and into default (cell 23), and the highest chance of recovery to being up to date if they do become delinquent (cells 10, 20). Surprisingly if they do become two payments in arrears they also have the lowest probability of recovering to one payment behind. Students have the highest probability of moving from two payments behind into default.

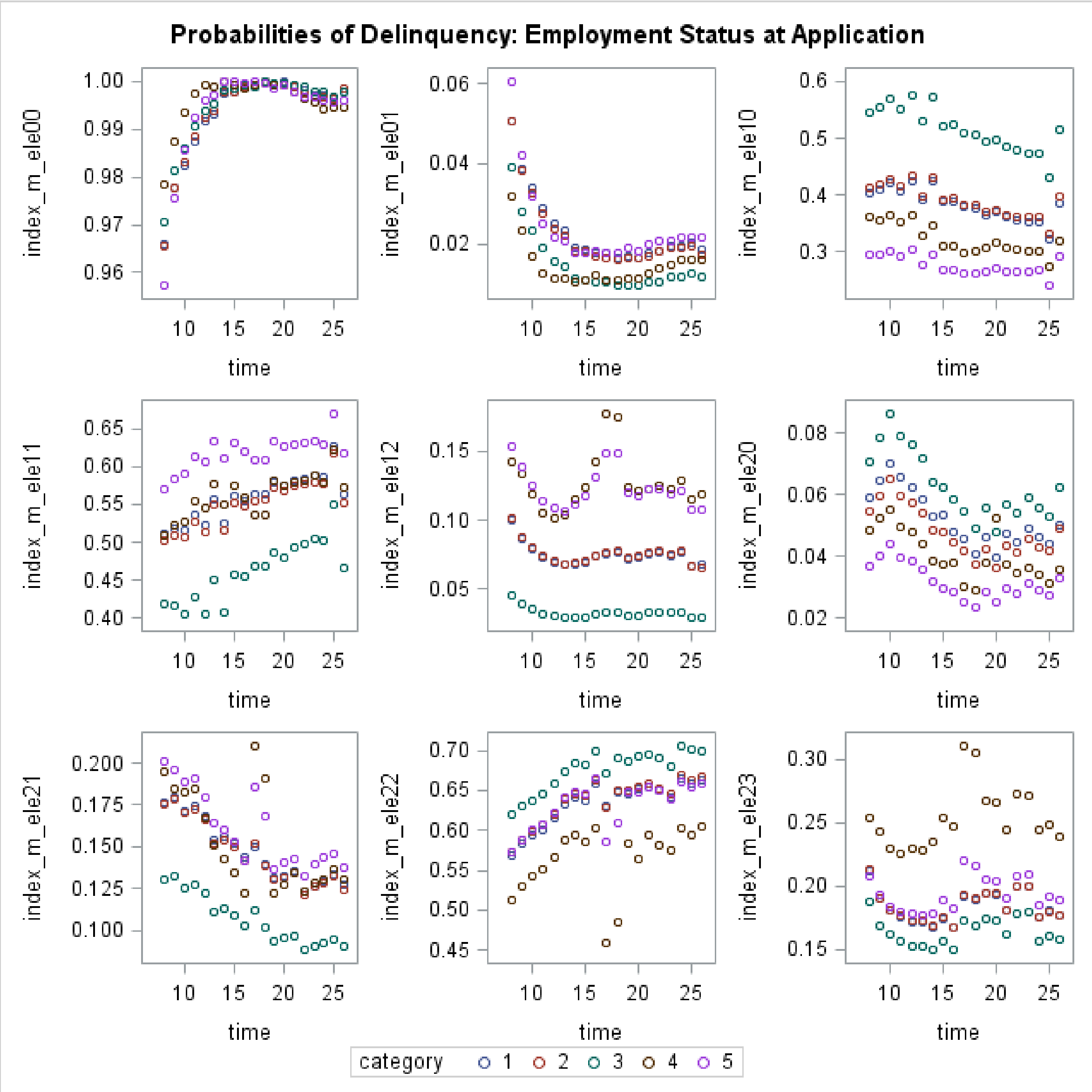

Figure 3: Mean indexed probabilities of transitions segmented by employment status at application for test set. Indexed probabilities of transition from L-R for top row: transition 00, transition 01, transition 10; for second row: transition 11, transition 12, transition 20; for

bottom row: transition 21, transition 22, transition 23. Details about the groups are given in the appendix.

