European Journal of Operational Research 236 (2014) 685–694

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Intensity models and transition probabilities for credit card loan delinquencies

Mindy Leow [1], Jonathan Crook *

Credit Research Centre, Business School, University of Edinburgh, United Kingdom
Business School, University of Edinburgh, 29 Buccluech Place, Edinburgh EH8 9JS, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 24 January 2012
Accepted 17 December 2013
Available online 7 January 2014

Keywords:
Risk analysis
Probability of default
Intensity modelling
Time-varying covariates
State space modelling
Retail loans

ABSTRACT

We estimate the probability of delinquency and default for a sample of credit card loans using intensity models, via semi-parametric multiplicative hazard models with time-varying covariates. It is the first time these models, previously applied for the estimation of rating transitions, are used on retail loans. Four states are defined in this non-homogenous Markov chain: up-to-date, one month in arrears, two months in arrears, and default; where transitions between states are affected by individual characteristics of the debtor at application and their repayment behaviour since. These intensity estimations allow for insights into the factors that affect movements towards (and recovery from) delinquency, and into default (or not). Results indicate that different types of debtors behave differently while in different states. The probabilities estimated for each type of transition are then used to make out-of-sample predictions over a specified period of time.

1. Introduction

Risk models for retail portfolios of financial institutions, as well as within the academic literature, have not been developed as extensively as they have been in the corporate sector, mainly due to the availability and inaccessibility of the necessary data. However, with the financial crisis in 2008, awareness and the importance of credit risk management have increased, and new insights were gained, especially in terms of how correlated loans losses, debtor behaviour and the economic climate can be. In that, there has been work in the corporate sector, estimating Probability of Default (PD) and Loss Given Default (LGD) models with the inclusion of macroeconomic variables (for example, see Frye (2000a, 2000b) for PD; Gupton and Stein (2002, 2005) for LGD), but only recently has this been undertaken for retail loan credit models (for example, see Bellotti and Crook (2010), Pennington-Cross (2003) for PD; Bellotti and Crook (2012), Leow, Mues, and Thomas (2011) for LGD).

Using a large dataset of credit card loan accounts provided by a major UK bank, we develop intensity models to predict delinquency and default. Our work differs from existing work in a number of ways. The majority of retail loans PD models currently in the literature are of static regression models (see Crook and Bellotti (2010) and Leow and Mues (2012)), where models predicting default are developed using loan application characteristics and are valid only within a specified outcome period, e.g. within 12 months of opening. Such models are also unable to handle accounts that are active but have not (yet) experienced any event (known as censoring) or closed, so such accounts are usually deleted from the dataset used to develop such models. Furthermore, these models are only able to account for time-varying covariates at any single snapshot in time yet these indicators essentially change over time, so are unable to adequately incorporate the effect of macroeconomic variables. Subsequent work has been based on the use of survival models (see Banasik, Crook, and Thomas (1999), Stepanova and Thomas (2002) and Bellotti and Crook (2010)) for default risk, which will allow for a more dynamic prediction of events. Such models will predict not just the probability of whether an event will occur (and not limited to a pre-defined outcome period), but also the (conditional) probabilities of that event occurring over time. Although survival models can account for different types of events (via competing risks), they are based on the assumption that the risk of each event occurring is independent right up to when any event occurs, which might not necessarily hold true. For example, in the months leading up to the default event, there would be certain behavioural indicators which take on values to indicate two or more missed monthly repayments, and this should mean that the risk of default increases more than the risk of early prepayment. There have also been papers suggesting the use of Markov chains (see Ho, Thomas, Pomroy, and Scherer (2004) and Malik and Thomas (2012)), and although these have been useful in trying to quantify the behaviour of consumers, they have the complication of having to assume stationarity and finding the

* Corresponding author at: Credit Research Centre, Business School, University of Edinburgh, United Kingdom. Tel.: +44 131 650 3802.
E-mail addresses: mindy.leow@ed.ac.uk (M. Leow), j.crook@ed.ac.uk (J. Crook).
[1] Tel.: +44 131 650 9850.

0377-2217/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.12.026

appropriate first, second or third order chain. We propose the use of intensity models to predict for delinquency and default, which have been previously applied to the estimation of rating transitions of corporate loans (see Jarrow, Lando, and Turnbull (1997), Duffie, Saita, and Wang (2007) and Lando and Skødeberg, 2002), but have not been used on retail loans yet. Also, other approaches for estimating transition probabilities have been applied to corporate loans, for example the standard unobserved latent factor model and Bayesian methods (see for example Stefanescu, Tunaru, and Turnbull (2009) and Kadam and Lenk (2008)) but in these papers only aggregated data was used.

