Monotonic Estimation for the Survival Probability over a Risk-Rated Portfolio by Discrete-Time Hazard Rate Models

Yang, Bill Huajian

18 March 2019

Monotonic Estimation for the Survival Probability over a Risk-Rated Portfolio by Discrete-Time Hazard Rate Models

Bill Huajian Yang

***Abstract*—Monotonic estimation for the survival probability of a loan in a risk-rated portfolio is based on the observation arising, for example, from loan pricing that a loan with a lower credit risk rating is more likely to survive than a loan with a higher credit risk rating, given the same additional risk covariates. Two probit-type discrete-time hazard rate models that generate monotonic survival probabilities are proposed in this paper. The first model calculates the discrete-time hazard rate conditional on systematic risk factors. As for the Cox proportion hazard rate model, the model formulates the discrete-time hazard rate by including a baseline component. This baseline component can be estimated outside the model in the absence of model covariates using the long-run average discrete-time hazard rate. This results in a significant reduction in the number of parameters to be otherwise estimated inside the model. The second model is a general form model where loan level factors can be included. Parameter estimation algorithms are also proposed. The models and algorithms proposed in this paper can be used for loan pricing, stress testing, expected credit loss estimation, and modeling of the probability of default term structure.**

***Index terms*—loan pricing, survival probability, Cox proportion hazard rate model, baseline hazard rate, forward probability of default, probability of default term structure**

I. INTRODUCTION

Monotonic learning is a learning process based on the prior knowledge of the monotone relationship between input and output. For example, we expect the loss for a portfolio to be higher in an economic downturn, and we also expect a loan with a lower credit risk rating to survive more likely than a loan with a higher credit risk rating, given the same additional risk covariates. Examples of monotonic learning include isotonic regression [1, 2], classification trees [3], rule learning [4], binning [5], and deep lattice network [6].

Let $\{R_i\}_{i=1}^{G+1}$ denote a rating system for a portfolio ranking the credit risk (likelihood to default) of loans. Assume that higher index ratings carry a higher default risk; that is, loans with a rating R_i are less likely to default than loans with a rating R_{i+1}, given the same additional risk covariates. Rating R_{G+1} denotes the worst rating (i.e., the default rating).

Let $x^j = (x_1^j, x_2^j, \dots, x_m^j)$ denote a vector of covariates for the risk profile of a loan in the period $(t_{j-1}, t_j]$. For a loan with an initial rating R_i at the initial observation time t_0, let $p_{ik} = p_{ik}(x^k)$ denote the forward probability of default (PD) in the period $(t_{k-1}, t_k]$ conditional on x^k, given that the loan survives the period $[t_0, t_{k-1}]$. Forward PD is also called discrete-time hazard rate [7]. Hereafter, we will use the terms "forward PD" and "discrete-time hazard rate" interchangeably.

Let s_{ik} denote the probability that the loan survives the period $(t_0, t_k]$, given the multivariate information time series $H_k = \{x^j, 1 \le j \le k\}$, and let $cp_{i\,k}$ denote the cumulative PD over the period $(t_0, t_k]$ given H_k. The survival probability is said to be monotonic over the rating system if (1.1) holds for each period index $k \ge 1$,

$$s_{1k} \ge s_{2k} \ge \cdots \ge s_{Gk}, \tag{1.1}$$

given the same H_k between ratings. That is, a loan with a lower index rating is more likely to survive, given the same time series H_k. This monotonicity is a fundamental requirement for a PD term structure model.

Under the assumption that the forward PD $p_{ij}(x^j)$, given the covariate vector x^j, is the same as the forward PD p_{ij} conditional on the time series H_j, the marginal PD for the period $(t_{k-1}, t_k]$ given H_k, for a loan with an initial rating R_i at the initial observation time t_0, is equal to $(1 - cp_{i\,k-1})p_{ik}$. Therefore, we have

$$\begin{aligned} cp_{i\,k} &= cp_{i\,k-1} + (1 - cp_{i\,k-1})p_{ik} \\ &\Rightarrow 1 - cp_{i\,k} = (1 - cp_{i\,k-1})(1 - p_{ik}). \end{aligned} \tag{1.2}$$

Then, by induction on the time index k using the relation $s_{ik} = 1 - cp_{ik}$, we have the following equation [8]:

$$s_{ik} = (1 - p_{i1})(1 - p_{i2}) \dots (1 - p_{ik}). \tag{1.3}$$

This means that (1.1) holds whenever (1.4) holds for forward PD for each period index $k \ge 1$:

$$p_{1k} \le p_{2k} \le \cdots \le p_{Gk}. \tag{1.4}$$

One of the most important hazard rate models is the Cox proportion hazard rate model [9], which is implemented by SAS procedure PROC PHREG [10]. One can use this SAS procedure, with rating as a class variable, to estimate forward PD between ratings, hence the survival probability by (1.3). Nevertheless, the baseline component of this model is in this procedure estimated either by the Kaplan–Meier method or by the Breslow method [11]. Monotonicity (1.4) is generally not guaranteed, without additional monotonic constraints being imposed for the baseline component.

