



Article

Profile and Non-Profile MM Modeling of Cluster Failure Time and Analysis of ADNI Data

Xifen Huang ^{1,*}, Jinfeng Xu ^{1,*}  and Yunpeng Zhou ² ¹ School of Mathematics, Yunnan Normal University, Kunming 650092, China; 190004@ynnu.edu.cn² Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong, China; u3514104@connect.hku.hk

* Correspondence: 204203@ynnu.edu.cn

Abstract: Motivated by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) data, the objective of integration of important biomarkers for the early detection of Mild Cognitive Impairment (MCI) to Alzheimer’s disease (AD) as a therapeutic intervention is most likely to be beneficial in the early stages of disease progression. Developing predictors for MCI to AD comes down to genotype variables such that the dimension of predictors increases as the sample becomes large. Thus, we consider the sparsity concept of coefficients in a high-dimensional regression model with clustered failure time data such as ADNI, which enables enhancing predictive performances and facilitates the model’s interpretability. In this study, we propose two MM algorithms (profile and non-profile) for the shared frailty survival model firstly and then extend the two proposed MM algorithms to regularized estimation in sparse high-dimensional regression model. The convergence properties of our proposed estimators are also established. Furthermore simulation studies and analysis of ADNI data are illustrated by our proposed methods.

Keywords: clustering; frailty model; sparsity; MM algorithm; ADNI

Citation: Huang, X.; Xu, J.; Zhou, Y. Profile and Non-Profile MM Modeling of Cluster Failure Time and Analysis of ADNI Data. *Mathematics* **2022**, *10*, 538. <https://doi.org/10.3390/math10040538>

Academic Editors: Min Wang, Haijun Gong, Liucang Wu and Songfeng Zheng

Received: 8 January 2022

Accepted: 31 January 2022

Published: 9 February 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In biomedical research, we often encounter clustered failure time data in which individuals from the same cluster (e.g., family) share common genetic and/or environmental factors. An illustrative example comes from the Alzheimer’s Disease Neuroimaging Initiative study where each participant may develop Mild Cognitive Impairment (MCI) and/or develop Alzheimer’s disease (AD). The times to these two events for each individual are expected to be highly associated.

In order to account for the correlation of associated failure times, shared frailty or random effect models are commonly used by researchers ([1–4]). In particular, the gamma frailty model ([5–8]) is numerically convenient for such analysis since the likelihood function exhibits a closed form. For example, ref. [2] proposed a frailty model where individuals within the same cluster share a common random effect. Ref. [9–11] provided a good and comprehensive review of the applications of different frailty models for clustered survival data. The authors of [12] compared several commonly used methods and demonstrated the advantages of the gamma frailty model. Ref. [13,14] also explored the nonparametric frailty model and the between-within frailty model for correlated survival data.

For high-dimension regression analysis, an important and useful strategy is to exploit sparsity and assume that the true parameters lie in a low-dimensional subspace. In the past years, sparsity-restricted estimation has attracted a great deal of attention in high-dimensional regression models. That is, only a few regression coefficients are assumed to be nonzero ([15,16]). In addition to the widely used LASSO penalty, there exists other types of regularization methods such as ridge regression ([17]), bridge regression ([18]), and elastic net ([19]). In order to improve the performance of the LASSO, many modifications

such as the adaptive Lasso, the smoothly clipped absolute deviation (SCAD, [20]), and minimax concave penalty (MCP, [21]) have also been proposed.

In this paper, we develop fast and efficient algorithms for regularized estimation in the general frailty model for clustered failure time data. As the model parameters consists of both the regression coefficients and the unknown nonparametric baseline hazard function, computation in the frailty model with survival data is usually intensive, especially when frailty does not result in a closed-form likelihood function. The regularized estimation poses even more challenges for the problem and further complicates computational burden. The existing approaches of general frailty model almost rely on the EM algorithm. The EM algorithm adopts Newton's method in its maximization step, which involves matrix inversion and may not have good performance in a high-dimensional environment. The minorization-maximization (MM) algorithm possesses an ascent property, which driving the target likelihood function to increase and reliably converges to the maximum from well-chosen initial values ([22–24]). In particular, ref. [25] proposed three different MM algorithms for the gamma frailty model and further demonstrated their utility for the estimation in high-dimensional situations via decomposing a high-dimensional objective function into separable low-dimensional functions. In this paper, we first develop a pair of MM algorithms, including profile and non-profile, for the general frailty model. As demonstrated by [25] in numerical studies, the decomposition inoculated with regularized estimation well in sparse high-dimensional settings. For illustration, we chose to use concave penalties such as the smoothly clipped absolute deviations penalty (SCAD, [20]) and the minimax concave penalty (MCP, [21]) to explore the sparsity because they both possess the good property of unbiasedness.

The rest of the paper is organized as follows. In Section 2, we provide the model description, an overview of MM principle, and then propose a pair of MM algorithms. In Section 3, we derive regularized estimation methods via profile and non-profile MM algorithms in sparse high-dimensional regression setting. The convergence properties of the proposed algorithms are provided in Section 4. In Section 5, a series of simulation studies was conducted to assess the finite-sample performances of the proposed methods. Section 6 provides a real application to the ADNI data.

2. The Model and Estimation

2.1. The Model

Consider datasets from some population that contains $M_i \geq 1$ individuals in i -th subgroup of the population, $i = 1, \dots, B$. Individuals within the i -th subgroup have dependent event times due to some unobserved covariate information summarized in a frailty, ω_i . Let Y_{ij} be the event time, let C_{ij} be the censoring time, and let $\mathbf{X}_{ij}^\top = (X_{ij1}, \dots, X_{ijp})$ denote the potential covariates for the j -th individual in the i -th subgroup. The censoring time C_{ij} is assumed to be independent of Y_{ij} , given covarites \mathbf{X}_{ij} and frailty ω_i . Define $t_{ij} = \min(Y_{ij}, C_{ij})$ and $\delta_{ij} = I(Y_{ij} \leq C_{ij})$, where $I(\cdot)$ denotes the indicator function. Suppose that censorship is noninformative. Conditional on frailty ω_i , the hazard rate function is of the following form:

$$\lambda(t_{ij} | \mathbf{X}_{ij}, \omega_i) = \omega_i \lambda_0(t_{ij}) \exp\{\mathbf{X}_{ij}^\top \boldsymbol{\beta}\}, \quad (1)$$

where $\lambda_0(\cdot)$ is an arbitrary baseline hazard rate, and $\boldsymbol{\beta}$ is a vector of unknown parameters. Assume that $\omega_i, i = 1, \dots, B$ are independent and identically distributed with density function $f(\omega_i | \boldsymbol{\theta})$ on the domain \mathbb{W} . Denote $\boldsymbol{\alpha} = (\boldsymbol{\theta}, \boldsymbol{\beta}, \Lambda_0)$; we then propose profile MM and non-profile MM methods to estimate parameters based on the minorization-maximization (MM) principle.

2.2. An Overview of MM Principle

Before introducing the two proposed methods, let us review the MM principle. Assume that Y_{obs} is the observed data, $\ell(\boldsymbol{\alpha} | Y_{obs})$ is the log-likelihood function with unknown

parameter $\alpha = (\alpha_1, \dots, \alpha_q)^T$, and the maximum likelihood estimate of α is $\hat{\alpha} = \operatorname{argmax} \ell(\alpha | Y_{obs})$. The MM principle involves two M steps: One is a minorization step, and the other is maximization step. In a maximization problem, the first step is minorization, which aims to construct a minorization/surrogate function for the objective log-likelihood function $\ell(\alpha | Y_{obs})$ to be maximized through a series of inequalities satisfying the following two conditions:

$$\begin{cases} Q(\alpha | \alpha^{(k)}) \leq \ell(\alpha | Y_{obs}), \\ Q(\alpha^{(k)} | \alpha^{(k)}) = \ell(\alpha^{(k)} | Y_{obs}), \end{cases} \tag{2}$$

where $\alpha^{(k)}$ is the k -th approximation of $\hat{\alpha}$. Once the minorization function $Q(\alpha | \alpha^{(k)})$ is successfully constructed for objective function $\ell(\alpha | Y_{obs})$, the following maximization step is to maximize the surrogate function $Q(\alpha | \alpha^{(k)})$ to obtain the $(k + 1)$ -th approximation of $\hat{\alpha}$ rather than the objective function, i.e.,

$$\alpha^{(k+1)} = \operatorname{argmax} Q(\alpha | \alpha^{(k)}).$$

By MM principle, we may have $\ell(\alpha^{(k+1)} | Y_{obs}) \geq Q(\alpha^{(k+1)} | \alpha^{(k)}) \geq Q(\alpha^{(k)} | \alpha^{(k)}) = \ell(\alpha^{(k)} | Y_{obs})$, and the values of the objective function continue to increase until convergence.

