



Research article

Promote sign consistency in cure rate model with Weibull lifetime

Chenlu Zheng^{1,2}, and Jianping Zhu^{1,2,*}

¹ School of Management, Xiamen University, Xiamen, China

² Data Mining Research Center, Xiamen University, Xiamen, China

* **Correspondence:** Email: xmjpzhu@xmu.edu.cn.

Abstract: In survival analysis, the cure rate model is widely adopted when a proportion of subjects have long-term survivors. The cure rate model is composed of two parts: the first part is the incident part which describes the probability of cure (infinity survival), and the second part is the latency part which describes the conditional survival of the uncured subjects (finite survival). In the standard cure rate model, there are no constraints on the relations between the coefficients in the two model parts. However, in practical applications, the two model parts are quite related. It is desirable that there may be some relations between the two sets of the coefficients corresponding to the same covariates. Existing works have considered incorporating a joint distribution or structural effect, which is too restrictive. In this paper, we consider a more flexible model that allows the two sets of covariates can be in different distributions and magnitudes. In many practical cases, it is hard to interpret the results when the two sets of the coefficients of the same covariates have conflicting signs. Therefore, we proposed a sign consistency cure rate model with a sign-based penalty to improve interpretability. To accommodate high-dimensional data, we adopt a group lasso penalty for variable selection. Simulations and a real data analysis demonstrate that the proposed method has competitive performance compared with alternative methods.

Keywords: sign consistency; cure rate model; survival analysis; regularization; Weibull distribution

Mathematics Subject Classification: 62N01, 62N02

1. Introduction

Survival analysis is a method to analyze time to event of interest assuming that all subjects will eventually experience the event [1]. It has been widely applied in many fields such as medicine, engineering, credit scoring, etc. [2–4]. However, in practice, a proportion of subjects may not

experience the event of interest. They are considered ‘cured’, or in other words, the long-term survivors. For instance, in medical practice, there may be a fraction of patients cured of their disease. Therefore, the cure rate model, an extension of survival analysis, is introduced to modeling data with long-term survivors [5].

Let \mathbf{x}_i be the covariate vector of subject i , and $S_i(t|\mathbf{x}_i)$ be the survival probability in time t . The cure rate model can be written as

$$S_i(t|\mathbf{x}_i) = \pi_i(\mathbf{x}_i) + (1 - \pi_i(\mathbf{x}_i))S_i(t|U_i = 1, \mathbf{x}_i), \quad (1.1)$$

where U_i is the random binary variable which indicates the cure status of subject i . The cure status $U_i = 0$ denotes subject i is cured and will not experience the event, and $U_i = 1$ otherwise. Here $\pi_i(\mathbf{x}_i)$ is the probability function of being cured with the vector of regression coefficients $\boldsymbol{\alpha}$, and $S_i(t|U_i = 1, \mathbf{x}_i)$ is the survival function with the vector of regression coefficients $\boldsymbol{\beta}$.

As shown in (1.1), the cure rate model is composed of two parts. The first part is the incident part, which predicts the probability of being cured. The second part is the latency part, which describes the conditional survival of the uncured subjects. Since the cure rate model can predict whether subjects are cured and the time to event of the uncured subjects, it is commonly adopted when a proportion of subjects have long-term survivors [6].

There are numerous studies in the literature regarding many extensions of the cure rate model. Cooner et al. [7] proposed a flexible hierarchical cure rate model to distinguish among underlying mechanisms that lead to cure. Rodrigues et al. [8] assumed the number of risk factors to follow the Conway–Maxwell Poisson distribution and proposed a Conway–Maxwell Poisson cure rate model to unify some cure rate models. Li et al. [9] considered a mixture of linear dependent tail-free processes as the prior for the distribution of the cure rate parameter to develop a latent promotion time cure rate model. Results showed that the cure rate model incorporated penalized spline has better performance in prediction [10]. Georgiana proposed a Bayesian spatial cure rate model with Weibull lifetime to model spatial variability in the censoring mechanism [11]. Pal et al. [12] proposed a projected non-linear conjugate gradient algorithm for the cure rate model under a competing risks scenario. Besides, many works have developed semi-parametric and nonparametric methods to investigate the effects of covariates on the outcome. For instance, Li et al. [13] proposed a semi-parametric additive predictor consisting of a sum of linear and nonparametric terms in the incident part. Chen et al. [14] modeled the covariate effects by a nonparametric form.

However, many existing works tend to pay much less attention to the relations between the vector of regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the two model parts. Generally, they assume that there are no direct constraints on the relation of the vector of regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ corresponding to the same covariates [15,16]. In other words, these works assume that the probability of being cured and the time to event are independent, and there are no direct constraints between the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Notably, in practical applications, the two model parts are quite related. The incident part describes the probability of cure (infinity survival), and the latency part describes the conditional survival of the uncured subjects (finite survival). It is desirable that there may be some connections between the vector of regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ corresponding to the same covariates. Theoretical derivations and case studies also suggest that relaxing the assumption of no direct

constraints on regression coefficients can improve the performance of the model [17,18]. Liu et al. [19] relax the assumption of no direct constraints by establishing a joint distribution of the covariates and the logarithm of the hazard rate. Fan et al. [20] incorporate structural effects of α and β in the cure rate model to improve the estimation accuracy and interpretability. A joint distribution or structural effects of α and β are crude yet effective way to impose the relations between the coefficients. However, the two parts of the model still describe two different aspects. The assumption of a joint distribution or structural effects are too restrictive constraints and might not work well. So we consider a more flexible model that allows the coefficients α and β can be in different distributions and magnitudes. Sign consistency penalty is proposed to promote the similarity in sign to get more interpretable results by Zhang [21]. In many practical cases, it is hard to interpret the results when coefficients α and β corresponding to the same covariates have conflicting signs. In this paper, we consider a sign consistency cure rate model with a sign-based penalty. In addition, the models may suffer from bad performance due to the high dimension of the data, and grouping structure arises naturally in many practical cases. A group lasso penalty is also imposed for group variable selection.

In this paper, we propose a cure rate model with group selection and sign consistency (CRGS), which can select important groups of covariates and promote similarity in the sign of coefficients. The proposed method can promote the similarity in the signs of coefficients α and β to improve interpretability. Compared to the individual variable selection approaches such as the sign consistency method in [22], the proposed method with the group selection approach takes into consideration the grouping structure and can lead to better prediction. Compared with previously employed approaches such as joint distribution or structural effects of coefficients, the CRGS method can avoid too strict constraints between the coefficients and lead to more consistency and hence more interpretable results.

The paper is organized as follows. In section 2, the sign consistency cure rate model with Weibull lifetime and the algorithm is introduced. Simulation is presented in section 3. Section 4 displays a real data application. Finally, section 5 discusses the conclusions.

