mathematics

MDPI

Article

Regression Analysis of Multivariate Interval-Censored Failure Time Data under Transformation Model with Informative Censoring

Mengzhu Yu [1] and Mingyue Du [1,2,*]

[1] School of Mathematics, Jilin University, Changchun 130012, China
[2] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China
* Correspondence: mingydu@jlu.edu.cn

Abstract: We consider a regression analysis of multivariate interval-censored failure time data where the censoring may be informative. To address this, an approximated maximum likelihood estimation approach is proposed under a general class of semiparametric transformation models, and in the method, the frailty approach is employed to characterize the informative interval censoring. For the implementation of the proposed method, we develop a novel EM algorithm and show that the resulting estimators of the regression parameters are consistent and asymptotically normal. To evaluate the empirical performance of the proposed estimation procedure, we conduct a simulation study, and the results indicate that it performs well for the situations considered. In addition, we apply the proposed approach to a set of real data arising from an AIDS study.

Keywords: case K interval-censored data; informative censoring; semiparametric transformation model; sieve approach

MSC: 62N02; 62H12; 62G20

check for updates

Citation: Yu, M.; Du, M. Regression Analysis of Multivariate Interval-Censored Failure Time Data under Transformation Model with Informative Censoring. *Mathematics* **2022**, *10*, 3257. https://doi.org/10.3390/math10183257

Academic Editors: Yiqiang Zhan and Xingrong Liu

Received: 31 July 2022
Accepted: 30 August 2022
Published: 7 September 2022

1. Introduction

In this paper, we consider a regression analysis of multivariate interval-censored failure time data where the censoring may be informative. Interval-censored data arise when the failure time of interest is known or observed only to belong to an interval instead of being observed exactly (Finkelstein, 1986 [1]; Sun, 2006 [2]). It is apparent that one can treat right-censored data as a special case of interval-censored data. Multivariate interval-censored data occur if a failure time study involves more than one related failure time of interest for which only interval-censored data are available. Among others, one can often face multivariate interval-censored data in epidemiological studies and clinical trials. Informative censoring occurs if the censoring mechanism or the underlying process generating observations is related to the failure times of interest (Kalbfleisch and Prentice, 2002 [3]; Sun, 2006 [2]).

An example of informative censoring is given by a clinical trial or periodic study on a failure event such as death for which some symptoms may occur before the occurrence of the event. For the situation, the study subject may tend to pay more clinical visits when the symptoms occur rather than following the pre-specified schedule. Many authors have pointed out' that with informative censoring, the analysis that ignores it could lead to serious biased estimators or analysis results (Wang et al., 2010 [4]; Sun, 2006 [2]; Zhang et al., 2005 [5], 2007 [6]). For example, Sun (1999) [7] studied the issue for univariate current status data, a special case of interval-censored data where the observed interval includes either zero or infinity, and showed that the analysis could yield misleading results if the informative censoring is ignored or treated to be non-informative censoring.

A large amount of literature has been established for the regression analysis of multivariate interval-censored failure time data or their special cases, multivariate current status data and bivariate interval-censored data (Chen et al., 2007 [8], 2009 [9], 2013 [10]; Goggins and Finkelstein, 2000 [11]; Shen, 2015 [12]; Tong et al., 2008 [13]; Wang et al., 2008 [14]; Zeng et al., 2017 [15]; Zhang et al., 2009 [16]; Zhou et al., 2017 [17]). For this, three types of methods are commonly used, including the copula model approach, the marginal model-based approach and the frailty model-based method. The first employs various copula models to characterize the relationship between correlated failure times of interest (Wang et al., 2008 [14]; Zhang et al., 2009 [16]), and among others, Sun and Ding (2019) [18] discussed this for bivariate cases under the framework of the two-parameter Archimedean copula model. The second mainly focuses on the marginal distribution and puts no assumption on the correlation between the failure times of interest (Wei et al. 1989 [19]). The authors who developed such methods include Chen et al. (2007) [8], Chen et al. (2013) [10] and Tong et al. (2008) [13].

The frailty model-based approach generally employs the frailty or latent variable to model the correlation between the correlated failure times. It has the advantage of allowing one to directly estimate the correlation. One main shortcoming of most of the existing methods for multivariate interval-censored data is that they assume independent or non-informative interval censoring, and it is apparent that this may not hold in practice as discussed above. In this paper, we will adopt the frailty model-based approach to develop a new estimation procedure that allows for dependent or informative interval censoring.

Several authors have considered a regression analysis of univariate informatively interval-censored failure time data. For example, Zhang (2005) [5], Wang et al. (2010) [4] and Wang et al. (2018) [20] investigated the problem for current status data, case II interval-censored data and case K interval-censored data, respectively. Case II means that each study subject is observed only twice, while case K refers to the situation where each subject is observed at a sequence of observation times, which is much more general than others (Sun, 2006) [2]. As mentioned above, most of the existing methods for multivariate interval-censored data apply only to the situation with independent interval censoring, except Yu et al. (2022) [21]. Yu et al. (2022) [21] only considered case II interval-censored data under the additive hazards model. In this paper, the focus will be on case K multivariate interval-censored data with informative censoring and the proposed methods apply to much more general situations than Yu et al. (2022) [21].

More specifically, in Section 2, some notation and assumptions will be first introduced as well as the data structures. In the proposed method, we will focus on the case where the failure time of interest marginally follows a general class of semiparametric transformation models. The proposed approximated maximum likelihood estimation procedure will be presented in Section 3, and for the implementation of the proposed method, a novel EM algorithm will be developed. The asymptotic properties of the resulting estimators of the regression parameters will be given in Section 4. Section 5 will present some simulation results obtained from a study performed to evaluate the performance of the proposed method, and they indicate that it performs as expected. In Section 6, we apply the proposed methodology to a set of real data arising from an AIDS clinical trial, and Section 7 contains some discussion and concluding remarks.

2. Assumptions and Background

In this section, we first introduce some notation and background and then describe the model and data structure. Suppose that there is a failure time study consisting of n independent subjects and concerning M failure events of interest that may be related. Define T_{im} to be the failure time of interest and X_{im} a p-dimensional vector of covariates both related to the ith subject and the event m. Furthermore, for each subject, suppose that there exists a sequence of potential observation times $U_{i0} = 0 < U_{i1} < \ldots < U_{iK_i^*}$ and a follow-up or stopping time τ_i, where K_i^* denotes the number of potential observations, $i = 1, \ldots, n$. For simplicity, we assume that for each subject, the observation times for different failure events are the same and the proposed method below can be easily generalized to more general situations.

For subject i, define the point process $\tilde{N}_i(t) = \sum_{j=1}^{K_i^*} I(U_{ij} \leq t)$, describing the observation process on the subject that jumps only at the observation times, $i = 1, \ldots, n$. Note that for the situation considered here, we have $M + 1$ processes, the M underlying failure time processes of interest and the observation process $\tilde{N}_i(t)$, and as mentioned above, the focus below will be on the case where they may be related (Ma et al., 2015 [22]; Wang et al., 2016 [23]; Zhang et al., 2007 [6]). To describe their relationships and the possible covariate effects on them, we assume that there exists a vector of latent variables b_i and another latent variable u_i with mean zero, and given X_{im}, Z_{im}, b_i and u_i, the cumulative hazard function of T_{im} has the form

$$\Lambda_{im}(t|X_{im}, Z_{im}, b_i, u_i) = G_m\{\exp(X_{im}^T \beta_{xm} + Z_{im}^T b_i + u_i \beta_{um})\Lambda_m(t)\}\,. \tag{1}$$

Here, $G_m(.)$ is a known strictly increasing transformation function, $\Lambda_m(.)$ is an unknown baseline cumulative hazard function, Z_{im} contains 1 and part of the covariates X_{im}, and $\beta_m = (\beta_{xm}^T, \beta_{um})^T$ denotes the vector of unknown regression parameters.

For the observation process, it will be assumed that $\tilde{N}_i(t)$ is a non-homogeneous Poisson process satisfying

$$\lambda_{ih}(t|X_i, u_i) = \lambda_{0h}(t)\,\exp\left(\sum_{m=1}^{M} X_{im}^T \alpha_m + u_i\right) = \lambda_{0h}(t)\,\exp\left(X_i^T \alpha + u_i\right) \tag{2}$$

for the intensity function given X_{im} and u_i. Here, λ_{0h} is an unknown continuous baseline intensity function, $X_i = (X_{i1}^T, \ldots, X_{iM}^T)^T$, and $\alpha^T = (\alpha_1^T, \ldots, \alpha_M^T)$, which is a vector of regression parameters as β_m. In the following, it will be assumed that given X_{im}, b_i and u_i, $T_{i1}, \ldots, T_{iM}$ are independent, and given X_{im} and u_i, T_i and $\tilde{N}_i$ are independent. Moreover, τ_i is independent of T_i and $\tilde{N}_i$. We point out that models (1) and (2) with $u_i = 0$ have been commonly used in the analysis of failure-time data (Klein and Moeschberger, 2003 [24]) and the analysis of event history data (Cook and Lawless, 2007 [25]), respectively. The parameter β_{um} denotes the degree of the correlation between the failure-time process and the observation process, and they will be independent if $\beta_{um} = 0$.

The semiparametric transformation model (1) with $b_i = 0$ and $u_i = 0$ is quite general and can give many specific models. In particular, one can express it as a class of frailty-induced transformations

$$G_m(x) = -\log \int_0^{\infty} \exp(-xt)\, f_m(t) dt\,.$$

In the above, $f_m(t)$ denotes the density function of a frailty variable with support $[0, \infty]$. By setting $f_m(t)$ to be the gamma density with mean 1 and variance r_m, it gives the class of logarithmic transformations $G_m(x) = r_m^{-1} \log(1 + r_m x)$ with $r_m > 0$ (Chen et al., 2002 [26]). In particular, it yields the proportional odds model with $r_m = 1$ or $G_m(x) = \log(1 + x)$ and gives the proportional hazards model with $r_k = 0$ or $G_m(x) = x$.