In this work, we do not just focus on the prediction of default. Instead, using both application and behavioural variables, we model time to delinquency, and then to default, based on how debtors have behaved throughout their loan period as well as how they might have handled previous experiences of periods in arrears. As such, we use an alternative definition of default here: three months of missed payments, but not necessarily consecutive; instead of the conventional one of defining default to occur when the debtor has missed three consecutive months (or 90 days' worth) of payments (The Financial Services Authority (2009), BIPRU 4.3.56 and 4.6.20). This then allows us to define four states chronicling the progression from up-to-date to default (to be defined in the following section), as well as when accounts move towards (or away) from default. Credit card accounts are tracked over a period of time where transitions between the various states could be affected by the individual characteristics of the debtor at time of application and how the debtor has managed their finances since gaining the credit account. Other external factors, like macroeconomic variables, could also be included but are not considered in this work. Each possible transition in this intensity model is modelled separately via a semi-parametric multiplicative hazard model with time-varying covariates (see Andersen, Borgan, Gill, and Keiding (1993) and Andersen, Hansen, and Keiding (1991)), which are then calibrated to get the probabilities of moving (or staying) between states. Although this methodology has been detailed in a number of academic papers including Andersen et al. (1991), Jarrow et al. (1997), Lando and Skødeberg (2002) and Berd (2005) among others, they focus mainly on the estimation of parameter estimates, and where predictions were done, they were only in the time-homogenous case. The models we advance allow one to estimate a complete matrix of transition probabilities between any two repayment states between any two duration time periods for each case (in our application each account). By applying cut-offs they allow the prediction of the numbers of cases (in a sample) that would be expected to transit from one repayment state in one time period to any other state in any other time period the analyst may choose.

The models have several important advantages over cross-section regression models and over survival models. Compared with the former, the user can gain predictions of the probability of moving between states in any future time period, not just the time period chosen at the time of model estimation. Also, time varying covariates can be incorporated. Compared with simple survival models, intensity models give many additional types of predictions, for example predictions of entire transition probability matrices in any future time period for each borrower, rather than merely the hazard probability of default occurring in any specified duration time.

Intensity models have several potential uses by practitioners in financial institutions. First, since intensity models allow the predictions of transition probability matrices in any future time period, they would be crucially useful to the computation of a financial institution's economic capital in any future time period. Second, by developing a model that can predict the different states of delinquency, not only are we be able to get predictions for default over time, we are also be able to get more intricate predictions for *each state of delinquency* leading up to default. This would enable a lender to attain insights into the factors that affect movements towards, and recovery from, delinquencies, as well as factors contributing towards a move from delinquency into default. So a lender could identify those types of borrowers that are likely to progress to default, those likely to progress to merely two or just one payment in arrears and also those likely to *recover* from being in arrears. The models also allow the identification of *when* each borrower who is in arrears is likely to recover. Third, although we do not go into detail here, this work could also be used by a lender to predict default risk in low- or zero-default portfolios. The analysis conducted on a low-default portfolio could underestimate default risk, but this might be mitigated by taking into account, by using the models we advance, the episodes where accounts go into arrears but not default.

The rest of this paper is structured as follows: The data and notation are described in Section 2. Section 3 describes the methodology, and Sections 4 and 5 detail results and predictions respectively. Section 6 concludes.

2. Data and definitions

Data was supplied by a major UK bank and were active credit card accounts from all parts of the UK. This large dataset of more than 49,000 unique accounts is a random sample of credit cards that were issued from January 2002 up to June 2005, as well as their monthly histories since the account was opened, up to June 2006 or the time at which the credit card account was closed, whichever is earlier. Accounts that were still active in June 2006 are said to be censored at that time. As part of pre-processing, accounts that do not have consecutive monthly observations were removed, as well as accounts where their history did not start from the time the credit card was issued. Account and debtor characteristics available are common application variables, including type of employment, length of time the debtor had been with the bank, time at address and age; and behavioural variables available on a monthly basis, including spending and repayment amounts, credit limit and outstanding balance.

From accounts that were active during the period of May 2005 to June 2006, a random sample of unique accounts (20%), as well as all their respective monthly observations, were selected and kept separate as the validation sample. All other accounts were included in the training sample. Thus, the training and validation sets are kept completely separate.

2.1. Minimum repayment amount

A minimum repayment amount is required for the assignment of states but this information was not directly available from the provider of the data. We define τ as duration time since an account was opened and we define the minimum repayment amount in duration month τ, M_τ, to be the higher of 1% of the outstanding balance in month $\tau - 1$ or £5. It is possible for the minimum repayment amount to be £0, if there is zero outstanding balance on the account. Also, if the account is in credit, the minimum repayment amount required is also defined to be £0.

2.2. Definition of states

Four states are defined: state 0 means that the account is up to date; states 1–3 mean that the account is in one, two and three months in arrears, respectively. Note that these months in arrears do not necessarily have to be consecutive. State 3 is also known as the default state, so any account that enters state 3 is said to have entered default. For the purpose of this work, the default state is