STATISTICAL MODELS FOR THE BASEL II INTERNAL RATINGS-BASED APPROACH TO MEASURING CREDIT RISK OF RETAIL PRODUCTS

by

Tze Leung Lai
Samuel Po-Shing Wong

Technical Report No. 2008-6
July 2008

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
STATISTICAL MODELS FOR THE BASEL II INTERNAL RATINGS-BASED APPROACH TO MEASURING CREDIT RISK OF RETAIL PRODUCTS

by

Tze Leung Lai
Department of Statistics
Stanford University

Samuel Po-Shing Wong
Department of Statistics
The Chinese University of Hong Kong

Technical Reports No. 2008-6
July 2008

This research was supported in part by grant from National Science Foundation, DMS - 0305749

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Statistical Models for the Basel II Internal Ratings-Based Approach to Measuring Credit Risk of Retail Products

TZE LEUNG LAI* AND SAMUEL PO-SHING WONG†

Abstract

The Basel II Accord is a financial risk management standard recently adopted by many financial institutions and regulators around the world. The general spirit of the accord is to develop a systematic approach to evaluating and controlling risks based on timely data and their analysis and interpretation. The interface between statistical modeling and the financial application is of pivotal importance in the development of the internal ratings-based (IRB) approach recommended by the Basel II Accord. This article reviews the IRB requirements and develops new empirical Bayes models for modeling probability of default and loss given default, which are the key ingredients in the IRB approach to credit risk analysis of retail exposures.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62P05; secondary 62J12, 62M05.

KEYWORDS AND PHRASES: probability of default, loss given default, empirical Bayes, Markov chain, generalized linear mixed models, credit scoring.

1. Introduction

The Basel Committee on Banking Supervision (BCBS) is an institution founded by the central bank governors of the G-10 nations (Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, the United Kingdom, and the United States) in 1974. Its current membership is composed of senior representatives of banking supervisory authorities and central banks not only from the G-10 countries but also from Luxembourg and Spain. BCBS meets four times a year, usually at the Bank for International Settlements (BIS) in Basel, and proposes supervisory standards and

*Research sponsored by NSF grant DMS-0305740.
†Research sponsored by RGC of Hong Kong.
guidelines in banking with the expectation that members’ and other nations’ banks and regulatory agencies will implement therein. These international standards are referred to as the Basel Accords. The Basel I Accord was proposed by BCBS in 1988 and was legally enforced in the G-10 countries in 1992. Basel I focused on the risk caused by loans or interest not being paid back to the banks by the obligors. Such risk is called credit risk.

In 2001, BCBS proposed a new accord, which is known as the Basel II Accord and was finalized in 2005, for more comprehensive risk evaluation. In Section 2 we give a brief review of the Basel II Accord, and in particular the IRB (internal ratings-based) approach and the decomposition of the expected credit loss \( E(L_e) \) for a loan:

\[
E[L_e] = EAD \times LGD \times PD,
\]

(1.1)

where \( PD \) is the probability of default, \( LGD \) is the expected loss given default, which is expressed as a rate between 0 and 1, and \( EAD \) is the exposure at default. There are many statistical issues concerning how \( PD \) and \( LGD \) should be estimated. In this paper we consider retail loans (e.g., mortgages, automobile loans and personal loans) and propose new statistical models to address these issues. Since the baseline data of these loans are typically credit scores, we give in Section 3 an overview of the statistical methods used in credit scoring. The obligor’s record of payments on the loan and the interest provides longitudinal data, which we combine with the baseline data to build a Markov chain model in Section 4 for estimating an obligor’s default probability. We use an empirical Bayes approach to make use of the data from other obligors while allowing for inter-subject variability. Moreover, the Bayesian model can readily incorporate macroeconomic variables and their forecasts that are important exogenous variables in estimating next period’s default probabilities. The empirical Bayes approach is also used in Section 5 to estimate a loan’s LGD for the next period. Some concluding remarks are given in Section 6.

2. Basel II and its IRB Requirements

We begin by reviewing how Basel I measured credit risk. Under Basel I, the assets of banks were classified into five categories, with corresponding risk weights 0, 10%, 20%, 50% and 100%. For example, G-10 government debt had zero weight, G-10 bank debt had weight 20%, and other debts (which include corporate debt and the debt of non-G-10 governments) had weight 100%. Basel I required banks to put aside at least 8% of the total
weighted credit exposure to guard against various financial risks. Thus the capital requirement of a particular debt under Basel I was $EAD \times RW \times 0.08$, where $EAD$ is the exposure at default of the debt and $RW$ is the risk weight.

Unlike Basel I which focuses on credit risk exposures, Basel II also requires banks to measure market risk and operational risk; see [1]. With all these risks being included, Basel II still maintains the 8% capital requirement on the total weighted risk exposures. For many commercial banks, credit risk, which is the focus of this paper, is the major component of the risk exposures. Basel II recommends each bank to measure its credit risk progressively, first by the standardized approach and then changing to the IRB approach. The standardized approach is very similar to the Basel I risk weighting scheme except that $RW$ can now be determined by using external ratings (Moody’s KMV, or Standard & Poor, or other rating agencies acceptable to the regulators). Basel II recommends the IRB approach to proceed in two stages, starting with the foundation IRB approach (FIRB) and then switching to the advanced IRB approach (AIRB). BCBS provides $EAD$ and $LGD$ for corporate loans to FIRB banks and requires their internal rating systems to compute $PD$ for corporate loans and $EAD$, $LGD$ and $PD$ for all retail loans. AIRB banks have to estimate $EAD$, $PD$ and $LGD$ for all loans and use them in the minimum capital requirement formulas of BCBS[2].

As an example, consider the FIRB formula $RW = k \times 12.5$ in BCBS [2, §§328-330] for the risk weight of retail loans, where 12.5 is the reciprocal of 8% and

$$k = \left( \Phi \left[ \frac{1}{\sqrt{1 - \rho}} \Phi^{-1}(PD) + \left\{ \sqrt{\frac{\rho}{1 - \rho}} \Phi^{-1}(0.999) \right\} \right] - PD \right) \times LGD, \quad (2.1)$$

in which $\Phi$ is the standard normal cumulative distribution function and $\rho$ is related to the pairwise correlation of the asset levels of obligors belonging to the same risk class. This formula is derived from Schonbucher’s one-factor model of the obligors’ asset values; see [3, 4, 5, 6, 7]. Let $V_i$ be the standardized asset level of the $i$th obligor in a given risk class and assume that

$$V_i = \sqrt{\rho}Z + \sqrt{1 - \rho}\varepsilon_i \quad (2.2)$$

where $\rho > 0$, $Z$ and $\varepsilon_i$ are i.i.d. standard normal for $i = 1, \ldots, M$. Default is assumed to occur when $V_i$ falls below a threshold $K$. Let $I_i = 1_{\{V_i < K\}}$ and note that $E(I_i)$ is the same for all obligors in the same risk class. Moreover, since $Z$ is the common hidden factor, $V_i$ and $V_j$ are correlated and so are $I_i$ and $I_j$ for $i \neq j$. For a large number $M$ of obligors, Schonbucher[3] has
shown that
\[
\lim_{M \to \infty} \Pr\left\{ \frac{\sum_{i=1}^{M} I_i}{M} \leq x \right\} = \Phi\left( \frac{\sqrt{1 - \rho} \Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}} \right),
\]
(2.3)
where \( p = E(I_i) \). Setting (2.3) equal to 0.999 gives \( x = x_{0.999} \), the 99.9th percentile of the limiting distribution in (2.3), which can be expressed as:
\[
x_{0.999} = \Phi\left[ \frac{\sqrt{\rho}}{\sqrt{1 - \rho}} \Phi^{-1}(0.999) + \frac{1}{\sqrt{1 - \rho}} \Phi^{-1}(p) \right].
\]
This is the first term inside the parentheses of (2.1). The second term inside the parentheses of (2.1) is related to the expected credit loss which has a minus sign attached because the interest on each loan should be able to cover such expected loss; see the discussion in [8, p.1461]. Although the above argument assumes the \( PD_i \) to be equal for the \( M \) obligors, Vasicek[5] has shown that if the magnitudes of \( EAD_i \) and \( LGD_i \) are similar for all \( M \) obligors and \( M \) is large, formula (2.1) still provides adequate capital requirement to cover the portfolio credit loss 99.9% of the time.