2.3. Profile MM Estimation Procedure

For general shared frailty models, we can write the observed log-likelihood function $\ell(\alpha | Y_{obs})$ as follows:

$$\ell(\alpha | Y_{obs}) = \sum_{i=1}^B \log \int_{\mathbb{W}} \tau_i(\omega_i | \alpha) d\omega_i, \tag{3}$$

where the following is the case.

$$\tau_i(\omega_i | \alpha) = f(\omega_i | \theta) \prod_{j=1}^{M_i} \left\{ [\lambda_0(t_{ij}) \omega_i \exp(\mathbf{X}_{ij}^T \beta)]^{\delta_{ij}} \exp[-\Lambda_0(t_{ij}) \omega_i \exp(\mathbf{X}_{ij}^T \beta)] \right\}.$$

Generally, the Laplace transform of the frailty’s distribution is theoretically intractable. Hence, the explicit form of marginal hazard is not available to us. However, that has not stopped us from developing the profile MM approach for estimation in general shared frailty model. Define the following:

$$v_i(\omega_i | \alpha^{(k)}) = \frac{\tau_i(\omega_i | \alpha^{(k)})}{\int_{\mathbb{W}} \tau_i(\omega_i | \alpha^{(k)}) d\omega_i},$$

and rewrite the objective function as follows.

$$\ell(\alpha | Y_{obs}) = \sum_{i=1}^B \log \left[\int_{\mathbb{W}} \frac{\tau_i(\omega_i | \alpha)}{v_i(\omega_i | \alpha^{(k)})} \cdot v_i(\omega_i | \alpha^{(k)}) d\omega_i \right]. \tag{4}$$

By Jensen’s inequality, we have the following:

$$\varphi \left[\int_{\mathbb{X}} h(x) \cdot g(x) dx \right] \geq \int_{\mathbb{X}} \varphi[h(x)] \cdot g(x) dx,$$

where \mathbb{X} is a subset of the real line \mathbb{R} , $\varphi(\cdot)$ is the concave function, $h(\cdot)$ is an arbitrary real-valued function defined on \mathbb{X} , and $g(\cdot)$ is a density function defined on \mathbb{X} . In Equation (4), $v_i(\omega_i | \alpha^{(k)})$ is a density function, and choosing $h(x)$ as $\tau_i(\omega_i | \alpha) / v_i(\omega_i | \alpha^{(k)})$, we can apply the above Jensen’s inequality and construct the following surrogate function for $\ell(\alpha | Y_{obs})$:

$$Q_1(\alpha | \alpha^{(k)}) = Q_{11}(\theta | \alpha^{(k)}) + Q_{12}(\beta, \Lambda_0 | \alpha^{(k)}), \tag{5}$$

where the following is the case:

$$Q_{11}(\boldsymbol{\theta}|\boldsymbol{\alpha}^{(k)}) = \sum_{i=1}^B \int_{\mathbb{W}} \log[f(\omega_i|\boldsymbol{\theta})] \cdot v_i(\omega_i|\boldsymbol{\theta}^{(k)}, \boldsymbol{\alpha}^{(k)}) d\omega_i, \tag{6}$$

which only consists of parameter $\boldsymbol{\theta}$. Moreover, we have the following:

$$Q_{12}(\boldsymbol{\beta}, \Lambda_0|\boldsymbol{\alpha}^{(k)}) = \sum_{i=1}^B \sum_{j=1}^{M_i} \left[\delta_{ij} \log(\lambda_0(t_{ij})) + \delta_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - A_i^{(k)} \Lambda_0(t_{ij}) \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}) \right], \tag{7}$$

where $A_i^{(k)} = \int_{\mathbb{W}} \omega_i \cdot v_i(\omega_i|\boldsymbol{\alpha}^{(k)}) d\omega_i, i = 1, \dots, B$. The minorizing function $Q_1(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)})$ separates parameters $\boldsymbol{\theta}$ and $(\boldsymbol{\beta}, \Lambda_0)$ into (6) and (7), respectively. In the second M-step, the updated estimates of $\boldsymbol{\theta}$ involve maximizing (6) numerically. Due to the presence of the nonparametric component Λ_0 , updating $(\boldsymbol{\beta}, \Lambda_0)$ is still a big challenge. Following [7], we apply the profile estimation method to profile out Λ_0 in $Q_{12}(\boldsymbol{\beta}, \Lambda_0|\boldsymbol{\alpha}^{(k)})$, which results in the estimate of Λ_0 given $\boldsymbol{\beta}$.

$$d\hat{\Lambda}_0(t_{ij}) = \frac{\delta_{ij}}{\sum_{r=1}^B \sum_{s=1}^{M_r} I(t_{rs} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rs}^\top \boldsymbol{\beta})}. \tag{8}$$

Substituting (8) into $Q_{12}(\boldsymbol{\beta}, \Lambda_0|\boldsymbol{\alpha}^{(k)})$, we obtain the following:

$$Q_{13}(\boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)}) = \sum_{i=1}^B \sum_{j=1}^{M_i} \left\{ \delta_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - \delta_{ij} \log \left[\sum_{r=1}^B \sum_{s=1}^{M_r} I(t_{rs} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rs}^\top \boldsymbol{\beta}) \right] \right\}, \tag{9}$$

which involves $\boldsymbol{\beta}$ only. In Equation (9), Newton’s method and large matrix inversion are required to update $\boldsymbol{\beta}$ in this procedure when there exist a large number of covariates. Here, we further construct minorizing functions for $Q_{13}(\boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$ to separate the regression parameters β_1, \dots, β_q from each other under the MM principle. We first use the supporting hyperplane inequality:

$$-\log(x) \geq -\log(x_0) - \frac{x - x_0}{x_0}$$

to minorize $Q_{13}(\boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$; then, we have the following surrogate function:

$$Q_{14}(\boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)}) = \sum_{i=1}^B \sum_{j=1}^{M_i} \left[\delta_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - \frac{\delta_{ij} \sum_{r=1}^B \sum_{s=1}^{M_r} I(t_{rs} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rs}^\top \boldsymbol{\beta})}{\sum_{r=1}^B \sum_{s=1}^{M_r} I(t_{rs} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rs}^\top \boldsymbol{\beta}^{(k)})} \right] + c,$$

where c is a constant not depending on $\boldsymbol{\beta}$. We apply Jensen’s inequality to the concave function $-\exp(\cdot)$ in $Q_{14}(\boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$ by rewriting the following:

$$\mathbf{X}_{rs}^\top \boldsymbol{\beta} = \sum_{p=1}^q \pi_{prs} [\pi_{prs}^{-1} X_{prs} (\beta_p - \beta_p^{(k)}) + \mathbf{X}_{rs}^\top \boldsymbol{\beta}^{(k)}],$$

where $\pi_{prs} = |X_{prs}| / \sum_{p=1}^q |X_{prs}|$. Finally, the minorizing function for $Q_{14}(\boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$ is as follows:

$$Q_{15}(\beta_1, \dots, \beta_q|\boldsymbol{\alpha}^{(k)}) \triangleq \sum_{p=1}^q Q_{15p}(\beta_p|\boldsymbol{\alpha}^{(k)}), \tag{10}$$

where the following is the case.

$$Q_{15p}(\beta_p | \alpha^{(k)}) = \sum_{i=1}^B \sum_{j=1}^{M_i} \left\{ \delta_{ij} X_{pij} \beta_p \right. \\ \left. - \frac{\delta_{ij} \sum_{r=1}^B \sum_{s=1}^{M_r} I(t_{rs} \geq t_{ij}) A_r^{(k)} \pi_{pr} \exp [\pi_{pr}^{-1} X_{prs} (\beta_p - \beta_p^{(k)}) + \mathbf{X}_{rs}^\top \boldsymbol{\beta}^{(k)}]}{\sum_{r=1}^B \sum_{s=1}^{M_r} I(t_{rs} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rs}^\top \boldsymbol{\beta}^{(k)})} \right\}, p = 1, \dots, q. \tag{11}$$

In general, the structure of minorizing function for the objective log-likelihood function $\ell(\alpha | Y_{obs})$ in (3) is the following:

$$Q_{pro}(\theta, \beta | \alpha^{(k)}) = Q_{11}(\theta | \alpha^{(k)}) + \sum_{p=1}^q Q_{15p}(\beta_p | \alpha^{(k)}), \tag{12}$$

with an explicit form update of $d\Lambda_0$ by (8). We may observe that the objective function to be maximized is decomposed into a sum of $q + 1$ univariate functions from (12) as $Q_{11}(\theta | \alpha^{(k)})$ only consists of one parameter usually. The next maximization step of this MM algorithm only involves $q + 1$ separate univariate optimizations and matrix inversion is unnecessary. Note that the success of the above profile MM algorithm requires the convergence of two improper integrals such as $\int_{\mathbb{W}} \omega_i \cdot v_i(\omega_i | \alpha^{(k)}) d\omega_i$ and $\int_{\mathbb{W}} \log[f(\omega_i | \theta)] \cdot v_i(\omega_i | \alpha^{(k)}) d\omega_i$. Moreover, the convergences of these improper integrals are obvious when we assume the distribution of random effect comes from an exponential distribution family. The estimation proceeds by profile MM algorithm are summarized as follows:

- Step 1.** Given initial values for θ , β , and Λ_0 ;
- Step 2.** Update θ via (6). Update each β_p via (11) for $p = 1, \dots, q$;
- Step 3.** Based on the update of β , compute the estimates of $\Lambda_0(t_{ij})$ via (8);
- Step 4.** Iterate steps 2 and 3 until convergence.

2.4. Non-Profile MM Estimation Procedure

In this subsection, we bypass the profile estimation procedure in previous subsection and continue to develop new MM procedures for Equation (7) to separate parameters β and nuisance baseline hazard rate Λ_0 . Actually, to separate β and Λ_0 of (7) is to deal with the last term $-\Lambda_0(t_{ij}) \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta})$. As in [26], we use the following arithmetic-geometric mean inequality.

$$-\prod_{i=1}^n x_i^{a_i} \geq -\sum_{i=1}^n \frac{a_i}{\|\mathbf{a}\|_1} x_i^{\|\mathbf{a}\|_1} \tag{13}$$

Here, x_i and a_i are non-negative. Choosing $x_1 = \Lambda_0(t_{ij}) / \Lambda_0^{(k)}(t_{ij})$ and $x_2 = \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}) / \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)})$ in inequality (13), we obtain the following surrogate function for (7):

$$Q_2(\beta, \Lambda_0 | \alpha^{(k)}) \\ = \sum_{i=1}^B \sum_{j=1}^{M_i} \left[\delta_{ij} \log(\lambda_0(t_{ij})) + \delta_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - \frac{A_i^{(k)} \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)})}{2\Lambda_0^{(k)}(t_{ij})} \Lambda_0(t_{ij})^2 - \frac{A_i^{(k)} \Lambda_0^{(k)}(t_{ij})}{2 \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)})} \exp(2\mathbf{X}_{ij}^\top \boldsymbol{\beta}) \right], \\ \doteq Q_{21}(\Lambda_0 | \alpha^{(k)}) + Q_{22}(\beta | \alpha^{(k)}) \tag{14}$$

where the following is the case.

