International Journal of Forecasting 28 (2012) 171–182

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

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

Loss given default models incorporating macroeconomic variables for credit cards

Tony Bellotti, Jonathan Crook *

University of Edinburgh – Business School, 50 George Square Edinburgh Mid Lothian EH8 9JY, United Kingdom

ARTICLE INFO

Keywords:
Loss given default
Credit cards
Basel II

ABSTRACT

Based on UK data for major retail credit cards, we build several models of Loss Given Default based on account level data, including Tobit, a decision tree model, a Beta and fractional logit transformation. We find that Ordinary Least Squares models with macroeconomic variables perform best for forecasting Loss Given Default at the account and portfolio levels on independent hold-out data sets. The inclusion of macroeconomic conditions in the model is important, since it provides a means to model Loss Given Default in downturn conditions, as required by Basel II, and enables stress testing. We find that bank interest rates and the unemployment level significantly affect LGD.

1. Introduction

Loss Given Default (LGD) is the loss incurred by a financial institution when an obligor defaults on a loan, given as the fraction of exposure at default (EAD) unpaid after some period of time. It is usual for LGD to have a value between 0 and 1, where 0 means that the balance is fully recovered and 1 means the total loss of EAD. LGD is an important value that, for several reasons, banks need to estimate accurately. Firstly, it can be used along with the probability of default (PD) and EAD to estimate the expected financial loss. Secondly, a forecast of LGD for an individual can help to determine the collection policy to be used for that individual following default. For example, if a high LGD is expected, then more effort may be employed to help reduce this loss. Thirdly, an estimate of LGD, and therefore of the portfolio financial risk, is an integral part of the operational calculation of capital requirements to cover credit loss during extreme economic conditions. The Basel II Capital Accord (Basel Committee on Banking Supervision, 2006) allows banks the opportunity to estimate LGD using their own models via the advanced internal ratings based (IRB) approach.

In this paper we focus on modelling and forecasting LGD for UK retail credit cards based on account variables (AVs), and also on the inclusion of macroeconomic variables (MVs). Our prior expectation is that as interest rates rise, so the cost of mortgages and other debts will increase, making it more difficult for an obligor to repay outstanding credit card balances, thus increasing the mean LGD. Equally, an increase in the level of unemployment means that more people find themselves in circumstances where they cannot repay credit, which also increases the mean LGD. On the other hand, an increase in earnings means that more people have more income available to pay off debts, and therefore decreases the mean LGD. In addition, some defaulters will be less able to repay than others when the state of the economy changes. For example, those who are unemployed at the time of a credit card application may be particularly sensitive to interest rate increases, as may home owners with mortgages. Similarly, borrowers with higher default balances may be particularly sensitive to increases in interest rates. For this reason we also consider interactions between MVs and account data. We consider four key research questions:

Q1. Which credit card application and default variables are the key drivers of retail LGD?

* Corresponding author. Tel.: +44 131 650 3802; fax: +44 131 650 8337.
E-mail address: j.crook@ed.ac.uk (J. Crook).

0169-2070/$ – see front matter
doi:10.1016/j.ijforecast.2010.08.005

Q2. What is the best modelling approach for retail LGD?
Q3. How well do the models perform for forecasting LGD?
Q4. Does the inclusion of MVs lead to improved models of retail LGD?

We investigate these questions by building several alternative models of LGD. We find that there are many important drivers of LGD taken from application details and default information. Given that the distribution of LGD is a bimodal U-shape, we consider a Tobit model and a decision tree model, along with various transformations of the dependent variable. Although LGD is not easy to model and a poor model fit is typical, we nevertheless find that models can be built which provide improved estimates of LGD and good forecasts of the mean LGD across a portfolio of accounts. Surprisingly, we find that the best forecasting model is Ordinary Least Squares (OLS) regression. Economic conditions are included as values of MVs for the bank interest rate, the level of unemployment and the earnings growth at the time of the account default. We find that the first two MVs are statistically significant explanatory variables that give rise to improved forecasts of LGD in hold-out tests at both the account and portfolio levels. Building LGD models with MVs also addresses the Basel II requirement to estimate the "downturn LGD", since stressed values of MVs can be used in the model to forecast LGD in poor economic conditions. This can be done by stressing interest rate values, as we explain in our conclusions.

The modelling and forecasting of LGD for retail credit using macroeconomic conditions is a new area of study. There is an extensive body of literature regarding LGD models for corporate loans (see for example Altman, Resti, & Sironi, 2005). However, there is less about *forecasting* LGD. An exception is the work of Gupton and Stein (2005), who describe a predictive LGD model for corporate loans using Moody-KMV's Losscalc© software. There is also very little in the literature regarding retail credit LGD, even though this is a large financial market: total lending in the UK consumer credit market reached over £1.4 trillion in 2009 (source: Bank of England). Grippa, Iannotti, and Leandri (2005) published empirical LGD models for a sample of 20,724 Italian accounts, including both small businesses and households. They observed differences in LGD and recovery periods across different geographic regions and recovery channels. They also conducted a multivariate analysis that showed a statistically significant negative relationship between the presence of a collateral or personal guarantee and LGD, and a positive relationship with the size of the loan. However, the range of variables used is far more limited than that available to financial institutions that have made credit card or personal loans, and the study did not attempt to forecast LGD. Dermine and de Carvalho (2005) model LGD for loans to small and medium-sized firms in Portugal. They apply mortality analysis and include the annual GDP growth as an explanatory variable. However, they find that the GDP growth is not significant, and suggest that this may be due to the fact that the period of analysis, 1995–2005, did not include a significant recession. We may also note that their training sample size (374 defaults) was relatively small, and may not have been large enough for a significant relationship between the economy and LGD to be discovered. Querci (2005) provides an LGD model for loans to small businesses and individuals by an Italian bank. The author shows the importance of regional differences in LGD variation, but does not include time varying macroeconomic conditions. Figlewski, Frydman, and Liang (2007) model the effect of macroeconomic factors on corporate default, with a detailed study of numerous economic conditions including the level of unemployment, inflation, GDP and a production index. They found that many of these MVs were significant explanatory variables. Saurina and Trucharte (2007) model PD for retail mortgage portfolios in Spain. They show that the GDP growth rate is a significant cyclical variable in the regression and has a negative sign, as we would expect. That is, during downturns (low GDP growth), PD increases. However, they also include an interest rate variable, and although it has a positive sign and is significant, report that including interest rates does not improve the accuracy.