For retail loans, estimating \( \rho \) from some observed “asset-type” levels is difficult and impractical. This led BCBS[4, p.13] to “reverse engineer” the asset correlations from (i) economic capital figures from large international banks and (ii) historical loss data from supervisory databases of the G-10 countries. Based on these studies, [2] recommends choosing \( \rho = 0.15 \) for mortgage loans, \( \rho = 0.04 \) for revolving unsecured exposures to individuals up to 100,000 Euros, and
\[
\rho = 0.03 \times \frac{1 - \exp(-35 \times PD)}{1 - \exp(-35)} + 0.16 \times \left[ 1 - \frac{1 - \exp(-35 \times PD)}{1 - \exp(-35)} \right]
\]
for other retail exposures. In Section 6, we comment on these recommendations of BCBS.

3. Credit scoring

Credit scoring has been used in banking industry since the 1940s. Durand’s [9] seminal paper uses Fisher’s linear discriminant analysis to distinguish bad loans from the good ones. The basic idea is to assign scores to the loan applicants based upon their attributes such as gender, education level, income level, profession, marital status and credit history. Such scoring systems,
called scorecards, are usually developed by in-house credit risk departments and/or consulting firms.

The training sample for building a credit scorecard consists of \((x_i, I_i)\), \(1 \leq i \leq n\), from \(n\) loans, where \(x_i\) a \(d\)-dimensional vector of predictor variables (attributes) of \(i\)th loan, and \(I_i\) is a binary response variable defined by \(I_i = 1\) (or 0) if the \(i\)th loan defaults (or stays clean) within the last 12 or 18 or 24 months after at least one year since the loan was offered [10, p.153]. Regarding \((x_1, I_1), \ldots, (x_n, I_n)\) as a training sample of independent replicates of \((X, I)\), the statistical problem is to estimate \(P(I = 1 | X = x)\) and the methods commonly used in credit scoring include linear or quadratic discriminant analysis, logistic or probit regression, neural networks, classification and regression trees, and support vector machines. These methods are summarized below. They estimate \(P(I = 1 | X = x)\) by some function of \(\beta^T x\), in which \(\hat{\beta}\) is estimated from the training sample. The linear combination \(\beta^T x\) therefore provides the “score” of an obligor with attribute vector \(x\).

Linear discriminant analysis (LDA) assumes that there is a prior probability \(\pi\) of default, i.e., \(\Pr\{I = 1\} = \pi\), and that the conditional distribution of \(X\) given \(I = i\) \((i = 0, 1)\) is \(N(\mu_i, \Sigma)\). Then by Bayes’ theorem,

\[
\Pr\{I = 1 | X = x\} = \frac{\pi \exp\{-\frac{1}{2}d_1(x)\}}{\pi \exp\{-\frac{1}{2}d_1(x)\} + (1 - \pi) \exp\{-\frac{1}{2}d_0(x)\}} = \frac{1}{1 + e^{-s(x)}},
\]

where \(d_j(x) = (x - \mu_j)^T \Sigma^{-1} (x - \mu_j)\) is the Mahalanobis distance from \(x\) to \(\mu_j\) \((j = 0, 1)\) and \(s(x) = \beta^T x + \gamma\) with \(\beta = \Sigma^{-1}(\mu_1 - \mu_0)\) and \(\gamma = \log \frac{\pi}{1 - \pi} - \frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1 + \frac{1}{2} \mu_0^T \Sigma^{-1} \mu_0\), which can be estimated from the data. The assumption of equal variance matrices conditional on \(I = 1\) or on \(I = 0\) can be relaxed, resulting in quadratic discriminant analysis (QDA); see [11, pp.88-90].

Logistic regression assumes that \(\Pr\{I = 1 | X = x\} = \psi(\beta^T x + \gamma)\) with \(\psi(t) = 1/(1 + e^{-t})\), where \(\beta\) and \(\gamma\) are unknown parameters that can be estimated by maximum likelihood. Probit regression uses \(\psi = \Phi\), the standard normal distribution instead of the logistic function.

Letting \(Y = 1 - 2I\) so that default corresponds to \(Y = -1\), a support vector machine (SVM) considers the model \(\Pr\{I = 1 | X = x\} = \beta^T x + \gamma\) and estimates the unknown parameters \(\beta\) and \(\gamma\) by solving the quadratic programming problem of minimizing

\[
\|\beta^2\|^2/2 + \lambda \sum_{i=1}^n \xi_i \text{ subject to } \xi_i \geq 0, \ y_i(x_i^T \beta + \gamma) \geq 1 - \xi_i \text{ for all } 1 \leq i \leq n,
\]

(5)
where $\lambda > 0$ controls how much overlap in the $x$-space is allowed between the classes corresponding to $y_i = -1$ and $y_i = 1$; see [11, pp.372-375], where the separating hyperplane background (in the case $\lambda \to \infty$) of SVM classifiers and computational algorithms for minimizing (5).

Instead of a single score $\beta^T x$, a neural network with $J$ hidden units involves $J$ scores $\beta_j^T x$ and assumes that $P(I = 1|X = 1) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \psi(x^T \beta_j + \gamma_j)$, in which $\alpha_0$, $\alpha_j$, $\beta_j$ and $\gamma_j$ are unknown parameters that can be estimated by maximum likelihood, similar to logistic regression. We can use BIC to estimate $J$; see [11, p.206]. Another variant is to replace $x$ by a transformed variate vector $h(x) = (h_1(x), \ldots, h_m(x))^T$, as in the case of a classification and regression tree (CART) that assumes the model $P(I = 1|X = x) = \beta^T h(x)$ with $h_j(x) = 1_{x \in R_j}$, where $R_j$ is a rectangle in $\mathbb{R}^d$ that is determined from the training sample by using a minimization algorithm to choose the "splitting variables" and "split points"; see [11, pp.267-271]. Such transformations have also been used to achieve better class separation of SVMs, with $x_i$ in (5) replaced by $h(x_i)$; see [11, pp.377-379].

One of the most widely used credit scoring systems was developed by Fair Isaac & Company (FICO), which created credit scoring systems for U.S. banks and retail stores in the 1960s and 1970s and then expanded to meet the needs of other industries and to evaluate the credit of borrowers. It assesses credit reports and credit history to determine a score that ranges between 300 and 850, but it ignores salary and occupation and other characteristics such as race, gender and marital status that may bias a lender. In the 1980s the success of credit scoring in credit cards led banks to use scoring for different loan-related tasks. Application scoring systems are used to distinguish the “bad” from the “good” borrowers. A borrower is classified as “bad” if $P(I = 1|X = x)$ exceeds some cut-off value. After the loan is offered, the borrower’s payment record provides additional information, and “behavioral scoring” systems can be used for making decisions to manage existing accounts, such as what credit limits to set, whether to market new products to these clients and how to manage recovery if the account should turn bad; see [10, pp.161-163].