Main Results. In this paper, we propose two probit-type discrete-time hazard rate models. Both models generate monotonic discrete-time hazard rates in the sense of (1.4). The first model (i.e., model (3.1) in Section III) estimates the discrete-time hazard rate conditional on systematic risk factors, with default points as the baseline component, whereas the second model (i.e., model (3.4) in Section III) is a general form model where loan level factors can be included. Monotonicity (1.4) is achieved by appropriate monotonic constraints being imposed for the baseline component for the first model and for the intercepts for the second model. Algorithms for parameter estimation are proposed.

The advantage of the first model is that the baseline hazard rate component can be estimated outside the model using the long-run average discrete-time hazard rate, in the absence of model covariates. This leads to a significant reduction in the number of parameters to be otherwise estimated inside the model.

The key ideas for the proposed algorithms are based on the reparameterization of the baseline component for the first model (see Algorithm 5.1) and the intercept component for the second model (see Algorithm 5.4) so that the required monotonic constraints for these components [i.e., (3.2) and (3.5)] are automatically satisfied. This transforms the original constrained optimization into a simpler tractable mathematical programming problem.

This paper is organized as follows. In Section II, we briefly review the hazard rate models. Two probit-type discrete-time hazard rate models are proposed in Section III. Log-likelihood functions are shown in Section IV. Model parameter estimation algorithms based on the maximum likelihood are proposed in Section V. An empirical example is provided in Section VI, where we train a discrete-time hazard rate model for a wholesale portfolio using the first proposed model.

II. A Brief Review of the Discrete-Time Hazard Rate Models

In [7], Allison proposed a discrete-time hazard rate model:

$$p_{ik}(x) = F(b_{ik} + a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m),$$

where F denotes the cumulative density function for logistic distribution. The intercept is time-dependent, whereas variable coefficients are time-independent and are differentiated between ratings. One can use the SAS logistic regression procedure [10], with rating and term number as two class variables, to train this model for a given sample. However, the survival probability generated by this model is not necessarily monotonic, without additional monotonic constraints (e.g., (3.5) in Section III) being imposed for the intercepts.

With the Cox proportion hazard rate model [9], the continuous-time hazard rate is estimated by

$$h(t) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m),$$

where covariate coefficients can be estimated robustly using the partial likelihood method in the absence of $h_0(t)$ (i.e., even when $h_0(t)$ is unspecified). One can use the SAS hazard rate regression procedure PROC PHREG [10], with rating as a class variable, to estimate the baseline and covariate coefficients. With this procedure, the baseline is estimated by either the Kaplan–Meier method or the Breslow method [11]. Monotonicity (1.4) is not necessarily satisfied, without additional monotonic constraints being imposed for the baseline component.

A discrete-time hazard rate model derived under the Merton model framework was proposed in [8] when scenarios are given by systematic risk factors (common to all loans), as described below.

For a loan with a nondefault risk rating R_i at initial time t_0, we assume that the loan has survived the period $[t_0, t_{k-1}]$, and we consider its forward PD in the period $(t_{k-1}, t_k]$. Assume that the default risk for the loan in the period $(t_{k-1}, t_k]$ is driven by a latent normalized random variable $z_{ik}(t)$ that splits into two parts:

$$z_{ik}(t) = s(t)\sqrt{\rho_{ik}} + \varepsilon_{ik}(t)\sqrt{1-\rho_{ik}}, \qquad (2.1)$$

where $0 < \rho_{ik} < 1$ and $s(t) \sim N(0,1)$ represents the systematic risk (time-varying, common to all ratings) at time t, whereas $\varepsilon_{ik}(t) \sim N(0,1)$ is the idiosyncratic risk, independent of $s(t)$.

By Merton's model [12, 13], there exists a threshold value c_{ik}, called default point, for initial rating R_i, such that the loan will default in the kth period $(t_{k-1}, t_k]$ when $z_{ik}(t)$ falls below the threshold value c_{ik}. Here we assume that loans within the same initial rating are riskhomogeneous in the sense that the default point c_{ik} in the kth period is the same for all these loans.

For simplicity, we suppress the time label t from $z_{ik}(t)$, $s(t)$, and $\varepsilon_{ik}(t)$ and write them as z_{ik}, s, and ε_{ik}, respectively. Denote by $E_e[\Phi(a+be)]$ the expected value of $\Phi(a+be)$ with respect to a random variable e. The following lemma is important.

Lemma 2.1. ([14]) $E_e[\Phi(a+be)] = \Phi\left(a/\sqrt{1+b^2}\right)$, where $e \sim N(0,1)$. □

For a loan with an initial risk rating R_i at time t_0, let $p_{ik}(s)$ denote the kth forward PD given the systematic risk s as a latent variable in the period $(t_{k-1}, t_k]$. By applying Lemma 2.1 to (2.1), we have the following equation [8]:

$$p_{ik}(s) = \Phi(c_{ik}\sqrt{1+r_{ik}^2} - r_{ik}s), \qquad (2.2)$$

where $r_{ik} = \sqrt{\rho_{ik}}/\sqrt{1-\rho_{ik}}$. The default point c_{ik} satisfies the equation $E_s[p_{ik}(s)] = \Phi(c_{ik})$, by Lemma 2.1. Thus, c_{ik} can be estimated by $c_{ik} = \Phi^{-1}(p_{ik}^0)$, where p_{ik}^0 denotes the long-run average of $p_{ik}(s)$.