4.3. Simulation I

In each run, we have the predicted state at each duration time for each individual account. By tabulating the number of predicted transitions at each duration time, we are able to get a distribution of predicted events and predicted transition rates.

Figure 4 consists of 6 graphs each representing the 6 predicted transition rates of the test set over 50 runs, where the horizontal axis represents duration time of the loan. Each blue dot on the graph represents a single run of the simulation, and the green and red lines represent the observed transition rates for the training and test set respectively. Results of the simulation show that our model seems to be able to capture the trend of the observed transition rates over the duration time of the loan. However, it tends to over-estimate the transition rates for delinquency transitions (c.f. graphs in the top row in Figure 4), where most of the predictions (represented by the blue dots) lie above the observed transition rates. The opposite is the case for the recovery transitions (c.f. graphs in the bottom row in Figure 4), where we see the predicted transition rates lie below the observed transition rates[7].

[7] The results are preliminary and we need to reconsider the random walk assumption to improve the fit of the simulated values.

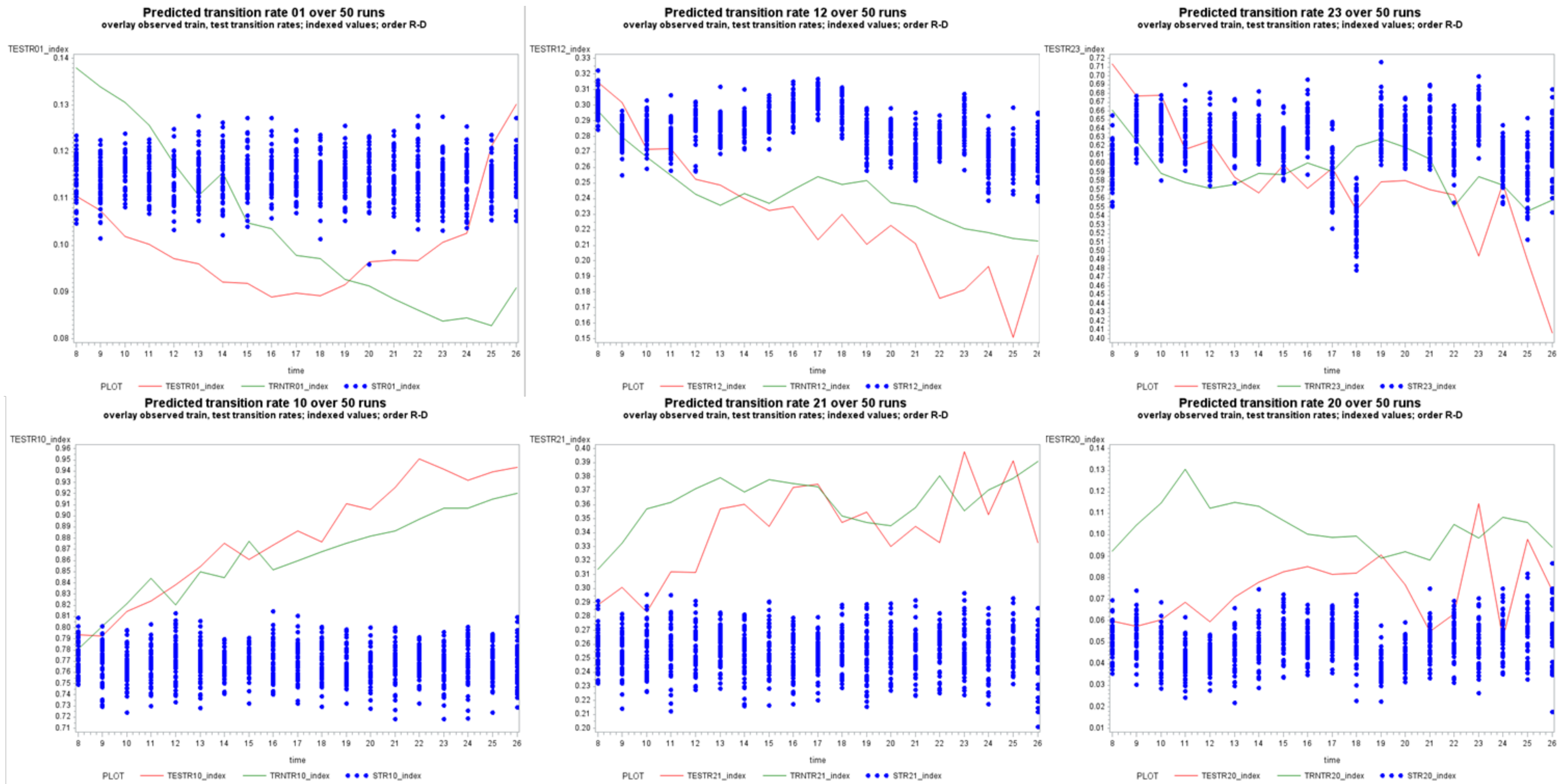