2. Methods

2.1. Sign consistency cure rate model with Weibull lifetime

Consider data with n subjects and p covariates. Denote Y_i as the time to event of subject i , that is, the time until the event of interest is observed to occur. Let C_i be the time of right censoring, and $\delta_i = I(Y_i < C_i)$ be the censoring indicator of subject i , where $\delta_i = 0$ for censored and $\delta_i = 1$ for uncensored. Denote $y_i = \min(T_i, C_i)$. Note that the censored subjects include the cured subjects and the uncured subjects for whom the event has not occurred at censoring time. So the cure status U_i is unobservable. The observable data is $\{(y_i, \delta_i, \mathbf{x}_i), i = 1, \dots, n\}$.

Denote \mathbf{x} as the covariate vector with grouping structure. Let $\mathbf{x} = (\mathbf{x}_1^T, \dots, \mathbf{x}_J^T)^T$ be the covariate vector with J nonoverlapping subgroups. $\mathbf{x}_j = (x_{j1}, \dots, x_{jp_j})^T$ is the j -th subgroup of covariate vector, with $\sum_{j=1}^J p_j = p$. Grouping structure arises naturally in many practical cases. Examples include the expression of a multi-level factor by a group of dummy covariates and the expression of an addictive model by several basis functions [23]. In addition, grouping structure can

also be introduced into a model by taking advantage of prior knowledge [24]. For example, genes belong to the same biological pathway [25].

In the incident part of the cure rate model, we adopted logistic regression, a generalized linear model, to describe the probability of cure $\pi_i(\mathbf{x}_i) = 1 / (1 + \exp(\alpha_0 + \mathbf{x}_i^T \boldsymbol{\alpha}))$. Here $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_j^T)^T$ is the vector of regression coefficients, $\boldsymbol{\alpha}_j = (\alpha_{j1}, \dots, \alpha_{jp_j})^T$ is the j -th subgroup of the coefficient vector, and α_0 is the intercept.

In the latency part, let λ_i be the link function. The survival function is the chance an individual survives to time t given the individual will eventually experience the event of interest, while λ_i is the probability of an individual experiencing the event in the next instant of time t .

We assume the time to event t follow the Weibull distribution. Weibull distribution is a considerable flexibility distribution to model lifetime data [26]. It had been justified to be a valid lifetime distribution within the broad family of generalized gamma models [27,28]. Referring to [1,29], the survival function for the uncured subjects following the Weibull distribution can be written as

$$S_i(t | U_i = 1) = \exp\left(-(\lambda_i t)^r\right), \quad (2.1)$$

with the probability density function $f(t | U_i = 1) = (r/t)(\lambda_i t)^{r-1} \exp\left(-(\lambda_i t)^r\right)$. Here $r > 0$ is the shape parameter (more discussions below). The link function can be written as

$$\lambda_i = \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}), \quad (2.2)$$

where β_0 is the intercept, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_j^T)^T$ is the vector of regression coefficients, and the j -th subgroup of the coefficient vector is $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp_j})^T$.

For observable data $\{(y_i, \delta_i, \mathbf{x}_i), i = 1, \dots, n\}$, the log-likelihood function is

$$\begin{aligned} L(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta}) = & \sum_{i=1}^n \left(\delta_i \left(\log(1 - \pi_i) + \log(r) - \log(t) + r \log(\lambda_i y_i) - (\lambda_i y_i)^r \right) \right) \\ & + \sum_{i=1}^n \left((1 - \delta_i) \log \left((1 - \pi_i) \exp\left(-(\lambda_i y_i)^r\right) + \pi_i \right) \right). \end{aligned} \quad (2.3)$$

For promoting sign consistency and variable selection, we propose the following objective function

$$Q(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta}) = -L(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta}) + P_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) + P_2(\boldsymbol{\alpha}, \boldsymbol{\beta}). \quad (2.4)$$

Here $P_1(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the group variable selection penalty function, and $P_2(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the sign consistency penalty function as follows.

$$P_2(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\mu_2}{2} \sum_{j=1}^J \sum_{k=1}^{p_j} (\text{sign}(\alpha_{jk}) - \text{sign}(\beta_{jk}))^2, \quad P_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mu_1 \sum_{j=1}^J \sqrt{p_j} (\|\boldsymbol{\alpha}_j\| + \|\boldsymbol{\beta}_j\|), \quad (2.5)$$

where $\|\bullet\|$ is the l_2 norm, $\text{sign}(\bullet)$ is the sign function, $\mu_1 > 0$ and $\mu_2 > 0$ are tuning parameters. μ_1 and μ_2 control the degree of penalty.

The first penalty $P_1(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is a group lasso penalty, which can conduct group variable selection and generate more accurate estimation. Group lasso has the advantage in group variable selection performance and is commonly used in literature. It has been justified that the group lasso is more robust to noise compared with lasso when the underlying signal is strongly group-sparse [30]. In addition, different from methods in some existing works [20,22], group lasso can consider grouping structures and it has better performance compared with the group LARS and the group non-negative garrotte [31]. The second penalty $P_2(\boldsymbol{\alpha}, \boldsymbol{\beta})$ promotes the similarity in signs of coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which can lead to more interpretable results. The proposed method is more flexible than the methods with structural effect in [20] or joint distribution in [19] of coefficients.

2.2. EGCD algorithm

In this section, we propose the Expectation Group Coordinate Descent (EGCD) algorithm to optimize the objective function. In the E-step, we introduce a latent unobserved U_i to obtain a complete log-likelihood function. In the GCD-step, group coordinate descent is adopted to iteratively update a subgroup of parameters with the remaining parameters fixed at their most recent values. Since the sign function $\text{sign}(\bullet)$ is not differentiable and continuous, and it is hard to optimize. So we introduce the following approximation for computational feasibility [21,22].

$$(\text{sign}(\alpha_{jk}) - \text{sign}(\beta_{jk}))^2 \approx \left(\frac{\alpha_{jk}}{|\alpha_{jk}| + \tau} - \frac{\beta_{jk}}{|\beta_{jk}| + \tau} \right)^2, \quad (2.6)$$

with τ is a small positive constant.

The EGCD algorithm iteratively updates α_0 , $\boldsymbol{\alpha}$, β_0 , and $\boldsymbol{\beta}$ in the m -th iteration.