The novelties of our paper are that, unlike published work, we (1) consider forecasts of LGD for retail credit cards, (2) report the results of model comparisons, (3) include macroeconomic conditions in our models, and (4) do so using a very large sample across several different credit card products. In Section 2 we describe our modelling and performance assessment methods. In Section 3 we discuss the application and the macroeconomic data used. In Section 4 we provide model comparisons and test results, along with a description of an explanatory model with MVs. Finally, in Section 5 we provide some conclusions and discussion.

2. Method

We consider several models as combinations of different variables, modelling frameworks and data transformations.

2.1. Models

In general, for retail credit, there are five categories of circumstances that will affect the amount an individual repays on a defaulted loan and can be used to build models of LGD:

(1) individual details, some of which can be collected at the time of application, such as age, income, employment, housing status and address;
(2) account information at default: date or age of account at default and outstanding balance;
(3) changes in personal circumstances of an obligor over time;
(4) macroeconomic or business conditions at the date of default, or possibly with either a lag or lead on the date of default;
(5) operational decisions made by the bank, such as the level of risk they were willing to accept on the credit product and the process they used to follow up bad debts.

Of these, the richest source of explanatory variables we have is the information provided at the time of the application for credit, along with the credit bureau score collected by the bank at the time of application. These data fall into category (1). We also include category (2) data, account information at default. Including these data implies that the model is conditional on default. It is possible to build models unconditionally, but this is outside the scope of this paper. It is difficult for a lender to extract category (3) data. It cannot easily keep track of an individual's employment status, still less his or her personal difficulties, such as divorce or illness, that may lead him or her to be unable to repay the debt fully. It is possible to use account behaviour data or a behavioural score, but we do not do this in this study, since such information is not homogeneous within the data we have. It is understood that LGD is likely to be time dependent, varying over the business cycle (Schuermann, 2005), and we therefore include macroeconomic conditions (4). Including a bank's operational decisions (5) for each credit card product over time could also be fruitful; however, this information was not available for our study.

To answer Q4, we build and compare models with and without MVs. The question can be explored further by contrasting the results with models that also include interaction terms between AVs and MVs. To help answer Q3, we contrast our models with a simple model built with no variables. Within the context of OLS, this simple model effectively forecasts LGD as the mean value of the training data set. If the more complex models produce better forecasts than this simple model, then the variables we include provide useful information for LGD estimation. We therefore have four model structures based on including different explanatory variables:

- *Simple*: no covariates in the model.
- *AV*: account variables only.
- *AV&MV*: account and macroeconomic variables.
- *AV&MV with interactions*: also includes interaction terms between AVs and MVs. It is not feasible to include all interaction terms, so variable selection is used, as is described in Section 3.3.

We do not restrict our models to OLS, but consider three alternatives as well. Tobit and a decision tree model are considered, since they have a structure which is better suited to the bimodal nature of LGD. Least absolute value regression is also considered, since it may be that the absolute error is a more sensible criterion for estimating LGD than least square errors.

Since LGD has a truncated distribution, with a large number of cases at the extreme values 0 and 1, the Tobit model may be more suitable, since it takes bounds on a dependent variable y_i into account through truncation. The two-tailed Tobit model uses a latent variable y^* to model boundary cases, such that $y_i^* = \boldsymbol{\beta} \cdot \mathbf{x}_i + \varepsilon_i$, where $y_i = \min\left(1, \max\left(0, y_i^*\right)\right)$. Assuming that the distribution of the residuals conditional on $\mathbf{x}$ is normal, the following log-likelihood function is constructed for maximum likelihood estimation of $\boldsymbol{\beta}$ and variance of residuals σ^2:

$$\log(\boldsymbol{\beta}, \sigma) = \sum_{0<y_i<1} \log\left[\phi\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) \Big/ \sigma\right] + \sum_{y_i=0} \log\left[1 - \Phi\left(\frac{\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right] + \sum_{y_i=1} \log\left[\Phi\left(\frac{\mathbf{x}_i^T\boldsymbol{\beta} - 1}{\sigma}\right)\right],$$

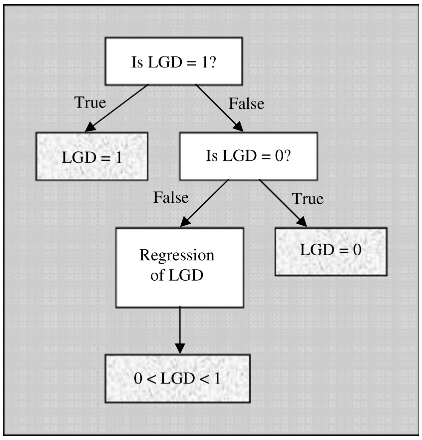

Fig. 1. Decision tree model. The first two binary decision boxes (white) are modelled by logistic regression, whilst the final decision box models LGD using OLS regression.

where ϕ and Φ are the probability and cumulative density functions for the standard normal distribution, respectively. This likelihood function is constructed by considering the probabilities of the dependent variable being between the boundaries, and also on each boundary separately (Greene, 1997, pp. 962–966).

The decision tree model uses two logistic regression sub-models to model the special cases for total loss and no loss, i.e., LGD $=$ 1 and 0 respectively, as binary classification problems. Then, if $0 < \text{LGD} < 1$, an OLS regression model is used. This decision tree model is illustrated in Fig. 1. We might expect this decision tree to be good at modelling LGD, due to the large number of boundary cases at 0 and 1, which allow us to naturally approach the problem as a hybrid of two classification problems and a regression problem. This approach is meaningful because there could be special conditions which would make a customer either pay back the full amount of the debt or pay back nothing, rather than just paying a portion. LGD is forecast for an account i as the expected value given the three sub-models, i.e. $(1 - p_{0i})\,(p_{1i} + (1 - p_{1i})\,L_i)$, where p_{0i} is the probability of LGD $=$ 0 for account i computed from the second estimated logistic regression model, p_{1i} is the probability of LGD $=$ 1 from the first estimated logistic regression model, and L_i is the estimate of LGD, assuming that the loss is fractional and computed from the regression model.

As is conventional in the literature, we model LGD in terms of the *recovery rate* (RR) rather than modelling LGD directly, where RR $= 1 -$ LGD. Our working definition of RR is

$$\text{RR} = \frac{\text{sum of repayments made over a period } t \text{ following default}}{\text{outstanding balance at date of default}}.$$

The choice of the recovery period is a business decision. We consider a recovery period of $t = 12$ months following default. Banks are often interested in longer periods or want to estimate the recovery at close of account, but, as we discover, a 12 month model can be used to produce estimates for longer recovery periods.