Figure 4: Simulation result of predicted transition rates for test set, over 50 runs. All horizontal axes represent duration time and vertical axes represent the indexed predicted transition rate for that transition. Results of the simulation are represented by the blue dots; the green and red lines represent the observed transition rates for the training and test set respectively. The top row consists of delinquency transitions, L-R: transition 01, transition 12, transition 23; the bottom row consists of recovery transitions, L-R: transition 10, transition 21, transition 20.

4.4. Simulation II

We present the distributions of (indexed) probabilities for three transitions between duration time 9 months and duration time 24 months. These transitions are from state 0 to state 3, from state 1 to state 3 and from state 2 to state 3, shown in Figures 5 to 7. Our preliminary results are for about 9,000 sample accounts (10%) from the test set and 100 runs only. From the simulations we find that the 99.9 percentile value for transition 03 is 6.6 times the mean; and 3.6 times and 1.4 times the mean for transitions 13 and 23 respectively (see Table 2). This suggests that when accounts are already in arrears (state 1 and 2), the probability of going into default is much higher (compared to if account is up-to-date), so the 99.9 percentile value is not many times larger than the mean.

	Transition 03	**Transition 13**	**Transition 23**
Mean (indexed)	0.0732	0.1569	0.5779
99.9 percentile (indexed)	0.4826	0.5674	0.8314

Table 2: Indexed values of mean and 99.9 percentile based on results from simulation containing 10% sample of test set for time 9 to 24 months, 100 runs

Notice also that the distributions for transitions from states 1 or 2 to 3 are much more symmetric than those for the transition from 0 to 3.

