Review Article

Credit Portfolios, Credibility Theory, and Dynamic Empirical Bayes

Tze Leung Lai

Department of Statistics, Stanford University, Stanford, CA 94305, USA

Correspondence should be addressed to Tze Leung Lai, lait@stanford.edu

We begin with a review of (a) the pricing theory of multiname credit derivatives to hedge the credit risk of a portfolio of corporate bonds and (b) current approaches to modeling correlated default intensities. We then consider pricing of insurance contracts using credibility theory in actuarial science. After a brief discussion of the similarities and differences of both pricing theories, we propose a new unified approach, which uses recent advances in dynamic empirical Bayes modeling, to evolutionary credibility in insurance rate-making and default modeling of credit portfolios.

1. Introduction

Credit markets provide liquidity to the underlying financial system. They include not only banks that give loans to consumers and businesses, but also financial institutions and corporations that develop and trade securities and derivative contracts derived from these loans and credit-sensitive investments such as corporate bonds, and insurance companies and their subsidiaries that issue insurance and reinsurance contracts to protect against adverse credit events. The recent financial crisis in the USA was associated with unpreparedly high default rates of mortgage loans and defaultable bonds in credit markets, beginning with subprime mortgage loans in 2007 and culminating in the collapse of large financial institutions such as Bear Stearns and Lehman Brothers in 2008. In Sections 2 and 3, we give a brief review of the conventional approaches to modeling default risk and of widely used derivative contracts to hedge credit risk, respectively. In particular, we point out in Section 2 the similarities and differences between the intensity models in credit risk and those in survival analysis.

Parallel to the increasing volume of subprime mortgage loans whose value was estimated to be $1.3 trillion by March 2007, an important development in financial markets during the past decade was the rapid growth of credit derivatives, culminating in $32 trillion worth of notional principal for outstanding credit derivatives by December 2009. These derivative contracts are used to hedge against credit loss of either a single one or a portfolio of corporate bonds. A "credit default swap" (CDS) is a contract between a buyer, who obtains the right to sell bonds issued by a company or country (a "single name" that is called the *reference entity*), and a seller (of the CDS contract) who agrees to buy the bonds for their face values when a credit event occurs. The face value of a coupon-bearing bond is the principal that the issuer repays at maturity if it does not default. The total face value of the bonds is the notional principal of the CDS. The buyer of the CDS makes periodic payments (usually quarterly) to the seller until the expiration date T of the CDS or until a credit event occurs. If there is no credit event up to time T, the buyer receives no payoff. If there is a credit event at time $\tau < T$, the buyer's payoff is the difference between the face value and the mid-market value of the cheapest deliverable bond determined by an auction process several days after the credit event if the contract specifies cash settlement, or between the face value and what the buyer receives by selling the bond if the contract specifies physical settlement. The credit event is defined in the CDS contract and includes failure to pay according to schedule, agency downgrade, bankruptcy, and debt restructuring. The credit risk of a portfolio of corporate bonds can be mitigated by using a multiname credit derivative. Stochastic models of default intensities are used to price these derivatives. For multiname credit derivatives, joint modeling of the default intensities has been a challenging statistical problem. The review in Section 3.3 describes the correlated default models that have been widely used prior to the financial crisis and points out some of their limitations and deficiencies.

In July 2007, Bear Stearns disclosed that two of its subprime hedge funds, the High-Grade Structured Credit Fund and the High-Grade Structured Credit Strategies Enhanced Leverage Master Fund, which were invested in CDOs, had lost nearly all their value following a rapid decline in the subprime mortgage market. Lawsuits were brought by investors against former managers of the hedge funds, and Standard & Poor's (S&P) downgraded the company's credit rating. In March 2008, the Federal Reserve Bank of New York initially agreed to provide a $25 billion collateralized 28-day loan to Bear Stearns, but subsequently changed the deal to make a $30 billion loan to J. P. Morgan Chase to purchase Bear Stearns at $2 a share, or less than 7% of Bear Stearns' market value just two days before. After a class action suit was filed on behalf of the shareholders, J. P. Morgan Chase raised the offer to $10 per share, which was approved by the shareholders in May 2008. Lehman Brothers also suffered unprecedented loss for its large positions in subprime and other lower-rated mortgage-backed securities in 2008. After attempts to sell it to Korea Development Bank and then to Bank of America and to Barclays failed, it filed for Chapter 11 bankruptcy protection on September 15, 2008, making the largest bankruptcy filing, with over $600 billion in assets, in US history.

A day after Lehman's collapse, American International Group (AIG) needed bailout by the Federal Reserve Bank, which gave the insurance company a secured credit facility of up to $85 billion to enable it to meet collateral obligations after its credit ratings were downgraded below AA, in exchange for a stock warrant for 79.9% of its equity. AIG's London unit had sold credit protection in the form of CDS and CDO to insure $44 billion worth of securities originally rated AAA. As Lehman's stock price was plummeting, investors found that AIG had valued its subprime mortgage-backed securities at 1.7 to 2 times the values used by Lehman and lost confidence in AIG. Its share prices had fallen over 95% to just

$1.25 by September 16, 2008, from a 52-week high of $70.13. The "contagion" phenomenon, from increased default probabilities of subprime mortgages to those of counterparties in credit derivative contracts whose values vary with credit ratings, was mostly neglected in the models of joint default intensities that were used to price CDOs and mortgage-backed securities. These models also failed to predict well the "frailty" traits of latent macroeconomic variables that underlie the mortgages and mortgage-backed securities. Section 4 reviews recent works in the literature to incorporate frailty and contagion in modeling joint default intensities.

The major business of AIG is insurance, and insurance contracts rather than credit derivatives contracts are what it sells primarily. These insurance contracts provide protection against loss due to automobile accidents and other unexpected events such as fire and floods in home insurance, just like credit derivatives that provide protection against loss due to default or other credit events. On the other hand, as will be explained in Section 5, whereas pricing credit derivatives involves the risk-neutral stochastic dynamics of the credit events that are priced by the market, pricing insurance contracts involves real-world (rather than risk-neutral) statistical models of the events. Section 5 discusses the similarities and differences between both pricing theories and also reviews *credibility theory* in actuarial science, which aims at deriving the premium of an insurance contract that balances the experience of an individual risk with the class risk experience. In particular, pricing a CDO that provides CDS protection for a portfolio of companies and setting the premium of automobile insurance for drivers in the same risk class both involve predicting the probability of occurrence and the loss incurred for certain future events. Estimation of the probability and the loss can pool information from the individual (company or driver) and the collective (portfolio of companies or risk group) levels.

Such pooling of information is the essence of empirical Bayes (EB) methodolgy in statistics, introduced by Robbins [1] and Stein [2]. The review of credibility theory in Section 5 also reviews EB and points out their connections. Lai et al. [3] have recently shown how the predictive performance of conventional EB methods in estimating the accident-proneness of a driver in a future period, or other insurance applications such as workers' compensation losses due to permanent partial disability, can be improved by using a *dynamic EB* approach to longitudinal data. Section 6 describes this new approach and relates it to generalized linear mixed models (GLMM) in biostatistics introduced by Breslow and Clayton [4]. It also reviews recent work applying the dynamic EB approach via GLMM to credibility theory, credit portfolios, and default probability forecasts of subprime mortgage loans. Thus the dynamic EB methodology provides a unified statistical approach to credit portfolios and credibility theory, and Section 7 provides further discussion and concluding remarks.

2. Intensity and Structural Models of Default Risk

2.1. Survival Analysis, Intensity Processes, and Intensity Models for Default Risk

We begin with a brief review of survival analysis in biostatistics and reliability theory and then show its connections with widely used intensity models for pricing corporate bonds and other credit-sensitive securities. To pinpoint these connections, our review focuses on the martingale approach to survival analysis, as described in [5, 6]. A major theme of the statistical problems in survival analysis is to study the distribution of a failure time τ, based

on possibly censored data from either a homogeneous population or a regression model with covariates. The survival function of τ is $S(t) = P(\tau \geq t) = 1 - F(t-)$, where $F(s) = P(\tau \leq s)$ and $F(t-) = \lim_{s \to t-} F(s) = P(\tau < t)$. Clearly, $S(t)$ is a nonincreasing left-continuous function of t with $S(0) = 1$ and $\lim_{t \to \infty} S(t) = 0$. If τ is absolutely continuous, the *probability density function* of τ is $f(t) = -dS(t)/dt$, and the *hazard* (or *intensity*) function is

$$\lambda(t) = \lim_{h \to 0+} \frac{P(t \leq \tau < t + h \mid \tau \geq t)}{h} = \frac{f(t)}{S(t)}. \tag{2.1}$$

Integrating (2.1) with respect to t yields

$$S(t) = \exp\left[-\int_0^t \lambda(u)du\right] = \exp[-\Lambda(t)], \tag{2.2}$$

where $\Lambda(t) = \int_0^t \lambda(s)ds$ is called the *cumulative hazard function*. Without assuming F to be continuous, we can define the cumulative hazard function by the Stieltjes integral

$$\Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s-)}. \tag{2.3}$$

Regarding the indicator functions $I_{\{\tau \leq t\}}$, $t \geq 0$, as a stochastic process, an important property of Λ is that

$$\left\{I_{\{\tau \leq t\}} - \Lambda(\tau \wedge t), t \geq 0\right\} \text{ is a martingale,} \tag{2.4}$$

where $\tau \wedge t = \min(\tau, t)$. Fleming and Harrington [5] and Andersen et al. [6] give overviews of continuous-time martingales and their applications, via (2.4) and its analogs, to survival analysis. Based on n independent observations τ_i sampled from F, the nonparametric maximum likelihood estimate of F is the empirical distribution function $\hat{F}(t) = (1/n) \sum_{i=1}^{n} I_{\{\tau_i \leq t\}}$. However, in applications of survival analysis to biomedical and econometric data, the τ_i may not be fully observable because of censoring. Some subjects (or firms) may not fail during the observation period or may have been lost in followup during the period; the data on these individuals are said to be *right-censored*.

Right-censoring can be formulated by introducing censoring variables c_i that indicate the time of loss to followup. The censoring variable c_i can also indicate the length of the observation period. The observations, therefore, are (T_i, δ_i), $i = 1, \ldots, n$, where $T_i = \min(\tau_i, c_i)$ and $\delta_i = I_{\{\tau_i \leq c_i\}}$ is the censoring indicator that indicates whether T_i is an actual failure time or is censored. Subject i is "at risk" at time s if $T_i \geq s$ (i.e., has not failed and has not been lost to followup prior to s). Let

$$Y(s) = \sum_{i=1}^{n} I_{\{T_i \geq s\}}, \qquad N(s) = \sum_{i=1}^{n} I_{\{T_i \leq s, \delta_i = 1\}}. \tag{2.5}$$

Note that $Y(s)$ is the risk set size and $\Delta N(s) = N(s) - N(s-)$ is the number of observed deaths at time s, and that $\Delta N(s)/Y(s)$ is the analog of $\hat{p}_j$. We use $\Delta N(s)$ to denote the jump size of N

at s (i.e., $\Delta N(s) = N(s) - N(s-)$) and also use the convention $0/0 = 0$. The cumulative hazard function $\Lambda(t)$ can be estimated by

$$\widehat{\Lambda}(t) = \sum_{s \le t} \frac{\Delta N(s)}{Y(s)} = \int_0^t \frac{I_{\{Y(s)>0\}}}{Y(s)} dN(s) \tag{2.6}$$

for the censored data (T_i, δ_i), $1 \le i \le n$. This is called the *Nelson-Aalen estimator* [7, 8]. Analogous to (2.4),

$$\left\{ \int_0^t U(s)[dN(s) - Y(s)d\Lambda(s)], \, t \ge 0 \right\} \text{ is a martingale,} \tag{2.7}$$

for every left-continuous stochastic process $U(s)$; note that $Y(s)$ is left-continuous. From this and the martingale central limit theorem,

$$\frac{\widehat{\Lambda}(t) - \Lambda(t)}{\left\{ \int_0^t (I_{\{Y(s)>0\}}/Y^2(s))dN(s) \right\}^{1/2}} \xrightarrow{\mathcal{D}} N(0,1) \tag{2.8}$$

as $n \to \infty$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution; see [5, 6].

Altman [9] and others have developed mortality tables for loans and bonds by using methods in actuarial science. These mortality tables (also called life tables) are used by actuaries to set premiums for life insurance policies. Partition time into disjoint intervals $I_1 = (0, t_1]$, $I_2 = (t_1, t_2]$, and so forth. A life table summarizes the mortality results of a large cohort of n subjects as follows:

n_j = number of subjects alive at the beginning of I_j,

d_j = number of deaths during I_j, l_j = number lost to followup during I_j.

It estimates $p_j = P$ (dying during I_j | alive at the beginning of I_j) by $\widehat{p}_j = d_j/(n_j - l_j)$ so that the actuarial (life-table) estimate of $P(\tau > t_k)$ is the product

$$\prod_{j=1}^k (1 - \widehat{p}_j) = \prod_{j=1}^k \left(1 - \frac{d_j}{n_j - l_j} \right). \tag{2.9}$$

Without discretizing the failure times, Kaplan and Meier [10] introduced the *product-limit* estimator of $S(t) = P(\tau \ge t)$:

$$\widehat{S}(t) = \prod_{s<t} \left(1 - \frac{\Delta N(s)}{Y(s)} \right). \tag{2.10}$$

Since $N(s)$ has at most n jumps, the product in (2.10) is finite. Note that $\widehat{S}(t) = \prod_{s<t}(1 - \Delta\widehat{\Lambda}(s))$ by (2.6) and (2.10). As shown in [6],

$$S(t) = \prod_{s<t} (1 - d\Lambda(s)) \quad \forall t \text{ such that } \Lambda(t) < \infty. \tag{2.11}$$

The product in (2.11) is called the *product integral* of the nondecreasing function $\Lambda(\cdot)$ on $[0, t]$; it is defined by the limit of $\prod\{1 - [\Lambda(t_i) - \Lambda(t_{i-1})]\}$, where $0 = t_0 < t_1 < \cdots < t_m = t$ is a partition of $[0, t]$ and the product is in the natural order from left to right, as the mesh size $\max_{1 \leq i \leq m} |t_i - t_{i-1}|$ approaches 0. Moreover, by making use of the martingale central limit theorem, it is shown in [5, 6] that

$$\frac{\widehat{S}(t) - S(t)}{\left[\widehat{S}(t)\left\{\int_0^t (I_{\{Y(s)>0\}} / Y^2(s)) dN(s)\right\}^{1/2}\right]} \xrightarrow{\mathfrak{D}} N(0, 1). \tag{2.12}$$

2.1.1. Regression Models for Hazard Functions with Covariates

In applications, one often wants to use a model for τ to predict future failures from a vector $\mathbf{x}(t)$ of predictors based on current and past observations; $\mathbf{x}(t)$ is called a time-varying covariate. When $\mathbf{x}(t) = \mathbf{x}$ does not depend on t, it is called a time-independent (or baseline) covariate. In practice, some predictors may be time-independent, while other components of $\mathbf{x}(t)$ may be time-varying. Since prediction of future default from $\mathbf{x}(t)$ is relevant only if $\tau > t$ (i.e., if default has not occurred at or before t), one is interested in modeling the conditional distribution of τ given $\tau > t$, for example, by relating the hazard function $\lambda(t)$ to $\mathbf{x}(t)$. Noting that $\lambda(t)dt$ is the probability that $\tau \in [t, t + dt]$, Cox [11] introduced the *proportional hazards* (or *Cox regression*) model

$$\lambda(t) = \lambda_0(t) \exp\left(\boldsymbol{\beta}^T \mathbf{x}(t)\right). \tag{2.13}$$

Putting $\mathbf{x}(t) = 0$ in (2.13) shows that $\lambda_0(\cdot)$ is also a hazard function; it is called the *baseline hazard*. Cox regression fits (2.13) to the observed data $(T_i, \delta_i, \mathbf{x}_i(t)), t \leq T_i, 1 \leq i \leq n$. Lane et al. [12] apply Cox regression to predict bank default, using time-independent covariates.

The parameters of (2.13) are $\boldsymbol{\beta}$ and the cumulative hazard function Λ_0, which is an infinite-dimensional parameter. One way to estimate $(\boldsymbol{\beta}, \Lambda_0)$ is to specify Λ_0 via a parameter vector $\boldsymbol{\theta}$ so that $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can be estimated by maximum likelihood. Since an observed failure contributes a factor $f_i(T_i)$ to the likelihood, where $f_i = \lambda_i S_i$ is the density function of the τ_i while a censored failure contributes a factor $S_i(T_i)$ to the likelihood, and since $\log S_i(t) = -\Lambda_i(t)$ and $\Lambda_i(t) = \Lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_i(t))$, the log-likelihood function can be written as

$$\begin{aligned}
l(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \sum_{i=1}^{n} \left\{\delta_i \log f_i(T_i) + (1 - \delta_i) \log S_i(T_i)\right\} \\
&= \sum_{i=1}^{n} \left\{\delta_i \left[\log \lambda_0(T_i; \boldsymbol{\theta}) + \boldsymbol{\beta}^T \mathbf{x}_i(T_i)\right] - \Lambda_0(T_i; \boldsymbol{\theta}) e^{\boldsymbol{\beta}^T \mathbf{x}_i(T_i)}\right\};
\end{aligned} \tag{2.14}$$

see [13]. Commonly used parametric models for failure times are the exponential density $\theta e^{-\theta t}$ that has constant hazard $\theta > 0$ and the Weibull density $\theta t^{\theta-1} \exp(-t^\theta)$ that has hazard function $\theta t^{\theta-1}$ and cumulative hazard function t^θ, with $\theta > 0$. Lee and Urrutia [14] use a Weibull distribution of default times to predict insolvency in the property-liability insurance industry. Hazard models with time-varying covariates have been used by McDonald

and Van de Gucht [15] to model the timing of high-yield bond default and of exercising call options. Related survival analysis to predict bankruptcy has been carried out in [16–19].