To describe the observed data, define $\delta_{imj} = I(U_{ij-1} < T_{im} \leq U_{ij})$, indicating if the failure time T_{im} belongs to the interval $(U_{ij-1}, U_{ij}]$. In the following, it will be assumed that the observed data have the form

$$\mathcal{O} = \{\, O_i = (\tau_i, X_{im}, Z_{im}, U_{ij}, \delta_{imj},\ m = 1, \ldots, M, j = 1, \ldots, K_i^*);\ i = 1, \ldots, n \,\}\,,$$

where $K_i^* = \tilde{N}_i(\tau_i)$. That is, we observe case K interval-censored data (Sun, 2006).

3. Maximum Likelihood Estimation

3.1. Estimation Procedure

Now, we discuss inference about models (1) and (2), and for this, we will propose a two-step or an approximate maximum likelihood estimation procedure by following Huang and Wang (2004) [27] and Wang et al. (2016) [23]. More specifically, we will first consider the estimation of model (2) and then estimation of $\phi^T = (\beta_1^T, \ldots, \beta_M^T)$, the parameter of interest. The first step will be based on the following two facts. One is that one can easily show that K_i^* follows the Poisson distribution with the mean

$$\Lambda_{ih}(\tau_i; X_i, u_i) = \Lambda_{0h}(\tau_i) \exp(X_i^T \alpha + u_i)$$

given X_i and u_i, $i = 1, \ldots, n$. The other is that the observation times $U_{i1}, \ldots, U_{iK_i^*}$ can be seen as the order statistics of a set of i.i.d. random variables with the density function

$$\pi(t) = \frac{\lambda_{0h}(t) \exp(X_i^T \alpha + u_i)}{\Lambda_{0h}(\tau_i) \exp(X_i^T \alpha + u_i)} I(0 \le t \le \tau_i) = \frac{\lambda_{0h}(t)}{\Lambda_{0h}(\tau_i)} I(0 \le t \le \tau_i) \,.$$

One can see that the function above does not depend on neither X_i nor u_i, which suggests that the function $\Lambda_{0h}(t)$ can be estimated by

$$\hat{\Lambda}_{0h}(t) = \prod_{s(l)>t} (1 - \frac{d_{(l)}}{R_{(l)}}) \,.$$

In the above, the $s_{(l)}$'s denote the ordered and distinct values of observation times $\{U_{ik}\}$, $d_{(l)}$ the number of the observation times equal to $s_{(l)}$, and $R_{(l)}$ the number of observation times satisfying $U_{ik} \le s_{(l)} \le \tau_i$ among all subjects.

Under the assumptions above, it is easy to show that $E[K_i^*; X_{i.}, u_i, \tau_i] = \Lambda_{0h}(\tau_i) \exp$ $(X^T \alpha + u_i)$. This yields

$$E_{u_i}[E[K_i^* \Lambda_{0h}^{-1}(\tau_i); X_{im}, u_i, \tau_i]] = E(e^{u_i}) \exp(X_i^T \alpha) \,,$$

and a class of estimating equations

$$\sum_{i=1}^{n} \omega_i X_i (K_i^* \hat{\Lambda}_{0h}^{-1}(\tau_i) - E(e^{u_i}) \exp(X_i^T \alpha)) = 0$$

for estimation of α_m, $m = 1, \ldots, M$ with the ω_i's being some weights. Let $\hat{\alpha}_m$ denote the estimator of α_m given by the estimating equations above, which suggests that one can naturally estimate u_i by

$$\hat{u}_i = \log \left\{ \frac{K_i^*}{\hat{\Lambda}_{0h}(\tau_i) \exp(X_i^T \alpha)} \right\}.$$

Now, consider estimation of ϕ as well as model (1). For this, note that if the u_i's were known, it would be natural to maximize the likelihood function

$$\begin{aligned} L_n(\phi, \Lambda, \gamma \,|\, u_i's) = \quad & \prod_{i=1}^{n} \int \prod_{m=1}^{M} \prod_{j=1}^{K_i^*} \Big\{ \exp(-G_m[\int_0^{U_{i,j-1}} \exp\{x_{im}^{*T} \beta_m + Z_{im}^T b_i\} d\Lambda_m(t)]) - \\ & \exp(-G_m[\int_0^{U_{ij}} \exp\{x_{im}^{*T} \beta_m + Z_{im}^T b_i\} d\Lambda_m(t)]) \Big\}^{\Delta_{imj}} \\ & \exp(-G_m[\int_0^{U_{iK_i^*}} \exp\{x_{im}^{*T} \beta_m + Z_{im}^T b_i\} d\Lambda_m(t)])^{1 - \sum_{j=1}^{K_i^*} \Delta_{imj}} f_b(b_i|\gamma) db_i \,, \end{aligned}$$

where $\Lambda = (\Lambda_1, \ldots, \Lambda_M)$, $x_{im}^* = (X_{im}^T, \hat{u}_i)^T$, and f_b denotes the density function of the b_i's assumed to be known up to a vector of parameters γ. Define $L_{im} = max\{U_{ij} : U_{ij} < T_{im}, j = 0, \ldots, K_i^*\}$ and $R_{im} = min\{U_{ij} : U_{ij} \ge T_{im}, j = 1, \ldots, K_i^* + 1\}$, where $U_{i0} = 0$ and $U_{i,K_i^*+1} = \infty$. Then, $(L_{im}, R_{im}]$ represents the shortest time interval that brackets T_{im} and the likelihood function above can be rewritten as

$$\begin{aligned} L_n(\phi, \Lambda, \gamma \,|\, u_i's) = \quad & \prod_{i=1}^{n} \int \prod_{m=1}^{M} \Big\{ \exp(-G_m[\int_0^{L_{im}} \exp(x_{im}^{*T} \beta_m + Z_{im}^T b_i) d\Lambda_m(t)]) \\ & - \exp(-G_m[\int_0^{R_{im}} \exp(x_{im}^{*T} \beta_m + Z_{im}^T b_i) d\Lambda_m(t)]) \Big\} f_b(b_i \mid \gamma) db_i \,. \end{aligned}$$

By following Huang and Wang (2004) [27] and others, it is natural to estimate ϕ and Λ by their values that maximize the approximated likelihood function $L_n(\phi, \Lambda, \gamma \mid \hat{u}_i' s)$.

For the maximization of $L_n(\phi, \Lambda, \gamma \mid \hat{u}_i' s)$, note that it involves the unknown functions Λ_m's and integrations. To deal with them, for the former, we propose to adopt the non-parametric approach. More specifically, for each $m = 1, \ldots, M$, let $0 = t_{m0} < t_{m1} < \ldots < t_{mk_m} < \infty$ denote the ordered sequence of all L_{im} and R_{im} with $R_{im} < \infty$ and assume that Λ_m is a step function that jumps only at the t_{mq}'s with the jump sizes λ_{mq}'s. Then, $L_n(\phi, \Lambda, \gamma \mid u_i' s)$ can be expressed as

$$L_n(\phi, \Lambda, \gamma \mid u_i' s) = \prod_{i=1}^{n} \int \prod_{m=1}^{M} \left\{ \exp\left(-G_m \left[\sum_{t_{mq} \le L_{im}} \exp\{x_{im}^{*T} \beta_m + Z_{im}^T b_i\} \lambda_{mq} \right] \right) - \exp\left(-G_m \left[\sum_{t_{mq} \le R_{im}} \exp\{x_{im}^{*T} \beta_m + Z_{im}^T b_i\} \lambda_{mq} \right] \right) \right\} f_b(b_i \mid \gamma) db_i \,. \quad (3)$$

In the following, we will develop an EM algorithm for the maximization with the focus on the situation where f_b is a multivariate normal distribution with the covariance matrix $\Sigma(\gamma)$ depending on the q-dimensional unknown parameter γ. The algorithm is valid for other distributions and some comments on this will be given below. It is worth to point out that as mentioned above, the idea discussed above has been used by Huang and Wang (2004) [27] and Wang et al. (2016) [23], among others. However, the problem discussed here is different or much more general than the existing literature.