4. A Markov chain approach to modeling PD

4.1. Baseline and time-varying covariates

We first consider the data available and the statistical issues in building models to estimate the probability of default ($PD_i$) of the $i$th obligor in the next year, which we denoted by $t+1$ and use $t$ to denote the current year.
The $i$th obligor's credit score at the time of loan application summarizes the baseline information, and his/her payment history represents time-varying information. The Basel II Accord recommends using at least 5 years of past data [2, §463]. Because of the possibility of changes in the model parameters and in the bank itself over time, it is unwise to go back in time for substantially more than 5 years. In the sequel, we assume for definiteness that 5 years of past data are used. For the $i$th obligor, we use credit score $x_i$ as the baseline covariate and the time series $I_i(t), I_i(t - 1), \ldots$ as the outcome variable and time-varying covariates to build a model for estimating the default probabilities $PD_i = E[I_i(t + 1)]$, where $I_i(s)$ is 1 if the $i$th obligor has a default (i.e., overdue payment for longer than 90 days in the case of retail products) in year $s$, and 0 otherwise.

Although many statistical methods for modeling $P(I = 1|X = x)$ have been used in credit scoring, as reviewed in Section 3, there are two difficulties in applying them to the present problem of estimating $E[I_i(t + 1)]$. First, obligors may begin (or pay back) their loans any time during the previous 5-year period. Second, most of the obligors do not default, i.e., most $I_i(t)$ are 0, and therefore the well-known difficulty of using maximum likelihood to estimate the probabilities of rare events also applies to logistic regression, which is essentially maximum likelihood estimation of probabilities as a function of the covariates. Whereas this is not an important issue in loan application scoring, the goal of which is to identify the "bad" borrowers whose probability of default exceeds some cut-off value, more precision is required for the estimation of $PD_i$ in evaluating the overall capital requirement covering all these loans.

The first difficulty can be resolved by using a Markov chain model to accommodate the different information sets of obligors. To address the difficulty caused by the relative sparsity of defaults, we use a Bayesian approach to modify the maximum likelihood estimates of the transition probabilities of the Markov chain model, which also incorporates macroeconomic factors for forecasting default in the next period. Details are given in the next section.

4.2. A Markov chain model for $PD_i$

Based on the credit scores, the applicants are grouped into various classes called risk buckets, within which the obligors can be regarded as having the same risk (or more precisely, the same prior probability of default); see [12, p.311] for an example. According to BCBS [2, p.91], each bank has to use at least seven risk buckets for borrowers who have not defaulted and at least one for those who have defaulted previously at the time of their loan application.
Risk bucketing is tantamount to stratifying the obligors into classes so that the baseline covariate (credit score) of an individual obligor can be replaced by the class membership. Therefore, instead of considering the $i$th obligor whose credit score is $x_i$, we now use $(i,j)$ to label the $i$th obligor in the $j$th class, $j = 1, \ldots, J$. Using the past 5 years of data, the bank has the record of the obligor’s time series $I_{ij}(t), I_{ij}(t-1), \ldots$. However, as noted in Section 4.1, not all obligors have all 5 years of data, depending on the year $t_{ij}$ that the loan started. Let $C_j^{(s)}$ be the set of the $j$th-class obligors who have available credit histories in year $s$ with $s = t, t-1, \ldots, t-4$, where $t$ is the current year. We begin by considering a simple model for this time series, namely, a 2-state Markov chain $I_{ij}(s)$ (taking values 1 and 0) with initial state 0 and transition probabilities

$$P\{I_{ij}(s) = 1|I_{ij}(s-1) = 0\} = p_j,$$

$$P\{I_{ij}(s) = 1|I_{ij}(s-1) = 1\} = \tilde{p}_j,$$  

for $i \in C_j^{(s)} \cap C_j^{(s-1)}$. The likelihood function of the Markov chain model (6)-(7) can be expressed as

$$\prod_{j=1}^{J} \{p_j N_{01,j}^{N_{00,j}} \} \{\tilde{p}_j N_{11,j}^{N_{10,j}} \}.$$

where $N_{ii',j} = \sum_{s=t-4}^{t} N_{ii',j}(s)$ for $i, i' \in \{0, 1\}$ and

$$N_{00,j}(s) = \sum_{i \in C_j^{(s)} \cap C_j^{(s-1)}} \{1 - I_{ij}(s-1)\} \{1 - I_{ij}(s)\},$$

$$N_{01,j}(s) = \sum_{i \in C_j^{(s)} \cap C_j^{(s-1)}} \{1 - I_{ij}(s-1)\} I_{ij}(s),$$

$$N_{10,j}(s) = \sum_{i \in C_j^{(s)} \cap C_j^{(s-1)}} I_{ij}(s-1) \{1 - I_{ij}(s-1)\},$$

$$N_{11,j}(s) = \sum_{i \in C_j^{(s)} \cap C_j^{(s-1)}} I_{ij}(s) I_{ij}(s-1).$$

The likelihood function (8) has the same form as that based on independent Binomial($n_j, p_j$) and Binomial($\bar{n}_j, \bar{p}_j$) observations $N_{01,j}$ and $N_{01,j}$, $1 \leq j \leq J$, where $n_j = N_{00,j} + N_{01,j}$ and $\bar{n}_j = N_{10,j} + N_{11,j}$. When $N_{01,j}$ is small (in particular, when $N_{01,j} = 0$), the maximum likelihood estimate
MLE $\hat{p}_j = \frac{N_{01,j}}{N_{j}}$ is an unreliable estimate of $p_j$. A commonly used method to deal with such “data-poor” situations is to introduce a Beta($a, b$) prior on $p_j$ and a Beta($\tilde{a}, \tilde{b}$) prior on $\hat{p}_j$. These are conjugate families and the posterior distributions of $p_j$ and $\hat{p}_j$ are also beta; in particular, the posterior distribution of $p_j$ is Beta($a + N_{01,j}, b + N_{00,j}$). When $n_j$ is large relative to $a + b$ and $N_{01,j}$ is substantially larger than $a$, the posterior distribution is concentrated at $\hat{p}_j$ and there is little difference between the MLE and the Bayes estimate $(N_{01,j} + a)/(n_j + a + b)$. To choose the “best fitting” Bayesian model in the sense of Kullback-Leibler divergence, we determine $\theta := (a, b)$ by maximizing the likelihood

$$\prod_{j=1}^{J} \left\{ \int p^{N_{01,j}}(1-p)^{N_{00,j}} g_{\theta}(p) dp \right\} \left\{ \int \tilde{p}^{N_{11,j}}(1-\tilde{p})^{N_{10,j}} g_{\theta}(\tilde{p}) d\tilde{p} \right\},$$

in which $g_{\theta}$ is the density function of the Beta distribution. This is similar to the approach of Breslow and Clayton [13] that will be discussed below.