Instead of assuming a parametric model to estimate the baseline hazard function $\lambda_0(\cdot)$, Cox [11, 20] introduced a semiparametric method to estimate the finite-dimensional parameter $\boldsymbol{\beta}$ in the presence of an infinite-dimensional nuisance parameter $\lambda_0(\cdot)$; it is semiparametric in the sense of being nonparametric in λ_0 but parametric in $\boldsymbol{\beta}$ (in terms of $\boldsymbol{\beta}^T \mathbf{x}_t$). Cox's partial likelihood method decomposes the likelihood function into two factors, with one involving only $\boldsymbol{\beta}$ and the other involving both $\boldsymbol{\beta}$ and Λ. It estimates $\boldsymbol{\beta}$ by maximizing the following partial likelihood, which is a factor that only involves $\boldsymbol{\beta}$. Order the observed censored failure times as $\tau_{(1)} < \cdots < \tau_{(m)}$, with $m \le n$. Let C_j denote the set of censored T_i's in the interval $[\tau_{(j-1)}, \tau_{(j)})$, and let (j) denote the individual failing at $\tau_{(j)}$, noting that with probability 1 there is only one failure at $\tau_{(j)}$ because the failure time distributions have density functions. Let $R_{(j)} = \{i : T_i \ge \tau_{(j)}\}$ denote the risk set at $\tau_{(j)}$. Cox's regression estimator $\widehat{\boldsymbol{\beta}}$ is the maximizer of the partial log-likelihood

$$l(\boldsymbol{\beta}) = \sum_{j=1}^{m} \left\{ \boldsymbol{\beta}^T \mathbf{x}_{(j)}(\tau_{(j)}) - \log\left(\sum_{i \in R_{(j)}} \exp\left(\boldsymbol{\beta}^T \mathbf{x}_{(i)}(\tau_{(j)}) \right) \right) \right\}, \qquad (2.15)$$

or equivalently, is the solution of $(\partial / \partial \boldsymbol{\beta})l(\boldsymbol{\beta}) = 0$. Martingale theory can be used to show that

$$\left(-\ddot{l}(\boldsymbol{\beta}) \right)^{1/2} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \text{ has a limiting standard normal distribution as } n \longrightarrow \infty, \qquad (2.16)$$

where $\ddot{l}(\boldsymbol{\beta})$ is the Hessian matrix of second partial derivatives $(\partial^2 / \partial\beta_k \partial\beta_h)l(\boldsymbol{\beta})$. One, therefore, can perform usual likelihood inference, treating the partial likelihood as a likelihood function. Moreover, even though $\widehat{\boldsymbol{\beta}}$ is based on partial likelihood, it has been shown to be asymptotically efficient; see [6]. Modifying $Y(s)$ in (2.5) to

$$Y(s) = \sum_{i \in R_{(j)}} \exp\left(\boldsymbol{\beta}^T \mathbf{x}_i(s) \right) \quad \text{at } s = \tau_{(j)} \qquad (2.17)$$

leads to the Breslow estimator [21] of Λ_0, which is again given by (2.6) but with $Y(s)$ defined by (2.17).

2.1.2. Intensity-Based Models for Pricing Default Risk

Besides the important roles they have played in the statistical analysis of failure-time data as explained in the preceding section, intensity modeling has also become a standard approach to pricing corporate bonds and other credit-sensitive securities since the 1990s. Intensity models provide an elegant way to combine term-structure modeling for default-free bonds with the default risk and recovery features of corporate bonds. Chapter 10 of [22] gives an overview of interest rate markets and the term structure of interest rates, statistical methods to estimate the zero-coupon yield curve from the current prices of default-free Treasury bonds, arbitrage-free pricing theory of interest rate derivatives, short rate and other models under the risk-neutral (instead of the real-world) measure entailed by arbitrage-free

pricing, and calibration of these models to market data involving actively traded interest rate derivatives. For corporate bonds, besides the interest rate processes, we also include an additional stochastic process for the default intensity of the bond. All the stochastic processes are defined on the same probability space in which the probability measure Q is a risk-neutral (or martingale) measure associated with arbitrage-free pricing. Expectations taken under this measure will be denoted by $\widehat{E}$ or E_Q, while E is used to denote expectation under the real-world (or physical) measure P.

A *Cox process* (also called *stochastic Poisson process*) is often used to model the default intensity of the bond's issuer. The default intensity λ_t is assumed to be governed by an exogenous stochastic process $\mathbf{X}_t$, $t \geq 0$, so that $\lambda_t = \lambda(\mathbf{X}_t)$ and the stochastic dynamics of λ_t is specified through $\mathbf{X}_t$. Let τ be the default time and $\Lambda(t) = \int_0^t \lambda(\mathbf{X}_s)ds$. Then $\Lambda(\tau)$ is distributed as an exponential random variable e_1 with mean 1 that is independent of $\{\mathbf{X}_s, s \geq 0\}$ since $P\{\Lambda(\tau) > t\} = P\{\tau > \Lambda^{-1}(t)\} = S(\Lambda^{-1}(t)) = \exp\{-\Lambda(\Lambda^{-1}(t))\} = e^{-t}$. Hence we can use $\lambda(\mathbf{X}_s)$ to generate τ by $\tau = \inf\{t : \int_0^t \lambda(\mathbf{X}_s)ds \geq e_1\}$. Consider a zero-coupon bond, with maturity date T and face value 1, issued by a firm at time 0. Assume that there is a short-rate process $r(\mathbf{X}_s)$ under the risk-neutral measure Q such that the default-free bond price is given by

$$p(0,T) = \widehat{E}\left\{\exp\left(-\int_0^T r(\mathbf{X}_s)ds\right)\right\}; \tag{2.18}$$

see [22, page 257]. Assuming $\lambda(\mathbf{X}_s)$ to be the intensity process for the default time τ of the firm and assuming zero recovery at default, the price of the defaultable bond at time 0 is

$$\pi(0,T) = \widehat{E}\left\{I_{\{\tau>T\}}\exp\left(-\int_0^T r(\mathbf{X}_s)ds\right)\right\} = \widehat{E}\left\{\widehat{E}[\cdots \mid \mathbf{X}_s, \, s \leq T]\right\}$$

$$= \widehat{E}\left\{\exp\left(-\int_0^T r(\mathbf{X}_s)ds\right)\widehat{E}[I_{\{\tau>T\}} \mid \mathbf{X}_s, \, s \leq T]\right\} \tag{2.19}$$

$$= \widehat{E}\left\{\exp\left(-\int_0^T (r+\lambda)(\mathbf{X}_s)ds\right)\right\},$$

since $\widehat{E}[I_{\{\tau>T\}} \mid \mathbf{X}_s, \, s \leq T] = Q\{\int_0^T \lambda(\mathbf{X}_s) < e_1 \mid \mathbf{X}_s, \, s \leq T\} = \exp(-\int_0^T \lambda(\mathbf{X}_s)ds)$, noting that $P(e_1 > x) = e^{-x}$ and that e_1 is independent of $\{\mathbf{X}_s, s \geq 0\}$. Thus (2.19) replaces the short rate in the default-free bond pricing formula by the default-adjusted short rate $(r+\lambda)(\mathbf{X}_s)$. Moreover, closed-form expressions are available for $\pi(0,T)$ if one uses affine models for both r and λ; see [23, 24]. The short rate process and the intensity process have unknown parameters which can be encoded in a vector $\boldsymbol{\theta}$ and can be estimated from the bond prices by using least squares in the same way as that for Treasury bonds; see [22]. Note that unlike survival analysis in biostatistics that uses censored failure-time data to estimate the intensity in (2.13) or its integral in (2.3), estimation of the parameters of the intensity process is based not on observed defaults but rather on the bond prices that involve the risk-neutral (rather than the real-world) measure.

In practice, recovery rates also need to be considered and the vector $\mathbf{X}_s$ may include predictors that make $r(\mathbf{X}_s)$ and $\lambda(\mathbf{X}_s)$ much more complex than the affine model that gives an explicit formula for the integral in (2.19). Monte Carlo methods are used instead to evaluate

(2.19) in these more complex situations. Since corporate bonds typically promise cash flows, which mimic those of Treasury bonds, it is natural to consider yields when one compares the effects of different recovery assumptions at default. The yield at date t of a zero-coupon bond with price B_t, face value A, and maturity T is defined by

$$y(t,T) = \frac{1}{T-t} \log \frac{A}{B_t},$$

(2.20)

and the difference between the yield of a defaultable bond and a corresponding Treasury bond is referred to as the *credit spread* or *yield spread*. The first type of recovery assumptions is *recovery of market value*. This measures the change in market prices before and after the default, representing the loss in the bond's value associated with the default. More generally, consider the price process of a defaultable contingent claim V promising a payoff of $f(X_T)$ at T. The claim is said to have a fractional recovery of market value of δ at default time τ if the amount recovered in the event of default is equal to

$$h(\tau) = \delta V(\tau-) \quad \text{for } \tau \leq T,$$

(2.21)

where $V(\tau-)$ is the value of the claim just prior to default and $\delta \in [0,1)$. With this recovery assumption, the price at t of the defaultable contingent claim (if there has not been a default up to date t) is

$$V(t) = \widehat{E}\left[f(\mathbf{X}_T) \exp\left(-\int_t^T (r + (1-\delta)\lambda)(\mathbf{X}_s) ds \right) \mid \mathbf{X}_s, \ s \leq T \right],$$

(2.22)

which is an extension of (2.19). A variant of recovery of market value is *recovery of face value*, which is the closest to legal practice in the sense that debt with the same priority is assigned a fractional recovery depending on the outstanding notional amount but not on maturity or coupon. It is also the measure typically used in rating-agency studies. The third type of recovery assumptions is *recovery of treasury*, under which the corporate bond in default is replaced with a treasury bond with the same maturity but a reduced payment. Unlike recovery of face value, it tries to correct for the fact that amounts of principal with long maturity should be discounted more than principal payments with short maturity.

2.2. Econometric Modeling of Default Risk

The intensity-based models in the preceding section are sometimes called *reduced-form* models, using a terminology in econometrics that refers to models of the relationship of an economic variable to exogenous variables. In contrast, *structural* models attempt to derive the relationship from some underlying economic theory. Structural models for the default of a firm assume that default is triggered when the firm's asset and debt values enter some region whose boundary is called a "default barrier."

2.2.1. Merton's Model of Debt and Equity Values and the KMV Credit Monitor Model

As in the Black-Scholes model described in Section 8.1 of [22], Merton [25] assumes a market with continuous trading, no transaction costs, unlimited short selling and perfectly divisible assets, and with a risk-free asset that has constant interest rate r. The asset value A_t of the firm is assumed to follow a geometric Brownian motion (GBM) $dA_t/A_t = \mu dt + \sigma dw_t$. Suppose at time 0 the firm has issued two kinds of contingent claims: equity and debt, in which debt is assumed to be a zero-coupon bond with face value D and maturity date T. We can regard the firm as being run by equity owners. At maturity date T of the bond, equity owners pay D to retain ownership if $A_t \geq D$, and would not pay D if $A_t < D$ since they have limited liability. In the latter case, bond holders receive a recovery of A_T instead of the promised payment D. Therefore, the equity value at time t is given by the Black-Scholes formula

$$V_e(t) = E_Q\left\{e^{-r(T-t)}(A_T - D)^+ \mid A_t\right\} = A_t\Phi(d_1) - De^{-r(T-t)}\Phi(d_2), \tag{2.23}$$

where Q is the risk-neutral measure under which the drift of the GBM is r instead of μ, and

$$d_1 = \frac{\log(A_t/D) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2 = d_1 - \sigma\sqrt{T-t}; \tag{2.24}$$

see [22, page 186]. Similarly, the debt value at time t is

$$V_d(t) = E_Q\left\{e^{-r(T-t)}\min(A_T, D) \mid A_t\right\} = A_t - V_e(t), \tag{2.25}$$

since $\min(A_t, D) = A_t - (A_t - D)^+$.

In the 1990s, the KMV Corporation of San Francisco (now merged into Moody's KMV) introduced its Credit Monitor Model for default prediction using Merton's framework. To begin with, the GBM for the asset value can be expressed in terms of the Brownian motion w_t by $A_t = A_0 \exp\{(\mu - (1/2)\sigma^2)t + \sigma w_t\}$; see [22, page 66]. Therefore, the probability of default on the debt at maturity is

$$
\begin{aligned}
P(A_T < D) &= P\left\{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma w_T < \log\left(\frac{D}{A_0}\right)\right\} \\
&= \Phi\left(-\frac{\log A_0 + (\mu - \sigma^2/2)T - \log D}{\sigma\sqrt{T}}\right) \\
&= \Phi\left(-\frac{E(\log A_t) - \log D}{\sigma\sqrt{T}}\right).
\end{aligned} \tag{2.26}
$$

This leads to KMV's *distance to default* defined by

$$\mathrm{DD} = \frac{\{E(\log A_T) - \log D\}}{\sigma\sqrt{T}} \tag{2.27}$$

so that the predicted probability of default under Merton's model is $\Phi(-DD)$. Since $\sigma\sqrt{T}$ is the standard deviation of $\log A_T$, DD can be regarded as the number of standard deviations of the log asset value from the log liability. Instead of using the normal distribution for $\log A_T$ to predict the default probability as in (2.26), the KMV Credit Monitor Model uses a large proprietary database of firms and firm defaults to construct an EDF (expected default frequency) score for a firm with κ standard deviations away from the default (i.e., $DD = \kappa$) at the beginning of a future period. The idea is to stratify the firms based on their DD's and estimate empirically the default probability of the firm from its associated stratum. The firm's DD is evaluated from its balance sheet and option models for its debt.

2.2.2. Black-Cox Framework of Default Barriers and Other Extensions of Merton's Model

Black and Cox [26] extend the Merton model by allowing defaults to occur prior to the maturity of the bond; specifically, when the asset value A_t hits a nonrandom, continuous boundary $g(t) \leq D$. They also allow the equity owners to be paid dividends, at rate q, so that under the risk-neutral measure Q the GBM for A_t has drift $r - q$; see [22, page 186]. Let $\tau = \inf\{0 \leq t \leq T : A_t = g(t)\}$. The payoff to bond holders is

$$\min(A_T, D)I_{\{\tau > T\}} + g(\tau)I_{\{\tau \leq T\}}, \tag{2.28}$$

in which the first summand reflects that the bond holders receive $\min(A_T, D)$ at the bond's maturity if default has not occurred up to that time, while the second summand reflects that the bond holders receive the recovery of $A_\tau = g(\tau)$ upon default. Similarly, the payoff to equity holders is $(A_T - D)^+ I_{\{\tau > T\}}$.

Stochastic models for the default-free interest rate dynamics (see [22, Chapter 10]) have been used to replace the constant r in Merton's model. Another extension is to allow coupon payments to the bond holders. More general jump diffusion models have been introduced to replace GBM in Merton's model. Lando [24] provides an overview of these extensions, which also have been considered in conjunction with default barriers.

2.3. Credit Ratings and Credit Value at Risk

Credit scoring has been used in the banking industry since the 1940s to assess the creditworthiness of loan applicants. The basic idea is to assign scores to loan applications based on their attributes such as gender, marital status, education level, income, profession, and credit history. Such scoring systems, called *scorecards*, are usually developed by in-house credit departments and/or consulting firms. One of the most widely used credit scoring systems was developed by Fair Isaac & Company (FICO), which built credit scoring systems for US banks and retail stores in the 1960s and 1970s and then expanded those systems to meet the needs of other industries and to evaluate the credit of borrowers. Credit scoring systems have now been developed for virtually all types of credit analysis, from consumer credit to commercial loans. The idea is to preidentify certain key factors that determine the probability of default and combine or weight them into a quantitative score.

Credit ratings of the creditworthiness of corporate bonds are provided by rating agencies such as Moody's, S&P, and Fitch. Using Moody's system, the ratings are Aaa, Aa, A, Baa, Ba, B, Caa, Ca, and C and "investment grade" ratings are Baa or above.

The corresponding S&P and Fitch ratings are AAA, AA, A, BBB, BB, B, CCC, CC, and C. For finer ratings, Moody's divides the Aa rating category into Aa1, Aa2, Aa3, the A category into A1, A2, A3, and so forth, and S&P and Fitch divide AA into AA+, AA, and AA−, A into A+, A, and A−, and so forth. Although methodologies and standards differ from one rating agency to another, regulators generally do not make distinctions among the agencies. Furthermore, although there is a strong similarity between the rating systems of Moody's and S&P, different agencies might assign slightly different ratings for the same bond. The ratings given by rating agencies are used by several credit risk software models such as CreditMetrics of J. P. Morgan, a system that evaluates risks individually or across an entire portfolio. Most corporations approach rating agencies to request a rating prior to sale or registration of a debt issue. S&P assigns and publishes ratings for all public corporate debt issues over $50 million, with or without a request from the issuer; but in all instances, S&P analytical staff first contact the issuer to call for meetings.