3.2. EM Algorithm

In this subsection, we will develop an EM algorithm for the maximization of $L_n(\phi, \Lambda, \gamma \mid \hat{u}_i' s)$, and for this, we will first discuss the data augmentation. Let the ξ_{im}'s denote the random sample of size n from the density $f_m(t)$. Then, we can rewrite the observed likelihood function as

$$L_n(\phi, \Lambda, \gamma \mid u_i' s) = \prod_{i=1}^{n} \int \left\{ \prod_{m=1}^{M} \int_{\xi_{im}} \exp\{-\xi_{im} \sum_{t_{mq} \le L_{im}} \lambda_{mq} \exp(x_{im}^{*T} \beta_m + Z_{im}^T b_i)\} \left[1 - \exp\{-\xi_{im} \sum_{L_{im} < t_{mq} \le R_{im}} \lambda_{mq} \exp(x_{im}^{*T} \beta_m + Z_{im}^T b_i)\} \right]^{I(R_{im} < \infty)} f_m(\xi_{im}) d\xi_{im} \right\} \times f_b(b_i \mid \gamma) db_i \,. \quad (4)$$

Moreover, let the W_{imq}'s denote the random sample of size n from the Poisson distributions with means $\xi_{im} \lambda_{mq} \exp(x_{im}^{*T} \beta_m + Z_{im}^T b_i)$ given ξ_{im}, and define $A_{im} = \sum_{t_{mq} \le L_{im}} W_{imq}$ and $B_{im} = I(R_{im} < \infty) \sum_{L_{im} < t_{mq} \le R_{im}} W_{imq}$ such that

$$P(A_{im} = 0, B_{im} > 0 | L_{im}, R_{im}, x_{im}^*) = \exp\{-\xi_{im} \sum_{t_{mq} \le L_{im}} \lambda_{mq} \exp(x_{im}^{*T} \beta_m + Z_{im}^T b_i)\} \left[1 - \exp\{-\xi_{im} \sum_{L_{im} < t_{mq} \le R_{im}} \lambda_{mq} \exp(x_{im}^{*T} \beta_m + Z_{im}^T b_i)\} \right]^{I(R_{im} < \infty)} .$$

It is easy to see that the maximization of (4) is equivalent to maximizing the likelihood function based on the data $(L_{im}, R_{im}, x_{im}^*, A_{im} = 0, B_{im} > 0)$ $(i = 1, \ldots, n; m = 1, \ldots, M)$. Based on this, for the development of the EM algorithm, it is natural to use the W_{imq}'s, ξ_{im}'s and b_i's to augment the observed data. As a consequence, one can derive the resulting pseudo complete data log-likelihood function as

$$
\begin{aligned}
l_c(\phi, \Lambda, \gamma \mid u_i's) = \sum_{i=1}^{n} \Bigg\{ \sum_{m=1}^{M} \Bigg(\sum_{q=1}^{k_m} I(t_{mq} \le R_{im}^*) \Big[W_{imq} \log\{\xi_{im}\lambda_{mq} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i)\} \\
- \xi_{im}\lambda_{mq} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i) - \log(W_{imq}!) \Big] + \log f_m(\xi_{im}) \Bigg) \\
- \frac{d_i}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_i(\gamma)| - \frac{b_i^T \Sigma_i(\gamma)^{-1} b_i}{2} \Bigg\},
\end{aligned} \tag{5}
$$

where $R_{im}^* = L_{im} I(R_{im} = \infty) + R_{im} I(R_{im} < \infty)$.

Now, we consider the E-step of the EM algorithm. At the $(s+1)$th iteration and given $(\phi^s, \Lambda^s, \gamma^s)^T$, we need to determine

$$
\begin{aligned}
Q(\phi, \Lambda, \gamma | \phi^s, \Lambda^s, \gamma^s) = E[l_c(\phi, \Lambda, \gamma \mid u_i's, \mathcal{O}, \phi^s, \Lambda^s, \gamma^s)] \\
= \sum_{i=1}^{n} \Bigg\{ \sum_{m=1}^{M} \Bigg(\sum_{q=1}^{k_m} I(t_{mq} \le R_{im}^*) \Big[E\Big[W_{imq} \log\Big\{ \xi_{im}\lambda_{mq} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i) \Big\} \\
- \xi_{im}\lambda_{mq} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i) \Big] - E\big[\log(W_{imq}!)\big] \Big] + \log f_m(\xi_{im}) \Bigg) \\
- \frac{d_i}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_i(\gamma)| - \frac{E\big[b_i^T \Sigma_i(\gamma)^{-1} b_i\big]}{2} \Bigg\}
\end{aligned}
$$

under the multivariate normal distribution with the covariance matrix $\Sigma_i(\gamma)$. To calculate the conditional expectations $E[\xi_{im} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i)]$, $E[W_{imq}]$ and $E[b_i^T \Sigma^{-1}(\gamma) b_i]$ given the observed data, we need to employ the joint density of ξ_{im} and b_i given the observed data, which is proportional to

$$
\begin{aligned}
\prod_{m=1}^{M} \Bigg[\exp\Bigg\{ -\xi_{im} \sum_{t_{mq} \le L_{im}} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i)\lambda_{mq} \Bigg\} \\
- I(R_{im} < \infty) \exp\Bigg\{ -\xi_{im} \sum_{t_{mq} \le R_{im}} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i)\lambda_{mq} \Bigg\} \Bigg] \\
\times f_m(\xi_{im}) (2\pi)^{-d_i/2} |\Sigma_i(\gamma)|^{-1/2} \exp\{ -\frac{b_i^T \Sigma_i(\gamma)^{-1} b_i}{2} \}.
\end{aligned}
$$

Note that the conditional expectation of W_{imq} for $t_{mq \le R_{im}^*}$ given $\xi_{im} (m = 1, \ldots, M)$, b_i and the observed data is given by

$$
\hat{E}(W_{imq} | \xi_{im}, b_i) = I(L_{im} < t_{mq} \le R_{im}) \frac{\lambda_{mq}\xi_{im} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i)}{1 - \exp\{ -\sum_{L_{im} < t_{mq'} \le R_{im}} \lambda_{mq'} \xi_{imq'} \exp(x_{im}^{*T}\beta + Z_{im}^T b_i) \}}.
$$

In the M-step, we can employ the Newton–Raphson method to update β_m based on the equation

$$
\sum_{i=1}^{n} \sum_{m=1}^{M} \sum_{q=1}^{k_m} I(t_{mq} \le R_{im}^*) \hat{E}(W_{imq}) \Bigg\{ x_{im}^* - \frac{\sum_{i'=1}^{n} I(t_{mq} \le R_{i'm}^*) \hat{E}\{\xi_{i'm} \exp(x_{i'm}^{*T}\beta_m + Z_{i'm}^T b_{i'})\} x_{i'm}^*}{\sum_{i'=1}^{n} I(t_{mq} \le R_{i'm}^*) \hat{E}\{\xi_{i'm} \exp(x_{i'm}^{*T}\beta_m + Z_{i'm}^T b_{i'})\}} \Bigg\} = 0.
$$

For estimation of λ_{mq}, we have the closed form expression

$$
\lambda_{mq} = \frac{\sum_{i=1}^{n} I(t_{mq} \le R_{im}^*) \hat{E}(W_{imq})}{\sum_{i=1}^{n} I(t_{mq} \le R_{im}^*) \hat{E}\{\xi_{im} \exp(x_{im}^{*T}\beta_m + Z_{im}^T b_i)\}}, \tag{6}
$$

for $q = 1, \ldots, k_m$ and $m = 1, \ldots, M$. To estimate γ, one can maximize $-\log\|\Sigma_i(\gamma)\| - \hat{E}\{b_i^T \Sigma_i^{-1}(\gamma) b_i\}$ with the Σ_i's given by $\Sigma = n^{-1} \sum_{i=1}^{n} \hat{E}(b_i^T b_i)$.

Now, we summarize the EM algorithm described above as follows.

Step 1. Choose initial estimates $\phi^{(0)}, \Lambda^{(0)}, \gamma^{(0)}$ of ϕ, Λ, γ, respectively.
Step 2. In the $(s+1)$th iteration, calculate $\hat{E}[\xi_{im} \exp(x_{im}^{*T} \beta^{(s)} + Z_{im}^T b_i)]$, $\hat{E}[W_{imq}]$ and $\hat{E}[b_i^T \Sigma^{-1}(\gamma^{(s)}) b_i]$ by using, for example, the Gaussian quadrature method. $s = 0, 1, 2, \ldots$
Step 3. Update $\Lambda^{(s+1)}$ by (6) with current $\phi^{(s)}, \gamma^{(s)}$, and then update $\gamma^{(s+1)}$ by maximizing $-\log \|\Sigma_i(\gamma)\| - \hat{E}\{b_i^T \Sigma_i^{-1}(\gamma) b_i\}$. In addition, estimate $\phi^{(s+1)}$ by employing the one-step Newton–Raphson method.
Step 4. Repeat Steps 2–3 until the convergence such that the absolute difference of the log-likelihood values between two consecutive iterations is less than a given positive value ε such as 10^{-3}.

4. Asymptotic Properties

Let $\hat{\theta} = \left(\hat{\beta}_1^{\mathrm{T}}, \ldots, \hat{\beta}_M^{\mathrm{T}}, \hat{\gamma}^{\mathrm{T}}, \hat{\Lambda}_1, \ldots, \hat{\Lambda}_M\right)^{\mathrm{T}}$ denote the estimator of $\theta = (\beta_1^{\mathrm{T}}, \ldots, \beta_M^{\mathrm{T}}, \gamma^{\mathrm{T}}, \Lambda_1, \ldots, \Lambda_M)^{\mathrm{T}}$ defined above and $\theta_0 = (\beta_{01}^{\mathrm{T}}, \ldots, \beta_{0M}^{\mathrm{T}}, \gamma_0^{\mathrm{T}}, \Lambda_{01}, \ldots, \Lambda_{0M})^{\mathrm{T}}$ the true value of θ. Define $\zeta_0 = \left(\beta_{01}^{\mathrm{T}}, \ldots, \beta_{0M}^{\mathrm{T}}, \gamma_0^{\mathrm{T}}\right)^{\mathrm{T}}$ and $\hat{\zeta} = \left(\hat{\beta}_1^{\mathrm{T}}, \ldots, \hat{\beta}_M^{\mathrm{T}}, \hat{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}}$. In this section, we will establish the asymptotic properties of $\hat{\theta}$, and for this, we first describe the regularity conditions needed.

Define

$$Q_m^*(t, b; \beta_m, \Lambda_m) = \exp\left(-G_m\left[\int_0^t \exp\{\beta_m^T x_m^* + b^T Z_m\}\right]\right) d\Lambda_m(s)\,,$$

$D_m(U_m, b; \beta_m, \Lambda_m) = \sum_{l=0}^{K_m} \Delta_{ml}\{Q_m^*(U_{ml}, b; \beta_m, \Lambda_m) - Q_m^*(U_{m,l+1}, b; \beta_m, \Lambda_m)\}$, $U_m = (U_{m1}, \ldots, U_{m,K_m})$, $\Delta_{ml} = I(U_{ml} \leqslant T_m < U_{m,l+1})$, and $p(b|\gamma) = (2\pi)^{-d/2} |\Sigma(\gamma)|^{-1/2} \exp(-b^{\mathrm{T}} \Sigma(\gamma)^{-1} b/2)$. For the asymptotic properties of $\hat{\theta}$, we need the following regularity conditions.