2.2.1. E-step

In the E-step of the m -th iteration, let the observation of the latent U_i be $U_i^{[m]}$. Consider the complete data $\{(y_i, \delta_i, U_i^{[m]}, \mathbf{x}_i), i = 1, \dots, n\}$ with the latent $U_i^{[m]}$. The complete log-likelihood is

$$\begin{aligned} L^{[m]} &= \sum_{i=1}^n \left((1 - U_i^{[m]}) \log(\pi_i^{[m]}) + U_i^{[m]} \log(1 - \pi_i^{[m]}) \right) + \sum_{i=1}^n \left(\delta_i \lambda_i^{[m]} - U_i^{[m]} \left(y_i \exp \lambda_i^{[m]} \right)^r \right) \\ &:= L_1^{[m]} + L_2^{[m]}, \end{aligned} \quad (2.7)$$

where

$$\pi_i^{[m]} = \frac{1}{1 + \exp(\alpha_0^{[m]} + \mathbf{x}_i^T \boldsymbol{\alpha}^{[m]})}, \quad (2.8)$$

$$\lambda_i^{[m]} = \exp(\beta_0^{[m]} + \mathbf{x}_i^T \boldsymbol{\beta}^{[m]}). \quad (2.9)$$

Regarding the expectation of U_i , there are three possible situations of subjects. (a) $\delta_i=0$ and $U_i=0$: censored and cured, indicates long-term survivors; (b) $\delta_i=0$ and $U_i=1$: censored and uncured, indicates subjects who will eventually experience the event, and the event have not occurred in censoring time C_i ; (c) $\delta_i=1$ and $U_i=1$: uncensored and uncured, indicates subjects who have experienced the event. Therefore, the expectation of U_i is 1 when uncensored ($\delta_i=1$). When censored ($\delta_i=0$), the expectation of U_i is related to the probability of cure and the proportion that uncured subjects for whom the event has not occurred at time t . Denote the expectation of $U_i^{[m]}$ as $u_i^{[m]}$,

$$u_i^{[m]} = \mathbb{E}(U_i^{[m]}) = \begin{cases} 1 & \delta_i = 1 \\ \frac{(1 - \pi_i^{[m]}) \exp(-(\lambda_i^{[m]} y_i)^r)}{\pi_i^{[m]} + (1 - \pi_i^{[m]}) \exp(-(\lambda_i^{[m]} y_i)^r)} & \delta_i = 0. \end{cases} \quad (2.10)$$

Given the complete data $\{(y_i, \delta_i, U_i^{[m]}, \mathbf{x}_i), i = 1, \dots, n\}$, we take the expectation of $L^{[m]}$ in (2.7) with respect to $u_i^{[m]}$

$$\mathbb{E}(L^{[m]}) = l_1^{[m]} + l_2^{[m]}, \quad (2.11)$$

where

$$l_1^{[m]} = \mathbb{E}(L_1^{[m]}) = \sum_{i=1}^n \left((1 - u_i^{[m]}) \log(\pi_i^{[m]}) + u_i^{[m]} \log(1 - \pi_i^{[m]}) \right), \quad (2.12)$$

$$l_2^{[m]} = \mathbb{E}(L_2^{[m]}) = \sum_{i=1}^n \left(\delta_i (\beta_0^{[m]} + \mathbf{x}_i^T \boldsymbol{\beta}^{[m]}) - u_i^{[m]} \left(y_i \exp(\beta_0^{[m]} + \mathbf{x}_i^T \boldsymbol{\beta}^{[m]}) \right)^r \right). \quad (2.13)$$

The objective function can be written as

$$-(l_1 + l_2) + \mu_1 \sum_{j=1}^J \sqrt{p_j} (\|\boldsymbol{\alpha}_j\| + \|\boldsymbol{\beta}_j\|) + \frac{\mu_2}{2} \sum_{j=1}^J \sum_{k=1}^{p_j} \left(\frac{\alpha_{jk}}{|\alpha_{jk}| + \tau} - \frac{\beta_{jk}}{|\beta_{jk}| + \tau} \right)^2. \quad (2.14)$$

2.2.2. GCD-step

In the GCD-step, group coordinate descent is adopted to iteratively update α_0 , $\boldsymbol{\alpha}$, β_0 , and $\boldsymbol{\beta}$. Group coordinate descent algorithms optimize the objective function with respect to a group of

parameters at a time, and iteratively cycles through the parameter groups until convergence [25]. The intercept $\alpha_0^{[m+1]}$ is updated by

$$\alpha_0^{[m+1]} = \alpha_0^{[m]} - \left(\frac{\partial^2 l_1^{[m]}}{\partial (\alpha_0^{[m]})^2} \right)^{-1} \frac{\partial l_1^{[m]}}{\partial \alpha_0^{[m]}}, \quad (2.15)$$

where $\partial l_1^{[m]} / \partial \alpha_0^{[m]} = \sum_{i=1}^n (\pi_i^{[m]} + u_i^{[m]} - 1)$, and $\partial^2 l_1^{[m]} / \partial (\alpha_0^{[m]})^2 = \sum_{i=1}^n \pi_i^{[m]} (\pi_i^{[m]} - 1)$.

For $\alpha_j^{[m+1]} \in R^{p_j}$, we can obtain Taylor's quadratic expansion of the objective function in (2.14) respect to $\alpha_j^{[m+1]}$. Referring to the fast unified algorithm for group lasso [32], the upper bound of the objective function can be written as

$$\frac{M_{1j}^{[m]}}{2} (\alpha_j^{[m+1]} - \alpha_j^{[m]})^T (\alpha_j^{[m+1]} - \alpha_j^{[m]}) - (\alpha_j^{[m+1]} - \alpha_j^{[m]})^T \left(\frac{\partial l_1^{[m]}}{\partial \alpha_j^{[m]}} + \mu_2 \mathbf{V}_{1j}^{[m]} \right) + \mu_1 \sqrt{p_j} \|\alpha_j^{[m+1]}\|, \quad (2.16)$$

where $\partial l_1^{[m]} / \partial \alpha_j^{[m]} = \sum_{i=1}^n (\pi_i^{[m]} + u_i^{[m]} - 1) x_{jk}$. Here $\mathbf{V}_{1j}^{[m]}$ is p_j -length vector, and $M_{1j}^{[m]}$ is a constant as follows.

$$\mathbf{V}_{1j}^{[m]} = \left(\frac{1}{|\alpha_{jk}^{[m]}| + \tau} \left(\frac{\beta_{jk}^{[m]}}{|\beta_{jk}^{[m]}| + \tau} - \frac{\alpha_{jk}^{[m]}}{|\alpha_{jk}^{[m]}| + \tau} \right) \right)_{1 \leq k \leq p_j}, \quad (2.17)$$

$$M_{1j}^{[m]} = \psi \left(- \left(\frac{\partial^2 l_1^{[m]}}{\partial \alpha_{jk_1}^{[m]} \partial \alpha_{jk_2}^{[m]}} \right)_{1 \leq k_1, k_2 \leq p_j} \right) + \max_k \left(\mu_2 \left(\frac{1}{|\alpha_{jk}^{[m]}| + \tau} \right)^2 \right),$$

where $\partial^2 l_1^{[m]} / \partial \alpha_{jk_1}^{[m]} \partial \alpha_{jk_2}^{[m]} = \sum_{i=1}^n (1 - \pi_i^{[m]}) \pi_i^{[m]} x_{jk}^2$, and $\psi(\bullet)$ is the maximum eigenvalue function. By minimizing (2.16), we obtain $\alpha_j^{[m+1]}$:

$$\alpha_j^{[m+1]} = \frac{M_{1j}^{[m]} \alpha_j^{[m]} + \partial l_1^{[m]} / \partial \alpha_j^{[m]} + \mu_2 \mathbf{V}_{1j}^{[m]}}{M_{1j}^{[m]}} \left(1 - \frac{\mu_1 \sqrt{p_j}}{|M_{1j}^{[m]} \alpha_j^{[m]} + \partial l_1^{[m]} / \partial \alpha_j^{[m]} + \mu_2 \mathbf{V}_{1j}^{[m]}|} \right)_+, \quad (2.19)$$

where $(a)_+ = \max\{a, 0\}$.