2.3.1. CreditMetrics, Credit Ratings Migration, and Product-Limit Estimators

CreditMetrics was introduced in 1997 by J. P. Morgan and its cosponsors (including KMV, UBS, Bank of America, and others). The system measures the risk and potential losses of loans and portfolios of bonds and loans over a period (e.g., a year) by simulation and analytic methods. In particular, it uses Monte Carlo simulations to estimate credit rating changes of all counterparties. Historical data provided by rating agencies on credit rating changes can be used to estimate the transition probability matrix for credit rating changes. In the Basel Accord for bank regulation, banks are required to construct these transition probability matrices based on their own data to stress test their portfolios. Some risk management systems such as CreditMetrics use these credit rating transition matrices for Monte Carlo simulations of credit losses due to credit ratings migration in a future period. A Markov chain model of ratings migration assumes a finite state space $\mathcal{S} = \{1, \ldots, K\}$ consisting of different rating classes, in which state 1 denotes the best credit rating class and state K represents the default case. The $K \times K$ probability transition matrix of the Markov chain is

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ p_{K-1,1} & p_{K-1,2} & \cdots & p_{K-1,K} \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \tag{2.29}$$

in which $p_{ii} = 1 - \sum_{j \neq i} p_{ij}$ for all i and p_{ij} represents the actual probability of moving to state j from initial rating state i in one time step. Note that the default state K is an absorbing state. The estimates of credit transition matrices published by rating agencies use this Markov chain framework.

Suppose there are N_i firms in a given rating category i at the beginning of the period and that N_{ij} firms migrate to the category j at the end of the period. An obvious estimate of p_{ij} is $\widehat{p}_{ij} = N_{ij}/N_i$ for $j \neq i$. However, there are serious difficulties with this estimate in practice. First, if no transition from rating class i to j is observed in the period, then the estimated transition probability $\widehat{p}_{ij}$ becomes 0. Second, the assumption of homogeneity over time is too strong for credit transition matrices. The product-limit estimator (2.10)

for the survival function has been generalized by Aalen [27], Aalen and Johansen [28], and Fleming [29] to estimate the transition probability matrix $\mathbf{P}(s, t)$ of a nonhomogeneous continuous-time Markov process on a finite state-space. Analogous to (2.11), the transition probability matrix from time s to t can be expressed as the product integral $\mathbf{P}(s, t) = \prod_{(s,t]}(\mathbf{I} + d\mathbf{A}(u))$, and the components A_{ij} of the cumulative transition intensity matrix can be estimated by the Nelson-Aalen estimator of the form (2.6); see Section IV.4 of [6].

To simulate the distribution of credit losses due to credit rating changes of two counterparties, say one with initial Aaa rating (state 1) and the other with initial B rating, over a 1-year period, we can use the estimate $\widehat{\mathbf{P}}$ of the transition matrix (2.29) to sample the new ratings for both counterparties in the following way. Without assuming the rating changes of the two counterparties to be independent, a Gaussian copula model, which will be described in Section 3.3, is typically used to simulate the joint probability distribution of the rating changes. Therefore, we first generate two standard normal random variables Z_A and Z_B with correlation ρ. Noting that $\Phi(Z_A)$ is uniform, the new rating of the Aaa company remains Aaa if $\Phi(Z_A) < \widehat{p}_{11}$, moves to states 2 (i.e., Aa) if $\widehat{p}_{11} \leq \Phi(Z_A) < \widehat{p}_{11} + \widehat{p}_{12}$, and so forth. This is tantamount to generating the state transitions of the Aaa company by the first row of the probability transition matrix $\widehat{\mathbf{P}}$. While the state transitions of the B company are likewise generated by the corresponding row of $\widehat{\mathbf{P}}$, the state transitions of the two companies are not independent and their dependence structure is inherited from the correlation of Z_A and Z_B. From the simulated credit rating changes of the two counterparties we can revalue their outstanding contracts to determine the total credit losses during the year. From a large number of such simulations, a probability distribution for credit losses is obtained, which can be used to evaluate credit VaR (Values at Risk) for risk management.

2.3.2. Credit VaR and a One-Factor Default Model

A $100(1 - \alpha)\%$ credit VaR over a period of T years is the $(1 - \alpha)$th quantile of the credit loss within the period. J. P. Morgan's CreditMetrics defines credit losses as those arising not only from defaults but also from credit ratings downgrade. This involves estimating the probability distribution of credit losses by Monte Carlo simulation of defaults and credit rating changes of all counterparties. Instead of Monte Carlo, a closed-form approximation, which does not involve downgrade in credit ratings, to credit VaR for a large portfolio of similar loans is used in the minimum capital requirement formulas of the Basel Committee on Banking Supervision. Suppose the portfolio consists of M loans in the same risk class. Letting τ_i denote the time to default of the ith loan, we can express the proportion of loans that default during the period as $\sum_{i=1}^{M} I_{\{\tau_i \leq T\}} / M$. Since the loans belong to the same risk class, their τ_i are identically distributed, with common distribution function G that is assumed to be continuous. Then $G(\tau_i)$ is uniform and $Z_i = \Phi^{-1}(G(\tau_i))$ is standard normal, where Φ is the standard normal distribution function. Vasicek [30, 31] and Schonbucher [32] assume the one-factor Gaussian model

$$Z_i = \sqrt{\rho}X + \sqrt{1 - \rho}\epsilon_i, \quad i = 1, \ldots, M, \tag{2.30}$$

in which X and ϵ_i are i.i.d. standard normal. Since $\tau_i \leq T$ if and only if $Z_i \leq \Phi^{-1}(G(T))$, it follows that

$$P(\tau_i \leq T \mid X) = P\left\{ \epsilon_i \leq \frac{\Phi^{-1}(G(T)) - \sqrt{\rho}X}{\sqrt{1-\rho}} \mid X \right\}$$

$$= \Phi\left(\frac{\Phi^{-1}(G(T)) - \sqrt{\rho}X}{\sqrt{1-\rho}} \right). \tag{2.31}$$

Conditional on X, $I_{\{\epsilon_i \leq [\Phi^{-1}(G(T)) - \sqrt{\rho}X]/\sqrt{1-\rho}\}}$ are i.i.d. and therefore by the strong law of large numbers,

$$M^{-1}\sum_{i=1}^{M} I_{\{\tau_i \leq T\}} = M^{-1}\sum_{i=1}^{M} I_{\{\epsilon_i \leq [\Phi^{-1}(G(T)) - \sqrt{\rho}X]/\sqrt{1-\rho}\}}$$

$$\longrightarrow E\left[I_{\{\epsilon_i \leq [\Phi^{-1}(G(T)) - \sqrt{\rho}X]/\sqrt{1-\rho}\}} \mid X \right] \tag{2.32}$$

$$= \Phi\left(\frac{\Phi^{-1}(G(T)) - \sqrt{\rho}X}{\sqrt{1-\rho}} \right)$$

as $M \to \infty$, with probability 1.

Let L be the size of the loan portfolio and R the recovery rate. The proportional loss given default is $1 - R$ and therefore the loss due to loan default in the period is $L(1 - R)\sum_{i=1}^{M} I_{\{\tau_i \leq T\}}/M$. Using the approximation (2.32), the $100(1-\alpha)\%$ credit VaR over the period is

$$L(1-R)\Phi\left(\frac{\Phi^{-1}(G(T)) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}} \right), \tag{2.33}$$

since the $(1-\alpha)$th quantile of X is $\Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$.

3. Credit Derivatives: An Overview

3.1. Pricing and Hedging with Credit Default Swaps

Credit default swaps (CDSs) have been described in the second paragraph of Section 1. The total amount paid per year by the buyer of the CDS, as a percentage of the notional principal per annum, is called the *CDS spread*. Suppose the reference bond has coupon dates $t_1, \ldots, t_M = T$ and the protection buyer pays a constant premium c at these dates prior to τ, the time of

the credit event. Assume that the face value of the bond is 1 and that the protection seller pays $1 - \delta$ in the case of a credit event. Then the value v_b of the premium leg of the swap is

$$
\begin{aligned}
v_b &= c\widehat{E}\left\{ \sum_{i=1}^{M} \exp\left(- \int_0^{t_i} r_s ds \right) I_{\{\tau > t_i\}} \right\} \\
&= c\widehat{E}\left\{ \sum_{i=1}^{M} \exp\left(- \int_0^{t_i} (r_s + \lambda_s) ds \right) \right\} = c \sum_{i=1}^{M} \pi(0, t_i),
\end{aligned}
\tag{3.1}
$$

where r_s is the short rate process, λ_s is the default intensity process, and $\pi(0, t_i)$ is the price of the defaultable bond with zero recovery and maturity t_i, using the same argument as that in (2.19). Similarly the value v_s of the default leg of the swap, to be paid by the protection seller, is

$$
\begin{aligned}
v_s &= (1 - \delta)\widehat{E}\left\{ \exp\left(- \int_0^{\tau} r_s ds \right) I_{\{\tau \leq T\}} \right\} \\
&= (1 - \delta) \int_0^T \widehat{E}\left\{ \exp\left(- \int_0^{\tau} r_s ds \right) \right\} \widehat{f}(t) dt,
\end{aligned}
\tag{3.2}
$$

where $\widehat{f}$ is the density function of τ under the risk-neutral measure Q, assuming that the short rate process and the default intensity process are independent under Q. Note that the survival function of τ under Q is

$$
\widehat{S}(t) = Q(\tau \geq t) = \widehat{E} \exp\left(- \int_0^t \lambda_s ds \right).
\tag{3.3}
$$

Letting $\widehat{\lambda} = \widehat{f}/\widehat{S}$ be the hazard function of τ under Q, it then follows from (3.2) that

$$
v_s = (1 - \delta) \int_0^T p(0, t)\widehat{\lambda}(t)\widehat{S}(t) dt,
\tag{3.4}
$$

where $p(0, t)$ is the price of the default-free bond; see (2.18). Note that under this independence assumption on the short rate and default intensity processes, we can further simplify $\pi(0, t_i)$ in (3.1) to

$$
\pi(0, t_i) = \widehat{E}\left[\exp\left(- \int_0^{t_i} r_s ds \right) \right] \widehat{E}\left[\exp\left(- \int_0^{t_i} \lambda_s ds \right) \right] = p(0, t_i)\widehat{S}(t_i).
\tag{3.5}
$$

In the absence of arbitrage, the premium c_{ds} of the default swap is the premium c for which $v_b = v_s$. Therefore, combining (3.1) with (3.4) yields

$$
c_{\text{ds}} = \frac{(1 - \delta) \int_0^T p(0, t)\widehat{S}(t)\widehat{\lambda}(t) dt}{\sum_{i=1}^{M} p(0, t_i)\widehat{S}(t_i)}.
\tag{3.6}
$$

3.2. Multiname Credit Derivatives

The contingent claim of a multiname credit derivative is on credit loss of the portfolio involving K reference entities, with respective default times $\tau^1, \ldots, \tau^K$. The default process N_t counts the number of defaults up to time t, while the loss process L_t records the cumulative financial loss due to default up to time t, for example,

$$N_t = \sum_{k=1}^{K} I_{\{\tau^k \leq t\}}, \qquad L_t = \sum_{k=1}^{K} l^k I_{\{\tau^k \leq t\}}, \tag{3.7}$$

where l^k is the loss due to default of the kth reference entity. Modeling the dependence of the intensity processes for the τ^i will be discussed in Section 3.3. The risk-neutral measure is used to price the derivative and the physical measure to predict future losses. Once the model is specified, one can use Monte Carlo simulations to evaluate the derivative price or the future expected loss.

3.2.1. Basket CDS

A basket CDS is an over-the-counter derivative involving a portfolio of n (typically between 5 and 15) reference entities. The protection buyer pays a premium, called the *basket spread* and expressed as a fraction of the notional principal per annum, until the kth credit event or maturity T, whichever is first; this is the premium leg of the basket CDS, similar to that of a regular CDS. The protection seller compensates the protection buyer for the credit loss associated with the kth credit event, constituting the default leg of the basket CDS. The fair basket spread equates the values of the premium and default legs, similar to a regular CDS. For $k = 1$, the basket swap is called a *first-to-default* CDS. For $k = n$, it is called an *add-up basket* CDS. More generally, we can have $1 \leq k \leq n$ for a *kth-to-default* CDS.

3.2.2. Credit Indices

Indices to track CDS spreads in credit markets have been developed by different providers who reached an agreement in 2004 to consolidate them. Two index portfolios that have become standard are CDX NA IG, which is a portfolio of 125 investment-grade companies in North America, and iTraxx, which is a portfolio of 125 investment-grade companies in Europe. An index portfolio consists of CDSs referenced on all companies in the portfolio, with common maturity T, common premium dates, and common notional principal. The principal leg consists of payments by the protection buyer that are proportional to the total notional principal on the surviving companies at the premium dates. The default leg is a stream of payments by the protection seller that cover the portfolio losses up to T as they occur.

3.2.3. Synthetic or Cash CDO

In a collateralized debt obligation (CDO), the sponsor chooses a portfolio of companies and a maturity T. The sponsor of a *synthetic* CDO sells CDS protection for each company in the portfolio with the CDS maturities equal to T. The synthetic CDO principal is the total

of the notional principals underlying the CDSs. The CDO sponsor issues a collection of tranches to investors and sets different rules for determining the cash inflows and outflows for different tranches. The tranche holders sell protection to the CDO sponsor and thereby earn the specified spreads on the outstanding tranche principal, but are responsible for the sponsor's payouts to the CDS protection buyers when the companies in the portfolio default. As an illustration, suppose there are three tranches: equity, mezzanine, and senior. The equity tranche earns the largest spread of the three tranches, say 1000 basis points per year, on the outstanding tranche principal. They are responsible for all payouts on the CDS up to a certain percentage, say 5%, of the CDO principal. Thus, if the equity, mezzanine, and senior tranches have initial capital of $5 million, $15 million, and $80 million, respectively, and the defaults by companies in the portfolio lead to payouts of $2 million after 1 year, then the equity tranche holders are responsible for these payouts and the equity tranche capital reduces to $3 million.

A *cash* CDO is created from a bond portfolio, and the sponsor issues a collection of tranches and requires an initial investment from the tranche holders to finance the underlying corporate bonds. Rules are designed to ensure that if one tranche is more senior than another it is more likely to receive the promised interest payments and repayments of principal, recognizing that there may be credit losses in these defaultable bonds.

3.3. Correlated Default Intensities of Multiple Obligors

For a multiname credit derivative involving k firms, it is important to model not only the individual default intensity process but also the joint distribution of these processes. Finding tractable models that can capture the key features of the interrelationships of the firms' default intensities has been an active area of research. The following are some traditional modeling approaches described by [24].

3.3.1. Gaussian Copula Approach

Let $\mathbf{X} = (X_1, \ldots, X_m^T)$ be a continuous random vector with given marginal distribution functions F_i for X_i. Since $U_i = F_i(X_i)$ is uniformly distributed, we can specify the joint distribution of $\mathbf{X}$ by a *copula function*, which is the joint distribution function of m uniform random variables. Specifically, the copula function $C(u_1, \ldots, u_m)$ defined on $0 \leq u_i \leq 1$ ($1 \leq i \leq m$) can be used to provide a joint distribution function of $\mathbf{X}$ via

$$
\begin{aligned}
C(F_1(x_i), \ldots, F_m(x_m)) &= P\{F_1(X_1) \leq F_1(x_1), \ldots, F_m(X_m) \leq F_m(x_m)\} \\
&= P\{X_1 \leq x_1, \ldots, X_m \leq x_m\}.
\end{aligned}
\tag{3.8}
$$

Using the representation

$$
\tau_i = \inf\left\{ t : \exp\left(-\int_0^t \lambda^i(\mathbf{X}_s) ds \right) \leq U_i \right\},
\tag{3.9}
$$

in which $U_i = e^{-\epsilon_i}$ is uniformly distributed on $[0, 1]$ (since ϵ_i is exponential with mean 1) and is independent of $\{\mathbf{X}_s, s \geq 0\}$, we can use a copula function on $[0, 1]^M$ to model the joint distribution of $U_1, \ldots, U_M$ and thereby also of $\tau_1, \ldots, \tau_M$ via (3.9).

Let G_i be the distribution function of default time τ_i for the ith obligor, $1 \le i \le M$. Then $Z_i = \Phi^{-1}(G_i(\tau_i))$ is standard normal. A Gaussian copula model assumes that $(Z_1, \ldots, Z_M)$ is multivariate normal and specifies its correlation matrix Γ. Li [33] introduced it in 2000 to model default correlations, and it quickly became a widely used tool to price CDOs and other multiname credit derivatives that were previously too complex to price. A convenient way is to use the correlations of the asset returns to specify Γ. A better way is to use a factor model

$$Z_i = \boldsymbol{\beta}_i^T \mathbf{X} + \sigma_i \epsilon_i \tag{3.10}$$

to generate Γ, in which $\mathbf{X}$ is a multivariate normal vector of common factors affecting defaults and ϵ_i is standard normal independent of $\mathbf{X}$. A special case is the one-factor model (2.30).

3.3.2. Dependence among Default Intensities

Whereas the factor model (3.10) is applied to the Gaussian $Z_i = \Phi^{-1}(G_i(\tau_i))$ that involves the distribution function G_i of τ_i, a more direct approach is to decompose the default intensity process λ_t^i of the ith firm as $\lambda_t^i = \mu_t + \nu_t^i$, in which μ_t is the default intensity of a common factor and ν_t^i is that of an idiosyncratic component such that $\mu_t, \nu_t^1, \ldots, \nu_t^M$ are independent processes. Instead of the preceding additive decomposition of λ_i^t, an alternative is to use a common multiplicative factor ξ so that $\lambda_t^i = \xi \nu_t^i$, in which ξ is a positive random variable (e.g., exponential) that is independent of the independent processes $\nu_t^1, \ldots, \nu_t^M$.