$$Q_{21}(\Lambda_0|\mathbf{a}^{(k)}) = \sum_{i=1}^B \sum_{j=1}^{M_i} \left[\delta_{ij} \log(\lambda_0(t_{ij})) - \frac{A_i^{(k)} \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)})}{2\Lambda_0^{(k)}(t_{ij})} \Lambda_0(t_{ij})^2 \right], \tag{15}$$

$$Q_{22}(\boldsymbol{\beta}|\mathbf{a}^{(k)}) = \sum_{i=1}^B \sum_{j=1}^{M_i} \left[\delta_{ij} \mathbf{X}_{ij}^\top \boldsymbol{\beta} - \frac{A_i^{(k)} \Lambda_0^{(k)}(t_{ij})}{2 \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)})} \exp(2\mathbf{X}_{ij}^\top \boldsymbol{\beta}) \right]. \tag{16}$$

In order to obtain the nonparametric estimate of Λ_0 , the maximization of (15) is required. For ease of computation, a one-step late skill is applied to the first order derivative of Equation (15); then, we obtain the estimate of λ_0 by the following:

$$d\hat{\Lambda}_0(t_{ij}) = \frac{\delta_{ij}}{\sum_{r=1}^B \sum_{s=1}^{M_r} I(t_{rs} \geq t_{ij}) A_r^{(k)} \exp(\mathbf{X}_{rs}^\top \boldsymbol{\beta}^{(k)})}, \tag{17}$$

which is same as (8) in the profile estimation method. To update $\boldsymbol{\beta}$, a similar technique as dealing with $Q_{14}(\boldsymbol{\beta}|\mathbf{a}^{(k)})$ is used. We apply Jensen’s inequality to the concave function $-\exp(\cdot)$ in $Q_{22}(\boldsymbol{\beta}|\mathbf{a}^{(k)})$ by rewriting the following:

$$2\mathbf{X}_{ij}^\top \boldsymbol{\beta} = \sum_{p=1}^q \pi_{pij} [2\pi_{pij}^{-1} X_{pij} (\beta_p - \beta_p^{(k)}) + 2\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)}],$$

where $\pi_{pij} = |X_{pij}| / \sum_{p=1}^q |X_{pij}|$. In the end, the minorizing function for $Q_{22}(\boldsymbol{\beta}|\mathbf{a}^{(k)})$ is as follows:

$$Q_{23}(\beta_1, \dots, \beta_q|\mathbf{a}^{(k)}) \hat{=} \sum_{p=1}^q Q_{23p}(\beta_p|\mathbf{a}^{(k)}), \tag{18}$$

where the following is obtained:

$$Q_{23p}(\beta_p|\mathbf{a}^{(k)}) = \sum_{i=1}^B \sum_{j=1}^{M_i} \left\{ \delta_{ij} X_{pij} \beta_p - \frac{\pi_{pij} A_i^{(k)} \Lambda_0^{(k)}(t_{ij}) \exp [2\pi_{pij}^{-1} X_{pij} (\beta_p - \beta_p^{(k)}) + 2\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)}]}{2 \exp(\mathbf{X}_{ij}^\top \boldsymbol{\beta}^{(k)})} \right\}, \tag{19}$$

for $p = 1, \dots, q$. As a result, we construct the surrogate function for the objective log-likelihood function via a non-profile MM principle as follows:

$$Q_{nonpro}(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{a}^{(k)}) = Q_{11}(\boldsymbol{\theta}|\mathbf{a}^{(k)}) + \sum_{p=1}^q Q_{23p}(\beta_p|\mathbf{a}^{(k)}), \tag{20}$$

with explicit form update of $d\Lambda_0$ by (17). From (20), we can find similar nice features as $Q_{pro}(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{a}^{(k)})$; that is, $Q_{nonpro}(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{a}^{(k)})$ is a sum of $q + 1$ univariate functions, which means that the next maximization (second M) step only involves $q + 1$ simple univariate optimizations. It is worth noting that the parameter separated feature in the proposed profile MM and non-profile MM algorithms will help incorporate with the existing simple off-the-shelf accelerators well and brings about great effectiveness in computation time, as discussed in [25]. The estimation proceeds by non-profile MM algorithm are summarized as follows:

Step 1. Given initial values of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, and Λ_0 ;

Step 2. Update the estimate of $\boldsymbol{\theta}$ via (6). Update the estimate of β_p based on (19) for $p = 1, \dots, q$;

Step 3. Using the updated estimate of β , compute the estimates of $\Lambda_0(t_{ij})$ via (17);

Step 4. Iterate steps 2 and 3 until convergence.

3. Regularized Estimation Methods via MM Methods

Followed by Section 3, both proposed estimation approaches created nice parameter-separated surrogate functions to be maximized which may incorporate with regularization cleverly. In this part, we propose regularized estimation approaches based on MM algorithms in regression analysis under general shared frailty model. Many variable selection criteria arise as special cases of the general formulation as discussed in [20], where the penalized likelihood function takes the form

$$\ell^P(\alpha|Y_{obs}) = \ell(\theta, \beta, \Lambda_0|Y_{obs}) - N \sum_{p=1}^q P(|\beta_p|, \lambda), \tag{21}$$

where $\ell(\theta, \beta, \Lambda_0|Y_{obs})$ is the log-likelihood function for the shared frailty model, q is the dimension of β , $P(\cdot, \lambda)$ is a given nonnegative penalty function, and $\lambda \geq 0$ is a tuning parameter, which is allowed to use λ_p in more general cases. This penalty shrinks some of the coefficients to zero. Under the scope of general frailty topic, the computation of MLEs is extremely complicated and hard in terms of accuracy as parameters of interests involve three blocks θ , β , and Λ_0 , and even more complicated when there exists a large numbers of coefficients. It is worth mentioning that the proposed profile and non-profile MM algorithms decomposed the coefficient vector β from the other two blocks θ , and Λ_0 and different coefficient parameters are separated from each other. This nice feature of the proposed profile and non-profile MM algorithms may mesh well with various regularization problems in (21) to produce more sparse and accurate estimates. Under profile MM algorithmic technique in (12), we obtain the corresponding minorization function for (21) as follows.

$$Q_{pro}(\theta, \beta|\alpha^{(k)}) - N \sum_{p=1}^q P(|\beta_p|, \lambda). \tag{22}$$

Under the non-profile MM algorithmic technique in (20), the minorization function for (21) is as follows.

$$Q_{nonpro}(\theta, \beta|\alpha^{(k)}) - N \sum_{p=1}^q P(|\beta_p|, \lambda). \tag{23}$$

When $P(\cdot, \lambda)$ is piecewise differentiable, nondecreasing, and concave on $(0, \infty)$ such as L_1 , MCP and SCAD penalties [27], explored a connection between the local quadratic approximation with MM algorithm. The penalty term $-P(\cdot, \lambda)$ can be minorized by a local quadratic approximation form as follows:

$$-P(|\beta_p|, \lambda) \geq -P(|\beta_p^{(k)}|, \lambda) - \frac{[\beta_p^2 - (\beta_p^{(k)})^2]P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|} \triangleq \Phi(\beta_p|\beta_p^{(k)}), \tag{24}$$

which is actually a one-step minorizing process. By combining function (24) with (12) or (20), respectively, we obtain the final surrogate functions for penalized likelihood function (21) as follows:

$$Q_{pro}^P(\theta, \beta|\alpha^{(k)}) = Q_{pro}(\theta, \beta|\alpha^{(k)}) - N \sum_{p=1}^q \frac{\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|} + c_1, \tag{25}$$

and the following is obtained.

$$Q_{nonpro}^P(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)}) = Q_{nonpro}(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)}) - N \sum_{p=1}^q \frac{\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|} + c_2. \tag{26}$$

Both Equations (25) and (26) are written as a sum of a series of univariate functions so that maximizing (25) and (26) will be easier than directly maximizing (21). Moreover, some simple off-the-shelf accelerators may also be used here to make the optimization problems more simpler and efficient. Regularized estimation proceeds by (profile/non-profile) MM algorithms and are summarized as follows:

Step 1. Given initial values of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and Λ_0 ;

Step 2. Update the estimate of $\boldsymbol{\theta}$ via (6);

Step 3. For profile MM method, update $\boldsymbol{\beta}$ by maximizing

$$\sum_{p=1}^q Q_{15p}(\beta_p|\boldsymbol{\alpha}^{(k)}) - N \sum_{p=1}^q \frac{\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|}.$$

For non-profile MM method, update $\boldsymbol{\beta}$ by maximizing

$$\sum_{p=1}^q Q_{23p}(\beta_p|\boldsymbol{\alpha}^{(k)}) - N \sum_{p=1}^q \frac{\beta_p^2 \cdot P'(|\beta_p^{(k)}|_+, \lambda)}{2|\beta_p^{(k)}|};$$

Step 4. Using the updated estimate of $\boldsymbol{\beta}$ in Step 3, compute the estimates of $\Lambda_0(t_{ij})$ via (8) for profile MM method and via (17) for non-profile MM method, respectively;

Step 5. Iterate steps 2 to 4 until convergence.

Model Selection

From the recent literature, tuning parameter λ may be selected by multiple data driven model selection criteria, such as Bayesian information criterion BIC ([28]) and generalized cross-validation GCV ([29]). In this paper, we consider a widely used BIC-type criterion, defined by the following:

$$BIC_\lambda = -2\ell(\hat{\boldsymbol{\alpha}}) + C_n(\hat{S} + 1) \log(N), \tag{27}$$

to select the tuning parameter λ , where $C_n = \max\{1, \log[\log(q + 1)]\}$, q is the dimension of $\boldsymbol{\beta}$, and the degrees of freedom \hat{S} are defined as the number of nonzero parameters in $\hat{\boldsymbol{\beta}}$.