Similarly, the intercept $\beta_0^{[m+1]}$ is updated by

$$\beta_0^{[m+1]} = \beta_0^{[m]} - \left(\frac{\partial^2 l_2^{[m]}}{\partial (\beta_0^{[m]})^2} \right)^{-1} \frac{\partial l_2^{[m]}}{\partial \beta_0^{[m]}}, \quad (2.20)$$

where $\partial l_2^{[m]} / \partial \beta_0^{[m]} = \sum_{i=1}^n (\delta_i - u_i^{[m]} r (y_i \lambda_i^{[m]})^r)$, and $\partial^2 l_2^{[m]} / \partial (\beta_0^{[m]})^2 = - \sum_{i=1}^n (u_i^{[m]} r^2 (y_i \lambda_i^{[m]})^r)$.

For $\beta_j^{[m+1]} \in R^{p_j}$, consider the optimization function

$$\frac{M_{2j}^{[m]}}{2} (\boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]})^T (\boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]}) - (\boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]})^T \left(\frac{\partial l_2^{[m]}}{\partial \boldsymbol{\beta}_j^{[m]}} + \mu_2 \mathbf{V}_{2j}^{[m]} \right) + \mu_1 \sqrt{p_j} \|\boldsymbol{\beta}_j^{[m+1]}\|, \quad (2.21)$$

where $\partial l_2^{[m]} / \partial \boldsymbol{\beta}_j^{[m]} = \sum_{i=1}^n \left(\delta_i x_{jk} - u_i^{[m]} r x_{jk} (y_i \lambda_i^{[m]})^r \right)$.

Here $\mathbf{V}_{2j}^{[m]}$ is p_j -length vector, and $M_{2j}^{[m]}$ is a constant as follows.

$$\mathbf{V}_{2j}^{[m]} = \left(\frac{1}{|\boldsymbol{\beta}_{jk}^{[m]}| + \tau} \left(\frac{\boldsymbol{\alpha}_{jk}^{[m]}}{|\boldsymbol{\alpha}_{jk}^{[m]}| + \tau} - \frac{\boldsymbol{\beta}_{jk}^{[m]}}{|\boldsymbol{\beta}_{jk}^{[m]}| + \tau} \right) \right)_{1 \leq k \leq p_j}, \quad (2.22)$$

$$M_{2j}^{[m]} = \psi \left(- \left(\frac{\partial^2 l_2^{[m]}}{\partial \boldsymbol{\beta}_{jk_1}^{[m]} \partial \boldsymbol{\beta}_{jk_2}^{[m]}} \right)_{1 \leq k_1, k_2 \leq p_j} \right) + \max_k \left(\mu_2 \left(\frac{1}{|\boldsymbol{\beta}_{jk}^{[m]}| + \tau} \right)^2 \right),$$

where $\partial^2 l_2^{[m]} / \partial \boldsymbol{\beta}_{jk_1}^{[m]} \partial \boldsymbol{\beta}_{jk_2}^{[m]} = - \sum_{i=1}^n u_i^{[m]} (r x_{jk})^2 (y_i \lambda_i^{[m]})^r$.

By minimizing (2.21), we obtain $\boldsymbol{\beta}_j^{[m+1]}$:

$$\boldsymbol{\beta}_j^{[m+1]} = \frac{M_{2j}^{[m]} \boldsymbol{\beta}_j^{[m]} + \partial l_2^{[m]} / \partial \boldsymbol{\beta}_j^{[m]} + \mu_2 \mathbf{V}_{2j}^{[m]}}{M_{2j}^{[m]}} \left(1 - \frac{\mu_1 \sqrt{p_j}}{\left| M_{2j}^{[m]} \boldsymbol{\beta}_j^{[m]} + \partial l_2^{[m]} / \partial \boldsymbol{\beta}_j^{[m]} + \mu_2 \mathbf{V}_{2j}^{[m]} \right|_+} \right). \quad (2.23)$$

Regarding the parameters μ_1, μ_2 , and r , Wang et al [33] had demonstrated that the tuning parameters selected by the Bayesian information criterion (BIC) type criterion can identify the true model consistently as long as the covariate dimension is fixed. So the parameters, μ_1, μ_2 , and r , are selected by Bayesian information criterion (BIC). The EGCD algorithm is shown in Table 1.

Table 1. EGCD algorithm.

Expectation Group Coordinate Descent Algorithm	
1.	Initialize $m = 0$, $\boldsymbol{\alpha}_0^{[m]} = \boldsymbol{\beta}_0^{[m]} = 0$, and $\boldsymbol{\alpha}^{[m]} = \boldsymbol{\beta}^{[m]} = 0_{J \times p}$;
2.	Repeat the following updates: <ul style="list-style-type: none"> 2.1 E-step: <ul style="list-style-type: none"> Update $\pi_i^{[m]}$ and from (2.8), $\lambda_i^{[m]}$ from (2.9), and $u_i^{[m]}$ from (2.10); 2.2 GCD-step: <ul style="list-style-type: none"> Update $\boldsymbol{\alpha}_0^{[m+1]}$ from (2.15); For $j = 1, \dots, p$, update $\boldsymbol{\alpha}_j^{[m+1]}$ from (2.19); Update $\boldsymbol{\beta}_0^{[m+1]}$ from (2.20); For $j = 1, \dots, p$, update $\boldsymbol{\beta}_j^{[m+1]}$ from (2.23);
	$m = m + 1$
	Until $\max \left\{ \left\ \boldsymbol{\alpha}_j^{[m+1]} - \boldsymbol{\alpha}_j^{[m]} \right\ , \left\ \boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]} \right\ \right\} \leq 5 \times 10^{-3}$.

3. Simulation

In this section, we perform a numerical study to evaluate our method, in terms of both variable selection and estimation performance. The variable selection is assessed in terms of the (1) true positive rate (TPR), and (2) false positive rate (FPR) of variable selection. Estimation is evaluated in terms of mean square error (MSE) of coefficient estimates.