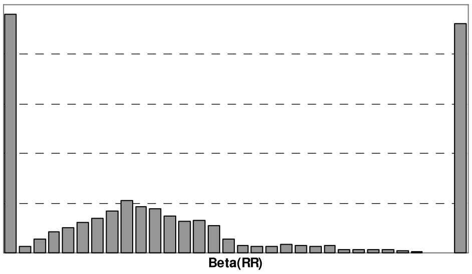

Fig. 2. Distribution of recovery rates following the application of a beta distribution transformation.[2]

It is possible to calculate LGD in alternative ways. For example, the time after default could be an event, such as account charge-off, rather than a fixed period; or the market value of the bad debt based on the sale of the exposure in the market could be taken into account along with repayments; or the administration costs of following up a default could be included in the LGD calculation. However, these alternatives are beyond the scope of this paper. The loss of interest payments could also be included in the definition of LGD, but this is not required by Basel II.

Since the distribution of RR is bimodal and U-shaped, we also consider modelling fractional logit, beta distribution and probit transformations of RR.

Fractional logit transformation:

$$T_{\mathrm{RR}} = \log(\mathrm{RR}) - \log(1 - \mathrm{RR}).$$

The fractional logit model is particularly attractive, since it deals specifically with response variables in the range 0–1, like RR, by transforming them into a larger range of values. It has been applied in several other econometric analyses (Papke & Wooldridge, 1996), and was applied in particular by Dermine and de Carvalho (2005) for modelling LGD for corporate loans.

Beta distribution transformation: $T_{RR} = \Phi^{-1}(\text{beta}(RR, \alpha, \beta, 0, 1))$, where Φ is the cumulative density function of the standard normal distribution and α, β are parameters estimated from training data using maximum likelihood estimation. The Beta distribution is particularly appealing because it is able to model bimodal variables with a U-shaped distribution over the interval 0–1. It is therefore particularly useful for RR and tends to transform RR into an approximately normal distribution. Fig. 2 shows a distribution of the Beta transformation of RR for the credit card default data we use. It illustrates how the transformation creates an approximately normal distribution between two extreme values. The Beta distribution has been used successfully in Moody's KMV Losscalc© software package for modelling RR (Gupton & Stein, 2005).

Probit transformation: $T_{RR} = \Phi^{-1}\left(\frac{|\{i: R_i \le RR\}|}{n}\right)$, where Φ is the cumulative density function of the standard normal distribution and $R_1, \ldots, R_n$ are observed RRs taken from the training data. This transformation uses a nonparametric approach to transform RR into a normal distribution based on the empirical distribution in the training data.

2.2. Model assessment

For OLS we report the *adjusted* R^2 of the model fit. There are two reasons for this. Firstly, the various models we consider are nested, so inevitably those with additional covariates will give an improved R^2 model fit. The adjusted R^2 compensates for the additional variables. Secondly, reporting the R^2 value is misleading, since it does not give a fair comparison between different studies with different sample sizes.

We report coefficient estimates for the OLS model with RR as the dependent variable. However, since RR is not normally distributed, the error terms may not be normally distributed either. Thus, conventional estimators of the standard errors may not be unbiased. Instead, we use a bootstrap to construct distributions for the coefficient estimates (Kennedy, 2003, Section 4.6). As Lam and Veall (2002) show, when OLS is used with non-normal, and in particular bimodal, distributed error terms, the bootstrap gives accurate estimates of the confidence intervals where the usual analytical method fails.

We report the relative effect of each MV within the model. The coefficient estimates for MVs are multiplied by their standard deviation over the training period to derive *standardized coefficient estimates*. They show the change in RR following a one standard deviation change in the covariate value, and are therefore comparable and give an indication of the relative importance of each MV within the model. This approach was taken by Figlewski et al. (2007) to study the effects of MVs on corporate default. Since our credit card data span the period 1999 to 2005, standard deviations for each MV are computed based on values within this period.

2.2.1. Hold-out test procedure for forecasts

To test the effectiveness of the LGD model for forecasting, we use a hold-out sample, testing on credit card default data that are independent and follow chronologically from the period of the training data used to build the models. This approach allows us to simulate the expected operational use of LGD models in retail credit when a financial institution may want to assess the LGD risk on a new batch of defaults based on the performance of past defaults. Specifically, we select cohorts of test data sets consisting only of accounts that default in a particular quarter. For each of these cohorts, we train using only default data available prior to that quarter. Since we need to measure LGD t months after default, we need to ensure that the dates of defaults in the training data are at least t months prior to the beginning of the test quarter. This test procedure is illustrated in Fig. 3. Thus, for example, if our test set was 2003Q3 and we are considering LGD after 12 months, our training data would consist of all cases that defaulted within the period 1999Q1 to 2002Q2. To estimate robust models with MVs, we should train over the whole business cycle, which is usually considered to be between 3 and 5 years. Thus, we consider using training data sets with a minimum of 3 years of defaults. This then gives us 10 quarters of test set data from 2003Q1 to 2005Q2. The results across these independent test sets then form a time series of forecast results.

The accuracy of forecasts relative to the observed true values is measured at the account level by the mean square error (MSE). However, we are also interested in how well the model is able to estimate the observed, or true, mean LGD or RR over a portfolio of accounts. Therefore, for each test set quarter we measure the difference between the forecast and observed mean RR across all test cases. If this difference is greater than zero, then the model is generally overestimating RR, whereas if it is less than zero, it is underestimating RR. The closer to zero the difference between the forecast and observed RR is, the better the estimate. The mean values of both the MSE and the absolute value of the difference between the forecast and observed mean RR (abs diff RR) across the several test quarters are reported in order to obtain an aggregate measure of performance.

Since RR must be between 0 and 1, we truncate all forecasts of RR to fall within that range, prior to measuring the performance for all cases and all models. Also, to generate comparable results, the performance is measured as the difference between the predicted and observed RR, regardless of which transformation of RR is modelled. Therefore, if a transformation of RR is used, the inverse transformation is applied to extract the predicted RR and the performance is measured based on this rather than the transformed value. This is reasonable, since ultimately the value we want to model is RR, and the transformation should be merely a means to that end.

2.2.2. Forecasting 24 month LGD using a 12 month LGD model

For these experiments, we will be assessing models of LGD after 12 months. However, financial institutions often follow bad debts over several years, and thus they are also interested in forecasting LGD over a longer period, say 24 or 48 months. For this reason, we also assess how well the 12 month model forecasts for longer periods relative to a 24 month model. If it does well, then this implies that a single LGD model may be used to forecast for any LGD period. A 12 month LGD model can be used to forecast the 24 month LGD by calibrating the 12 month forecasts to 24 month forecasts. A simple way to do this is to use OLS regression on the 24 month LGD training data to build a linear model of the 24 month RR with an intercept and the 12 month RR as the explanatory variable. This model is then used to convert 12 month to 24 month forecasts. We use the training and testing scheme set out above. For the 24 month LGD we use 6 quarters of test data, from 2003Q1 to 2004Q2.