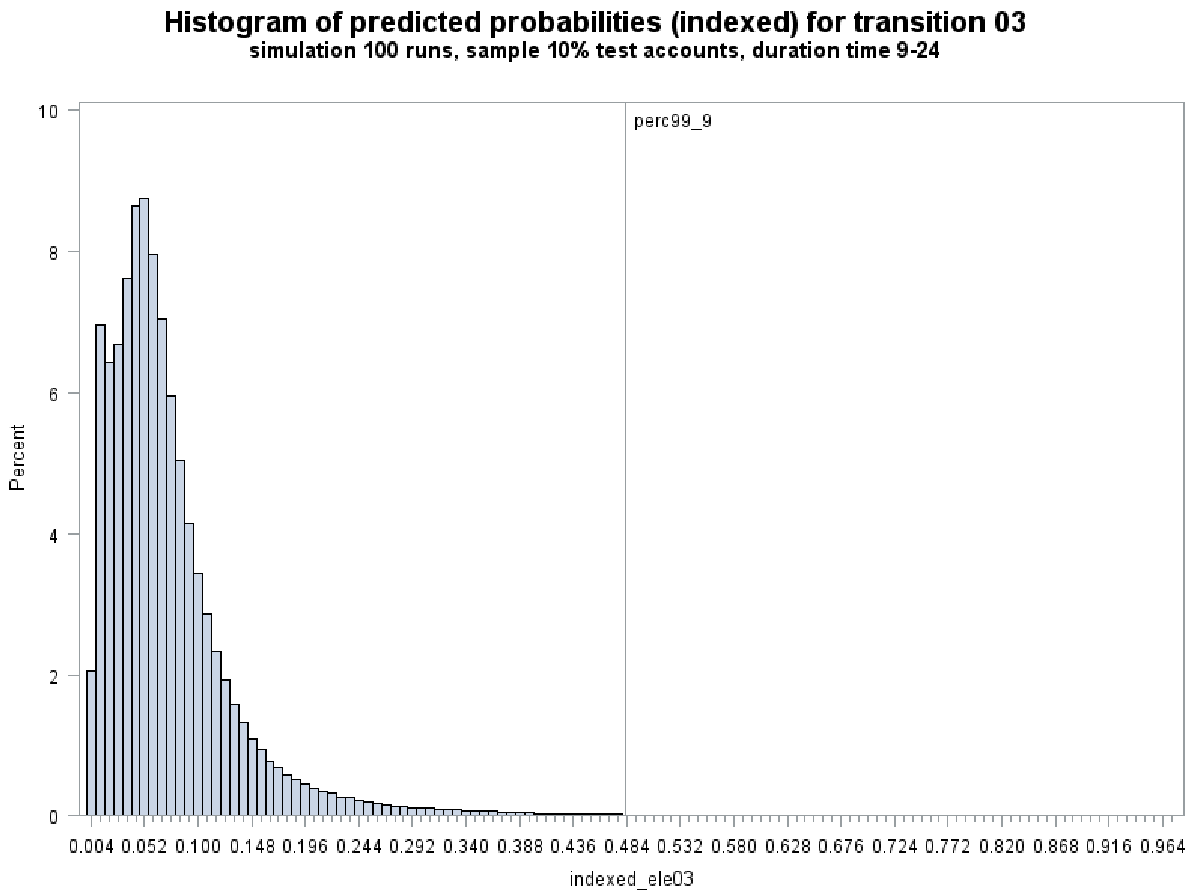

Figure 5: Histogram of indexed predicted probabilities for transition 03. Simulation of 100 runs based on 10% sample of test accounts, randomly sampled, for duration time 9 to 24 months, i.e. the probability of an account being in state 3 at time 24 given that it was in state 0 at time 9.