4. Theoretical Properties

We first consider the convergence properties for the profile and non-profile MM algorithms of maximizing $\ell(\boldsymbol{\alpha})$ and $\ell^P(\boldsymbol{\alpha})$. As discussed in Sections 3 and 4, $Q_{pro}(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$ and $Q_{nonpro}(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$ are the constructed minorizing functions of $\ell(\boldsymbol{\alpha})$ via the profile and non-profile MM approach, $Q_{pro}^P(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$ and $Q_{nonpro}^P(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{\alpha}^{(k)})$ are the constructed minorizing functions of penalized likelihood function $\ell^P(\boldsymbol{\alpha})$ via the profile and non-profile MM approach, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are the parameter vectors, and $\boldsymbol{\alpha}^{(k)}$ is its current estimate. In this paper, we assume that the common component $Q_{11}(\boldsymbol{\theta}|\boldsymbol{\alpha}^{(k)})$ of all minorizing functions is strictly concave with respect to $\boldsymbol{\theta}$, and the second M step of MM principle is based on the Newton–Raphson method. As strict concavity holds for all surrogate functions $Q_{pro}(\cdot|\boldsymbol{\alpha}^{(k)})$, $Q_{nonpro}(\cdot|\boldsymbol{\alpha}^{(k)})$, $Q_{pro}^P(\cdot|\boldsymbol{\alpha}^{(k)})$ and $Q_{nonpro}^P(\cdot|\boldsymbol{\alpha}^{(k)})$ by using concave penalties L_1 , MCP and SCAD, we denote their unique maximizers by $M_{pro}(\boldsymbol{\alpha}^{(k)})$, $M_{nonpro}(\boldsymbol{\alpha}^{(k)})$, $M_{pro}^P(\boldsymbol{\alpha}^{(k)})$, and $M_{nonpro}^P(\boldsymbol{\alpha}^{(k)})$, respectively. Following Proposition 15.4.3 of Lange [30], we provide the following convergence properties.

Proposition 1. Assume the differentiability and coerciveness of $-\ell(\alpha)$ hold, all stationary points of $-\ell(\alpha)$ are isolated and the subsets $\{\alpha \in \Omega : \ell(\alpha) \geq \ell(\alpha^{(k)})\}$ of parameter domain Ω are compact. Then, the profile iteration sequence $(\theta^{(k+1)}, \beta^{(k+1)}) = M_{\text{pro}}(\alpha^{(k)})$ together with $\Lambda_0^{(k+1)}$ in (8) and non-profile sequence $(\theta^{(k+1)}, \beta^{(k+1)}) = M_{\text{nonpro}}(\alpha^{(k)})$ together with $\Lambda_0^{(k+1)}$ in (17) converge to the stationary point of $\ell(\alpha)$. If the strict concavity of $\ell(\alpha)$ also hold, then the profile MM sequence and non-profile MM sequence converge to the same maximum point of $\ell(\alpha)$.

Proposition 2. Suppose $P(\cdot|\lambda)$ is piecewise differentiable, nondecreasing and concave on $(0, \infty)$, continuous at 0 and $P'(0_+|\lambda) < \infty$. Then, for all $\beta_p \neq 0$, $\Phi(\beta_q|\beta_q^{(k)})$ as defined in (24) minorizes $-P(|\beta_q||\lambda)$ at points $\pm|\beta_q|$. In particular, conditions in construction minorization function hold, that is, $-P(|\beta_q||\lambda) \geq \Phi(\beta_q|\beta_q^{(k)})$ for all β_q with equality at $\beta_q = \beta_q^{(k)}$ and the ascent property $-P(|\beta_q^{(k+1)}||\lambda) \geq \Phi(\beta_q^{(k+1)}|\beta_q^{(k)}) \geq \Phi(\beta_q^{(k)}|\beta_q^{(k)}) = -P(|\beta_q^{(k)}||\lambda)$ hold.

Based on Proposition 2, we can easily obtain the following convergence properties of profile MM and non-profile MM algorithms for maximizing penalized likelihood function $\ell^P(\alpha)$ defined in (21).

Proposition 3. Assume the differentiability and coerciveness of $-\ell^P(\alpha)$ hold, all stationary points of $-\ell^P(\alpha)$ are isolated and the subsets $\{\alpha \in \Omega : \ell^P(\alpha) \geq \ell^P(\alpha^{(k)})\}$ of parameter domain Ω are compact. Then, the profile iteration sequence $(\theta^{(k+1)}, \beta^{(k+1)}) = M_{\text{pro}}^P(\alpha^{(k)})$ together with $\Lambda_0^{(k+1)}$ in (8) and non-profile sequence $(\theta^{(k+1)}, \beta^{(k+1)}) = M_{\text{nonpro}}^P(\alpha^{(k)})$ together with $\Lambda_0^{(k+1)}$ in (17) converge to the stationary point of $\ell^P(\alpha)$. If the strict concavity of $\ell^P(\alpha)$ also hold, then the profile MM sequence and non-profile MM sequence converge to the same maximum point of $\ell^P(\alpha)$.

5. Numerical Examples

Example 1. We independently simulate data from three different frailty models:

$$\lambda(t|\mathbf{X}_{ij}, \omega_i) = \omega_i \lambda_0(t) \exp\{\mathbf{X}_{ij}^\top \beta\}, \quad \omega_i \sim \begin{cases} \text{Log-normal}(0, \theta), & \theta = 0.25, \\ \text{Inverse Gaussian}(\theta, \theta^2), & \theta = 1, \\ \text{Gamma}(1/\theta, 1/\theta), & \theta = 2, \end{cases}$$

with three different sample sizes $(B, M) = \{(15, 20), (30, 13), (50, 10)\}$. The true value of regression vector β is set to be $(-2_6^\top, -1_6^\top, 1_6^\top, 2_6^\top, 3_6^\top)$ with dimension $q = 30$ and all X_i 's are generated from independent uniform distribution between 0 and 0.5. The censoring times are generated from independent uniform distribution to yield censoring proportion at around 15% or 30%. In this example, we numerically illustrate the efficiency of two proposed profile and non-profile MM algorithms under three different (Log-normal, Inverse Gaussian, and Gamma) frailty models with three different sample sizes at two censoring situations. Furthermore, we compare the performance of the two proposed profile and non-profile MM algorithms with the existing estimation approach by coxph function of the survival R package under gamma and log-normal frailty models since the coxph function of the survival R package only allows gamma and log-normal frailty. The computation time of MM algorithm can be improved by using simple off-the-shelf accelerators ([31,32]); here, we implement the accelerated profile MM and non-profile MM algorithms with the squared iterative method (SqS1). We set the stopping criterion of iteration as follows.

$$\frac{|\ell(\alpha^{(k+1)}|Y_{\text{obs}}) - \ell(\alpha^{(k)}|Y_{\text{obs}})|}{|\ell(\alpha^{(k)}|Y_{\text{obs}})| + 1} < 10^{-6}.$$

Based on 500 replications, the average values of estimated frailty and regression parameters (MLE), their biases (Bias), their empirical standard deviations (SD), and run times (T) based on three estimation methods are summarized in Tables 1–6. In general, with a sample size increase, the biases and empirical standard deviations of almost all

parameters become smaller. Small sample size $(B, M) = (15, 20)$ always causes obvious biases for frailty parameters in Log-normal and Inverse Gaussian frailty models because the number of clusters is too small. Moreover, a larger censoring proportion usually results in greater empirical standard deviations for almostly all cases. From Tables 1–4, we also observe that the existing estimation approach using coxph function is the fastest among the three methods since the survival R package is optimised using the C language. In terms of estimation accuracy, the non-profile MM algorithm performs the best for both frailty and regression parameters, almost always exhibiting smallest biases and empirical standard deviations in all situations. Even in a small sample size such as $(B, M) = (15, 20)$, the non-profile MM estimates for frailty parameter still perform well, especially for Log-normal frailty and Gamma frailty models.

Table 1. Simulation results for Log-normal frailty model in Example 1 with different sample sizes at 15% censoring.

Par.	Scenario 1: $(B, M) = (15, 20)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
	MLE	8.5716 Bias	SD	MLE	5.7928 Bias	SD	MLE	0.0586 Bias	SD
θ	0.3438	0.0938	0.3359	0.2925	0.0425	0.3173	0.2955	0.0455	0.3458
β_1	-2.2157	-0.2157	0.5034	-2.0969	-0.0969	0.5175	-2.1697	-0.1697	0.5842
β_5	-2.1316	-0.1316	0.5725	-2.1392	-0.1392	0.5567	-2.2005	-0.2005	0.5747
β_{10}	-1.0976	-0.0976	0.5523	-1.0863	-0.0863	0.5232	-1.0854	-0.0854	0.5794
β_{15}	1.0720	0.0720	0.5160	1.0126	0.0126	0.5185	1.1232	0.1232	0.5392
β_{20}	2.1871	0.1871	0.5584	2.0966	0.0966	0.5286	2.0560	0.0560	0.5217
β_{25}	3.2619	0.2619	0.5576	3.1412	0.1412	0.5455	3.2396	0.2396	0.6022
β_{30}	3.2156	0.2156	0.5775	3.1507	0.1507	0.5468	3.2781	0.2781	0.5669
Par.	Scenario 2: $(B, M) = (30, 13)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
	MLE	5.7458 Bias	SD	MLE	7.5817 Bias	SD	MLE	0.0893 Bias	SD
θ	0.3222	0.0722	0.3006	0.2520	0.0020	0.1099	0.2813	0.0313	0.1065
β_1	-2.1110	-0.1110	0.4399	-2.0930	-0.0930	0.4531	-2.0783	-0.0783	0.4743
β_5	-2.1236	-0.1236	0.4634	-2.1208	-0.1208	0.4615	-2.1232	-0.1232	0.4384
β_{10}	-1.0788	-0.0788	0.4491	-1.0528	-0.0528	0.4517	-1.0598	-0.0598	0.4877
β_{15}	1.0507	0.0507	0.4467	0.9944	-0.0056	0.4416	1.0457	0.0457	0.4539
β_{20}	2.0663	0.0663	0.4679	2.0421	0.0421	0.4665	2.1415	0.1415	0.4872
β_{25}	3.1734	0.1734	0.4824	3.0586	0.0586	0.4499	3.2054	0.2054	0.4726
β_{30}	3.1833	0.1833	0.4574	3.0666	0.0666	0.4371	3.1279	0.1279	0.4558
Par.	Scenario 3: $(B, M) = (50, 10)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
	MLE	6.2193 Bias	SD	MLE	11.4968 Bias	SD	MLE	0.1281 Bias	SD
θ	0.2683	0.0183	0.2427	0.2478	-0.0022	0.0862	0.2866	0.0366	0.0884
β_1	-2.0630	-0.0630	0.3870	-2.0491	-0.0491	0.3831	-2.0548	-0.0548	0.4021
β_5	-2.0668	-0.0668	0.3975	-2.0430	-0.0430	0.3903	-2.0340	-0.0340	0.3716
β_{10}	-1.0286	-0.0286	0.3875	-1.0327	-0.0327	0.3799	-1.0583	-0.0583	0.4058
β_{15}	1.0613	0.0613	0.4001	1.0131	0.0131	0.3915	1.0136	0.0136	0.3867
β_{20}	2.0946	0.0946	0.3984	2.0315	0.0315	0.3853	2.0697	0.0697	0.4497
β_{25}	3.1535	0.1535	0.4104	3.0739	0.0739	0.3973	3.1178	0.1178	0.4147
β_{30}	3.1266	0.1266	0.4187	3.0378	0.0378	0.4088	3.1216	0.1216	0.3875

Table 2. Simulation results for Log-normal frailty model in Example 1 with different sample sizes at 30% censoring.