3. Data

3.1. Application data

For this study we have available a data set consisting of over 55,000 credit card accounts in default over the period 1999 to 2005 for customers across the whole of the UK. Account holders are expected to make a minimum payment of the outstanding balance each month. We define default as a case where a credit card holder is recorded as having failed to make the minimum payments for three consecutive months or more where the time window over which

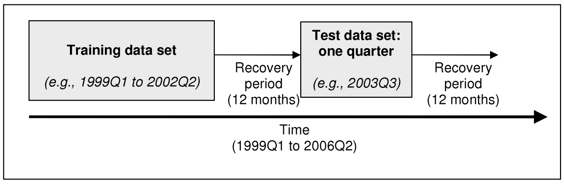

Fig. 3. Illustration of the hold out test procedure for one quarter of test data. The test data set holds all cases that default in the same financial quarter and the training data set comprises all cases that default and have a full recovery period prior to the test quarter.

the missing payments are recorded changes over time as payments are made. This is a typical definition of default for credit cards (Thomas, Edelman, & Crook, 2002, p. 123), and is in line with the default definition given by the Basel II Accord (Basel Committee on Banking Supervision, 2006). The data consist of four different credit card products which are a selection of those offered by a financial institution. As is typical for LGD, its distribution in our data set is between 0 and 1, and is approximately U-shaped. [1] The calculation of LGD should ideally include the administration costs involved in managing and implementing the collection procedure following account default. Unfortunately, this information was not available for the credit card data we used. The credit card data we use has many details extracted at the time of application, including the applicant's housing and employment status, age, income, total number of known credit cards and length of relationship with bank (time with bank). A credit bureau score is also provided and is from the same source, and therefore is homogeneous across products. Demographic information is provided to classify the area of residence of the credit card holder at the time of application. The demographic information was from the same source for all credit card products, and was coded into four broad categories: (1) council or poor housing, (2) rural, (3) suburban or wealthy area and (4) other. Additionally, information gathered at the time of default is also included in the data set. This consists of the balance outstanding at default and the age of the credit card account. We include the balance at default because there is strong evidence from past studies that it has an important effect (Dermine & de Carvalho, 2005; Grippa et al., 2005), and it makes sense to include it operationally, especially if it improves the forecasts of LGD. Table 2 gives the full list of the variables used. Some of the variables, namely time with bank, income and age, have a small percentage of missing values (less than 6% of accounts). For each of these variables, we code missing values as 0 and create a dummy variable to indicate missing values, in order to capture the mean value amongst accounts with missing values.

Several studies have found that PD and LGD are positively correlated (see Rösche & Scheule, 2006). We found this to be the case in our data too. Nevertheless, PD is not

[1] For reasons of commercial confidentiality, to protect our data supplier, we cannot reveal the exact distributions or values of LGD for our data, nor provide descriptive details of the other variables. This is usual in this area of research when large portfolios of live financial data are being studied. Nevertheless, since our focus is on reporting forecasting performances, this should not be a hindrance to scientific reporting.

included in our models since it is effectively represented by the inclusion of the application variables that are usually used to model it, along with macroeconomic variables that may explain the joint systematic risk to both PD and LGD (Altman et al., 2005).

Any single credit card portfolio is liable to have operational effects that will alter the overall risk over time for that specific product, such as changes in cut-offs on credit scores when accepting applications. This may lead to idiosyncratic links to economic conditions, and therefore poorer models when using MVs. By combining data across several products, the impact of these idiosyncratic effects will be reduced, and changes in risk over time are more likely to be linked to more objective effects such as the economy. In addition, combining several products in one data set will increase the size of the training set. These two factors should lead to stronger MV models, and we test this hypothesis by running experiments both for all products combined and for each product separately. When all products are included in the data set, a dummy variable is used to indicate which product the account belongs to, in order to model different levels of RR between products.

3.2. Macroeconomic variables

We consider three series of macroeconomic data for the UK which we believe to have a strong direct effect on the mean LGD for UK retail credit cards:

- Selected UK retail banks' base interest rates.
- UK unemployment level, measured as thousands of adults (16+) unemployed.
- UK earnings index (2000 = 100) for the whole economy, including bonuses as a ratio of the retail price index.

These are all available as monthly series from the UK Office for National Statistics. We use non-seasonally adjusted data for earnings, since we expect that seasonal changes in the economy may have some effect on abilities to repay. We would also have preferred to use non-seasonally adjusted data for the unemployment level, but unfortunately this was not available. GDP growth for the UK is a common indicator of economic conditions, but we have not included it since it is not available as monthly data, which is the granularity that lenders typically require, and therefore the granularity we require for our models. The MVs at the date of default are included for each case, but it may be that if there is a relationship between MVs and LGD then this is lagged or led. For example, in general, changes in interest rates may affect people's ability to pay several months later. We experimented with several different lag lengths and found a better performance for lags of 0 and 6 months. For this reason, we also consider LGD models with MVs lagging or leading by 6 months.

Each MV has a time trend: interest rates and unemployment level are generally falling over the period 1999–2005,

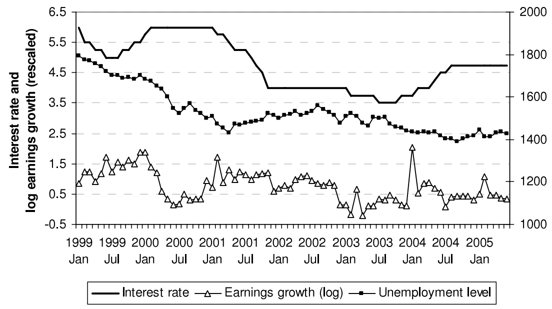

Fig. 4. UK macroeconomic data 1999–2005.

whilst real earnings are increasing steadily. Indeed, earnings generally increase exponentially with time, and therefore we include the growth in log earnings over 12 months to remove this obvious time trend. The three MV time series we use are shown in Fig. 4. We ensure that any model fit of MVs as explanatory variables is not simply because they follow a time trend that matches a trend in RR by explicitly including the date of default in the models. If a MV is a good explanatory variable simply because of a time trend, then the inclusion of the date of default should weaken its effect within the model. Including the date of default in the AV model also allows us to test whether any improvements in the forecasts are due simply to a time trend, rather than specifically to economic conditions.