The Markov chain (6)–(7) assumes stationary transition probabilities, which is reasonable when the economic environment is relatively stable in the past 5 years and is expected to continue in the coming year. However, if major shifts in macroeconomic variables like unemployment rate or GDP have occurred in the past 5 years or is expected to occur in the coming year, the transition probabilities of the Markov chain should accommodate these changes over calendar time, instead of being treated as stationary. Incorporating macroeconomic variables in a factor model yields the following modification of (6)–(7) for $i \in C_j^{(s)} \cap C_j^{(s-1)}$:

$$\logit(P(I_{ij}(s) = 1 | I_{ij}(s-1) = 0)) = \lambda_j + \alpha^T f(s),$$  \hspace{1cm} (10)

$$\logit(P(I_{ij}(s) = 1 | I_{ij}(s-1) = 1)) = \tilde{\lambda}_j + \beta^T f(s),$$  \hspace{1cm} (11)

where $\logit(p) = \log(p/(1-p))$ and $f(s)$ is a low-dimensional vector of macroeconomic factors in year $s$. A standard approach to estimating the parameters $\lambda_1, \ldots, \lambda_J, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_J$, $\alpha$ and $\beta$ is to use logistic regression, which is essentially maximum likelihood. Because most of the obligors do not default, i.e., most $I_{ij}(s)$ are equal to 0, it is desirable to use a Bayesian approach that puts a prior distribution on $\lambda_j$ with density function $g_{\theta}$ which has a parameter vector $\theta$, and a prior distribution on $\tilde{\lambda}_j$ with density function $g_{\tilde{\theta}}$, as we have done earlier for the simple model (6)–(7). This is tantamount to regarding $\lambda_j$ and $\tilde{\lambda}_j$ as “random effects” (instead of as fixed parameters).
in the generalized linear mixed models introduced by Breslow and Clayton [13], who propose to estimate $\theta, \hat{\theta}, \alpha$ and $\beta$ by maximizing the likelihood

\[
L = \prod_{j=1}^{J} \left\{ \int \left( \prod_{s=t-4}^{t} \left[ 1 - \psi(\lambda + \alpha^T f(s)) \right] \right)^{N_{00,j}(s)} \left[ \psi(\lambda + \alpha^T f(s)) \right]^{N_{01,j}(s)} g_0(\lambda) d\lambda \right\} \times \int \left( \prod_{s=t-4}^{t} \left[ 1 - \psi(\hat{\lambda} + \beta^T f(s)) \right] \right)^{N_{10,j}(s)} \left[ \psi(\hat{\lambda} + \beta^T f(s)) \right]^{N_{11,j}(s)} g_0(\hat{\lambda}) d\hat{\lambda} \right\},
\]

(12)

where $\psi(u) = 1/(1 + e^{-u})$ and we use the same notation as in (9).

Both (8) and (12) involves terms of the form $F(\alpha, \theta) = \int H_\alpha(\lambda) g_\theta(\lambda) d\lambda$. Following [13], we can evaluate such integrals by using the Laplace approximation

\[
F(\alpha, \theta) \approx H_\alpha(\hat{\lambda}) g_\theta(\hat{\lambda}) \sqrt{2\pi/V}
\]

(13)

where $\hat{\lambda} = \text{arg max}_\lambda H_\alpha(\lambda) g_\theta(\lambda)$ and $V = \left[ \frac{d^2}{d\lambda^2} \log(H_\alpha(\lambda) g_\theta(\lambda)) \right]_{\lambda=\hat{\lambda}}$. The accuracy of Laplace's approximation depends on the magnitude of $V$; see [14]. If $V$ is sufficiently large, (13) provides good approximation. When $V$ is not large, we can compute the integral by Gaussian quadrature:

\[
F(\alpha, \theta) = \int H_\alpha(\lambda) g_\theta(\lambda) d\lambda = \int \exp\{-\lambda^2/2\} \exp\{\lambda^2/2\} H_\alpha(\lambda) g_\theta(\lambda) d\lambda
\]

\[
= \int \exp\{-\lambda^2/2\} h(\lambda) d\lambda \approx \sqrt{2} \sum_{k=1}^{K} a_k h(\sqrt{2} s_k),
\]

(14)

where $h(\lambda) = \exp\{\lambda^2/2\} H_\alpha(\lambda) g_\theta(\lambda)$, $K$ is a positive integer and $s_k, a_k$ $(1 \leq k \leq K)$ are constants such that the approximation is exact if $h$ is a polynomial of degree less than $2K$. Note that Monte Carlo is used in [14] because a multivariate integral is involved. Using (13) or (14) to evaluate the integrals in (12), we propose following procedure to evaluate the MLE of $\Theta = (\alpha, \theta, \beta, \hat{\theta})$. Let $\psi = 1 - \psi$.

**Step 1.** Set $r = 0$ and $\Theta^{(0)}$ as a random starting value.

**Step 2.** Set $\Theta = \Theta^{(r)}$ and compute for $j = 1, \ldots, J$,

\[
\hat{\lambda}_j = \text{arg max}_\lambda \left( \prod_{s=t-4}^{t} \left[ \psi(\lambda + \alpha^T f(s)) \right] \right)^{N_{00,j}(s)} \left[ \psi(\lambda + \alpha^T f(s)) \right]^{N_{01,j}(s)} g_0(\lambda),
\]

where $\psi(u) = 1/(1 + e^{-u})$ and we use the same notation as in (9).
\[ \hat{\lambda}_j = \arg \max_{\lambda} \left( \prod_{s=t-4}^{t} \left[ \psi(\lambda + \beta^T f(s)) \right]^{N_{10,j}(s)} \left[ \psi(\lambda + \alpha^T f(s)) \right]^{N_{11,j}(s)} \right) g_{\theta}(\lambda). \]

**Step 3.** For each \( j \), compute

\[ V_j = - \left[ \frac{d^2}{d\lambda^2} \log \left( \prod_{s=t-4}^{t} \left[ \psi(\lambda + \alpha^T f(s)) \right]^{N_{00,j}(s)} \left[ \psi(\lambda + \alpha^T f(s)) \right]^{N_{01,j}(s)} g_{\theta}(\lambda) \right) \right]_{\lambda = \hat{\lambda}_j}, \]

and

\[ \bar{V}_j = - \left[ \frac{d^2}{d\lambda^2} \log \left( \prod_{s=t-4}^{t} \left[ \psi(\lambda + \beta^T f(s)) \right]^{N_{10,j}(s)} \left[ \psi(\lambda + \beta^T f(s)) \right]^{N_{11,j}(s)} g_{\theta}(\lambda) \right) \right]_{\lambda = \hat{\lambda}_j}. \]

If \( V_j > c \), use Laplace's approximation (13) to evaluate the first integral in (12), which will be denoted by \( I_j(\alpha, \theta) \). Otherwise, use (14) to evaluate \( I_j(\beta, \bar{\theta}) \). Similarly, by comparing \( V_j \) with the predetermined threshold \( c \), we can evaluate the second integral of (12), denoted by \( \bar{I}_j(\beta, \bar{\theta}) \).

**Step 4.** Compute \( \hat{\Theta}^{(r+1)} = \arg \max_{\Theta} \prod_{j=1}^{J} I_j(\alpha, \theta) \bar{I}_j(\beta, \bar{\theta}) \).

**Step 5.** If \( \left\| \hat{\Theta}^{(r+1)} - \hat{\Theta}^{(r)} \right\| < \text{tolerance} \), stop and output \( \hat{\Theta}^{(r+1)} \) as the MLE upon stopping. Otherwise, update \( r \) by \( r + 1 \), \( \hat{\Theta}^{(r)} \) by \( \hat{\Theta}^{(r+1)} \), and go to Step 2.

Note that the preceding algorithm not only computes the MLE of the parameters \( \alpha, \beta, \theta \) and \( \bar{\theta} \) of the Bayesian Markov chain model, but it also provides estimates of the random effects \( \lambda_j \) and \( \hat{\lambda}_j \) in (10) and (11) as posterior modes in Step 2 when the algorithm stops. We use this algorithm, with \( c = 5 \) and tolerance = 0.0001, in the simulation study below, following the suggestion in [14]. A maximum number of 1000 iterations is imposed in Step 5 of the algorithm.