In addition, for simplicity, we write a macroeconomic scenario $x^j = (x_1^j, x_2^j, \ldots, x_m^j)$ for the period $(t_{j-1}, t_j]$ as $x = (x_1, x_2, \ldots, x_m)$. Let $c(x) = a_1x_1 + a_2x_2 + \cdots + a_mx_m$, and

$$ci(x) = (c(x) - u)/v, \qquad (2.3)$$

where u and v denote, respectively, the mean and standard deviation of the linear combination $c(x)$. Assume that the systematic risk factor s splits as

$$s = -\lambda ci(x) - e\sqrt{1-\lambda^2}, \qquad (2.4)$$

where $e \sim N(0,1)$, $0 < \lambda < 1$. Then, $p_{ik}(s) = \Phi\{c_{ik}\sqrt{1+r_{ik}^2} + r_{ik}[\lambda ci(x) + \sqrt{1-\lambda^2}e]\}$. Let $g_{ik} = r_{ik}\lambda/\sqrt{1+r_{ik}^2(1-\lambda^2)}$. By Lemma 2.1 again, we have the following equation, assuming that e is independent of $ci(x)$ [8]:

$$\begin{aligned} p_{ik}(x) &= E_e[p_{ik}(s)|x] \\ &= \Phi\left[c_{ik}\sqrt{1+g_{ik}^2} + g_{ik}ci(x)\right]. \end{aligned} \qquad (2.5)$$

We write in the remaining of the paper g_{ik} by r_{ik}. Then, (2.5) becomes

$$p_{ik}(x) = \Phi\left[c_{ik}\sqrt{1+r_{ik}^2} + r_{ik}ci(x)\right]. \qquad (2.6)$$

Model (2.6) is the hazard rate model we proposed in [8]. This model formulates forward PD as being given by three risk components: the index score $ci(x)$ approximating the systematic risk for the portfolio, the baseline long-run forward PD (via the default point c_{ik}), and the sensitivity parameter r_{ik} that measures the responsiveness for a rating in responding to the changes of the systematic risk index $ci(x)$.

One advantage of model (2.6) is that the baseline component $\{\Phi(c_{ik})\}$ can be estimated outside the model using the long-run forward PDs, leading to a significant reduction in the number of parameters to be otherwise estimated inside the model. However, monotonicity (1.4) is not necessarily guaranteed, without additional monotonic constraints (see (3.2) in Section III) being imposed for the baseline component.

III. The Proposed Discrete-Time Hazard Rate Model

A. *The Proposed Forward PD Models with Systematic Risk Covariates Only*

In order to ensure that lower credit risk ratings are more likely to survive, given the same additional risk covariates, the following two conditions, (a) and (b), are imposed to model (2.6), for each term index k.

(a) The sensitivity parameter r_{ik} is constant between ratings. This is equivalent to the assumption that ρ_{ik} in (2.1) is constant across ratings. Then, (2.6) becomes

$$p_{ik}(x) = \Phi[c_{ik}\sqrt{1+r_k^2} + r_k ci(x)]. \qquad (3.1)$$

(b) $c_{1k} \le c_{2k} \le \cdots \le c_{G\,k}$. (3.2)

B. *The Proposed General Forward PD Models with Loan-Specific Covariates*

For a loan with an initial rating R_i at initial time t_0, let $\{x_1, x_2, \ldots, x_m\}$ denote the macroeconomic variables, which are common to all ratings, and let $\{x_{m+1}, x_{m+2}, \ldots, x_{m+p}\}$ denote the loan-specific variables. Let $x = (x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_{m+p})$. We assume that there exists a latent variable y_{ik} of the form

$$y_{ik} = -b_{ik} - r_k c(x) + \varepsilon,$$

such that a loan with an initial rating R_i will default in the period $(t_{k-1}, t_k]$ if $y_{ik} < 0$, where $\varepsilon \sim N(0,1)$, and $c(x) = a_1x_1 + a_2x_2 + \cdots + a_{m+p}x_{m+p}$ subject to

$$a_1^2 + a_2^2 + \cdots + a_{m+p}^2 = 1. \qquad (3.3)$$

Constraint (3.3) is imposed to prevent disturbances in parameter estimation caused by free switches for a scalar between the coefficient vector $(a_1, a_2, \ldots, a_m)$ and the sensitivity parameters $\{r_k\}$.We, thus, have

$$p_{ik}(x) = P(y_{ik} < 0 \,|x) = P[\varepsilon < b_{ik} + r_k c(x)].$$
$$\Longrightarrow p_{ik}(x) = \Phi[b_{ik} + r_k c(x)]. \quad (3.4)$$