Different effects on LGD over time could be captured by using dummy variables for cohorts at either the yearly or quarterly level. However, there are two reasons why we do not do this. Firstly, the freedom gained from using any time dummies could simply absorb the effect that we expect the MVs to explain. Secondly, although this is fine for explanatory models, it is not clear how such time dummies could be used for forecasting on a hold-out sample, since the dummy variables for the period of the hold-out cohort will necessarily have the value 0 for all cases in the training data, and thus will not have a coefficient estimate.

A high correlation between MVs is a potential problem, since this could lead to multicollinearity within the LGD model and therefore distort the parameter estimates. We can test for multicollinearity by measuring the variance inflation factor (VIF), given by $\left(1-R^2\right)^{-1}$, when each MV is regressed on all other model covariates (Kennedy, 2003). A high VIF indicates multicollinearity, and a VIF greater than 5 is an indication that there may be a problem.

3.3. Inclusion of interaction terms

Since there are many possible combinations of variables that could form interaction terms, the number included in the model is controlled using forward variable selection. All AVs are included in the model, but an iterative process is used to include MVs and interaction terms between AVs and MVs. At each step, all of the outstanding MVs and interaction terms not already in the model are added separately. A term that maximally increases the fit criterion is added to the model. The process is then repeated until no new interaction terms are found that

[2] The values of the axes are withheld for reasons of commercial confidentiality of the data provider.

improve the fit. There are several possible fit criteria that could be used, and it is common to use an F-test. However, since we are interested in forecasting, we use Akaike's information criterion (AIC) (Akaike, 1973). This has the advantage that it takes the parameter space of the model into account and discourages complex models with large numbers of variables. This in turn discourages over-fitting to the training data set. We approximate AIC by $n \ln(\text{MSE}) + 2p$, where n and p are the number of observations and the number of parameters in the model, respectively, and MSE is the mean square error of observations in the training data. We find that using the AIC produces better forecasts than using the standard F-test. A further discussion of variable selection methods and the use of the AIC for predictive models is given by Miller (1990). The variable selection procedure we use is further constrained so that, for each interaction term included, its constitutive terms are automatically included as well (Brambor, Clark, & Golder, 2005).

4. Results

Section 4.1 describes the forecast performances for a comparison of the different models. Section 4.2 then describes the best performing model for forecasts and its statistically significant explanatory variables. Models for longer recovery periods are considered in the third subsection.

4.1. Model comparisons

Table 1 shows the forecast results for different models. Focussing on the first section, which shows the results for a 12 month recovery period and for all products combined, it is clear that the standard OLS model with both AVs & MVs performs best for both measures of forecast performance. The more complex models, Tobit, decision tree and least absolute value regression, perform worse, as does using any of the transformations of RR. We may expect OLS to do well for the MSE measure, but we may also expect that one of the alternative models would perform better when estimating the mean RR over the portfolio. In particular, we may expect the least absolute value regression to be better, since it is a linear loss function. However, OLS produces the best estimates of the mean RR. This is a robust result which was obtained in many alternative experiments. The inclusion of interaction terms leads to slightly worse forecasts than the AV&MV model, so including interaction terms does not provide any benefit in estimating LGD. It is also notable that the simple model that effectively forecasts the mean RR from the training data set does well and outperforms many of the more complex models. Nevertheless, using both AVs and MVs leads to a considerable gain in performance over the simple model.

We also consider MVs with lags or leads of 6 months. We consider lags because it is possible that the effect of the economy on obligor behaviour may be delayed. We consider leads because this would give the values of MVs midway through the recovery period. Table 1 shows that using lags or leads of MVs produces worse forecasts than

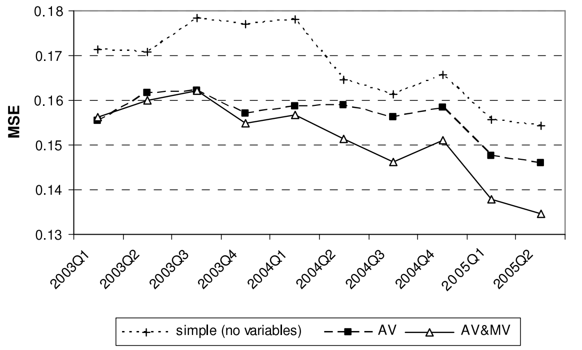

Fig. 5. MSE of the forecasts for each test data set from 2003Q1 to 2005Q2 for three OLS models with different variables.

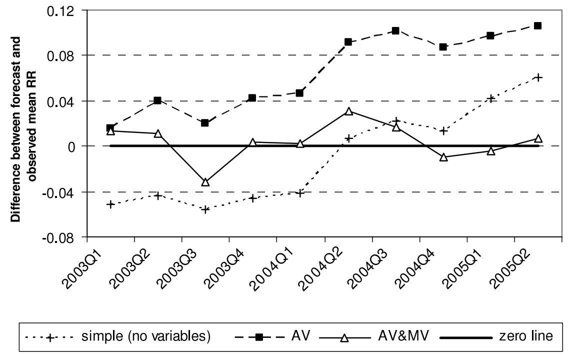

Fig. 6. Difference between the forecast and observed mean RR. These results relate to each test data set from 2003Q1 to 2005Q2 for three OLS models with different variables.

using MV values at the time of default, and is almost as bad as not including MVs at all.

The second section of Table 1 shows the results for separate products. It is clear that the AV&MV model does not consistently do well for separate products, and rarely does as well as when all products are combined. In many cases, the simple model with no explanatory variables performs best for forecasting the mean RR. This is due partly to operational peculiarities within each portfolio which may have a spurious link to macroeconomic movements, and also partly to the reduced training sample size. These results suggest that better LGD models can be built when data from different products are combined.

Figs. 5 and 6 show time series forecasting results for the first three models listed in Table 1 in more detail for each test quarter. They show that the AV&MV model performs consistently better over time. Fig. 5 shows that the MSE is lower for AV&MV than for either the simple or AV model, and improves over time, possibly as a result of having training data over a longer period of the business cycle for modelling the MV effects. Fig. 6 shows that overall the AV&MV model forecasts follow the mean RR much more closely than the other models, as is evidenced by the closeness to zero of the difference between the forecast and observed mean RR. In contrast, the AV model consistently over-estimates RR over time. In 2004Q2, the simple model achieves the best result, but this is purely

Table 1
Aggregate forecast results for different models.