### 4.3. A simulation study

We study the performance of the parameter estimates in the following model on the probabilities of default of 22,210 obligors. There are \( J = 5 \) risk buckets, with the following initial bucket sizes (in year 0) and the number of obligors who default in year 0 given in parentheses:

- 10000(10), 5000(20), 3000(30), 3000(50), 1000(100)
The first number (from left to right) corresponds to the lowest risk bucket, labeled \( j = 1 \), and the last one corresponds to the highest risk bucket, labeled \( j = 5 \). This study involves a 6-year period, with the data from the first 5 years being used as a training sample to estimate the model parameters and the probability of default in year 6.

Inspired by [15, p.282] and data from the second author’s consulting work with a bank in Hong Kong, we choose the baseline default probabilities \( p_j \) for those obligors who have not defaulted in the previous year, to belong to the following intervals for \( j = 1, \ldots, 5 \):

\[
[10^{-6}, 1.7 \times 10^{-4}], \quad [4 \times 10^{-4}, 8.9 \times 10^{-4}], \quad [1.86 \times 10^{-3}, 5.7 \times 10^{-3}], \\
[0.01, 0.027], \quad [0.05, 0.11]
\]

in which the first interval corresponds to \( j = 1 \) (lowest risk bucket), and the last to \( j = 5 \) (highest risk bucket). The baseline default probabilities \( \tilde{p}_j \) for those obligors who have defaulted in the previous year to belong to the following intervals for \( j = 1, \ldots, 5 \) (from left to right):

\[
[0.001, 0.01], [0.05, 0.1], [0.2, 0.3], [0.35, 0.4], [0.45, 0.5]
\]

Besides the baseline default propensities of the obligors, we also incorporate changes in a key macroeconomic variable that affects default during the period by assuming that the unemployment rates during the 6-year period are

\[
f(1) = 9\%, f(2) = 12\%, f(3) = 10\%, f(4) = 8\%, f(5) = 7\%, f(6) = 6%,
\]

in which \( f(6) \) for the future year 6 should be regarded as the unemployment rate forecast for year 6, while \( f(1), \ldots, f(5) \) are the actual unemployment rates.

With these choices of \( f(s) \), \( p_j \) and \( \tilde{p}_j \), we generate data from the model that has the form of

\[
P\{I_{ij}(s) = 1|I_{ij}(s-1) = 0\} = \frac{p_j}{p_j + (1 - p_j)e^{-3f(s)}}, \tag{18}
\]

\[
P\{I_{ij}(s) = 1|I_{ij}(s-1) = 1\} = \frac{\tilde{p}_j}{\tilde{p}_j + (1 - \tilde{p}_j)e^{-40f(s)}}, \tag{19}
\]

in which \( p_j \) and \( \tilde{p}_j \) are chosen from the intervals in (15) and (16), respectively. Note that (18) can be rewritten as

\[
\text{logit}(P\{I_{ij}(s) = 1|I_{ij}(s-1) = 0\}) = \text{logit}(p_j) + 3f(s),
\]
and (19) can also be expressed similarly. Thus, \( \lambda_j \) in (10) corresponds to 
\( \log(\frac{p_j}{\lambda_j}) \) and \( \bar{\lambda}_j \) in (11) corresponds to \( \log(\frac{\tilde{p}_j}{\lambda_j}) \). On the other hand, the working model (10)-(11) assumes the \( \lambda_j \) and \( \bar{\lambda}_j \) to be independent normal and therefore may be near each other with positive probability, whereas in the actual model \( p_1 < \cdots < p_5 \) and \( \tilde{p}_1 < \cdots < \tilde{p}_5 \), with clear separations between the adjacent intervals to show that the buckets indeed have different risks. After fitting the working model (10)-(11) to 5 years of data simulated from the actual model, we use it to calculate the predicted default probabilities \( \hat{p}_{ij}(6) \) of the obligor \((i, j)\) in year 6.

A commonly used approach to PD modeling is logistic regression, “pooling many obligors across many years” [15, p.273]. This yields \( \hat{p}_{ij}(6) \), which is obtained by fitting the logistic regression model with time-varying covariates \( I_{ij}(s-1) \) and \( f(s) \):

\[
\logit(p_{ij}(s)) = \lambda_j + \gamma I_{ij}(s-1) + \xi f(s). \tag{20}
\]

The model (20) is a special case of “point-in-time” (PIT) regression models for PD with explanatory variables that track the state of the credit cycle [15, p.273]. As pointed out in Section 4.1, different obligors may have different information sets because of drop-in and drop-out during the 5-year training period. To avoid such difficulties in fitting (20) to compare the predicted \( \hat{p}_{ij}(6) \) with \( \hat{p}_{ij}(6) \), we assume that no obligor drops out or drops in during the 6-year period. Besides \( \hat{p}_{ij}(6) \) and \( \hat{p}_{ij}(6) \), we also consider \( p_{ij}^{EST}(6) \) obtained by fitting logistic regression to (18) and (19) with \( \lambda_j = \logit(p_j) \) and \( \bar{\lambda}_j = \logit(\tilde{p}_j) \), which is tantamount to treating \( \lambda_j \) and \( \bar{\lambda}_j \) as fixed parameters (estimated by MLE) rather than random effects (estimated by Bayesian methods).

Our simulation study first considers the following choice of \( p_j \) and \( \tilde{p}_j \) in (15) and (16):

\[
\begin{align*}
p_1 &= 1.4 \times 10^{-5}, \quad p_2 = 4.5 \times 10^{-4}, \quad p_3 = 3.5 \times 10^{-3}, \quad p_4 = 0.027, \quad p_5 = 0.058, \\
\tilde{p}_1 &= 0.008, \quad \tilde{p}_2 = 0.0995, \quad \tilde{p}_3 = 0.26, \quad \tilde{p}_4 = 0.38, \quad \tilde{p}_5 = 0.466.
\end{align*}
\]

We use two measures to evaluate the performance of the estimate \( p_{ij}^{EST}(6) \). The first measures the “divergence” from the actual probability \( p_{ij}(6) \) and is defined by the Kullback-Leibler loss function

\[
D = \sum_{j=1}^{5} \sum_{i=1}^{n_j} [p_{ij}(6) \log p_{ij}(6) \frac{p_{ij}(6)}{p_{ij}^{EST}(6)} + (1 - p_{ij}(6)) \log \frac{1 - p_{ij}(6)}{1 - p_{ij}^{EST}(6)}]. \tag{21}
\]
The second is the $L_1$-loss

$$
L = \sum_{j=1}^{5} \sum_{i=1}^{n_j} |p_{ij}(6) - p_{ij}^{EST}(6)|.
$$

(22)

Figures 1 and 2 give the box-plots of $D$ and $L$ for the estimators $\hat{p}_{ij}(6), p_{ij}^{s}(6)$ and $\tilde{p}_{ij}(6)$ based on 50 simulations from the above model. The box-plot of $D$ for $p_{ij}^{s}(6)$ only involves 32 simulated data sets as the remaining 18 simulations have $p_{i1}^{s}(6) = 0$ (for the lowest risk bucket $j = 1$) and therefore $D = \infty$. The means of the $L_1$-loss for $\hat{p}_{ij}(6), p_{ij}^{s}(6)$ and $\tilde{p}_{ij}(6)$ are 22.45 (±1.19), 23.31 (±1.22) and 99.78 (±0.65), respectively, in which the standard errors are given in parentheses.