Forward PDs generated by (3.4) satisfy (1.4) when the constraints below are imposed for each term index k:

$$b_{1k} \le b_{2k} \le \cdots \le b_{Gk}. \quad (3.5)$$

IV. The Log-Likelihood Functions

A. The Log-Likelihood for Model (3.1) Subject to (3.2) with Macroeconomic Covariates Only

Let n_{ik} denote the number of loans that survive the period $(t_0\,, t_{k-1}]$ with an initial risk rating R_i at initial time t_0, and let d_{ik} denote the number of defaulters of these n_{ik} loans in the period $(t_{k-1}\,, t_k]$. For models including only macroeconomic variables, such as model (3.1), the log-likelihood for the kth forward term is

$$FL_{i,k} = \sum_{t_k} \{(n_{ik} - d_{ik})\log[1 - p_{ik}(x)] + d_{ik}\log([p_{ik}(x)]\}, \quad (4.1)$$

with $(t_{k-1}\,, t_k]$ sliding through the sample time window. Here, we assume that the term default count d_{ik} follows a binomial distribution given the systematic risk factors $x = (x_1, x_2, \ldots, x_m)$. Expression (4.1) holds, up to a constant given by the logarithms of some binomial coefficients. (4.1) is essentially the Bernoulli log-likelihood. We call $FL_{i,k}$ in (4.1) the forward log-likelihood.

B. The Log-Likelihood for Model (3.4) Subject to (3.5) with Loan-Specific Covariates

Similarly, let n_{ik} denote the number of loans in the portfolio that survive the interval $[t_0\,, t_{k-1}]$ with an initial rating R_i. Let y_{ikj} be an indicator, for the jth loan with an initial risk rating R_i, with value 1 if the loan defaults in the kth forward period $(t_{k-1}\,, t_k]$ and zero otherwise. For models with loan-specific covariates, such as model (3.4), the log-likelihood for the kth forward period is given by

$$FL_{i,k} = \sum_{t_k} \sum_j \{(1 - y_{ikj})\log[1 - p_{ik}(x)] + y_{ikj}\log([p_{ik}(x)]\}, \quad (4.2)$$

with $(t_{k-1}\,, t_k]$ sliding through the sample time window. We call $FL_{i,k}$ in (4.2) the forward log-likelihood at the loan level.

Let $L_i(h, h + k)$ denote the log-likelihood for loans with initial rating R_i at t_0 for the combined period $[t_h\,, t_{k+h}]$, given that the loans survive the period $[t_0\,, t_{h-1}]$. Here, the period $[t_h\,, t_{k+h}]$ slides through the sample time window. Similarly, let $L(h, h + k)$ be the log-likelihood for the period $[t_h\,, t_{k+h}]$ for all loans in the portfolio with a nondefault initial risk rating at time t_0, given that the loans survive the period $[t_0\,, t_{h-1}]$, where $[t_h\,, t_{k+h}]$ slides through the sample time window.

The following equation holds under the assumption that there is no withdrawal for the sample [8]:

$$L_i(h, h + k) = FL_{i,h+1} + FL_{i,h+2} + \cdots + FL_{i,h+k}. \quad (4.3)$$

V. Algorithms for Parameter Estimation by Maximum Likelihood

A. Algorithms for Model (3.1) Subject to (3.2) with Macroeconomic Covariates Only

Estimating Default Points $\{\mathbf{c_{ik}}\}$ Subject to $(\mathbf{3.2})$. Given the sample, the realized default rate in period $(t_{k-1}\,, t_k]$ for loans with an initial risk rating is $r_{ik} = d_{ik}/n_{ik}$. We estimate $\{c_{ik}\}$ subject to (3.2) by minimizing, for each term number k, the sum squares error below:

$$SSE_k = \sum_{t_k} \sum_{i=1}^{G} n_{ik}[(r_{ik} - \Phi(c_{ik})]^2, \quad (5.1)$$

with $(t_{k-1}\,, t_k]$ sliding through the sample time window.

Algorithm 5.1. (Monotonic estimation for default points). For each term number k, do the following:

(a) Parameterize c_{ik} as $c_{ik} = (\alpha_1 + \alpha_2 + \cdots + \alpha_i)$, where $\alpha_i \ge 0$ for $2 \le i \le G$. With this parametrization, $\{c_{ik}\}$ satisfies (3.2).
(b) Plug in $\Phi(c_{ik})$ and minimize (5.1) to obtain the estimates for $\{\alpha_i\}_{i=1}^{G}$ and, thus, the estimates for $\{c_{ik}\}$.

This algorithm can be implemented as the monotonic estimation algorithms proposed in [15] using, for example, the SAS procedure PROC NLMIXED [16].

The Variable Covariance Matrix. Given a list of macroeconomic variables $\{x_1, x_2, \ldots, x_m\}$, to be included in the models, calculate the corresponding sample covariance matrix. Let (v_{ij}) denote this covariance matrix.