Experiment					Result	
Recovery period	Product	RR transform	Model	Explanatory variables	MSE	Abs RR diff
12 month	All	None	OLS regression	None (simple)	0.168	0.0384
				AV	0.156	0.0645
				AV & MV	**0.151**	**0.0130**
				AV & MV & interaction terms	0.152	0.0169
				AV & MV lag 6	0.155	0.0529
				AV & MV lead 6	0.154	0.0445
			Tobit	AV & MV	0.163	0.0620
			Decision tree	AV & MV	0.221	0.1766
			Least absolute value regression	AV & MV	0.173	0.1144
		Beta	OLS regression	AV & MV	0.166	0.0748
		Logit	OLS regression	AV & MV	0.175	0.1086
		Probit	OLS regression	AV & MV	0.198	0.2041
		Beta	Tobit	AV & MV	0.188	0.1053
			Decision tree	AV & MV	0.192	0.0414
12 month	1	None	OLS regression	None (simple)	0.160	**0.0383**
				AV	0.159	0.0795
				AV & MV	**0.155**	0.0534
	2	None	OLS regression	None (simple)	0.175	**0.0391**
				AV	**0.163**	0.0732
				AV & MV	0.195	0.1361
	3	None	OLS regression	None (simple)	0.165	**0.0380**
				AV	0.155	0.0774
				AV & MV	**0.151**	0.0431
	4	None	OLS regression	None (simple)	0.156	0.0412
				AV	0.149	0.0771
				AV & MV	**0.142**	**0.0227**
24 month	All	None	OLS regression	AV & MV	0.194	0.0554
			OLS model built for 12 month recovery period	AV & MV	**0.183**	**0.0525**

Abs RR diff is the absolute difference between the observed and forecast mean RR over each test quarter. The results are given as mean values across the series of test quarters from 2003Q1 to 2005Q2 and 2004Q2 for the 12 and 24 month recovery periods respectively. The figures in bold indicate the best results for each recovery period and product group.

serendipitous, since its forecasts at that time are simply moving from underestimating to overestimating RR. These results show that including MVs is important for improving the LGD forecast results. It should be noted, however, that in our study, at least 3 years of training data are required. We found that using less than this led to unstable models, with MVs that occasionally extremely over- or underestimate LGD.

4.2. Explanatory model

Table 1 shows that the AV&MV model using OLS regression was the best performing forecaster, so we describe this model estimate in more detail. We report the model fit results for LGD models built using all data from 1999 to 2005. Including MVs in the LGD model improves the fit to the training data, and the MV coefficients are statistically significant. The AV model has an adjusted R^2 model fit of 0.105. When MVs are included, this increases to 0.110. When interaction terms between MVs and application variables are also included using forward selection, the adjusted R^2 is 0.111. This small increase indicates that adding interactions does not provide a noticeable improvement, reinforcing the results observed for forecasting. Table 2 shows coefficient estimates using OLS regression with the AV&MV model. Since the error residuals are non-normal, we have used bootstrapping to compute the statistical significance. The reported p-values are from a normal-based distribution imposed on bootstrap coefficient estimates. This is reasonable, since we found that the 95% confidence intervals for the normal-based distributions matched those for the empirical percentile distribution closely, never sharing less than 92% of the other's range.

Many of the model variables are statistically significant at the 0.01 level. In particular, housing status is important, with council and private tenants generally having a lower RR than home owners; a longer time of the customer being with the bank (time with bank) and a longer time of the individual holding the credit card prior to default (time on books at default) both tend to increase RR; individuals with higher incomes also tend to have higher RR values. All of these are indicators of *customer stability*, which we would expect to give lower risk. Higher credit bureau scores tend to give higher RR values, which again shows that individuals with low expected credit risks tend to pay

Table 2
Coefficent estimates for the OLS recovery rate model.

Variable	Coefficient estimate	Standard error	z	$P > \|z\|$
Intercept	1.43	0.162	8.83	0
Home status (D):				
Council tenant	−0.158	0.00544	−29.0	0
Private tenant	−0.146	0.00451	−32.4	0
Term time accommodation	−0.118	0.0410	−2.87	0.004
Other	−0.116	0.00702	−16.6	0
Lives with parents	−0.0643	0.00577	−11.1	0
Excluded category: Own home				
Time with bank	0.0812	0.00716	11.4	0
No data on time with bank (D)	0.0220	0.00774	2.84	0.004
Income (log)	0.0280	0.00770	3.63	0
Income unknown (D)	0.129	0.0328	3.94	0
Number of cards	−0.0153	0.00335	−4.56	0
Years at current address	0.00175	0.000238	7.36	0
Employment status (D):				
Home-maker	−0.0363	0.0109	−3.34	0.001
Unemployed	−0.0338	0.0230	−1.47	0.141
Retired	−0.00570	0.0102	−0.56	0.577
Part time	−0.00451	0.00904	−0.50	0.618
Self employed	−0.00339	0.00525	−0.65	0.519
Other	0.0230	0.00622	3.69	0
Student	0.0266	0.0120	2.21	0.027
Excluded category: Employed				
Age at default	−0.00348	0.000187	−18.6	0
Age unknown (D)	0.0586	0.0146	4.03	0
Credit bureau score	0.186	0.0117	15.9	0
Demographic group (D):				
Council or poor housing area	−0.0404	0.00385	−10.5	0
Rural area	0.0349	0.00933	3.74	0
Suburban or wealthy area	0.0355	0.00498	7.11	0
Excluded category: Other				
Credit card product (D): 1	0.00744	0.00523	1.42	0.155
2	0.00803	0.00660	1.22	0.223
3	0.0147	0.00505	2.90	0.004
Excluded category: 4				
Time on books at default (months)	0.000489	0.000133	3.68	0
Balance at default (log)	−0.165	0.00369	−44.8	0
Date of default	−0.0000161	0.000006	−2.74	0.006
Annual effect	*−0.00588*			
Macroeconomic variables:				
Bank interest rates	−0.0505	0.00296	−17.1	0
Standardized estimate	−0.0404			
Unemployment level	−0.000191	0.000048	−3.94	0
Standardized estimate	−0.0195			
Earnings growth (log)	0.839	0.484	1.73	0.083
Standardized estimate	+0.0032			

The training data set for the AV&MV model included all available cases from 1999 to 2005. The standard errors and p-values are computed via bootstrap estimation, with 10 000 repetitions. Values in *italics* show standardized estimates for MVs, and effect over one year for the coefficient on default date. Dummy variables taking the value 0 or 1, for false and true respectively, are indicated by (D).

back more of their bad debts. Also, the size of the initial balance at default has a negative effect on RR, as we would expect, since larger outstanding debts are more difficult to pay back. We note a positive correlation between the default balance and income, which suggests that people with higher incomes tend to build up larger balances on their credit card. Indeed, if we remove the default balance from the model, the sign on income becomes negative, since it becomes a surrogate for the missing default balance variable. However, when both income and default balance are in the model together, the sign on income becomes positive, as we would expect, since the availability of a higher income implies a greater capacity to repay the outstanding balance. Employment status has less impact, although we see that home-makers tend to have lower RR values. Demographic information was important, with those living in areas classified as council or poor housing tending to have lower RR values than those in rural, suburban or wealthy areas.