Condition 1. *The true value ζ_0 belongs to a known compact set $\mathbb{A} \bigotimes \mathbb{B} \bigotimes \mathbb{C}$, where $\mathbb{A}$ denotes a compact set of $\mathbb{R}^{\mathrm{pM}}$, $\mathbb{B}$ a compact set in $\mathbb{R}^{\mathrm{M}}$, and $\mathbb{C}$ a compact set of $\mathbb{R}^{\mathrm{q}}$ in the domain of γ such that $\Sigma(\gamma)$ is a positive-definite matrix with eigenvalues bounded away from zero and ∞. In addition, the true value $\Lambda_{0m}(\cdot)$ is continuously differentiable with positive derivatives in $[0, \tau_m]$.*

Condition 2. *The covariate vector X_m and Z_k are bounded in $[0, \tau_m]$.*

Condition 3. *For the transformation function G_m, assume that it is twice continuously differentiable on $[0, \infty)$ with $G_m(0) = 0$, $G_m'(x) > 0$ and $G_m(\infty) = \infty$.*

Condition 4. *Assume that $\sup_{\gamma \in \mathbb{C}} \int_b g(b) p^{(j)}(b|\gamma) db < \infty$ for any smooth function $g(\cdot)$ and $j = 0, 1, 2$. Here, $p^{(j)}(b|\gamma)$ denotes the jth derivative of $p(b|\gamma)$ with respect to γ.*

Condition 5. *If there exists a vector u and some constants $v_m, m = 1, \ldots, M$ such that*

$$\left(u^T \frac{\partial}{\partial \zeta} + \sum_{m=1}^{M} v_m \frac{\partial}{\partial y_m}\right)\Bigg|_{(\zeta, y_1, \ldots, y_M) = (\zeta_0, \Lambda_{10}(c_1), \ldots, \Lambda_{M0}(c_M))} \cdot \log \int_b \prod_{m=1}^{M} D_m(U_m, b; \beta, \beta_u, \Lambda_m) p(b \mid \gamma) \mathrm{d}b = 0$$

for each of these values, then $u = 0_{pM+M+q}$ and $v_m = 0$. In addition, 0_{pM+M+q} denotes a $(pM + M + q)$-dimensional vector of zeros.

Condition 6. *Assume that $P(\tau_m \geq \tau_0, \exp(u) > 0) > 0$ for the follow-up time τ_m and latent variable u, where τ_0 denotes the longest study time and the variance of $\exp(u)$ is bounded and there exists a positive small constant $\epsilon > 0$ such that $\exp(u) > \epsilon$ almost surely. Moreover, for τ_m and u, the function $F(s) = E[\exp(u) I(\tau_m \geq s)]$ is continuous for $s \in [0, \tau_0]$.*

Note that Conditions 1 and 2 are standard conditions in survival analysis, and it is easy to check that Condition 3 on the transformation function holds for the logarithmic family $G_r(x) = r^{-1}\log(1+rx)(r \geqslant 0)$ and the Box–Cox family $G_d(x) = d^{-1}\left\{(1+x)^d - 1\right\}(d \geqslant 0)$. Moreover, Condition 4 holds for modeling multivariate data with frailty models, and Condition 5 is required for the identifiability of the model. In addition, Condition 6 describes the relationship between the latent variable u and the parameters of interest. Most of the conditions above are purely for technique purposes and hold in general, in particular, for periodic follow-up studies.

Let $\|\cdot\|$ denote the Euclidean norm and define $\mathbb{P}f = \int f(x)\mathrm{d}P(x)$ and $\mathbb{P}_n f = n^{-1}\sum_{i=1}^{n} f(X_i)$ for a function f and a random variable X with distribution P. The following two theorems give the asymptotic properties of $\hat{\theta}$.

Theorem 1. *Suppose that Conditions 1–6 hold. Then, as $n \to \infty$, we have that $\|\hat{\zeta} - \zeta_0\| + \sum_{m=1}^{M} \sup_{t\in[0,\tau_m]} |\hat{\Lambda}_m(t) - \Lambda_{0m}(t)| \to 0$ almost surely.*

Theorem 2. *Suppose that Conditions 1–6 hold. Then, as $n \to \infty$, we have that $\sqrt{n}(\hat{\zeta}_n - \zeta_0) \to^d N(0, I_0^{-1})$, where $I_0 = \mathbb{P}\{\tilde{l}(\theta_0)\tilde{l}(\theta_0)^T\}$ with $\tilde{l}(\theta_0)$ given in the Appendix A.*

We will sketch the proof for the results described above in Appendix A. For inference about ζ, it is apparent that one needs to estimate the covariance matrix, and for this, one can see from Appendix A that it would be difficult to derive a consistent estimator of I_0. Thus, we propose to employ the profile likelihood approach to estimate the covariance matrix of $\hat{\zeta}$ (Murphy & van der Vaart, 2000) [28]. Specifically, let $\mathcal{C}$ denote the set of all step functions with nonnegative jumps at t_{mq} and define $\mathrm{pl}_n(\zeta) = \max_{\Lambda\in\mathcal{C}} \log L_n(\zeta,\Lambda)$, the profile log-likelihood. Then, one can estimate the covariance matrix of $\hat{\zeta}$ by the negative inverse of the matrix with the (j,k)th element given by

$$\frac{\mathrm{pl}_n(\hat{\zeta}) - \mathrm{pl}_n(\hat{\zeta} + h_n e_k) - \mathrm{pl}_n(\hat{\zeta} + h_n e_j) + \mathrm{pl}_n(\hat{\zeta} + h_n e_k + h_n e_j)}{h_n^2}.$$

In the above, e_j denotes the jth canonical vector in $\mathcal{R}^d$ and h_n is a constant of order $n^{-1/2}$. Note that to calculate $pl_n(\zeta)$ for each ζ, one can reuse the proposed EM algorithm with β held fixed and the only step in the EM algorithm is to explicitly evaluate $\hat{E}(W_{imq})$ and $\hat{E}(\xi_{im})$ to update λ_m using above. The iteration converges quickly in general by setting $\hat{\lambda}_m$ to be the initial value.

5. A Simulation Study

In this section, we give some of the simulation results obtained from a study performed to evaluate the finite sample performance of the proposed method with the focus on estimation of the β_m's. In the study, we considered the situation with $M = 2$ correlated failure times of interest and two covariates. For the covariates, it was assumed that the first covariate follows the Bernoulli distribution with the success probability of 0.5 and the second covariate, the uniform distribution over $(0,1)$. To generate the true failure times, we first set Z_{im} to be one and generated the latent variables b_i's from the normal distribution $N(0,\sigma^2)$ with $\sigma^2 = 0.25$ and the latent variables u_i's from the normal distribution with the mean 0 and variance 1. Then, given the X_{im}'s, Z_{im}'s, b_i's and u_i's, the T_{i1}'s and T_{i2}'s were generated under model (1) with $G_m(x) = r_m^{-1}\log(1+r_m x)$, $\Lambda_1(t) = \log(1+0.5t)$ and $\Lambda_2(t) = 0.5t$ for $r_1 = r_2 = 0$, $r_1 = r_2 = 0.5$ or $r_1 = r_2 = 1$, respectively.

For the generation of the observation process and the observed data, we first assumed that the $\tau_i' s$ follow the uniform distribution over the interval [2, 3] and generated the K_i^*'s from the Poisson distribution with the mean

$$\Lambda_{ih}(\tau_i; X_i, u_i) = \tau_i \exp(X_i^T\alpha + u_i)$$

given the X_i's and u_i's. Note that in the above, we took $\Lambda_{0h}(t) = t$ and $\alpha_m = 1$. Given the K_i^*'s, we took $U_{i1} < \ldots < U_{iK_i^*}$ to be the order statistics of the random sample of size K_i^* from the uniform distribution over $(0, \tau_i)$. In the following, we considered two sets of true values, $(0,0,0)^T$ and $(0.5,0.5,0.5)^T$, for the regression parameters $\beta_1 = (\beta_{x11}, \beta_{x12}, \beta_{u1})^T$ and $\beta_2 = (\beta_{x21}, \beta_{x22}, \beta_{u2})^T$, corresponding to T_{i1} and T_{i2}, respectively. The results given below are based on $n = 200$ or 400 with 1000 replications.

Table 1 gives the results on the estimation of β_1 and β_2 given by the proposed estimation procedure with $r_1 = r_2 = 0$, $r_1 = r_2 = 0.5$ and $r_1 = r_2 = 1$. Here, we calculated the estimated bias (Bias) given by the average of the estimates minus the true value, the sample standard error (SSE) of the estimates, the average of the estimated standard errors (ESE) and the 95% empirical coverage probability (CP). The results suggest that the proposed estimator of the regression parameters seems to be unbiased and the variance estimation based on the profile likelihood approach also seems to be reasonable. Furthermore, the results on the empirical coverage probabilities indicate that the normal approximation to the distribution of the proposed estimator of the regression parameters appears to be appropriate. In addition, the results got better in general with the increasing sample size, as expected.

Table 1. Simulation results on estimation of β with the b_i's generated from the normal distribution.