3.1. Design and assessment

In Scenario 1, the covariates x_{jk} , $j=1, \dots, J$, $k=1, \dots, p_j$, are generated from a multivariate normal distribution. The correlation coefficient of covariates x_{jk_m} and x_{jk_n} in the same group is $\rho = 0.1^{|k_m - k_n|}$, whereas $\rho = 0$ when in different groups. We consider the case with discrete covariates in Scenario 2. In Scenario 2, x_{jk} is defined as follows.

$$x_{jk} = \begin{cases} x_{jk} & j \leq J/3 \\ \mathbf{I}(x_{jk} > 0) & j > J/3. \end{cases} \quad (3.1)$$

The sample size is $n=500$. We consider low-dimension data with $p=40$, and high-dimension data with $p=200$. The censoring time is generated from a Weibull distribution with the shape parameter $r = \{0.25, 2.5\}$. We compare the proposed method (CRGS) with three alternative methods. The alternatives are the standard cure rate model without sign consistency and variable selection penalty (CR), the cure rate model with sign consistency (CRS), and the cure rate model with group lasso penalty (CRG), respectively. For comparison, we also consider the alternatives with the logistic regression in the incident part and the Weibull distribution in the latency part. The grouping structure and coefficients of the 2 scenarios are generated as listed in Table 2.

Table 2. Grouping structure and coefficients in Scenarios 1 and 2.

	Scenario 1	Scenario 2
Non-zero subgroups (α)	$\underbrace{0.5, \dots, 0.5}_5, \underbrace{0.5, \dots, 0.5}_5, \underbrace{0.5, \dots, 0.5}_5, \underbrace{-0.4, \dots, -0.4}_5$	$\underbrace{0.5, \dots, 0.5}_{15}, \underbrace{-0.4, \dots, -0.4}_5$
Non-zero subgroups (β)	$\underbrace{0.1, \dots, 0.1}_5, \underbrace{0.1, \dots, 0.1}_5, \underbrace{0.1, \dots, 0.1}_5, \underbrace{-0.3, \dots, -0.3}_5$	$\underbrace{0.1, \dots, 0.1}_{15}, \underbrace{-0.3, \dots, -0.3}_5$
Covariates	Continuous	Discrete and continuous

Let $\theta \in \{\alpha, \beta\}$, and $\hat{\theta}$ be the estimation of θ . We evaluate variable selection performance in terms of $TPR(\theta)$ and $FPR(\theta)$:

$$TPR(\theta) = \frac{TP}{TP + FN}, \quad FPR(\theta) = \frac{FP}{TN + FP}, \quad (3.2)$$

where

$$TP = \sum_{j=1}^p \mathbf{I}(\theta_j \neq 0 \cap \hat{\theta}_j \neq 0), \quad TP + FN = \sum_{j=1}^p \mathbf{I}(\theta_j \neq 0), \quad (3.3)$$

$$FP = \sum_{j=1}^p \mathbf{I}(\theta_j = 0 \cap \hat{\theta}_j \neq 0), \quad TN + FP = \sum_{j=1}^p \mathbf{I}(\theta_j = 0).$$

Estimates are evaluated by $MSE(\boldsymbol{\theta})$:

$$MSE(\boldsymbol{\theta}) = \frac{\sum_{j=1}^p (\theta_j - \hat{\theta}_j)^2}{\sum_{j=1}^p (\theta_j)^2}. \quad (3.4)$$

3.2. Results

Tables 3 and 4 summarize the mean and standard deviation in parentheses of the MSEs, TPRs, and FPRs for Scenarios 1 and 2. Both the scenarios are repeated 100 times.

As shown in Tables 3 and 4, the MSEs of methods with group lasso penalty (the proposed and CRG) are smaller than the CRS and CR methods. The estimates of CRS and CR methods have increasing MSEs with higher dimensions. The results indicate that higher dimension leads to less efficient estimation, and group lasso penalty can improve the estimation performance. Comparing the TPRs and FPRs of the proposed and CRG methods, the proposed method has better performance in terms of variable selection. Compared with alternatives, the proposed method has the lowest MSEs. Results of simulation reveal that compared with alternatives, the proposed method can improve the performance in terms of variable selection as well as estimation.

Table 3. Results of Scenario 1.

r	p		α				β			
			CR	CRS	CRG	CRGS	CR	CRS	CRG	CRGS
0.25	40	MSE	24.40	2.62	1.75	1.51	9.36	15.24	4.08	3.49
			(17.50)	(0.38)	(0.18)	(0.17)	(4.01)	(4.06)	(4.36)	(4.38)
		TPR	-	-	0.98	0.96	-	-	0.49	0.94
			-	-	(0.09)	(0.07)	-	-	(0.22)	(0.10)
		FPR	-	-	0.08	0.06	-	-	0.15	0.06
			-	-	(0.11)	(0.10)	-	-	(0.15)	(0.10)
200	200	ME	35.91	10.35	1.71	1.49	773.53	74.24	3.99	3.09
			(37.71)	(3.23)	(0.11)	(0.07)	(1396.13)	(22.85)	(0.71)	(0.47)
		TPR	-	-	0.99	0.96	-	-	0.44	0.94
			-	-	(0.04)	(0.06)	-	-	(0.19)	(0.09)
		FPR	-	-	0.02	0.01	-	-	0.09	0.01
			-	-	(0.02)	(0.01)	-	-	(0.05)	(0.01)
2.5	40	ME	24.41	3.34	1.73	1.50	9.06	8.16	3.64	3.06
			(17.48)	(0.44)	(0.13)	(0.07)	(2.18)	(1.31)	(0.70)	(0.41)
		TPR	-	-	0.98	0.96	-	-	0.48	0.94
			-	-	(0.09)	(0.07)	-	-	(0.21)	(0.10)

Continued on next page

	FPR	-	-	0.07	0.05	-	-	0.14	0.05
		-	-	(0.07)	(0.02)	-	-	(0.13)	(0.02)
200	ME	50.98	3.66	24.71	1.54	48.03	14.03	5.17	3.92
		(11.07)	(0.51)	(27.75)	(0.06)	(7.82)	(2.23)	(2.91)	(0.52)
	TPR	-	-	0.52	0.98	-	-	0.65	0.98
		-	-	(0.50)	(0.06)	-	-	(0.35)	(0.06)
	FPR	-	-	0.01	0.01	-	-	0.06	0.01
		-	-	(0.01)	(0.00)	-	-	(0.06)	(0.00)

Note: in each cell, mean (standard deviation).