Table 2 shows that the coefficient estimates for MVs have the expected signs. That is, the parameter estimate for interest rates is negative, meaning that higher interest rates at the time of default tend to give lower RR values. Similarly, higher unemployment levels are also linked to lower RR values. However, higher earnings growth, year-on-year, leads to an increased RR, which suggests that earnings growth leads to better recoveries. Bank interest rates and the unemployment level are both statistically significant at the 0.01 level, although earnings growth is not statistically significant in the model. The coefficient estimate for the date of default is statistically significant but has a relatively small effect when its annual effect is compared with standardized estimates for the significant MV coefficients, as is shown in Table 2. This implies that the effect of the MVs is not due to a simple time trend. The bank interest rate clearly has the largest magnitude, with the level of unemployment having less than half the interest rate effect. In addition, we find that the VIF for any of the MVs when regressed on all of the other covariates in model (1) was always less than 2, which is sufficiently small that we would not expect the results to be affected by multicollinearity. Since including interaction terms did not improve either the forecasts or the model fit, we do not report interaction terms in the explanatory model.

4.3. *Forecasting 24 month LGD with a 12 month LGD model*

We test whether it is possible to use a model built for a 12 month recovery period to forecast for a 24 month period. Following the procedure given in Section 2.2.2, we get

$$\text{RR}\,(24\text{ months}) = 0.13 + 0.92 \times \text{RR}\,(12\text{ months}),$$

with adjusted $R^2 = 0.70$. This model always gives a higher estimated RR after 24 months than after 12 months. This is intuitively correct, since we would not normally expect RR for an individual to decrease over time. This model is used to compare the AV&MV model built on the 12 month period with one built on a 24 month period using the procedure described in Section 2.2.2. The results are shown in the third section of Table 1. The 12 month MV model outperforms the 24 month model for forecasts of the 24 month RR for both reported forecast measures. However, this is natural, since the recovery period buffer between the training and test data (see Fig. 2) implies that the 12 month model is built from more recent data. These results indicate that working with a 12 month LGD model is sufficient, since the same model can also be used to model longer periods. However, with further investigation, we expect that some hybrid model combining the 12 and 24 month trained LGD models would be the most effective.

5. Conclusion

In the introduction we posed four main research questions. We discuss our conclusions relating to each of these questions in turn.

Q1. *Which credit card application and default variables are the key drivers of retail LGD?*

Our experiments have shown that several application variables can be used to model LGD. In particular, Table 2 shows that home status, time with bank and the credit bureau score are strong explanatory variables. In addition, we found that income and the balance at default form a joint effect. The negative correlation between the default balance and RR matches the finding of Dermine and de Carvalho (2005) that the loan size is a significant explanatory variable for RR. We found that the default balance contributes to forecasts of LGD, since the forecast performance worsens when it is removed from the AV&MV model. Other variables at the time of default, such as the age of obligor and age of account, also influence LGD. Interestingly, age has a positive effect on LGD (negative on RR), and we found that the linear relationship between age and LGD remained even when age was replaced by several age categories using dummy variables. This is surprising, since we would normally expect the risk to decrease with maturity. That is, for PD models, the positive effect of age typically peaks in the mid-30s.

Q2. *What is the best modelling approach for the retail LGD?*

Despite trying several different combinations of variables, models such as Tobit and a decision tree, and various transformations of the dependent variable — all of which should *in theory* be good models for the bimodal LGD — the best forecast model in our experiments turned out to be simple OLS. Why this is so is unclear, although we conjecture that since LGD is difficult to model, with a poor model fit, this implies that regression forecasts tend to fall in a narrow range away from the boundary cases of 0 and 1, and therefore models dealing carefully with the boundary cases are superfluous in practice.

Q3. *How well do the models perform at forecasting LGD?*

The model fit is weak, with $R^2 = 0.11$, but such low values are typical of modelling LGD. On a sample of 1118 defaulted financial leases, De Laurentis and Riani (2005) report R^2 values of between 0.20 and 0.45 after outliers had deliberately been removed. Using a data set of 374 defaulted loans to small and medium size firms, Dermine and de Carvalho (2005) report a pseudo-R^2 value of 0.13 when considering a 12-month recovery period. The lower R^2 value we report is partly a consequence of the large sample size of our data. In contrast, if we restrict our sample size to just 500 randomly selected cases we get $R^2 = 0.20$. This is more typical of other studies; nevertheless, it is misleading, and we get an adjusted $R^2 = 0.13$, which is closer to the value for the full sample. We therefore report adjusted R^2 values in our main results and suggest their use for the comparison of LGD models across studies with different sample sizes.

For forecasting, the simple model — which effectively forecasts the mean LGD from the training data set — does very well and outperforms many of the complex models, as can be seen in Table 1. Nevertheless, we still see a modest improvement in MSE when model AV&MV is compared with the simple model with no variables. For financial institutions, even a small improvement in estimating the risk is welcome. When we turn to estimates of LGD across the portfolio, however, we can see from Fig. 6 that AV&MV is plainly the better forecaster when compared with the simple model. In this way, these models may prove particularly valuable for the estimation of risk at the portfolio level.

Our experiments focussed on modelling LGD for a recovery period of 12 months. However, financial institutions may be interested in longer periods. Nevertheless, we have shown that a 12 month LGD model can be used successfully to forecast the 24 month LGD.