3.3.3. Mixture Models for Default Indicators

Vasicek [34] introduced a mixed binomial model to study the distribution of the number of defaults in a loan portfolio over a time horizon T. It played an important role in the Basel II rules and inspired subsequent work on using mixture models and factor variables; see [35, 36]. The mixed binomial model assumes that the indicator variables $X_i = I_{\{\tau_i \le T\}}$ are conditionally independent Bernoulli (p_i) given $p_1, \ldots, p_M$ and that the p_i are random variables having the same distribution Π with mean p and such that $\text{Cov}(p_i, p_j) = \tau$ for $i \ne j$. Then

$$E(X_i) = p, \quad \text{Var}(X_i) = p(1 - p), \quad \text{Cov}(X_i, X_j) = \tau \quad \text{for } i \ne j. \tag{3.11}$$

Therefore, the mixing distribution induces correlations among the M Bernoulli random variables that are conditionally independent given the unknown probabilities p_i that have common prior distribution Π. The mean and variance of the number of defaults up to time T, denoted by #, are

$$E(\#) = Mp, \qquad \text{Var}(\#) = Mp(1 - p) + M(M - 1)\tau. \tag{3.12}$$

4. Frailty and Contagion: HMM and Point Process Models

As noted in Section 3.3, the Gaussian copula was popular because of its ease of use. However, there is no convincing argument to connect the asset return correlations to the correlations of

the normally distributed transformed default times. In a 2009 commentary on "the biggest financial meltdown since the Great Depression," Salmon [37] mentioned that the Gaussian copula approach, which "looked like an unambiguously positive breakthrough," "was adopted by everybody from bond investors and Wall Street banks to rating agencies and regulators" and "became so deeply entrenched—and was making people so much money— that warnings about its limitations were largely ignored." In the wake of the financial crisis, it was recognized that better albeit less tractable models of correlated default intensities are needed for pricing CDOs and risk management of credit portfolios. It was also recognized that such models should include relevant firm-level and macroeconomic variables for default prediction and also incorporate frailty and contagion mentioned in Section 1.

4.1. Intensity Processes with Common Frailty

The mixture binomial model for the indicator variables $I_{\{\tau_i \leq T\}}$ with unknown means p_i, allowing "random effects" in the subject-specific probabilities p_i, described in Section 3.3 has more general analogs in survival analysis under the rubric of "frailty." In the case of the proportional hazards regression model (2.13), random effects can be introduced by including a positive random factor Y in $\lambda(t) = Y\lambda_0(t)\exp(\boldsymbol{\beta}^T \mathbf{x}(t))$, or equivalently, by including i.i.d. random effects Z_i in

$$\lambda_i(t) = \lambda_0(t)\exp\left(Z_i + \boldsymbol{\beta}^T \mathbf{x}_i(t)\right), \quad 1 \leq i \leq n. \tag{4.1}$$

Such ideas first arose in demography, in which Vaupel et al. [38] and Manton et al. [39, 40] discussed the impact of heterogeneity on the dynamics of mortality and introduced the term "frailty" for the latent random variable $Y = e^Z$ to account for the heterogeneity, which they model with a gamma distribution. Subsequently, Heckman and Singer [41, 42] and Hougaard [43, 44] considered other distributions to model frailty and applications to econometrics and biostatistics. Clayton and Cuzick [45] use the frailty approach to give an explicit formulation of the joint survival function with given marginals and to derive multivariate generalizations of survival analysis.

Taking $\lambda_0 \equiv 1$, Duffie et al. [46] propose an enhancement of (4.1) by including another unobserved stochastic process $z_{\boldsymbol{\theta}}(t)$, leading to the *dynamic frailty model*

$$\lambda_i(t) = \exp\left(Z_i + z_{\boldsymbol{\theta}}(t) + \boldsymbol{\beta}^T \mathbf{x}_i(t)\right), \quad 1 \leq i \leq n, \tag{4.2}$$

in which $z_{\boldsymbol{\theta}}(t)$ is an affine process, with unknown parameter vector $\boldsymbol{\theta}$, of the type used in (2.19). Because the process $Z_i + z_{\boldsymbol{\theta}}(t)$ undergoing Markovian dynamics is not directly observable and the observations are $(T_i, \delta_i, \mathbf{x}_i(t))$, $t \leq T_i$, $1 \leq i \leq n$, as in Section 2.1, (4.2) is a *hidden Markov model* (HMM) for which parameter estimation and prediction of future outcomes are much harder than the likelihood or partial likelihood methods in (2.14) or (2.15).

4.1.1. State and Parameter Estimation in HMM

Duffie et al. [46] use (4.2) to model default of US public nonfinancial firms, carrying out an empirical study of 2793 companies for the period from January 1979 to March 2004.

Fitting the model involves estimation of not only the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ but also the unobservable states $z_{\boldsymbol{\theta}}(t)$ and Z_i. Mergers and other exits of the companies are regarded as censoring variables. Estimation of the posterior distribution of $\lambda_i(t)$ for $0 \leq t \leq T_i$ given the observations is called the "smoothing" problem in HMM, in contrast to the "filtering" problem that involves observations up to time t; see [47]. Duffie et al. [46] use the stochastic EM algorithm [48, 49] to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ by maximum likelihood, noting that the complete data likelihood involves Z_i and $z_{\boldsymbol{\theta}}(t)$ that are not observed. Replacing $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ by their maximum likelihood estimates, they use Markov chain Monte Carlo (MCMC) to evaluate the posterior distribution of $\lambda_i(t)$, $0 \leq t \leq T_i$, $i = 1, \ldots, n$. The MCMC scheme they use involves both Gibbs sampling and random walk Metropolis-Hastings steps [50, 51]. Because of the computational complexity in the stochastic EM and MCMC schemes, they choose a simple Ornstein-Uhlenbeck process $dz_{\boldsymbol{\theta}}(t) = -\theta z_{\boldsymbol{\theta}}(t)dt + dw(t)$, with a univariate parameter θ and standard Brownian motion $w(t)$, for the affine process. The covariate vector $\mathbf{x}_i(t)$ in (4.2) is the same as used earlier by [52] without the frailty terms Z_i and $z_{\boldsymbol{\theta}}(t)$. It includes the following firm-specific covariates $x_{ij}(t)$ and macroeconomic covariates $u_k(t)$:

(i) the firm's distance to default,

(ii) the firm's trailing 1-year stock return,

(iii) the 3-month Treasury bill rate, which is a macroeconomic variable "consistent with the effect of a monetary policy that lowers short-term interest rates when the economy is likely to perform poorly,"

(iv) the trailing 1-year return on the S&P 500 index.

4.2. Contagion and Point Process Models

Davis and Lo [53] introduced a simple model to reflect default contagion among n homogeneous firms. Suppose prior to any default all firms have default intensity λ. After a default occurs, the default intensities of the surviving firms increase by a common amount μ. This is an example of the more general "top-down" approach to modeling portfolio loss and contagion. Consider a bond portfolio involving n companies and order their default times by $\tau_{(1)} \leq \tau_{(2)} \leq \cdots \leq \tau_{(n)}$. The number of defaults up to time t is the point process $N_t = \sum_{i=1}^{n} I_{\{\tau_{(i)} \leq t\}}$. The top-down approach models directly the intensity of the point process N_t. In contrast, the "bottom-up" approach models the intensities $\lambda_i(t)$ of the individual default times τ_i and aggregates them into the point process N_t with intensity $\sum_{i=1}^{n} \lambda_i(t) I_{\{\tau_i \geq t\}}$. One such contagion model is $\lambda_i(t) = \lambda_i^0(t) + \boldsymbol{\beta}_i \mathbf{N}_t$, where $\mathbf{N}_t = (I_{\{\tau_1 \leq t\}} \cdots I_{\{\tau_n \leq t\}})^T$, β_{ij} measures the increase in hazard for firm i after firm j defaults, and λ_i^0 is the Cox-Ingersoll-Ross process for default intensity that is given by

$$d\lambda_i^0(t) = \kappa_i\left(\theta_i - \lambda_i^0(t)\right)dt + \sigma_i\sqrt{\lambda_i^0(t)}dw_i(t), \tag{4.3}$$

in which $w_i(t)$, $t \geq 0$, is standard Brownian motion; see [54].

Additive regression models for hazard functions are often used to model contagion in biostatistics, generalizing the simple contagion model of [53] described above. Aalen [55] and Huffer and McKeague [56] provide the statistical counterparts of the multiplicative hazard model (2.13).

5. Insurance Contracts, Credibility Theory, and Linear EB

Credibility theory for pricing insurance contracts in actuarial science uses historical data from an individual contract and the risk class to which it belongs to arrive at a *credibility premium* of the form

$$\alpha\overline{Y} + (1 - \alpha)\mu, \quad \text{with } 0 < \alpha < 1, \tag{5.1}$$

where $\overline{Y}$ is the observed mean claim amount for the individual contract and μ is the overall mean claim per contract in the class during the previous period; see [57]. It can be derived as a Bayes estimate of θ under the assumption that $\overline{Y}$ is normal (resp., gamma, binomial, or Poisson) with mean θ and that θ has a prior normal (resp., inverse gamma, beta, or gamma) distribution with mean μ, for which the form of α is explicitly determined by the distributional parameters. Without assuming a prior distribution with specified parameters, Bühlmann [58, 59] considered the best linear estimator of the form $\alpha\overline{Y} + b$ based on observed claims $Y_1, \ldots, Y_n$. The best linear estimator involves the mean and variance of the Y_i, which can be estimated consistently by the method of moments, as will be explained below.

5.1. Linear Empirical Bayes and Credibility Factors

The empirical Bayes (EB) methodology introduced by Robbins [1] for n independent and structurally similar problems of statistical inference on unknown parameters θ_i based on observed data Y_i ($i = 1, \ldots, n$), where Y_i has probability density $f(y \mid \theta_i)$ and θ_i and Y_i may be vectors. The θ_i are assumed to have a common prior distribution G that has unspecified hyperparameters. Letting $d_G(y)$ be the Bayes decision rule (with respect to some loss function and assuming known hyperparameters) when $Y_i = y$ is observed, the basic principle underlying empirical Bayes is that d_G can often be consistently estimated from $Y_1, \ldots, Y_n$, leading to the empirical Bayes decision rule $d_{\widehat{G}}$. Thus, the n structurally similar problems can be pooled to provide information about unspecified hyperparameters in the prior distribution, thereby yielding $\widehat{G}$ and the decision rules $d_{\widehat{G}}(Y_i)$ for the independent problems. In particular, Robbins [1] considered Poisson Y_i with mean θ_i, as in the case of the number of accidents by the ith driver in a sample of size n (in a given year) from a population of drivers, with distribution G for the accident-proneness parameter θ. In this case the Bayes estimate (with respect to squared error loss) of θ_i when $Y_i = y$ is observed is

$$d_g(y) = \frac{(y+1)g(y+1)}{g(y)}, \quad y = 0, 1, \ldots, \tag{5.2}$$

where $g(y) = \int_0^\infty \theta^y e^\theta dG(\theta)/(y!)$. Using $\widehat{g}(k) = n^{-1} \sum_{i=1}^n I_{\{Y_i=k\}}$ to replace $g(k)$ in (5.2) yields the (EB) estimate $d_{\widehat{g}}(y)$. Stein [2] and subsequently James and Stein [60] considered the special case $Y_i \sim N(\theta_i, \sigma^2)$ with known σ but unknown θ_i.

Unlike the pricing of credit derivatives, credibility theory for pricing insurance contracts uses the physical (real-world) measure rather than the risk neutral measure. This extends to insurance derivatives such as CAT (catastrophic) bonds used by the insurance industry to hedge exposures to catastrophic risks due to hurricanes, earthquakes, or other severe unexpected (rare) events. For example, an insurance company that wants protection

for California earthquake losses between \$30 million and \$40 million can issue CAT bonds with \$10 million total principal. In the event that the losses exceed \$30 million, bondholders would lose some or all of their principal, or only their promised interest if the insurance company has made a bond issue large enough to cover this excess cost layer. Like insurance contracts, insurance derivatives are priced under the physical measure, using historical or actuarial data to estimate the parameters of future earthquake or other catastrophic losses. The value of the estimated expected payoff is discounted at the risk-free rate to price an insurance derivative; the uncertainties of their payoffs are not priced by the market.

Robbins [61] gives a general description of linear EB models, which include Stein's normal model as a special case and which is also related to Bühlmann's linear estimation approach to credibility premium. Suppose the decision rule $d(y)$ is restricted to linear functions of the form $A + By$ and assume that the family of density functions $f(y \mid \theta)$ satisfies

$$E(Y \mid \theta) = \theta, \qquad \text{Var}(Y \mid \theta) = a + b\theta + c\theta^2 \tag{5.3}$$

for some constants a, b, and $c \neq -1$. Then the Bayes rule $d_G(y)$ that minimizes the Bayes risk, with respect to squared error loss, among linear estimators of θ is given by the linear regression formula

$$d_G(y) = E\theta + \frac{\text{Cov}(\theta, Y)}{\text{Var}(Y)}(y - EY). \tag{5.4}$$

Moreover, from (5.3), it follows that

$$EY = E\theta, \qquad \text{Cov}(\theta, Y) = \text{Var}(\theta), \tag{5.5}$$

$$\text{Var}(Y) = E[\text{Var}(Y \mid \theta)] + \text{Var}[E(Y \mid \theta)]$$
$$= a + bE\theta + c(E\theta)^2 + (c + 1)\text{Var}(\theta), \tag{5.6}$$

and therefore

$$\text{Var}(\theta) = \frac{\text{Var}(Y) - \left\{a + bEY + c(EY)^2\right\}}{c + 1}. \tag{5.7}$$

Since $(\theta_1, Y_1), \ldots, (\theta_n, Y_n)$ are i.i.d. vectors from the distribution under which θ_i has prior distribution G and Y_i has conditional density $f(\cdot \mid \theta_i)$ given Y_i, the parameters $E\theta$, $\text{Cov}(\theta, Y)$, $\text{Var}(Y)$, and EY in the Bayes rule (5.4) can be estimated consistently by using (5.5), (5.7), and the method of moments. This yields the linear EB estimate

$$\widehat{d}(y) = \overline{Y} + \left[1 - \frac{cs^2 + a + b\overline{Y} + c\overline{Y}^2}{(c + 1)s^2}\right]_+ (y - \overline{Y}), \tag{5.8}$$

where $x_+ = \max(x, 0)$, $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ estimates the mean EY, and $s^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2/(n - 1)$ estimates the variance $\text{Var}(Y)$. In the case $Y_i \mid \theta_i \sim N(\theta_i, \sigma^2)$ with known σ^2, (5.3) holds with

$a = \sigma^2$ and $b = c = 0$, for which (5.8) corresponds to a variant of the James-Stein [60] estimator of $(\theta_1, \ldots, \theta_n)^T$ which has a smaller mean squared error than the maximum likelihood estimator $(Y_1, \ldots, Y_n)^T$. Moreover, if G is normal, then (5.4) minimizes the Bayes risk among all (not necessarily linear) estimators of θ.

The linear EB estimate (5.8) was used by Bühlmann [58] to determine the *credibility factor* α in (5.1), writing $\alpha = \alpha_n$ to indicate its dependence on the training sample $Y_1, \ldots, Y_n$ of size n. The monograph by Bühlmann and Gisler [62] describes a variety of extensions of (5.8) to more general settings. For example, suppose there are I risk classes and Y_{ij} denotes the jth claim of the ith class, with $E(Y_{ij} \mid \theta_i) = \theta_i$ and $\mathrm{Var}(Y_{ij} \mid \theta_i) = \sigma_i^2, 1 \le j \le n_i, 1 \le i \le I$. Assuming a normal prior $N(\mu, \tau^2)$ for θ_i, the Bayes estimate of θ_i is

$$E(\theta_i \mid Y_{i1}, \ldots, Y_{i,n_i}) = \alpha_i \overline{Y}_i + (1 - \alpha_i)\mu, \tag{5.9}$$

where $\alpha_i = \tau^2/(\tau^2 + \sigma_i^2/n_i)$ and $\overline{Y}_i = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$. Since $E(Y_{ij}) = E(\theta_i) = \mu$ and $\mathrm{Var}(Y_{ij}) = \mathrm{Var}(\theta_i) + E(\mathrm{Var}(Y_{ij} \mid \theta_i)) = \tau^2 + \sigma_i^2$, we can estimate μ, σ_i^2, and τ^2 by the method of moments when $n_i > 1$: $\hat{\mu} = (\sum_{i=1}^{I} \sum_{j=1}^{n_i} Y_{ij}) / \sum_{i=1}^{I} n_i$, $\hat{\sigma}_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i)^2/(n_i - 1)$, and $\hat{\tau}^2 = \sum_{i=1}^{I} n_i(\overline{Y}_i - \hat{\mu})^2 / \sum_{i=1}^{I} n_i$. Replacing μ, σ_i^2, and τ^2 in (5.9) by $\hat{\mu}$, $\hat{\sigma}_i^2$, and $\hat{\tau}^2$ yields the EB estimate

$$\hat{\theta}_i = \hat{\alpha}_i \overline{Y}_i + (1 - \hat{\alpha}_i)\hat{\mu}, \tag{5.10}$$

where $\hat{\alpha}_i = \hat{\tau}^2/(\hat{\tau}^2 + \hat{\sigma}_i^2/n_i)$ is the credibility factor for the ith class. Another important extension, introduced by Hachemeister [63], is the credibility regression model that relates claim sizes to certain covariates; the credibility factor is a matrix in this case.

Frees et al. [64, 65] unified various credibility models into the framework of linear mixed models (LMM) of the form

$$Y_{ij} = \boldsymbol{\beta}^T \mathbf{x}_{ij} + \mathbf{b}_i^T \mathbf{z}_{ij} + \epsilon_{ij}, \tag{5.11}$$

with fixed effects $\beta_1, \ldots, \beta_p$ forming the vector $\boldsymbol{\beta}$, subject-specific random effects $b_{i1}, \ldots, b_{iq}$ forming the vector $\mathbf{b}_i$ such that $E(\mathbf{b}_i) = \mathbf{0}$, and zero-mean random disturbances ϵ_{ij} that have variance σ^2 and are uncorrelated with the random effects and the covariates $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$. In particular, the credibility regression model $Y_i = \boldsymbol{\alpha}_i^T \mathbf{x}_i + \epsilon_i$, in which the subject-specific regression parameters $\boldsymbol{\alpha}_i$ have distribution G, is a special case with $j = 1 = n_i$, $\mathbf{z}_i = \mathbf{x}_i$, $\boldsymbol{\beta}$ being the mean of G, and $\mathbf{b}_i = \boldsymbol{\alpha}_i - \boldsymbol{\beta}$.