	True Value	$r_1 = r_2 = 0$				$r_1 = r_2 = 0.5$				$r_1 = r_2 = 1$			
		Bias	SSE	SEE	CP	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
						$n = 200$							
β_{x11}	0	0.030	0.248	0.252	0.945	0.006	0.289	0.288	0.951	0.031	0.251	0.259	0.945
β_{x12}	0	0.028	0.412	0.422	0.957	0.021	0.479	0.476	0.953	0.029	0.418	0.452	0.957
β_{u1}	0	−0.024	0.141	0.143	0.938	−0.026	0.164	0.163	0.945	−0.024	0.142	0.142	0.938
β_{x21}	0	0.044	0.236	0.236	0.953	0.001	0.278	0.276	0.948	0.044	0.242	0.252	0.953
β_{x22}	0	0.001	0.390	0.422	0.949	0.011	0.459	0.460	0.949	0.001	0.403	0.429	0.949
β_{u2}	0	−0.013	0.138	0.135	0.960	−0.026	0.159	0.158	0.945	−0.014	0.138	0.140	0.960
β_{x11}	0.5	0.040	0.233	0.246	0.944	0.024	0.291	0.290	0.955	0.033	0.314	0.317	0.950
β_{x12}	0.5	0.025	0.378	0.410	0.955	−0.007	0.485	0.501	0.951	0.020	0.508	0.533	0.948
β_{u1}	0.5	0.010	0.135	0.142	0.956	−0.026	0.165	0.168	0.942	−0.022	0.177	0.179	0.948
β_{x21}	0.5	0.028	0.227	0.238	0.949	0.017	0.280	0.282	0.949	0.062	0.307	0.316	0.947
β_{x22}	0.5	0.011	0.358	0.436	0.953	0.011	0.466	0.484	0.956	0.054	0.489	0.498	0.949
β_{u2}	0.5	−0.005	0.135	0.131	0.943	−0.027	0.160	0.160	0.950	−0.022	0.176	0.179	0.958
						$n = 400$							
β_{x11}	0	0.009	0.171	0.178	0.952	0.010	0.200	0.204	0.954	−0.001	0.226	0.228	0.949
β_{x12}	0	0.001	0.280	0.283	0.952	−0.003	0.327	0.329	0.954	0.017	0.370	0.374	0.946
β_{u1}	0	−0.023	0.097	0.104	0.949	−0.036	0.112	0.113	0.944	0.036	0.127	0.130	0.942
β_{x21}	0	0.009	0.162	0.180	0.941	0.007	0.191	0.194	0.946	0.009	0.219	0.222	0.942
β_{x22}	0	0.001	0.260	0.287	0.948	0.014	0.314	0.321	0.951	0.003	0.358	0.366	0.955
β_{u2}	0	−0.025	0.093	0.096	0.942	−0.037	0.108	0.109	0.948	−0.036	0.124	0.123	0.946
β_{x11}	0.5	0.005	0.158	0.163	0.949	0.012	0.187	0.190	0.949	0.031	0.215	0.22	0.942
β_{x12}	0.5	0.019	0.253	0.254	0.954	0.023	0.299	0.305	0.957	0.010	0.344	0.363	0.949
β_{u1}	0.5	−0.025	0.090	0.096	0.935	−0.031	0.105	0.105	0.937	−0.034	0.121	0.120	0.944
β_{x21}	0.5	0.012	0.153	0.166	0.952	0.008	0.183	0.185	0.942	0.024	0.211	0.219	0.946
β_{x22}	0.5	0.035	0.239	0.259	0.956	0.007	0.289	0.302	0.949	0.011	0.333	0.348	0.953
β_{u2}	0.5	−0.024	0.089	0.094	0.941	−0.032	0.104	0.104	0.94	−0.03	0.120	0.115	0.943

As mentioned before, the proposed estimation procedure can be applied to any distribution for the latent variables b_i's. To see this, we repeated the study above, except that we generated the b_i's from the uniform distribution over $(-1, 1)$, and Table 2 presents the obtained results on the estimation of β_1 and β_2 with $n = 200$ and $r_1 = r_2 = 0$. One can see that they are similar to those given in Table 1 and again suggest that the proposed approach seems to work well for the situations considered. To see the performance of the proposed approach with different types of covariates, we also repeated the study giving Table 1, except that both covariates were assumed to follow the standard normal distribution and give the obtained results with $n = 200$ and $r_1 = r_2 = 0$ in Table 3. They indicate that the proposed estimation procedure seems to be robust to different types of covariates.

Table 2. Simulation results on estimation of β with the b_i's generated from the uniform distribution and $r_1 = r_2 = 0$.

	True Value	Bias	SSE	ESE	CP
β_{x11}	0	−0.005	0.238	0.240	0.948
β_{x12}	0	−0.004	0.404	0.404	0.953
β_{u1}	0	0.001	0.134	0.134	0.946
β_{x21}	0	−0.001	0.223	0.232	0.950
β_{x22}	0	−0.004	0.374	0.412	0.945
β_{u2}	0	0.003	0.126	0.126	0.957
β_{x11}	0.5	0.031	0.223	0.231	0.948
β_{x12}	0.5	0.035	0.361	0.401	0.956
β_{u1}	0.5	−0.021	0.129	0.134	0.947
β_{x21}	0.5	0.022	0.215	0.231	0.948
β_{x22}	0.5	0.054	0.342	0.412	0.947
β_{u2}	0.5	−0.012	0.129	0.129	0.955

Table 3. Simulation results on estimation of β with the covariates generated from the normal distribution and $r_1 = r_2 = 0$.

	True Value	Bias	SSE	ESE	CP
β_{x11}	0	−0.003	0.127	0.125	0.949
β_{x12}	0	0.003	0.126	0.126	0.953
β_{u1}	0	−0.006	0.142	0.145	0.950
β_{x21}	0	0.003	0.120	0.119	0.938
β_{x22}	0	−0.001	0.120	0.119	0.941
β_{u2}	0	0.004	0.133	0.141	0.957
β_{x11}	0.5	0.037	0.142	0.141	0.946
β_{x12}	0.5	0.027	0.141	0.141	0.954
β_{u1}	0.5	−0.011	0.146	0.152	0.948
β_{x21}	0.5	0.037	0.137	0.135	0.945
β_{x22}	0.5	0.033	0.138	0.134	0.939
β_{u2}	0.5	−0.008	0.143	0.148	0.949

Note that in the proposed estimation procedure, it has been assumed that the observation process $\tilde{N}_i(t)$ is a non-homogeneous Poisson process and one may be interested in the performance of the proposed method if this assumption is not true. To see this, we repeated the study giving Table 1, except that the $\tilde{N}_i(t)$'s were assumed to be mixed Poisson processes with the intensity function

$$\lambda_{ih}(t|X_i, u_i) = v_i\, \lambda_{0h}(t)\, \exp\left(X_i^T \alpha + u_i\right)$$

given the v_i's, where the v_i's were generated from the gamma distribution. Table 4 presents the results on the estimation of β_1 and β_2 given by the proposed approach with $n = 200$ and $r_1 = r_2 = 0$, and they indicate that the approach seems to be robust with the processes $\tilde{N}_i(t)$'s.

Table 4. Simulation results on estimation of β with mixed Poisson observation processes and $r_1 = r_2 = 0$.

	True Value	Bias	SSE	ESE	CP
β_{x11}	0	0.002	0.246	0.250	0.952
β_{x12}	0	0.013	0.418	0.429	0.951
β_{u1}	0	0.001	0.137	0.136	0.951
β_{x21}	0	−0.002	0.233	0.241	0.948
β_{x22}	0	0.006	0.387	0.419	0.946
β_{u2}	0	−0.001	0.130	0.131	0.949
β_{x11}	0.5	0.037	0.233	0.241	0.948
β_{x12}	0.5	0.075	0.376	0.400	0.949
β_{u1}	0.5	−0.034	0.133	0.136	0.945
β_{x21}	0.5	0.037	0.226	0.257	0.953
β_{x22}	0.5	0.038	0.357	0.435	0.949
β_{u2}	0.5	−0.029	0.132	0.142	0.952

For the initial value in the EM algorithm here, we set $\phi^{(0)} = 0$, $\lambda_{mq}^{(0)} = \frac{1}{k_m}$, $q = 1, \ldots, k_m$, and $\gamma^{(0)} = 0.25$. It is worth to point out that we did try other initial values and the proposed EM algorithm seems to be robust with respect to the selection of the initial values. In other words, we did not encounter non-convergence issue in the simulation study. We also considered some other setups, including multivariate cases and the case with more than one covariate and obtained similar results.

6. An Application

In the section, the estimation procedure proposed in the previous sections is applied to the set of bivariate interval-censored data arising from an AIDS clinical trial, AIDS Clinical Trial Group 181, described in Goggins and Finkelstein (2000) [11]. The study concerns the opportunistic infection cytomegalovirus (CMV) and examined the study patients periodically. At each clinical visit or observation, among other information, the blood and urine samples were collected and tested to detect the existence of the CMV virus in the sample, which is also commonly referred to as the shedding of the virus. In addition, for each patient, the CD4 count, indicating the status of a person's immune system and being commonly used to measure the stage of HIV infection, was also recorded at the entry time. For the analysis here, we are mainly interested in if the baseline CD4 account, the indicator of the initial stage of HIV disease, is related to the CMV shedding risk in either blood or urine.

The data set consists of 204 subjects, and they belong to two groups based on their baseline CD4 counts, either less than 75 or otherwise. More specifically, the two groups have 111 and 93 patients, respectively. On the observation of the CMV shedding times, some patients gave left-censored observations and some right-censored observations. The others provided some intervals or interval-censored observations, given by the last negative and first positive test dates. That is, we have bivariate interval-censored data on the CMV shedding times in the blood and urine. The percentages of right-censored observations for the CMV shedding times in the blood and urine are about 85% and 43%, respectively, which indicate that the CMV shedding risk in the urine may be higher than that in the blood. For the application of the proposed estimation procedure, let T_{i1} and T_{i2} denote the CMV shedding times in the blood and urine associated with the ith patient, respectively, and define $X_i = 1$ if the ith subject's baseline CD4 count was less than 75 and 0 otherwise. As in the simulation study, we took $G_m(x) = r_m^{-1} \log(1 + r_m x)$ and set $Z_i = 1$ for all patients.