Q4. *Does the inclusion of MVs lead to improved models of retail LGD?*

Our database spanned the period 1999 to 2005. Fig. 4 shows that this period covered a range of economic conditions in the UK, with interest rates generally decreasing and an overall reduction in unemployment. Earnings generally rose, although growth was higher at some times than at others. This period does have the disadvantage for our analysis that there were no major recessions or downturns to train from, and towards 2005 the UK economy was stable and fairly unremarkable. This point was also noted by Dermine and de Carvalho (2005) with regard to their study. We speculate that a very good macroeconomic model of LGD should have training data across the entire business cycle. Unfortunately, due to practical reasons of data availability, we were unable to provide this. Nevertheless, given this limitation, we still found the MV model to be effective. We show that adding bank interest rates and the level of unemployment into a LGD model as MVs yields a better model fit, and that these variables are statistically significant explanatory variables. In addition, including these MVs improves the forecasts, with generally better MSE values and estimates of the mean RR across the test quarters. Although the improvement in MSE is modest, Fig. 5 suggests that the AV&MV models improve relative to the duration or size of the training data set. Comparing the AV&MV model with the AV model in Fig. 6 shows a forecast of LGD that is clearly better at the portfolio level. We also report results using the model for separate products, where we found that the AV&MV model was less effective, suggesting that several products are required to build effective LGD models based on macroeconomic conditions.

We found that the inclusion of interaction terms between AVs and MVs did not generally improve the performance, and indeed, led to slightly worse results. The poor performance of the model with interaction terms confirms the comment by Gayler (2006) that the main effects are believed to be more stable than interactions for prediction in credit scoring. Nevertheless, we feel that there are likely to be some useful interaction effects between MVs and application terms; for example, those with a high outstanding debt, say a mortgage on a property, are more likely to be affected by changes in bank interest rates. The problem is to determine which of them are important prior to modelling. The automated forward selection process we use is clearly not sufficient for this task. Gayler (2006) recommends that prior expert knowledge be used to determine stable interactions. Thus, useful future work could involve incorporating expert credit advice into the model building stage, prior to automated modelling.

Finally, we ran a simple experiment using MV models for stress testing with hypothetical changes in interest rates. For example, we substituted the maximum and minimum interest rate values that occurred during our training period (6% and 3.5%) into the AV&MV model for the last of our test periods, 2005Q2. The forecast mean RR changed by −17% and +24% respectively. These results are plausible, in the sense that the forecast mean LGDs were in the range that we would expect, given historic data. Nevertheless, further work on the use of models with MVs for stress testing is needed. Firstly, a method for calibrating stress test estimates is needed. Secondly, the linearity of the models, along with the truncated distribution of LGD, implies that extreme values of MVs — necessary for stress testing — will have an overly extreme effect on forecasts of LGD. A logit transformation of RR would help to dampen extreme predictions. However, our experiments show that this is not the best model in terms of forecasts. Alternatively, a logit transformation of MVs, prior to their use in the model, might also mitigate the problem of extreme forecasts of RR. This is another area for further work.

Acknowledgements

We would like to thank our commercial partners for their assistance and comments in preparing this paper. This research was funded by UK EPSRC grant number EP/D505380/1, working as part of the Quantitative Financial Risk Management Centre.

References

Akaike, H. (1973). Information theory and an extension of the maximum principle. In *Proceedings of the 2nd international symposium on information theory* (pp. 267–281). Budapest: Akademia Kiado.
Altman, E. I., Resti, A., & Sironi, A. (2005). Loss given default: a review of the literature. In E. Altman, A. Resti, & A. Sironi (Eds.), *Recovery risk*. Risk Books.
Basel Committee on Banking Supervision (BCBS) (2006). *Basel II: International convergence of capital measurement and capital standards*. Basel.
Brambor, T., Clark, W. R., & Golder, M. (2005). Understanding interaction models: improving empirical analyses. *Political Analysis*, *14*, 63–82.
De Laurentis, G., & Riani, M. (2005). Estimating LGD in the leasing industry: empirical evidence from a multivariate model. In E. Altman, A. Resti, & A. Sirona (Eds.), *Recovery risk*. London: Risk Books.
Dermine, D., & de Carvalho, C. N. (2005). Bank loan losses-given-default: a case study. *Journal of Banking and Finance*, *30*(4), 1219–1243.
Figlewski, S., Frydman, H., & Liang, W. (2007). *Modeling the effect of macroeconomic factors on corporate default and credit rating transitions*. NYU Stern Finance working paper.
Gayler, R. (2006). Comment on classifier technology and the illusion of progress – Credit scoring, by D. Hand. *Statistical Science*, *21*(1), 19–23.
Greene, W. H. (1997). *Econometric analysis*. London: Prentice Hall International.
Grippa, P. S., Iannotti, F., & Leandri, F. (2005). Recovery rates in the banking industry: stylised facts emerging from the Italian experience. In E. Altman, A. Resti, & A. Sirona (Eds.), *Recovery risk*. London: Risk Books.
Gupton, G. M., & Stein, R. M. (2005). *Losscalc V2: dynamic prediction of LGD, modelling methodology*. Moody's KMV Company.
Kennedy, P. (2003). *A guide to econometrics* (5th ed.). Blackwell.
Lam, J.-P., & Veall, M. R. (2002). Bootstrap prediction intervals for single period regression forecasts. *International Journal of Forecasting*, *18*, 125–130.
Miller, A. J. (1990). *Subset selection in regression*. NY: Chapman & Hall.
Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 410 (K) plan participation rates. *Journal of Applied Econometrics*, *11*, 619–632.
Querci, F. (2005). Loss given default on a medium-sized Italian bank's loans: an empirical exercise. In *European financial management association 2005*.
Rösche, D., & Scheule, H. (2006). A multifactor approach for systematic default and recovery risk. In B. Engelmann, & R. Rauhmeier (Eds.), *The Basel II risk parameters*. Berlin: Springer.

Saurina, J., & Trucharte, C. (2007). *An assessment of Basel II procyclicality in mortgage portfolios*. Banco de España.
Schuermann, T. (2005). What do we know about loss given default? In E. Altman, A. Resti, & A. Sirona (Eds.), *Recovery risk*. London: Risk Books.
Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). Credit scoring and its applications. In *SIAM monographs on mathematical modeling and computation*, Philadelphia, USA: SIAM.

Tony Bellotti is lecturer in statistics in the Department of Mathematics at Imperial College London. He received his PhD in computational learning from Royal Holloway, University of London in 2006 and his main research interest is the application of statistical and machine learning models in credit risk analysis. He is an Honorary Fellow of the Credit Research Centre at the University of Edinburgh.

Jonathan Crook is the Director of the Credit Research Centre and Professor of Business Economics at the University of Edinburgh. He is the joint author/editor of five books, including: *Readings in credit scoring*, OUP 2004, *Credit scoring and its applications*, SIAM 2002, and *Credit scoring and credit control*, OUP 1992. His main research focus is on credit risk analysis, and he has published a large number of papers relating to credit in many different journals. In 1999 he was joint winner of the Goodeve Medal for the best paper in any Operational Research Society (UK) journal.