Weibull Racing Time-to-event Modeling and Analysis of Online Borrowers' Loan Payoff and Default

Quan Zhang[*], Qiang Gao[†], Mingfeng Lin[‡] and Mingyuan Zhou[§]

November 1, 2019

Abstract

We propose Weibull delegate racing (WDR) to explicitly model surviving under competing events and to interpret how the covariates accelerate or decelerate the event time. It explains non-monotonic covariate effects by racing a potentially infinite number of sub-events, and consequently relaxes the ubiquitous proportional-hazards assumption which may be too restrictive. For inference, we develop a Gibbs-sampler-based MCMC algorithm along with maximum a posteriori estimations for big data applications. We analyze time to loan payoff and default on Prosper.com, demonstrating not only a distinguished performance of WDR, but also the value of standard and soft information.

Keywords: interpretable nonlinearity, FinTech, online P2P lending, soft information

[*]quan.zhang@mccombs.utexas.edu, McCombs School of Business, the University of Texas at Austin.

[†]qiang.gao@baruch.cuny.edu, Baruch college, the City University of New York.

[‡]mingfeng.lin@scheller.gatech.edu, Scheller College of Business, the Georgia Institute of Technology.

[§]mingyuan.zhou@mccombs.utexas.edu, McCombs School of Business, the University of Texas at Austin.

1 Introduction

The risk-return tradeoff is always an important concern to investors in financial markets. The tradeoff takes into account not only borrowers' creditworthiness that determines the risk of loan default, but also the time to default since a prolonged repayment of principal and interest can reduce, compensate for or even exceed the loss from default [Banasik et al., 1999]. Meanwhile, time to early payoff has analogous influence on return of investment (ROI). For example, if a loan is fully repaid in a lump sum at an early stage, the ROI will be remarkably lower than expected. One may argue that the annualized rate of return will be unchanged if the loan is paid off early and the money repaid continues to be invested. Actually, this is not true in practice due to the existence of transaction costs, such as the transaction fees and the time that investors spend in searching equally profitable bonds. Time to loan payoff and time to default hereby play a crucial role in classifying a loan into an appropriate level of risk and reward and should be an essential consideration in investment decisions.

Ideally, investors maximize returns by investing in loans where borrowers have a long time to complete repayment or to default. However, it can be hard to screen such borrowers by qualitatively examining a small amount of standard information because a borrower's characteristic that can reduce the chance of default or postpone the default time often contributes to early payoff. For example, a borrower with a high credit grade and high income is less likely to default but more likely to pay off the loan early to avoid the interest. Consequently, the effect of credit grades and income levels on ROI may be ambiguous without quantitative analysis. Such ambiguity gives prominence to the joint analysis of time to both endpoints. Survival analysis, or time-to-event modeling, provides a powerful tool to analyze how borrower's characteristics influence the time to payoff and time to default and thus ROI.

In survival analysis that explores the relationship between features and time to events, existing methods often parameterize the hazard function with a weighted linear combina-

tion of the features. One of the most popular methods is the proportional hazards models, like Cox model [Cox, 1992] which is semi-parametric in that it assumes a non-parametric baseline hazard multiplied by a covariate-dependent coefficient. The proportional hazards models are often applied to population-level studies that try to unveil the relationship between the risk factors and hazard functions, such as to what degree a unit increase in a covariate is multiplicative to the hazard rate. However, it may have several drawbacks when we care about interpreting the feature effects as well as predicting the event time. First, the feature effect is interpreted as the log of a multiplier of the nonparametric baseline hazard which is estimated separately, but the effect on the event time is often not evident and the time prediction requires further modeling. Second, the interpretability is obtained by sacrificing model flexibility, because the proportional-hazards assumption is violated when the covariate effects are non-monotonic. For example, both very high and very low ambient temperatures were related to high mortality rates in Valencia, Spain, 1991-1993 [Ballester et al., 1997], and a significantly increased mortality rate is associated with both underweight and obesity [Flegal et al., 2007]. Another source of nonlinearity is from interaction effects. For example, an increased interest rate may accelerate the time to default for borrowers with low income and a low credit grade, but such acceleration may be less remarkable or insignificant for borrowers of very high credit-worthiness. Data transformations like adding square and interaction terms can alleviate this problem but often requires expert opinions. Moreover, complex data transformations like using kernels or neural networks would undermine the model interpretability. Third, the proportional hazards models rank all subjects by their event time and are incapable of big data analysis as the rankings of subjects always change when a random mini-batch sample is drawn so that an unbiased gradient is hard to find and thus a stochastic-gradient-based optimization is not applicable.

It is very common in survival analysis to consider competing events (also known as competing risks) which are mutually exclusive, i.e., the occurrence of one event precludes

other events. Apart from modeling the time, in the presence of competing events, it is also important to model the event type, or which one of the events is likely to occur first. Two of the most classical models are the Cox model with competing risks [Wolbers et al., 2014, Austin et al., 2016] and the Fine-Gray (FG) subdistribution model [Fine and Gray, 1999] which model cause-specific and subdistribution hazard functions [Putter et al., 2007, Lau et al., 2009], respectively, with the proportional-hazards assumption. The former is often applied to studying the etiology of diseases, while the latter is favorable when developing prediction models and risk-censoring systems [Austin et al., 2016].

Loan payoff and default are two competing events and the time to both events needs to be jointly modeled. Compared to early payoff, more research has studied the default probability and default time alone. Divino et al. [2013] use the Cox proportional hazards model and time-varying covariates to study how the loan-specific and economy's basic interest rates affect the default probability in Brazil. Durovic [2017] stratifies data by loan terms and purposes, and use Kaplan-Meier estimate of survival function within the vintage framework to study the default probability of online loans on LendingClub.com. Miller [2012] explores risk factors of online loan default on Prosper.com using quantile regression of default time. Regarding loan default and early payoff as competing events, Stepanova and Thomas [2002] study a personal loan data set from a UK financial institution. They propose a way of coarse-classifying of features by survival analysis of early payoff and default using two separate Cox models. Concretely, when analyzing the time to default they treat the borrowers who have repaid the loan as censored observations, and analogously when analyzing the time to payoff they treat as censored the borrowers who have defaulted. Baesens et al. [2005] also analyze time to loan early payoff and default, compare Cox models and neural networks and have found their performances are comparable. Apart from Cox models' lack of modeling flexibility and neural networks' lack of interpretability, they also analyze default (payoff) separately by treating borrowers who have paid off the loan (default) as censored.

Although we can enable every survival model that can handle censoring to model competing events if we separately analyze one competing event at a time and regard observations having other events as censored, this trick of censoring is problematic because it violates the assumption of independent or noninformative censoring [Kalbfleisch and Prentice, 2011]. This assumption requires subjects who are exposed to the risk to have the same future risk for the occurrence of the event as the censored ones do, as if censoring is random and thus noninformative. The assumption often holds in the single-event analysis where censoring is due to random missing or lost to follow-up, but it is unreasonable to assume that borrowers who have paid off the loans can represent those who may default, or the other way around. As a result, violating this assumption may lead to over-estimation of the cumulative incidence [Austin et al., 2016, Tong et al., 2012]. A possible remedy is to adjust the cumulative incidence through the Kaplan-Meier estimate [Tong et al., 2012], which, however, may be inaccurate because the Kaplan-Meier estimate is for the whole population without considering individual feature effects. A better solution is to turn to models that are specifically designed for joint analysis of multiple competing events, such as the FG subdistribution hazard model. It is good at predicting individuals' risks [Lau et al., 2009], but its linear proportional-hazards assumption may narrow its scope of application.

The existence of non-monotonic covariate effects can easily challenge and break the proportional-hazards assumption as in the FG model. This barrier has been surmounted by nonparametric approaches, such as random survival forests [Ishwaran et al., 2014], Gaussian processes [Barrett and Coolen, 2013] and neural networks that discretize the survival time [Lee et al., 2018]. These approaches are specifically designed for the existence of competing events and used for studies at an individual level, such as predicting the survival time, but almost unable to tell which or how covariates influence the event-specific hazard or survival time. So these models may be not of interest when we aim to explain the feature effects on loan default and early payoff.

To this end, we need a model for survival analysis of competing events that achieves a good balance of model interpretability and flexibility and obeys the noninformative censoring assumption as well. So we construct Weibull delegate racing (WDR) survival model, a gamma-process-based nonparametric Bayesian hierarchical model for competing events. WDR utilizes the race among Weibull random variables to jointly model all the competing events, enables the nonlinear modeling capability which is determined data-adaptively, and has an interpretation as a race among latent sub-events. In addition, WDR is an accelerated failure time model where the features accelerate or decelerate the progression of time to each respective competing event, so that it is more appealing compared to proportional hazards models when feature effects on payoff time or default time are of interest.

We test our WDR model using data from Prosper.com, one of the largest debt-based crowdfunding or peer-to-peer (P2P) lending platforms in the U.S. We choose P2P as our testing context for the following two main reasons. First, we have rich and objective information on the borrowers, the loans and the loan performance during our study period. We know a large amount of borrower information disclosed from their personal and credit files. Such quantitative information is the basis for loan underwriting in offline debt finance. We also know when each loan was originated and would mature, when and how much each installment was paid, whether each loan was prepaid or not, whether each loan was fully repaid or defaulted, and, if defaulted, how much the principal was lost. In addition to the rich information, prepayment and default are two common phenomena in online P2P market [Duarte et al., 2012]. All aforementioned information is seldom available from traditional offline debt financial institutions. Online P2P lending provides a good opportunity to test the predictive power of our proposed model.

The proposed WDR outperforms other popular predictive models widely used by academics and industries in both time to event modeling and classification of loan payoff and default. Specifically, WDR's performance is better than the Fine-Gray model and the

random survival forests when predicting the cumulative incidence functions, and better than these two models and a neural network in terms of both classification accuracy and area under the ROC curve (AUC). Moreover, we find the inclusion of soft information (i.e., information cannot be easily quantified by credit history or financial and employment status but available on Prosper) leads to a remarkable improvement in prediction. We also demonstrate the wisdom of the crowd in distinguishing loans with a high risk of default, but not as much in avoiding loans tending to be paid off early.

The paper proceeds as follows. In Section 2 we propose WDR and show its properties. Section 3 shows how our Bayesian inference deals with missing event time or type and truncated event time including censoring. In Section 4 we use synthetic data to show WDR's nonlinear modeling capacity and its outstanding performance. In Section 5 we analyze the Prosper data to show the feature effects on loan default and early payoff and the value of soft information that may be overlooked by traditional financial institutions. Section 6 concludes. We defer all proofs and algorithms, and a part of experimental settings to the appendix.

2 Weibull racing survival model for competing events

We first introduce a property of Weibull racing describing the distribution of the minimum of Weibull-distributed random variables and show how this property can be directly used for time-to-event modeling of competing events. Then we propose Weibull racing model assuming monotonically accelerating or decelerating covariate effects on the time to competing events, and Weibull delegate racing model where covariates can have monotonic or arbitrarily non-monotonic effects on the event time and the survival and hazard functions.

2.1 Property of the minimum of Weibull random variables

Let $t \sim \text{Weibull}(a, \lambda)$ represent duration t following a Weibull distribution with the probability density function $f(t \mid a, \lambda) = a\lambda t^{a-1} e^{-\lambda t^a},\ t \in \mathbb{R}_+$ and the cumulative distribution function $F(t \mid a, \lambda) = 1 - e^{-\lambda t^a},\ t \in \mathbb{R}_+$ where $\mathbb{R}_+$ represents the nonnegative side of the real line, $a > 0$ is the shape parameter and $\lambda > 0$ such that $\mathbb{E}[t] = \lambda^{-1/a}\Gamma(1 + 1/a)$. Shown below is a property that characterizes a race among independent Weibull random variables.

Property 1 (Weibull racing)**.** *If $t_j \sim \text{Weibull}(a, \lambda_j)$, where $j = 1, \ldots, J$, are independent to each other, then $t = \min\{t_1, \ldots, t_J\}$ and the argument of the minimum $y = \underset{j \in \{1,\ldots,J\}}{\text{argmin}}\ t_j$ are independent and they satisfy*

$$t \sim\ \text{Weibull}\Big(a, \sum\nolimits_{j=1}^{J} \lambda_j\Big),\ y \sim \text{Categorical}\Big(\lambda_1 \Big/ \sum\nolimits_{j=1}^{J} \lambda_j, \cdots, \lambda_J \Big/ \sum\nolimits_{j=1}^{J} \lambda_j\Big). \quad (1)$$

Intuitively, Suppose there is a race among team $j = 1, \cdots, J$, whose completion time t_j follows $\text{Weibull}(a, \lambda_j)$, with the winner being the team with the minimum completion time. Property 1 shows the winner's completion time t still follows a Weibull distribution and is independent of which team wins the race. In the context of time-to-event modeling, we can regard a competing event as a team and the latent time to this event as the completion time of the team. Since the occurrence of one competing event precludes others, y indicating which is the first event to happen will be the observed event type, and t will be the observed event time. Weibull racing not only describes a natural mechanism of surviving under competing events but also provides an attractive modeling framework amenable to Bayesian inference; conditioning on a and λ_j's, the joint distribution of the event type y and time to event t becomes fully factorized as

$$P(y, t \mid a, \{\lambda_j\}_{1,J}) = \lambda_y a t^{a-1} e^{-t^a \sum_{j=1}^{J} \lambda_j} \quad (2)$$

which allows Bayesian inference by MCMC.