6. Dynamic Empirical Bayes and Generalized Linear Mixed Models

Linear credibility theory, with a credibility premium of the form (5.1), has been extended in two directions. The first is linear *evolutionary credibility* that assumes a dynamic linear Bayesian model for the prior means over time; see [62]. The second is *exact credibility* theory for which the claims are assumed to belong to an exponential family of distributions so that

generalized linear (or generalized linear mixed) models are used in lieu of linear regression (or linear mixed) models; see [66–72] that assume Y_{ij} to have a density function of the form

$$f(y \mid \mathbf{b}_i, \mathbf{z}_{it}, \mathbf{x}_{it}) = \exp\left\{ \frac{[y\theta_{it} - \psi(\theta_{it})]}{\sigma} + c(y, \sigma) \right\}, \tag{6.1}$$

in which σ is a dispersion parameter and $\mu_{it} = d\psi/d\theta|_{\theta=\theta_{it}}$ satisfies $\mu_{it} = g^{-1}(\boldsymbol{\beta}^T \mathbf{x}_{it})$ in the case of generalized linear models and

$$\mu_{it} = g^{-1}\left(\boldsymbol{\beta}^T \mathbf{x}_{it} + \mathbf{b}_i^T \mathbf{z}_{it}\right) \tag{6.2}$$

in the case of generalized linear mixed models (GLMM), where g^{-1} is the inverse of a monotone link function g. The case $g = d\psi/d\theta$ is called the "canonical link." The random effects $\mathbf{b}_i$ can contain an intercept term a_i by augmenting the covariate vector to $(1, \mathbf{z}_{it}^T)$ in case a_i is not included in $\mathbf{b}_i$; $\boldsymbol{\beta}$ is a vector of fixed effects and can likewise contain an intercept term. The density function (6.1) with $\sigma = 1$ is that of an exponential family, which includes the Bernoulli and normal distributions as special cases.

6.1. Linear Evolutionary Credibility, Kalman Filtering, and Dynamic EB

Noting that insurers observe claims of risk classes over successive periods, Frees et al. [64, 65] in fact consider the setting of longitudinal data in their LMM approach to credibility theory; the index j in (5.11) is replaced by t that denotes time so that the Y_{it} represent longitudinal data of claims. Bühlmann and Gisler [62] go further in assuming a dynamic Bayesian model in which the mean θ_{it} of Y_{it} has a prior distribution with mean μ_t that changes with time according to the AR(1) model

$$\mu_t = \rho\mu_{t-1} + (1 - \rho)\mu + \eta_t, \tag{6.3}$$

with i.i.d. zero-mean η_t having variance V. Thus, the μ_t are latent states undergoing AR(1) dynamics, yielding a linear state-space model for which the unobserved states μ_t can be estimated from the observations Y_{is}, $s \leq t$, by the Kalman filter $\widehat{\mu}_{t|t}$ given by the recursion

$$\widehat{\mu}_{t|t} = \widehat{\mu}_{t|t-1} + \rho^{-1}\mathbf{K}_t\left(\mathbf{Y}_t - \widehat{\mu}_{t|t-1}\mathbf{1}\right), \qquad \widehat{\mu}_{t+1|t} = \rho\widehat{\mu}_{t|t} + (1 - \rho)\mu, \tag{6.4}$$

in which $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{nt})^T$, $\mathbf{1} = (1, \ldots, 1)^T$, and $\mathbf{K}_t$ is the $1 \times n$ Kalman gain matrix defined recursively in terms of the hyperparameters $V = \text{Var}(\eta_t)$, $v_t = \text{Var}(Y_{it} \mid \mu_t)$, and ρ by

$$\mathbf{K}_t = \rho\Sigma_{t|t-1}\mathbf{1}^T\left(\Sigma_{t|t-1}\mathbf{1}\mathbf{1}^T + v_t\mathbf{I}\right)^{-1}, \tag{6.5}$$

$$\Sigma_{t+1|t} = \left(\rho - \mathbf{K}_t\mathbf{1}\right)^2\Sigma_{t|t-1} + V + v_t\mathbf{K}_t\mathbf{K}_t^T. \tag{6.6}$$

Under the assumption that $\theta_{it} \mid \mu_t$ has variance τ^2 and $Y_{it} \mid \theta_{it}$ has variance $a + b\theta_{it} + c\theta_{it}^2$ as in (5.3), $v_t = \tau^2 + a + b\theta_{it} + c\theta_{it}^2$. The Kalman filter is the minimum-variance linear estimator of μ_t; it is also the Bayes estimator if Y_{it} (conditional on μ_t) and η_t are normal.

As in the linear EB approach described in Section 5, we can use the method of moments to estimate μ_t and v_t for every $1 \le t \le T$. Specifically, $\overline{Y}_t = n^{-1} \sum_{i=1}^{n} Y_{it}$ is a consistent estimate of μ_t and $s_t^2 = \sum_{i=1}^{n}(Y_{it} - \overline{Y}_t)^2/(n-1)$ is a consistent estimate of v_t. Moreover, assuming $E(\mu_0) = \mu$, it follows from (6.3) that $E(\mu_t) = \mu$ for all t. Therefore, μ can be consistently estimated from the observations up to time t by $\widehat{\mu}(t) = (\sum_{s=1}^{t} \overline{Y}_s)/t$, and ρ and V can be consistently estimated by

$$\widehat{\rho}(t) = \frac{\sum_{s=2}^{t} \left(\overline{Y}_s - \widehat{\mu}(t)\right)\left(\overline{Y}_{s-1} - \widehat{\mu}(t)\right)}{\sum_{s=2}^{t} \left(\overline{Y}_s - \widehat{\mu}(t)\right)^2},$$

$$\widehat{V}(t) = \frac{\sum_{s=2}^{t} \left\{\overline{Y}_s - \rho\overline{Y}_{s-1} - (1-\rho)\widehat{\mu}(t)\right\}^2}{(t-2)},$$

(6.7)

for $t \ge 2$. Replacing the hyperparameters μ, ρ, V, and v_t in the Bayes estimate $\widehat{\mu}_{t|t}$ of μ_t by their estimates $\widehat{\mu}(t)$, $\widehat{\rho}(t)$, $\widehat{V}(t)$, and s_t^2 yields the EB estimate of μ_t. Of particular interest in applications of longitudinal data is the prediction of $Y_{i,t+1}$ based on the set of observations $\mathcal{Y}_t$ up to time t. The Bayes predictor, assuming normal priors and normal observations, is

$$E(Y_{i,t+1} \mid \mathcal{Y}_t) = E(\theta_{i,t+1} \mid \mathcal{Y}_t) = E(\mu_{t+1} \mid \mathcal{Y}_t) = \widehat{\mu}_{t+1|t}, \tag{6.8}$$

which is the Kalman predictor given by (6.4). This is also the best linear predictor without assuming normality. The corresponding EB predictor replaces the hyperparameters in the Kalman predictor by their method-of-moment estimates $\widehat{\mu}(t)$, $\widehat{\rho}(t)$, $\widehat{V}(t)$, and s_t^2.

Note that to estimate the hyperparameters in this EB approach via (6.3) and Kalman filtering, we have used the cross-sectional mean $\overline{Y}_{t-1}$ of n independent observations that have mean μ_{t-1}. Lai and Sun [73] propose an alternative and more direct approach that replaces μ_{t-1} in (6.3) by $\overline{Y}_{t-1}$, leading to

$$\mu_t = \rho\overline{Y}_{t-1} + \omega + \eta_t \tag{6.9}$$

with $\omega = (1 - \rho)\mu$, which can be written as an LMM for the longitudinal data Y_{ij}:

$$Y_{it} = \rho\overline{Y}_{t-1} + \omega + b_i + \epsilon_{it}, \tag{6.10}$$

by absorbing η_t into ϵ_{it}. Since (6.10) is in the form of a regression model, Lai and Sun [73] also include additional covariates to increase the predictive power of the model in the LMM

$$Y_{it} = \rho\overline{Y}_{t-1} + a_i + \boldsymbol{\beta}^T \mathbf{x}_{it} + \mathbf{b}_i^T \mathbf{z}_{it} + \epsilon_{it}, \tag{6.11}$$

where a_i and $\mathbf{b}_i$ are subject-specific random effects, $\mathbf{x}_{it}$ represents a vector of subject-specific covariates that are available prior to time t (for predicting Y_{it} prior to observing it at time t),

and $\mathbf{z}_{it}$ denotes a vector of additional covariates that are associated with the random effects $\mathbf{b}_i$. A simulation study in [71] to compare the performance of parameter estimation and one-step-ahead prediction using (6.11) with that of Kalman filtering shows that the dynamic LMM (6.11) yields results comparable to those of Kalman filtering when the data are generated by a linear state-space model of the type (6.5), the parameters of which are not assumed to be known and are estimated by (6.7).

6.2. Dynamic EB Modeling via GLMM

Lai and Sun [73] assume the prior distribution for which the μ_{it} are i.i.d. with mean μ_t and

$$g(\mu_t) = \sum_{j=1}^{p} \theta_j g\left(\overline{Y}_{t-j}\right) \quad \text{for } 1 \le t \le T, \text{ where } \overline{Y}_s = n^{-1} \sum_{i=1}^{n} Y_{is}. \tag{6.12}$$

The dynamic model (6.12) for μ_t is an EB version of the Markov model introduced by Zeger and Qaqish [74], who consider a single time series Y_t that has a density function of the form (6.1) and models μ_t by $g(\mu_t) = \sum_{j=1}^{p} \theta_j g(Y_{t-j})$, where g is a link function and $n = 1$ in this case. For general n, note that μ_s is the mean of μ_{is} and can be consistently estimated by $\overline{Y}_s$, which is the basic idea underlying the linear EB approach. This suggests using $g(\overline{Y}_{t-j})$ in lieu of $g(Y_{t-j})$ for $n \ge 1$ that leads to (6.12). As in the LMM (6.11), Lai and Sun [73] propose to increase the predictive power of the model by including fixed and random effects and other time-varying covariates of each subject i, thereby removing the dependence of $g(\mu_{it}) - g(\mu_t)$ on t in the GLMM

$$g(\mu_{it}) = \sum_{j=1}^{p} \theta_j g\left(\overline{Y}_{t-j}\right) + a_i + \boldsymbol{\beta}^T \mathbf{x}_{it} + \mathbf{b}_i^T \mathbf{z}_{it}, \tag{6.13}$$

in which $\theta_1, \ldots, \theta_p$ and $\boldsymbol{\beta}$ are the fixed effects and a_i and $\mathbf{b}_i$ are subject-specific random effects. Following [4], it is assumed that a_i and $\mathbf{b}_i$ are independent normal with zero means. Denote the unknown vector specifying the covariances of the random effects by $\boldsymbol{\alpha}$, and for notational simplicity, we can augment $\mathbf{b}_i$ to include a_i so that (6.13) can be written as $g(\mu_{it}) = \sum_{j=1}^{p} \theta_j g(\overline{Y}_{t-j}) + \mathbf{x}_{it}^T \boldsymbol{\beta} + (1, \mathbf{z}_{it}^T) \mathbf{b}_i$. Note that (6.13) is a natural extension of the dynamic linear EB model (6.11) to the exponential family. By including $g(\overline{Y}_{t-j})$, $j = 1, \ldots, p$, in $\mathbf{x}_{it}$, we can implement the dynamic EB model (6.13) by using the methods described for general GLMMs.

6.3. Implementation and Model Selection

Breslow and Clayton assume the $\mathbf{b}_i$ in (6.2) to have a common normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}$ that depends on an unknown parameter. Lai and Shih [75] have developed a nonparametric method to estimate the random effects distribution in the more general setting of nonlinear mixed effects models and have shown that there is very low resolution in estimating the distribution. This provides a heuristic explanation for why the choice of a normal distribution, with unspecified parameters, for the random effects distribution G does not result in a worse estimate of $\boldsymbol{\beta}$ and σ than

the semiparametric approach that estimates G nonparametrically. An advantage of assuming a normal distribution for the random effects $\mathbf{b}_i$ is that it only involves the covariance matrix of the random effects. The likelihood function of the GLMM defined by (6.1) and (6.2) is of the form $\prod_{i=1}^{n} L_i(\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$, where

$$L_i(\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \int \left\{ \prod_{t=1}^{T} f(y_{it}; \theta_{it}, \sigma) \right\} \phi_{\boldsymbol{\alpha}}(\mathbf{b}) d\mathbf{b}, \qquad (6.14)$$

in which $\phi_{\boldsymbol{\alpha}}$ denotes the normal density function with mean 0 and covariance matrix depending on an unknown parameter $\boldsymbol{\alpha}$. We review three methods to compute the likelihood function, the maximizer of which gives the MLE of σ, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$.

6.3.1. Laplace's Approximation

Letting $e^{l_i(\mathbf{b}|\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})}$ be the integrand in the right-hand side of (6.14), Laplaces's asymptotic formula for integrals yields the approximation

$$\int e^{l_i(\mathbf{b}|\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})} d\mathbf{b} \approx (2\pi)^{q/2} \left\{ \det\left[-\ddot{l}_i\left(\widehat{\mathbf{b}}_i \mid \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}\right)\right] \right\}^{-1/2} \exp\left\{ l_i\left(\widehat{\mathbf{b}}_i \mid \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}\right) \right\}, \qquad (6.15)$$

where q is the dimension of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is the maximizer of $l_i(\mathbf{b}\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$, and $\ddot{l}_i$ denotes the Hessian matrix consisting of second partial derivatives of l_i with respect to the components of $\mathbf{b}$. The R package lme4 computes the MLE by using the Laplace approximation to (6.14) or the restricted pseudolikelihood approach proposed by [76], as the user-specified option.

6.3.2. Gauss-Hermite Quadrature

Laplaces's asymptotic formula (6.15) is derived from the asymptotic approximation of l_i by a quadratic function of $\mathbf{b}$ in a small neighborhood of $\widehat{\mathbf{b}}_i$ as $\lambda_{\min}(-\ddot{l}_i(\widehat{\mathbf{b}}_i \mid \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})) \rightarrow \infty$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a symmetric matrix. Therefore, such formula may produce significant approximation error for L_i if the corresponding $\lambda_{\min}(-\ddot{l}_i(\widehat{\mathbf{b}}_i \mid \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}))$ is not sufficiently large. One way to reduce the possible approximation error is to compute L_i by using an adaptive Gauss-Hermite quadrature rule, as in [77]. The software package SAS uses adaptive Gauss-Hermite quadrature in the NLMIXED procedure to compute (6.14); the R function lmer($\cdot$) also uses Gaussian quadrature to compute (6.14) but only for certain special cases of the exponential family (6.1).

6.3.3. Hybrid Method

The numerical integration procedure demands a much higher computational effort and becomes computationally infeasible if n or q is large. To circumvent the issue of high-dimensional numerical integration, Antonio and Beirlant [72] propose to put prior distributions on the unknown parameters and to estimate them by using the Markov chain Monte Carlo (MCMC) method in a Bayesian way. The performance of the MCMC method,

however, depends on how the prior parameters are set as well as the convergence rate of the Markov chain to its stationary distribution, which may not even exist. Yafune et al. [78] use direct Monte Carlo integration but point out that the computational time may be too long to be of practical interest. Instead of relying solely on Monte Carlo, Lai and Shih [79] proposed the following hybrid approach. Let $\mathbf{V}_i = -\ddot{l}_i(\hat{\mathbf{b}}_i|\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Since Laplaces's asymptotic formula (6.15) may be a poor approximation to (6.14) when $\lambda_{\min}(V_i)$ is not sufficiently large, Monte Carlo integration, whose error is independent of q, can be used as an alternative method to evaluate (6.14).

Lai and Shih [79] suggest using Monte Carlo integration instead of Laplace's asymptotic formula for i whose $\lambda_{\min}(\mathbf{V}_i) < c$, where c is a positive threshold. Lai et al. [80, 81] further improve the hybrid method by using importance sampling in the Monte Carlo component of the hybrid method. Specifically, instead of sampling $\mathbf{b}^{(h)}$ directly from $\phi_{\boldsymbol{\alpha}}$ as in [79], sample it from a mixture of the prior normal distribution with density $\phi_{\boldsymbol{\alpha}}$ and the posterior normal distribution $N(\hat{\mathbf{b}}_i, [-\ddot{l}_i(\hat{\mathbf{b}}_i \mid \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \varepsilon\mathbf{I}]^{-1})$, where ε is a small positive number to ensure that the covariance matrix is invertible. We denote such choice of trial density as λ. It has the advantage of further incorporating the essence of Laplace's method in the Monte Carlo step such that the method is less dependent (than direct Monte Carlo) on the choice of the threshold c.

(i) If $\lambda_{\min}(\mathbf{V}_i) < c$, evaluate $L_i(\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ in (6.14) by the Monte Carlo approximation

$$\frac{\left\{ \sum_{h=1}^{H} \left[w_h \prod_{t=1}^{T} f\left(y_{it}; \theta_{it}(\mathbf{b}^{(h)}, \boldsymbol{\beta}), \sigma\right) \right] \right\}}{\left\{ \sum_{h=1}^{H} w_h \right\}} \tag{6.16}$$

in which $\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(H)}$ are independently sampled from the importance density λ and $w_h = \phi_{\boldsymbol{\alpha}}(\mathbf{b}^{(h)})/\lambda(\mathbf{b}^{(h)})$ are the importance weights.