Par.	Scenario 1: $(B, M) = (15, 20)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	MLE	4.8084 Bias	SD	MLE	5.7248 Bias	SD	MLE	0.0541 Bias	SD
θ	0.2712	0.0212	0.1446	0.2491	-0.0009	0.1434	0.3022	0.0522	0.1507
β_1	-2.2181	-0.2181	0.6137	-2.1211	-0.1211	0.5712	-2.1928	-0.1928	0.6084
β_5	-2.1717	-0.1717	0.6270	-2.1431	-0.1431	0.6230	-2.2133	-0.2133	0.5920
β_{10}	-1.1136	-0.1136	0.6001	-1.0786	-0.0786	0.5881	-1.0936	-0.0936	0.6158
β_{15}	1.0926	0.0926	0.6117	1.0452	0.0452	0.5603	1.0499	0.0499	0.6049
β_{20}	2.2211	0.2211	0.6234	2.0423	0.0423	0.5667	2.2068	0.2068	0.6567
β_{25}	3.2667	0.2667	0.6156	3.1369	0.1369	0.6192	3.2481	0.2481	0.5558
β_{30}	3.2490	0.2490	0.6192	3.1452	0.1452	0.6355	3.2599	0.2599	0.5502

Par.	Scenario 2: $(B, M) = (30, 13)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	MLE	5.56726 Bias	SD	MLE	7.7258 Bias	SD	MLE	0.0818 Bias	SD
θ	0.3189	0.0689	0.9271	0.2496	-0.0004	0.1150	0.2749	0.0249	0.1077
β_1	-2.0826	-0.0826	0.4974	-2.0987	-0.0987	0.4898	-2.0761	-0.0761	0.5023
β_5	-2.1135	-0.1135	0.5391	-2.0965	-0.0965	0.4950	-2.0974	-0.0974	0.5200
β_{10}	-1.0764	-0.0764	0.5065	-1.0743	-0.0743	0.5180	-1.0510	-0.0510	0.5055
β_{15}	1.0654	0.0654	0.5137	1.0433	0.0433	0.5028	1.0739	0.0739	0.5766
β_{20}	2.1250	0.1250	0.5033	2.0525	0.0525	0.4933	2.1029	0.1029	0.5137
β_{25}	3.1731	0.1731	0.5236	3.0821	0.0821	0.5105	3.1316	0.1316	0.5783
β_{30}	3.1966	0.1966	0.5207	3.0716	0.0716	0.5034	3.2236	0.2236	0.5268

Par.	Scenario 3: $(B, M) = (50, 10)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	MLE	4.7591 Bias	SD	MLE	8.4607 Bias	SD	MLE	0.1188 Bias	SD
θ	0.2729	0.0229	0.4236	0.2434	-0.0066	0.0865	0.2792	0.0292	0.0966
β_1	-2.0917	-0.0917	0.3814	-2.0714	-0.0714	0.3741	-2.1494	-0.1494	0.4376
β_5	-2.0692	-0.0692	0.4076	-2.0495	-0.0495	0.3990	-2.0617	-0.0617	0.4617
β_{10}	-1.0320	-0.0320	0.3822	-1.0341	-0.0341	0.3725	-1.0714	-0.0714	0.4174
β_{15}	1.0606	0.0606	0.3928	1.0163	0.0163	0.3843	1.0991	0.0991	0.3947
β_{20}	2.0530	0.0530	0.3851	1.9865	-0.0135	0.3773	2.0780	0.0780	0.4779
β_{25}	3.1058	0.1058	0.4000	3.0158	0.0158	0.3855	3.1531	0.1531	0.4395
β_{30}	3.1071	0.1071	0.3849	3.0182	0.0182	0.3752	3.1250	0.1250	0.4593

Table 3. Simulation results for Gamma frailty model in Example 1 with different sample sizes at 15% censoring.

Par.	Scenario 1: $(B, M) = (15, 20)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	22.3576			18.1613			0.0846		
	MLE	Bias	SD	MLE	Bias	SD	MLE	Bias	SD
θ	2.0584	0.0584	0.6494	2.1028	0.1028	0.7066	2.1197	0.1197	0.7023
β_1	-2.1343	-0.1343	0.5452	-2.1155	-0.1155	0.4974	-2.1027	-0.1027	0.6415
β_5	-2.1700	-0.1700	0.5426	-2.1303	-0.1303	0.4923	-2.1633	-0.1633	0.5354
β_{10}	-1.0561	-0.0561	0.5395	-1.0550	-0.0550	0.5505	-1.0768	-0.0768	0.6086
β_{15}	1.0515	0.0515	0.5749	1.0073	0.0073	0.5046	1.0883	0.0883	0.6006
β_{20}	2.1385	0.1385	0.542	2.1311	0.1311	0.5639	2.1685	0.1685	0.5608
β_{25}	3.1702	0.1702	0.5815	3.1928	0.1928	0.5893	3.2140	0.2140	0.5916
β_{30}	3.2185	0.2185	0.5623	3.1268	0.1268	0.6005	3.2433	0.2433	0.6320
Par.	Scenario 2: $(B, M) = (30, 13)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	23.2786			41.7191			0.1438		
	MLE	Bias	SD	MLE	Bias	SD	MLE	Bias	SD
θ	2.0934	0.0934	0.4866	2.0270	0.0270	0.4817	2.151	0.1510	0.6232
β_1	-2.1266	-0.1266	0.4648	-2.1027	-0.1027	0.4600	-2.1189	-0.1189	0.5060
β_5	-2.1197	-0.1197	0.4763	-2.0886	-0.0886	0.4549	-2.1842	-0.1842	0.4881
β_{10}	-1.0328	-0.0328	0.4703	-1.0685	-0.0685	0.4789	-1.0671	-0.0671	0.4818
β_{15}	1.0604	0.0604	0.4871	1.0212	0.0212	0.4847	1.0956	0.0956	0.4446
β_{20}	2.1313	0.1313	0.4696	2.0185	0.0185	0.4745	2.0961	0.0961	0.4848
β_{25}	3.1788	0.1788	0.4695	3.0747	0.0747	0.4954	3.1749	0.1749	0.4888
β_{30}	3.1934	0.1934	0.5120	3.1287	0.1287	0.4953	3.1908	0.1908	0.5070
Par.	Scenario 3: $(B, M) = (50, 10)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	27.4950			47.8408			0.2415		
	MLE	Bias	SD	MLE	Bias	SD	MLE	Bias	SD
θ	2.0792	0.0792	0.4091	2.0268	0.0268	0.3890	2.0905	0.0905	0.4287
β_1	-2.0896	-0.0896	0.4017	-2.0466	-0.0466	0.4406	-2.1372	-0.1372	0.4366
β_5	-2.0619	-0.0619	0.4060	-2.0388	-0.0388	0.4199	-2.1541	-0.1541	0.463
β_{10}	-1.0431	-0.0431	0.4186	-1.0190	-0.0190	0.4047	-1.0672	-0.0672	0.4016
β_{15}	1.0739	0.0739	0.4190	1.0363	0.0363	0.4025	1.0006	0.0006	0.4269
β_{20}	2.0655	0.0655	0.4141	2.0694	0.0694	0.4272	2.0781	0.0781	0.4213
β_{25}	3.1542	0.1542	0.3919	3.0680	0.0680	0.4461	3.1171	0.1171	0.4331
β_{30}	3.1085	0.1085	0.4201	3.0707	0.0707	0.4800	3.1198	0.1198	0.4200

Table 4. Simulation results for Gamma frailty model in Example 1 with different sample sizes at 30% censoring.