2.2 Weibull racing for linear covariate effects

We show how to model the covariate dependence of event time in the Weibull racing framework by introducing gamma-mixed Weibull distribution. Specifically, let $\lambda \sim \text{Gamma}(r, 1/b)$ represent a gamma distribution with $\mathbb{E}[\lambda] = r/b$ and $\text{Var}[\lambda] = r/b^2$. If $t \sim \text{Weibull}(a, \lambda)$ where $\lambda \sim \text{Gamma}(r, 1/b)$, we have the marginal density of t given a, r and b as

$$f(t \mid a, r, b) = \int_0^\infty \text{Weibull}(t; a, \lambda)\text{Gamma}(\lambda; r, 1/b)d\lambda = \frac{arb^r t^{a-1}}{(b+t^a)^{r+1}}.$$

This gamma-mixed Weibull distribution has a decreasing density $f(t \mid a, r, b)$ (or equivalently, uni-modal at 0) if $a \leq 1$, and has a uni-mode at $(\frac{ab-b}{1+ar})^{\frac{1}{a}}$ if $a > 1$. Note that this property of uni-modality is analogous to that of the regular Weibull distribution. We define Weibull racing model as follows.

Definition 1. *With competing events $j \in \{1, \ldots, J\}$ and their associated latent event time t_j, Weibull racing model has the observed event time t and event type y depending on covariates $\boldsymbol{x}$ as*

$$t = t_y,\ y = \text{argmin}_{j \in \{1,\ldots,J\}} t_j,\ t_j \sim \text{Weibull}(a, \lambda_j),\ \lambda_j \sim \mathit{Gamma}(r_j, e^{\boldsymbol{x}\boldsymbol{\beta}_j}). \tag{3}$$

We apply this definition to the scenario of borrowers' loan payoff ($j = 1$) and default ($j = 2$) that are two competing events. Before the outcome of the loan is observed, the payoff time t_1 and default time t_2 are latent. If the payoff happens earlier than default, i.e., $t_1 < t_2$, we observe the loan fully repaid (i.e., $y = 1$) at time $t = t_1$. Otherwise, we observe the default happens (i.e., $y = 2$) at time $t = t_2$. An alternative view of Weibull racing is from the perspective of a discrete choice model [Hanemann, 1984, Greene, 2003,

Train, 2009]. The observed event type y is corresponding to the event whose latent arrival time is the minimum among all the competing events. Distinct from ordinary discrete choice models where the decision is made to maximize latent utility, the event type y is determined by Weibull racing to minimize the waiting time for any one of the competing events, and this minimum waiting time t is also of interest.

For the observed event time $t = \min_j t_j$, the survival funcation and the hazard function, respectively, are

$$S(t) = \prod_j (e^{\boldsymbol{x}'\boldsymbol{\beta}_j} t^a + 1)^{-r_j}, \quad h(t) = \frac{-dS(t)/dt}{S(t)} = \sum_j \frac{a r_j t^{a-1}}{t^a + e^{-\boldsymbol{x}'\boldsymbol{\beta}_j}}. \tag{4}$$

For each competing event j, we can express its event time as $t_j \sim \text{Weibull}(a, e^{\boldsymbol{x}'\boldsymbol{\beta}_j}\lambda_{j0})$ where $\lambda_{j0} \sim \text{Gamma}(r_j, 1)$ so that $\log t_j = -\boldsymbol{x}'\boldsymbol{\beta}_j/a + \log t_{j0}$ where $t_{j0} \sim \text{Weibull}(a, \lambda_{j0})$. Thus Weibull racing is an accelerated failure time model [Kalbfleisch and Prentice, 2011] for each competing event j in that the features $\boldsymbol{x}$ accelerate or decelerate the baseline event time t_{j0} by $\boldsymbol{\beta}/a$. Furthermore, with $S_{j0}(t) = (t^a + 1)^{-r_j}$ and $h_{j0}(t) = \frac{a r_j t^{a-1}}{1+t^a}$, we can write the survival function for event j as $S_j(t_j) = (e^{\boldsymbol{x}'\boldsymbol{\beta}_j} t^a + 1)^{-r_j} = S_{j0}(t_j e^{\boldsymbol{x}'\boldsymbol{\beta}_j/a})$ and write the corresponding hazard function as $h_j(t_j) = \frac{a r_j t_j^{a-1}}{t_j^a + e^{-\boldsymbol{x}'\boldsymbol{\beta}_j}} = e^{\boldsymbol{x}'\boldsymbol{\beta}_j/a} h_{j0}(t_j e^{\boldsymbol{x}'\boldsymbol{\beta}_j/a})$.

Weibull racing can be restricted in that the survival function and the hazard function for event j are monotonic in $\boldsymbol{x}$. Consequently, Weibull racing can only model competing events whose respective event time is linearly dependent on features. An overall nonlinear feature effect may result from multiple latent mechanisms of one competing event which is often pre-defined without a fine-grained categorization. For example, in medical research where survival analysis is widely applied, the nosology of competing events is often subject to human knowledge, diagnostic techniques, and patient population. Diseases with the same phenotype, categorized into one competing event, might have distinct etiology and different impacts on survival, and thus require different therapies. An example of an event with subordinate categories is diabetes which can be divided into Type 1 and Type

2. Type 1 is ascribed to insufficient production of insulin from the pancreas whereas Type 2 arises from the cells' failure in responding properly to insulin Varma et al. [2014].

In areas of social sciences, fine-grained categorizations of a competing event can be hard to pre-define, and consequently, the feature effects can be nonlinear since the competing event may be an aggregation of many sub-events. In this regard, it is often necessary for a model to identify these sub-events, not only to improve the fit of event time, but also to explore the underlying mechanisms. We develop Weibull delegate racing, assuming that an event consists of several sub-events under each of which the latent sub-event time is log-linearly accelerated by covariates.

2.3 Weibull delegate racing for nonlinear covariate effects

Based on the idea of Weibull racing that an individual's observed event time is the minimum of latent time to competing events, we further propose *Weibull delegate racing* (WDR), assuming that the time to a competing event is the minimum of the latent time to a number of sub-events appertaining to this competing event. In particular, let us first denote $G_j \sim \Gamma\mathrm{P}(G_{0j}, 1/c_{0j})$ as a gamma process defined on the product space $\mathbb{R}_+ \times \Omega$, where G_{0j} is a finite and continuous base measure over a complete separable metric space Ω, and $1/c_{0j}$ is a positive scale parameter such that $G_j(A) \sim \mathrm{Gamma}(G_{0j}(A), 1/c_{0j})$ for each Borel set $A \subset \Omega$. A draw from the gamma process consists of countably infinite non-negatively weighted atoms, expressed as $G_j = \sum_{k=1}^{\infty} r_{jk}\delta_{\boldsymbol{\beta}_{jk}}$. Now we formally define WDR model as follows.

Definition 2 (Weibull delegate racing)**.** *With competing events $j \in \{1, \ldots, J\}$ and their associated latent event time t_j, and given a random draw of a gamma process $G_j \sim \Gamma\mathrm{P}(G_{0j}, 1/c_{0j})$, expressed as $G_j = \sum_{k=1}^{\infty} r_{jk}\delta_{\boldsymbol{\beta}_{jk}}$, for each $j \in \{1, \ldots, J\}$, Weibull delegate*

racing models the observed event time t and event type y given covariates $\boldsymbol{x}$ as

$$t = t_y,\ y = \underset{j\in\{1,\ldots,J\}}{\operatorname{argmin}}\ t_j,\ t_j = t_{j\kappa_j},\ \kappa_j = \underset{k\in\{1,\ldots,\infty\}}{\operatorname{argmin}}\ t_{jk},$$
$$t_{jk} \sim \text{Weibull}(a, \lambda_{jk}),\ \lambda_{jk} \sim \text{Gamma}(r_{jk}, e^{\boldsymbol{x}'\boldsymbol{\beta}_{jk}}). \tag{5}$$

Here we assume an infinite number of sub-events under a competing event j, and each sub-event k has a latent event time t_{jk}. WDR can be considered as a two-phase race. In the first phase, for a pre-specified competing event j, there is a race among its countably infinite sub-events $\{k \,|\, k = 1, \ldots, \infty\}$, and the winner, say sub-event κ_j, whose time $t_{j\kappa_j} = \min_k t_{jk}$, represents the event j by letting t_j be equal to $t_{j\kappa_j}$. In the second phase, J events with the associated event time $\{t_j\}_j$ compete with each other to eventually determine both the observed event time t and the observed event type y.

Although WDR assumes a potentially infinite number of sub-events for each competing event, the summation of the weights of these sub-events, $G_j = \sum_{k=1}^{\infty} r_{jk}\delta_{\boldsymbol{\beta}_{jk}}$, is finite according to the gamma process. Therefore, the gamma process not only admits a race among a potentially infinite number of sub-events, but also parsimoniously shrinks toward zero the weights of negligible ones [Zhou et al., 2016, Zhou, 2016], so that the non-monotonic covariate effects on the time to a competing event can be interpreted as the *minimum*, which is a nonlinear operation, of time to sub-events whose accelerating factor is log-linear in covariates $\boldsymbol{x}$.

Intuitively, the nonlinear modeling capacity of WDR is fulfilled by taking the minimum of the minimums which is a two-step nonlinear operation. In mathematics, the event time t_j, the survival function $S_j(t_j)$ and the hazard function $h_j(t_j)$ for competing event j are no longer monotonic in $\boldsymbol{x}$. Specifically, the following corollary explicitly shows an equivalent definition of WDR and its nonlinear modeling capacity.

Corollary 1. *Weibull delegate racing can also be expressed as*

$$t = t_y,\ y = \text{argmin}_j\, t_j, t_j \sim \text{Weibull}\left(a, \sum_{k=1}^{\infty} e^{\boldsymbol{x}'\boldsymbol{\beta}_{jk}} \tilde{\lambda}_{jk}\right),\ \tilde{\lambda}_{jk} \sim \text{Gamma}(r_{jk}, 1). \quad (6)$$

The survival function and the hazard function, respectively, for event j are

$$S_j(t_j) = \prod_k (1 + t_j^a e^{\boldsymbol{x}'\boldsymbol{\beta}_{jk}})^{-r_{jk}}, \quad h_j(t_j) = \frac{-dS_j(t_j)/dt_j}{S_j(t_j)} = \sum_k \frac{a r_{jk} t_j^{a-1}}{t_j^a + e^{-\boldsymbol{x}'\boldsymbol{\beta}_{jk}}}. \quad (7)$$

Corollary 1 represents WDR as a generalization of Weibull racing, where the scale of the time to a competing event j is proportional to a weighted summation of a countably infinite number of gamma random variables with covariate-dependent weights. Moreover, the survival and hazard functions are non-monotonic in the features; they can be adaptively flexible as the weights r_{jk}'s are learned from data. To study how the hazard function changes over time, we note that

$$\frac{dh_j(t_j)}{dt_j} = \sum_k \frac{a(a-1) r_{jk} e^{-\boldsymbol{x}'\boldsymbol{\beta}_{jk}} t^{a-2} - a r_{jk} t^{2a-2}}{(t^a + e^{-\boldsymbol{x}'\boldsymbol{\beta}_{jk}})^2}.$$

So if $a \leq 1$, $h_j(t_j)$ is always decreasing in t_j, and otherwise $h_j(t_j)$ can be increasing (until some time that is long enough) or arbitrarily non-monotonic in t_j with proper values of r_{jk}'s and $\boldsymbol{\beta}_{jk}$'s.

3 Bayesian inference

In this section, we propose an efficient MCMC inference for WDR. The fully factorized likelihood admits Gibbs sampling updates of all the parameters except the Weibull shape a whose full conditional distribution is unimodal so that an efficient slice sampling scheme [Damlen et al., 1999, Neal et al., 2003] can be used. Moreover, we show that our proposed MCMC inference is still applicable even if the data has missing event types

and truncated or missing event time. In section 3.1 we use notations of Weibull racing for brevity to illustrate how the likelihood can be written in a fully factorized form with data augmentation tricks, and the tricks apply to and the corresponding conclusions hold for WDR. Section 3.2 gives the full hierarchical representation of WDR that facilitates MCMC inference and parameter shrinkage to avoid overfit. Additionally, we propose maximum a posteriori (MAP) estimations using stochastic gradient descent for big data analysis. Note that we use reparameterization for feasible gradient calculation and use self normalization for variance reduction. The details of the MAP estimations are provided in the appendix.

3.1 Fully factorized likelihood

In survival analysis, it is rarely the case that both y and t are observed for all observations, and one often need to deal with missing data or right or left censoring. We write $t \sim \text{Weibull}_\Psi(a, \lambda)$ as a truncated Weibull random variable defined by the density function $f_\Psi(t \mid a, \lambda) = a\lambda t^{a-1} e^{-\lambda t^a} / \int_\Psi a\lambda u^{a-1} e^{-\lambda u^a} du$, where $t \in \Psi$ and Ψ is an open interval on $\mathbb{R}_+$ representing censoring. Specifically, Ψ can be $(T_{r.c.}, \infty)$ indicating right censoring with censoring time $T_{r.c.}$, $(0, T_{l.c.})$ indicating left censoring with censoring time $T_{l.c.}$, or a more general case (T_1, T_2), $T_2 > T_1$. If we don't observe y or t, or there exists censoring, we have the following two scenarios in both of which it is necessary to introduce appropriate auxiliary variables to achieve fully factorized likelihoods: 1) If we only observe y (or t), then we can draw t (or y) shown in (1) as an auxiliary variable, leading to the fully factorized likelihood as in (2); 2) If we don't observe t but know $t \in \Psi$ with $P(t \in \Psi \mid a, \{\lambda_j\}_j) = \int_\Psi a(\sum_j \lambda_j) u^{a-1} e^{-u^a \sum_j \lambda_j} du$, then we draw $t \sim \text{Weibull}_\Psi(a, \sum_j \lambda_j)$, resulting in the likelihood

$$P\left(t, t \in \Psi \mid a, \sum\nolimits_j \lambda_j\right) = f_\Psi\left(t \mid a, \sum\nolimits_j \lambda_j\right) P\left(t \in \Psi \mid a, \sum\nolimits_j \lambda_j\right) = a\left(\sum\nolimits_j \lambda_j\right) t^{a-1} e^{-t^a \sum_j \lambda_j}. \tag{8}$$

Together with y which can be drawn by (1) if missing, the likelihood $P(y, t, t \in \Psi \,|\, \{\lambda_j\}_{1,J})$ becomes the same as in (2). The procedure of sampling t and/or y which generates fully factorized likelihoods under different censoring conditions plays a crucial role as a data augmentation scheme that will be used for the MCMC inference of the proposed Weibull (delegate) racing model. While in the case of right censoring y is unknown and the likelihood $P(t > T \,|\, a, \{\lambda_j\}) = e^{-T^a \sum_j \lambda_j}$ is already fully factorized, we still need to augment t and y, and use (2) as the likelihood. For the convenience of implementation, as in Zhou and Carin [2015], we truncate the total number of atoms of a gamma process to be K that is large enough by choosing a finite and discrete base measure as $G_{0j} = \sum_{k=1}^{K} \frac{\gamma_{0j}}{K} \delta_{\boldsymbol{\beta}_{jk}}$.