Table 4. Results of Scenario 2.

r	p		α				β			
			CR	CRS	CRG	CRGS	CR	CRS	CRG	CRGS
0.25	40	MSE	6.01	2.17	1.57	1.42	67.18	19.26	9.83	6.70
			(2.47)	(0.85)	(0.12)	(0.09)	(33.69)	(16.42)	(3.33)	(2.36)
		TPR	-	-	0.95	0.95	-	-	0.95	0.94
			-	-	(0.08)	(0.06)	-	-	(0.06)	(0.07)
		FPR	-	-	0.05	0.06	-	-	0.35	0.19
			-	-	(0.00)	(0.06)	-	-	(0.21)	(0.15)
	200	ME	58.02	9.78	1.37	1.26	187.33	74.62	8.96	6.12
			(58.60)	(12.42)	(0.11)	(0.08)	(199.25)	(99.61)	(3.20)	(2.19)
		TPR	-	-	0.93	0.94	-	-	0.87	0.90
			-	-	(0.15)	(0.07)	-	-	(0.18)	(0.08)
		FPR	-	-	0.01	0.05	-	-	0.25	0.24
			-	-	(0.00)	(0.02)	-	-	(0.05)	(0.05)
2.5	40	ME	48.42	2.94	2.46	2.04	9.14	6.60	7.46	4.55
			(24.54)	(0.33)	(0.14)	(0.11)	(2.63)	(0.67)	(0.31)	(0.38)
		TPR	-	-	1.00	0.99	-	-	0.41	0.99
			-	-	(0.02)	(0.02)	-	-	(0.17)	(0.02)
		FPR	-	-	0.50	0.11	-	-	0.07	0.11
			-	-	(0.16)	(0.07)	-	-	(0.08)	(0.07)
	200	ME	81.13	3.43	2.57	1.95	73.28	9.83	8.02	4.42
			(18.85)	(0.43)	(0.17)	(0.09)	(12.06)	(2.50)	(0.37)	(0.33)
		TPR	-	-	1.00	0.99	-	-	0.53	0.99
			-	-	(0.02)	(0.02)	-	-	(0.14)	(0.02)
		FPR	-	-	0.54	0.13	-	-	0.06	0.13
			-	-	(0.04)	(0.03)	-	-	(0.03)	(0.03)

Note: in each cell, mean (standard deviation).

4. Analysis of credit data in China

In this section, we apply the proposed method to credit data. The data comes from a retail business of a commercial bank in China. It contains 16 covariates of 1213 customers in a personal

loan from 2014 to 2019. The primary interest is to assess the credit risk of a credit loan and find the important covariates to predict the time to default of the credit loan customers. The mean observed time is 1.38 years with a standard deviation of 0.69. Customers with missing data of annual household income are removed from our analysis. After preprocessing, the covariates and their descriptions are summarized in Table 5. By transforming the multi-level covariates, we have 24 covariates in the credit model. Censoring time C_i is the interval between the value date and either default or the end of observation (June 1, 2019). Due to the different value dates of the loans, the censoring time vary from individual to individual. Customers whose time to event Y_i is longer than the censoring time C_i are censored ($\delta_i = 0$). In this data, 1201 out of 1213 customers are censored.

Table 5. Covariates and their descriptions.

Covariates	Descriptions
Interest rate	[0.037, 0.087]
Loan line	(0, 7000, 000)
Loan term	(0, +∞)
Business type	consumer durables, housing decoration loans, and other personal consumption loans
Entrusted payment	yes, no
Early repayment	yes, no
Age	[20, 70]
Gender	male, female
Education	master/doctor, bachelor, vocational education, high school and below
Medical insurance	yes, no
Housing status	self-purchasing (with a mortgage), self-purchasing (without a mortgage), others
Annual household income (RMB)	≤ 200,000 , 200,000-400,000 , 400,000-600,000 , ≥ 600,000
Employment	employed, others
Type of workplace	government organization/institution, firm, others
Occupation	managers, commercial and service workers, others
Professional title	advanced, intermediate, primary, no professional title

The data are randomly divided into the training set and test set by 7:3. The training data is used for fitting the model and the test data is used to verify the performance of the fitted model. The parameters μ_1 , μ_2 , and r are selected by BIC. Different from the simulations, the real coefficient is unknown for the real data. Therefore, in this section, we adopt the negative log-likelihood to evaluate the performance of the methods. The mean negative log-likelihood (standard error) of the proposed method is 25.91 (11.61), compared with 56.33 (258.73), 31.04 (27.34), and 26.17 (13.02) for the CR, CRS, and CRG method respectively. For stability, all the results are based on 100 duplicates. The results indicate that the proposed method has competitive prediction performance than alternatives.

The coefficients are estimated based on 100 duplicates. With the median estimates of coefficients, we can compute the probability of cure (non-default) $\pi_i(\mathbf{x}_i)$ for all the customers. We dichotomize $\pi_i(\mathbf{x}_i)$ at the median and get two different groups of customers. One group with lower

$\pi_i(\mathbf{x}_i)$ is denoted as “high risk”, whereas another with higher $\pi_i(\mathbf{x}_i)$ is denoted as “low risk”. Figure 1 presents the Kaplan-Meier curves of the survival of the customers belonging to different groups. Kaplan-Meier curve describes the change of the survival probability over time. It is commonly used in survival analysis, see Rodrigues et al [8], and Pal [34]. As indicated in Figure 1, the “low risk” group has higher survival probability than the “high risk” group.

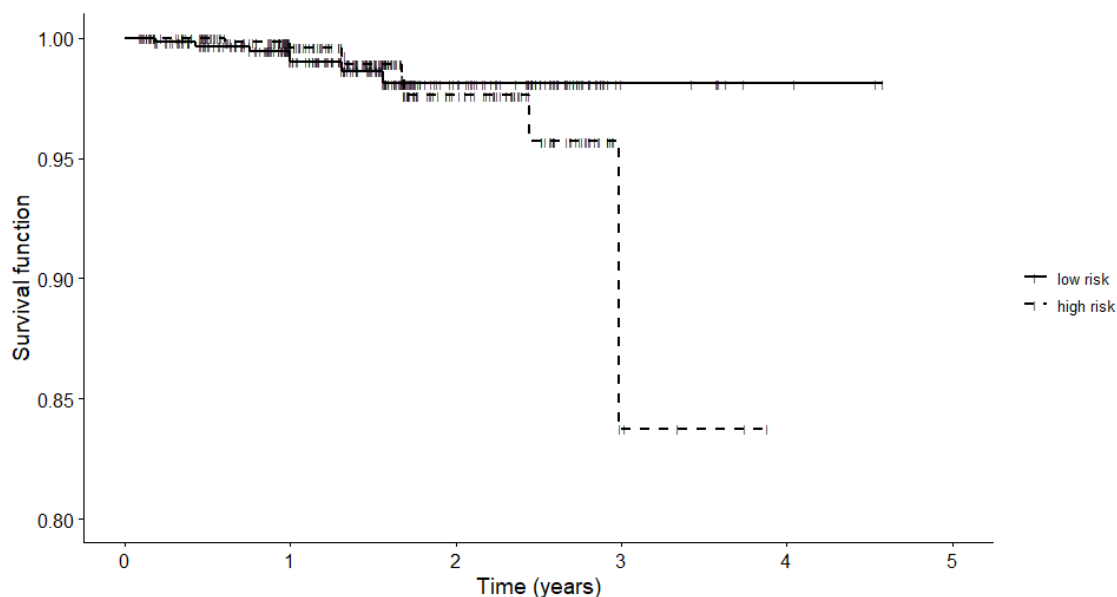


Figure 1. Kaplan-Meier curves stratified by different groups.