Figures 1 and 2 about here

We next generate 50 simulated samples in which $p_j$ and $\tilde{p}_j$ are sampled from the uniform distributions over the respective intervals in (15) and (16). Figures 3 and 4 give the box-plots of $D$ and $L$ for the estimators $\hat{p}_{ij}(6), p_{ij}^{s}(6)$ and $\tilde{p}_{ij}(6)$ based on these 50 simulated samples. The box-plot $D$ of $p_{ij}^{s}(6)$ involves 46 of these simulations, with the other four giving unacceptably large values of $D$. The means of the $L_1$-loss for $\hat{p}_{ij}(6), p_{ij}^{s}(6)$ and $\tilde{p}_{ij}(6)$ are 22.37 (±1.17), 23.60 (±1.25) and 89.66 (±1.63), respectively, in which the standard errors are given in parentheses. The preceding simulation results show that the Bayesian approach gives somewhat smaller $L_1$-loss in estimating $p_{ij}(6)$ than the likelihood approach and avoids difficulties of the latter when there are relatively few observed defaults. Moreover, using the Markov chain model (10)–(11) gives substantially better estimates of $p_{ij}(6)$ than fitting the logistic regression model (20) with time-varying covariates as in the standard PIT approach.

Figures 3 and 4 about here

4.4. Embedding $PD_i$ in a more informative Markov chain

Because the Markovian structure only involves one-step transition probabilities, the above Markov chain models are able to circumvent the difficulty that different obligors have different information sets which involve from less than one to more than five years of past data. However, the Markovian structure imposes the restriction that the conditional distribution of $I_{ij}(s)$ given $I_{ij}(s-1), I_{ij}(s-2), ...$, can only depend on $I_{ij}(s-1)$. Although one can remove this restriction by using a $K$th-order Markov chain as in [8, p.1451],
one has to further divide the obligors according to their information sets into sub-classes for which (a) only the baseline covariate is available, or (b) only $I_{ij}(s-1)$ is available as the time-varying covariate, or (c) $I_{ij}(s-1)$ and $I_{ij}(s-2)$ are available as time-varying covariates, etc., so that different $K$’s can be used for different sub-classes.

Instead of using a $K$th-order Markov chain for $I_{ij}(s)$, we propose to augment the 0-1 state encoding whether the loan has been overdue for more than 90 days in that year, as in (6)–(7) or (10)–(11), by incorporating how long the loan payment has been overdue in year $s$. Specifically, letting $\tau_{ij}(s)$ denote the duration of the longest overdue payment of obligor $(i,j)$ in year $s$, we divide the possible values of $\tau_{ij}(s)$ into $K$ groups, e.g. (i) less than 3 months (corresponding to $I_{ij}(s) = 0$), (ii) 3–12 months, (iii) 12–24 months and (iv) over 24 months. Letting $X_{ij}(s) = k$ if $\tau_{ij}(s)$ belongs to the $k$th group $(0 \leq k < K)$, we model $X_{ij}(s)$ as a finite-state Markov chain with $K$ states so that state 0 corresponds to no default (i.e., $I_{ij}(s) = 0$). This is tantamount to subdividing \{\{I_{ij}(s) = 1\}\} into $K-1$ states, representing the “severity” of the delinquency (“default”). It is straightforward to extend (10)–(11) for the 2-state case to the $K$-state model that has more parameters. Choosing the states appropriately in the $K$-state model is important to keep $K$ reasonably small (and the number of parameters manageable) which incorporating the severity of the previous delinquency as a predictor of $PD$ in the next period.

5. A generalized linear mixed model for $LGD$

5.1. Regression model in LossCalc

A commonly used method to estimate LGD is to combine historical averages with expert opinions in a look-up table; see [16]. A shortcoming of this method is that it does not adapt to changing macroeconomic conditions. More sophisticated statistical modeling is needed to better predict LGD for the next period. A difficulty with modeling LGD lies in the complexity of the recovery process. For corporate loans, Moody’s KMV has developed a software package LossCalc to calculate the LGD. After a review of how LossCalc uses recovery data of corporate loans to model LGD, we introduce new models for the LGD of retail products.

According to [16], LossCalc is built from a global database of 3,026 recovery observations that include defaulted corporate loans, bonds and preferred stock from 1981 to 2004. This training sample is used to construct a regression model for the recovery rate (RR) that is implied by the market value of
the loan (bid-side market quote) one month after default; 1−RR is called the economic LGD. The regression model can be used to predict the economic LGD of a defaulted loan and is based on the following simplification of the actual recovery process. The loans are usually most liquid 15 to 60 days after default and the prices reflect how much the market anticipates recoveries. Approximately 50% of the defaults are expected to be charged off the accounting books some time (with an average of 1.75 years) after default, with ultimate court-ordered resolution payments in the form of cash, assets, new debt extensions, etc. Because of the complexity of the recovery process and its duration, which ranges from 1.25 to 5 years, it is prohibitively difficult to predict such accounting LGD.

The regression model in LossCalc is a linear regression model of a transformed recovery rate \( y \) on nine predictors chosen from the following five groups of variables:

(i) collateral, which includes cash, assets, property plant and equipment and support from subsidiaries;

(ii) debt type/seniority class, with corporate loan, corporate bond or preferred stock as the debt type, and secured, senior unsecured, subordinated as seniority classes;

(iii) the firm’s business-cycle-adjusted leverage, relative seniority standing and (for public companies) Moody’s KMV “distance-to-default”;

(iv) historical average of recoveries aggregated across many firms;

(v) industry-level distance-to-default within countries and geographical regions.

Letting \( F_{\alpha, \beta} \) denote the distribution function of the \( \text{Beta}(\alpha, \beta) \) distribution, the response variable \( y \) in the regression model of LossCalc is \( y = \Phi^{-1}(F_{\alpha, \beta}(RR)) \), where \( \Phi \) is the standard normal distribution function. The method of least squares is used to estimate \( \alpha, \beta \) and the regression parameters. A year-by-year out-of-sample prediction error criterion is used to select the regressors; see Appendix B of [16].

The assumption underlying the Box-Cox-type transformation \( \Phi^{-1}(F_{\alpha, \beta}(RR)) \) in LossCalc is that RR, which takes values in the unit interval \((0, 1)\), has a beta distribution whose parameters depend on the predictors. The transformation is used to convert the beta random variable into one whose support is the entire real line so that ordinary linear regression can be applied. A better approach to regression with beta-distributed responses involves generalized linear models, as illustrated in the following.
Example 1. Suppose one has a sample of \( n = 100 \) independent random variables \( y_i \) having the beta distribution with density function

\[
\frac{\Gamma(\phi)}{\Gamma(\mu_i \phi)\Gamma((1-\mu_i) \phi)} y^{\mu_i \phi - 1} (1 - y)^{(1-\mu_i) \phi - 1}, \quad 0 < y < 1,
\]

in which \( \mu_i \) is a function of a predictor \( x_i \). Beta regression \([17]\), which is a special case of generalized linear models, uses the logit link function to relate \( \mu_i \) to a linear function of \( x_i \), i.e., \( \text{logit}(\mu_i) = a + bx_i \). We first consider this model and generate a sample \( (x_i, y_i), \; i = 1, \ldots, 100 \), from this model with \( a = 0.1, b = 0.5, \phi = 30 \) and with \( x_i \) evenly spaced in \([0, 5]\). The MLEs based on this sample are \( \hat{a} = 0.12, \hat{b} = 0.502, \phi = 35.6 \), which can be used to estimate \( \mu_i \) by \( \hat{\mu}_i = \hat{a} + \hat{b} x_i \), yielding \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 = 0.0017 \). If we apply least squares to the transformed \( \Phi^{-1}(F_{\alpha, \beta}(y_i)) \) as in LossCalc, we obtain \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 = 0.01 \), where \( \hat{\mu}_i \) is the estimate of \( \mu_i \) based on the least squares estimate of the transformed model. Figure 5 gives the box-plots of the cumulative squared errors of the two methods based on 50 simulated samples. The p-value of the pairwise t-test comparing \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 \) and \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 \), with respective mean values 0.0093 and 0.0158, for these 50 simulated samples, is \( 4.176 \times 10^{-6} \), showing that beta regression performs much better than the LossCalc transformation.