Par.	Scenario 1: $(B, M) = (15, 20)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	18.1609			27.0720			0.0601		
	MLE	Bias	SD	MLE	Bias	SD	MLE	Bias	SD
θ	2.0325	0.0325	0.8620	2.0844	0.0844	0.7522	2.0332	0.0332	0.6807
β_1	-1.9628	0.0372	0.8105	-2.2108	-0.2108	0.6094	-2.1984	-0.1984	0.5954
β_5	-1.9076	0.0924	0.8051	-2.1330	-0.1330	0.5744	-2.1072	-0.1072	0.6045
β_{10}	-1.0357	-0.0357	0.5701	-1.0732	-0.0732	0.5423	-1.1172	-0.1172	0.5989
β_{15}	0.9478	-0.0522	0.6464	1.0731	0.0731	0.5596	1.0585	0.0585	0.5717
β_{20}	1.9482	-0.0518	0.7783	2.0924	0.0924	0.5450	2.1451	0.1451	0.6499
β_{25}	2.9937	-0.0063	1.0379	3.1667	0.1667	0.6197	3.2788	0.2788	0.5763
β_{30}	2.9867	-0.0133	1.0485	3.2437	0.2437	0.6203	3.2631	0.2631	0.5981

Par.	Scenario 2: $(B, M) = (30, 13)$								
	Profile MM approach			Non-Profile MM Approach			Coxph		
T	16.7308			30.7202			0.0995		
	MLE	Bias	SD	MLE	Bias	SD	MLE	Bias	SD
θ	2.0880	0.0880	0.5207	2.0962	0.0962	0.4792	2.1913	0.1913	0.6610
β_1	-2.1346	-0.1346	0.5360	-2.1102	-0.1102	0.4817	-2.1127	-0.1127	0.5501
β_5	-2.0743	-0.0743	0.5299	-2.1301	-0.1301	0.5327	-2.0796	-0.0796	0.5362
β_{10}	-1.0846	-0.0846	0.5289	-1.0862	-0.0862	0.5136	-1.0690	-0.0690	0.4962
β_{15}	1.0980	0.0980	0.5242	1.0001	0.0001	0.5169	1.0932	0.0932	0.5306
β_{20}	2.1329	0.1329	0.5216	2.0354	0.0354	0.5278	2.1117	0.1117	0.5432
β_{25}	3.2089	0.2089	0.5183	3.1365	0.1365	0.5155	3.2383	0.2383	0.5587
β_{30}	3.1702	0.1702	0.5345	3.1695	0.1695	0.5495	3.2301	0.2301	0.5380

Par.	Scenario 3: $(B, M) = (50, 10)$								
	Profile MM Approach			Non-Profile MM Approach			Coxph		
T	16.3097			29.1778			0.1647		
	MLE	Bias	SD	MLE	Bias	SD	MLE	Bias	SD
θ	2.0999	0.0999	0.4480	2.0462	0.0462	0.3831	2.0948	0.0948	0.3700
β_1	-2.1151	-0.1151	0.4605	-2.0508	-0.0508	0.4187	-2.1324	-0.1324	0.4554
β_5	-2.0767	-0.0767	0.4197	-2.0876	-0.0876	0.4615	-2.1001	-0.1001	0.4394
β_{10}	-1.0963	-0.0963	0.4629	-1.0712	-0.0712	0.4393	-1.0513	-0.0513	0.4306
β_{15}	1.0443	0.0443	0.4313	0.9910	-0.0090	0.4234	2.0347	0.0347	0.4476
β_{20}	2.0813	0.0813	0.4775	2.0330	0.0330	0.4630	2.1039	0.1039	0.4538
β_{25}	3.1089	0.1089	0.4848	3.0978	0.0978	0.4660	3.1829	0.1829	0.4484
β_{30}	3.1216	0.1216	0.4805	3.1462	0.1462	0.5005	3.1834	0.1834	0.5308

Table 5. Simulation results for Inverse Gaussian model in Example 1 with different sample sizes at 15% censoring.

Par.	Scenario 1: $(B, M) = (15, 20)$					
	Profile MM Approach			Non-Profile MM Approach		
T	MLE	3.0369 Bias	SD	MLE	7.9534 Bias	SD
θ	0.7060	-0.2940	0.6920	0.7130	-0.2870	0.6511
β_1	-2.1516	-0.1516	0.5664	-2.1002	-0.1002	0.4646
β_5	-2.1783	-0.1783	0.5366	-2.1062	-0.1062	0.5173
β_{10}	-1.1028	-0.1028	0.5526	-1.0881	-0.0881	0.4719
β_{15}	1.1284	0.1284	0.5523	0.9661	-0.0339	0.5069
β_{20}	2.2162	0.2162	0.5225	1.9742	-0.0258	0.4936
β_{25}	3.2591	0.2591	0.5269	2.9877	-0.0123	0.5332
β_{30}	3.2139	0.2139	0.5553	3.0247	0.0247	0.5250
Par.	Scenario 2: $(B, M) = (30, 13)$					
	Profile MM Approach			Non-Profile MM Approach		
T	MLE	5.2872 Bias	SD	MLE	27.2494 Bias	SD
θ	0.7537	-0.2463	0.5623	0.9178	-0.0822	0.4438
β_1	-2.1515	-0.1515	0.4592	-2.1441	-0.1441	0.4455
β_5	-2.1300	-0.1300	0.4646	-2.1146	-0.1146	0.4907
β_{10}	-1.0873	-0.0873	0.4743	-1.0830	-0.0830	0.4316
β_{15}	1.0681	0.0681	0.4755	1.0223	0.0223	0.4445
β_{20}	2.1454	0.1454	0.4600	2.0591	0.0591	0.4625
β_{25}	3.2323	0.2323	0.4972	3.1031	0.1031	0.4807
β_{30}	3.2223	0.2223	0.4698	3.0855	0.0855	0.4907
Par.	Scenario 3: $(B, M) = (50, 10)$					
	Profile MM Approach			Non-Profile MM Approach		
T	MLE	54.9905 Bias	SD	MLE	36.7796 Bias	SD
θ	0.8461	-0.1539	0.4143	0.9724	-0.0276	0.3678
β_1	-2.0933	-0.0933	0.4056	-2.0688	-0.0688	0.4135
β_5	-2.1509	-0.1509	0.4375	-2.0664	-0.0664	0.4318
β_{10}	-1.0878	-0.0878	0.3992	-1.0790	-0.0790	0.3824
β_{15}	1.0437	0.0437	0.4217	0.9728	-0.0272	0.4111
β_{20}	2.1137	0.1137	0.4077	2.0257	0.0257	0.4005
β_{25}	3.1511	0.1511	0.4311	3.0709	0.0709	0.4129
β_{30}	3.1846	0.1846	0.4137	3.0540	0.0540	0.4219

Table 6. Simulation results for Inverse Gaussian model in Example 1 with different sample sizes at 30% censoring.

Par.	Scenario 1: $(B, M) = (15, 20)$					
	Profile MM Approach			Non-Profile MM Approach		
T	MLE	9.7444 Bias	SD	MLE	16.1049 Bias	SD
θ	0.5343	-0.4657	0.6665	0.7017	-0.2983	0.6431
β_1	-2.1693	-0.1693	0.6153	-2.1795	-0.1795	0.6054
β_5	-2.2288	-0.2288	0.5688	-2.1677	-0.1677	0.6601
β_{10}	-1.0228	-0.0228	0.5690	-1.1006	-0.1006	0.5843
β_{15}	1.1635	0.1635	0.5542	1.0577	0.0577	0.6157
β_{20}	2.2779	0.2779	0.6479	2.0902	0.0902	0.6168
β_{25}	3.2790	0.2790	0.6760	3.1711	0.1711	0.6672
β_{30}	3.3969	0.3969	0.6867	3.1939	0.1939	0.6412

Par.	Scenario 2: $(B, M) = (30, 13)$					
	Profile MM Approach			Non-Profile MM Approach		
T	MLE	22.7787 Bias	SD	MLE	22.0124 Bias	SD
θ	0.6564	-0.3436	0.5263	1.0203	0.0203	0.5297
β_1	-2.1801	-0.1801	0.5395	-2.1915	-0.1915	0.4992
β_5	-2.0213	-0.0213	0.4727	-2.1508	-0.1508	0.5142
β_{10}	-1.1390	-0.1390	0.5202	-1.0678	-0.0678	0.5140
β_{15}	1.0824	0.0824	0.5276	0.9845	-0.0155	0.5005
β_{20}	2.1354	0.1354	0.4639	2.0252	0.0252	0.5088
β_{25}	3.1809	0.1809	0.6421	3.1362	0.1362	0.5233
β_{30}	3.2375	0.2375	0.5542	3.0828	0.0828	0.5183

Par.	Scenario 3: $(B, M) = (50, 10)$					
	Profile MM Approach			Non-Profile MM Approach		
T	MLE	20.9400 Bias	SD	MLE	32.6620 Bias	SD
θ	0.8502	-0.1498	0.3856	0.9809	-0.0191	0.3556
β_1	-2.1046	-0.1046	0.4252	-2.0688	-0.0688	0.4124
β_5	-2.1536	-0.1536	0.4224	-2.1123	-0.1123	0.4065
β_{10}	-1.0597	-0.0597	0.3893	-1.0542	-0.0542	0.3806
β_{15}	1.0692	0.0692	0.3973	1.0067	0.0067	0.3841
β_{20}	2.1129	0.1129	0.4162	2.0184	0.0184	0.4005
β_{25}	3.1770	0.1770	0.4159	3.0489	0.0489	0.4005
β_{30}	3.2102	0.2102	0.4277	3.0832	0.0832	0.4120

Example 2. In this example, we simulated 200 realizations consisting of $B = 50$ clusters and $M = 6$ subjects in each cluster from the frailty model:

$$\lambda(t|\mathbf{X}_{ij}, \omega_i) = \omega_i \lambda_0(t) \exp\{\mathbf{X}_{ij}^\top \boldsymbol{\beta}\}, \tag{28}$$

where the frailty terms are $\omega_i \sim \text{Gamma}(1/\theta, 1/\theta)$ with $\theta = 0.5$ for $i = 1, \dots, B$, $\boldsymbol{\beta} = (1, 3, 0_{46}^\top, 2, 4)^\top$, $q = 50$, $\lambda_0(t) = 5$. The X_i were marginally standard normal, and the correlation between X_i and X_j is $\rho^{|i-j|}$ with $\rho = 0.25, 0.75$, respectively. The censoring times were generated from uniform distribution to yield a censoring proportion around 30%. In this simulation experiment, the utility of our proposed profile MM and non-profile MM estimation methods for regularized estimation was illustrated in a sparse high-dimensional regression model (21) with three different penalties (L_1 , MCP, and SCAD). The model error (ME) and relative model error (RME), which is the ratio of the model error of the regularized estimator and that of the ordinary maximum likelihood estimator, are calculated to evaluate the estimation accuracy. Based on 200 replications, we report the median of relative model errors (MRME) and the average number of correctly and

incorrectly identified zero coefficients in Table 7 in which the column labeled “Correct” presents the average restricted only to the true zero coefficients, while the column labeled “Incorrect” depicts the average of coefficients erroneously set to 0. From Table 4, we can observe that the proposed profile and non-profile MM algorithms mesh very well with MCP and SCAD penalties and provide good results in both parameter estimation and variable selection, especially for MCP penalty. For L_1 penalty, we observe that it meshes well with profile and non-profile MM algorithms under lower correlation case at $\rho = 0.25$ but tends to yield biased estimates and inaccurate variable selection results when the correlation between X_i and X_j becomes stronger. Furthermore, we report the average values of estimated parameters (MLE), their biases (Bias), and their empirical standard deviations (SD) under MCP and SCAD penalties based on 200 repetition in Table 8. It can be observed that both profile and non-profile MM methods perform similarly well in varying ρ and different penalties, always showing small biases and empirical standard deviations.