The median estimates of coefficients based on 100 duplicates are listed in Table 6. A positive coefficient α indicates that the covariate is positively related to the probability of default, and a positive coefficient β indicates that the covariate is negatively related to default time. Both the probability of default and default time are two quite relevant credit aspects. Customers with a higher probability of default are likely to default earlier. Compared with the alternative method, the signs of the α and β of the proposed method are promoted to be more consistent, whereas many covariates such as the housing status in the CR method have an opposite effect on the probability and time to default.

The coefficient results of the proposed method reveal that loan line, early repayment, gender, housing status, annual household income, employment status, type of workplace, and occupation are important covariates for credit risk assessment. The impact of business type and education on credit is not clear.

The loan line has a positive effect. One possible explanation is that customers with better credit status are more likely to obtain a higher loan line. Customers with housing loans are more likely to default. The increasing annual household income leads to better credit status. Employed customers are less likely to default. Compared with other employment groups such as self-employed, freelance, and unemployed, the employed group has a more stable income and is less likely to default. Customers who work in a government organization and institution have better credit status. Customers who are managers or commercial and service workers are less likely to default. Customers with early payment records tend to maintain good credit records and are less likely to default. Compared with women, men are more likely to default. This is consistent with the results of [35] and the personality characteristics of men’s risk preference [36].

Table 6. Estimates of coefficients.

	CR		CRS		CRG		CRGS	
	α	β	α	β	α	β	α	β
α_0 / β_0	-0.73	-9.15	-0.95	-7.89	-2.40	-5.82	-2.40	-5.82
Interest rate	0.00	4.68	0.00	0.00	0.00	0.00	0.00	0.00
Loan line	-0.33	-11.55	-1.45	-5.01	-1.47	-2.38	-1.47	-2.38
Loan term	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Business type								
Consumer durables	0.00	4.44	1.22	1.07	0.91	0.57	0.91	0.57
Housing decoration loans	-0.67	-5.12	-1.24	-2.77	-1.08	-1.27	-1.08	-1.27
Entrusted payment (yes)	0.00	-2.91	-0.18	-0.25	0.00	0.00	0.00	0.00
Early repayment (yes)	0.00	-28.92	-1.51	-7.29	-1.42	-3.78	-1.42	-3.78
Age	0.00	-0.10	0.00	-0.05	0.00	-0.07	0.00	-0.07
Gender (male)	0.71	9.50	1.16	6.04	0.98	4.96	0.98	4.96
Education								
Master/Doctor	-0.30	-3.13	-0.28	-0.37	0.00	0.00	0.00	0.00
Bachelor	0.77	-0.71	0.08	0.01	0.05	0.05	0.05	0.05
Vocational education	0.00	1.55	0.35	0.38	0.00	0.00	0.00	0.00
Medical insurance (yes)	0.24	4.18	1.02	1.42	0.91	0.72	0.91	0.72
Housing status								
Self-purchasing (with a mortgage)	0.59	-0.02	0.33	0.29	0.24	0.19	0.24	0.19
Self-purchasing (without a mortgage)	-0.75	1.40	0.00	0.00	0.00	0.00	0.00	0.00
Annual household income (RMB)								
200,000-400,000	0.47	-3.23	0.00	-0.75	0.00	0.00	0.00	0.00
400,000-600,000	-0.85	-5.66	-0.95	-1.74	-0.61	-0.64	-0.61	-0.64
$\geq 600,000$	-1.18	-4.12	-0.76	-1.66	-0.50	-0.60	-0.50	-0.60
Employment (employed)	0.00	-2.03	-0.54	-0.70	-0.52	-0.44	-0.52	-0.44
Type of workplace								
Government organization and institution	-0.36	-7.53	-0.74	-2.61	-0.76	-0.84	-0.76	-0.84
Firm	0.00	-0.39	0.00	0.00	0.00	0.00	0.00	0.00
Occupation								
Managers	0.28	-3.69	-0.38	-0.83	-0.52	-0.41	-0.52	-0.41
Commercial and service workers	-0.36	-1.95	-0.95	-0.76	-0.76	-0.40	-0.76	-0.40
Professional title								
Advanced	1.27	4.77	1.12	2.90	0.88	0.72	0.88	0.72
Intermediate	0.63	5.03	1.31	2.42	0.97	1.02	0.97	1.02
Primary	0.17	4.95	0.88	2.25	0.00	0.00	0.00	0.00

5. Conclusions

The cure rate model is commonly used when the data has long-term survivors. The model is composed of two parts. The incident part describes the probability of cure and the latency part describes the survival function of the uncured group. The drawback of the standard cure rate model is that it assumes that there are no direct constraints between the coefficients corresponding to the same covariates in the two parts of the model. This may lead to conflicting results of covariate effects on the probability of cure and conditional survival of the uncured group. In fact, the two parts of the model describe quite related aspects. It is desirable that there may be some connections between coefficients corresponding to the same covariates. Existing works have considered joint distribution or structural effect of the two sets of covariates, which is too strict.

In this paper, we consider a more flexible cure rate model that allows the two sets of covariates can be in different distributions and magnitudes. In the proposed method, a sign consistency cure rate model is proposed to promote the similarity in the sign of coefficients in the two model parts to improve interpretability. In addition, we also impose a group lasso penalty for variable selection. Simulation results show that compared with alternatives, the proposed method has better performance in terms of variable selection and estimation. An analysis of credit data in China illustrates that the proposed method can improve prediction performance as well as interpretability.

Acknowledgments

We are grateful to the reviewers and the editor for their helpful comments and suggestions. This work was supported by the National Office for Philosophy and Social Sciences of China under Grant 20&ZD137.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. P. Klein, M. L. Moeschberger, *Survival analysis: techniques for censored and truncated data*, 2 Eds., New York: Springer-Verlag, 2003. doi: 10.1007/b97377.
2. M. Stepanova, L. Thomas, Survival analysis methods for personal loan data, *Oper. Res.*, **50** (2002), 277–289. doi: 10.1287/opre.50.2.277.426.
3. V. B. Djeundje, J. Crook, Dynamic survival models with varying coefficients for credit risks, *Eur. J. Oper. Res.*, **275** (2019), 319–333. doi: 10.1016/j.ejor.2018.11.029.
4. Q. Zhang, S. Zhang, J. Liu, J. Huang, S. Ma, Penalized integrative analysis under the accelerated failure time model, *Stat. Sin.*, **26** (2016), 492–508. doi: 10.5705/ss.2014.194.
5. J. Berkson, R. P. Gage, Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.*, **47** (1952), 501–515. doi: 10.1080/01621459.1952.10501187.
6. J. Rodrigues, V. G. Cancho, M.D. Castro, F. Louzada-Neto, On the unification of long-term survival models, *Stat. Probability Letters*, **79** (2009), 753–759. doi: 10.1016/j.spl.2008.10.029.
7. F. Cooner, S. Banerjee, B. P. Carlin, D. Sinha, Flexible cure rate modeling under latent activation schemes, *J. Am. Stat. Assoc.*, **102** (2007), 560–572. doi: 10.1198/016214507000000112.