![Figure 5 is about here.](image)

We next consider the model \( \mu_i = \Phi(F_{1,1}(a + bx_i)) \) as assumed by LossCalc. Since \( F_{1,1} \) is the distribution of the Beta(1,1) (=uniform over \([0,1]\)) distribution, this corresponds to using the probit link (instead of the logit link) in the beta regression model. Figure 6 gives the box-plots of \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 \) and \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 \) based on 50 simulated samples, in which \( \hat{\mu}_i \) and \( \mu_i \) are the same as before, with \( \hat{\mu}_i \) still using the logit link. The means of \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 \) and \( \sum_{i=1}^{100}(\mu_i - \hat{\mu}_i)^2 \) are 0.0149 and 0.0179, respectively, and the p-value of the pairwise t-test comparing these two cumulative squared errors is 0.0047, showing that beta regression still performs better than the LossCalc transformation model even though the data are generated from \( \mu_i = \Phi(a + bx_i) \) that is consistent with LossCalc which, however, does not use the MLE.

![Figure 6 is about here.](image)

5.2. Modeling \( LGD_i \) using Bernoulli and Beta mixed models

Unlike corporate loans which are still quite liquid after default, personal loans, auto loans and mortgage loans lack such liquidity and the usual recovery process involves (a) in-house departments, (b) selling off the debt
at some fraction of the face value, and (c) collection or foreclosure agencies which take a fixed percentage of the recovered loan as commission. In a simplified description of the recovery process, the bank first tries (a) and then switches to (b) or (c) if (a) is unsuccessful after, say, 3 months. In the case of mortgage loans, (b) means selling the mortgage to another mortgage company (e.g., Fannie Mae or Freddie Mac in the U.S.), while (c) means foreclosure of the mortgage and repossession and sale of the property. The shortfall is the residual loan value plus planned interest minus the sales price of the mortgage or the repossessed property. For occasional missed payments, the “defaulted” obligors who have forgotten to pay or have been away usually pay within one month after they are contacted by the in-house department. Such scenario corresponds to complete recovery and leads to an atom at 0 in the marginal distribution of the economic LGD.

Example 2. To illustrate the importance of including the atom at 0, consider the following model. For $i = 1, \ldots, n$, let $I_i$ be independent Bernoulli($\tau_i$), where $\pi_i = 1/(1 + \exp(0.1 + 0.2x_i))$. Let $Y_i$ be independent beta random variables having density functions (16) with $\phi_i = 30$ and $\mu_i = 1/(1 + \exp(0.1 + 0.5x_i))$ which are the same as in Example 1. Let $\bar{\mu}_i$ and $\tilde{\mu}_i$ be the same as in Example 1, ignoring the atom of LGD at 0. It is more appropriate to use logistic regression to estimate $\pi_i$ by $\tilde{\pi}_i$ and beta regression to estimate the mean $\mu_i$ of $Y_i|Y_i > 0$ by $\bar{\mu}_i$, so that $\pi_i\bar{\mu}_i$ is used to estimate $E(LGD_i|x_i) = \pi_i\mu_i$. Figure 7 gives the box-plots of $\sum_{i=1}^{100}(\pi_i\bar{\mu}_i - \tilde{\pi}_i\bar{\mu}_i)$, $\sum_{i=1}^{100}(\pi_i\mu_i - \bar{\mu}_i)^2$ and $\sum_{i=1}^{100}(\pi_i\mu_i - \tilde{\pi}_i\bar{\mu}_i)^2$, based on 50 simulations from the Bernoulli-beta model. The means of $\sum_{i=1}^{100}(\pi_i\bar{\mu}_i - \tilde{\pi}_i\bar{\mu}_i)$, $\sum_{i=1}^{100}(\pi_i\mu_i - \bar{\mu}_i)^2$ and $\sum_{i=1}^{100}(\pi_i\mu_i - \tilde{\pi}_i\bar{\mu}_i)^2$ are 0.1061, 0.0561 and 0.0231, respectively, showing that $\tilde{\pi}_i\bar{\mu}_i$ gives a substantially smaller mean squared error than the other two estimates.

BCBS [2] requires LGD estimates to be “grounded in historical recoveries and, where applicable, must not solely based on the collateral’s estimated market values.” Business cycles should also be taken into account since at times of economic downturn both PD and LGD tend to increase; the database should ideally cover at least one business cycle and must not be shorter than 5 years. In addition, other attributes of the obligor’s creditworthiness and collateral should also be included. We propose the following generalized linear mixed model to estimate LGD using these data.

Let $D_s$ denote the set of obligors who default in year $s$ or before and who have available credit and collateral information in year $s - 1$, with $s = t, t - 1, \ldots, t - 4$. For $i \in D_s$, let $Y_i(s)$ be the economic LGD in year $s$. 

Figure 7 is about here.
As noted above, \( Y_i(s) \) has an atom at 0. Let \( \pi_i(s) = P\{Y_i(s) > 0\} \), which we model by
\[
\text{logit}(\pi_i(s)) = \xi_i + \eta^T x_i(s-1) + \gamma^T f(s),
\]
where \( x_i(s-1) \) is a vector of attributes of the obligor's creditworthiness and collateral in year \( s-1 \) and \( f(s) \) is a vector of macroeconomic factors in year \( s \). Note that our goal is to predict LGD\(_i(t+1)\) of the \( i \)th obligor for year \( t+1 \) using \( x_i(t) \) and the forecast of \( f(t+1) \), and that LGD\(_i(t+1)\) = \( \pi_i(t+1)E(Y_i(t+1)|Y_i(t+1) > 0) \). The conditional density of \( Y_i(s)|Y_i(s) > 0 \) is assumed to belong to the exponential family
\[
f(y|\mu_i(s), \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_i(s))\phi} y^{\mu_i(s)-1}(1-y)^{(1-\mu_i(s))\phi-1}
\]
for \( 0 < y < 1 \), where
\[
\mu_i(s) = E(Y_i(s)|Y_i(s) > 0), \quad \text{logit}(\mu_i(s)) = \zeta_i + \kappa^T x_i(s-1) + \nu^T f(s),
\]
\[
\text{Var}(Y_i(s)|Y_i(s) > 0) = \frac{\mu_i(s)(1-\mu_i(s))}{1+\phi},
\]
and \( \xi_i \) are i.i.d. having density function \( h_\theta \) and are independent of \( \zeta_i \) which are i.i.d having density function \( k_\phi \). As in (10)–(11), we introduce inter-subject random effects \( \xi_i \) and \( \zeta_i \) in (23) and (25), yielding a generalized linear mixed model of the type introduced by Breslow and Clayton [13]. Hamele, Knapp and Wildenauer [18, pp.129-130] have also used mixed effects models to analyze the recovery rates of corporate loans. They assume the one-factor model (3), in which \( V_i \) is replaced by \( \{\text{logit}(RR) - \mu_{i,s}\}/\sigma \) for the \( i \)th obligor and \( Z \) is replaced by \( Z_s \). Then \( \mu_{i,s} \) is like \( \text{logit}(\mu_i(s)) \) in (25) but uses a constant \( \zeta \) instead of the subject-specific \( \zeta_i \). The random effects in their linear mixed effects model arise from the unobservable factors \( Z_s \).