Table 5 presents the estimation results given by the proposed approach for different combinations of $r_1 = 0, 0.5, 1$ and $r_2 = 0, 0.5, 1$, and they include the estimated covariate effects, $\hat{\beta}_{blood}$ and $\hat{\beta}_{urine}$, the estimated standard errors (SE) and the p-values for testing no covariate effect (P). In addition, we have calculated the Akaike Information Criterion (AIC, Akaike, 1973 [29]) and Bayesian Information Criterion (BIC, Schwarz, 1978 [30]) for the selection of the optimal model. One can see from the table that the AIC and BIC values are quite close for all combinations of r_1 and r_2, and the same is true for the estimated effects. By choosing $r_1 = r_2 = 0$, which correspond to the proportional hazards models for both the T_{i1}'s and T_{i2}'s, we have $\hat{\beta}_{blood} = 2.312$ and $\hat{\beta}_{urine} = 1.143$ with the estimated standard errors being 0.396 and 0.142, respectively. They suggest that the patients with lower CD4 at the baseline experienced CMV shedding in both blood and urine significantly early. To provide a graphical view about the difference between the CMV shedding in the blood and urine, Figure 1 presents the estimates of the baseline marginal survival functions given by the proposed method with $r_1 = r_2 = 0$ for the CMV shedding times in the blood and urine, respectively. They suggest that as discussed above, the CMV shedding in the urine occurred much earlier than in the blood.

Table 5. Analysis results for the AIDS clinical trials data.

r_1	r_2	β_{blood}	β_{urine}	SE_{blood}	SE_{urine}	P_{blood}	P_{urine}	AIC	BIC
0	0	2.312	1.143	0.396	0.142	0.000	0.001	727.917	744.507
	0.5	2.333	1.291	0.401	0.208	0.000	0.002	733.972	750.563
	1	2.381	1.437	0.413	0.256	0.000	0.002	738.959	755.550
0.5	0	2.483	1.141	0.424	0.141	0.000	0.001	727.327	743.918
	0.5	2.504	1.288	0.430	0.206	0.000	0.002	733.405	749.996
	1	2.552	1.435	0.442	0.255	0.000	0.002	738.426	755.016
1	0	2.652	1.140	0.451	0.140	0.000	0.001	726.826	743.416
	0.5	2.670	1.285	0.457	0.205	0.000	0.002	732.924	749.514
	1	2.717	1.433	0.469	0.254	0.000	0.002	737.981	754.572

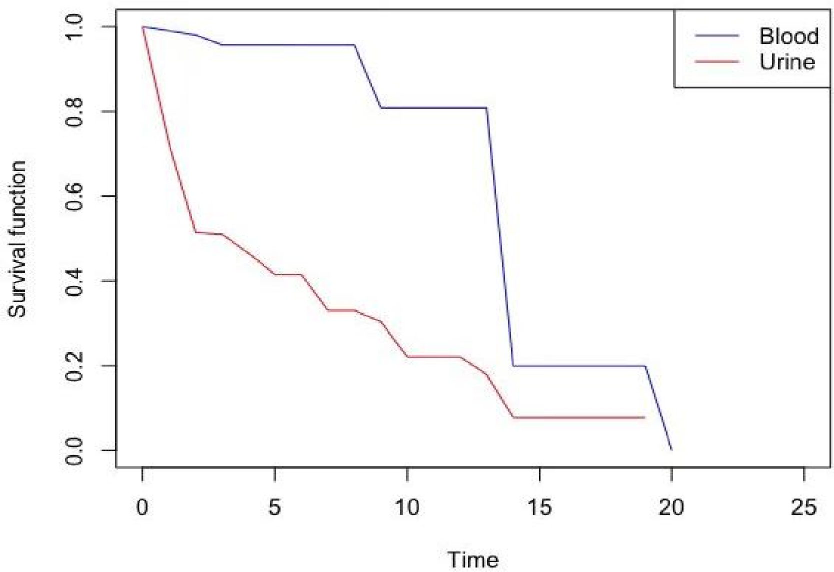

Figure 1. Estimated marginal survival functions for the CMV shedding times in the blood and urine.

In addition, with $r_1 = r_2 = 0$, the proposed method yielded $\hat{\beta}_{u1} = 2.845$ and $\hat{\beta}_{u2} = 0.958$ with the estimated standard errors of 0.112 and 0.116, respectively. They indicate that the observation process was significantly correlated with the CMV shedding times in both blood and urine. That is, we had dependent or informative censoring. To investigate the effects of informative censoring on the covariate effects, we assumed that $\beta_{u1} = \beta_{u2} = 0$, meaning independent interval censoring, and obtained $\hat{\beta}_{blood} = 1.560$ and $\hat{\beta}_{urine} = 1.306$ with the estimated standard errors being 0.514 and 0.326, respectively. They would correspond to the p-values of 0.015 and 0.011 for testing $\beta_{blood} = 0$ and $\beta_{urine} = 0$, respectively. Although these results are similar to those given above, it is apparent that they underestimated the effects of the baseline CD4 on the risks of the CMV shedding times.

7. Discussion and Concluding Remarks

In the preceding sections, the regression analysis of case K multivariate interval-censored failure-time data was discussed under a general class of semiparametric transformation models in the presence of informative censoring. For the problem, an approximate maximum likelihood estimation procedure was proposed and the resulting estimators of the regression parameters were shown to be consistent and asymptotically normal. In the method, the frailty approach was employed to characterize the informative censoring as well as the relationship among the correlated failure times of interest. To implement the proposed approach, a novel EM algorithm was developed and the numerical studies indicated that the proposed method works well in practical situations. In addition, it was applied to a set of real bivariate interval-censored data arising from an AIDS clinical trial.

The proposed approach can be seen as a generalization of the method given by Zeng et al. (2017) [15] to allow for informative interval censoring, which can occur quite often, as discussed above and in the literature. In particular, it has been shown that in the presence of informative censoring, the analysis that ignores it could lead to biased or misleading results and conclusions. The proposed method has the advantages that it does not need or impose an assumption on the distribution of the latent variables and it is quite flexible and can be easily implemented. Moreover, the type of the data considered here includes most types of the failure-time data discussed in the literature as special cases and the model (1) gives many commonly used models.

As discussed above, although model (1) is quite flexible, it may not be straightforward to choose an optimal model for a given set of data, and one commonly used procedure for this is to apply the AIC or BIC. As an alternative, one may prefer to develop a model-checking or data-driven technique. However, this may be difficult and such a method does not seem to exist even for simple types of multivariate interval-censored data. It is worth noting that instead of the proposed approximation maximum likelihood estimation method, one may consider a full maximum likelihood estimation procedure. For this, however, one would need to specify or postulate some distributions for the latent variables b_i's, which may be hard to be verified, and also the implementation would be much more complicated.

Author Contributions: Formal analysis, M.Y.; Conceptualization, M.Y. and M.D.; Software, M.Y.; Supervision, M.D.; Funding acquisition, M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Natural Science Foundation of China Grant (12101522).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the finding of this study are openly available in "ACTG 181" at https://doi.org/10.1007/0-387-37119-2 [2].

Acknowledgments: We want to thank the editor as well as the two reviewers for their many insightful and helpful comments and suggestions that greatly helped the paper. The R code about the implementation of the proposed method is available upon request to the first author.

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

Appendix A. Proof of Theorems 1 and 2

In this Appendix, we will sketch the proof of the two theorems given above.

Proof of Theorem 1. To prove the consistency, we can verify the condition of Theorem 5.7 of Van der Vaart (1998) [31]. $BV[0, \tau_m]$ denotes the functions whose total variation in $[0, \tau_m]$ are bounded by a given constant. Define $\mathbb{M} = \{\theta : \theta \in \mathbb{A} \bigotimes \mathbb{B} \bigotimes \mathbb{C} \bigotimes BV^{\bigotimes M}\}$, where $BV^{\bigotimes M} = BV[0, \tau_1] \bigotimes BV[0, \tau_2] \bigotimes \ldots \bigotimes BV[0, \tau_M]$ and $\mathbb{M}_0$ is a similar space with $\mathbb{M}$ containing θ_0. Moreover, define the metric $\rho(\theta, \theta_0)$ on the parameter space $\mathbb{M}$ as $\rho(\theta, \theta_0) =$

$\sum_{m=1}^{M}\|\beta_{xm}-\beta_{0xm}\|^2+\sum_{m=1}^{M}\|\beta_{um}-\beta_{0um}\|^2+\|\gamma-\gamma_0\|^2+\sum_{m=1}^{M}\sup_{t\in[0,\tau_m]}|\Lambda_m(t)-\Lambda_{0m}(t)|$. Let $L(\theta)$ be the likelihood, so the log-likelihood is

$$l(\theta)=\log\int\left\{\prod_{m=1}^{M}D_m(U_m,b;\beta_m,\Lambda_m)\right\}p(b|\gamma)\mathrm{d}b$$

Then, the class of function $D_m(U_m,b;\beta_m,\Lambda_m)$ is a Donsker class. By condition 4, we know that $l(\theta)$ is bounded away from zero. Therefore, $l(\theta,\mathcal{O})$ belongs to some Donsker class due to the preservation property of the Donsker class under the Lipschitz-continuous transformations. Then, we can conclude that $\sup_{\theta\in\mathbb{M}}|\mathbb{P}_n l(\theta,\mathcal{O})-\mathbb{P}l(\theta_0,\mathcal{O})|$ converges in probability to 0 as $n\to\infty$.