8. J. Rodrigues, M. Castro, V.G. Cancho, N. Balakrishnan, COM-Poisson cure rate survival models and an application to a cutaneous melanoma data, *J. Stat. Plan. Infer.*, **139** (2009), 3605–3611. doi: 10.1016/j.jspi.2009.04.014.
9. L. Li, J. H. Lee, A latent promotion time cure rate model using dependent tail-free mixtures, *J. R. Statist. Soc. A*, **180** (2017), 891–905. doi: 10.1111/rssa.12226.
10. L. Dirick, G. Claeskens, B. Baesens, Time to default in credit scoring using survival analysis: a benchmark study, *J. Oper. Res. Soc.*, **68** (2017), 652–665. doi: 10.1057/s41274-016-0128-9.
11. O. Georgiana, A. B. Lawson, Bayesian cure-rate survival model with spatially structured censoring, *Spatial Stat.*, **28** (2018), 352–364. doi: 10.1016/j.spasta.2018.08.007.
12. S. Pal, S. Roy, A new non-linear conjugate gradient algorithm for destructive cure rate model and a simulation study: illustration with negative binomial competing risks, *Commun. Stat.-Simul. Comput.*, 2020. doi: 10.1080/03610918.2020.1819321.
13. C. Li, J. M. G. Taylor, Smoothing covariate effects in cure models, *Commun. Statist.- Theory Meth.*, **31** (2002), 477–493. doi: 10.1081/STA-120002860.
14. T. Chen, P. Du, Promotion time cure rate model with nonparametric form of covariate effects, *Stat. Sin.*, **37** (2018): 1625–1635. doi: 10.1002/sim.7597.
15. E. N. C. Tong, C. Mues, L. C. Thomas, Mixture cure models in credit scoring: if and when borrowers default, *Eur. J. Oper. Res.*, **218** (2012), 132–139. doi: 10.1016/j.ejor.2011.10.007.
16. C. Jiang, Z. Wang, H. Zhao, A prediction-driven mixture cure model and its application in credit scoring, *Eur. J. Oper. Res.*, **277** (2019), 20–31. doi: 10.1016/j.ejor.2019.01.072.
17. C. Han, R. Kronmal, Two-part models for analysis of Agatston scores with possible proportionality constraints, *Commun. Stat.-Theory Meth.*, **35** (2006), 99–111. doi: 10.1080/03610920500438614.
18. K. Fang, X. Wang, B.C. Shia, S. Ma, Identification of proportionality structure with two-part models using penalization, *Comput. Stat. Data Anal*, **99** (2016), 12–24. doi: 10.1016/j.csda.2016.01.002.
19. F. Liu, Z. Hua, A. Lim, Identifying future defaulters: a hierarchical Bayesian method, *Eur. J. Oper. Res.*, **241** (2015), 202–211. doi: 10.1016/j.ejor.2014.08.008.
20. X. Fan, M. Liu, K. Fang, Y. Huang, S. Ma, Promoting structural effects of covariates in the cure rate model with penalization, *Stat. Methods Med. Res.*, **26** (2017), 2078–2092. doi: 10.1177/0962280217708684.
21. Q. Zhang, S. Ma, Y. Huang, Promote sign consistency in the joint estimation of precision matrices, *Comput. Stat. Data Anal.*, **159** (2021), 107210. doi: 10.1016/j.csda.2021.107210.
22. X. Shi, S. Ma, and Y. Huang, Promoting sign consistency in the cure model estimation and selection, *Stat. Methods Med. Res.*, **29** (2020), 15–28. doi: 10.1177/0962280218820356.
23. M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, *J. R. Statist. Soc. B*, **68** (2006), 49–67. doi: 10.1111/j.1467-9868.2005.00532.x.
24. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2 Eds., New York: Springer, 2009. doi: 10.1007/978-0-387-84858-7.
25. J. Huang, P. Breheny, S. Ma, A selective review of group selection in high-dimensional models, *Stat. Sci.*, **27** (2012), 481–499. doi: 10.1214/12-STS392.
26. N. Balakrishnan, S. Pal, Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes, *Stat. Methods Med. Res.*, **25** (2016), 1535–1563. doi: 10.1177/0962280213491641.

27. M. Omer, M. Bakar, M. B. Adam, M. S. Mustafa, Cure models with exponentiated Weibull exponential distribution for the analysis of melanoma patients, *Mathematics*, **8** (2021), 1926. doi: 10.3390/math8111926.
28. S. Pal, N. Balakrishnan, Likelihood inference based on EM algorithm for the destructive length-biased Poisson cure rate model with Weibull lifetime, *Commun. Stat. Simulation Computation*, **47** (2018), 644–660. doi: 10.1080/03610918.2015.1053918.
29. X. Li, Y. Tang, A. Xu, Objective Bayesian analysis of Weibull mixture cure model, *Qual. Eng.*, **32** (2020), 449–464. doi: 10.1080/08982112.2020.1757706.
30. J. Huang, T. Zhang, The benefit of group sparsity, *Ann. Stat.*, **38** (2010), 1978–2004. doi: 10.1214/09-AOS778.
31. L. Meier, S. V. D. Geer, P. Bhlmann, E. T. H. Zrich, The group lasso for logistic regression, *J. R. Statist. Soc. B*, **70** (2008), 53–71. doi: 10.1111/j.1467-9868.2007.00627.x.
32. Y. Yang, H. Zou, A fast unified algorithm for solving group-lasso penalize learning problems, *Stat. and Comput.*, **25** (2015), 1129–1141. doi: 10.1007/s11222-014-9498-5.
33. H. Wang, B. Li, C. Leng, Shrinkage tuning parameter selection with a diverging number of parameters, *J. R. Statist. Soc. B*, **71** (2009), 671–683. doi: 10.1111/j.1467-9868.2008.00693.x.
34. S. Pal, A simplified stochastic EM algorithm for cure rate model with negative binomial competing risks: an application to breast cancer data, *Stat. Med.*, 2021. doi: 10.1002/sim.9189.
35. Y. Li, Y. Li, Y. Li, What factors are influencing credit card customer's default behavior in China? A study based on survival analysis, *Physica A*, **526** (2019), Article ID 120861. doi: 10.1016/j.physa.2019.04.097.
36. Y. Shu, Q. Y. Yang, Research on auto loan default prediction based on large sample data model, *Manage. Rev.*, **29** (2017), 59–71.



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)