Generalized linear mixed models are marginal regression models. Because time-varying covariates are used in the Bernoulli mixed model for \( 1_{\{Y_i(s) > 0\}} \) in (23) and the beta mixed model (24)–(25) for \( Y_i(s)|Y_i(s) > 0 \), we use the working assumption of independence among \( Y_i(s), Y_i(s-1), \ldots , \) as in [13] and [14]; see also [19]. Let \( D = \cup_{t-4 \leq s \leq t} D_s \). The likelihood function for the beta mixed model (24)–(25) is
\[
\prod_{i \in D} \int_{t-4 \leq s \leq t} \prod_{i \in D_s} f(Y_i(s)|\mu_i(s; \zeta, \kappa, \nu)) h_\theta(\zeta) d\zeta,
\]
in which we write \( \mu_i(s; \zeta, \kappa, \nu) \) instead of \( \mu_i(s) \) that depends on the parameter \( \kappa \) and \( \nu \) besides the random effect \( \zeta_i \) with density function \( h_\theta \). Similarly,
letting \( I_i(s) = 1_{\{Y_i(s) > 0\}} \), the likelihood function for the Bernoulli mixed model (23) is

\[
\prod_{i \in C} \int_{t-4 \leq s \leq t} \prod_{i \in C_s} \left( \pi_i(s; \xi, \eta, \gamma) \right)^{I_i(s)} \left( 1 - \pi_i(s; \xi, \eta, \gamma) \right)^{1-I_i(s)} k_\theta(\xi) d\xi, \quad (28)
\]

where \( C_s \) denotes the set of obligors who have available credit and collateral information in year \((s-1)\), similar to the \( C_j^{(s)} \) in (9), and \( C = \bigcup_{t-4 \leq s \leq t} C_s \). The integral in (27) or (28) can be evaluated by Gaussian quadrature as in (14).

6. Discussion

This paper studies modeling PD and LGD for retail loans, for which "default" means delinquency. There is a much more extensive literature on credit risk of corporate loans, for which default means bankruptcy (an absorbing state) and requires hazard modeling. In particular, as indicated in Section 2, factor models have been developed to incorporate "default correlations" and extended to retail products in an unnatural and somewhat arbitrary way. Difficulties with multivariate hazard modeling can be circumvented for retail products by modeling \( PD_i \) directly as the transition probability of a non-homogeneous Markov chain. This Markov chain formulation also circumvents difficulties with drop-in and drop-out of obligors in the training sample, which correspond to staggered entry and censoring for hazard modeling in survival analysis and which would require stratification of the obligors according to their information sets if one should use logistic regression with baseline and time varying covariates, as pointed out in Section 4.2. We use an empirical Bayes model for the transition probabilities of the Markov chain to handle the relatively small number of defaults in the training sample, especially for low-risk buckets, and to incorporate macroeconomic factors. Note that the macroeconomic factors provide some linkage among the obligors, similar to the factor models (3) for corporate loans. However, unlike (3) that uses the factor \( Z \) and idiosyncratic risk \( \varepsilon_i \) to model the obligor’s standardized asset level whose drop below some theoretical threshold triggers default, the macroeconomic factors \( f(s) \) relate directly to the default probabilities (10) and (11) of all obligors, making the model much easier to estimate than (3). Moreover, econometric forecasts of these factors, such as unemployment rate and GDP, for the next period are easily accessible, unlike the abstract unobservable factor \( Z \) in [18]. In this connection, we comment
that the choice (5) or 0.15, 0.04 of \( \rho \) for retail loans by BCBS is unnecessarily fettered by the inherent difficulties of multivariate default (bankruptcy) modeling of corporate loans, and that a more direct statistical approach can be used to set capital requirement for retail products.

Note that we have used (10)–(11) with random effects for \( \lambda_j \) and \( \bar{\lambda}_j \) only as a working model. As explained in the second paragraph of Section 4.2 and illustrated in the simulation study in Section 4.3, these random effects are used to circumvent difficulties, particularly from the regulatory perspective, caused by relatively few observed defaults for those obligors in low risk buckets when one uses maximum likelihood that treats \( \lambda_j \) and \( \bar{\lambda}_j \) as fixed parameters instead of the Bayesian approach that imposes prior distributions on them. On the other hand, we have only put convenient marginal prior distributions on \( \lambda_j \) (and \( \bar{\lambda}_j \)) and have combined them by a working independence assumption, instead of using a joint prior distribution that incorporates the order constraint \( \lambda_1 < \cdots < \lambda_J \) for the risk buckets. The independence assumption results in one-dimensional integrals for the algorithm in Section 4.2, and it suffices for modifying the MLE of \( \lambda_j \) (or \( \bar{\lambda}_j \)) with a Bayes estimator when \( N_{01,j} \) (or \( N_{11,j} \)) is small, whereas using a joint prior distribution that incorporates the order constraint will lead to a considerably more complicated algorithm that involves multivariate integrals.

The empirical Bayes approach is also used to model LGD, in the form of generalized linear mixed models, with a Bernoulli component to model the probability that \( \text{LGD}=0 \) and a Beta component to model LGD given that it is positive. Since its focus is on these empirical Bayes models for PD and LGD, this paper does not address other important issues concerning statistical models for the IRB approach to measuring credit risk of retail loans. One such issue is the performance assessment of a bank’s internal models. As noted by Oung in BCBS[4, p.96], “At the current stage of knowledge, unambiguous and complete statistical tests enabling a formal acceptance or rejection of an IRB system do not appear to be available. Difficulties mainly relate to the effect of default correlation, data constraints, and the definition of meaningful and robust target criteria for validating IRB systems.” Compounding the difficulties in coming up with “meaningful and robust” performance measures to assess the internal models, banks and their regulators have obviously different objectives which need to be reconciled in the development and assessment of the internal models. We are currently studying backtesting of the models proposed herein and exploring new performance measures to assess these and other models in practice.
References


Tze Leung Lai
Department of Statistics
Stanford University
Stanford, CA 94305, U.S.A.
E-mail address: lait@stanford.edu

Samuel Po-Shing Wong
Department of Statistics
The Chinese University of Hong Kong
Shatin, Hong Kong.
E-mail address: samwong@sta.cuhk.edu.hk
Figure 1: Box-plots of divergence (21) for $p_{ij}^{EST}(6)$, in which the estimate (EST) is point-in-time (PIT) logistic regression $\hat{p}_{ij}(6)$, or MLE or Bayes modeling of (10)--(11) that yields $p_{ij}^*(6)$ or $\hat{p}_{ij}(6)$. 
Figure 2: Box-plots of $L_1$-loss (22) for $p_{ij}^{EST}(6)$, in which the estimate (EST) is point-in-time (PIT) logistic regression $\bar{p}_{ij}(6)$, or MLE or Bayes modeling of (10)–(11) that yields $p_{ij}^*(6)$ or $\hat{p}_{ij}(6)$. 
Figure 3: Box-plots of divergence (21) in the case of randomly generated $p_j$ and $\tilde{p}_j$ ($1 \leq j \leq J$).
Figure 4: Box-plots of $L_1$-loss (22) in the case of randomly generated $p_j$ and $\tilde{p}_j$ ($1 \leq j \leq J$).
Figure 5: Box-plots of cumulative squared errors of two methods.
Figure 6: Box-plots of cumulative squared errors under probit link.
Figure 7: Box-plots of cumulative squared errors under the Bernoulli-Beta model.