(ii) If $\lambda_{\min}(\mathbf{V}_i) \geq c$, use Laplace approximation (6.15). Lai et al. [80] propose to use $c = 10, H = 100$, and a mixture distribution that assigns 0.2 probability to $\phi_{\boldsymbol{\alpha}}$. Note that choosing $\lambda = \phi_{\boldsymbol{\alpha}}$ yields $w_h = 1$, for which (6.16) reduces to direct Monte Carlo evaluation of $L_i(\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ in [79]. The likelihoods $L_i(\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$ thus computed then yield the likelihood function $\prod_{i=1}^{n} L_i(\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta})$, whose logarithm can be maximized by using iterative gradient-type schemes.

6.3.4. Model Selection

As shown by Breslow and Clayton [4], a consistent estimator of the asymptotic covariance matrix of $(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is $\hat{\mathbf{V}}^{-1} = \{-\sum_{i=1}^{n} \ddot{R}_i(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})\}^{-1}$, where $R_i = \log L_i$ and $\ddot{R}_i$ denotes the Hessian matrix of second derivatives of R_i with respect to σ, $\boldsymbol{\alpha}$, and the components of $\boldsymbol{\beta}$. As in [79], we compute the observed information matrix $\hat{\mathbf{V}} = -\sum_{i=1}^{n} \ddot{R}_i(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ by taking numerical derivatives and using the hybrid method to evaluate L_i. Lai and Shih [79] have shown that $\hat{\mathbf{V}}^{-1/2}(\hat{\sigma} - \sigma, \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$ has a limiting standard multivariate normal distribution as $n \to \infty$ under certain regularity conditions, and that the sandwich estimator

$$\left\{ -\sum_{i=1}^{n} \ddot{R}_i\left(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\right) \right\}^{-1} \left\{ \sum_{i=1}^{n} \dot{R}_i\left(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\right) \dot{R}_i^T\left(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\right) \right\} \left\{ -\sum_{i=1}^{n} \ddot{R}_i\left(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\right) \right\}^{-1} \tag{6.17}$$

is more robust than $\hat{\mathbf{V}}^{-1}$, where $\dot{R}_i$ denotes the gradient vector of partial derivatives.

   To choose the covariates in GLMM defined by (6.1) and (6.2), Lai et al. [80, 81] propose to use a model selection procedure that consists of stepwise forward addition followed by stepwise backward elimination. A forward addition step aims at adding to the current model the most significant covariate among all possible candidates, while a backward elimination step aims at excluding from the current model the least significant candidate covariate. Specifically, for forward addition, they propose to choose one explanatory variable, amongst K possible candidates, to add to a fitted model with $k - 1$ covariates, with coefficients $\tilde{\beta}_1, \ldots, \tilde{\beta}_{k-1}$, $\tilde{\sigma}$, and $\tilde{\boldsymbol{\alpha}}$. For the jth candidate, consider the kth component $s_k^{(j)}$ of the score vector $\sum_{i=1}^n \dot{R}_i(\tilde{\sigma}, \tilde{\boldsymbol{\alpha}}, (\tilde{\beta}_1, \ldots, \tilde{\beta}_{k-1}, 0)^T)$, called the "Rao statistic," that includes the $k - 1$ chosen covariates and the jth candidate. Let $\tilde{v}_k^{(j)}$ be the kth diagonal element of $- \sum_{i=1}^n \ddot{R}_i(\tilde{\sigma}, \tilde{\boldsymbol{\alpha}}, (\tilde{\beta}_1, \ldots, \tilde{\beta}_{k-1}, 0)^T)$. The candidate that maximizes $|s_k^{(j)}|/\tilde{v}_k^{(j)}$ is chosen for forward addition. For backward elimination, they propose to eliminate one covariate from a fitted model with k covariates, with coefficients $\tilde{\beta}_1, \ldots, \tilde{\beta}_k$, $\tilde{\sigma}$, and $\tilde{\boldsymbol{\alpha}}$. They consider the "Wald statistic" $|\hat{\beta}_j|/\hat{v}_j$ for testing $\beta_j = 0$, where $\hat{v}_j$ is the square root of the jth diagonal element of $\hat{V}^{-1} = \{- \sum_{i=1}^n \ddot{R}_i(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})\}^{-1}$. If the elimination of the covariate with the smallest Wald statistic leads to a new model with a smaller information criterion (e.g., BIC), then the new model is preferred. The stepwise forward addition part of the model selection procedure terminates when the chosen information criterion for model selection does not improve with the addition of a covariate. It also terminates at the current model if there is no more candidate covariate to be included. At that point, backward elimination starts and proceeds until the selected information criterion does not improve the current model by deleting an explanatory variable. Since the estimation algorithm is likelihood based, Schwarz's Bayesian Information Criterion (BIC) can still be used for model selection:

$$\text{BIC} = -2 \sum_{i=1}^n R_i\left(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\right) + (\log n) \text{ (number of parameters)}. \tag{6.18}$$

If the sandwich estimator (6.17) is to be used instead of $\hat{V}^{-1}$, then in the forward addition and backward elimination procedure above, $\hat{v}_j$ should be replaced by the square root of the jth diagonal entry of (6.17).

6.4. Prediction in GLMM

The main goal of default modeling in many other applications in econometrics is to predict certain characteristics of a specific subject given its covariate vector, for example, to forecast $\mu_{i,t+1} = g^{-1}(\boldsymbol{\beta}^T \mathbf{x}_{i,t+1} + \mathbf{b}_i^T \mathbf{z}_{i,t+1})$ at time t with observed covariate vector $(\mathbf{x}_{i,t+1}^T, \mathbf{z}_{i,t+1}^T)^T$. More generally, we would like to estimate a function of the unobservable $\mathbf{b}_i$ and we denote it by $h(\mathbf{b}_i)$. Now, if $\sigma, \boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ are known, then $h(\mathbf{b}_i)$ should be estimated by the posterior mean $E_{\sigma,\boldsymbol{\alpha},\boldsymbol{\beta}}[h(\mathbf{b}_i)|i\text{th subject's data}]$. Without assuming $\sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}$ being known, the empirical Bayes approach replaces them in the posterior mean by their estimates $(\hat{\sigma}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ so that h is estimated by

$$\hat{h} = E_{\hat{\sigma},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}}\left[h(\mathbf{b}_i) \mid i\text{th subject's data}\right]. \tag{6.19}$$

The expectation in (6.19) can be evaluated by the hybrid method, which we have used for likelihood calculation; see [80, 81]. First note that Laplace's approximation in (6.15) is based on the Taylor expansion

$$l_i(\mathbf{b}) \approx l_i\left(\widehat{\mathbf{b}}_i\right) + \frac{\left(\mathbf{b} - \widehat{\mathbf{b}}_i\right)\ddot{l}_i\left(\widehat{\mathbf{b}}_i\right)\left(\mathbf{b} - \widehat{\mathbf{b}}_i\right)^T}{2}. \tag{6.20}$$

Since the density function of $\mathbf{b}_i$ given the ith risk class data is proportional to $e^{l_i(\mathbf{b})}$, it follows from (6.20) that the conditional distribution of $\mathbf{b}_i$ given the data is approximately normal with mean $\widehat{\mathbf{b}}_i$ and covariance matrix $\mathbf{V}_i^{-1}$, where $\mathbf{V}_i = -\ddot{l}_i(\widehat{\mathbf{b}}_i)$. Hence for a subject with informative data (i.e., satisfying $\lambda_{\min}(\mathbf{V}) \geq c$), for whom we drop the subscript i in l_i, $\mathbf{V}_i$, and $\widehat{\mathbf{b}}_i$, we can use the preceding normal density to evaluate $\widehat{h}$ in (6.19). Consider

$$h(\mathbf{b}) \approx h\left(\widehat{\mathbf{b}} + \mathbf{V}^{-1/2}\mathbf{Z}\right) \approx h\left(\widehat{\mathbf{b}}\right) + \dot{h}\left(\widehat{\mathbf{b}}\right)^T \mathbf{V}^{-1/2}\mathbf{Z} + \frac{1}{2}\mathbf{Z}^T \mathbf{V}^{-1/2}\ddot{h}\left(\widehat{\mathbf{b}}\right)^T \mathbf{V}^{-1/2}\mathbf{Z}, \tag{6.21}$$

where $\mathbf{Z}$ is a q-dimensional standard normal distribution. Using the preceding approximation gives

$$\widehat{h} = h\left(\widehat{\mathbf{b}}\right) + \frac{1}{2}\mathbf{V}^{-1/2}\ddot{h}\left(\widehat{\mathbf{b}}\right)^T \mathbf{V}^{-1/2} \tag{6.22}$$

for the informative subject.

When the eigenvalue criterion in the hybrid method fails (i.e., when $\lambda_{\min}(\mathbf{V}) < c$), (6.19) can be evaluated by using importance sampling for the Monte Carlo approximation

$$\widehat{h} \approx \frac{\sum_{h=1}^{H}\left[h(\mathbf{b}^{(h)})w_h\right]}{\sum_{h=1}^{H} w_h}, \tag{6.23}$$

where $\{\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(H)}\}$ are independent samples from the trial density $\lambda(\cdot)$ and the importance weights are $w_h = e^{l(\mathbf{b}^{(h)})}/\lambda(\mathbf{b}^{(h)})$. Similar to the hybrid scheme presented in Section 6.3, $\lambda(\cdot)$ is chosen to be a mixture of the prior normal distribution with density ϕ_α and the posterior normal distribution $N(\widehat{\mathbf{b}}_i, [-\ddot{l}_i(\widehat{\mathbf{b}}_i \mid \sigma, \alpha, \beta) + \varepsilon\mathbf{I}]^{-1})$, where ε is a small positive number to ensure that the covariance matrix is invertible.

6.5. Information Sets and Selection of Variables for Prediction

We have assumed so far that the observations $(Y_{i,t}, \mathbf{x}_{i,t}, \mathbf{z}_{i,t})$ are available at every $1 \leq t \leq T$, for all $1 \leq i \leq n$. In longitudinal data in biostatistics, however, there is often between-subject variations in the observation times. Lai et al. [82] recently addressed this difficulty by using a prediction approach that customizes the predictive model for an individual by choosing predictors from the individual's information set and using the data from all subjects whose information sets contain these predictors to estimate the parameters of the predictive model. Let T_i be the set of times when the ith subject is observed, and let $\mathbf{u}_{i,t}$

be the subject's covariate vector at time $t \in T_i$. The marginal regression approach in longitudinal data analysis models $E(Y_{i,t} \mid \mathbf{u}_{i,t})$, $t \in T_i$, by using generalized linear models, under the independence assumption over time or other working dependence assumptions (such as exchangeability, or autoregression when T_i is regularly spaced) as in the GEE (generalized estimating equation) methodology introduced by Liang and Zeger [83]. It has been shown by Pepe and Anderson [84] that in ignorance of the true dependence structure over time and in the presence of time-varying covariates, the GEE parameter estimates may be biased unless the working independence assumption is used. In contrast to the marginal regression approach, the transitional modeling approach incorporates outcome and covariate histories to model $E(Y_{i,t} \mid Y_{i,s}, \mathbf{u}_{i,\tau}; s \in T_i$ and $s < t, \tau \in T_i$ and $\tau \leq t)$ via generalized linear models, by estimating the model parameters from all subjects who are observed at all times $t \in T_i$. A compromise between marginal regression and transitional modeling is the partly conditional mean approach proposed by Pepe and Couper [85]. Instead of working with the subject-specific set T_i, this approach considers the set T of all possible times and introduces an indicator variable $R_i(t)$ and a stratifying variable $K_i(t)$ such that $R_i(t) = 1$ (or 0) if the ith subject is observed (or not observed) at time t and $K_i(t)$ takes the values $1, \ldots, K$ to represent the stratum to which subject i belongs. Lai et al. [82, Section 2.1] have pointed out that partly conditional mean modeling can be viewed as a special case of a more general approach that chooses predictors from a subject's information set. This more general approach allows one to use ideas similar to those used in information criteria such as BIC or AIC to choose predictors from a subject's information set. More importantly, it provides a framework for Lai et al. [3] to extend and refine the dynamic EB approach as follows.

   To begin with, instead of K strata in the partly conditional mean approach, we divide the subjects into K structurally similar subgroups. In many applications, subjects belonging to the same subgroup have similar observation times because of their structural similarity. For example, patients who have more serious ailments are monitored more frequently than others in a study cohort, causing the irregularity of observation times over different subgroups. Lai et al. [3] assume the cross-sectional dynamics (6.12) separately for each subgroup, that is, with μ_t and $\overline{Y}_{t-j}$ replaced by $\mu_t^{(k)}$ and $\overline{Y}_{t-j}^{(k)}$ for the kth subgroup, in which $\overline{Y}_s^{(k)}$ is the sample average from all subjects (from the subgroup) who are observed at time s. Concerning (6.13) for $\mu_{i,t}$ with i belonging to the kth subgroup, we choose only predictors from the information set of subject i that are common to all subjects in the kth subgroup. Lai et al. [3] propose to use the BIC, details of which are given in [80, page 58-59], for the GLMM (6.13) with i belonging to the kth subgroup and $\overline{Y}_{t-j}$ replaced by $\overline{Y}_{t-j}^{(k)}$, and give examples from the problem of predicting the batting performance of baseball players studied previously by Efron and Morris [86, 87] and Brown [88] to illustrate this point.

6.6. Evolutionary Credibility via GLMM (6.13)

In Section 6.1 we have used the dynamic EB approach to elucidate the close connection between Kalman filtering in the linear evolutionary credibility theory in [62] and the linear mixed model (6.11). Section 6.2 extends (6.11) to the GLMM (6.13) to handle generalized linear models of the form (6.1). The usefulness of (6.1) for insurance claims data has been pointed out by Hsiao et al. [89] who note that these claims data often exhibit "excess zeros,"

meaning that a large fraction of policies during a given year do not have claims. They propose to handle excess zeros by using a two-part model of the form

$$Y_{it} = I_{it} Z_{it}, \quad i = 1, \ldots, n; \ t = 1, \ldots, T, \tag{6.24}$$

where $I_{it} = I_{\{Y_{it} \neq 0\}}$ and Z_{it} has the conditional distribution of Y_{it} given that $Y_{it} \neq 0$. In this two-part model, if Y_{it} is the claim size of the ith policy at time t, then Z_{it} is the claim size given that the ith policy files a claim at time t.

6.6.1. Two-Part Nonlinear State-Space Models

The state-space model for linear evolutionary credibility in Section 6.1 can be extended to the two-part model (6.24) as follows:

$$I_{it} \sim \text{Bernoulli}(\pi_{it}), \qquad Z_{it} \sim \text{Gamma}\left(\kappa, \frac{\mu_{it}}{\kappa}\right) \quad \text{if } I_{it} = 1, \tag{6.25}$$

where κ is the shape parameter and link functions of $\pi_{it} = E(I_{it})$ and $\mu_{it} = E(Z_{it})$ undergo AR(1) dynamics of the form

$$\text{logit}(\pi_{it}) = \rho \text{logit}(\pi_{i,t-1}) + (1 - \rho)\text{logit}(\pi_i) + \eta_{it},$$
$$\log(\mu_{it}) = \rho' \log(\mu_{i,t-1}) + (1 - \rho') \log(\mu_i) + \epsilon_{it}. \tag{6.26}$$

The function $\text{logit}(\pi) = \log(\pi/(1 - \pi))$ is the canonical link function for the exponential family of Bernoulli random variables. The η_{it} and ϵ_{it} in (6.26) are assumed to be independent normal random variables with mean 0, $\text{Var}(\eta_{it}) = V$, and $\text{Var}(\epsilon_{it}) = V'$.

Whereas Kalman filtering can be used to estimate the unobserved states μ_t in the linear state-space model for evolutionary credibility in Section 6.1, the state-space model (6.25)-(6.26) is nonlinear, for which the Kalman filter is replaced by nonlinear filters to estimate the unobservable states π_{it} and μ_{it}. These nonlinear filters are highly complex and require Markov chain Monte Carlo methods for their implementation when the model hyperparameters $\rho, \rho', V, V', \pi_1, \ldots, \pi_n, \mu_1, \ldots, \mu_n$ are known. Since they are unknown in practice, there is the additional complexity of estimating them, for which a stochastic EM approach implemented by Monte Carlo methods can be used to address the computational difficulties, as we have discussed in Section 4.1 for another application.

6.6.2. Two-Part Dynamic Modeling of Claims via GLMM

Since the Bernoulli and gamma distributions are special cases of the exponential family, we can apply the GLMM introduced in Section 6.2 to obtain a more tractable two-part model of evolutionary credibility, replacing (6.26) by

$$\text{logit}(\pi_{it}) = \rho_1 \text{logit}\left(\overline{I}_{t-1}\right) + \mu' + a_i,$$
$$\log(\mu_{it}) = \rho_2 \log\left(\overline{Z}_{t-1}\right) + \mu + b_i, \tag{6.27}$$

in which $a_i \sim N(0, v')$ and $b_i \sim N(0, v)$ are random effects, $\bar{I}_s = n^{-1} \sum_{i=1}^{n} I_{is}$, and

$$\bar{Z}_s = \frac{\sum_{i:I_{is} \neq 0} Z_s}{\sum_{i:I_{is} \neq 0} 1} \tag{6.28}$$

is the average claim size over all positive claims at time s. The parameters of the GLMM, which uses the cross-sectional means $\bar{I}_{t-1}$ and $\bar{Z}_{t-1}$ to substitute for the unobservable states $\pi_{i,t-1}$ and $\mu_{i,t-1}$ in (6.26), are μ, v, μ', and v'. As pointed out in Section 6.2, the GLMM can also incorporate additional covariates $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ as in (6.13) to improve the predictive power of the model.