Initial Parameter Values. Initially, the values for all sensitivity parameters $\{r_k\}$ are set to 1. For macroeconomic coefficients $\{a_j\}$, let $p_k(D\,|x)$ denote the conditional forward PD for a loan in the kth forward period $(t_{k-1}, t_k]$ given *x*. Fit a simple model of the form below:

$$p_k(D\,|x) = \Phi(d_k + a_1\, x_1 + a_2 x_2 + \cdots + a_m x_m), \tag{5.2}$$

targeting the default event for the portfolio in the kth period $(t_{k-1}\,, t_k]$ for some consecutive terms (e.g., $1 \le k \le 4$). This model can be fitted similarly by a simple logistic regression as proposed by Allison in [7] (here, with probit as the link function), using the SAS logistic regression procedure, with term number k as a class variable. When this is done, rescale each a_j by $1/v$, where v is the standard deviation of $a_1\, x_1 + a_2 x_2 + \cdots + a_m x_m$, calculated as

$$v = \sqrt{\textstyle\sum_i \sum_j a_i a_j v_{ij}}, \tag{5.3}$$

where (v_{ij}) is the variable covariance matrix.

Algorithm 5.2 (Estimating parameters in model (3.1) other than default points). Assume that the sample mean for each macroeconomic variable has been removed (i.e., all macroeconomic variables have sample mean zero). Given the default points $\{c_{ik}\}$ satisfying (3.2) and the initial values for $\{r_k\}$ and $\{a_j\}$, do the following.

1a. Given $\{r_k\}$, fit for $\{a_j\}$ by maximizing the pooled log-likelihood below at the portfolio level:

$$LL = \textstyle\sum_{i=1}^{G} \sum_k FL_{i,k}, \tag{5.4}$$

where $FL_{i,k}$ is as (4.1), and $p_{ik}(x)$ is given by (3.1), that is,

$$p_{ik}(x) = \Phi[c_{ik}\sqrt{1 + r_k^2} + r_k ci(x)]. \tag{5.5}$$

We will perform an unconstrained search for the new values for $\{a_j\}$ using their current values. Before the search, the score $c(x)$ (i.e.,., $a_1\, x_1 + a_2 x_2 + \cdots + a_m x_m$), with the current values for $\{a_j\}$, is normalized. $c(x)$ has standard deviation 1. For any new set of values for $\{a_j\}$, the standard deviation for the new score $c(x)$ is given by v in (5.3). For this new score $c(x)$, we have

$$r_k c(x) = (r_k v)[\tfrac{c(x)}{v}] = (r_k v) ci(x).$$

Here, we use the relationship $ci(x) = c(x)/v$ (as the mean of $c(x)$ is zero). This means that, under model (3.1), the sensitivity parameter r_k is scaled up by v in response to the new set values of $\{a_j\}$; that is, $p_{ik}(x)$ in (5.5) becomes

$$\begin{aligned} p_{ik}(x) &= \Phi[c_{ik}\sqrt{1 + (r_k v)^2} + (r_k v) ci(x)] \\ &= \Phi[c_{ik}\sqrt{1 + (r_k v)^2} + r_k c(x)]. \end{aligned} \tag{5.6}$$

Estimate $\{a_j\}$ by maximizing (5.4) with $p_{ik}(x)$ being given by (5.6) and v being given by (5.3). When this is done, rescale $\{a_j\}$ by scalar $1/v$ and rescale $\{r_k\}$ by scalar v accordingly. Note that this rescaling does not change the current value of $p_{ik}(x)$.

1b. Given $\{a_j\}$, fit for $\{r_k\}$ by maximizing the pooled log-likelihood (5.4) at the portfolio level.

1c. Repeat steps 1a and 1b until convergence is reached.

Remark 5.3. In the simplest case when the sensitivity parameter r_k is assumed to be the same for all forward term numbers $k's$, steps 1a, 1b, and 1c can be combined (i.e., run 1a to get $\{a_j\}$ and v). Then, v is the value of the unique sensitivity parameter. Actually, model (3.1) in this case reduces to

$$p_{ik}(x) = \Phi[c_{ik}\sqrt{1 + v^2} + (a_1\, x_1 + a_2 x_2 + \cdots + a_m x_m)]. \tag{5.7}$$

There is no need for an independent sensitivity parameter and no need for step 1c for the iteration.

This algorithm differs from the algorithm proposed in [8], for fitting $\{a_j\}$ and $\{r_k\}$ in model (5.5). The algorithm in [8] fits the macroeconomic coefficients $\{a_j\}$ separately by a separate model, whereas Algorithm 5.2 simply fits both $\{a_j\}$ and $\{r_k\}$ in the same model (5.6).

B. Algorithms for Model (3.4) Subject to (3.5) with Loan-Specific Covariates

Initial Values for Variable Coefficients $\{a_j\}$. Let $p_k(D|x)$ denote the conditional forward PD for a loan in the portfolio for the kth forward period $(t_{k-1}, t_k]$, given $x = (x_1, x_2, \ldots, x_{m+p})$. Fit a simple model by logistic regression with term number k as a class variable (e.g. for $1 \le k \le 4$):

$$p_k(D\,|x) = \Phi\big(d_k + a_1\, x_1 + a_2 x_2 + \cdots + a_{m+p} x_{m+p}\big), \tag{5.8}$$

targeting the default event for the portfolio in the kth forward period $(t_{k-1}, t_k]$. When this is done, rescale $\{a_j\}$ according to (3.3) by scalar $1/\sqrt{a_1^2 + a_2^2 + \cdots + a_{m+p}^2}$.