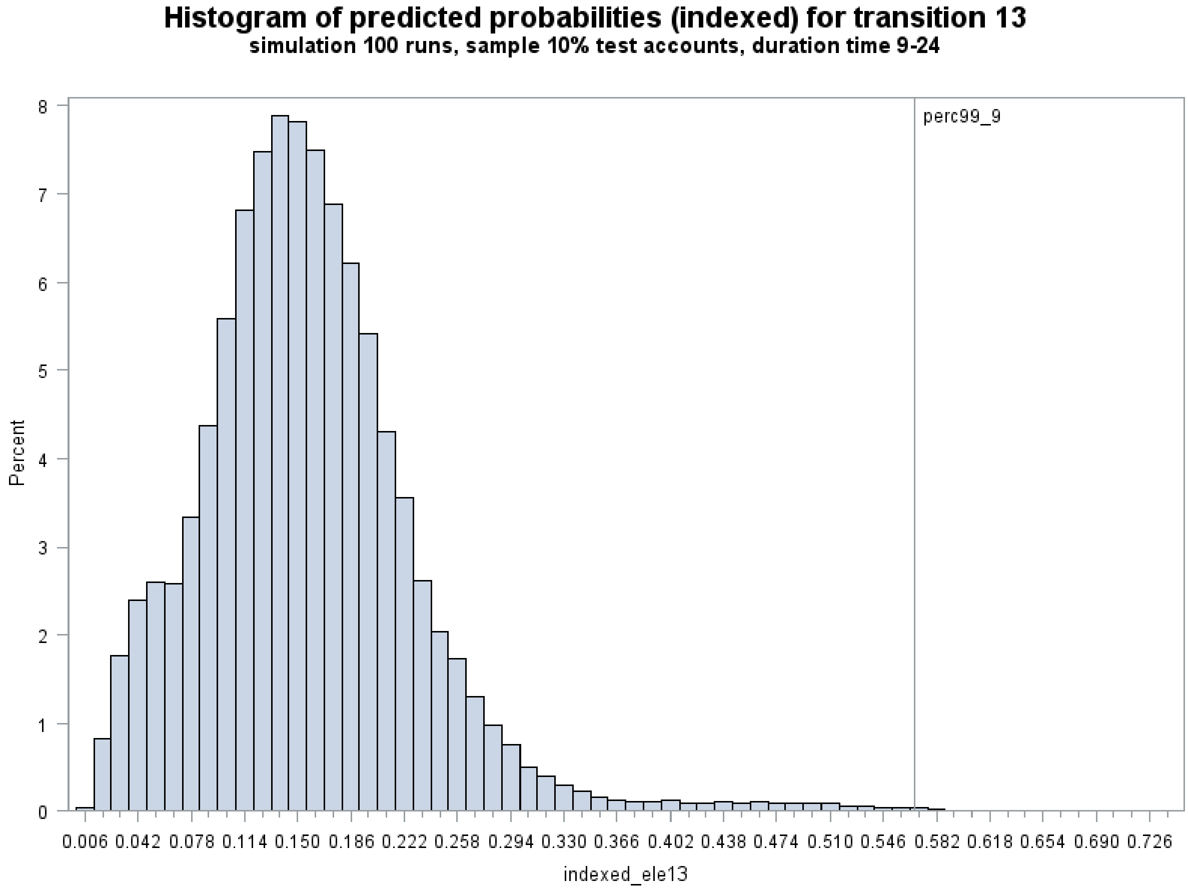

Figure 6: Histogram of indexed predicted probabilities for transition 13. Simulation of 100 runs based on 10% sample of test accounts, randomly sampled, for duration time 9 to 24 months, i.e. the probability of an account being in state 3 at time 24 given that it was in state 1 at time 9.