In the state-space model (6.25)-(6.26), the conditional distribution of the unobservable state vector (π_{it}, μ_{it}) given the observed data D_t is difficult to represent analytically or algorithmically, unlike that in the linear Gaussian state-space model for which the conditional distribution is normal and its mean and covariance matrix have a closed-form recursive formula, that is, the Kalman filter. Instead of specifying the stochastic dynamics of the state vector precisely, West et al. [90] assume a conjugate family of prior distributions of the exponential family in the generalized linear model (6.1) so that the parameters of the conjugate family can be updated similarly to the Kalman filter. Lai and Sun [73] have compared the prediction performance of this approximation to the nonlinear HMM filters for the two-part state-space model (6.25)-(6.26) with that of the GLMM (6.27) in a simulation study, in which the data are generated by the two-part state-space model, and have found that the two methods perform similarly. An advantage of using GLMM to model evolutionary credibility over nonlinear state-space models or their approximations proposed in [90] is that GLMM has a relatively complete and easy-to-use methodology for likelihood inference and model selection; see [3, 80]. In particular, we can use model selection criteria such as BIC to choose the order p of an AR(p) model for μ_{it}, or to determine which of the variables in $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ should be included in the model to avoid overfitting.

State-space models have traditionally been considered in the context of a single d-dimensional series. A distinctive feature of evolutionary credibility is that it involves many structurally similar time series, called "dynamic panel data" in econometrics, and we can exploit this feature to derive a *dynamic empirical Bayes* approach to evolutionary credibility. As pointed out in [91, page 335], traditional credibility theory can be called "empirical Bayes credibility" since it basically amounts to replacing the hyperparameters in the Bayes credibility (assuming Gaussian claims and a Gaussian prior) by their method-of-moment estimates. Extending this empirical Bayes approach to the Bayesian model of linear evolutionary credibility, which is a linear state-space model, yields the LMM in Section 6.1. Similarly, the GLMM in this section can be interpreted as a dynamic empirical Bayes approach to evolutionary credibility in generalized linear models.

6.7. Competing Risks and Retail Loans

We now return to the problem of predicting the default probabilities and loss given default of mortgages and corporate bonds that we have mentioned in Section 1. The prediction problem is important for monitoring and management of a bank's credit risk. In 2001, the Basel Committee on Banking Supervision [92] proposed a new accord for risk evaluation,

which was finalized in 2005 and is known as the Basel II Accord. Using the internal ratings-based (IRB) approach, the expected credit loss $E(L_c)$ for a loan is decomposed into

$$E(L_c) = \text{EAD} \times \text{LGD} \times \text{PD}, \tag{6.29}$$

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. We first consider default probabilities of retail loans (such as personal loans, mortgages, credit cards, and car loans). "Retail" in banking means consumer-related activities, while "wholesale" means business-related loan transactions which include commercial loans, commercial real estate loans, and leases to businesses.

Retail loans are typically grouped into several classes called *risk buckets*, within which the obligors can be regarded as having the same prior probability of default. Risk bucketing is tantamount to stratifying the obligors into classes so that the baseline covariate (such as credit score at the time of loan application) of an individual obligor can be replaced by the class membership. For retail loans, default means overdue payment for longer than 90 days. Thus, unlike a corporation that files for bankruptcy and the LGD of the loan is determined later by the recovery rate, a retail loan borrower can default and incur a loss to the lender while still holding the loan until it is foreclosed and taken off the balance sheet. Moreover, besides the possibility of default, another possible event that the lender faces is prepayment.

6.7.1. Competing Risks

Calhoun and Deng [93] regard default and prepayment as *competing risks* and introduce a multilogit regression model of the trinomial variable $Y(t)$, taking the values 2, 1, and 0 that correspond to prepayment, foreclosure due to default, and still surviving at time t, respectively. In our setting that considers the trinomial variable $Y_{ij}(t)$ of the ith obligor in the jth risk class, the multilogit regression model relates $Y_{ij}(t)$ to a covariate vector $\mathbf{u}_{ij}(t)$ by

$$P\left(Y_{ij}(t) = k \mid Y_{ij}(s) = 0 \text{ for } s < t, \mathbf{u}_{ij}(t)\right) = \frac{\exp\left(\alpha_{jk} + \beta_{jk}^T \mathbf{u}_{ij}(t)\right)}{1 + \sum_{h=1}^{2} \exp\left(\alpha_{jh} + \beta_{jh}^T \mathbf{u}_{ij}(t)\right)} \tag{6.30}$$

for $k = 1, 2$. Competing-risks models can be viewed as "marked" failure-time models, in which a *mark* is assigned to a failure event, indicating the type of the event that has occurred.

Competing-risks data are also prevalent in cancer studies for which the response to treatment can be classified in terms of failure from disease processes and non-disease-related causes. For example, after treatments for leukemia and lymphoma, patients may either relapse or die in remission. In quality-of-life studies, the risks of the different failure types need to be considered for cost-effectivness analyses. Although Cox's proportional hazards model has routinely been used to analyze competing-risks data with covariates, what it models is the *cause-specific hazard rate*

$$\lambda_j(t; \mathbf{Z}) = \lim_{h \to 0} h^{-1} P\{t \le Y \le t + h, J = j \mid Y \ge t, \mathbf{Z}\}, \tag{6.31}$$

where $\mathbf{Z}$ is a vector of covariates, Y is the failure time, and J is the cause of failure, with the causes labeled $1, 2, \ldots, j^*$. This approach treats each cause separately as the only cause for the observed failures that can be attributed to it and treats the observed failures due to other causes as censoring outcomes. Viewing competing risks as multiple endpoints each of which precludes the occurrence of the others, Fine and Gray [94] proposed to consider, instead of (6.31), the hazard function of the cumulative incidence due to cause j:

$$
\begin{aligned}
\widetilde{\lambda}_j(t; \mathbf{Z}) &= \lim_{h \to 0} h^{-1} P\big[t \le Y \le t + h, J = j \mid \{Y \ge t\} \cup \{Y < t, J \ne j\}, \mathbf{Z}\big] \\
&= -\frac{d}{dt} \log\big(1 - F_j(t; \mathbf{Z})\big),
\end{aligned}
\tag{6.32}
$$

where $F_j(t; \mathbf{Z})$ is the *cumulative incidence function*

$$
F_j(t; \mathbf{Z}) = P\{Y \le t, J = j \mid \mathbf{Z}\}
\tag{6.33}
$$

introduced by Gray [95]. Note that $1 - F_j(t-; \mathbf{Z}) = 1 - P\{Y < t, J = j \mid \mathbf{Z}\} = P\{Y \ge t \mid \mathbf{Z}\} + P\{Y < t, J \ne j \mid \mathbf{Z}\}$. In the absence of covariates, the cumulative incidence function can be estimated by the Aalen-Johansen estimator [27, 28] based on a finite-state Markov chain model of competing risks; see the survey by Klein and Shu [96]. Although the proportional hazards model $\lambda_j(t; \mathbf{Z}) = \lambda_j^0(t) e^{\boldsymbol{\beta}^T \mathbf{Z}}$ can readily be extended to the hazard functions $\widetilde{\lambda}_j$ of the competing risks, regression analysis of $\widetilde{\lambda}_j$ under the model $\widetilde{\lambda}_j(t; \mathbf{Z}) = \widetilde{\lambda}_j^0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z})$ becomes much more difficult. Three approaches to this problem have been developed. First, Fine and Gray [94] used the following modification of risk sets to extend partial likelihood to competing-risks data. The risk set at time t includes the subjects that are naturally at risk and those who have failed prior to time t due to other competing risks and are not censored at time t. A subject who has failed prior to time t due to other competing risks is weighted by the probability that it is not censored at time t given that it has not been censored at the failure time. The second method, proposed by Scheike et al. [97], uses weighted counting processes as responses in a binomial regression model for cumulative incidence functions. Both methods assume a common censoring distribution. Their performance is highly sensitive to how well the censoring distribution can be estimated. The third method, proposed by Klein and Andersen [98], uses jackknife pseudovalues to model the cumulative incidence functions directly. These pseudo-values are used in a generalized estimating equation to obtain estimates on model parameters, and Graw et al. [99] have recently shown this method to be consistent and asymptotically normal.

An important difference between the retail loans data $Y_{ij}(t)$ in (6.30) and those in cancer biostatistics is that the latter is subject to right-censoring due to patient withdrawal whereas the former remain in the lender's balance sheet until prepayment or default leading to foreclosure. Right-censoring results in much greater complexity than multilogit regression in modeling competing risks.

6.7.2. GLMM for Loan Default Probabilities

Calhoun and Deng [93] fitted (6.30) to about 1.3 million conforming fixed-rate and adjustable-rate residential mortgages originated between 1979 and 1993. Their covariate

vector includes loan age, dummy variables for origination years, the loan-to-value ratio (LTV), seasonality, occupancy status of the property, the original loan size, the ratio of the 10-year Constant Maturity Treasury rate to the 1-year Constant Maturity Treasury rate, mortgage premium value, and the probability of negative equity based on estimated drift and volatility of housing prices. Instead of the multilogit model (6.30) that generalizes logistic regression, Lai et al. [100] use the GLMM in Section 6.2 to model the trinomial variable $Y_{ij}(t)$ that represents continuing payment, or default, or prepayment at time t of the ith obligor in the jth risk class. Whereas [93] does not divide the borrowers into risk classes, the statistical analysis of subprime mortgages in [100] conforms to risk bucketing that is commonly used for risk management of retail loans and is based on loan-level mortgage data consisting of 10,000 first-lien, 2–28 adjustable-rate subprime loans originated between January 2004 and June 2006; these loans are randomly selected from a database maintained by LoanPerformance, Inc., a mortgage information and analytics provider. The 10,000 loans are grouped into $I = 5$ risk classes according to the FICO score when a loan was originated; the class with the lowest FICO scores is labeled 1 and the highest labeled 5. These data involve two time scales, calendar time t, and loan age s, which also arise in biostatistics in the context of time-sequential clinical trials with failure-time endpoints and staggered entry, where $s = (t - \tau)_+$ and τ is the entry time of the patient into the study (or in the case of mortgages the initiation date of the loan); see [101]. The case $s = 0$ denotes that the loan has not been originated at calendar time t.

Lai et al. [100] use the link function

$$\eta_{ijt}^{(k)} = \log\left(\frac{P\{Y_{ij}(t) = k \mid Y_{ij}(t-1) = 0, \mathbf{u}_{ij}(t)\}}{P\{Y_{ij}(t) = 0 \mid Y_{ij}(t-1) = 0, \mathbf{u}_{ij}(t)\}} \right), \quad k = 1, 2, \tag{6.34}$$

to express (6.30) in the generalized linear form, which can then be extended to the GLMM in Section 6.2:

$$\eta_{ijt}^{(k)} = \rho^{(k)} \log\left(\frac{\overline{Y}_{t-1}^{(k)}}{\overline{Y}_{t-1}^{(0)}} \right) + a_j^{(k)} + \boldsymbol{\beta}^{(k)T} \mathbf{X}_{ijt} + \mathbf{b}_j^{(k)T} \mathbf{Z}_{ijt}, \tag{6.35}$$

where $\mathbf{X}_{ijt}$ and $\mathbf{Z}_{ijt}$ are covariates that are available prior to t and $\overline{Y}_{t-1}^{(k)}$ is the cross-sectional mean of the indicator variables $I_{\{Y_{ij}(t-1)=h, Y_{ij}(t-2)=0\}}$ for $k = 0, 1, 2$. If the loan has not been initiated at time $t - 1$, then the conditional probabilities in (6.34) are taken to be $P\{Y_{ij}(t) = \cdot \mid \mathbf{u}_{ij}(t)\}$ and the cross-sectional term involving $\overline{Y}_{t-1}^{(k)}/\overline{Y}_{t-1}^{(0)}$ is also removed from (6.35). The covariate vector $\mathbf{X}_{ijt}$ includes the loan age, occupancy status, initial mortgage rate, purpose (for purchase or refinance), a dummy variable for full documentation, or otherwise (important for reflecting the lender's standards in issuing the subprime loan), and $\mathbf{Z}_t$ includes mortgage premium value and equity position as defined in [100], which involve the macroeconomic variables 1-year Constant Maturity Treasury rate and zip-code-based house price index, together with the loan's size, unpaid rate, mortgage rate, and LTV. These data do not contain the loan's payment history; ideally $Y_{ij}(t)$ should also indicate if the payment is overdue for longer than 90 days even when the loan is not foreclosed, but such information is not available in the data obtained from LoanPerformance, only showing the trinomial variable if the loan is prepaid, foreclosed, or active at time t. The advantages of (6.35) over the multilogit model (6.30) in predicting defaults are discussed in [102].

6.7.3. Loss Given Default

Lai and Wong [103, Section 5.2] have described how GLMM can be used to model the LGD of retail loans. They first review how LGD is calculated for corporate loans and bonds in the software package LossCalc of Moody's KMV, which is built from a training sample of recovery data of the defaulted corporate loans, bonds, and preferred stock. The training sample is used to fit a linear regression model of $y = \Phi^{-1}(F_{\alpha,\beta}(\text{RR}))$ on nine predictors chosen from a set of regressions using a year-by-year out-of-sample prediction error criterion, where RR is the recovery rate implied by the market value (bid-side market quote) one month after default, Φ is the standard normal distribution function, and $F_{\alpha,\beta}$ is the distribution function of the Beta(α, β) distribution. The transformation $\Phi^{-1} \circ F_{\alpha,\beta}$ is used to convert a beta random variable into a normal one so that linear regression can be applied. As shown in [103, page 237], a better approach is to fit the beta regression model, which is a generalized linear model.

As pointed out in [103], unlike corporate loans 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. To describe simply the recovery process, the bank first tries (a) and switches to (b) or (c) if (a) is unsuccessful after, say, three 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 US), 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 a scenario corresponds to complete recovery and leads to an atom at 0 in the marginal distribution of the economic LGD. Lai and Wong [103] point out the importance of including the atom 0 and demonstrate the advantages of using beta regression over linear regression of the transformed recovery rate on covariates. They also note that the Basel II Accord requires LGD estimates to be grounded in historical recoveries and to incorporate business cycles 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 five years. They propose the following GLMM 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 $s = t, t - 1, \dots, t - 4$. For $i \in D_s$, let $Y_i(s)$ be the economic LGD in year s. As noted above, $Y_i(s)$ has an atom at 0. Let $\pi_i(s) = P\{Y_i(s) > 0\}$, which can be modeled by

$$\text{logit}(\pi_i(s)) = \xi_i + \boldsymbol{\eta}^T \mathbf{x}_i(s - 1) + \boldsymbol{\gamma}^T \mathbf{f}(s), \tag{6.36}$$

where $\mathbf{x}_i(s - 1)$ is a vector of attributes of the obligor's creditworthiness and collateral in year $s - 1$ and $\mathbf{f}(s)$ is a vector of macroeconomic factors in year s. The conditional density of $Y_i(s) \mid Y_i(s) > 0$ is assumed to belong to the exponential family of beta density functions

$$f(y \mid \mu_i(s), \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_i(s)\phi)\Gamma((1 - \mu_i(s))\phi)} y^{\mu_i(s)\phi - 1} (1 - y)^{(1 - \mu_i(s))\phi - 1} \tag{6.37}$$

for $0 < y < 1$, where $\mu_i(s) = E(Y_i(s) \mid Y_i(s) > 0)$,

$$\text{logit}(\mu_i(s)) = \zeta_i + \boldsymbol{\kappa}^T \mathbf{x}_i(s-1) + \boldsymbol{\nu}^T \mathbf{f}(s), \qquad (6.38)$$

$\text{Var}(Y_i(s) \mid Y_i(s) > 0) = \mu_i(s)(1 - \mu_i(s))/(1 + \phi)$, and ξ_i are i.i.d. normal and are independent of ζ_i which are also i.i.d. normal. This GLMM includes loan-specific random effects ξ_i and ζ_i in (6.36) and (6.38) and is similar to the two-part model in Section 6.6. It enables one to predict directly $\text{LGD}_i(t+1)$ of the ith obligor for year $t+1$ using $\mathbf{x}_i(t)$ and the forecast of $\mathbf{f}(t+1)$, noting that $\text{LGD}_i(t+1) = \pi_i(t+1)E(Y_i(t+1) \mid Y_i(t+1) > 0)$.

6.8. GLMM for Modeling Frailty and Contagion of Credit Portfolios

Similarly to the applications to evolutionary credibility and mortgage loans in the preceding two sections, the dynamic EB approach via GLMM can be applied to provide a new model for correlated default intensities of multiple firms in a portfolio of corporate bonds or a multiname credit derivative. This model can incorporate both frailty and contagion via the covariates of the GLMM and is considerably easier to implement. The details are given in [3, 100]. Note that the use of a mixture distribution for firm-specific random effects also appears in Vasicek's mixed binomial model (3.11) for the number of defaults in a loan portfolio up to a horizon T. It is shown in [3] that the prediction performance of the nonlinear state-space model of Duffie et al. [46] is similar to that of the much simpler GLMM model of dynamic frailty even when the data are actually generated by the nonlinear state-space model (4.2).

7. Conclusion

This paper shows the methodological connections between (a) default modeling and prediction of mortgage loans and portfolios of corporate bonds, a topic of timely interest in the wake of the financial crisis following the events in 2007 and 2008 described in Section 1, (b) insurance rate-making in evolutionary credibility theory, and (c) analysis of longitudinal data in biostatistics via GLMM. The underlying statistical principle linking these disparate fields is dynamic empirical Bayes developed in [3] on the foundational works on empirical Bayes by Robbins [1] and Stein [2]. Although survival analysis may seem to be a more natural candidate from biostatistics to connect to bond or loan default—and indeed we have provided such connections in Section 2—a major difference between survival data in biostatistics and default data in finance lies in the event rate, for which the latter typically has very low rate. In fact, importance sampling is often needed to simulate the "rare events" of loan default in managing portfolio credit risk; see [104]. Another difference lies in the objectives of the statistical analysis. In biostatistics the goal is to compare survival patterns of different groups (e.g., one receiving an experimental treatment and the other receiving standard of care in cancer patients), whereas in finance the goal is to predict default and to use the prediction for pricing, hedging, and risk management. It should be noted that the complications in the analysis of survival data in biostatistics are mostly caused by censoring. On the other hand, with loans, censoring is not an issue if one introduces competing risks, as shown in Section 6.7.