Initial Values for Sensitivity Parameters $\{r_k\}$. Given the initial values for $\{a_j\}$, form $c(x) = a_1x_1 + a_2x_2 + \cdots + a_{m+p}x_{m+p}$. Use $c(x)$ as the single variable to run for each forward term k a logistic regression model with the initial rating index as a class variable:

$$p_i(D\,|x) = \Phi[\alpha_{ik} + \beta_k c(x)], \quad (5.9)$$

targeting the default event in this forward term period for loans with an initial rating R_i. Set the initial values for r_k by β_k.

Initial Values for Intercepts $\{b_{ik}\}$. Assume that the sample mean for each covariate has been removed. Initialize b_{ik} as $c_{ik}\sqrt{1 + (r_k v)^2}$, where $\{c_{ik}\}$ are the monotonic threshold values in (3.2) and v is the standard deviation of $c(x)$, whereas $\{r_k\}$ are the initial values obtained previously.

Algorithm 5.4 (Parameter estimation for model (3.4)). Assume that the sample mean for each covariate has been removed. Given the initial values for all parameters, do the following.

2a. Given $\{b_{ik}\}$ and $\{r_k\}$, fit for $\{a_j\}$ by maximizing the pooled log-likelihood,

$$LL = \textstyle\sum_{i=1}^{G}\sum_k FL_{i,k}, \quad (5.10)$$

at the portfolio level, where $FL_{i,k}$ is as in (4.2). Rescale $\{a_j\}$ for (3.3) by scalar $1/\sqrt{a_1^2 + a_2^2 + \cdots + a_{m+p}^2}$, and rescale the current values for $\{r_k\}$ by scalar $\sqrt{a_1^2 + a_2^2 + \cdots + a_{m+p}^2}$ accordingly.

2b. Given $\{a_j\}$ and $\{b_{ik}\}$, fit for $\{r_k\}$ by maximizing the pooled log-likelihood (5.10).

2c. Given $\{r_k\}$ and $\{a_j\}$, fit for $\{b_{ik}\}$. For each forward term k, parameterize b_{ik} as $b_{ik} = (\alpha_1 + \alpha_2 + \cdots + \alpha_i)$, where $\alpha_i \geq 0$ for $2 \leq i \leq G$, as in Algorithm 5.1. Then, (3.5) is automatically satisfied under this parameterization. Estimate $\{\alpha_i\}$ by maximizing the pooled log-likelihood:

$$FL_k = FL_{1,k} + FL_{2,k} + \cdots + FL_{G,k}. \quad (5.11)$$

When this is done, we will have estimates for $\{\alpha_i\}$ and, thus, $\{b_{ik}\}$ for the fixed k.

2d. Repeat steps 2a, 2b, and 2c until convergence is reached.

Remark 5.5. In the case when the sensitivity parameter r_k is assumed to be the same for all forward terms $k's$, steps 2a and 2b can be combined (i.e., run step 2a to get $\{a_j\}$), and then $r = \sqrt{a_1^2 + a_2^2 + \cdots + a_{m+p}^2}$ is the value of the unique sensitivity parameter. Actually, model (3.4) in this case reduces to

$$p_{ik}(x) = \Phi[b_{ik} + (a_1\,x_1 + a_2x_2 + \cdots + a_mx_m)]. \quad (5.12)$$

There is no need for an independent sensitivity parameter and no need to rescale $\{a_j\}$ by $1/\sqrt{a_1^2 + a_2^2 + \cdots + a_{m+p}^2}$ in step 2a.

VI. An Empirical Example

In this section, we show an empirical example where we estimate the monotonic survival probability for a wholesale portfolio by a discrete-time hazard rate model (3.1) subject to (3.2). A logistic regression model is trained as a benchmark.

The sample includes the historical data between 2002Q3 and 2016Q3 for a wholesale portfolio of commercial and industrial loans. There are six ratings, with R_6 as the default rating and R_1 as the best quality rating. The sample contains the risk ratings of loans at the end of each quarter between 2002Q3 and 2016Q3. Loans with a nondefault initial risk rating at initial time t_0 are kept for observation of default events for the next 16 quarters. The charts below show the quarterly default rate by rating during the period between 2006Q3 and 2016Q3. Default risk intensified during the financial crisis period between 2008Q1 and 2010Q1. Only a few defaults are observed for the best credit quality ratings R_1 and R_2 at quarters 2008Q3 and 2009Q2, where the realized default rate for R_1 is slightly higher than that for R_2. The overall average quarterly sample default rate is 0.8% for R_1 and 0.6% for R_2. As such, a general logistic regression model, without additional monotonic constraints being imposed for the intercepts, could lead to a counterintuitive prediction between ratings R_1 and R_2.