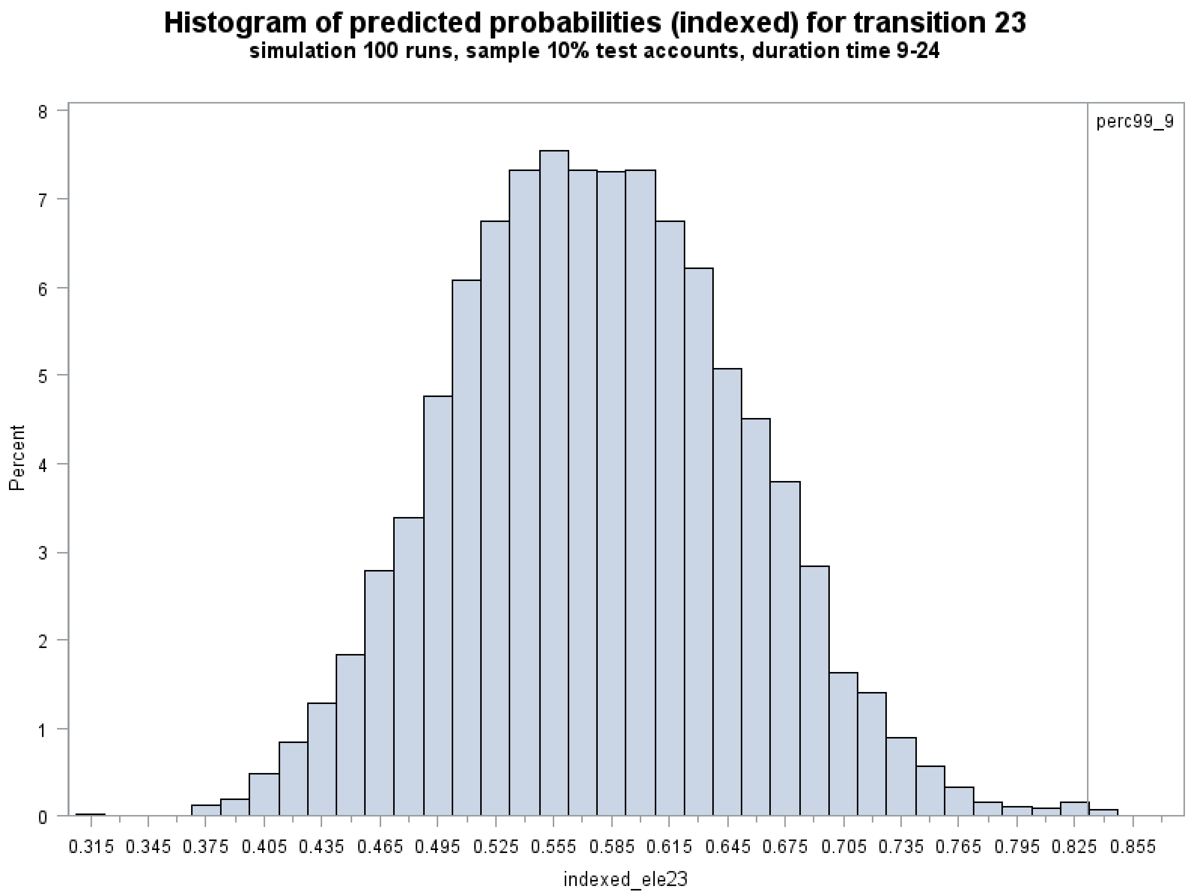

Figure 7: Histogram of indexed predicted probabilities for transition 23. Simulation of 100 runs based on 10% sample of test accounts, randomly sampled, for duration time 9 to 24 months, i.e. the probability of an account being in state 3 at time 24 given that it was in state 2 at time 9.

4. Concluding remarks

Our results show that various macroeconomic variables significantly influence transition intensity of credit card holders between delinquency states. The variables include indicators of nominal wealth, income and confidence. However complex dynamics appear to at play. The direction of effects of account level and behavioural variables on transition probabilities generally accord with expectations. We find the probabilities of remaining up to date increase until around 16 months of the loan and the probability of moving to more delinquent states decreases with time since account opening, this is especially marked for mild delinquency. The probability of recovery also decreases with account age especially from moderate delinquency to full recovery. Older card holders have a lower chance of becoming delinquent and moving into default and a higher chance of recovery. The youngest borrowers are the opposite. Lower income borrowers have the highest chance of delinquency and lowest chance of recovery and higher income borrowers have the highest chance of recovery. Consistently, the retired have the same patterns as the oldest age groups. We show a methodology to simulate delinquency events recursively over account duration time but more work on the exact form of the model is needed. We find that the VaR of the transition into default from up to date, from one payment behind and from two payments behind, decrease respectively.

References

Andersen, P.K., Borgan, Ø., Gill, R.D. & Keilding, N. (1993) *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Andersen, P.K., Hansen, L.S. & Kielding, N. (1991) non and semi-parametric estimation of transition oprobabilities from censored observation of a non-homogeneous Markov process. *Scandinavian Journal of Statistics*, 18, 154-167.

Duffie, D., Saita, L., & Wang, K.(2007) Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83, 635-665.

Grimshaw, S.D. & Alexander, W.P. (2011) Markov chain models for delinquency: transition matrix estimation and forecasting. *Applied Stochastic Models in Business and Industry*, 27, 267-279.

Ho, J. & Thomas, L.C., Pomtoy, T.A. & Scherer, W.T. (2004) Segmentation in Markov chain consumer credit behavioural models. In L. C. Thomas, J. N. Crook, & D. B. Edelman (Eds.), *Readings in credit scoring*. Oxford: Oxford University Press.

Kadam & Lenk (2008) Bayesian inference for issuer heterogeneity in credit ratings migration. *Journal of Banking and Finance*, 32, 2267-2274.

Lando D. & Skodeberg, T.M. (2002) Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance*, 26, 423-444.

Leow, M. & Crook, J. N. (2014) Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies. *European Journal of Operational Research,* 236**,** 685-694.

Malik, M. & Thomas, L.C. (2012) Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, 28, 183-195.

Stefanescu, C., Tunaru, R. & Turnbull, S (2009) The credit rating process and estimation of transition probabilities: A bayesian approach. *Journal of Empirical Finance,* 16, 216-234.