3.2 Hierarchical model of WDR and MCMC inference

With $\boldsymbol{x}_i$ denoting the covariates, y_i event type, and t_i the time to event of observation i, we express the full hierarchical form of WDR defined in (5), as

$$
\begin{aligned}
&t_i = t_{iy_i},\ y_i = \underset{j\in\{1,\ldots,J\}}{\operatorname{argmin}}\ t_{ij},\ t_{ij} = t_{ij\kappa_{ij}},\ \kappa_{ij} = \underset{k\in\{0,\ldots,K\}}{\operatorname{argmin}}\ t_{ijk},\\
&t_{ijk} \sim \text{Weibull}(a, \lambda_{ijk}),\ \lambda_{ijk} \sim \text{Gamma}(r_{jk}, e^{\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}}),\ k = 1, \cdots, K,\\
&\boldsymbol{\beta}_{jk} \sim \prod_{g=1}^{P} \mathcal{N}(0, \alpha_{gjk}^{-1}),\ \alpha_{gjk} \sim \text{Gamma}(a_0, 1/b_0),\ r_{jk} \sim \text{Gamma}(\gamma_{0j}/K, 1/c_{0j}),
\end{aligned}
$$

where $k = 1, \cdots, K$, $i = 1, \cdots, n$, and $j = 1, \cdots, J$. We further let $\gamma_{0j} \sim \text{Gamma}(e_0, 1/f_0)$, $c_{0j} \sim \text{Gamma}(e_1, 1/f_1)$, $r_0 \sim \text{Gamma}(e_0, 1/f_0)$, and set $e_0 = f_0 = e_1 = f_1 = 0.01$. The Gamma prior on the precision parameter α for each element of $\boldsymbol{\beta}$ is imposed to penalize large absolute values of $\boldsymbol{\beta}$ to avoid overfit [Tipping, 2001].

Since left censoring is uncommon and not shown in the Prosper data to be analyzed, we only consider right censoring in our inference and leave to readers other types of censoring which can be analogously done. All the parameters can be inferred by Gibbs

sampling except the Weibull shape parameter a. To the best of our knowledge, there does not exist a Gibbs sampling scheme for the Weibull shape parameter. Alternatively, we use an efficient slice sampling update for a without any tuning parameter as its full conditional distribution is unimodal. The proof of this unimodality and the complete MCMC algorithm are provided in the appendix.

4 Model validation by synthetic data

In this section, we validate the proposed WDR model using synthetic data by comparing WDR with some benchmark approaches whose implementations and experiment settings are deferred to the appendix for brevity. Specifically, we compare the proposed WDR model, Fine-Gray proportional subdistribution hazards model (FG) [Fine and Gray, 1999] and random survival forests (RF) [Ishwaran et al., 2014] that are all designed for survival modeling of competing events. We show that WDR not only can tell whether the covariate effects are monotonic or not by inferring the number of remarkable sub-events but also performs uniformly well in both situations.

4.1 Quantification measurements

We quantify the survival model performance by the cause-specific concordance index [Wolbers et al., 2014] and the Brier score [Gerds et al., 2008, Steyerberg et al., 2010]. Specifically, the Brier score (BS) for risk j at time t is calculated as

$$\mathrm{BS}_j(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}(t_i \le t, y_i = j) - P(t_i \le t, y_i = j)\right]^2,$$

with a smaller value indicating a better model fit. Cause-specific concordance index (C-index) of event j at time t is computed as

$$\mathcal{C}_j(t) = P\left(Score_j(\boldsymbol{x}_i, t) > Score_j(\boldsymbol{x}_{i'}, t) \,|\, y_i = j \text{ and } [t_i < t_{i'} \text{ or } y_{i'} \neq j]\right),$$

where $i \neq i'$ and $Score_j(\boldsymbol{x}_i, t)$ is a prognostic score at time t depending on $\boldsymbol{x}_i$ such that its higher value reflects a higher risk of event j. Wolbers et al. [2014] write C-index as a weighted average of time-dependent AUC that is related to sensitivity, specificity, and ROC curves for competing events [Saha and Heagerty, 2010]. So a higher value of C-index indicates a better model fit and a value around 0.5 implies a model failure.

A good choice of the prognostic score is the cumulative incidence function (CIF), *i.e*, $Score_j(\boldsymbol{x}_i, t) = \text{CIF}_j(i, t) = P(t_i \leq t, y_i = j)$ [Fine and Gray, 1999, Kalbfleisch and Prentice, 2011, Crowder, 2001]. Distinct from a survival function that measures the probability of surviving beyond some time, CIF estimates the probability that an event occurs by a specific time in the presence of competing events. For WDR given a, $\{r_{jk}\}$ and $\{\boldsymbol{\beta}_{jk}\}$,

$$\text{CIF}_j(i, t) = P(t_i \leq t, y_i = j) = \mathbb{E}\left[\frac{\sum_k \lambda_{ijk}}{\sum_{j',k} \lambda_{ij'k}} \left(1 - e^{-t^a \sum_{j',k} \lambda_{ij'k}}\right) \,|\, a, \{r_{jk}\}, \{\boldsymbol{\beta}_{jk}\}\right],$$

where the expectation is taken over $\{\lambda_{ijk}\}_{j,k}$ and $\lambda_{ijk} \sim \text{Gamma}(r_{jk}, e^{\boldsymbol{x}_i' \boldsymbol{\beta}_{jk}})$. The expectation can be evaluated by Monte-Carlo estimation if we have point estimates or a collection of post-burn-in MCMC samples of r_{jk} and $\boldsymbol{\beta}_{jk}$.

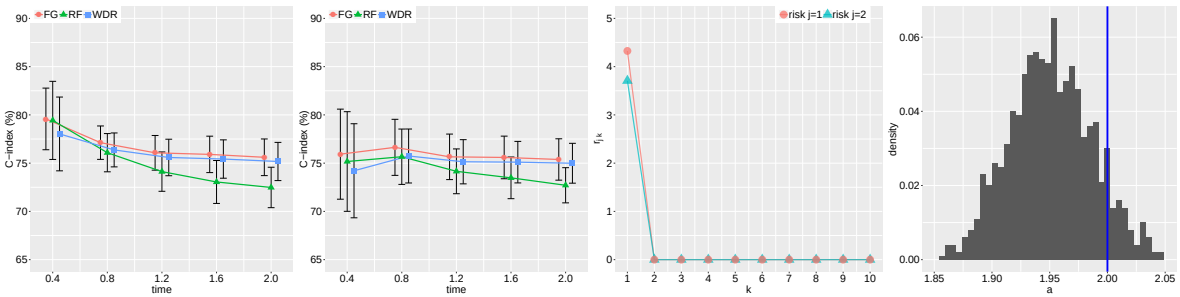
4.2 Synthetic data analysis

We simulate two data sets according to the data generating process in Table 1 where $\boldsymbol{x}_i \in \mathbb{R}^3$ and each coordinate is sampled from Uniform$(0, 1)$, and use them to validate WDR and to illustrate its nonlinear modeling capability. In Table 1, t_{ij} denotes the latent time to event (risk) j, $j = 1, 2$ and t_i is the observed time to event of observation i. The

Table 1: Synthetic data generating process.

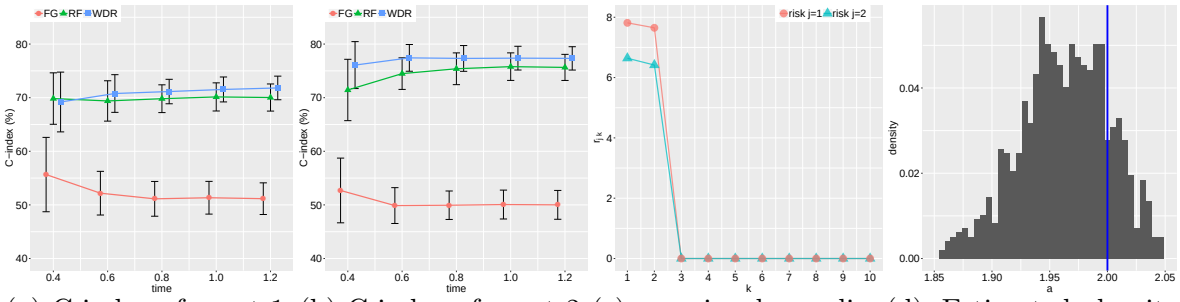
Synthetic data 1	Synthetic data 2
$t_{i1} \sim \text{Weibull}(a = 2, e^{\boldsymbol{x}_i'\boldsymbol{\beta}_1})$	$t_{i1} \sim \text{Weibull}(a = 2, \cosh(\boldsymbol{x}_i'\boldsymbol{\beta}_1))$
$t_{i2} \sim \text{Weibull}(a = 2, e^{\boldsymbol{x}_i'\boldsymbol{\beta}_2})$	$t_{i2} \sim \text{Weibull}(a = 2, \lvert\sinh(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\rvert)$
$t_i = \min(t_{i1}, t_{i2}, 2.1)$	$t_i = \min(t_{i1}, t_{i2}, 1.3)$

observed event type $y_i = \arg\min_j t_{ij}$ if $t_i < T_{r.c.}$, and $y_i = 0$ indicates right censoring if $t_i = T_{r.c.}$ where the censoring time $T_{r.c.} = 2.1$ for data 1 and 1.3 for data 2. We simulate 2,000 random observations, and use 1800 for training and the remaining 200 for testing withoug censored observations. We randomly take 20 training/testing partitions for each of which we evaluate the C-indices and Brier scores on the testing set at time $0.4, 0.8, 1.2, 1.6, 2.0$ for data 1 and at time $0.4, 0.6, 0.8, 1.0, 1.2$ for data 2. Displayed in panels (a) to (d), respectively, of Figure 1, are the sample mean $\pm$ standard deviation of the estimated C-indices of event $j = 1$ and 2 by WDR, the estimated r_{jk}'s of both events, and the empirical density estimation of the Weibull shape a by a histogram using the post-burn-in MCMC samples with the blue vertical line indicating the true value for data 1. Note that Panel (c) and (d) show results of one training/testing partition but without loss of generality. Analogous plots of results by WDR for data 2 are shown in Figure 2.



(a) C-index of event 1. (b) C-index of event 2. (c) r_{jk} in descending order by WDR. (d) Estimated density of a by WDR.

Figure 1: Cause-specific C-indices, shrinkage of r_{jk} and estimation of a by WDR for synthetic data 1.



(a) C-index of event 1. (b) C-index of event 2. (c) r_{jk} in descending order by WDR. (d) Estimated density of a by WDR.

Figure 2: Cause-specific C-indices, shrinkage of r_{jk} and estimation of a by WDR for synthetic data 2.

For data 1 where the time to each of the two events depends on the covariates monotonically, WDR has comparable performance with FG and slightly outperforms RF in terms of the mean values of C-indices when $t \geq 1.2$. The underperformance of RF in the case of monotonic covariate effects has also been observed in the original paper Ishwaran et al. [2014]. For data 2 where the event time and covariates are not monotonically related, WDR and RF at any time evaluated significantly outperform FG. Furthermore, FG fails on this dataset as its C-indices are around 0.5 for both events as in Figure 2 (a) and (b), respectively. Moreover, WDR slightly outperforms RF especially for event 2. Panels (c) of Figure 1 and 2 show r_{jk} inferred from data 1 and 2, respectively. Specifically, both events consist of only one sub-risk for data 1, indicating linear covariate effects. By contrast, the racing between two sub-events can approximate the complex data generating process of each of the two competing events in data 2. Panel (c) in each of the two figures also demonstrates WDR's shrinkage of unneeded modeling capacity. For both synthetic data sets, Panel (d) shows a good recovery of the Weibull shape a as well as its uncertainty by WDR. The comparable performance to RF and the shrinkage of r_{jk}'s justifies WDR's data-adaptively flexible modeling capacity. In addition, we provide in the appendix the Brier scores by FG, RF and WDR in Table 8 to 11; the Brier scores also show comparable performances of the three models on synthetic data 1 and that WDR and RF dominate FG on synthetic data 2.

5 Prosper data analysis

In this section, we analyze a real data set from Prosper.com to show how features of standard and soft information affect the time to loan payoff and default. First, we introduce the data set that contains loans that have been fully funded after lenders' bidding. Subsequently, we show the feature effect estimations by WDR and compare its model performance to the Fine-Gray model and the random survival forests. We also show the predictability of different groups of partial information and demonstrate the wisdom of the crowd in distinguishing loans with a high risk of default.