Next, we need to verify another condition of Theorem 5.7 of Van der Vaart (1998) [31], for any $\epsilon>0$,

$$\sup_{\rho(\theta,\theta_0)>\epsilon}\mathbb{P}l(\theta,\mathcal{O})<\mathbb{P}l(\theta_0,\mathcal{O}).$$

Following Gibbs' inequality, we have that $\mathbb{P}l(\theta,\mathcal{O})\le\mathbb{P}l(\theta_0,\mathcal{O})$ for all $\theta\in\mathbb{M}$ with equality holds if and only if $l(\theta,\mathcal{O})=l(\theta_0,\mathcal{O})$ almost surely. Assume that $\sup_{\rho(\theta,\theta_0)>\epsilon}\mathbb{P}l(\theta)=\mathbb{P}l(\theta_0)$. Then, there exists a sequence θ_j such that $\mathbb{P}l(\theta_j)\to\mathbb{P}l(\theta_0)$ and $\rho(\theta_j,\theta_0)>\epsilon$. Because $\mathbb{A}\bigotimes\mathbb{B}\bigotimes\mathbb{C}$ are compact and $BV^{\bigotimes M}$ are uniformly bounded such that θ_{jm} converges to θ_{j0}, and θ_{jm} is the subsequence of θ_j, where θ_{j0} may or not be in $\mathbb{M}$, but in $\mathbb{M}_0$. Clearly, $\mathbb{P}l(\theta)$ is continuous with respect to θ, such that $\mathbb{P}l(\theta_{j0})=\mathbb{P}l(\theta_0)$. By Condition 5 and similar arguments to the proof of Theorem 2.1 of Chang et al. (2007) [32], we can show the identifiability of the model parameters, so that $\theta_{j0}=\theta_0$. However, $\rho(\theta_{jm},\theta_0)>\epsilon$, so θ_{jm} cannot converge to θ_0. This is a contradiction. Therefore, $\sup_{\rho(\theta,\theta_0)>\epsilon}\mathbb{P}l(\theta)<\mathbb{P}l(\theta_0)$. Following Theorem 5.7 of Van der Vaart (1998) [31], we have $\rho(\hat{\theta},\theta_0)=o_p(1)$, which completes the proof of Theorem 1. □

Proof of Theorem 2. Define

$$S_{\beta_{xm}}(\theta)=\frac{\partial l(\theta)}{\partial\beta_{xm}},\ S_{\beta_{um}}(\theta)=\frac{\partial l(\theta)}{\partial\beta_{um}},\ S_{\gamma}(\theta)=\frac{\partial l(\theta)}{\partial\gamma},$$

the score functions with respect to β_{xm}, β_{um} and γ, respectively. For $m=1,\dots,M$, let $h_m(t)$ be a nonnegative and nondecreasing function on $[0,\tau_m]$. Define $\mathcal{H}=\{h=(h_1(t),\dots,h_M(t))\}$, $\Lambda_\epsilon(t)=(\Lambda_{1,\epsilon}(t),\dots,\Lambda_{M,\epsilon}(t))$, and

$$H_{ml}(t;\theta)=\frac{\int B_m(t,U_{ml},U_{m,l+1},b;\beta_m,\Lambda_m)\left\{\prod_{m'=1,m'\neq m}^{M}D_{m'}(U_{m'},b;\beta_{m'},\Lambda_{m'})\right\}p\{b\mid\gamma\}\mathrm{d}b}{\int\left\{\prod_{m'=1}^{M}D_{m'}(U_{m'},b;\beta_{m'},\Lambda_{m'})\right\}p\{b\mid\gamma\}\mathrm{d}b},$$

where $\Lambda_{m,\epsilon}(t)=\Lambda_m(t)+\epsilon h_m(t)$ and

$$\begin{aligned}B_m(t,s_1,s_2,b;\beta_m,\Lambda_m)=&\exp\left\{\beta_{xm}^{\mathrm{T}}X_m+u\beta_{um}+b^{\mathrm{T}}Z_m\right\}\\&\times\Big(Q_m(s_2,b;\beta_m,\Lambda_m)G_m'\left[\int_0^v\exp\left\{\beta_m^{\mathrm{T}}x_m^*+b^{\mathrm{T}}Z_m\right\}\mathrm{d}\Lambda_m(s)\right]I(s_2\geqslant t)\\&-Q_m(s_1,b;\beta_m,\Lambda_m)G_m'\left[\int_0^u\exp\left\{\beta_m^{\mathrm{T}}x_m^*+b^{\mathrm{T}}Z_m\right\}\mathrm{d}\Lambda_m(s)\right]I(s_1\geqslant t)\Big).\end{aligned}$$

It follows that

$$S_{\beta_x}(\theta) = \sum_{m=1}^{M}\sum_{l=0}^{K_m}\int_0^{\tau_m} H_{ml}(t;\theta,\mathcal{A})X_m \mathrm{d}\Lambda_m(t)\,,$$
$$S_{\beta_u}(\theta) = \sum_{m=1}^{M}\sum_{l=0}^{K_m}\int_0^{\tau_m} H_{ml}(t;\theta,\mathcal{A})u \mathrm{d}\Lambda_m(t)\,,$$
$$S_{\gamma}(\theta) = \frac{\int\left\{\prod_{m=1}^{M} D_m(U_m,b;\beta_m,\Lambda_m)\right\}p'_\gamma\{b\mid\gamma\}\mathrm{d}b}{\int\left\{\prod_{m=1}^{M} D_m(U_m,b;\beta_m,\Lambda_m)\right\}p\{b\mid\gamma\}\mathrm{d}b}\,,$$

where $p'_\gamma\{b\mid\gamma\}$ is the first-order derivative of $p\{b\mid\gamma\}$ with respect to γ, $\beta_x = (\beta_{x1}^T,\ldots,\beta_{xM}^T)^T$ and $\beta_u = (\beta_{u1},\ldots,\beta_{uM})^T$.

To obtain the score operator for $\mathcal{A}$, we consider submodels $\mathcal{A}_\epsilon(h)$, where $h = (h_1,\ldots,h_M)^{\mathrm{T}}$ is a vector of functions in $L_2[0,\tau_m]$. Then, we have that $\mathrm{d}\Lambda_{m,\epsilon,h_m} = (1+\epsilon h_m)\mathrm{d}\Lambda_m$, and the score function along the mth submodels for every $\Lambda_m, m = 1,\ldots,M$ has the form

$$S_{\Lambda_m}(\theta)(h) = \sum_{l=0}^{K_m}\Delta_{ml}\int_0^{\tau_m} H_{ml}(t;\theta)h_m(t)\mathrm{d}\Lambda_m(t)\,.$$

The efficient score for ζ at (ζ_0,Λ_0) is $\tilde{l}(\zeta_0,\Lambda_0) = S_\zeta(\zeta_0,\Lambda_0) - \sum_{m=1}^{M} S_{\Lambda_m}(\zeta_0,\Lambda_0)[h_m^*]$, where $S_\zeta(\zeta_0,\Lambda_0) = \left(S_{\beta_{x1}}(\theta_0),\ldots,S_{\beta_{xM}}(\theta_0),S_{\beta_u}(\theta_0),S_\gamma(\theta_0)\right)^T$, h_m^* is a $(pM+M+q)$-vector function satisfying

$$\mathbb{P}\left[\left(S_\zeta(\zeta_0,\Lambda_0) - \sum_{m=1}^{M} S_{\Lambda_m}(\zeta_0,\Lambda_0)[h_m^*]\right)^T\left(\sum_{m=1}^{M} S_{\Lambda_m}(\zeta_0,\Lambda_0)[h_m]\right)\right] = 0,$$

for each h_m in $\mathcal{H}$.

By following similar calculations in Section 3 of Chang et al. (2007) [32], we can establish the existence of h_m^* in the above equation. The efficient Fisher information matrix I_0 for ζ at (ζ_0,Λ_0) is defined as $\mathbb{P}\left(\tilde{l}(\zeta_0,\Lambda_0)\tilde{l}(\zeta_0,\Lambda_0)^T\right)$. In the following, we will show that I_0 is positive definite. If the I_0 is singular, then there exists a nonzero vector $\nu \in R^{(pM+M+q)}$ such that $\nu^T I_0 \nu = 0$. It follows that, with probability one, the score function along the submodel $\{\zeta_0+\epsilon\nu, \Lambda_{10}+\epsilon\nu^T h_1^*,\ldots,\Lambda_{M0}+\epsilon\nu^T h_M^*\}$ is zero. Therefore,

$$\nu^T\left(\frac{\partial}{\partial\zeta} + \sum_{m=1}^{M} h_m^*\frac{\partial}{\partial y_m}\right)\Bigg|_{(\zeta,y_1,\ldots,y_M)=(\zeta_0,\Lambda_{10}(c_1),\ldots,\Lambda_{M0}(c_M))}$$
$$\cdot\log\int_b\prod_{m=1}^{M}\{D_m(U_m,b,\beta_m,\Lambda_m)\}p(b\mid\gamma)\mathrm{d}b = 0.$$

Using Condition 5, we know that $\nu = 0$, and this is a contradiction. Therefore, we can conclude that $\nu^T I_0 \nu = 0$ implies $\nu = 0$. That is, the efficient Fisher information matrix is positive.