Table 7. The median of relative model errors for gamma frailty model by L_1 , MCP, and SCAD penalties with sample size $(B, M) = (50, 6)$ based on 200 realizations in Example 2.

Penalty	MRME	Zeros		MRME	Zeros	
		Correct	Incorrect		Correct	Incorrect
		$\rho = 0.25$			$\rho = 0.75$	
<i>Profile MM method</i>						
L_1	0.159	45.775	0	0.163	44.39	0
MCP ($\gamma = 3$)	0.091	46	0	0.066	46	0
SCAD ($\gamma = 3.7$)	0.143	45.915	0	0.101	45.965	0
<i>Non-profile MM method</i>						
L_1	0.089	45.74	0	0.188	44.225	0
MCP ($\gamma = 3$)	0.051	46	0	0.085	46	0
SCAD ($\gamma = 3.7$)	0.077	45.885	0	0.106	45.88	0

Table 8. The average values of estimated parameters (MLE), their biases (Bias), and their empirical standard deviations (SD) under MCP and SCAD penalties in Example 2.

Penalty	Par.	Profile MM Approach			Non-Profile MM Approach		
		MLE	Bias	SD	MLE	Bias	SD
MCP ($\rho = 0.25$)	θ	0.4835	−0.0165	0.1120	0.4820	−0.0180	0.1398
	β_1	0.9993	−0.0007	0.0950	0.9988	−0.0012	0.1028
	β_2	3.0090	0.0090	0.1984	2.9856	−0.0144	0.1971
	β_{49}	1.9940	−0.0060	0.1513	1.9791	−0.0209	0.1512
	β_{50}	3.9697	−0.0303	0.2498	3.9868	−0.0132	0.2566
SCAD ($\rho = 0.25$)	θ	0.4967	−0.0033	0.1267	0.4800	−0.0200	0.1335
	β_1	1.0133	0.0133	0.0960	0.9864	−0.0136	0.0975
	β_2	3.0167	0.0167	0.2023	2.9467	−0.0532	0.1644
	β_{49}	2.0183	0.0183	0.1417	1.9816	−0.0184	0.1255
	β_{50}	4.0315	0.0315	0.2657	3.9300	−0.0700	0.2334
MCP ($\rho = 0.75$)	θ	0.4886	−0.0114	0.1258	0.4842	−0.0158	0.1480
	β_1	0.9774	−0.0226	0.1331	1.0071	0.0071	0.1369
	β_2	3.0252	0.0252	0.2020	2.9743	−0.0257	0.2570
	β_{49}	2.0083	0.0083	0.1831	1.9857	−0.0143	0.2203
	β_{50}	4.0134	0.0134	0.2606	4.0026	0.0026	0.3055
SCAD ($\rho = 0.75$)	θ	0.4869	−0.0131	0.1277	0.4839	−0.0161	0.1341
	β_1	1.0119	0.0119	0.1364	1.0152	0.0152	0.1560
	β_2	2.9815	−0.0185	0.2063	3.0035	0.0035	0.2422
	β_{49}	2.0012	0.0012	0.2238	1.9984	−0.0016	0.2052
	β_{50}	4.0119	0.0119	0.2633	4.0021	0.0021	0.2855

6. Real Data Analysis

Alzheimer Disease's (AD) is the most common dementia that causes progressive memory and other body function losses. According to Alzheimer Association, Alzheimer's is the sixth leading cause of death in the United States. Many researches have studied the development of AD from Mild Cognitive Impairment (MCI). The authors of [33] analyzed the AUC score of whether MCI will transfer to AD using LASSO penalized logistic regression. The authors of [34] applied cognitive scores, ApoE genotype, and CSF biomarkers to predict the transition time from MCI to AD. Moreover, there are research studies such as [35] that used image data to analyze the transition time to dementia. In this paper, we will apply clinical data and selected SNP genotypes to conduct feature selection under three different frailty models (Gamma, Inverse Gaussian, and Log-normal) using SCAD and MCP penalties. The dataset was obtained from ADNI database (adni.loni.usc.edu). In this dataset, 267 people were recorded as cognitively normal (CN) during the first visit. Among these people, 78 of them were diagnosed with MCI before the last visit. Eventually, we observed 22 people transferred from the MCI stage to dementia. To predict these two conversion times, 19 clinical predictors were applied mainly based on the age, marriage status, education level, and test score such as Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-Cog) and Functional Activities Questionnaire (FAQ). Other than these predictors, we also included genotypes of SNP from GWAS such as ApoE- ϵ 4, which have a relationship with early onset dementia or late-onset dementia; 132 covariates are selected for model training.

In general, using the same notation from simulation, we have $(B, M) = (276, 2)$. Individuals are independent, and two events of the same individual are grouped into the same cluster that share the same frailty. The dataset contains the following information. For individual $(i = 1, 2, \dots, 276)$ and event $(j = 1, 2)$, t_{ij} is minimum value of event time Y_{ij} and censoring time C_{ij} . As shown by Table 9, censoring time C_{ij} in this dataset is the time difference between the first observation date of certain event (CN or MCI) and last observation date of this patient. The event time is the state transition time, and Y_{i1} is the time difference between the first observation date of this patient in state CN and date of transition from CN to MCI. If no transition is observed, censoring occurs where $\delta_{i1} = 0$ and $t_{i1} = C_{i1}$, otherwise, $\delta_{i1} = 1$ and $t_{i1} = Y_{i1}$. Similarly, Y_{i2} is the time difference between the first observation date of this patient in state MCI and the date of transition from MCI to Dementia. If no transition is observed, $\delta_{i2} = 0$ and $t_{i2} = C_{i2}$, otherwise, $\delta_{i2} = 1$ and $t_{i2} = Y_{i2}$. $\mathbf{X}_{ij}^T = (X_{ij1}, \dots, X_{ijp})$ denote the potential covariates (such as SNP genotypes and clinical predictors) where $p = 132$. According to data description, it is reasonable to assume that censoring time C_{ij} is independent of event time Y_{ij} . Moreover, the two events from each individual are considered to be highly associated due to common genetic factors and/or certain habits. Thus, the frailty model is suitable for analyzing this dataset.

Three frailty models are applied for variable selection using SCAD and MCP penalties. According to the results presented by Tables 10–12, around 70 covariates were selected for Gamma Frailty and Log-normal Frailty model, and 27 covariates were selected by an Inverse Gaussian model. Similar covariates were selected by SCAD or MCP penalty for both Profile MM and Non-profile MM under the same model, which verifies the result from simulation study. For selected covariates, the ApoE- ϵ 4 is selected as a negative effect to the survival time as expected. In addition to ApoE- ϵ 4, single SNPs such as rs2333227 and rs669 from the MPO and A2M gene are associated with the development of Alzheimer's Disease. We can observe similar results from [36]. However, many selected SNPs are not recognized from other studies, and it needs to be analyzed further whether these SNPs can help to identify the risk of Alzheimer's.

We use the model with the lowest BIC score for prediction by constructing prognostic index in order to see whether the selected covariates can help identify the individual with low risk or high risk. Let \mathbf{X} be a collection of \mathbf{X}_{ij}^T . The full dataset is divided into 10 parts ($\mathbf{X}^{(k)}$ ($k = 1, 2, \dots, 10$)). Similarly to the cross-validation process, without the part of dataset $\mathbf{X}^{(k)}$, covariates are selected using BIC criteria and the vector of parameters $\hat{\beta}_k$ is estimated

correspondingly. The risk factor $\hat{r}_k = \mathbf{X}^{(k)} \hat{\beta}_k$ ($k = 1, 2, \dots, 10$) is calculated using dataset $\mathbf{X}^{(k)}$ without considering the individuals' Latent variable. Individuals with estimated risk factors below median are classified into to low risk group, and and individual with estimated risk above median is classified into the high risk group. The survival plots for estimated high risk group and low risk group are constructed based on their true survival times. Figure 1 shows that the selected model has conducted a good estimation for the transition time from stage CN to stage MCI. Due to the lack of data for the second transition, their confidence intervals are overlapped. Therefore, the prediction result, especially from stage CN to stage MCI, has shown good estimation using the corresponding model, and two risk groups are distinguished by predicted risk levels.

Table 9. Illustration of survival years and censoring date for some patients from the ADNI dataset.

Patient ID	From CN to MCI				
	First Observation Date of State CN	Date of Transition to MCI	Last Date of Observation	t_{ij} (yr)	δ_{ij}
011_S_0002	8 September 2005	26 September 2012	18 October 2017	7.05	1
011_S_0021	24 October 2005	-	27 November 2017	12.10	0
100_S_0035	8 November 2005	8 December 2010	8 December 2010	5.08	1
131_S_0123	7 February 2006	23 February 2012	10 February 2016	6.05	1
127_S_0259	28 March 2006	9 April 2014	22 September 2017	8.04	1

Patient ID	From MCI to Dementia				
	First Observation Date of State MCI	Date of Transition to Dementia	Last Date of Observation	t_{ij} (yr)	δ_{ij}
011_S_0002	26 September 2012	-	18 October 2017	5.06	0
011_S_0021	-	-	-	-	-
100_S_0035	8 December 2010	-	8 December 2010	0.00	0
131_S_0123	23 February 2012	12 February 2014	10 February 2016	1.97	1
127_S_0259	9 April 2014	10 April 2015	22 September 2017	1.00	1

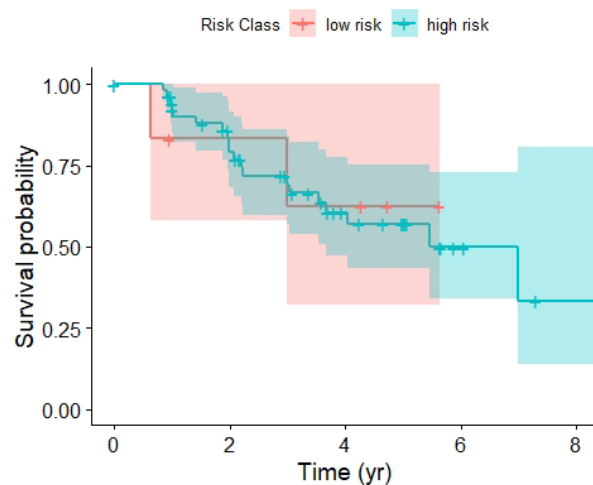


Figure 1. Survival curves of the transition times from CN to MCI and from MCI to AD for high-risk and low risk groups constructed using prognostic index.