5.1 Context and data

Our data comes from Prosper.com[1] that is one of the largest P2P lending sites in the US with over sixteen billion dollars in funded loans belonging to more than 970,000 borrowers by 2019. The data set contains fully funded loan listings that are personal, fixed-rate and unsecured between November 9th, 2005 and December 19th, 2010[2] with the term of 36 months[3]. We describe a typical lending process, especially features directly related to our study. More detailed information can be found from studies of Zhang and Liu [2012] and Lin et al. [2013].

The process starts with the creation of a loan request (known as a "listing") that contains information about the borrower and their request. In the posted listings, the borrowers mainly specify the requested amount and the offered interest rates. The listing page also includes borrowers' standard credit information, such as debt-to-income ratio, delinquency history, and credit lines, and borrowers self-reported information such as income and employment status. The standard information is often called verified hard

[1] https://www.prosper.com/about

[2] Prosper.com changed its bidding mechanism on December 20th, 2010. Before this date, the platform implemented an auction mechanism. The borrowers specified the interest rates they were willing to accept. After this date, the platform decides the borrowers' interest rates that are calculated using borrowers' information such as credit scores, income levels, and debt-to-income ratios.

[3] Very few loans had 12 and 72 months maturity and we exclude these observations.

information which was extracted from the borrower's credit report provided by a credit agency. Lenders can view this information and place bids on listings. They can bid as little as $25 and need to specify the minimum interest rates at which they would fund the listing.

Successful listings that attract 100% of the requested funds become loans after the website's verification. Lenders with lower reservation rates win the bidding, and the contract interest rate is determined by the highest reservation rates among the successful bidders. Every month during the life cycle of the loan (usually 36 months), Prosper.com will automatically debit the borrower's bank account and repay the investors after deducting fees. As long as the borrower makes the monthly payment in time, the loan status is marked as "current." Otherwise, the status will incrementally change to "late," "1 month late," "2 months late," etc. If the borrower fails to make payment for more than four months, the loan will be marked as "defaulted." Defaulted loans will be transferred to third-party collection agencies designated by Prosper.com, and will negatively affect the borrower's personal credit score.

The data set includes features of standard information that are often used by traditional financial institutions, and features of non-standard or soft information like loan purposes. To maintain the interpretability of coefficients $\boldsymbol{\beta}$ of WDR, we keep the original scale of the continuous features, and use dummy variables for both binomial and multinomial categorical features such as credit grades, income levels, employment status and loan purposes. So all features have non-negative values. All the multinomial variables have a class of "not available" which is used as the baseline category and not included in the design matrix to avoid multicolinearity. The detailed feature description is given in Table 2. Features containing soft information is indicated by the bold font. Such features include `GroupMember`, `OrderOfListing`, `TotalListingOfBorrowers`, `FundingOption`, `BorrowerMaximumRate`, `Duration`, `Images` and the loan purpose which is represented by one-hot encoding for 20 purposes, such as debt consolidation, home im-

provement and so on (shown in the third column of Table 2). Other features containing standard information include borrowers' credit grades, credit and financial history, employment status, contract interest rate and so on. There are 32,904 fully funded listings in total and they are partitioned into a training set of 29,904 and a testing set of 3,000. All the quantifications including Brier scores, C-indices and AUCs report the results on the testing data. Following Stepanova and Thomas [2002], we treat the loans that are fully repaid on the 36th month as censored.

Table 2: Prosper data feature description (bold font for soft information)

Feature	Description	Feature	Description
DebtToIncomeRatio	Continuous	Income:50k-75k	Dummy
BankcardUtilization	Continuous	Income:75k-100k	Dummy
AmountDelinquent	In thousand dollars	Income:>100k	Dummy
DelinquenciesLast7Years	Number of delinquencies	Retired	Dummy
InquiriesLast6Months	Number of inquiries	Full-time	Dummy
PublicRecordsLast10Years	Number of public records	Self-employed	Dummy
PublicRecordsLast12Months	Number of public records	Part-time	Dummy
CurrentCreditLines	Number of current credit lines	NotEmployed	Dummy
OpenCreditLines	Number of open credit lines	OtherEmployment	Dummy
TotalCreditLines	Number of total credit lines	**Debt consolidation**	Dummy
RevolvingCreditBalance	in thousand dollars	**Home improvement**	Dummy
GroupMember	1 if a group member and 0 otherwise	**Business**	Dummy
IsBorrowerHomeowner	1 if a home owner and 0 otherwise	**Personal loan**	Dummy
OrderOfListing	Borrower's order of the current listing	**Student use**	Dummy
TotalListingOfBorrowers	Borrower's total number of listings	**Auto**	Dummy
AmountRequested	Listing amount in thousand dollars	**Other**	Dummy
FundingOption	0 if close when funded and 1 if open for duration	**Baby** & **adoption**	Dummy
BorrowerMaximumRate	Maximum interest rate the borrower offered	**Boat**	Dummy
ContractInterest	Contract interest rate	**Cosmetic**	Dummy
Duration	Number of days in which the listing is valid for	**Engagement ring**	Dummy
Images	1 if borrower has an image and 0 otherwise	**Green loans**	Dummy
CreditGrade: AA	Dummy	**Household expenses**	Dummy
CreditGrade: A	Dummy	**Large purchase**	Dummy
CreditGrade: B	Dummy	**Medical/dental**	Dummy
CreditGrade: C	Dummy	**Motorcycle**	Dummy
CreditGrade: D	Dummy	**RV**	Dummy
CreditGrade: E	Dummy	**Taxes**	Dummy
CreditGrade: HR	Dummy	**Vacation**	Dummy
Income:1-25k	Dummy	**Wedding**	Dummy
Income:25k-50k	Dummy		

5.2 Feature effect estimations

We found only one sub-event by WDR for both early payoff and default, respectively, which implies that the features log-linearly accelerate or decelerate the event time and no remarkable interaction effects on the time to payoff or the time to default. For the

remaining analysis of the Prosper data we omit the subscript k for brevity. Since $\log t_j = -\boldsymbol{x}'\boldsymbol{\beta}_j/a + \log t_{j0}$ where $t_{j0} \sim \text{Weibull}(a, \lambda_{j0})$ and $\lambda_{j0} \sim \text{Gamma}(r_j, 1)$ as discussed in Section 2.2, $e^{\boldsymbol{\beta}_j/a}$ is interpreted as how $\boldsymbol{x}$ accelerate the time to event j. Concretely, a positive coefficient indicates the corresponding covariate accelerating the event time, whereas negative decelerating the event time. Quantitatively, with one unit increase of x_v which is the vth coordinate of $\boldsymbol{x}$, t_j will be reduced by $(100 \times \beta_{vj}/a)\%$, where β_{vj} is the coefficient of x_v for event j.

We show the estimated coefficient effects, $(\boldsymbol{\beta}_j/a)$, $j = 1, 2$ with standard errors by WDR in Figure 3 below and in Table 14 in the appendix as well. Features whose coefficient is negative for both events (or negative for one event and insignificantly different from 0 for the other event) would decelerate the time to early payoff and (or) default, so that a loan listing with big values of such features will have a prolonged time of repaying the principal and interest until payoff or default and thus leads to a high ROI. Meanwhile, investors should avoid listings with features whose values are big and coefficients are positive for both events (or whose coefficients are positive for one event and insignificantly different from 0 for the other event), because they expedite payoff and (or) default and thus results in a low ROI. Features whose coefficients are of different signs for the two events have opposite effects on the time to payoff and to default and their effects on ROI need to be analyzed quantitatively.

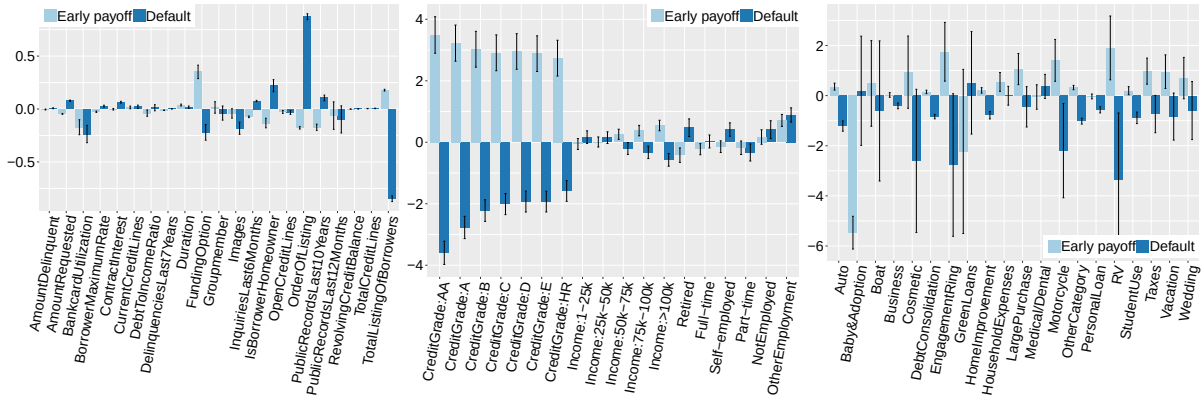
We summarize in Table 3 the features with a significant effect on the time to both early payoff and default. For example, a higher number of total credit lines accelerates the time to both events, while a higher bank card utilization decelerates the time to both events. Instead of accelerating or decelerating payoff and default simultaneously, many features have opposite effects on the time to the two events, as reported in the top-right and the bottom-left cells of Table 3. For example, a borrower with an annual income over 100,000 dollars has their default time significantly postponed but they would be more likely to pay off the debt early, compared to the condition that their income was lower. It

is also implied that if a borrower would like to accept a high-interest rate when proposing the loan listing, they are likely to pay off the loan late or to default early. A possible explanation is that borrowers who propose a high-interest rate on Prosper have either a high risk of default and late repayment or an urgent need for money that cannot be funded quickly by a traditional financial institution. So in addition to facilitating the loan funding procedure, proposing a high-interest rate helps the borrowers with a high risk of default to attract risk-seeking investors to make an adverse selection. In comparison, a high contract interest rate accelerates the time to default but does not influence the time to payoff when fixing all other features. The effect and the predictability of the contract interest rate will be discussed in detail in Section 5.4.

Table 3: Features with significant effects by WDR.

	Accelerating early payoff time	Decelerating early payoff time
Accelerating default time	`TotalCreditLines`, `OtherEmployment`	`AmountRequested`, `PublicRecordsLast10Years`, `IsBorrowerHomeowner`, `OrderOfListing`, `InquiriesLast6Months`, `BorrowerMaximumRate`
Decelerating default time	`TotalListingOfBorrowers`, `FundingOption`, `CreditGrade:AA, A, B, C, D, E and HR`, `Income:>100k`, `Category:Debt Consolidation, Auto and Other`	`BankcardUtilization`

It is worth to note that the accelerating and decelerating effects on the time to early payoff and default, respectively, are basically decreasing with borrowers' credit grades changing from high to low (compared to the baseline of unknown credit grades), as shown in Figure 3 (b). This also results in a trade-off between early payoff and early default when lenders select loan listings by the borrowers' credit grades. Additionally, soft information is important in predicting not only the chance of payoff and default [Iyer et al., 2015] but also the time. For example, Figure 3 (a) shows that if a borrower provides an image (`image=1`), their default time will be postponed. Panel (c) shows that a loan for auto or home improvement has significantly postponed default time if defaulted, potentially because borrowers applying for such loans have anticipated a better financial status and thus are less likely to default.

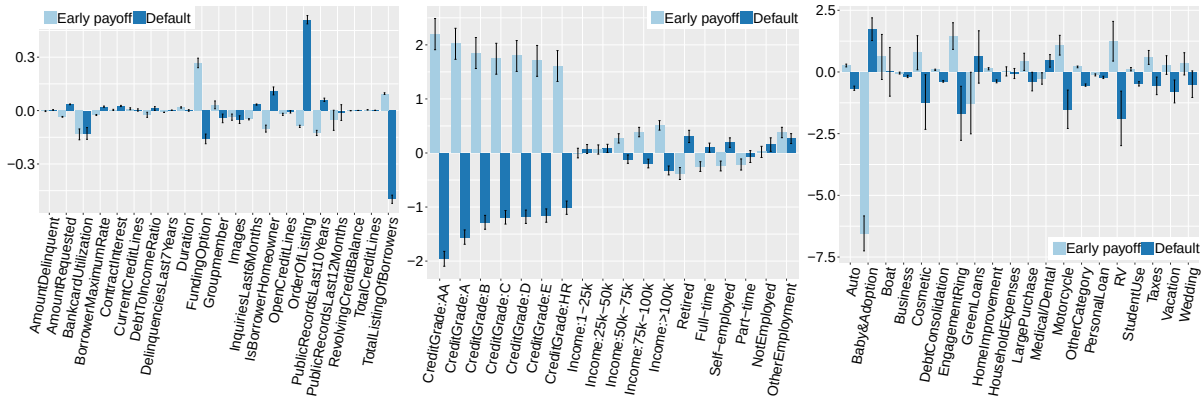


(a) Continuous and binary features.

(b) Credit grades, income levels and employment status.

(c) Loan purposes.

Figure 3: Feature effects ($\boldsymbol{\beta}/a$) with standard errors from WDR.



(a) Continuous and binary features.

(b) Credit grades, income levels and employment status.

(c) Loan purposes.

Figure 4: Feature coefficients ($\boldsymbol{\theta}$) with standard errors from FG model.