Appendix A Key for Figures X to XX
Age at application

Category in Figure	Income group at application
1	£0 or unknown
2	£0 to £7,335 inclusive
3	£7,335 to £12,000 inclusive
4	£12,000 to £15,000 inclusive
5	£15,000 to £17,625 inclusive
6	£17,625 to £20,460 inclusive
7	£20,460 to £24,000 inclusive
8	£24,000 to £30,000 inclusive
9	£30,000 to £42,000 inclusive
10	£42,000 to £60,000 inclusive
11	Above £60,000

Income group

Category in Figure	Income group at application
1	£0 or unknown
2	£0 to £7,335 inclusive
3	£7,335 to £12,000 inclusive
4	£12,000 to £15,000 inclusive
5	£15,000 to £17,625 inclusive
6	£17,625 to £20,460 inclusive
7	£20,460 to £24,000 inclusive
8	£24,000 to £30,000 inclusive
9	£30,000 to £42,000 inclusive
10	£42,000 to £60,000 inclusive
11	Above £60,000

Employment Status at Application

Category on graph	Translated: Employment Code
1	Employed
2	Self-employed, homemaker, part time
3	Retired, unemployed
4	Student

5	Not available

Appendix A
Key for graphs

Income group

Number of cards

Age at application

Employment status at application

Appendix B. Elements of the Intensity Model

The transition matrix, $\alpha_{hji}(\tau)$, representing the transition intensities, i.e. the rate of change in the number of observations, between states *h* and *j* for account *i* at time τ, is given in its general form in Equation 1.

$$\alpha_{hji}(\tau) = Y_{hi}(\tau)\alpha_{hj0}(\tau)\exp(\beta_{hj} Z_{hji}(\tau - l)) \quad (1)$$

where $Y_{hi}(\tau)$ is an indicator for whether individual *i* was in state *h* at time τ, $\alpha_{hj0}(\tau)$ is the baseline transition intensity for the state *h* to state *j* transition at time τ, β_{hj} is a vector of unknown regression coefficients, $Z_{hji}(\tau - l)$ is the vector of time-independent account-specific covariates (application variables), time-dependent account-specific covariates with lag *l* (lagged behavioural variables) and time-dependent account-independent covariates with lag *l* (lagged macroeconomic variables) for individual *i* who transit from state *h* to state *j* at time τ, and where $h \neq j$.

The generator matrix, $\mathbf{A}$, cumulates the intensities of experiencing an event up to any time τ. The non-diagonal and diagonal elements are computed according to Equations 2 and 3 respectively.

$$\begin{aligned} \hat{A}_{hji}(\tau, \hat{\beta}_{hj}, Z_{hji}(\tau - l)) &= \int_0^{\tau} \alpha_{hji}(u) du \\ &\cong \sum_0^{\tau} Y_{hi}(u)\alpha_{hj0}(u)\exp(\hat{\beta}_{hj} Z_{hji}(u - l)) \qquad h \neq j \end{aligned} \quad (2)$$

$$\hat{A}_{hhi}(\tau, \hat{\beta}_{hj}, Z_{hji}(\tau - l)) = -\sum_{h \neq j} \hat{A}_{hji}(\tau) \qquad h = 0...3 \quad (3)$$

Finally, we compute the transition matrix according to Equation 4, usually denoted $P_i\left(s,t,Z_{hji}(\tau-l)\right)$, which give the probabilities of individual *i* being in each state at time *t* given the state it was in at time *s*. By defining $t = s+1$, we get predicted probabilities for transitions at each time step:

$$
\begin{aligned}
\hat{P}_i\left(s,s+1,Z_{hji}(\tau-l)\right) &= \prod_{(s,s+1]}\left(\mathrm{I}+d\hat{A}_i\left(u;Z_{hji}(u-l)\right)\right) \\
&= \prod_{(s,s+1]}\left(\mathrm{I}+\hat{A}_i\left(u;Z_{hji}(u-l)\right)-\hat{A}_i\left(u-1;Z_{hji}(u-1-l)\right)\right)
\end{aligned} \tag{4}
$$

where $\prod$ denotes a product integral.