Figure 1. Realized quarterly default rate for ratings 1, 2, 3

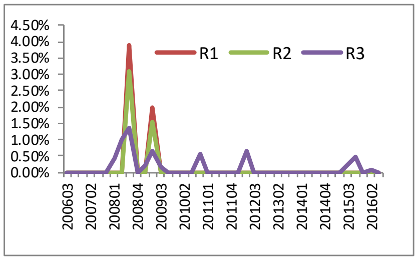
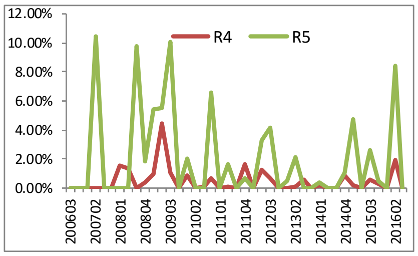


Figure 2. Realized quarterly default rate for ratings 4, 5



Seasonally adjusted macroeconomic data is downloaded from the Federal Reserve website and then appended to this term structure data by matching the calendar quarter in the macroeconomic data with the calendar quarter in the term structure data. Data with quarter time key between 2006Q3 and 2016Q3 is selected. This results in a sample with the following characteristics.

(1) For each nondefault rating R_i and term number $1 \le k \le 16$, the time series sample $\{(d_{ik}, n_{ik})\}$ has 41 data points for 41 quarters between 2006Q3 and 2016Q3, with the time interval $(t_{k-1}, t_k]$ sliding through this time window.
(2) The macroeconomic data is the same for all loans at each specific quarter.

For each macroeconomic variable, its four lagged versions are included: current (L0), lagged one quarter (L1), lagged two quarters (L2), and lagged three quarters (L3). The sample mean is removed from each of these variables.

We fit as follows two probit-type discrete-time hazard rate models.

(1) The logistic regression model served as a benchmark with probit function Φ as the link function. This is a model formulated as follows with rating and term index as two class variables:

$$p_{ik}(x) = \Phi(b_i + c_k + a_{i1}x_1 + a_{i2}x_2 + \dots + a_{im}x_m), \qquad (6.1)$$

where b_i is the intercept corresponding to nondefault rating R_i, $1 \le i \le 5$, and c_k is the intercept corresponding to term index k for the period $(t_{k-1}, t_k]$. The model is fitted using the SAS procedure PROC LOGISTIC [10].

(2) The proposed model (3.1) subject to (3.2) served as the champion model. The sensitivity parameter r_k is kept the same for all terms $1 \le k \le 16$. By Remark 5.3, the model reduces to

$$p_{ik}(x) = \Phi[c_{ik}\sqrt{1+v^2} + (a_1x_1 + a_2x_2 + \dots + a_mx_m)], \qquad (6.2)$$

where v denotes the standard deviation for the linear score $c(x) = a_1x_1 + a_2x_2 + \dots + a_mx_m$. This model is trained using the SAS procedure PROC NLMIXED [16].

We consider models that contain at least two variables but no more than four. Model selection is based on the value -2 log-likelihood (labeled as "2NLK;" lower values are better). The top model based on 2NLK consists of the same three variables below for both models (6.1) and (6.2):

(1) *L0 (Current) GDP*. Growth rate of the US gross domestic product (quarter over quarter annualized by compounding).
(2) *L0 (Current) Unemployment Rate*. Increase of the US civilian unemployment rate (quarter over quarter annualized).
(3) *L3 (Lagged Three Quarters) Volatility Index*. US implied volatility (maximum of daily values per quarter).

Table 1 shows the statistics for model estimation. The value of 2NLK for the proposed champion model is slightly better (lower) than that for the benchmark model.

The risk factor weight w_i for the ith variable in the above sequence is calculated as $w_i = (a_iv_i)/(|a_1v_1| + |a_2v_2| + \dots + |a_mv_m|)$, where v_i denotes the sample standard deviation for the ith variable. The risk factor weight measures the relative contribution for the variable (when standardized to have a standard deviation of one) in the model. As shown in Table 1, the risk factor weight is distributed more evenly between the unemployment rate (Variable 2) and the volatility index (Variable 3) for the proposed model.

Table 1. Model statistics

Model	Risk Factor Weight w1	w2	w3	P-Value pv1	pv2	pv3	2NLK[3]
Logistic[1]	3%	11%	86%	0.13	0.00	0.00	5804
Proposed[2]	10%	44%	45%	0.09	0.00	0.00	5704

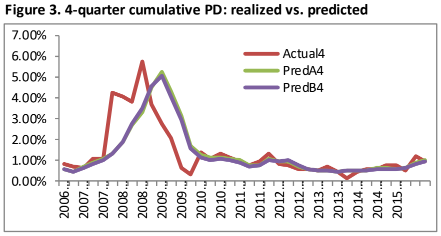
[1] The benchmark model by logistic regression

[2] The simple proposed model

[3] -2 log likelihood

The cumulative realized default rate and cumulative predicted PD are calculated using the formula $cp_{i\,k} = cp_{i\,k-1} + (1 - cp_{i\,k-1})p_{ik}$ at the rating level and then aggregated to the portfolio level. Figures 3 and 4 show plots of the performance charts for the predicted cumulative PD against the actual cumulative default rate at the portfolio level for cumulating 4 and 16 quarters. The RSQ for the predicted cumulative PDs for cumulating 4, 8, 12, and 16 quarters is, respectively, 0.46, 0.68, 0.77, and 0.78 for the benchmark model and 0.44, 0.67, 0.77, and 0.78 for the proposed model.