5.3 Model comparison

We compare our results by WDR with other survival and classification models to examine the credibility of the findings by WDR. First, we compare the calibration of feature effects by FG model which formulates subdistribution hazard function η_j for event j with a proportional-hazards assumption, i.e.,

$$\begin{aligned}\eta_j(i,t\,|\,\boldsymbol{x}_i) &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}Pr(t\le t_i<t+\Delta t, y_i=j\,|\,t_i\ge t \text{ or } (t_i\le t \text{ and } y_i\ne j),\boldsymbol{x}_i)\\ &= \eta_{j0}(t)\exp(\boldsymbol{x}_i'\boldsymbol{\theta}_j),\end{aligned}$$

where $\eta_{j0}(t)$ is a nonparametric baseline and $\boldsymbol{\theta}_j$ is the covariate coefficients to be estimated. A positive value of x_v with a positive coefficient θ_{vj} will increase the subdistribution hazard. In comparison, for WDR with one sub-event for each of the competing events, the hazard function for event j is increasing in x_v if $\beta_{vj} > 0$ as implied by (4). Therefore, if the parameters of WDR is correctly estimated, the signs of $\boldsymbol{\beta}_j$ should be the same as those of $\boldsymbol{\theta}_j$ for each j unless insignificant. We show the estimated $\boldsymbol{\theta}_j$ with the standard errors in Figure 4 and Table 14 in the appendix, and see that all the point estimates of feature effects are of the same sign, respectively, as those by WDR if significant, and the relative scales of the coefficients are also similar between WDR and FG. In addition, the model fit by WDR which finds linear covariate effects is comparable to, and for some time evaluated slightly better than the model fits by FG and random RF in terms of C-indices and Brier scores as shown in Table 12 and 13, respectively, in the appendix.

We can predict $P(y_i = j)$, the over-all probability of event j for observation i using $\text{CIF}_j(i, t = \infty)$, and consequently, a survival model for competing events like WDR, FG and RF can be used for classification of the event types. We show how well WDR does in distinguishing loans to be fully paid off from those to be defaulted, and compare it with FG and RF along with classification approaches including L_2-regularized logistic regression and a neural network with fully connected layers. Note that neural networks are arguably the most powerful classification models in the existence of nonlinearity if the data is large enough. We use the probability of 0.5 as the classification threshold and report in Table 4 the prediction accuracy and area under the receiver operating characteristic (ROC) curve which is denoted as AUC. We can see WDR gives very good prediction accuracy and AUC that are higher than all the other approaches except for slightly lower than logistic regression. It is worth to mention that the neural network classification may have over-fit the training data since the training accuracy has achieved as high as 97.6% whereas the testing accuracy and AUC are the lowest among all the

Table 4: Prediction accucary and AUC for classification of early payoff and default.

	logistic	neural network	FG	RF	WDR
Accuracy	**0.7373**	0.6893	0.7350	0.7200	0.7353
AUC	**0.8049**	0.7638	0.7995	0.8002	0.8046

methods compared.

5.4 Prediction by partial information

Table 5: Brier scores by WDR using different features.

	Early payoff				Default			
time	Interest rate	Credit grade	Standard info	All info	Interest rate	Credit grade	Standard info	All info
5	0.0778	0.0775	0.0761	**0.0760**	0.0151	**0.0147**	0.0163	0.0172
10	0.1472	0.1459	0.1353	**0.1323**	0.0806	0.0807	0.0782	**0.0744**
15	0.1887	0.1866	0.1680	**0.1618**	0.1414	0.1434	0.1362	**0.1249**
20	0.2143	0.2114	0.1891	**0.1800**	0.1767	0.1794	0.1698	**0.1534**
25	0.2289	0.2255	0.2006	**0.1889**	0.1952	0.1982	0.1879	**0.1694**
30	0.2368	0.2331	0.2091	**0.1951**	0.2078	0.2109	0.2000	**0.1802**
35	0.2396	0.2360	0.2145	**0.1962**	0.2168	0.2204	0.2073	**0.1864**

Table 6: C-indices by WDR using different features.

	Early payoff				Default			
time	Interest rate	Credit grade	Standard info	All info	Interest rate	Credit grade	Standard info	All info
5	0.5385	0.5537	0.6744	**0.7091**	0.7991	0.7395	0.8127	**0.8424**
10	0.5539	0.5761	0.6982	**0.7290**	0.7314	0.7114	0.7557	**0.8115**
15	0.5608	0.5765	0.6970	**0.7296**	0.7065	0.6829	0.7245	**0.7860**
20	0.5687	0.5817	0.6918	**0.7268**	0.6930	0.6713	0.7138	**0.7756**
25	0.5784	0.5887	0.6897	**0.7279**	0.6804	0.6582	0.7019	**0.7729**
30	0.5830	0.5910	0.6828	**0.7229**	0.6747	0.6535	0.6959	**0.7945**
35	0.5926	0.5951	0.6756	**0.7199**	0.6697	0.6472	0.6928	**0.7504**

Table 7: Classification accuracy and AUC by WDR using different features.

	Interest rate	Credit grade	Standard info	All info
Accucary	0.6280	0.6257	0.6623	**0.7353**
AUC	0.6825	0.6546	0.7345	**0.8046**

In previous sections, we have validated WDR from the modeling perspective by showing its data-adaptive nonlinear capacity and comparable performance to other popular methods in survival analysis and classification. In this section, we study the predictability of the Prosper features containing standard and soft information using WDR. We use the contract interest rate, credit grades, all standard information and all information

(whose prediction results have already been discussed in Section 5.3), respectively, to fit the WDR model and compare the predictability of these groups of features by examining the goodness of CIF estimations that is quantified by Brier scores and C-indices, and how accurately WDR distinguishes loan payoff and default. As the contract interest rate reflects lenders' evaluation of a loan listing, its predictability suggests how well lenders are making decisions. The credit grades stratify borrowers by their past behavior and summarize the borrowers' overall creditworthiness. By comparing the predictability of the two features we can see if lenders can acquire information of borrowers in addition to their credit history. Furthermore, compared to only using standard information we examine the improvement by using soft information which can be overlooked by traditional financial institutions.

Only one sub-event for early payoff and default, respectively, has been found by WDR for each of these four groups of features. So the feature effects are linear. We report the Brier scores in Table 5, C-indices in Table 6 and the classification accuracy and AUC in Table 7, respectively. Overall, the prediction using all information dominates those using the other groups of partial information. Compared to using the standard information, using additional soft information leads to a remarkable improvement in the prediction of the default probability and time. In particular, using additional soft information results in an increase of 0.0701 or 9.54% in AUC compared to using standard information alone. Note that an increase of 0.01 in AUC is a considerable improvement when predicting the loan default [Iyer et al., 2015]. The CIF is also better estimated using additional soft information as shown by the smaller values of Brier scores and the larger values of C-indices. The only exception lies in the Brier score at time = 5, indicating a default at the fifth month can be slightly better predicted by the credit grade or the contract interest rate alone. The quantification by C-indices does not show such an exception.

When predicting the CIF for loan default, WDR using the contract interest rate slightly outperforms WDR using the credit grade in terms of both Brier scores and C-

indices. Meanwhile, the prediction of the overall default probability using the contract interest rate alone is better than that using the credit grades; the difference of AUC is as large as 0.028. Moreover, the predictability of the contract interest rate on the CIF for loan default achieves, on average, 97.2% of standard information and 89.5% of all (standard and soft) information in terms of C-indices. So the contract interest rate determined by bidding contains a tremendous proportion of all information available and reflects lenders' accurate judgment of loan default. When predicting the CIF for loan payoff, however, the contract interest rate and the credit grades do not capture as much useful information as they do in predicting the default. Concretely, the C-indices by using the contract interest rate, which are all below 0.6 at the time evaluated, only achieve 82.7% and 78.5% of those using standard information and all information, respectively. This indicates that the crowd wisdom cares about the loan default (time and probability) more than the early payoff.

6 Conclusion and future work

The proposed Weibull delegate racing model for survival analysis with competing events uses a two-phase race among sub-events, not only maintaining the interpretability, but also allowing for non-monotonic feature effects. We use a gamma process to support a potentially countably infinite number of sub-events, and rely on its inherent shrinkage mechanism to remove unneeded model capacity, making WDR capable of detecting underlying mechanisms data-adaptively. Moreover, WDR does not break the noninformative censoring assumption and thus reduces the bias in CIF estimations. Analysis of the Prosper data shows WDR's outstanding performance and intriguing findings that can inspire investors to screen online loan listings using standard and soft information. The interpretation of WDR as an accelerated failure time model enables us to both qualitatively and quantitatively summarize the feature effects on time to default and payoff.

Therefore, WDR is an attractive alternative to existing methods for not only studying loan payoff and default but also many other applications that require interpretable nonlinear modeling capacity.

Many different model extensions and applications have been left for the future. From the modeling perspective, WDR assumes the (sub-)events are conditionally independent given the features and has not considered an intrinsic correlation. A possible improvement is to add random effects to capture the correlation among competing events and among their associated sub-events, respectively. We have tried this on the prosper data but did not find significant random effects, possibly because the standard and soft information already capture the potential correlation between the time to payoff and the time to default. This conditional-independence assumption may not hold in more complex scenarios where the features are sparse or do not have much predictability. For example, if a patient had multiple diseases and they died of one disease, their survival time may be different from those who had the same values of features and died of the same disease. In this case, the interaction between diseases cannot be reflected by measurable pathology indices and thus the latent survival times to these diseases are not conditionally independent given the features. We believe that adding random effects can remarkably improve WDR's performance in such cases.

We do not find nonlinear feature effects as WDR has learned only one sub-event for payoff and default, respectively, from the data. This finding is valuable as it shows borrowers' features on Prosper.com are informative for lenders to easily make good decisions to avoid loans with high risks of default. As we have selected the data of fully funded loans in the time period when the bidding system was in use, the data may be of high homogeneity and thus interactions among features may not be significant; nonlinear feature effects might emerge if we analyze more loans in a longer period of time. Another application of WDR using the Prosper data is to study the loan funding process. Specifically, within the funding period a proposed listing may have two endpoints as two competing

events; one is fully funded and the other is withdrawal by borrowers.

Another WDR's application area is in biomedical studies where the pre-specification of competing events is subject to human knowledge and diagnostic techniques, so that a competing event is very likely to include several unknown sub-events. We have analyzed a microarray gene-expression data of 240 patients with diffuse large B-cell lymphoma (DLBCL) [Rosenwald et al., 2002]. Multiple unsuccessful treatments to increase the survival rate suggest that there exist several subtypes of DLBCL that differ in responsiveness to chemotherapy. In the DLBCL dataset, Rosenwald et al. [2002] identify three gene-expression subgroups, including activated B-cell-like, germinal-center B-cell-like, and "type 3" which may be related to three different diseases as a result of distinct mechanisms of malignant transformation. They also suspect that "type 3" may be associated with more than one such mechanism. We use WDR to study the survival under the three types of DLBCL, and have found that "type 3" actually consists of two sub-types. This finding demonstrate the huge success of WDR in discovering new diseases.

References

P. C. Austin, D. S. Lee, and J. P. Fine. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, 133(6):601–609, 2016.

B. Baesens, T. Van Gestel, M. Stepanova, D. Van den Poel, and J. Vanthienen. Neural network survival analysis for personal loan data. *Journal of the Operational Research Society*, 56(9): 1089–1098, 2005.

F. Ballester, D. Corella, S. Pérez-Hoyos, M. Sáez, and A. Hervás. Mortality as a function of temperature. A study in Valencia, Spain, 1991-1993. *International Journal of Epidemiology*, 26(3):551–561, 1997.

J. Banasik, J. N. Crook, and L. C. Thomas. Not if but when will borrowers default. *Journal of the Operational Research Society*, 50(12):1185–1190, 1999.

J. E. Barrett and A. C. Coolen. Gaussian process regression for survival data with competing risks. *arXiv preprint arXiv:1312.1591*, 2013.

D. R. Cox. Regression models and life-tables. In *Breakthroughs in Statistics*, pages 527–541. Springer, 1992.

M. J. Crowder. *Classical competing risks*. CRC Press, 2001.

P. Damlen, J. Wakefield, and S. Walker. Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):331–344, 1999.

J. A. Divino, E. S. Lima, and J. Orrillo. Interest rates and default in unsecured loan markets. *Quantitative Finance*, 13(12):1925–1934, 2013.

J. Duarte, S. Siegel, and L. Young. Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, 25(8):2455–2484, 2012.

A. Durovic. Estimating probability of default on peer to peer market–survival analysis approach. *Journal of Central Banking Theory and Practice*, 6(2):149–167, 2017.

J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

R. G. FitzJohn. Diversitree: Comparative phylogenetic analyses of diversification in r. *Methods in Ecology and Evolution*, in press, 2012. doi: 10.1111/j.2041-210X.2012.00234.x.

K. M. Flegal, B. I. Graubard, D. F. Williamson, and M. H. Gail. Cause-specific excess deaths associated with underweight, overweight, and obesity. *Jama*, 298(17):2028–2037, 2007.

T. A. Gerds. *pec: Prediction Error Curves for Risk Prediction Models in Survival Analysis*, 2017. URL `https://CRAN.R-project.org/package=pec`. R package version 2.5.4.

T. A. Gerds, T. Cai, and M. Schumacher. The performance of risk prediction models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(4):457–479, 2008.

B. Gray. *cmprsk: Subdistribution Analysis of Competing Risks*, 2014. URL `https://CRAN.R-project.org/package=cmprsk`. R package version 2.2-7.

W. H. Greene. *Econometric analysis*. Pearson Education India, 2003.

W. M. Hanemann. Discrete/continuous models of consumer demand. *Econometrica: Journal of the Econometric Society*, pages 541–561, 1984.

H. Ishwaran and U. Kogalur. *Random Forests for Survival, Regression, and Classification (RF-SRC)*, 2018. URL `https://cran.r-project.org/package=randomForestSRC`. R package version 2.6.0.

H. Ishwaran, T. A. Gerds, U. B. Kogalur, R. D. Moore, S. J. Gange, and B. M. Lau. Random survival forests for competing risks. *Biostatistics*, 15(4):757–773, 2014.

R. Iyer, A. I. Khwaja, E. F. Luttmer, and K. Shue. Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577, 2015.

J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

B. Lau, S. R. Cole, and S. J. Gange. Competing risk regression models for epidemiologic data. *American Journal of Epidemiology*, 170(2):244–256, 2009.