Define

$$S_{\zeta,m}(\theta)[h_m] = \frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0} S_\zeta(\theta;\Lambda_m = \Lambda_{m\epsilon})\,,$$

and

$$S_{m,j}(\theta)\left[\tilde{h}_m,h_j\right] = \frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0} S_{\Lambda_m}(\theta;\Lambda_j = \Lambda_{j\epsilon})\left[\tilde{h}_k\right]$$

for $m = 1, \ldots, M$ and $j = 1, \ldots, M$, where $\partial / \partial\epsilon|_{\epsilon=0} \Lambda_{j\epsilon} = h_j$. By Taylor expansion, we can obtain

$$\mathbb{P}\tilde{l}(\zeta_0, \Lambda) = \mathbb{P}\tilde{l}(\zeta_0, \Lambda_0) + \mathbb{P}\left\{ \sum_{m=1}^{M} S_{\zeta m}(\theta)[\Lambda_m - \Lambda_{m0}] - \sum_{m=1}^{M}\sum_{j=1}^{M} S_{m,j}(\theta)[h_m^*, \Lambda_m - \Lambda_{m0}] \right\} + O_p\left(\sum_{m=1}^{M} \|\Lambda_m - \Lambda_{m0}\|^2 \right).$$

Note that $\mathbb{P}\tilde{l}(\zeta_0, \Lambda_0) = 0$, $\mathbb{P}\left(S_\zeta(\theta) S_{\Lambda_m}(\theta)[h_m]\right) = -\mathbb{P}\left(S_{\zeta,m}(\theta)[h_m]\right)$, $\mathbb{P}\left(S_{\Lambda_m}(\theta)\left[\tilde{h}_m\right] S_{\Lambda_j}(\theta)\left[h_j\right]\right) = -\mathbb{P}\left(S_{m,j}(\zeta)\left[\tilde{h}_m, h_j\right]\right)$. By the consistency and the proof of Theorem of Zeng et al. (2017), we can conclude that $\mathbb{P}\tilde{l}(\zeta_0, \hat{\Lambda}_n) = O_p\left(n^{-2/3}\right)$, which implies $\sqrt{n}\mathbb{P}\tilde{l}(\zeta_0, \hat{\Lambda}_n) = o_p(1)$.

We know from Example 19.11 of Van der Vaart (1998) [31] that the class of uniformly bounded functions with bounded variations is a Donsker class. By using Theorem 2.10.6 of Van der Vaart and Wellner (1996) [33], we can verify that $\tilde{l}(\zeta, \Lambda)$ is a uniformly bounded Donsker class. In addition, we have proved that $\hat{\theta}_n$ is consistent. Therefore, $\sqrt{n}(\mathbb{P}_n - \mathbb{P})(\tilde{l}(\hat{\zeta}_n, \hat{\Lambda}_n) - \tilde{l}(\zeta_0, \Lambda_0)) = o_p(1)$. Due to the fact that $\mathbb{P}_n\tilde{l}(\hat{\theta}_n) = \mathbb{P}\tilde{l}(\theta_0) = 0$ and $\mathbb{P}\tilde{l}(\zeta_0, \hat{\Lambda}_n) = o_p(1)$, we can have

$$-\sqrt{n}\mathbb{P}\left(\tilde{l}(\hat{\theta}_n) - \tilde{l}(\zeta_0, \hat{\Lambda}_n)\right) = \sqrt{n}\mathbb{P}_n\tilde{l}(\theta_0) + o_p(1).$$

By the mean value theorem, we have

$$-\sqrt{n}\mathbb{P}\frac{\partial}{\partial\zeta}\tilde{l}(\zeta', \hat{\Lambda}_n)(\hat{\zeta}_n - \zeta_0) = \sqrt{n}\mathbb{P}_n\tilde{l}(\theta_0) + o_p(1),$$

where ζ' is a point between $\hat{\zeta}_n$ and ζ_0. Because $\hat{\theta}_n$ is consistency and $\mathbb{P}\left(-\frac{\partial}{\partial\zeta}\tilde{l}(\theta_0)\right) = \mathbb{P}\left(\tilde{l}(\theta_0)\tilde{l}(\theta_0)^T\right) = I_0$, we can conclude that

$$\sqrt{n}(\hat{\zeta}_n - \zeta_0) = I_0^{-1}\sqrt{n}\mathbb{P}_n\tilde{l}(\theta_0) + o_p(1) \xrightarrow{d} N\left(0, I_0^{-1}\right).$$

This completes the proof of Theorem 2. □

References

1. Finkelstein, D.M. A proportional hazards model for interval-censored failure time data. *Biometrics* **1986**, *42*, 845–854. [CrossRef] [PubMed]
2. Sun, J. *Statistical Analysis of Interval-Censored Failure Time Data*; Springer: New York, NY, USA, 2006.
3. Kalbfleisch, J.D.; Prentice, R.L. *The Statistival Analysis of Failure Time Data*, 2nd ed.; Wiley: New York, NY, USA, 2002.
4. Wang, L.; Sun, J.; Tong, X. Regression analysis of case II interval censored failure time data with the additive hazards model. *Stat. Sin.* **2010**, *20*, 1709–1723. [PubMed]
5. Zhang, Z.; Sun, J.; Sun, L. Statistical analysis of current status data with informative observation times. *Stat. Med.* **2005**, *24*, 1399–1407. [CrossRef] [PubMed]
6. Zhang, Z.; Sun, L.; Sun, J.; Finkelstein, D. Regression analysis of failure time data with informative interval censoring. *Stat. Med.* **2007**, *26*, 2533–2546. [CrossRef]
7. Sun, J. A nonparametric test for current status data with unequal censoring. *J. R. Stat. Soc. B* **1999**, *61*, 243–250. [CrossRef]
8. Chen, M.; Tong, X.; Sun, J. The proportional odds model for multivariate interval-censored failure time data. *Stat. Med.* **2007**, *26*, 5147–5161. [CrossRef]
9. Chen, M.; Tong, X.; Sun, J. A frailty model approach for regression analysis of multivariate current status data. *Stat. Med.* **2009**, *28*, 3424–3436. [CrossRef]
10. Chen, M.; Tong, X.; Zhu, L. A linear transformation model for multivariate interval-censored failure time data. *Can. J. Stat.* **2013**, *41*, 275–290. [CrossRef]
11. Goggins, W.B.; Finkelstein, D.M. A proportional hazards model for multivariate interval-censored failure time data. *Biometrics* **2000**, *56*, 940–943. [CrossRef] [PubMed]

12. Shen, P. Additive transformation models for multivariate interval-censored data. *Commun. Stat. Theory Methods* **2015**, *44*, 1065–1079. [CrossRef]
13. Tong, X.; Chen, M.H.; Sun, J. Regression analysis of multivariate interval censored failure time data with application to tumorigenicity experiments. *Biom. J.* **2008**, *33*, 364–374. [CrossRef]
14. Wang, L.; Sun, J.; Tong, X. Efficient estimation for the proportional hazards model with bivariate current status data. *Lifetime Data Anal.* **2008**, *14*, 134–153. [CrossRef]
15. Zeng, D.; Gao, F.; Lin, D. Maximum likelihood estimation for semiparametric regression models with multivatiate interval-censored data. *Biometrika* **2017**, *104*, 505–525. [CrossRef]
16. Zhang, B.; Tong, X.; Sun, J. Efficient Estimation for the proportional odds model with bivariate current status data. *Far East J. Theor. Stat.* **2009**, *27*, 113–132.
17. Zhou, Q.; Hu, T.; Sun, J. A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *J. Am. Stat. Assoc.* **2017**, *112*, 664–672. [CrossRef]
18. Sun, T.; Ding, Y. Copula-based semiparametric regression method for bivariate data under general interval censoring. *Biostatistics* **2021**, *2*, 315–330. [CrossRef]
19. Wei, L.J.; Lin, D.Y.; Weissfeld, L. Regression analysis of multivariate incomplere failure time data by modeling marginal distributions. *J. Am. Stat. Assoc.* **1989**, *84*, 1065–1073. [CrossRef]
20. Wang, S.; Wang, C.; Wang, P.; Sun, J. Semiparametric analysis of the additive hazards model with informatively interval-censored failure time data. *Comput. Stat. Data Anal.* **2018**, *125*, 1–9. [CrossRef]
21. Yu, M.; Feng, Y.; Duan, R.; Sun, J. Regression analysis of multivariate interval-censored failure time data with informative censoring. *Stat. Methods Med. Res.* **2022**, *31*, 391–403. [CrossRef] [PubMed]
22. Ma, L.; Hu, T.; Sun, J. Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika* **2015**, *102*, 731–738. [CrossRef]
23. Wang, P.; Zhao, H.; Sun, J. Regression analysis of case k interval-censored failure time data in the presence of informative censoring. *Biometrics* **2016**, *72*, 1103–1112. [CrossRef]
24. Klein, J.P.; Moeschberger, M.L. *Survival Analysis: Techniques for Censored and Truncated Data*; Springer: New York, NY, USA, 2003.
25. Cook, R.J.; Lawless, J. *The Statistical Analysis of Recurrent Events*; Springer: New York, NY, USA, 2007.
26. Chen, K.; Jin, Z.; Ying, Z. Semiparametric Analysis of Transformation Models with Censored Data. *Biometrika* **2002**, *89*, 659–668. [CrossRef]
27. Huang, C.Y.; Wang, M.C. Joint modeling and estimation for recurrent event processes and failure time data. *J. Am. Stat. Assoc.* **2004**, *99*, 1153–1165. [CrossRef]
28. Murphy, S.A.; Van Der Vaart, A.W. On profile likelihood. *J. Am. Stat. Assoc.* **2000**, *95*, 449–465. [CrossRef]
29. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*; Petrov, B., Csaki, F., Eds.; Academiae Kiado: Budapest, Hungary, 1973; pp. 267–281.
30. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
31. Van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: New York, NY, USA, 1998.
32. Chang, I.S., Wen, C.C., Wu, Y.J. A profile likelihood theory for the correlated gamma-frailty model with current status family data. *Stat. Sin.* **2007**, *17*, 1023–1046.
33. Van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.