International Journal of Forecasting 29 (2013) 563–574

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Forecasting and stress testing credit card default using dynamic models

Tony Bellotti [a], Jonathan Crook [b,*]

[a] Imperial College - Mathematics and Statistics, London, United Kingdom

[b] University of Edinburgh - Business School, 50 George Square, Edinburgh, Mid Lothian EH8 9JY, United Kingdom

ARTICLE INFO

Keywords:
Survival analysis
Stress testing
Credit risk

ABSTRACT

We present discrete time survival models of borrower default for credit cards that include behavioural data about credit card holders and macroeconomic conditions across the credit card lifetime. We find that dynamic models which include these behavioural and macroeconomic variables provide statistically significant improvements in model fit, which translate into better forecasts of default at both account and portfolio levels when applied to an out-of-sample data set. By simulating extreme economic conditions, we show how these models can be used to stress test credit card portfolios.

1. Introduction

Application consumer credit scoring models use details about obligors or potential customers that are static. Such models are used to determine whether an applicant should be granted credit, based on data which are collected at the time of application and then remain fixed. Behavioural consumer credit scoring models use both information collected at the time of application and behavioural variables, the values of which have changed over time, but which are fixed at the time of estimation. Both are cross-sectional models and allow the prediction of a probability of default within a specified time window, like eighteen months. However, models that answer more specific questions can also be estimated from credit portfolios, since they provide panel data (Crook & Bellotti, 2010) for a sample of obligor accounts. Panel data allow one to estimate hazard models which predict the probability of an event (such as a default) occurring in the next instant of time, conditional on the event not having occurred before that time, for any future time period one chooses. Unlike cross-sectional models, in a panel model one can include variables whose values change over the estimation period. Of particular relevance here are common economic risk factors that affect all obligors in a portfolio in generally the same way. For example, we would expect that a large increase in interest rates would cause, *ceteris paribus*, a general increase in the probability of default (PD). Time varying behavioural variables may also be included. We call these sorts of models that include time varying covariates (TVCs) 'dynamic models'. Furthermore, static models typically only have value in assessing the riskiness of applicants and obligors. However, if we want a complete picture we should be looking at the *return* alongside risk, which requires the use of dynamic rather than static models (Ma, Crook, & Ansell, 2009; Thomas, Ho, & Scherer, 2001). In this paper, we present dynamic models of default which include time varying behavioural variables (BVs) and macroeconomic variables (MVs), in addition to application variables (AVs).

The inclusion of MVs also enables us to perform stress tests, since extreme economic conditions can be simulated and included in the model in order to generate a measure of stressed loss (default rates). Accurate stress tests are becoming increasingly important in evaluating the risk to banks, as is evident from the evaluation of US banks (Board of Governors, 2009) and the recognition by the Financial Services Authority (2008) that stress testing is a key tool in helping financial institutions to make business strategy, risk management and capital planning decisions.

Our paper contributes to the literature in four ways. Firstly, for a large portfolio of credit card accounts, we show

* Corresponding author. Tel.: +44 131 650 3802; fax: +44 131 668 3053.
E-mail address: J.Crook@ed.ac.uk (J. Crook).

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.04.003

that including behavioural variables improves the model fit in a discrete time hazard model, and that their inclusion improves the forecast accuracy. Secondly, we find that, while several MVs are statistically significant explanatory variables of default, this does not translate into improved forecasts at the account level. Thirdly, we show that including MVs can improve the estimation of loss (default rate) at the portfolio level. Fourthly, using account level data, we demonstrate the use of MVs for stress testing and report the distribution of expected default rates based on a Monte Carlo simulation of economic conditions.

In Section 2, we provide a literature review. In Section 3, we outline the methods we use, describing the discrete survival model, our test procedures and the stress testing methodology. We then describe our data in Section 4, and present some results in Section 5. Finally, we offer conclusions and a discussion in Section 6.