C. Lee, W. R. Zame, J. Yoon, and M. van der Schaar. DeepHit: A deep learning approach to survival analysis with competing risks. AAAI, 2018.

M. Lin, N. R. Prabhala, and S. Viswanathan. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, 59(1):17–35, 2013.

S. Miller. Risk factors for consumer loan default: a censored quantile regression analysis. Technical report, Working paper. University of Illinois, 2012.

P. G. Moschopoulos. The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, 37(1):541–544, 1985.

R. M. Neal et al. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.

H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430, 2007.

A. Rosenwald, G. Wright, W. C. Chan, J. M. Connors, E. Campo, R. I. Fisher, R. D. Gascoyne, H. K. Muller-Hermelink, E. B. Smeland, J. M. Giltnane, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New England Journal of Medicine*, 346(25):1937–1947, 2002.

P. Saha and P. Heagerty. Time-dependent predictive accuracy in the presence of competing risks. *Biometrics*, 66(4):999–1011, 2010.

M. Stepanova and L. Thomas. Survival analysis methods for personal loan data. *Operations Research*, 50(2):277–289, 2002.

E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan. Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, 21(1):128, 2010.

M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1(Jun):211–244, 2001.

E. N. Tong, C. Mues, and L. C. Thomas. Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, 218(1):132–139, 2012.

K. E. Train. *Discrete choice methods with simulation*. Cambridge University Press, 2009.

R. Varma, N. M. Bressler, Q. V. Doan, M. Gleeson, M. Danese, J. K. Bower, E. Selvin, C. Dolan, J. Fine, S. Colman, et al. Prevalence of and risk factors for diabetic macular edema in the United States. *JAMA Ophthalmology*, 132(11):1334–1340, 2014.

M. Wolbers, P. Blanche, M. T. Koller, J. C. Witteman, and T. A. Gerds. Concordance for prognostic models with competing risks. *Biostatistics*, 15(3):526–539, 2014.

J. Zhang and P. Liu. Rational herding in microloan markets. *Management Science*, 58(5): 892–912, 2012.

M. Zhou. Softplus regressions and convex polytopes. *arXiv:1608.06383*, 2016.

M. Zhou and L. Carin. Negative binomial process count and mixture modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(2):307–320, 2015.

M. Zhou, Y. Cong, and B. Chen. Augmentable gamma belief networks. *Journal of Machine Learning Research*, 17(163):1–44, 2016.

Weibull Racing Time-to-event Modeling and Analysis of Online Borrowers' Loan Payoff and Default: Appendix

A Theorems and proofs

A.1 Proof of Proposition 1

Proof of Proposition 1. Since $P(t > \tau) = \prod_j P(t_j > \tau) = e^{-\tau^a \sum_j \lambda_j}$, then $t = \min_j t_j \sim$ Weibull$(a, \sum_j \lambda_j)$. Assuming $t_h = \min_j t_j$, we have

$$\begin{aligned} P(\tau < \min_j t_j,\ t_h = \min_j t_j) &= \prod_{j \neq h} P(\tau < t_h < t_j) \\ &= \int_\tau^\infty f(t_h \mid a, \lambda_h) \prod_{j \neq h} P(t_h < t_j) dt_h \\ &= \int_\tau^\infty e^{-t_h^a \sum_{j \neq h} \lambda_j} a \lambda_h t_h^{a-1} e^{-\lambda_h t_h^a} dt_h \\ &= \frac{\lambda_h}{\sum_j \lambda_j} e^{-\tau^a \sum_j \lambda_j}. \end{aligned}$$

Let $\tau = 0$. We have $P(t_h = \min_j t_j) = \frac{\lambda_h}{\sum_j \lambda_j}$. This proves the categorical distribution of $y = \operatorname{argmin}_j t_j$. Consequently,

$$P(\tau < \min t_j,\ t_h = \min_j t_j) = P(\tau < t,\ y = h) = P(\tau < t) P(y = h).$$

This proves the independence between t and y. □

A.2 Marginal distribution of the WDR event time

Theorem A.1. *If* $t_i \sim$ *Weibull*$(a, \lambda_{i\bullet\bullet})$ *with* $\lambda_{i\bullet\bullet} = \sum_{j,k} \lambda_{ijk}$ *and* $\lambda_{ijk} \sim$ *Gamma*$(r_{jk}, 1/b_{ijk})$, *the PDF of* t_i *given* a, $\{r_{jk}\}$ *and* $\{b_{ijk}\}$ *is*

$$f(t_i \,|\, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k}) = a t_i^{a-1} c_i \sum_{m=0}^{\infty} \frac{(\rho_i + m)\delta_{im} b_{i(1)}^{\rho_i+m}}{(t_i^a + b_{i(1)})^{1+\rho_i+m}},$$

and the cumulative density function (CDF) is

$$P(t_i < q \,|\, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k}) = 1 - c_i \sum_{m=0}^{\infty} \frac{\delta_{im} b_{i(1)}^{\rho_i+m}}{(q^a + b_{i(1)})^{\rho_i+m}}, \tag{9}$$

where $c_i = \prod_{j,k} \left(\frac{b_{ijk}}{b_{i(1)}}\right)^{r_{jk}}$, $b_{i(1)} = \max_{j,k} b_{ijk}$, $\rho_i = \sum_{j,k} r_{jk}$, $\delta_{i0} = 1$, $\delta_{im+1} = \frac{1}{m+1} \sum_{h=1}^{m+1} h\gamma_{ih}\delta_{im+1-h}$ *for* $m \geq 1$, *and* $\gamma_{ih} = \sum_{j,k} \frac{r_{jk}}{h} \left(1 - \frac{b_{ijk}}{b_{i(1)}}\right)^h$.

It is difficult to utilize the PDF or CDF of t_i in the form of series, but we can use a finite truncation as an approximate. Concretely, as $P(t_i < \infty \,|\, a, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k}) = c_i \sum_{m=0}^{\infty} \delta_{im} = 1$, we find an M so large that $c_i \sum_{m=0}^{M} \delta_{im}$ close to 1 (say no less than 0.9999), and use $1 - c_i \sum_{m=0}^{M} \frac{\delta_{im} b_{i(1)}^{\rho_i+m}}{(q^a + b_{i(1)})^{\rho_i+m}}$ as an approximation. Consequently, sampling t_i is feasible by inverting the approximated CDF for general cases. We have tried prediction by finite truncation on some synthetic data where $a = 1$ and found M is mostly between 10 and 30, which is computationally acceptable.

Proof. We first study the distribution of gamma convolution. Specifically, if $\lambda_t \overset{ind}{\sim}$ Gamma$(r_t, 1/b_t)$ with $r_t, b_t \in \mathbb{R}_+$, then the PDF of $\lambda = \sum_{t=1}^{T}$ can be written in a form of series Moschopoulos [1985] as

$$f(\lambda \,|\, r_1, b_1, \cdots, r_T, b_T) = \begin{cases} c \sum_{m=0}^{\infty} \frac{\delta_m \lambda^{\rho+m-1} e^{-\lambda b_{(1)}}}{\Gamma(\rho+m)/b_{(1)}^{\rho+m}} & \text{if } \lambda > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $c = \prod_{t=1}^{T} \left(\frac{b_t}{b_{(1)}}\right)^{r_t}$, $b_{(1)} = \max_t b_t$, $\rho = \sum_{t=1}^{T} r_t$, $\delta_0 = 1$, $\delta_{m+1} = \frac{1}{m+1} \sum_{h=1}^{m+1} h\gamma_h \delta_{m+1-h}$

and $\gamma_h = \sum_{t=1}^{T} r_t \left(1 - \frac{b_t}{b_{(1)}}\right)^h / h$. Moschopoulos [1985] proved that $0 < \gamma_{ih} \leq \rho_i b_{i0}^h / h$ and $0 < \delta_{im} \leq \frac{\Gamma(\rho_i+m) b_{i0}^m}{\Gamma(\rho_i) m!}$ where $b_{i0} = max_{j,k}(1 - \frac{b_{ijk}}{b_{i(1)}})$. We want to show the PDF of t_i,

$$
\begin{aligned}
& f(t_i \,|\, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k}) \\
=& \int_0^\infty f(t_i \,|\, \lambda_{i\bullet\bullet}) f(\lambda_{i\bullet\bullet} \,|\, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k}) d\lambda_{i\bullet\bullet} \\
=& \int_0^\infty \sum_{m=0}^{\infty} \frac{c_i \delta_{im} a t_i^a \lambda_{i\bullet\bullet}^{\rho_i+m} \exp(-t_i^a \lambda_{i\bullet\bullet} - b_{i(1)} \lambda_{i\bullet\bullet})}{\Gamma(\rho_i+m)/b_{i(1)}^{\rho_i+m}} d\lambda_{i\bullet\bullet} \\
=& \sum_{m=0}^{\infty} \int_0^\infty \frac{c_i \delta_{im} a t_i^a \lambda_{i\bullet\bullet}^{\rho_i+m} \exp(-t_i^a \lambda_{i\bullet\bullet} - b_{i(1)} \lambda_{i\bullet\bullet})}{\Gamma(\rho_i+m)/b_{i(1)}^{\rho_i+m}} d\lambda_{i\bullet\bullet} \qquad (10) \\
=& a t_i^{a-1} c_i \sum_{m=0}^{\infty} \frac{(\rho_i+m) \delta_{im} b_{i(1)}^{\rho_i+m}}{(t_i^a + b_{i(1)})^{1+\rho_i+m}},
\end{aligned}
$$

which suffices to prove the equality in (10). Note that

$$
\begin{aligned}
& f(t_i \,|\, n_i, \lambda_{i\bullet\bullet}) f(\lambda_{i\bullet\bullet} \,|\, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k}) \\
=& a t_i^{a-1} c_i \lambda_{i\bullet\bullet}^{\rho_i} b_{i(1)}^{\rho_i} \exp(-t_i^a \lambda_{i\bullet\bullet} - b_{i(1)} \lambda_{i\bullet\bullet}) \sum_{m=0}^{\infty} \frac{\delta_{im} b_{i(1)}^m \lambda_{i\bullet\bullet}^m}{\Gamma(\rho_i+m)} \\
\leq& a t_i^{a-1} c_i \lambda_{i\bullet\bullet}^{\rho_i} b_{i(1)}^{\rho_i} \exp(-t_i^a \lambda_{i\bullet\bullet} - b_{i(1)} \lambda_{i\bullet\bullet}) \sum_{m=0}^{\infty} \frac{(b_{i0} b_{i(1)} \lambda_{i\bullet\bullet})^m}{\Gamma(\rho_i) m!} \\
=& a t_i^{a-1} c_i \lambda_{i\bullet\bullet}^{\rho_i} b_{i(1)}^{\rho_i} \exp(-t_i^a \lambda_{i\bullet\bullet} - b_{i(1)} \lambda_{i\bullet\bullet} + b_{i0} b_{i(1)} \lambda_{i\bullet\bullet}),
\end{aligned}
$$

which shows the uniform convergence of $f(t_i \,|\, n_i, \lambda_{i\bullet\bullet}) f(\lambda_{i\bullet\bullet} \,|\, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k})$. So the integration and countable summation are interchangeable, and consequently, (10) holds. Next we want to calculate the CDF of t_i,

$$
\begin{aligned}
P(t_i < q \,|\, n_i, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k}) &= \int_0^q a t_i^{a-1} c_i \sum_{m=0}^{\infty} \frac{(\rho_i+m) \delta_{im} b_{i(1)}^{\rho_i+m}}{(t_i^a + b_{i(1)})^{1+\rho_i+m}} dt_i \\
&= \sum_{m=0}^{\infty} \int_0^q a t_i^{a-1} c_i \frac{(\rho_i+m) \delta_{im} b_{i(1)}^{\rho_i+m}}{(t_i^a + b_{i(1)})^{1+\rho_i+m}} dt_i. \qquad (11)
\end{aligned}
$$

It suffices to show (11). Note that

$$
\begin{aligned}
&\sum_{m=0}^{\infty} \frac{(\rho_i+m)\delta_{im} b_{i(1)}^{\rho_i+m}}{(t_i^a+b_{i(1)})^{1+\rho_i+m}} \\
=&\sum_{m=0}^{\infty} \frac{\Gamma(1+\rho_i+m)\delta_{im} b_{i(1)}^{\rho_i+m}}{\Gamma(\rho_i+m)(t_i^a+b_{i(1)})^{n_i+\rho_i+m}} \\
\leq&\sum_{m=0}^{\infty} \frac{\Gamma(1+\rho_i+m) b_{i(1)}^{\rho_i+m}\Gamma(1+\rho_i)}{\Gamma(\rho_i+m)(t_i^a+b_{i(1)})^{1+\rho_i+m}\Gamma(\rho_i)m!} \\
=&\frac{\Gamma(\rho_i+1) b_{i(1)}^{\rho_i}}{\Gamma(\rho_i)(t_i^a+b_{i(1)})^{1+\rho_i}} \sum_{m=0}^{\infty}\left[\frac{\Gamma(1+\rho_i+m)}{\Gamma(1+\rho_i)m!}\left(\frac{b_{i(1)}}{t_i^a+b_{i(1)}}\right)^m\right] \\
=&\frac{\Gamma(\rho_i+1) b_{i(1)}^{\rho_i} t_i^{a(1+\rho_i)}}{\Gamma(\rho_i)(t_i^a+b_{i(1)})^{2(1+\rho_i)}}.
\end{aligned}
$$

The last equation holds because the summation of a negative binomial probability mass function is 1. So $f(t_i \,|\, a, \{r_{jk}\}_{j,k}, \{b_{ijk}\}_{j,k})$ is uniformly convergent and (11) holds. Calculating the integration, we obtain the CDF of t_i. □

B MCMC algorithm for WDR

Let us denote T_i and T_{ic} as the observed failure time and the right censoring time, respectively, for observation i. Since left censoring is uncommon and not shown in the Prosper data to be analyzed, we only consider right censoring in our inference and leave to readers other types of censoring which can be analogously done. The inference by MCMC accommodating missing event time or missing event types proceeds by iterating the following steps.