Figure 3. 4-quarter cumulative PD: realized vs. predicted

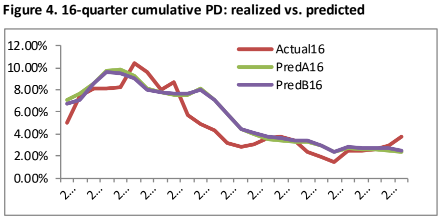


Actual4 - 4-quarter reazlied cumulative PD

PredA4 - 4-quarter predicted cumulative PD by (6.1)

PredB4 - 4-quarter predicted cumulative PD by (6.2)

Figure 4. 16-quarter cumulative PD: realized vs. predicted



Actual16 - 16-quarter reazlied cumulative PD

PredA16 - 16-quarter predicted cumulative PD by (6.1)

PredB16 - 16-quarter predicted cumulative PD by (6.2)

We observed that the proposed model (6.2) performs as good as the benchmark model (6.1). Given its compatible performance, model (6.2) generates monotonic forward PDs (hence, monotonic survival probabilities) between ratings, whereas (6.1) does not. The average quarterly PD predicted over the sample by the benchmark model is 0.9% for rating R_1 and 0.7% for R_2. This is in contrast to 0.7% for R_1 and 0.8% for R_2 in the average quarterly PD predicted by the proposed model.

VII. Conclusions and Future Work

The two probit-type discrete-time hazard rate models proposed in this paper generate monotonic survival probabilities between ratings. The first model focuses on systematic risks and includes only macroeconomic variables. Factorization of the intercepts via the default point results in a baseline hazard rate component, as the Cox proportion hazard rate model. This baseline component can be estimated outside the model in the absence of model covariates, leading to a significant reduction in the number of parameters to be otherwise estimated inside the model. A practical benefit for this proposed model is that, at times when model recalibration is imminent, due to, for example, the buildup of the latest portfolio data, one can simply recalibrate for the default points using the updated long-run forward PDs, assuming that the responsiveness for a risk rating with respect to the macroeconomic variables remains the same.

Two interesting future researches are the applications of reinforcement learning in optimal investment strategies and the discriminative restricted Boltzmann machine for detecting the default risk for a credit card portfolio, where a large number of risk covariates are generally involved.

Acknowledgment

The author thanks Carlos Lopez for his consistent supports in this research. Special thanks to Clovis Sukam for his critical reading of this manuscript, and to Glenn Fei, Wallace Law, Kaijie Cui, Biao Wu, Zunwei Du, Lan Gong, Wilson Kan, and Amada Huang for many valuable conversations.

The views expressed in this article are not necessarily those of Royal Bank of Canada or any of its affiliates. Please direct any comments to the author at: h_y02@yahoo.ca.

References

[1] Barlow, R. E.; Bartholomew, D. J.; Bremner, J. M. and Brunk, H. D. (1972). Statistical Inference Under Order Restrictions; The Theory and Application of Isotonic Regression. New York: Wiley. ISBN 0-471-04970-0.

[2] Friedman, J. and Tibshirani, R. (1984). The Monotone Smoothing of Scatterplots. *Technometrics*, *Vol. 26 (3),* pp. 243-250. DOI: 10.2307/1267550

[3] Potharst, R. and Feelders, A. J. (2002). Classification Trees for Problems with Monotonicity Constraints, *SIGKDD Explorations*, *Vol. 14 (1)*, 1-10, 2002

[4] Kotlowski, W. and Slowinski, R. (2009). Rule Learning with Monotonicity Constraints, Proceedings of the 26th Annual International Conference on Machine Learning,

pp. 537-544, 2009
[5] Eichenberg, T. (2018). Supervised Weight of Evidence Binning of Numeric Variables and Factors, R-Package Woebinning.
[6] You, S.; Ding, D.; Canini, K.; Pfeifer, J. and Gupta, M. (2017). Deep Lattice Networks and Partial Monotonic Functions, 31st Conference on Neural Information Processing System (NIPS), 2017
[7] Allison, P. D. (1992). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology, Vol. 13*, pp. 61-98.
[8] Yang, B. H. (2017). Point-in-Time PD-Term Structure Models for Multi-Period Scenario Loss Projections. *Journal of Risk Model Validation, Vol. 11 (1)*, pp. 73-94.
[9] Cox, D. R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B, 20*, pp.187–220.
[10] SAS Institute Inc. (2015). *SAS/STAT(R) 14.1 User's Guide.*
[11] Breslow, N. E. (1972). Discussion of Professor Cox's Paper. *J. Royal Stat. Soc. B, 34*, pp.216–217.
[12] Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance, Vol. 29 (2)*, pp. 449-470.
[13] Vasicek, O. (2002). Loan Portfolio Value. *RISK, December 2002*, pp. 160 - 162.
[14] Rosen, D. S. (2009). Analytical Methods for Hedging Systematic Credit Risk with Linear Factor Portfolios. *Journal of Economic Dynamics & Control, 33*, pp. 37-52.
[15] Yang, B. H. (2018). Smoothing Algorithms by Constrained Maximum Likelihood. *Journal of Risk Model Validation, Vol. 12 (2)*, pp. 89-102.
[16] Wolfinger, R. (2008). Fitting Nonlinear Mixed Models with the New NLMIXED Procedure. *SAS Institute Inc.*