Table 10. Training results for Gamma Frailty Model.

Penalty	Profile MM		Non-Profile MM	
	BIC	No. of Non-Zero β	BIC	No. of Non-Zero β
MCP	1154.70	70	1154.75	70
SCAD	1150.59	69	1150.82	69

Table 11. Training results for Inverse Gaussian Model.

Penalty	Profile MM		Non-Profile MM	
	BIC	No. of Non-Zero β	BIC	No. of Non-Zero β
MCP	1198.76	27	1198.82	27
SCAD	1203.97	27	1204.50	27

Table 12. Training results for Log-normal Frailty Model.

Penalty	Profile MM		Non-Profile MM	
	BIC	No. of Non-Zero β	BIC	No. of Non-Zero β
MCP	1149.10	71	1148.86	72
SCAD	1149.10	71	1148.86	71

7. Discussion

The profile and non-profile MM algorithms are proposed for high-dimension regression analysis with clustered failure time data where a general frailty is used to model within-cluster dependence and the penalty such as the SCAD and MCP is used for inducing sparsity. The proposed methods can separate the high-dimensional minorizing function into a sum of univariate functions after a sequence of minorization steps. These approaches avoid matrix inversion and provide a toolkit for developing more efficient algorithms in a broad range of in statistical optimization problems. Meshing well with sparsity penalties such as SCAD or MCP, the two regularized MM algorithms are further shown to exhibit certain numerical advantages in sparse high-dimensional regression analyses. The shared frailty model only represents a special and relatively simple model among the widely used frailty models for multivariate survival data. For example, the standard shared frailty model assumes that all subjects in the same cluster share a common frailty. This assumption can be relaxed to the correlated frailty terms among subjects in the same cluster. Correlated frailty models present the limitation that shared frailty models may only be used to fit positively correlated event times. Furthermore, frailty is assumed to be time-constant. However, unobserved heterogeneity may also be time dependent, which can be explained by an unobserved random process that unfolds over time. Based on this idea, several approaches have been proposed such as diffusion processes modeling or Levy processes modeling for frailty. As an approach based on birth-death Poisson or simpler, piecewise constant, frailty models have recently been proposed. It would be worthwhile to extend the proposed two MM algorithms in these applications. Lastly, spatially correlated survival data present another important and useful setting where the proposed MM algorithms can be further extended to accommodate.

Author Contributions: Conceptualization, J.X.; Data curation, X.H., J.X. and Y.Z.; Formal analysis, X.H. and Y.Z.; Investigation, Y.Z.; Methodology, X.H. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to the Editor and referees for many helpful comments. This work has been partially supported by the National Natural Science Foundation of China (11901515).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Clayton, D.G. A model for association in bivariate life tables and its application in epidemiologic studies of familial tendency in chronic disease incidence. *Biometrika* **1978**, *65*, 141–151. [[CrossRef](#)]
2. Clayton, D.G.; Cuzick, J. Multivariate generalizations of the proportional hazards model. *J. R. Stat. Soc. Ser. B* **1985**, *47*, 82–117. [[CrossRef](#)]
3. Oakes, D. Bivariate survival models induced by frailties. *J. Am. Stat. Assoc.* **1989**, *84*, 487–493. [[CrossRef](#)]
4. Zeng, D.; Chen, Q.; Ibrahim, J. Gamma frailty transformation models for multivariate survival times. *Biometrika* **2009**, *96*, 277–291. [[CrossRef](#)] [[PubMed](#)]
5. Andersen, P.K.; Klein, J.P.; Knudsen, K.M.; Palacios, R.T. Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics* **1997**, *53*, 1475–1484. [[CrossRef](#)] [[PubMed](#)]
6. Cox, D.R. Regression models and life tables (with discussion). *J. R. Stat. Soc. Ser. B* **1972**, *34*, 187–220.
7. Klein, J.P. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **1992**, *48*, 795–806. [[CrossRef](#)]
8. Nielsen, G.G.; Gill, R.D.; Andersen, P.K.; Sorensen, T.I.A. A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Stat.* **1992**, *19*, 25–43.
9. Balan, T.A.; Putter, H. A tutorial on frailty models *Stat. Methods Med Res.* **2020**, *29*, 3424–3454. [[CrossRef](#)]
10. Do Ha, I.; Lee, Y. A review of h-likelihood for survival analysis. *Jpn. J. Stat. Data Sci.* **2021**, *4*, 1157–1178.
11. Duchateau, L.; Janssen, P. *The Frailty Model*; Springer Science & Business Media: New York, NY, USA, 2007.
12. Glidden, D.V.; Vittinghoff, E. Modelling clustered survival data from multicentre clinical trials. *Stat. Med.* **2004**, *23*, 369–388. [[CrossRef](#)] [[PubMed](#)]
13. Dahlqvist, E.; Pawitan, Y.; Sjölander, A. Regression standardization and attributable fraction estimation with between-within frailty models for clustered survival data. *Stat. Methods Med Res.* **2019**, *28*, 462–485. [[CrossRef](#)] [[PubMed](#)]
14. Manda, S.O. A nonparametric frailty model for clustered survival data. *Commun. Stat. Methods* **2011**, *40*, 863–875. [[CrossRef](#)]
15. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic Decomposition by Basis Pursuit. *Siam J. Sci. Comput.* **1998**, *20*, 33–61. [[CrossRef](#)]
16. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
17. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
18. Frank, L.E.; Friedman, J.H. A Statistical View of Some Chemometrics Regression Tools. *Technometrics* **1993**, *35*, 109–135. [[CrossRef](#)]
19. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [[CrossRef](#)]
20. Fan, J.; Li, R. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [[CrossRef](#)]
21. Zhang, C.-H. Nearly Unbiased Variable Selection Under Minimax Concave Penalty. *Ann. Stat.* **2010**, *38*, 894–942. [[CrossRef](#)]
22. Hunter, D.R.; Lange, K. A tutorial on MM algorithms. *Am. Stat.* **2004**, *58*, 30–37. [[CrossRef](#)]
23. Lange, K.; Hunter, D.R.; Yang, I. Optimization transfer using surrogate objective functions (with discussions). *J. Comput. Graph. Stat.* **2000**, *9*, 1–20.
24. Becker, M.P.; Yang, I.; Lange, K. EM algorithms without missing data. *Stat. Methods Med. Res.* **1997**, *6*, 38–54. [[CrossRef](#)] [[PubMed](#)]
25. Huang, X.F.; Xu, J.F.; Tian, G.L. On profile MM algorithms for gamma frailty survival models. *Stat. Sin.* **2019**, *29*, 895–916. [[CrossRef](#)]
26. Lange, K.; Zhou, H. MM algorithms for geometric and signomial programming. *Math. Program. Ser. B* **2014**, *143*, 339–356. [[CrossRef](#)]
27. Hunter, D.R.; Li, R. Variable selection using MM algorithms. *Ann. Stat.* **2005**, *33*, 1617–1642. [[CrossRef](#)]
28. Schwarz, C. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
29. Craven, P.; Wahba, G. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **1979**, *31*, 377–403. [[CrossRef](#)]
30. Lange, K. *Numerical Analysis for Statisticians*, 2nd ed.; Statistics and Computing; Springer: New York, NY, USA, 2010.
31. Varadhan, R.; Roland, C. Simple and globally convergent methods for accelerating the convergence of any EM algorithms. *Scand. J. Stat.* **2008**, *35*, 335–353. [[CrossRef](#)]
32. Zhou, H.; Alexander, D.; Lange, K. A quasi-Newton acceleration for high-dimensional optimization algorithms. *Stat. Comput.* **2011**, *21*, 261–273. [[CrossRef](#)]
33. Ye, J.P.; Farnum, M.; Yang, E.; Verbeeck, R.; Lobanov, V.; Raghavan, V.; Novak, G.; DiaBernardo, A.; Narayan, V.A. Sparse learning and stability selection for predicting mci to ad conversion using baseline adni data. *BMC Neurol.* **2012**, *12*, 46. [[CrossRef](#)] [[PubMed](#)]
34. Da, X.; Toledo, J.; Toledo, J.; Zee, J.; Wolk, D.A.; Xie, S.X.; Ou, Y.; Shacklett, A.; Parmpi, P.; Shaw, L.; et al. Integration and relative value of biomarkers for prediction of mci to ad progression: Spatial patterns of brain atrophy, cognitive scores, apoe genotype, and csf markers. *Neuroimage Clin.* **2014**, *4*, 164–173. [[CrossRef](#)] [[PubMed](#)]

35. Kong, D.H.; Ibrahim, J.G.; Lee, E.; Zhu, H.T. Flcrm: Functional linear cox regression model. *Biometrics* **2018**, *74*, 109–117. [[CrossRef](#)] [[PubMed](#)]
36. Zappia, M.; Manna, I.; Serra, P.; Cittadella, R.; Andreoli, V.; La Russa, A.; Annesi, F.; Spadafora, P.; Romeo, N.; Nicoletti, G.; et al. Increased risk for alzheimer disease with the interaction of mpo and a2m polymorphisms. *Arch. Neurol.* **2004**, *61*, 341–344. [[CrossRef](#)] [[PubMed](#)]