Step 1: If y_i is observed, we first sample κ_{iy_i} by

$$
P(\kappa_{iy_i} = k \,|\, y_i, \cdots) = \frac{\lambda_{iy_ik}}{\sum_{k'=1}^{K}\lambda_{iy_ik'}}.
$$

If y_i is unobserved which means a missing event type, we sample (y_i, κ_{iy_i}) by

$$P(y_i = j, \kappa_{iy_i} = k \,|\, \cdots) = \frac{\lambda_{ijk}}{\sum_{j'=1}^{S}\sum_{k'=1}^{K}\lambda_{ij'k'}}.$$

We then denote $m_{jk} = \sum_{i:y_i=j} \mathbf{1}(\kappa_{iy_i} = k)$. Define $n_{ijk} = 1$ if $y_i = j$ and $\kappa_{iy_i} = k$, and otherwise $n_{ijk} = 0$. The above sampling procedure means that given the event type y_i, we sample the index of the sub-event that has the minimum event time.

Step 2: Update t_i for $i = 1, \cdots, n$, $j = 1, \cdots, J$ and $k = 1, \cdots, K$.

(a) If the event time T_i is observed, we set $t_i = T_i$.

(b) Otherwise, we sample $t_i \sim \text{Weibull}_{(T_{ic},\infty)}(a, \sum_{j=1}^{S}\sum_{k=1}^{K}\lambda_{ijk})$ where $\text{Weibull}_{(T_{ic},\infty)}(\cdot,\cdot)$ is a truncated weibull distribution so that $t_i \in (T_{ic}, \infty)$. Note $T_{ic} = 0$ if both event time and censoring time are missing for observation i.

Step 3: Sample $(\lambda_{ijk} \,|\, -) \sim \text{Gamma}\left(r_{jk} + n_{ijk}, \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}}}{1+t_i^a e^{\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}}}\right)$, for $i = 1, \cdots, n$, $j = 1, \cdots, J$ and $k = 1, \cdots, K$.

Step 4: Sample a by slice sampling. Since a determines how the hazard varies with time, we assume an improper prior of $a \in \mathbb{R}_+$, i.e., $p(a) \propto \mathbf{1}(a > 0)$ to reduce the impact of the prior on the posterior.

$$p(a \,|\, \ldots) \propto p(a)p(t_i, y_{ii} \,|\, a, \ldots) = a^n \prod_i \left[t_i^{a-1} \prod_{j,k} \left(1 + t_i^a e^{\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}}\right)^{n_{ijk}+r_{jk}} \right]$$

The uni-modality of $p(a \,|\, \ldots)$ can be shown so that slice sampling can be implemented efficiently. Concretely,

$$d \log p(a|\ldots)/da = \frac{n}{a} + \sum_i \log t_i - \sum_{i,j,k}(n_{ijk} + r_{jk})\frac{e^{a\log t_i + \boldsymbol{x}_i'\boldsymbol{\beta}_{jk}} \log t_i}{1 + e^{a\log t_i + \boldsymbol{x}_i'\boldsymbol{\beta}_{jk}}}.$$

Since $\frac{n}{a}$ in decreasing in a while $\sum_{i,j,k}(n_{ijk}+r_{jk})\frac{e^{a\log t_i+\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}}\log t_i}{1+e^{a\log t_i+\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}}}$ is increasing in a, there must be at most one $a\in\mathbb{R}_+$ so that $d\log p(a|\ldots)/da=0$. So $p(a|\ldots)=0$ is uni-modal. We use the `mcmc` function in `R` package `diversitree` [FitzJohn, 2012].

Step 5: Sample $\boldsymbol{\beta}_{jk}$, for $j=1,\cdots,J$ and $k=1,\cdots,K$, by Pólya Gamma (PG) data augmentation. First Sample $(\omega_{ijk}\,|-)\sim \mathrm{PG}(r_{jk}+n_{ijk},\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}+a\log t_i)$. Then sample $(\boldsymbol{\beta}_{jk}\,|-)\sim\mathrm{MVN}(\boldsymbol{\mu}_{jk},\boldsymbol{\Sigma}_{jk})$ where $\boldsymbol{\Sigma}_{jk}=(V_{jk}+\boldsymbol{X}'\Omega_{jk}\boldsymbol{X})^{-1}$, $\boldsymbol{X}=[\boldsymbol{x}_1',\cdots,\boldsymbol{x}_N']'$, $\Omega_{jk}=\mathrm{diag}(\omega_{1jk},\cdots,\omega_{njk})$ and $\boldsymbol{\mu}_{jk}=\boldsymbol{\Sigma}_{jk}\left[-\sum_{i=1}^N\left(a\omega_{ijk}\log t_i+\frac{r_{jk}-n_{ijk}}{2}\right)\boldsymbol{x}_i\right]$. Note to sample from the Pólya-Gamma distribution, we use a fast and accurate approximate sampler of Zhou [2016] that matches the first two moments of the original distribution; we set the truncation level of that sampler as five.

Step 6: Sample $(\alpha_{vjk}\,|-)\sim\mathrm{Gamma}\left(a_0+0.5,1/(b_0+0.5\beta_{vjk}^2)\right)$ for $v=0,\cdots,V$, $j=1,\cdots,J$ and $k=1,\cdots,K$.

Step 7: Sample r_{jk} and γ_{0j}, for $j=1,\cdots,J$ and $k=1,\cdots,K$, by Chinese restaurant table (CRT) data augmentation Zhou and Carin [2015].

First sample $(n_{ijk}^{(2)}\,|-)\sim\mathrm{CRT}(n_{ijk},r_{jk})$, and $(l_{jk}\,|-)\sim\mathrm{CRT}(\sum_{i=1}^N n_{ijk}^{(2)},\gamma_{0j}/K)$. Then sample $(r_{jk}\,|-)\sim\mathrm{Gamma}\left(\sum_{i=1}^N n_{ijk}^{(2)}+\gamma_{0j}/K,\frac{1}{c_{0j}+\sum_{i=1}^N\log(1+t_i^a e^{\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}})}\right)$, and $(\gamma_{0j}\,|-)\sim\mathrm{Gamma}\left(e_0+\sum_{k=1}^K l_{jk},\frac{1}{f_0-\frac{1}{K}\sum_{k=1}^K\log(1-p_{jk})}\right)$,
where $p_{jk}=\frac{\sum_{i=1}^N\log(1+t_i^a e^{\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}})}{c_{0j}+\sum_{i=1}^N\log(1+t_i^a e^{\boldsymbol{x}_i'\boldsymbol{\beta}_{jk}})}$.

Step 8: Sample $(c_{0j}\,|-)\sim\mathrm{Gamma}\left(e_1+\gamma_{0j},\frac{1}{f_1+\sum_{k=1}^K r_{jk}}\right)$ for $j=1,\cdots,J$.

Step 9: For $j=1,\cdots,J$ and $k=1,\cdots,K$, prune sub-risk k of risk j for all observations if $m_{jk}=0$, by setting $\lambda_{ijk}\equiv 0$ and $t_{ijk}\equiv\infty$ for $\forall i$.

Step 1 through 8 are MCMC updates of parameters. Step 9 is used to explicitly prune unneeded nonlinear modeling capacity. Although the item weights in the gamma process

is almost surely positive, the latent count allocation of n_{ijk} and $n^{(2)}_{ijk}$ in Step 1 and 7 make it possible to prune a sub-event which does not significantly represent a latent mechanism of a competing event.

C Maximum a posteriori estimations of WDR

With the reparameterization that $\lambda_{ijk} = \tilde{\lambda}_{ijk} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_{jk}}$ where $\tilde{\lambda}_{ijk} \overset{iid}{\sim} \text{Gamma}(r_{jk}, 1)$ we first find p_i, the likelihood of observation i having event type y_i at event time t_i.

$$p_i = \mathbb{E}\left(P(t_i, y_i \,|\, \boldsymbol{\lambda}_i)\right) \equiv \int \left(p_{t_i} \times p_{y_i}\right) p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) d\tilde{\boldsymbol{\lambda}}_i$$

where $\tilde{\boldsymbol{\lambda}}_i = \{\tilde{\lambda}_{ijk}\}_{j,k}$, $p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) = \prod_{j,k} \text{Gamma}(r_{jk}, 1)$, $\boldsymbol{r} = \{r_{jk}\}_{j,k}$, $\text{Gamma}(r_{jk}, 1)$ is the pdf of a gamma distribution with shape r_{jk} and scale 1, and

$$p_{t_i} = \begin{cases} a t_i^{a-1} (\sum_{j,k} \tilde{\lambda}_{ijk} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_{jk}}) \exp\left\{-t_i^a \sum_{jk} \tilde{\lambda}_{ijk} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_{jk}}\right\} & \text{if } t_i \text{ is uncensored and observed,} \\ \exp\left\{-T_{ic}^a \sum_{jk} \tilde{\lambda}_{ijk} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_{jk}}\right\} & \text{if } t_i \text{ is right censored at } T_{ic}, \text{ i.e., } t_i > T_{ic}, \\ 1 & \text{if } t_i \text{ is missing, but } y_i \text{ is not,} \end{cases}$$

$$p_{y_i} = \begin{cases} \frac{\sum_k \tilde{\lambda}_{iy_ik} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_{y_ik}}}{\sum_{j,k} \tilde{\lambda}_{ijk} e^{\boldsymbol{x}_i' \boldsymbol{\beta}_{jk}}} & \text{if } y_i \text{ is not missing,} \\ 1 & \text{if } y_i \text{ is missing, but } t_i \text{ is not.} \end{cases}$$

Note that we do not define $P(t_i, y_i \,|\, \boldsymbol{\lambda}_i)$ if both t_i and y_i are missing and remove such observations from data. We write $p_{t_i} \equiv p_t(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r})$ and $p_{y_i} \equiv p_y(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r})$.

Imposing a prior $p(a)$ on a, $p(\boldsymbol{\beta}_{jk})$ on $\boldsymbol{\beta}_{jk}$ and $p(r_{jk})$ on r_{jk}, the log posterior is

$$\log P = \sum_i \log p_i + \log p(a) + \sum_{j,k} \log p(\boldsymbol{\beta}_{jk}) + \sum_{j,k} \log p(r_{jk}) + C \tag{12}$$

where C is a constant function of a, $\{\boldsymbol{\beta}_{jk}\}$ and $\{r_{jk}\}$. In practice we assume an improper

prior $p(a) \propto 1$ on a, a Student's t distribution with degrees of freedom α on each element of $\boldsymbol{\beta}_{jk}$ and a Gamma$(0.01/K, 1/0.01)$ prior on r_{jk}. We also found a Gamma$(1/K, 1)$ prior on r_{jk} or an L_2-regularizer, $0.001||\boldsymbol{r}||_2$, is more numerically stable. Then we have

$$\log P = \sum_i \log p_i + \sum_{v,j,k} -\frac{\alpha+1}{2} \log\left(1 + \beta_{vjk}^2/\alpha\right) + \sum_{j,k} \left[(0.01/K - 1)\log r_{jk} - 0.01 r_{jk}\right] + c$$

where c is also a constant function of a, $\{\boldsymbol{\beta}_{jk}\}$ and $\{r_{jk}\}$. For simplicity, we define $\boldsymbol{\beta} = \{\boldsymbol{\beta}_{jk}\}_{j,k}$. We want to maximize $\log P$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{r}$. The difficulty lies in p_i being the expectation of $p_{t_i} \times p_{y_i}$ over $\tilde{\boldsymbol{\lambda}}_i$ which is a random variable parameterized by $\boldsymbol{r}$. Now we show how to approximate the derivatives of $\log p_i$ by Monte-Carlo simulation and score function gradients. Specifically,

$$\nabla_{a,\boldsymbol{\beta}} \log p_i = \frac{\int \left[\nabla_{a,\boldsymbol{\beta}} \left(p_{t_i} \times p_{y_i}\right)\right] p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) d\tilde{\boldsymbol{\lambda}}_i}{\int \left(p_{t_i} \times p_{y_i}\right) p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) d\tilde{\boldsymbol{\lambda}}_i} \approx \frac{\frac{1}{M}\sum_{m=1}^M \nabla_{a,\boldsymbol{\beta}} \left[p_t(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r}) \times p_y(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r})\right]}{\frac{1}{M}\sum_{m=1}^M \left[p_t(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r}) \times p_y(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r})\right]} \tag{13}$$

where M is a reasonably large number, say 10, $\tilde{\boldsymbol{\lambda}}_i^{(m)} = \{\tilde{\lambda}_{ijk}^{(m)}\}_{jk}$ and $\tilde{\lambda}_{ijk}^{(m)} \overset{iid}{\sim} \text{Gamma}(r_{jk}, 1)$, $\forall i = 1, \cdots, n$ and $m = 1, \cdots, M$. With the fact that $\nabla_{\boldsymbol{r}} p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) = p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) \nabla_{\boldsymbol{r}} \log p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r})$,

$$\begin{aligned}
\nabla_{\boldsymbol{r}} \log p_i &= \frac{\int \nabla_{\boldsymbol{r}} \left[\left(p_{t_i} \times p_{y_i}\right) p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r})\right] d\tilde{\boldsymbol{\lambda}}_i}{\int \left(p_{t_i} \times p_{y_i}\right) p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) d\tilde{\boldsymbol{\lambda}}_i} \\
&= \frac{\int \left(p_{t_i} \times p_{y_i}\right) \nabla_{\boldsymbol{r}} \log p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) d\tilde{\boldsymbol{\lambda}}_i}{\int \left(p_{t_i} \times p_{y_i}\right) p(\tilde{\boldsymbol{\lambda}}_i \,|\, \boldsymbol{r}) d\tilde{\boldsymbol{\lambda}}_i} \\
&\approx \frac{\frac{1}{M}\sum_{m=1}^M p_t(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r}) \times p_y(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r}) \nabla_{\boldsymbol{r}} \log p(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r})}{\frac{1}{M}\sum_{m=1}^M \left[p_t(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r}) \times p_y(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r})\right]} \\
&= \sum_{m=1}^M \frac{p_t(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r}) \times p_y(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r})}{\sum_{m'=1}^M \left[p_t(\tilde{\boldsymbol{\lambda}}_i^{(m')} \,|\, \boldsymbol{r}) \times p_y(\tilde{\boldsymbol{\lambda}}_i^{(m')} \,|\, \boldsymbol{r})\right]} \nabla_{\boldsymbol{r}} \log p(\tilde{\boldsymbol{\lambda}}_i^{(m)} \,|\, \boldsymbol{r}).
\end{aligned} \tag{14}$$

Therefore, we can approximate the derivatives of $\log P$ with respect to a, $\boldsymbol{\beta}$ and $\boldsymbol{r}$ by plugging in (13) and (14), respectively, and minimize $-\log P$ by (stochastic) gradient

descent.