2. Literature review

Several modelling techniques have been proposed for developing a dynamic model of credit (see Crook & Bellotti, 2010, for a review). Thomas et al. (2001) describe how a Markov chain stochastic process can be used as a dynamic model of delinquency. However, the approach they describe does not allow for model covariates, although models can be built on separate segments to allow the modelling of different risk groups. They also describe survival analysis as a means of building dynamic models, since this readily allows the inclusion of BVs and MVs as time-varying covariates (TVCs). Bellotti and Crook (2009) follow this path, using the Cox proportional hazard survival model to model the time to default for a large database of credit cards. They include MVs, but not BVs, as TVCs, and find a modest improvement in predictive performance in comparison to a static logistic regression. We take a similar approach using a survival model here, but with *discrete* time survival analysis. Discrete survival analysis can also be understood as a logistic regression on a panel data set, with the data arranged so that default is conditional on no prior default having occurred on that account. Since credit data are usually in the form of panel data, and in particular account records are discrete (e.g. monthly records), this is a more natural choice than continuous time survival analysis. It also has the advantage of being more computationally efficient, since probability forecasts involve simple summations over time periods, rather than an integration which may be complex when TVCs are included in the model.

Discrete survival models have been applied successfully to the analysis of personal bankruptcy and delinquency in the USA (Gross & Souleles, 2002), mortgage terminations (Calhoun & Deng, 2002), and competing risks of foreclosure and sales in the US subprime market (Gerardi, Shapiro, & Willen, 2008). Gross and Souleles (2002) used several different BVs and MVs. In particular, they included the outstanding account balance and repayments, and found that the former had a positive effect on bankruptcies, while the latter had a negative effect. They also found that the local unemployment rate had a statistically significant positive effect on bankruptcy, which is what we would expect, since an increase in unemployment is likely to affect some obligors adversely. Calhoun and Deng (2002) derived dynamic variables measuring the probability of negative equity and the mortgage premium. Both change over time and have a positive effect on default. They also included the ratio of 10-year to 1-year Constant Maturity Treasury yields, and found it to be statistically significant for models of early repayment. For fixed-rate mortgages, the coefficient increases for higher ratios, with the rationale that mortgagors will be moving to adjustable-rate mortgages, in order to take advantage of the short-term relatively low interest rates. Gerardi et al. (2008) found that interest rates (the 6-month libor rate) and the unemployment rate are statistically significant explanatory variables for both mortgage default and sales, with a positive effect on default, as we would expect, and a negative effect on sales. These studies have shown that both BVs and MVs are useful explanatory covariates for consumer credit risk. We therefore extend this work by using these dynamic models for forecasts and stress testing. Ultimately, financial institutions and regulators are interested in consumer credit risk models for the estimation of future losses at both the account and portfolio levels, either in normal (expected) circumstances or when considering adverse conditions. For this reason, we focus primarily on using the models for forecasting PD and the default rate.

The literature on stress testing is growing rapidly. In an early paper, Berkowitz (2000) proposed a stress testing methodology in which two separate forecast distributions are generated using a risk model: one for normal conditions and another reflecting stressed conditions, based on changes in an underlying factor. Our approach is rather different to that of Berkowitz (2000), in that we do not choose an initial distribution under stressed conditions; instead, we generate a single distribution of expected default rates and focus on the upper percentiles for stress testing.

Stress-testing models may be divided into macro stress-testing models and micro stress-testing models. Macro stress-testing models concern the implications of stressed states of the economy for the capital of *groups* of institutions, and the aim is to examine the capital adequacy of the financial sector of an economy in the event of adverse shocks. In contrast, a micro stress test relates to a specific portfolio of one lender. Sorge and Virolainen (2006) divide macro stress tests into (i) those that relate aspects of banks' balance sheets to macroeconomic activity, and (ii) value at risk models where macroeconomic factors are related to aggregate default rates (not the probabilities of default of individual obligors). For examples of balance sheet models, see Delgado and Saurina (2004) and Drehmann, Sorensen, and Stringa (2010).

Macro Value at Risk (VaR) stress tests and micro stress tests follow similar methodologies. First, a model that relates macroeconomic variables to each other, often a Vector Autoregressive Regression model, is estimated. Second, a default rate (macro) or probability of default (PD) (micro) model that incorporates macroeconomic variables is parameterised. Third, a macroeconomic scenario is chosen and its implications for the distributions of PD (or expected loss) are predicted, or impulse response functions are estimated. For examples of such macro stress tests,