Acknowledgment

References

[1] H. Robbins, "An empirical Bayes approach to statistics," in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, vol. I*, pp. 157–163, University of California Press, Los Angeles, Calif, USA, 1956.

[2] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, vol. I*, pp. 197–206, University of California Press, Los Angeles, Calif, USA, 1956.

[3] T. L. Lai, Y. Su, and K. H. Sun, "Dynamic empirical Bayes models and their applications to longitudinal data analysis and prediction," *Statistica Sinica*. In press.

[4] N. E. Breslow and D. G. Clayton, "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 9–25, 1993.

[5] T. R. Fleming and D. P. Harrington, *Counting Processes and Survival Analysis*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1991.

[6] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding, *Statistical Models based on Counting Processes*, Springer Series in Statistics, Springer, New York, NY, USA, 1993.

[7] O. Aalen, "Nonparametric inference in connection with multiple decrement models," *Scandinavian Journal of Statistics*, vol. 3, no. 1, pp. 15–27, 1976.

[8] W. Nelson, "Theory and applications of hazard plotting for censored failure data," *Technometrics*, vol. 14, no. 4, pp. 945–966, 1972.

[9] E. I. Altman, "Measuring corporate bond mortality and performance," *Journal of Finance*, vol. 44, no. 4, pp. 909–922, 1989.

[10] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, vol. 53, pp. 457–481, 1958.

[11] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society Series B*, vol. 34, pp. 187–220, 1972.

[12] W. R. Lane, S. W. Looney, and J. W. Wansley, "An application of the Cox proportional hazards model to bank failure," *Journal of Banking & Finance*, vol. 10, no. 4, pp. 511–531, 1986.

[13] J. D. Kalbeisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data-Chichester-Brisbane*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 1980.

[14] S. H. Lee and J. L. Urrutia, "Analysis and prediction of insolvency in the propertyliability insurance industry: a comparison of logit and hazard models," *Journal of Risk and Insurance*, vol. 63, no. 1, pp. 121–130, 1996.

[15] C. G. McDonald and L. M. Van de Gucht, "High-yield bond default and call risks," *The Review of Economics and Statistics*, vol. 81, no. 3, pp. 409–419, 1999.

[16] W. H. Beaver, "Market prices, financial ratios, and the prediction of failure," *Journal of Accounting Research*, vol. 6, no. 2, pp. 179–192, 1968.

[17] T. Shumway, "Forecasting bankruptcy more accurately: a simple hazard model," *The Journal of Business*, vol. 74, no. 1, pp. 101–124, 2001.

[18] S. Chava and R. A. Jarrow, "Bankruptcy prediction with industry effects," *Review of Finance*, vol. 8, pp. 537–569, 2004.

[19] S. A. Hillegeist, E. K. Keating, D. P. Cram, and K. G. Lundstedt, "Assessing the probability of bankruptcy," *Review of Accounting Studies*, vol. 9, pp. 5–34, 2004.

[20] D. R. Cox, "Partial likelihood," *Biometrika*, vol. 62, no. 2, pp. 269–276, 1975.

[21] N. E. Breslow, "Covariance analysis of censored survival data," *Biometrics*, vol. 30, no. 1, pp. 89–99, 1974.

[22] T. L. Lai and H. Xing, *Statistical Models and Methods for Financial Markets*, Springer Texts in Statistics, Springer, New York, NY, USA, 2008.

[23] D. Duffe and K. J. Singleton, *Credit Risk: Pricing, Measurement, and Management*, Princeton Series in Finance, Princeton University Press, 2003.

[24] D. Lando, *Credit Risk Modeling: Theory and Applications*, Princeton Series in Finance, Princeton University Press, 2004.

[25] R. C. Merton, "On the pricing of corporate debt: the risk structure of interest rates," *Journal of Finance*, vol. 29, no. 2, pp. 449–470, 1974.

[26] F. Black and J. C. Cox, "Valuing corporate securities: some effects of bond indenture provisions," *Journal of Finance*, vol. 31, no. 2, pp. 351–367, 1976.

[27] O. Aalen, "Nonparametric estimation of partial transition probabilities in multiple decrement models," *The Annals of Statistics*, vol. 6, no. 3, pp. 534–545, 1978.

[28] O. O. Aalen and S. Johansen, "An empirical transition matrix for non-homogeneous Markov chains based on censored observations," *Scandinavian Journal of Statistics*, vol. 5, no. 3, pp. 141–150, 1978.

[29] T. R. Fleming, "Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks," *The Annals of Statistics*, vol. 6, no. 5, pp. 1057–1070, 1978.

[30] O. A. Vasicek, "Probability of loss on a loan portfolio," Working Paper, KMV Corporation, 1987.

[31] O. A. Vasicek, "The distribution of loan portfolio value," *RISK*, vol. 15, no. 12, pp. 160–162, 2002.

[32] P. Schonbucher, "Factor models for portfolio credit risk," Working Paper, Department of Statistics, Bonn University, 2000.

[33] D. X. Li, "On default correlation: a copula function approach," *The Journal of Fixed Income*, vol. 9, no. 4, pp. 43–54, 2000.

[34] O. A. Vasicek, "Limiting loan loss probability distribution," Working Paper, KMV Corporation, 1991.

[35] M. B. Gordy, "A comparative anatomy of credit risk models," *Journal of Banking & Finance*, vol. 24, no. 1-2, pp. 119–149, 2000.

[36] R. Frey and A. J. McNeil, "Dependent defaults in models of portfolio credit risk," *Journal of Risk*, vol. 6, no. 1, pp. 59–92, 2003.

[37] F. Salmon, "The formula that killed Wall Street," *Significance*, vol. 9, no. 1, pp. 16–20, 2012.

[38] J. W. Vaupel, K. G. Manton, and E. Stallard, "The impact of heterogeneity in individual frailty on the dynamics of mortality," *Demography*, vol. 16, no. 3, pp. 439–454, 1979.

[39] K. G. Manton, E. Stallard, and J. W. Vaupel, "Methods for comparing the mortality experience of heterogeneous populations," *Demography*, vol. 18, no. 3, pp. 389–410, 1981.

[40] K. G. Manton, E. Stallard, and J. W. Vaupel, "Alternative models for the heterogeneity of mortality risks among the aged," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 635–644, 1986.

[41] J. J. Heckman and B. Singer, "Econometric duration analysis," *Journal of Econometrics*, vol. 24, no. 1-2, pp. 63–132, 1984.

[42] J. Heckman and B. Singer, "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, vol. 52, no. 2, pp. 271–320, 1984.

[43] P. Hougaard, "Life table methods for heterogeneous populations: distributions describing the heterogeneity," *Biometrika*, vol. 71, no. 1, pp. 75–83, 1984.

[44] P. Hougaard, "Survival models for heterogeneous populations derived from stable distributions," *Biometrika*, vol. 73, no. 2, pp. 387–396, 1986.

[45] D. Clayton and J. Cuzick, "Multivariate generalizations of the proportional hazards model," *Journal of the Royal Statistical Society Series A*, vol. 148, no. 2, pp. 82–117, 1985.

[46] D. Duffie, A. Eckner, G. Horel, and L. Saita, "Frailty correlated default," *Journal of Finance*, vol. 64, no. 5, pp. 2089–2123, 2009.

[47] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, Springer Series in Statistics, Springer, New York, NY, USA, 2005.

[48] G. Celeux and J. Diebolt, "The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem," *Computational Statistics Quarterly*, vol. 2, pp. 73–82, 1985.

[49] G. C. G. Wei and M. A. Tanner, "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 699–704, 1990.

[50] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[51] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, New York, NY, USA, 2nd edition, 2004.

[52] D. Duffie, L. Saita, and K. Wang, "Multi-period corporate default prediction with stochastic covariates," *Journal of Financial Economics*, vol. 83, no. 3, pp. 635–665, 2007.

[53] M. Davis and V. Lo, "Infectious defaults," *Quantitative Finance*, vol. 1, no. 4, pp. 382–387, 2001.

[54] S. Deng, K. Giesecke, and T. L. Lai, "Sequential importance sampling and resampling for dynamic portfolio credit risk," *Operations Research*, vol. 60, no. 1, pp. 78–91, 2012.

[55] O. Aalen, "A model for nonparametric regression analysis of counting processes," in *Mathematical Statistics and Probability Theory*, vol. 2 of *Lecture Notes in Statistics*, pp. 1–25, Springer, New York, NY, USA, 1980, Proceedings of the 6th International Conference, Wisła, Poland, 1978.

[56] F. W. Huffer and I. W. McKeague, "Weighted least squares estimation for Aalen's additive risk model," *Journal of the American Statistical Association*, vol. 86, no. 413, pp. 114–129, 1991.

[57] A. T. Payandeh Najafabadi, "A new approach to the credibility formula," *Insurance: Mathematics & Economics*, vol. 46, no. 2, pp. 334–338, 2010.

[58] H. Bühlmann, "Experience rating and credibility," *The Astin Bulletin*, vol. 4, no. 3, pp. 199–207, 1967.

[59] H. Bühlmann, "Experience rating and credibility," *The Astin Bulletin*, vol. 5, no. 2, pp. 157–169, 1969.

[60] W. James and C. Stein, "Estimation with quadratic loss," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 361–379, University of California Press, Berkeley, Calif, USA, 1961.

[61] H. Robbins, "Some thoughts on empirical Bayes estimation," *The Annals of Statistics*, vol. 11, no. 3, pp. 713–723, 1983.

[62] H. Bühlmann and A. Gisler, *A Course in Credibility Theory and Its Applications*, Universitext, Springer, Berlin, Germany, 2005.

[63] C. A. Hachemeister, "Credibility for regression models with application to trend," in *Credibility: Theory and Applications (Proceedings of the Berkeley Actuarial Research Conference on Credibility, University of California, Berkeley, Calif, USA, 1974; dedicated to E. A. Lew)*, P. M. Kahn, Ed., pp. 129–169, Academic Press, New York, 1975, With a discussion by Al Quirin.

[64] E. W. Frees, V. R. Young, and Y. Luo, "A longitudinal data analysis interpretation of credibility models," *Insurance: Mathematics & Economics*, vol. 24, no. 3, pp. 229–247, 1999.

[65] E. W. Frees, V. R. Young, and Y. Luo, "Case studies using panel data models," *North American Actuarial Journal*, vol. 5, no. 4, pp. 24–42, 2001.

[66] W. S. Jewell, "Credible means are exact Bayesian for exponential families," *The Astin Bulletin*, vol. 8, pp. 77–90, 1974.

[67] A. E. Renshaw, "Actuarial graduation practice and generalised linear and non-linear models," *Journal of the Institute of Actuaries*, vol. 118, no. 2, pp. 295–312, 1991.

[68] S. Haberman and A. E. Renshaw, "Generalized linear models and actuarial science," *Journal of the Royal Statistical Society Series D*, vol. 45, no. 4, pp. 407–436, 1996.

[69] J. A. Nelder and R. J. Verrall, "Credibility theory and generalized linear models," *The Astin Bulletin*, vol. 27, no. 1, pp. 71–82, 1997.

[70] Z. Landsman and U. E. Makov, "Credibility evaluation for the exponential dispersion family," *Insurance: Mathematics & Economics*, vol. 24, no. 1-2, pp. 23–29, 1999.

[71] Z. Landsman and U. E. Makov, "On credibility evaluation and the tail area of the exponential dispersion family," *Insurance: Mathematics & Economics*, vol. 27, no. 3, pp. 277–283, 2000.

[72] K. Antonio and J. Beirlant, "Actuarial statistics with generalized linear mixed models," *Insurance: Mathematics & Economics*, vol. 40, no. 1, pp. 58–76, 2007.

[73] T. L. Lai and K. H. Sun, "Evolutionary credibility theory: a generalized linear mixed modeling approach," *The North American Actuarial Journal*. In press.

[74] S. L. Zeger and B. Qaqish, "Markov regression models for time series: a quasi-likelihood approach," *Biometrics*, vol. 44, no. 4, pp. 1019–1031, 1988.

[75] T. L. Lai and M.-C. Shih, "Nonparametric estimation in nonlinear mixed effects models," *Biometrika*, vol. 90, no. 1, pp. 1–13, 2003.

[76] R. Wolfinger and M. O'Connell, "Generalized linear mixed models a pseudo-likelihood approach," *Journal of Statistical Computation and Simulation*, vol. 48, no. 3-4, pp. 233–243, 1993.

[77] Q. Liu and D. A. Pierce, "A note on Gauss-Hermite quadrature," *Biometrika*, vol. 81, no. 3, pp. 624–629, 1994.

[78] A. Yafune, M. Takebe, and H. Ogata, "A use of Monte Carlo integration for population pharmacokinetics with multivariate population distribution," *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 26, pp. 103–123, 1998.

[79] T. L. Lai and M.-C. Shih, "A hybrid estimator in nonlinear and generalised linear mixed effects models," *Biometrika*, vol. 90, no. 4, pp. 859–879, 2003.

[80] T. L. Lai, M.-C. Shih, and S. P. Wong, "A new approach to modeling covariate effects and individualization in population pharmacokinetics-pharmacodynamics," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 33, no. 1, pp. 49–74, 2006.

[81] T. L. Lai, M.-C. Shih, and S. P.-S. Wong, "Flexible modeling via a hybrid estimation scheme in generalized mixed models for longitudinal data," *Biometrics*, vol. 62, no. 1, pp. 159–318, 2006.

[82] T. L. Lai and S. P.-S. Wong, "Statistical models for the Basel II internal ratings-based approach to measuring credit risk of retail products," *Statistics and its Interface*, vol. 1, no. 2, pp. 229–241, 2008.

[83] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[84] M. S. Pepe and G. L. Anderson, "A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data," *Communications in Statistics B*, vol. 23, no. 4, pp. 939–951, 1994.

[85] M. S. Pepe and D. Couper, "Modeling partly conditional means with longitudinal data," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 991–998, 1997.

[86] B. Efron and C. Morris, "Data analysis using Stein's estimator and its generalizations," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 311–319, 1975.

[87] B. Efron and C. Morris, "Stein's paradox in statistics," *Scientific American*, vol. 236, no. 5, pp. 119–127, 1977.

[88] L. D. Brown, "In-season prediction of batting averages: a field test of empirical Bayes and Bayes methodologies," *The Annals of Applied Statistics*, vol. 2, no. 1, pp. 113–152, 2008.

[89] C. Hsiao, C. Kim, and G. Taylor, "A statistical perspective on insurance rate-making," *Journal of Econometrics*, vol. 44, no. 1-2, pp. 5–24, 1990.

[90] M. West, P. J. Harrison, and H. S. Migon, "Dynamic generalized linear models and Bayesian forecasting," *Journal of the American Statistical Association*, vol. 80, no. 389, pp. 73–83, 1985.

[91] A. T. Payandeh Najafabadi, "A new approach to the credibility formula," *Insurance: Mathematics & Economics*, vol. 46, no. 2, pp. 334–338, 2010.

[92] Basel Committee on Banking Supervision, "An explanatory note on the Basel II IRB risk weight functions," 2005, http://www.bis.org/bcbs/irbriskweight.pdf.

[93] C. A. Calhoun and Y. Deng, "A dynamic analysis of fixed- and adjustable-rate mortgage terminations," *Journal of Real Estate Finance and Economics*, vol. 24, no. 1-2, pp. 9–33, 2002.

[94] J. P. Fine and R. J. Gray, "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 496–509, 1999.

[95] R. J. Gray, "A class of K-sample tests for comparing the cumulative incidence of a competing risk," *The Annals of Statistics*, vol. 16, no. 3, pp. 1141–1154, 1988.

[96] J. P. Klein and Y. Shu, "Multi-state models for bone marrow transplantation studies," *Statistical Methods in Medical Research*, vol. 11, no. 2, pp. 117–139, 2002.

[97] T. H. Scheike, M.-J. Zhang, and T. A. Gerds, "Predicting cumulative incidence probability by direct binomial regression," *Biometrika*, vol. 95, no. 1, pp. 205–220, 2008.

[98] J. P. Klein and P. K. Andersen, "Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function," *Biometrics*, vol. 61, no. 1, pp. 223–229, 2005.

[99] F. Graw, T. A. Gerds, and M. Schumacher, "On pseudo-values for regression analysis in competing risks models," *Lifetime Data Analysis*, vol. 15, no. 2, pp. 241–255, 2009.

[100] T. L. Lai, Y. Su, and K. H. Sun, "A new approach to dynamic panel data and its applications to default modeling," Working Paper, Department of Statistics, Stanford Unviersity, 2012.

[101] M. Gu and T. L. Lai, "Repeated significance testing with censored rank statistics in interim analysis of clinical trials," *Statistica Sinica*, vol. 8, no. 2, pp. 411–428, 1998.

[102] Y. Jin, T. L. Lai, and S. Yong, "Evaluating econometric forecasts, with applications to mortgage default prediction and forecasting insurance claims," Working Paper, Department of Statistics, Stanford Unviersity, 2012.

[103] T. L. Lai and S. P.-S. Wong, "Statistical models for the Basel II internal ratings-based approach to measuring credit risk of retail products," *Statistics and its Interface*, vol. 1, no. 2, pp. 229–241, 2008.

[104] S. Deng, K. Giesecke, and T. L. Lai, "Sequential importance sampling and resampling for dynamic portfolio credit risk," *Operations Research*, vol. 60, no. 1, pp. 78–91, 2012.