D Experiment settings

We run $200,000$ interations of Gibbs sampler for WDR with the gamma process truncated at $K = 10$ for all experiments on the Prosper data, take the first $195,000$ as burn-in and estimate CIF by averaging its estimators from the last $5,000$ iterations. For the synthetic data analysis, we run $100,000$ iterations and collect the last 2000 MCMC samples. For all experiments of random survial forests, we set the number of trees equal to 100 and the number of splits equal to 2 as suggested by Ishwaran et al. [2014]. We use `R` to implement the FG model by package `cmprsk` Gray [2014] and the RF model by `randomForestSRC` Ishwaran and Kogalur [2018], and to calulate C-indices by package `pec` Gerds [2017].

We compare the classification of loan early payoff and default by FG, RF and WDR with three other models, including L_2 regularized logistic regression and a neural network. For the logistic regression, we use 5-fold cross validation to select the tuning parameter for the L_2 penalty of the parameters from $(2^{-10}, 2^{-9}, \ldots, 2^{10})$. The neural network has three fully connected hidden layers with the width as 150, 300, and 150, respectively, rectifier linear units as the activation function between hidden layers and a softmax output. For all the classification methods, an observation in the testing set is classified to the category associated with the predictive probability greater than 0.5.

E Additional experimental results

Table 8 to 11 show the Brier scores (mean $\pm$ standard error) for each event of the synthetic data 1 and 2, respectively. The model performance quantified by Brier score is basically consistent with that quantified by C-indices. For synthetic data 1 where covariates are linearly influential, the Brier scores are comparable among FG, RF and WDR. For synthetic data 2 that has nonlinear covariate effects, the Brier scores by WDR

and RF are smaller than those by FG. Note that the Brier scores by WDR is slightly larger than those by RF in terms of mean for synthetic data 2 when $t \leq 0.8$, but the difference is insignificant when the standard errors are taken into account.

Table 8: Brier score for event 1 of synthetic data 1.

	t=0.4	t=0.8	t=1.2	t=1.6	t=2.0
FG	0.090±0.008	**0.164±0.012**	**0.173±0.014**	**0.170±0.014**	0.172±0.013
RF	**0.089±0.010**	0.168±0.014	0.185±0.012	0.183±0.014	0.182±0.014
WDR	0.108±0.014	0.181±0.017	0.177±0.012	**0.170±0.013**	**0.170±0.012**

Table 9: Brier score for event 2 of synthetic data 1.

	t=0.4	t=0.8	t=1.2	t=1.6	t=2.0
FG	0.077±0.011	**0.158±0.014**	**0.169±0.013**	**0.166±0.013**	**0.167±0.013**
RF	**0.076±0.011**	0.165±0.014	0.178±0.015	0.175±0.014	0.176±0.013
WDR	0.086±0.015	0.173±0.018	0.174±0.012	0.167±0.013	**0.167±0.012**

Table 10: Brier score for event 1 of synthetic data 2.

	t=0.4	t=0.6	t=0.8	t=1.0	t=1.2
FG	0.115±0.017	0.208±0.011	0.250±0.002	0.248±0.003	0.243±0.004
RF	**0.110±0.014**	**0.178±0.015**	**0.198±0.012**	0.183±0.015	0.172±0.014
WDR	0.112±0.021	0.186±0.019	0.203±0.013	**0.174±0.007**	**0.169±0.006**

Table 11: Brier score for event 2 of synthetic data 2.

	t=0.4	t=0.6	t=0.8	t=1.0	t=1.2
FG	0.083±0.015	0.187±0.015	0.228±0.007	0.239±0.005	0.242±0.004
RF	**0.080±0.014**	**0.153±0.018**	0.167±0.018	0.165±0.016	0.166±0.015
WDR	0.084±0.017	0.164±0.023	**0.162±0.016**	**0.161±0.015**	**0.165±0.013**

Table 12 and 13, respectively, provide Brier scores and C-indices at month $5, 10, \cdots, 35$ by FG, RF and WDR for the Prosper data analysis using all the features in Table 2. We can see the three models deliver comparable Brier scores and C-indices at the time evaluated. Table 14 shows the point estimate ± standard error of the Prosper feature effects by WDR ($\boldsymbol{\beta}/a$) and FG ($\boldsymbol{\theta}$). All the point estimates by the two models are of the same sign unless insignificant, and this demonstrate the credibility of feature effect estimation by WDR.

Table 12: Brier scores for Prosper data

	Early payoff			Default		
time	FG	RF	WDR	FG	RF	WDR
5	0.0745	**0.0717**	0.0760	0.0110	**0.0106**	0.0172
10	0.1332	**0.1274**	0.1323	0.0765	**0.0731**	0.0744
15	0.1634	**0.1587**	0.1618	0.1275	0.1307	**0.1249**
20	0.1818	**0.1783**	0.1800	**0.1509**	0.1606	0.1534
25	0.1910	0.1909	**0.1889**	**0.1632**	0.1771	0.1694
30	0.1982	0.2001	**0.1951**	**0.1713**	0.1876	0.1802
35	0.2034	0.2052	**0.1962**	**0.1764**	0.1948	0.1864

Table 13: C-indices for Prosper data

	Early payoff			Default		
time	FG	RF	WDR	FG	RF	WDR
5	0.7151	**0.7205**	0.7091	0.7648	0.8406	**0.8424**
10	0.7168	**0.7327**	0.7290	0.7795	0.7864	**0.8115**
15	0.6920	0.7257	**0.7296**	**0.7874**	0.7542	0.7860
20	0.7090	0.7202	**0.7268**	0.7576	0.7419	**0.7756**
25	0.7140	0.7133	**0.7279**	**0.7745**	0.7284	0.7729
30	0.7158	0.7049	**0.7229**	0.7920	0.7201	**0.7945**
35	0.7031	0.6977	**0.7199**	**0.7539**	0.7146	0.7504

Table 14: Estimates of Prosper feature effects by WDR ($\boldsymbol{\beta}/a$) and FG ($\boldsymbol{\theta}$)

	WDR ($\boldsymbol{\beta}/a$)		FG ($\boldsymbol{\theta}$)	
feature	Early payoff	Default	Early payoff	Default
DebtToIncomeRatio	-0.039±0.030	0.012±0.029	-0.024±0.014	0.012±0.011
BankcardUtilization	-0.172±0.071	-0.236±0.082	-0.134±0.030	-0.129±0.033
AmountDelinquent	-0.005±0.004	0.007±0.004	-0.003±0.002	0.003±0.002
DelinquenciesLast7Years	-0.012±0.002	0.002±0.002	-0.010±0.001	0.002±0.001
InquiriesLast6Months	-0.074±0.008	0.076±0.006	-0.049±0.004	0.034±0.003
PublicRecordsLast10Years	-0.172±0.029	0.104±0.027	-0.126±0.015	0.060±0.009
PublicRecordsLast12Months	-0.062±0.13	-0.098±0.128	-0.055±0.057	-0.011±0.044
CurrentCreditLines	0.015±0.013	0.021±0.016	0.010±0.005	0.003±0.006
OpenCreditLines	-0.028±0.014	-0.033±0.018	-0.021±0.006	-0.007±0.007
TotalCreditLines	0.004±0.002	0.006±0.002	0.004±0.001	0.001±0.001
RevolvingCreditBalance	-0.001±0.001	0.001±0.001	-0.001±0.001	0.001±0.001
Groupmember	0.015±0.054	-0.038±0.067	0.032±0.022	-0.044±0.024
IsBorrowerHomeowner	-0.132±0.045	0.22±0.057	-0.101±0.020	0.110±0.022
OrderOfListing	-0.178±0.011	0.872±0.028	-0.089±0.006	0.510±0.023
TotalListingOfBorrowers	0.178±0.008	-0.848±0.027	0.095±0.005	-0.498±0.023
AmountRequested	-0.047±0.005	0.078±0.006	-0.035±0.002	0.035±0.002
FundingOption	0.351±0.064	-0.221±0.074	0.268±0.027	-0.159±0.027
BorrowerMaximumRate	-0.026±0.006	0.027±0.009	-0.025±0.003	0.022±0.003
ContractInterest	-0.002±0.007	0.064±0.009	0.004±0.003	0.026±0.003
Duration	0.038±0.009	0.017±0.012	0.017±0.004	0.001±0.004
Images	-0.042±0.044	-0.182±0.057	-0.037±0.018	-0.051±0.022
CreditGrade:AA	3.491±0.593	-3.596±0.378	2.199±0.288	-1.958±0.138
CreditGrade:A	3.225±0.588	-2.771±0.363	2.020±0.287	-1.556±0.133
CreditGrade:B	3.027±0.582	-2.222±0.352	1.849±0.287	-1.283±0.128
CreditGrade:C	2.910±0.584	-2.012±0.343	1.743±0.286	-1.190±0.125
CreditGrade:D	2.953±0.579	-1.929±0.34	1.795±0.285	-1.180±0.123
CreditGrade:E	2.883±0.580	-1.929±0.338	1.704±0.286	-1.158±0.123
CreditGrade:HR	2.737±0.582	-1.584±0.337	1.609±0.285	-1.012±0.123
Income:1-25k	-0.057±0.176	0.166±0.206	-0.001±0.089	0.074±0.082
Income:25k-50k	0.009±0.164	0.150±0.190	0.064±0.085	0.084±0.076
Income:50k-75k	0.260±0.165	-0.202±0.193	0.270±0.086	-0.114±0.077
Income:75k-100k	0.379±0.170	-0.328±0.202	0.387±0.087	-0.195±0.081
Income:>100k	0.546±0.171	-0.573±0.205	0.511±0.088	-0.323±0.082
Retired	-0.420±0.235	0.473±0.286	-0.380±0.111	0.306±0.113
Full-time	-0.225±0.180	0.026±0.211	-0.251±0.090	0.098±0.084
Self-employed	-0.143±0.189	0.414±0.218	-0.241±0.091	0.192±0.088
Part-time	-0.178±0.219	-0.332±0.277	-0.213±0.103	-0.066±0.110
NotEmployed	0.171±0.235	0.410±0.291	0.020±0.103	0.155±0.125
OtherEmployment	0.711±0.194	0.889±0.231	0.379±0.096	0.266±0.092

Table 11 continued: Estimates of Prosper feature effects by WDR ($\boldsymbol{\beta}/a$) and FG ($\boldsymbol{\theta}$)

	WDR		FG	
feature	Early payoff	Default	Early payoff	Default
DebtConsolidation	0.147±0.068	-0.838±0.087	0.086±0.027	-0.379±0.034
HomeImprovement	0.208±0.109	-0.783±0.148	0.130±0.044	-0.380±0.060
Business	0.026±0.092	-0.410±0.113	-0.045±0.039	-0.182±0.046
PersonalLoan	-0.034±0.096	-0.566±0.123	-0.119±0.042	-0.220±0.047
StudentUse	0.197±0.161	-0.887±0.228	0.112±0.068	-0.480±0.089
Auto	0.359±0.138	-1.215±0.204	0.262±0.057	-0.659±0.086
OtherCategory	0.321±0.085	-1.020±0.117	0.205±0.035	-0.532±0.047
Baby&Adoption	-5.469±0.654	0.193±2.178	-6.542±0.709	1.733±0.461
Boat	0.486±1.709	-0.613±2.795	0.608±0.913	0.005±0.988
Cosmetic	0.934±1.442	-2.604±2.855	0.779±0.691	-1.220±1.107
EngagementRing	1.749±1.172	-2.768±2.845	1.455±0.540	-1.676±1.095
GreenLoans	-2.227±3.274	0.510±2.045	-1.260±1.248	0.608±1.061
HouseholdExpenses	0.546±0.374	-0.006±0.375	0.023±0.186	-0.063±0.204
LargePurchase	1.059±0.622	-0.447±0.805	0.411±0.348	-0.389±0.378
Medical/Dental	-0.056±0.488	0.368±0.479	-0.264±0.233	0.449±0.254
Motorcycle	1.406±0.836	-2.197±1.880	1.084±0.401	-1.512±0.778
RV	1.902±1.274	-3.347±2.655	1.245±0.805	-1.882±1.104
Taxes	0.976±0.520	-0.736±0.745	0.589±0.283	-0.563±0.356
Vacation	0.956±0.670	-0.837±0.939	0.280±0.379	-0.789±0.464
Wedding	0.698±0.826	-0.596±1.158	0.330±0.451	-0.495±0.539